Data Sharing and Incentives*

Annie Liang† Erik Madsen‡

November 11, 2019

Abstract

Many organizations, such as banks and insurers, determine what services to offer based on a perceived quality of the recipient, e.g. their creditworthiness. With new access to detailed data on individual consumers, organizations are increasingly estimating quality not only from a given consumer's interactions with the organization, but also from interactions with comparable individuals. What are the consequences for consumer incentives to exert effort in their interactions with the firm, e.g. to maintain a good credit rating? To answer this question, we study a multiple-agent career concerns model in which agents choose whether to interact with a principal, who provides a service and aggregates data across all participating agents. Individuals' interactions create an informational externality on others, shaping participation rates and effort provision in equilibrium. We show that whether data sharing is welfare-improving depends crucially on how the actions of individuals affect inferences about related consumers, specifically on whether information across consumers is "complementary" or "substitutable."

*We are grateful to Alessandro Bonatti, Yash Deshpande, Ben Golub, and Steven Matthews for very helpful conversations, and to Changhwa Lee for valuable research assistance on this project.

†Department of Economics, University of Pennsylvania

‡Department of Economics, New York University

1 Introduction

Organizations have long used data to personalize their offerings to consumers. For instance, banks determine creditworthiness based on past debt and payment histories, and car insurance companies predict accident risk based on driver covariates such as age and vehicle model. The scale and diversity of data now available to organizations, however, are qualitatively new: The subprime lender CompuCredit was recently investigated by the Federal Trade Commission for failing to disclose that it reduced credit lines based on usage of the card at various "red flag" establishments, including marriage counselors and nightclubs.[1] Health insurance companies have begun experimenting with use of personal details acquired from data brokers—such as past purchases of plus-size clothing—to help predict health risks.[2] Car insurance companies offer drivers the option of installing tracking devices that continuously monitor driving behavior. And perhaps most strikingly, China's "social credit" score determines whether an individual is a good citizen based on detailed attributes ranging from the size of their social network to how often they play video games.[3]

This abundance of data can improve predictions. But credit scores and monitoring devices not only predict a consumer's type—they also have the effect of shaping incentives for good behavior, e.g. building credit and driving more attentively. These incentives may be reshaped when organizations make judgments based on data provided by other individuals in the system, and/or sources of auxiliary information.[4] It is thus important to understand the implications of "big data" predictions both for consumer willingness to participate in the system, and also for the way they behave within it.

To study this question, we build a simple model involving a population of agents with unknown *characteristics* or *types* (e.g. creditworthiness), which a principal (e.g. a bank) would like to predict. Each agent chooses whether to opt-in to interaction with the principal (e.g. sign up for a credit card). The principal observes a signal (e.g. the agent's past purchasing and repayment behavior) from each agent who opts in, which is informative about the agent's underlying type. The agent can manipulate this signal via costly effort (e.g. by budgeting or deferring expenditures). Crucially, the principal bases their prediction of the agent's type both on the agent's own data, as well as on the data generated by participation

[1] https://www.bloomberg.com/news/articles/2008-06-18/your-lifestyle-may-hurt-your-credit.

[2] https://www.pbs.org/newshour/health/why-health-insurers-track-when-you-buy-plus-size-clothes-or-binge-watch-tv.

[3] https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/

[4] For example, Jin and Vasserman (2019) finds that claim rates for drivers who opt-in to installation of a monitoring device reduce by 30%.

of other agents. Thus, there is a social externality to data sharing: each agent's data affects the principal's perception of other agents' types. Formally, our framework is a multiple-agent version of the career concerns model (Holmstrom, 1982), where signals are correlated across agents.

We highlight and contrast two ways in which data from one agent can be used to learn about another agent's type. First, social data can be used for *interpreting* or *unconfounding* information about any given agent. For example, if we interpret signals as repayment behavior, an unknown shock to the real interest rate affects repayment rates across all agents. Understanding this shock better allows the principal to extract more information from any given observation. Formally, we can model this externality by supposing that consumers have idiosyncratic types, but their signals are distorted by a common confound.

Alternatively, social data can be used to help the principal *refine* their prediction for a particular group. For example, more data about repayment rates for individuals who frequent nightclubs will help the principal to better estimate the average creditworthiness of this group. We can model this externality by supposing that consumers share a common and unknown type. In this second model, consumer data are *substitutable*—the presence of auxiliary data reduces the informativeness of each agent's signal about his own type—while in the first model, consumer data are *complementary*.

Our main results show that there are substantial differences in the equilibrium implications of data sharing across the two models. In the common confound model, data sharing encourages effort, and equilibrium effort is *higher* than it would have been in a single-agent model (without aggregation of data across consumers). In the common type model, data sharing depresses effort. This difference is exaggerated as the population size grows: Equilibrium effort in the common type model vanishes, while equilibrium effort in the common confound model is weakly increasing (but bounded).

We also compare the fraction of agents who opt-in across the two models. If the population is small, then equilibrium involves full entry in both models. But if the population is large, then there is full entry in the common type model and only partial entry in the common confound model. Intuitively, in the common type model, data from other agents depresses the sensitivity of the principal's beliefs to the realization of the agent's signal, decreasing the incentive for costly effort. Opt-in decisions encourage further opt-in, supporting a unique full-entry equilibrium. In contrast, in the common confound model, data from other agents increases the sensitivity of the principal's beliefs to the realization of the agent's signal, increasing the incentive for effort. Opt-in decisions thus discourage opt-in by other agents, so that only partial entry is supported in equilibrium (for large populations).

The key technical lemma we prove towards this result shows that in a population of *exogenous* size, the equilibrium effort level is monotonically decreasing in the population size in the common type model (with limit equal to zero) but increasing in the common confound model (with a finite limit). To show this, we need to consider the marginal value of exerting (unexpected) effort, and how this depends on the total number of auxiliary signals available to the principal. The proof of the lemma provides a new bound on the sensitivity of the principal's posterior's expectation to the realization of a *single* signal, and a characterization of the rate at which this sensitivity converges to its asymptotic value as the number of auxiliary signals grows large.

Finally, we use these equilibrium results to study welfare implications. We show that in both models and for all population sizes, equilibrium actions are inefficient relative to the first-best (this extends a well-known result established in Holmstrom (1982) for Gaussian signals to more general information structures). We additionally compare equilibrium against a "no data sharing" benchmark corresponding to equilibrium when the principal is only permitted to use an agent's *own* past data to predict that agent's type. When players share a common type, aggregation of data across agents *always* leads to a reduction in social welfare. In contrast, in the common confound model, the welfare implications of data sharing depend on the population size: In small populations, data sharing leads to an improvement in social welfare, while in sufficiently large populations, it leads to a reduction. These results suggest that data sharing does not have an unambiguous welfare implication—the way in which social data is correlated across agents is a crucial determinant for the direction of its welfare effect.

Our paper contributes to an emerging literature regarding the welfare consequences of data markets and algorithmic scoring. This literature has tackled several important social questions, such as whether predictive algorithms discriminate (Chouldechova, 2017; Kleinberg, Mullainathan and Raghavan, 2017; Kearns et al., 2018), how to protect consumers from loss of privacy (Acquisiti, Brandimarte and Loewenstein, 2015; Dwork and Roth, 2014; Fainmesser, Galeotti and Momot, 2019), how to price data (Bergemann, Bonatti and Smolin, 2018; Agarwal, Dahleh and Sarkar, 2019), and whether price discrimination based on big data harms consumers (Anderson et al., 2018; Jullien, Lefouili and Riordan, 2018; Gomes and Pavan, 2018; Ichihashi, 2019; Bonatti and Cisternas, 2018). There is additionally a growing literature about strategic interactions with machine learning algorithms: see Eliaz and Spiegler (2018) on the incentives to truthfully report characteristics to a machine learning algorithm, and Olea et al. (2018) on how economic markets select certain models for making predictions over others.

In particular, Acemoglu et al. (2019) and Bergemann, Bonatti and Gan (2019) also consider externalities created by social data. These papers study data sharing in environments where consumers may sell their data. In Bergemann, Bonatti and Gan (2019), other agents' information allows a firm to set more tailored (and possibly personalized) prices, which can decrease consumer surplus. In Acemoglu et al. (2019), agents value privacy, and thus information collected about one agent imposes a direct negative externality on other agents when types are correlated. The externality of interest in the present paper is how information provided by other agents reshapes incentives to exert costly effort. As we show, this externality is not always negative (i.e. when agents share a common type, their equilibrium payoffs are *increasing* in other agents' participation).

Our model formally builds on the career concerns model of Holmstrom (1982). The literature following Holmstrom (1982) has largely focused on signal extraction about a single agent's type in dynamic settings,[5] while we are interested in the externalities of social data in a multiple-agent setting. Our paper is most closely related to Dewatripont, Jewitt and Tirole (1999), which studies how auxiliary data impacts agents incentives for effort. Dewatripont, Jewitt and Tirole (1999) consider the externality of a single auxiliary signal, while we endogenize the auxiliary data as information from other players, who strategically decide whether or not to provide data. Thus, the number of auxiliary signals is determined in equilibrium, and may also be uncertain; this requires comparison of equilibrium actions across various information structures.

Finally, our paper contributes to work on strategic manipulation of information. Recent papers in this category include Frankel and Kartik (2019), which characterizes the degree to which a principal with commitment power should link his decision to a manipulated signal about the agent's type; Hu, Immorlica and Vaughan (2019), which shows that heterogeneous manipulation costs across different social groups can lead to inequities in outcome; and Georgiadis and Powell (2019), which studies optimal information acquisition for a designer setting a wage contract. Our paper contributes to this literature by exploring the role of correlations across data for an individual's incentives to manipulate.

[5] There is a limited number of papers, e.g. Auriol, Friebel and Pechlivanos (2002), which study career concerns in a multiple agent setting. These papers typically look at effort externalities instead of informational externalities.

2 Model

2.1 Setup

There is a single *principal* or *organization* and $N < \infty$ *agents*. Every agent i has an unknown and unobservable type $\theta_i \in \mathbb{R}$, which is commonly believed to be distributed according to F_θ, with mean $\mathbb{E}[\theta_i] = \mu > 0$ and variance $\mathbb{E}[(\theta_i - \mu)^2] = \sigma_\theta^2 > 0$. The timeline of the game is as follows:

$t = 0$**:** Each agent i first chooses whether to *opt-in* (I) or *opt-out* (O) of an interaction with the principal, where this decision is observed by the principal, but not by other agents. Opting out yields a payoff equal to the prior expectation of the agent's type, μ, and ends the game.[6]

An agent who opts in interacts with the principal in the following way. Following the classic career concerns model (Holmstrom, 1982; Dewatripont, Jewitt and Tirole, 1999), there are two subsequent periods:

$t = 1$**:** First, the agent privately chooses an effort level $a_i \in \mathbb{R}_+$. The agent's unobservable effort choice and his unobservable type θ_i jointly generate an output

$$S_i = \theta_i + a_i + \eta_i + \varepsilon_i,$$

where $\eta_i \sim F_\eta$ and $\varepsilon_i \sim F_\varepsilon$ are noise shocks satisfying $\mathbb{E}[\eta_i] = \mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\eta_i^2] = \sigma_\eta^2 \in \mathbb{R}_{++}$, $\mathbb{E}[\varepsilon_i]^2 = \sigma_\varepsilon^2 \in \mathbb{R}_{++}$. It will simplify the statement of the subsequent results to assume that all variances are the same, $\sigma_\theta^2 = \sigma_\eta^2 = \sigma_\varepsilon^2$, although all results can be extended if this doesn't hold.

We assume that $\theta_i, \eta_i, \varepsilon_i$ are jointly independent for each agent i. Further, each ε_i is independent of all other agents' types and noise shocks, and the vector of types (θ_i) is independent of the vector of shocks (η_i). However, output is correlated across agents through either the types or the shocks. We capture these two possibilities in the following models:

[6] It is not important that the outside option is precisely μ, as the subsequent results can be modified in a straightforward way for any constant outside option. Our choice of μ is motivated by the applications we have in mind: for example, if we interpret the principal as a car insurance company, and opt-in as installation of a monitoring device, then the outside option may be the principal's expectation of the agent's type in the absence of additional information via the monitoring device.

Common Confound. Agents' types θ_i are mutually independent, while $\eta_i = \eta$ for all agents i. In this model, we refer to η as the *common shock*, with each ε_i the *idiosyncratic shock* for agent i.[7]

Common Type. The shocks (η_i) are mutually independent, while $\theta_i = \theta$ for all agents i. In this model the aggregate shock $\eta_i + \varepsilon_i$ is purely idiosyncratic, but we retain the decomposition to ease comparison across models.[8] In particular, the marginal distribution of each agent's output is the same in the two models.

Each agent i's payoff in this period is

$$t_1 - C(a_i)$$

where $t_1 > 0$ is a constant reward, and $C(a_i)$ is the cost to choosing effort a_i.[9] We suppose that the cost function satisfies $C'(a) > 0$, $C(0) = 0$, and is twice-differentiable.

$t = 2$**:** In the second period, the agent receives

$$t_2 = \mathbb{E}(\theta_i \mid I, \mathbf{S}_I; \mathbf{a}_I^*)$$

which is the principal's expectation of the agent's unknown type θ_i, conditional on the set of agents $I \subseteq \{1, \ldots, N\}$ who chose to opt-in, the vector $\mathbf{S}_I = \{S_i\}_{i \in I}$ of outputs created by agents who opted-in, and the effort choices $\mathbf{a}_I^* = \{a_i^*\}_{i \in I}$ that the principal believes those agents to have taken. Each agent i's total payoff is the sum of his expected payoffs across the two periods, i.e.

$$t_1 + \mathbb{E}(t_2) - C(a_i).$$

We look for symmetric equilibria with non-random effort, where each agent chooses an action in $\Delta(\{I, O\}) \times \mathbb{R}_+$.

[7]The case of $\sigma_\theta^2 = 0$ and $\sigma_\eta^2 > 0$ in Bergemann, Bonatti and Gan (2019) returns the same model of correlation across the outputs.

[8]The case of $\sigma_{\theta_i}^2 = 0$ and $\sigma_\eta^2 = 0$ in Bergemann, Bonatti and Gan (2019) returns the same model of correlation across agent outputs. This model can also be interpreted as a special case of the signal structure in Acemoglu et al. (2019), where types are perfectly correlated.

[9]For example, some car insurance companies offer drivers a discount on their insurance premium in return for installing a tracking device (Jin and Vasserman, 2019). In China, opting-in to certain social credit score systems gives users a wide range of benefits including qualification for personal credit loans, preferential treatment at hospitals, and fast-tracked visa applications (Kostka, 2019).

2.2 Distributional Assumptions

We now state several regularity conditions on the distributions $F_\theta, F_\eta, F_\varepsilon$, which we maintain throughout the paper. Assumptions 1 through 3 are purely technical, and ensure that all distributions have full support and are smooth enough for appropriate derivatives and conditional expectations to exist. Assumptions 4 and 5 are substantive, and ensure monotonicity of inferences about latent variables in output and the sufficiency of the first-order approach.

Assumption 1. *The distribution functions F_θ, F_η, and F_ε admit strictly positive density functions f_θ, f_ε, f_η on $\mathbb{R}$.*

Assumption 2. *For every population size N, model $M \in \{T, C\}$, and agent $i \in \{1, ..., N\}$, $\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}]$ is twice differentiable wrt s_i for every $(\mathbf{s}, \mathbf{a})$, and $\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i]$ is twice differentiable wrt s_i for every (s_i, t, a_i).*

Assumption 3. *There exists a $\overline{\Delta} > 0$ and a function $J : \mathbb{R} \to \mathbb{R}_+$ such that*

$$\left(\frac{1}{\Delta}\frac{f_\varepsilon(z-\Delta) - f_\varepsilon(z)}{f_\varepsilon(z)}\right)^2 \leq J(z)$$

for all $z \in \mathbb{R}$ and $\Delta \in (0, \overline{\Delta})$ and

$$\int dF_\varepsilon(z)\, J(z) < \infty.$$

Assumption 3 is a slight strengthening of the assumption that the Fisher information of S_i about $\theta_i + \eta_i$ is finite. It ensures, roughly, that finite-difference approximations to the Fisher information are also finite and uniformly bounded as the approximation becomes more precise.[10]

Assumption 4. *The density functions f_θ, f_η, and f_ε are strictly log-concave.*[11]

[10] A sufficient condition for Assumption 3 is that the density does not vanish at the tails "much faster" than its derivative: specifically, there should exist a $K > 0$ and $\overline{\Delta} > 0$ such that:

$$\max_{\varepsilon \in \mathbb{R}, \Delta \in [0, \overline{\Delta}]} \left|\frac{f'_\varepsilon(\varepsilon - \Delta)}{f_\varepsilon(\varepsilon)}\right| \leq K.$$

This sufficient condition is satisfied, for example, by the t-distribution and the logistic distribution. It is *not* satisfied by the normal distribution, although we show in Appendix D.1 using other methods that the normal distribution does satisfy Assumption 3. We have not been able to find any commonly-used distributions that fail Assumption 3.

[11] A function $g > 0$ is *strictly log-concave* if $\log g$ is strictly concave.

In general, given three random variables X, Y, Z such that $X = Y + Z$ and Y and Z are independent, strict log-concavity of the density function of Z is both necessary and sufficient for the distribution of X to satisfy a strict monotone likelihood-ratio property in Y (Saumard and Wellner, 2014):

$$\frac{f_{X|Y}(x' \mid y')}{f_{X|Y}(x \mid y')} > \frac{f_{X|Y}(x' \mid y')}{f_{X|Y}(x \mid y)} \quad \text{if and only if} \quad x' > x, y' > y.$$

Following Milgrom (1981), this monotone likelihood-ratio property is the canonical sufficient condition ensuring monotonicity of the conditional expectation of Y in the observed value of X. Assumption 4 guarantees that the appropriate monotone likelihood-ratio properties are satisfied in our model; see Appendix A.1 for details.

Finally, we assume the cost function is "sufficiently convex" so that effort choices satisfying an appropriate first-order condition are globally optimal. The assumption is jointly imposed on the cost function and the distribution of the output, since the required amount of convexity depends on how sensitive the posterior expectation is to the realization of individual outputs.

Assumption 5. *There exists a $K \in \mathbb{R}$ such that $C''(x) > K$ for every $x \in \mathbb{R}_+$, and for each model $M \in \{T, M\}$, population size N, and agent $i \in \{1, ..., N\}$:*

- $\frac{\partial^2}{\partial s_i^2}\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] \leq K$ *for every* $(\mathbf{s}, \mathbf{a})$,
- $\frac{\partial^2}{\partial s_i^2}\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i] \leq K$ *for every* (s_i, t, a_i).

One important set of models satisfying these regularity conditions is the class of Gaussian models, in which each θ_i, η_i, and ε_i is normally distributed, with any strictly concave cost function. We verify in Appendix D.1 that Assumptions 1 through 5 are all met in this case, and provide additional results for this class in Section 5.

3 Preliminary Results: Exogenous Opt-In

We begin our analysis by studying a related model, in which the number of agents who opt-in, N, is exogenously given and commonly known. In this section we describe the equilibrium of this model. An accompanying formal derivation is provided in Appendix A.2.

3.1 Marginal Value of Effort

In equilibrium, agents choose effort such that the marginal impact of effort on the second-period reward t_2, which we will refer to as the *marginal value of effort*, equals its marginal cost. In this subsection, we characterize the marginal value of effort and describe some of its properties.

Let Ω denote the state space of realizations of types θ_i, noise terms η_i and ε_i, and output S_i for each agent $i = 1, ..., N$. The principal and all agent forms prior beliefs over Ω based on conjectures about the action taken by each agent. In any equilibrium, all players share a common conjecture and thus a common prior over Ω. But if some agent i deviates to a non-equilibrium action, then he no longer shares the principal's prior over Ω. Specifically, following such a deviation, the two players agree on the the marginal distribution of the set of types, and on the joint distribution of other agents' outputs conditional on their types, but they disagree about the joint distribution of agent i's output S_i, his type θ_i, and the confound η_i.

This fact means that an agent's expected period-2 reward following a deviation is an iterated expectation with respect to two different probability measures over Ω. Fix an equilibrium action profile $(a_1^*, ..., a_N^*)$, and let $\mathbb{E}^*$ denote expectations over Ω when all agents take their equilibrium action profile. For any profile of realized outputs, agent i's period-2 reward is

$$t_2 = \mathbb{E}^*[\theta_i \mid S_1, \ldots, S_N].$$

The agent can affect the distribution of t_2 via his choice of action. Specifically, if agent i takes action $a_i = a_i^* + \Delta$, then his expected period-2 reward $\mu_N(\Delta)$ is

$$\mu_N(\Delta) \equiv \mathbb{E}^\Delta[t_2] = \mathbb{E}^\Delta[\mathbb{E}^*[\theta_1 \mid S_1, \ldots, S_N]].$$

Note in particular that because $\mathbb{E}^0 = \mathbb{E}^*$, the agent's expected reward when choosing his equilibrium action a_i^* is

$$\mu_N(0) = \mathbb{E}^*[\mathbb{E}^*[\theta_1 \mid S_1, \ldots, S_N]] = \mu \tag{1}$$

by the law of iterated expectations, reflecting the usual martingale property of posterior expectations. That is, in the absence of distortion away from the equilibrium effort a_i^*, the agent expects that the principal expects his type to be μ.

However, when $\Delta \neq 0$, posterior expectations under the principal's beliefs are *not* a martingale from agent 1's perspective. In fact, as we show in Appendix A.2, $\mu_N(\Delta)$ is strictly increasing in Δ. That is, increasing effort beyond the expected effort level always leads to a higher expected value of the principal's expectation. (This monotonicity is a consequence of the log-concavity property imposed in Assumption 4.)

We define the marginal value of effort $MV(N)$ to be

$$MV(N) \equiv \mu_N'(0).$$

Our notation reflects the fact that $\mu_N(\Delta)$, thus also $MV(N)$, is independent of the levels of the equilibrium actions $a_1^*, ..., a_N^*$, due to the additive dependence of outputs on actions. (We establish this property formally in Appendix A.2.)

The quantity $MV(N)$ captures the *local* value of distorting effort near the equilibrium value. Assumption 5 ensures that no deviation from equilibrium effort can yield higher returns, net of costs, than a local deviation, justifying the focus on local deviations from equilibrium effort. Throughout, we use $MV_T(N)$ to denote the marginal value function in the common type model and $MV_C(N)$ to denote the marginal value function in the common confound model, dropping the subscript when a statement holds in both models. Note finally that $MV_T(1) = MV_C(1)$; that is, with a single agent, the incentives for effort in the two models are identical.

3.2 Equilibrium Effort

The symmetry of our model ensures that all agents share the same marginal value and marginal cost of effort. There is therefore a unique action $a^*(N)$ satisfying each agent's equilibrium first-order condition

$$MV(N) = C'(a^*(N)) \tag{2}$$

equating the marginal value of effort $MV(N)$ with its equilibrium marginal cost $C'(a^*(N))$. This condition is both necessary and sufficient to ensure that each agent's optimal action — given that the principal expects all agents to exert effort $a^*(N)$ — is indeed $a^*(N)$. The unique equilibrium of the exogenous-entry model then consists of choice of the effort level $a^*(N)$ by every agent. When we wish to describe the equilibrium action level for a given model, we will write $a_T^*(N)$ for the common type model, and $a_C^*(N)$ for the common confound model.

Note that the marginal cost of effort (at a fixed a) is independent of the population size N, while the marginal value of effort in general does depend on the population size. Thus the comparative statics of equilibrium effort in N are determined by the behavior of the marginal value of effort $MV(N)$ as N changes. The following key lemma characterizes this behavior:

Lemma 1. *(a) $MV_T(N)$ is nonincreasing in N and* $\lim_{N\to\infty} MV_T(N) = 0$.

(b) $MV_C(N)$ is nondecreasing in N and $\lim_{N\to\infty} MV_C(N) < \infty$.

Intuitively, the number of datapoints N influences how sensitive the principal's expectation of θ_i is to the realization of S_i. All else equal, the stronger the dependence on the realization of i's output, the stronger the incentive to manipulate its realization. In the common confound model, other agents' data (which inform about the common noise term η) complements agent i's output, improving its informativeness. Thus, the larger N is, the more weight the principal puts on i's output. This incentivizes effort. However, even in the limit of large N, agent i's output is still confounded by the error term ε_i, so incentives for effort are bounded above. By contrast, in the common type model, other agents' data (which inform about the common type θ) substitutes for i's signal; thus, the larger N is, the less weight the principal puts on the realization of i's output. This de-incentivizes effort. In the limit as $N \to \infty$, the principal can extract θ perfectly from the outputs of other agents and places vanishing weight on agent i's output, so the incentive to exert effort shrinks to zero.

Although this intuition is straightforward, we do not in general have access to the distribution of the principal's posterior expectation, so we cannot directly quantify the "dependence" of the posterior expectation on the output. (One important exception is when the model is Gaussian, in which case the principal's posterior expectation is a convex combination of his prior mean and the realized outputs, with weights that may be computed in closed form as a function of N. See Section 5 for details.) The most technically demanding part of the proof establishes that the expected impact of increasing effort by Δ, i.e. $\mu_N(\Delta) - \mu_N(0)$, can be bounded by an expression that shrinks (for Part (a)) or grows (for Part (b)) in N uniformly for all Δ.[12] This bound is necessary for the limiting results. Note that it is not enough to demonstrate pointwise convergence—for example, if each $\mu_N(\Delta)$ vanishes to zero as we vary N, this would still allow for $\lim_{N\to\infty} MV_T(N) = \lim_{N\to\infty} \mu'_N(0) > 0$. See Appendix B.2 for details.

The following corollary is immediate from Lemma 1:

Corollary 1. *(a) $a^*_T(N)$ is nonincreasing in N and* $\lim_{N\to\infty} a^*(N) = 0$.

[12] An implication of part (a) of Lemma 1 is that as $N \to \infty$, the agent's expectation of the principal's expectation converges to the agent's own expectation; that is, μ. This implication has the flavor of the classic Blackwell and Dubins (1962) result on merging of opinions, which states that if two agents have different prior beliefs which are absolutely continuous with respect to one another, then given sufficient information, their posterior beliefs must converge. However, the Blackwell and Dubins (1962) result demonstrates almost-sure convergence, while we are interested in l_1-convergence under a shifted measure—that is, whether the agent's expectation of the principal's expectation converges to the agent's own expectation given sufficient data, where the agent and principal use different priors. Neither of these two notions of convergence directly imply the other.

(b) $a^*_C(N)$ is nondecreasing in N and $\lim_{N\to\infty} a^*_C(N) < \infty$.

That is, equilibrium effort is weakly decreasing in population size in the common type model, and weakly increasing in the common confound model.[13]

4 Main Results

We now return to the main model, where N is endogenously determined in equilibrium by agents' opt-in decisions.

4.1 Equilibrium

First observe that in equilibrium, the principal correctly de-biases the observed outputs. Therefore, the equilibrium payoff t_2 in the second period is μ (see (1)), and opting-in is preferred over opting-out if and only if the agent's equilibrium action a^* satisfies $t_1 + \mu - C(a^*) > \mu$, or equivalently,

$$t_1 - C(a^*) > 0.$$

Throughout, we impose the following assumption, which guarantees that a single agent would find it strictly optimal to opt-in.

Assumption 6 (Individual Entry)**.** *$t_1 > C(a^*(1))$, where $a^*(1)$ is the equilibrium effort in the exogenous-entry game with a single agent (as defined in (2) with $N = 1$).*

Note that there is always a trivial no-entry equilibrium in which every agent chooses to opt-out, and the principal expects high effort from any agent who deviates to opting in. To refine away equilibria supported by potentially unreasonable beliefs, we require that if a principal observes deviation to opt-in by a single agent, he expects the action $a^*(1)$ that would have been chosen in the single-player exogenous-entry game.

Our main results explain how the equilibrium implications of data sharing differ across the two models:

Theorem 1. *In the common type model, there is a unique symmetric equilibrium for all population sizes N. In this equilibrium, each agent opts-in and chooses effort level $a^*_T(N) \in [0, a^*(1))$, where $a^*_T(N)$ is nonincreasing in N and* $\lim_{N\to\infty} a^*_T(N) = 0$.

[13]Further, the limiting effort level in the common confound model as the population size grows large can be shown to be exactly the equilibrium effort level in a single-agent model in which the agent's output is $S_i = \theta_i + \varepsilon_i$. This is because the confound can be completely removed in the limit.

Theorem 2. *In the common confound model, there exists an $N^* \in \mathbb{R}_+ \cup \{\infty\}$ such that:*

- *If $N \leq N^*$, there is a unique symmetric equilibrium in which each agent opts-in and chooses effort $a^*_C(N) > a^*(1)$, where $a^*_C(N)$ is nondecreasing in N.*
- *If $N > N^*$, there is a unique symmetric equilibrium in which each agent opts-in with probability $p(N) \in (0,1)$ and chooses effort $a^{**} \in [a^*_C(N^*), a^*_C(N^*+1))$. The effort level a^{**} is independent of N, while the opt-in probability $p(N)$ is strictly decreasing in N.*

When the population size is small, all agents opt-in in both models, and the equilibrium effort levels $a^*_T(N)$ and $a^*_C(N)$ are the same as in the exogenous-entry model with a fixed and known population size N (see Corollary 1). Thus, the equilibrium effort levels inherit the properties described in the previous section. Specifically, in the common type model, equilibrium effort is lower than it would have been in a single-agent model. In contrast, in the common confound model, the equilibrium effort level is higher than the single-agent benchmark (and consequently larger also than the equilibrium in the common type model).

As the population size grows large, agents continue to fully opt-in in the common type model, and the equilibrium effort level $a^*_T(N)$ vanishes to zero. But for large populations, (symmetric) pure-strategy equilibria can fail to exist in the common confound model.[14] The threshold population size N^* is pinned down by the equation $C(a^*_C(N^*)) = t_1$, which makes agents indifferent between opting-in at effort level $a^*_C(N^*)$ (i.e. the equilibrium effort in a fixed population of size N^*) and opting out. The threshold N^* is monotonically increasing in t_1, and $N^* = \infty$ for large t_1.

If however N^* is finite (correspondingly, the opt-in reward t_1 is not too large), and the population size exceeds N^*, then full entry cannot be supported in equilibrium. Instead, there is a unique equilibrium in mixed strategies in which agents randomize between opting-in and opting-out. This randomization leads to an opt-in population of uncertain size, where the opt-in probability $p(N)$ is the unique one that satisfies

$$\mathbb{E}\left[MV\left(\widetilde{N} \sim \text{Bin}(N, p(N))\right)\right] = C(a^{**}).$$

In general, this probability $p(N)$ is different from the probability $p^*(N)$ satisfying

$$MV\left[p^*(N) \cdot N\right] = C(a^{**}) \tag{3}$$

namely, the opt-in fraction such that equilibrium effort is a^{**} given *deterministic* entry of $p^*(N) \cdot N$ agents. We show in the subsequent Section 5 that $p(N) \geq p^*(N)$ for all N in

[14] Pure-strategy *asymmetric* equilibria do exist: any strategy profile in which exactly N^* agents opt-in and choose effort level a^{**} is an equilibrium.

the case of normally distributed outputs; that is, uncertainty in the realized population size increases entry.

4.2 Welfare Implications

Following Holmstrom (1982)—in which social welfare is measured as the sum of the expected output and the (single) agent's expected payoff—we use as our measure the expected *total* sum of outputs (for all agents who opt-in) and agent payoffs. If each agent opts-in with probability p and chooses effort a, the implied social welfare is then:[15]

$$\begin{aligned} W(p, a, N) &= pN \cdot (\mathbb{E}(S_i) + t_1 + \mathbb{E}(t_2) - C(a)) \\ &= pN \cdot (2\mu + a + t_1 - C(a)). \end{aligned}$$

From Theorems 1 and 2, equilibrium in the common type model yields welfare

$$W_T(N) \equiv W(1, a_T^*(N), N)$$

while equilibrium in the common confound model yields welfare

$$W_C(N) \equiv \begin{cases} W(1, a_C^*(N), N) & \text{if } N \leq N^* \\ W(p(N), a^{**}, N) & \text{otherwise} \end{cases}$$

Below we compare these quantities against two benchmarks:

First-Best Benchmark. Social welfare is maximized when all agents opt-in and choose the effort level a_{FB} satisfying $C'(a_{FB}) = 1$. This yields

$$W_{FB}(N) \equiv W(1, a_{FB}, N)$$

As in Holmstrom (1982), since the equilibrium action a^* satisfies $C'(a^*) = MV(N)$, it is inefficient whenever $MV(N) \neq 1$. Different from Holmstrom (1982), there is additionally a dimension of endogenous entry, which generates a separate potential source of inefficiency.

No Data Sharing Benchmark. We consider a second benchmark, of particular interest to our motivation, that corresponds to equilibrium in a modified environment in which the

[15] We could alternatively define $W(p, a, N) = pN \cdot (2\mu + a - C(a))$, dropping the agents' first-period reward t_1. This would be reasonable if we interpreted t_1 as a transfer from the principal to the agent, rather as surplus created through the interaction. This modification would not affect any of our results in this section.

principal is not permitted to use data from one agent to predict another's type. That is, let each agent i's reward in the second period be

$$t_2 = \mathbb{E}(\theta_i \mid S_i),$$

so that the principal's expectation of agent i's type is based on i's output alone. In equilibrium, each agent opts-in (by Assumption 6), and chooses effort level

$$a_{NS} \equiv a^*(1) \tag{4}$$

i.e. the action that would be taken (in either the common type or common confound model) for a population of size 1. So our no-sharing welfare benchmark is given by:

$$W_{NS}(N) \equiv W(1, a_{NS}, N)$$

Comparison. The proposition below shows that the equilibrium actions $a_T^*(N)$ and $a_C^*(N)$, as well as the no-sharing action $a_{NS}(N)$, are all below the efficient benchmark a_{FB}. The proposition extends a classic result from Holmstrom (1982), which showed that $a^*(1) < a_{FB}$ in the special case of Gaussian outputs.

Proposition 1. *For every $N \in \mathbb{Z}_{\geq 0}$,*

$$a_T^*(N) \leq a_{NS}(N) \leq a_C^*(N) < a_{FB}.$$

Both inequalities hold with equality if and only if $N = 1$.

As $N \to \infty$, the equilibrium action in the common confound model, $a_C^*(N)$, gets closer to the efficient level but remains bounded away from it. As $N \to \infty$, the equilibrium action in the common type model, $a_T^*(N)$, becomes increasingly inefficient.

The result follows directly from Lemma A.4 in the appendix, which demonstrates that $MV(N) \leq 1$ in both models for all population sizes N. Further taking into account the probability of entry, the subsequent corollary provides a comparison of welfare:

Corollary 2. *For all $N \in \mathbb{Z}_{>0}$,*

$$W_T^*(N) \leq W_{NS}^*(N) < W_{FB}$$

where all inequalities strict whenever $N > 1$. Moreover, there exist $1 \leq \underline{N} \leq \overline{N} < \infty$ s.t.

$$W_T^*(N) < W_{NS}(N) < W_C^*(N) < W_{FB}$$

for all $1 < N \leq \underline{N}$[16] *while*

$$W_C^*(N) < W_T^*(N) < W_{NS}^*(N) < W_{FB}$$

for all $N \geq \overline{N}$.

Equilibrium welfare is always strictly below the first best. Additionally, for all population with $N \geq 2$ agents, the no-sharing welfare benchmark $W_{NS}(N)$ exceeds equilibrium welfare under the common type model. Thus, if players share a common type, aggregation of consumer data *always* leads to a reduction in social welfare. This follows directly from Proposition 1, since there is full entry in the no-sharing benchmark as well as in the common type equilibrium, so the welfare comparison is completely determined by the relative sizes of the equilibrium actions $a_{NS}(N) > a_T^*(C)$.

In contrast, if the correlation across outputs is described via the common confound model, then the comparison is not so clear-cut: the equilibrium action exceeds the no-sharing action, but entry may be only partial. Thus the welfare implications of data sharing depend on the population size. In small populations, there is full-entry, so again the action comparison completely determines welfare, and data sharing leads to an improvement in social welfare. In sufficiently large populations, depressed entry eventually results in lower social welfare despite increased effort levels. These results suggest that blanket regulations for data-sharing may not be appropriate; in particular, policymakers should take into account the way in which the data is aggregated for predictions—i.e. whether the data is aggregated for prediction of a common type, or if it is used to de-bias other outputs.

5 Gaussian Uncertainty

Consider now the special case in which all unknowns are jointly normal, i.e.

$$\begin{pmatrix} \theta_i \\ \eta_i \\ \varepsilon_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\theta^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} \right).$$

In this case sharper results are possible. To simplify the statement of the closed-form expressions, we assume that costs are quadratic throughout this section: $C(a_i) = \frac{1}{2}a_i^2$. None of the results we derive depend qualitatively on this choice of cost function.

[16] If t_1 is sufficiently large, then $\underline{N}$ may be taken to be at least 2. Otherwise, there may be no N for which the chain of inequalities is satisfied with strict inequalities.

Suppose that each member of a population of size N opts-in, and the principal believes that each agent takes action a^*. If some agent i exerts effort $a^* + \Delta$ while expecting every other agent $j \neq i$ to take action a^*, then agent i's expectation of the principal's expectation of θ_i is:

$$\mu_N^T(\Delta) = \mu + \beta_T(N) \cdot \Delta$$
$$\mu_N^C(\Delta) = \frac{\sigma_\varepsilon^2 + \sigma_\theta^2}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} \cdot \mu + \beta_C(N) \cdot \Delta$$

where

$$\beta_T(N) = \frac{\sigma_\theta^2}{N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2} \qquad \beta_C(N) = \sigma_\theta^2 / \left(\sigma_\theta^2 + \frac{\sigma_\eta^2(\sigma_\varepsilon^2 + \sigma_\theta^2)}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} + \sigma_\varepsilon^2\right). \tag{5}$$

See Appendix D.2 for the derivations. These expressions are *linear* in Δ, so that every unit of extra effort improves the agent's expected payoff by the same amount. The marginal value of extra effort at the equilibrium effort level, $\mu_N'(0)$, is simply this constant slope; that is,

$$MV_T(N) = \beta_T(N)$$
$$MV_C(N) = \beta_C(N)$$

The equilibrium condition $MV(N) = C'(a^*)$ simplifies to $a^* = MV(N)$, so it follows from the expression above and Theorems 1 and 2 that:

Corollary 3. *(a) In the common type model, there is a unique symmetric equilibrium for all population sizes N, where each agent opts-in and chooses effort level*

$$a_T^*(N) = \sigma_\theta^2 / (N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2).$$

(b) In the common confound model, there is a unique symmetric equilibrium for all population sizes N. Define

$$N^* \equiv \frac{\sqrt{2t_1}(\sigma_\varepsilon^2 + \sigma_\theta^2)}{\sigma_\theta^2 - \sqrt{2t_1}(\sigma_\theta^2 + \sigma_\varepsilon^2)} - \frac{\sigma_\varepsilon^2 + \sigma_\theta^2}{\sigma_\eta^2} + 1. \tag{6}$$

If $N \leq N^*$, then each agent opts-in and chooses effort level

$$a_C^*(N) = \sigma_\theta^2 / \left(\sigma_\theta^2 + \frac{\sigma_\eta^2(\sigma_\varepsilon^2 + \sigma_\theta^2)}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} + \sigma_\varepsilon^2\right).$$

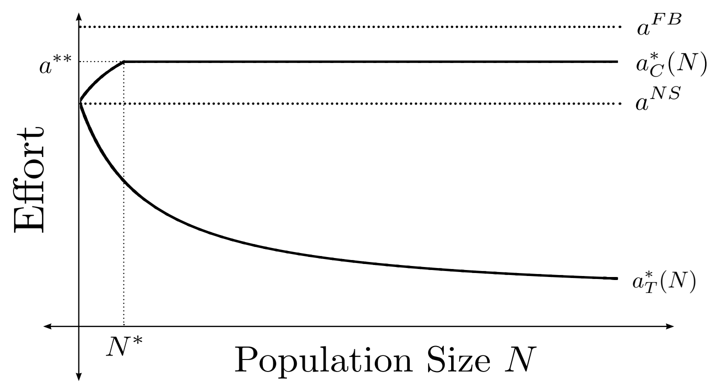
Otherwise, there is a mixed equilibrium in which each agent opts-in with probability $p(N) \in (0,1)$ and chooses effort $\sqrt{2t_1} = a_C^*(N^*)$.

Additionally, it follows from the expressions in (5) that:

Lemma 2. *(a) $MV_T(N)$ is everywhere convex.*

(b) $MV_C(N)$ is everywhere concave.

Thus, uncertainty about the number of entering agents has a dampening effect on incentives in the common confound but strengthens the effect on incentives in the common type model. This implies that $a_T^*(N)$ is convex while $a_C^*(N)$ is concave. The following figure compares these equilibrium actions with the first-best action of $a_{FB} = 1/2$ and the no data sharing action $a_{NS} = a^*(1)$ (as described in Section 4.2).



Additionally, Lemma 2 allows us to make stronger statements about the probability of entry in the mixed equilibrium. Let $p^*(N)$ be the opt-in fraction such that given *deterministic* entry of $p^*(N) \cdot N$ agents, then equilibrium effort is a^{**} (see (3)). Then:

Proposition 2. *Fix any $N > N^*$. In the unique symmetric equilibrium of the common confound model:*

(a) The expected number of agents opting in, $p(N) \cdot N$, is strictly increasing in N.

(b) The probability of entry $p(N)$ strictly exceeds $p^*(N)$.

(c) The equilibrium effort a^{**} is less than the equilibrium effort $a_C^*[p(N) \cdot N]$ associated with full-entry in a deterministic population of size $p(N) \cdot N$.

Part (a) of Proposition 2 says that even though the probability of entry decreases in N, the *expected* number of entrants $p(N) \cdot N$ is increasing in the total population size. Part (b) says that in equilibrium, the probability of entry $p(N)$ exceeds $p^*(N)$. Thus, uncertainty in the number of entrants has the effect of *increasing* the number of expected entrants.

Since any profile in which exactly $p^*(N) \cdot N$ agents opt-in and choose equilibrium effort a^{**} is an asymmetric equilibrium, this further implies that there is strictly less entry in those pure-strategy asymmetric equilibria than in the unique symmetric mixed equilibrium. Finally, part (c) says that the equilibrium effort in this mixed equilibrium is smaller than the effort level that would be chosen given a deterministic population of size $p(N) \cdot N$. That is, uncertainty in the number of entrants has the effect of decreasing effort.

6 Extensions

6.1 Improvements in Monitoring Technology

Suppose that the principal gains access to an improved monitoring technology allowing for further de-noising of each agent's output. How does this affect incentives for effort and equilibrium rates of entry?

To model this question precisely, we assume that each noise term ε_i can be decomposed as

$$\varepsilon_i = \sum_{j=1}^{M} \varepsilon_i^j,$$

where $M \geq 2$ and ε_i^1 through ε_i^M are independent, strictly log-concave, and independent across agents. (They need not be identically distributed.) This decomposition of the error term ε_i can be interpreted as separately accounting for the contribution of a number of different fluctuating environmental variables which affect the outcome of service provision. For instance, a driver's insurance claims may depend on the number of miles he has driven in the last coverage period, the roads on which this driving took place, and the weather and traffic conditions at the time of driving.

We model an improved monitoring technology as direct observation by the principal of ε_i^j for every j in some set $\mathcal{M} \subsetneq \{1, ..., M\}$ and each agent i. Under an improved monitoring technology, the principal gains the ability to directly observe several of the idiosyncratic error shocks confounding each agent's type.[17]

[17] One might alternatively attempt to model improved monitoring via a scaling of the idiosyncratic error term, so that the principal observes $S_i' = \theta_i + \eta_i + \varepsilon_i / M$ for some scale factor M. However, S_i' is not guaranteed to be a more informative signal of θ_i than S_i in the Blackwell order unless ε_i can be decomposed into the sum of a random variable with distribution ε_i / M and another independent random variable. Our construction bypasses this difficulty by imposing the desired decomposability directly. Note that for Gaussian models ε_i is always decomposable, so that the two approaches are equivalent for that class of models.

The following claim summarizes how an improvement in monitoring technology impacts equilibrium outcomes.

Claim 1. *(a) In the common confound model, improved monitoring increases equilibrium effort and decreases entry. Welfare is raised when the population size N is sufficiently small, and is lowered when N is sufficiently large.*

(b) In the common type model, the impact of improved monitoring on equilibrium effort and entry is ambiguous. In the Gaussian model, improved monitoring decreases equilibrium effort, does not impact entry, and reduces welfare for any population size N.

Consider first the common confound model. In this model improved monitoring reduces the effective confound obscuring both η and θ_i. The first effect allows for improved inference of η from all outputs, reducing the effective confound obscuring θ_i due to the common noise term. The second effect directly reduces the remaining confound obscuring θ_i from the idiosyncratic noise term. A shift in the output due to improved effort is then attributed in larger part to a higher θ_i, increasing the marginal value of effort. Equilibrium effort therefore at least weakly increases, and strictly increases whenever all agents enter. Above the full-entry threshold, equilibrium entry rates must drop to maintain the zero-profit equilibium effort level given the higher marginal value of effort at every population size. Below the full-entry threshold, raising effort levels moves effort closer to the first-best level, improving welfare. By contrast, above the full-entry threshold, effort is unchanged but entry rates drop, reducing welfare.

Now consider the common type model. In this model, improved monitoring has two competing effects. First, it makes S_i a more precise signal of θ, thus increasing the impact of an increase in effort on the estimated type. On the other hand, it also reduces the confound on θ in all other agents' outputs, reducing the contribution of S_i to the estimate of θ. The net effect of an improvement in monitoring depends on a balancing of these two forces. In the Gaussian model (as described in Section 5), the net effect is that the expectation of θ is less sensitive to S_i as monitoring improves. It follows that equilibrium effort drops. As all agents enter for any population size in equilibrium, entry rates do not change with lower effort, and the effect of better monitoring is a drop in welfare for all population sizes.

6.2 Comparative Statics in Cost

Consumers in different subpopulations often have different costs to exerting effort. For example, teenage males may have a higher cost to driving safely than women over the age of 40. In China, wealthy and well-connected individuals may have an easier time engaging in pro-social actions (e.g. donating money) towards obtaining a high Social Credit score, compared to more economically disadvantaged individuals. In this subsection we explore how these different costs translate into different equilibrium actions and participation rates.

For this exercise, we restrict to cost functions of the form

$$C(a;\beta) = \beta \cdot \frac{a^\gamma}{\gamma}$$

where $\beta > 0$ is a disutility scale parameter which may vary across subpopulations, while $\gamma > 1$ is common across all agents.[18] Both parameters are commonly known to each agent and the principal. For each β, define $a_T^*(\beta, N)$ and $a_C^*(\beta, N)$ to be the equilibrium actions in the respective models when the cost parameter is β and the opt-in population size is known to be N. (We assume that the cost parameter is homogenous across the subpopulation and solve for an equilibrium of that subpopulation.) These equilibrium action levels satisfy a natural comparative static in β :

Claim 2. *The equilibrium actions $a_T^*(\beta, N)$ and $a_C^*(\beta, N)$ are strictly decreasing in the cost parameter β for each population size N.*

Thus, in populations where the cost of exerting effort is higher, we expect to see lower equilibrium effort, and hence lower outputs. This implies for example that wealthy individuals in China will have higher Social Credit scores compared to economically disadvantaged individuals. Note that in equilibrium those higher scores will be correctly de-biased; that is, other individuals recognize that since the cost to obtaining a high Social Credit score is lower for wealthy individuals, a high score is less indicative of a high-quality individual.

The equilibrium effort levels $a_T^*(\beta, N)$ and $a_T^*(\beta, N)$ also satisfy the same comparative statics in N for fixed β as $a_T^*(N)$ and $a_C^*(N)$ do in the baseline model. So each model $M \in \{T, C\}$, define $N_M^*(\beta)$ to be the unique value of N satisfying

$$C(a_M^*(\beta, N); \beta) = t_1.$$

In the common type model, entry is optimal for an agent with cost parameter β if and only if the number of entrants is at least $N_T^*(\beta)$. Conversely, in the common confound model entry is optimal if and only if the number of entrants is no more than $N_C^*(\beta)$. Below and above these population thresholds, respectively, agents must mix over entry in equilibrium.

These entry thresholds satisfy the following comparative static in β :

[18]Our results also hold for any disutility of effort function C satisfying the "superconvexity" condition $\frac{d}{da} \log \frac{d}{da} \log C(a) > 0$ for all a. All strictly convex power functions satisfy this condition.

Claim 3. *The entry threshold $N_T^*(\beta)$ is decreasing in β, while $N_C^*(\beta)$ is increasing in β.*

This result follows from a calculation of the impact of increasing β on the equilibrium cost-of-effort $C(a_M^*(\beta, N); \beta)$ for a fixed subpopulation size N. On the one hand, increasing β increases the total cost of a given effort level; on the other hand, it depresses equilibrium effort. When C is a power function, the decline in effort dominates, and total effort costs decrease. (More generally, this comparative static holds whenever C satisfies the "super-convexity" condition stated in footnote 18.) A change in β therefore decreases the population threshold for full entry in the common type model, and increases it in the common confound model. Another implication of this result is that in both models, the minimum reward t_1 necessary to incentivize participation by all agents is decreasing in β.

7 Conclusion

As organizations and governments move towards collecting and using large datasets of consumer transactions and behavior for decision-making, the question of whether and how to regulate data sharing has emerged as an important policy question.

The welfare implications of data sharing depend on the criterion one has in mind. Recent regulations, such as the European Union's General Data Protection Regulation (GDPR), have focused on protecting consumers' privacy and improving transparency regarding what kind of data is being collected. These policies are motivated by a goal of preserving basic rights, e.g. a right to privacy or a right to know what others know about you.

A different perspective is that the regulations should be designed based on externalities or market consequences of data sharing. For example, we might care about the extent to which firms can use data for price discrimination, as in Bergemann, Bonatti and Gan (2019). In the present paper, we pose a new question, which is what effect data sharing has on consumer incentives for effort. This is a first-order question because many organizations engaged in data collection either explicitly or implicitly value certain behaviors—banks prefer for agents to pay back their loans; insurance companies prefer for individuals to engage in non-risky behaviors; and the Chinese Social Credit system is ostensibly motivated by a goal of encouraging pro-social behaviors.

We find that the behavioral consequences of data sharing are linked to how this data is correlated across individuals. When an organization aggregates data for the purpose of learning a common average type, then this aggregation results in *lower* effort provision in equilibrium. Thus consumers may prefer for the organization to aggregate data, while the organization may prefer instead to commit to a policy where each agent's data is used to predict their type alone. In contrast, if agents' data can be used to help the organization better interpret or extract information from other agents' data, then the consequence of data sharing is an *increase* in effort. This implies also that rates of participation with the organization may decrease under data sharing, since outcomes in which all agents participate and exert high effort are not sustainable.

These results suggest that regulations should take into account not just whether individual data is informative about other consumers, but *how* it is informative. That relationship is crucial to how data sharing reshapes incentives. In practice, whether data is used to predict a common type, or used for de-biasing other observations, is likely to differ across different domains, and may have as much to do with the underlying correlation structure of the data as it does with the algorithms used to aggregate that data. We do not explicitly model those algorithms here, although that is an interesting question for subsequent work.

Appendix

A Preliminary Results

A.1 Regularity of Posterior Distributions

In this appendix we establish important technical properties of the posterior distribution and mean of each agent's type, conditioning on the set of outputs. These results provide important technical tools underpinning the results of the main text.

For the results of this section, fix a population size N, and assume that all agents opt in. (All results extend immediately to any set of agents $I \subset \{1, ..., N'\}$ of size N entering from a population of size $N' > N$.) Let G_i^M be the marginal distribution of agent i's output in model $M \in \{T, C\}$, with $M = T$ the common type model and $M = C$ the common confound model. We will write g_i^M for the density function associated with G_i^M, which is guaranteed to exist given that f_θ, f_η, and f_ε exist.

Lemma A.1. *For each model $M \in \{T, C\}$, the conditional density functions $g_i^M(S_i \mid \theta_i, \eta_i; a_i)$ and $g_i^M(S_i \mid \theta_i; a_i)$ have the strict monotone likelihood ratio property (MLRP)*[19] *in θ_i, and the density $g_i^M(S_i \mid \eta_i; a_i)$ has the MLRP in η_i. The density $g_i^C(S_i \mid \theta_i, \mathbf{S}_{-i}; \mathbf{a})$ has the MLRP in θ_i.*

Proof. Consider first MLRP of $g_i^M(S_i \mid \theta_i, \eta_i; a_i)$ wrt θ_i. Write

$$g_i^M(S_i = s \mid \theta_i = t, \eta_i = u; a_i) = f_\varepsilon(s - t - u - a_i),$$

and fix $s' > s$ and $t' > t$. Define $z = s - t - u - a_i$, $z' = s' - t' - u - a_i$, $w = s - t' - u - a_i$, and $w' = s' - t - u - a_i$. Then MLRP is equivalent to

$$\frac{f_\varepsilon(z')}{f_\varepsilon(w')} > \frac{f_\varepsilon(w)}{f_\varepsilon(z)},$$

or

$$\frac{1}{2}\log f_\varepsilon(z) + \frac{1}{2}\log f_\varepsilon(z') > \frac{1}{2}\log f_\varepsilon(w) + \frac{1}{2}\log f_\varepsilon(w').$$

[19] A family of density functions $\{f(\cdot \mid \theta)\}$ has the strict monotone likelihood ratio property if for every $x > y$ and $\theta' > \theta$,

$$\frac{f(x \mid \theta')}{f(x \mid \theta)} > \frac{f(y \mid \theta')}{f(y \mid \theta)}.$$

Now, note that $w' > z, z' > w$ and $(z + z')/2 = (w + w')/2$. Thus the distribution $\frac{1}{2}w \oplus \frac{1}{2}w'$ is a mean-preserving spread of $\frac{1}{2}z \oplus \frac{1}{2}z'$, so that the desired inequality is an implication of strict log-concavity of f_ε.

Now consider MLRP of $g_i^M(S_i \mid \theta_i; a_i)$ wrt θ_i. Note that

$$g_i^M(S_i = s \mid \theta_i = t; a_i) = f_{\eta+\varepsilon}(s - t - a_i),$$

where $f_{\varepsilon+\eta}$ is the density function of $\eta_i + \varepsilon_i$. As η_i and ε_i are independent, f_η and f_ε are strictly log-concave, and the convolution of two strictly log-concave concave functions is itself strictly log-concave, $f_{\varepsilon+\eta}$ is itself a strictly log-concave function. The desired MLRP property then follows from an argument essentially identical to the one of the previous paragraph. MLRP of $g_i^M(S_i \mid \eta_i; a_i)$ wrt η_i follows similarly from the fact that the density function of $\theta_i + \varepsilon_i$ is strictly log-concave given that both f_θ and f_ε are strictly log-concave and θ_i and ε_i are independent.

Finally, consider MLRP of $g_i^C(S_i \mid \theta_i, \mathbf{S}_{-i}; \mathbf{a})$ in θ_i. Fix $(\mathbf{S}_{-i}, \mathbf{a}_{-i})$, and define $\widetilde{f}_\eta$ to be the density of η conditional on $(\mathbf{S}_{-i}, \mathbf{a}_{-i})$, with associated distribution function $\widetilde{F}_\eta$. Conditional on S_{-i}, S_i can be written

$$S_i = a_i + \theta_i + \widetilde{\eta}_i + \varepsilon_i,$$

where $\theta_i \sim F_\theta$, $\varepsilon_i \sim F_\varepsilon$, and $\widetilde{\eta}_i \sim \widetilde{F}_\eta$. It is therefore sufficient to show that $\widetilde{f}_\eta$ is strictly log-concave.

Note that the elements of $\mathbf{S}_{-i}$ are mutually independent conditional on η. Then by Bayes' rule,

$$\widetilde{f}_\eta(t) = \frac{f_\eta(\eta = t) \prod_{j \neq i} g_j^C(S_j \mid \eta = t; a_j)}{g^C(\mathbf{S}_{-i} \mid \mathbf{a}_{-i})} = \frac{f_\eta(\eta = t) \prod_{j \neq i} f_{\theta+\varepsilon}(S_j - t - a_j)}{g^C(\mathbf{S}_{-i} \mid \mathbf{a}_{-i})},$$

where $f_{\theta+\varepsilon}$ is the convolution of f_θ and f_ε. Taking logarithms yields

$$\log \widetilde{f}_\eta(t) = \log f_\eta(\eta = t) - \log g^C(\mathbf{S}_{-i} \mid \mathbf{a}_{-i}) + \sum_{j \neq i} \log f_{\theta+\varepsilon}(S_j - t - a_j).$$

As noted previously, $f_{\theta+\varepsilon}$ is strictly log-concave. It follows that $\log f_{\theta+\varepsilon}(S_j - t - a_j)$ is strictly log-concave in t, since for every $t' \neq t$ and $\lambda \in (0, 1)$,

$$\begin{aligned}
&\log f_{\theta+\varepsilon}(S_j - (\lambda t + (1 - \lambda)t') - a_j) \\
= \; &\log f_{\theta+\varepsilon}(\lambda(S_j - t - a_j) + (1 - \lambda)(S_j - t' - a_j)) \\
> \; &\lambda \log f_{\theta+\varepsilon}(S_j - t - a_j) + (1 - \lambda) \log f_{\theta+\varepsilon}(S_j - t' - a_j).
\end{aligned}$$

Therefore $\log \widetilde{f}_\eta$ is a sum of concave and strictly concave functions (with the constant term $-\log g_i^C(\mathbf{S}_{-i} \mid \mathbf{a}_{-i})$ being trivially weakly concave in t), and is thus itself strictly concave, as desired. $\square$

Given a distribution function $H(x \mid z)$ over $\mathbb{R}^K$ conditioned on some vector $z \in \mathbb{R}^J$, we will say that $H(x \mid z)$ satisfies first-order stochastic dominance, or FOSD, in z if $z > z'$ (in the vector partial order on $\mathbb{R}^J$) implies that $H(x \mid z) \leq H(x \mid z')$ for every $x \in \mathbb{R}^K$, with the inequality strict for at least one x.

Lemma A.2. *Fix an action profile* $\mathbf{a}$. *Under Assumption 4:*

(a) $F_\theta^T(\theta \mid \mathbf{S}; \mathbf{a})$ *satisfies FOSD in the realization of* $\mathbf{S}$.

(b) $F_\eta^C(\eta \mid \mathbf{S}; \mathbf{a})$ *satisfies FOSD in the realization of* $\mathbf{S}$.

(c) $F_{\theta_1}(\theta_1 \mid S_1, \eta_1; a_1)$ *satisfies FOSD in the realization of* S_1.

(d) $F_{\theta_1}^C(\theta_1 \mid \mathbf{S}; \mathbf{a})$ *satisfies FOSD in the realization of* $(S_1, -\mathbf{S}_{-1})$.

(e) $G_N^C(S_N \mid \mathbf{S}_{-N}; \mathbf{a})$ *and* $G_N^C(S_N \mid \mathbf{S}_{-N}; \mathbf{a})$ *satisfy FOSD in the realization of* S_1 *for every* $\mathbf{S}_{2:N-1}$.

Proof. Throughout this proof, all action profiles are held constant. We therefore suppress explicit conditioning on the action profile in what follows.

(a) Fix $i \in \{1, 2, ..., N\}$ and $\mathbf{S}_{-i}$. It is enough to show that $F_\theta^T(\theta \mid \mathbf{S})$ satisfies FOSD in S_i. By Proposition 1 of Milgrom (1981), this result holds so long as $g_i^T(S_i \mid \theta, \mathbf{S}_{-i})$ satisfies the MLRP in θ. Note that conditional on θ, the distribution of S_i is independent of $\mathbf{S}_{-i}$. So $g_i^T(S_i \mid \theta, \mathbf{S}_{-i}) = g_i^T(S_i \mid \theta)$, which satisfies the MLRP by Lemma A.1.

(b) Fix $i \in \{1, 2, ..., N\}$. It is enough to show that $F_\eta^C(\eta \mid \mathbf{S})$ satisfies FOSD in S_i. By Proposition 1 of Milgrom (1981), this result holds so long as $g_i^C(S_i \mid \eta, \mathbf{S}_{-i})$ satisfies the MLRP in η. Note that conditional on η, the distribution of S_i is independent of $\mathbf{S}_{-i}$. So $g_i^C(S_i \mid \eta, \mathbf{S}_{-i}) = g_i^C(S_i \mid \eta)$, which satisfies the MLRP by Lemma A.1.

(c) We establish the slightly stronger fact, used in the proof of part (d), that $F_{\theta_1}(\theta_1 \mid \eta_1, S_1)$ is strictly decreasing in S_1 for any (θ_1, η). This is implied by the proof of Proposition 1 in Milgrom (1981)[20] so long as $g_1(S_1 \mid \theta_1, \eta_1)$ satisfies the MLRP in θ_1, a property established in Lemma A.1.

(d) We first consider changing some output S_i for $i > 1$. Fix $i \in \{2, ..., N\}$ and $\mathbf{S}_{-i}$. It is enough to show that $F_{\theta_1}^C(\theta_1 \mid \mathbf{S})$ satisfies FOSD in $-S_i$. The distribution function may be written

$$F_{\theta_1}^C(\theta_1 \mid \mathbf{S}) = \int F_{\theta_1}^C(\theta_1 \mid \eta, \mathbf{S}) \, dF_\eta^C(\eta \mid \mathbf{S}).$$

[20]The proof of that proposition establishes that the posterior distribution function $G(\theta \mid x)$ is strictly decreasing in x for every θ at which $G(\theta) \in (0, 1)$. Since in our setting θ has full support, this implication holds everywhere.

Note that conditional on (η, S_1), the distribution of θ_1 is independent of $\mathbf{S}_{-1}$, i.e. $F^C_{\theta_1}(\theta_1 \mid \eta, \mathbf{S}) = F_\theta(\theta_1 \mid \eta_1, S_1)$. The proof of part (b) established that $F^C_\eta(\eta \mid \mathbf{S})$ satisfies FOSD wrt S_i. It is therefore enough to show that $F_{\theta_1}(\theta_1 \mid \eta_1, S_1)$ is strictly increasing in η_1 for any (θ_1, S_1). Note that because $S_1 = \theta_1 + \eta_1 + \varepsilon_1$ and $\theta_1, \eta_1, \varepsilon_1$ are mutually independent,

$$F_{\theta_1}(\theta_1 \mid \eta_1 = t, S_1 = s_1) = F_{\theta_1}(\theta_1 \mid \eta_1 = t', S_1 = s_1 + t' - t).$$

It is therefore sufficient that $F_{\theta_1}(\theta_1 \mid \eta, S_1)$ is strictly decreasing in S_1 for any (θ_1, η), a fact established in the proof of part (c).

Now consider changing S_1, holding fixed $\mathbf{S}_{-1}$. The desired FOSD relation holds so long as $g^C_1(S_1 \mid \theta_1, \mathbf{S}_{-1}; \mathbf{a})$ satisfies the MLRP in θ_1, which was established in Lemma A.1.

(e) Fix $\mathbf{S}_{2:N-1}$. Consider first the common type model. Note that

$$\begin{aligned} G^T_N(S_N \mid \mathbf{S}_{-N}) &= \int G^T_N(S_N \mid \theta, \mathbf{S}_{-N})\, dF^T_\theta(\theta \mid \mathbf{S}_{-N}) \\ &= \int G^T_N(S_N \mid \theta)\, dF^T_\theta(\theta \mid \mathbf{S}_{-N}), \end{aligned}$$

where $G^T_N(S_N \mid \theta, \mathbf{S}_{-N}) = G^T_N(S_N \mid \theta)$ given that S_N is independent of $\mathbf{S}_{-N}$ conditional on θ. We will show that $F^T_\theta(\theta \mid \mathbf{S}_{-N}; a_1)$ satisfies FOSD in S_1 and $G^T_N(S_N \mid \theta)$ is a strictly decreasing function of θ, which together imply the desired FOSD property.

First observe that the logic of the proof of part (a) did not depend on the number of outputs conditioned upon. Hence $F^T_\theta(\theta \mid \mathbf{S}_{-N})$ satisfies FOSD in $\mathbf{S}_{-N}$, and in particular in S_1. Meanwhile, $G^T_N(s_N \mid \theta = t) = \Pr(\eta_N + \varepsilon_N \le s_N - t - a_N)$, which is strictly decreasing in t given that $\eta_N + \varepsilon_N$ has full support over $\mathbb{R}$.

Now consider the common confound model. Note that

$$\begin{aligned} G^C_N(S_N \mid \mathbf{S}_{-N}) &= \int G^C_N(S_N \mid \eta, \mathbf{S}_{-N})\, dF^C_\eta(\eta \mid \mathbf{S}_{-N}) \\ &= \int G^C_N(S_N \mid \eta)\, dF^C_\eta(\eta \mid \mathbf{S}_{-N}), \end{aligned}$$

where $G^C_N(S_N \mid \eta, \mathbf{S}_{-N}) = G^C_N(S_N \mid \eta)$ given that S_N is independent of $\mathbf{S}_{-N}$ conditional on η. We will show that $F^C_\eta(\eta \mid \mathbf{S}_{-N})$ satisfies FOSD in S_1 and $G^C_N(S_N \mid \eta)$ is a strictly decreasing function of η, which together imply the desired FOSD property.

First observe that the logic of the proof of part (b) did not depend on the number of outputs conditioned upon. Hence $F^C_\eta(\eta \mid \mathbf{S}_{-N})$ satisfies FOSD in $\mathbf{S}_{-N}$, and in particular in S_1. Meanwhile, $G^C_N(s_N \mid \eta = t) = \Pr(\theta_N + \varepsilon_N \le s_N - t - a_N)$, which is strictly decreasing in t given that $\theta_N + \varepsilon_N$ has full support over $\mathbb{R}$. $\square$

Lemma A.3. *For each model $M \in \{T, C\}$ and every agent i :*

(a) $0 \leq \frac{\partial}{\partial s_i}\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] \leq 1$ *for every* $(\mathbf{s}, \mathbf{a})$*. Further, for every* $s_i' > s_i$,

$$0 < \mathbb{E}[\theta_i \mid \mathbf{S} = (s_i', \mathbf{s}_{-i}); \mathbf{a}] - \mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] < s_i' - s_i.$$

(b) $0 \leq \frac{\partial}{\partial s_i}\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i] \leq 1$ *for every* $(s_i, t; a_i)$*. Further, for every* $s_i' > s_i$,

$$0 < \mathbb{E}[\theta_i \mid S_i = s_i', \eta_i = t; a_i] - \mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i] < s_i' - s_i.$$

Proof. Fix an agent i. By Lemma A.1, for each model M the distribution $F^M_{\theta_1}(\theta_1 \mid \mathbf{S}; \mathbf{a})$ satisfies FOSD in S_1, implying that $\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}]$ is strictly increasing in s_1 and hence that

$$\frac{\partial}{\partial s_1}\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] \geq 0$$

and

$$\mathbb{E}[\theta_i \mid \mathbf{S} = (s_i', \mathbf{s}_{-i}); \mathbf{a}] - \mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] > 0$$

for every $s_i' > s_i$. The same lemma established that the distribution $F_{\theta_i}(\theta_i \mid S_i, \eta_i; a_i)$ satisfies FOSD in S_i, implying that $\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i]$ is strictly increasing in s_i and so

$$\frac{\partial}{\partial s_1}\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i] \geq 0$$

and

$$\mathbb{E}[\theta_i \mid S_i = s_i', \eta_i = t; a_i] - \mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i] > 0$$

for every $s_i' > s_i$.

Next, take conditional expectations of $S_i = a_i + \theta_i + \eta_i + \varepsilon_i$ given $(\mathbf{S}, \mathbf{a})$ to obtain the identity

$$S_i = a_i + \mathbb{E}[\theta_i \mid \mathbf{S}; \mathbf{a}] + \mathbb{E}[\eta_i + \varepsilon_i \mid \mathbf{S}; \mathbf{a}].$$

Suppose that

$$\frac{\partial}{\partial s_i}\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] > 1$$

for some s_i. This implies that there must be some $s_i' > s_i$ such that

$$\mathbb{E}[\theta_i \mid \mathbf{S} = (s_i, \mathbf{s}_{-i}); \mathbf{a}] > \mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}] + (s_i' - s_i).$$

The desired upper bound on $\frac{\partial}{\partial s_i}\mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}]$ is therefore implied by the desired upper bound on $\mathbb{E}[\theta_i \mid \mathbf{S} = (s_i', \mathbf{s}_{-i}); \mathbf{a}] - \mathbb{E}[\theta_i \mid \mathbf{S} = \mathbf{s}; \mathbf{a}]$. If this bound did not hold for some s_i and s_i', then it would follow that

$$\mathbb{E}[\eta_i + \varepsilon_i \mid \mathbf{S} = (s_i', \mathbf{s}_{-i}); \mathbf{a}] \leq \mathbb{E}[\eta_i + \varepsilon_i \mid \mathbf{S} = \mathbf{a}; \mathbf{a}].$$

The desired bound therefore holds if we can establish that $\mathbb{E}[\varepsilon_i \mid \mathbf{S}; \mathbf{a}]$ and $\mathbb{E}[\eta_i \mid \mathbf{S}; \mathbf{a}]$ are each strictly increasing in s_i everywhere, comparative statics that hold if $F^M_{\varepsilon_i}(\varepsilon_i \mid \mathbf{S}; \mathbf{a})$ and $F^M_{\eta_i}(\eta_i \mid \mathbf{S}; \mathbf{a})$ each satisfy FOSD in S_i.

The result for $F^M_{\varepsilon_i}(\varepsilon_i \mid \mathbf{S}; \mathbf{a})$ holds because ε_i is independent of $\mathbf{S}_{-i}$ and $g^M(S_i \mid \varepsilon_i; a_i)$ satisfies MLRP in ε_i given the strict log-concavity of f_θ and f_η, and therefore also of $f_{\theta+\eta}$. The same logic follows for $F^T_{\eta_i}(\eta_i \mid \mathbf{S}; \mathbf{a})$, as η_i is also independent of $\mathbf{S}_{-i}$ in the common type model. And the result for $F^C_{\eta_i}(\eta_i \mid \mathbf{S}; \mathbf{a})$ was established in Lemma A.2.

Finally, take conditional expectations of $S_i = a_i + \theta_i + \eta_i + \varepsilon_i$ given $(S_i, \eta; a_i)$ to obtain the identity

$$S_i = a_i + \mathbb{E}[\theta_i \mid S_i, \eta_i; a_i] + \eta_i + \mathbb{E}[\varepsilon_i \mid S_i, \eta_i; a_i].$$

By logic similar to the previous paragraphs, the desired upper bounds on $\frac{\partial}{\partial s_i}\mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i]$ and $\mathbb{E}[\theta_i \mid S_i = s_i', \eta_i = t; a_i] - \mathbb{E}[\theta_i \mid S_i = s_i, \eta_i = t; a_i]$ hold if we can show that $F_{\varepsilon_i}(\varepsilon_i \mid S_i = s_i, \eta_i = t; a_i)$ satisfies FOSD wrt s_i. This function is identical to the function $F_{\varepsilon_i}(\varepsilon_i \mid \theta_i + \varepsilon_i = s_i + t - a_i)$, so it is sufficient to show that $F_{\varepsilon_i}(\varepsilon_i \mid \theta_i + \varepsilon_i)$ satisfies FOSD wrt $\theta_i + \varepsilon_i$. This follows so long as the density of $\theta_i + \varepsilon_i$ conditional on ε_i satisfies the MLRP wrt ε_i, which is implied by strict log-concavity of f_θ. □

A.2 Sufficiency of the First-Order Approach

In this appendix we formally establish that there exists a unique equilibrium to the exogenous-entry model, which is characterized by the first-order condition described in the body of the paper. Fix a population size N, and assume all agents in the population enter in the first period. For every $\alpha \in \mathbb{R}^N_+$ and $\Delta \geq -\alpha_1$, define

$$\mu(\Delta; \alpha) \equiv \mathbb{E}[\mathbb{E}[\theta_1 \mid \mathbf{S}; \mathbf{a} = \alpha] \mid \mathbf{a} = (\alpha_1 + \Delta, \alpha_{-1})]$$

to be agent 1's expected second-period payoff from exerting effort $\alpha_1 + \Delta$ when the principal expects each agent $i \in \{1, ..., N\}$ to exert effort α_i.

Lemma A.4. *The value function $\mu(\Delta; \alpha)$ and its derivatives satisfy the following properties:*

(a) $\mu(\Delta; \alpha)$ is independent of α and is continuous and strictly increasing in Δ.

(b) $\mu'(\Delta; \alpha)$ exists, is continuous in Δ, and satisfies $0 \leq \mu'(\Delta; \alpha) \leq 1$ for every Δ.

(c) $D^+\mu'(\Delta; \alpha) \leq K$for every Δ.[21]

[21] Given a function $f : \mathbb{R} \to \mathbb{R}$, the Dini derivative D^+ is a generalization of the derivative existing for arbitrary functions and defined by $D^+f(x) = \limsup_{h\downarrow 0}(f(x+h) - f(x))/h$. When f is differentiable at a point x, $D^+f(x) = f'(x)$.

Proof. We prove the result for the common type case, with the result for the common confound model following from nearly identical work. The quantity $\mu(\Delta;\alpha)$ can be written explicitly as

$$\mu(\Delta;\alpha) = \int dG^T(\mathbf{S} = \mathbf{s} \mid \mathbf{a} = (\alpha_1 + \Delta, \alpha_{-1}))\, \mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha)].$$

Further,

$$\mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha] = \int \theta \, dF_\theta^T(\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha),$$

and by Bayes' rule

$$f_\theta^T(\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha) = \frac{g^T(\mathbf{S} = \mathbf{s} \mid \theta; \mathbf{a} = \alpha) f_\theta(\theta)}{g^T(\mathbf{S} = \mathbf{s} \mid \mathbf{a} = \alpha)}.$$

Since effort affects the output as an additive shift, $g^T(\mathbf{S} = \mathbf{s} \mid \mathbf{a} = \alpha) = g^T(\mathbf{S} = \mathbf{s} - \alpha \mid \mathbf{a} = \mathbf{0})$ and $g^T(\mathbf{S} = \mathbf{s} \mid \theta; \mathbf{a} = \alpha) = g^T(\mathbf{S} = \mathbf{s} - \alpha \mid \theta; \mathbf{a} = \mathbf{0})$. So

$$\begin{aligned} f_\theta^T(\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha) &= \frac{g^T(\mathbf{S} = \mathbf{s} - \alpha \mid \theta; \mathbf{a} = \mathbf{0}) f_\theta(\theta)}{g^T(\mathbf{S} = \mathbf{s} - \alpha \mid \mathbf{a} = \mathbf{0})} \\ &= f_\theta^T(\theta \mid \mathbf{S} = \mathbf{s} - \alpha, \alpha = \mathbf{0}). \end{aligned}$$

Thus

$$\mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \alpha] = \int \theta \, dF_\theta^T(\theta \mid \mathbf{S} = \mathbf{s} - \alpha; \mathbf{a} = \mathbf{0}) = \mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s} - \alpha; \mathbf{a} = \mathbf{0}].$$

Then $\mu(\Delta;\alpha)$ may be equivalently written

$$\mu(\Delta;\alpha) = \int dG^T(\mathbf{S} = \mathbf{s} - \alpha \mid \mathbf{a} = (\Delta, \mathbf{0}))\, \mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s} - \alpha; \mathbf{a} = \mathbf{0})].$$

Using the change of variables $\mathbf{s}' = \mathbf{s} - \alpha$ then reveals that $\mu(\Delta;\alpha) = \mu(\Delta;\mathbf{0})$, so μ is indeed independent of α.

Now fix Δ and $\Delta' < \Delta$. Since effort affects the output as an additive shift, $G^T(\mathbf{S} = \mathbf{s} \mid \mathbf{a} = (\alpha_1 + \Delta, \alpha_{-1})) = G^T(\mathbf{S} = (s_1 - (\Delta - \Delta'), \mathbf{s}_{-1}) \mid \mathbf{a} = (\alpha_1 + \Delta, \alpha_{-1}))$ for every s_1. Then defining a change of variables via $s_1' = s_1 - (\Delta - \Delta')$ and $\mathbf{s}_{-i}' = \mathbf{s}_{-i}$, the previous integral expression for $\mu(\Delta;\alpha)$ may be equivalently written

$$\mu(\Delta;\alpha) = \int dG^T(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = (\alpha_1 + \Delta', \alpha_{-1}))\, \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + (\Delta - \Delta'), \mathbf{s}_{-1}'); \mathbf{a} = \alpha)].$$

Now by Lemma A.2, $F_\theta^T(\theta \mid \mathbf{S}; \mathbf{a})$ satisfies FOSD in the realization of $\mathbf{S}$, so $\mathbb{E}[\theta \mid \mathbf{S}; \mathbf{a} = \alpha]$ is strictly increasing in S_1. Hence

$$\mu(\Delta;\alpha) > \int dG^T(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = (\alpha_1 + \Delta', \alpha_{-1}))\, \mathbb{E}[\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a} = \alpha)] = \mu(\Delta';\alpha).$$

So μ is a strictly increasing function of Δ.

Next, setting $\Delta' = 0$ in the previous expression for $\mu(\Delta; \alpha)$ yields

$$\mu(\Delta; \alpha) = \int dG^T(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = \alpha)\, \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)].$$

Now, by Assumption 2,

$$\frac{\partial}{\partial \Delta} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)]$$

exists, and by Lemma A.3 it is bounded in the interval $[0, 1]$ everywhere. Then by the Leibniz integral rule, $\mu'(\Delta; \alpha)$ exists and

$$\mu'(\Delta; \alpha) = \int dG^T(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = \alpha)\, \frac{\partial}{\partial \Delta} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)],$$

and in particular $\mu'(\Delta; \alpha) \in [0, 1]$. An immediate corollary is that $\mu(\Delta; \alpha)$ is continuous everywhere.

Meanwhile, by Assumption 5

$$\frac{\partial^2}{\partial \Delta^2} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)]$$

exists and is bounded in the interval $(-\infty, K]$ everywhere. This implies that $\frac{\partial}{\partial \Delta} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)]$ is continuous in Δ everywhere, which along with uniform boundedness implies that $\mu'(\Delta; \alpha)$ is continuous in Δ everywhere by the bouned convergence theorem. Further, for each $\delta > 0$ and $(\mathbf{s}, \mathbf{a}, \Delta)$, the mean value theorem implies that

$$\begin{aligned}
&\frac{1}{\delta} \left(\frac{\partial}{\partial \Delta} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta + \delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)] - \frac{\partial}{\partial \Delta} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta, \mathbf{s}_{-1}'); \mathbf{a} = \alpha)] \right) \\
&= \frac{\partial^2}{\partial \Delta^2} \mathbb{E}[\theta \mid \mathbf{S} = (s_1' + \Delta + \delta', \mathbf{s}_{-1}'); \mathbf{a} = \alpha)] \leq K
\end{aligned}$$

for some $\delta' \in [0, \delta]$. Reverse Fatou's lemma then implies that $D^+ \mu'(\Delta; \alpha) \leq K$ and $D^- \mu'(\Delta; \alpha) \leq K$. □

Lemma A.5. $\mu(\Delta; \alpha) - C(\alpha_1 + \Delta)$ *is a strictly concave function of* Δ *for any* α.

Proof. Fix an α, and define $\phi(\Delta) \equiv \mu(\Delta; \alpha) - C(\alpha_1 + \Delta)$. By Lemma A.4, ϕ' exists and is continuous everywhere. We establish the necessary and sufficient condition for strict concavity that ϕ' is strictly decreasing. We invoke the basic monotonicity theorem from analysis that any function f which is continuous and satisfies $D^+ f \geq 0$ everywhere is nondecreasing everywhere. We apply this result to $-\mu'(\Delta; \alpha) + K\Delta$. Using basic properties of the Dini derivatives D^+ and D_+, we have $D^+(-\mu'(\Delta; \alpha)) = -D_+ \mu'(\Delta; \alpha) \geq -D^+ \mu'(\Delta; \alpha)$. Then since

$K\Delta$ is differentiable and $D^+\mu'(\Delta;\alpha) \le K$ from Lemma A.4, we have $D^+(-\mu'(\Delta;\alpha)+K\Delta) = D^+(-\mu'(\Delta;\alpha)) + K \ge 0$. So $\mu'(\Delta;\alpha) - K\Delta$ is nonincreasing everywhere. So choose any Δ and $\Delta' > \Delta$. Then

$$\phi'(\Delta') = \mu'(\Delta';\alpha) - K\Delta' + K\Delta' - C'(\alpha_1 + \Delta') \le \mu'(\Delta;\alpha) + K(\Delta' - \Delta) - C'(\alpha_1 + \Delta').$$

But also by Assumption 5, $C''(\alpha_1 + \Delta'') > K$ for every $\Delta'' \in (\Delta, \Delta')$, so $C'(\alpha_1 + \Delta') > C'(\alpha_1 + \Delta) + K(\Delta' - \Delta)$. Thus

$$\phi'(\Delta') < \mu'(\Delta;\alpha) - C'(\alpha_1 + \Delta) = \phi'(\Delta),$$

as desired. □

Proposition A.1. *There exists a unique equilibrium action profile characterized by $a_i = a_i^*(N)$ for each player i, where $a_i^*(N)$ is the unique solution to*

$$\mu'(0;\mathbf{a}^*(N)) = C'(a^*(N)).$$

Proof. Lemma A.4 established that $\mu'(0;\mathbf{a}^*(N))$ is well-defined, independent of $a^*(N)$, and bounded in the interval $[0,1]$. Then as C' is continuous, strictly increasing, and satisfies $C'(0) = 0$ and $C'(\infty) = \infty$, there exists a unique solution to the stated first-order condition. This solution constitutes an equilibrium so long as $\Delta = 0$ maximizes the objective function $\mu(\Delta;\mathbf{a}^*(N)) - C(a^*(N) + \Delta)$, which is guaranteed by the fact, established in Lemma A.5, that this function is strictly concave in Δ. □

Define $MV(N) \equiv \mu'(0;\mathbf{a}^*(N))$ for each N. When we wish to make the model clear, we will write $MV_M(N)$ for $M \in \{T, C\}$. We conclude this appendix by establishing key bounds on the marginal value of effort which will be useful in what follows.

Lemma A.6. *For every population size N, $0 < MV(N) < 1$. Further, $0 < \lim_{N\to\infty} MV_C(N) < 1$.*

Proof. Fix a model M and population size N. As noted in the proof of Lemma A.4, $MV(N) = \mu'(0;\mathbf{a}^*(N))$ is bounded in $[0,1]$ and may be written

$$MV(N) = \int dG^M(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = \mathbf{a}^*(N))\, \frac{\partial}{\partial s_1}\mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)].$$

Now, suppose that $MV(N) = 0$. Fix $\underline{s}$ and $\overline{s} > \underline{s}$. By Lemma A.3, $\frac{\partial}{\partial s_1}\mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)]$ is everywhere non-negative. Then the expression for $MV(N)$ stated previously implies that

$$0 = \int_{\underline{s}}^{\overline{s}} ds'_N \ldots \int_{\underline{s}}^{\overline{s}} d\mathbf{s}'_1\, g^M(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = \mathbf{a}^*(N))\, \frac{\partial}{\partial s_1}\mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)],$$

and since $g^M > 0$ everywhere, also

$$0 = \int_{\underline{s}}^{\overline{s}} ds'_N ... \int_{\underline{s}}^{\overline{s}} d\mathbf{s}'_1 \frac{\partial}{\partial s_1} \mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)] = 0.$$

Evaluating the inner integral yields

$$0 = \int_{\underline{s}}^{\overline{s}} ds'_N ... \int_{\underline{s}}^{\overline{s}} d\mathbf{s}'_2 \left(\mathbb{E}[\theta_1 \mid \mathbf{S} = (\overline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)] - \mathbb{E}[\theta_1 \mid \mathbf{S} = (\underline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)] \right).$$

But by Lemma A.3, $(\mathbb{E}[\theta_1 \mid \mathbf{S} = (\overline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)] > \mathbb{E}[\theta_1 \mid \mathbf{S} = (\underline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)]$ for every $\mathbf{s}_{-1}$, a contradiction. So it must be that $MV(N) > 0$.

Next, suppose that $MV(N) = 1$. Then the previous expression for $MV(N)$ implies

$$0 = \int dG^M(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = \mathbf{a}^*(N)) \left(1 - \frac{\partial}{\partial s_1} \mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)] \right).$$

As Lemma A.3 implies that $1 - \frac{\partial}{\partial s_1} \mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a} = \mathbf{a}^*(N)]$ is everywhere non-negative, a chain of reasoning very similar to that performed above yields

$$0 = \int_{\underline{s}}^{\overline{s}} ds'_N ... \int_{\underline{s}}^{\overline{s}} d\mathbf{s}'_2 \left(\overline{s} - \underline{s} - (\mathbb{E}[\theta_1 \mid \mathbf{S} = (\overline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)] - \mathbb{E}[\theta_1 \mid \mathbf{S} = (\underline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)]) \right),$$

contradicting the result from Lemma A.3 that

$$\overline{s} - \underline{s} > \mathbb{E}[\theta_1 \mid \mathbf{S} = (\overline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)] - \mathbb{E}[\theta_1 \mid \mathbf{S} = (\underline{s}, \mathbf{s}_{-1}); \mathbf{a} = \mathbf{a}^*(N)]$$

for every $\mathbf{s}_{-1}$. So it must be that $MV_N(0) < 1$.

To obtain bounds on $\lim_{N\to\infty} MV_C(N)$, We use the result derived in the proof of Lemma 1 that $\lim_{N\to\infty} MV_C(N) = \mu'_\infty(0)$, where $\mu'_\infty(0)$ is the equilibrium marginal value of effort in a one-agent model where η_1 is observed by the principal. Since Lemma A.3 establishes comparative statics for $\mathbb{E}[\theta_1 \mid S_1, \eta_1; a_1]$ wrt S_1 identical to those for $\mathbb{E}[\theta_1 \mid \mathbf{S} = \mathbf{s}; \mathbf{a}]$ in the N-agent model, the same arguments used above may be applied to establish the desired bounds. □

B Proofs for Section 3 (Exogenous Entry)

B.1 Proof of Lemma 1

Throughout this proof, we will without loss of generality consider agent 1's problem. Notationally, we will write $\mathbf{S} = (S_1, ..., S_N)$ to denote the vector of all outputs and $\mathbf{S}_{-i} = (S_1, ..., S_{i-1}, S_{i+1}, ..., S_N)$ to denote the vector of outputs generated by all agents except agent $i \in \{1, ..., N\}$. We will also use the notation $\mathbf{S}_{i:j}$ to indicate the vector of outputs of agent i through agent j.

B.1.1 Monotonicity in N

We now establish the monotonicity portion of the lemma. It is sufficient to prove that for each $\Delta > 0$, $\mu_N(\Delta)$ is strictly decreasing in N in the common type model, and strictly increasing in N in the common confound model. For then the fact that $MV(N) = \mu'_N(0)$ exists, established in Lemma A.4, implies that for $N' > N$, in the common type model

$$\mu'_{N'}(0) = \lim_{\Delta \downarrow 0} \frac{\mu_{N'}(\Delta) - \mu}{\Delta} \leq \lim_{\Delta \downarrow 0} \frac{\mu_N(\Delta) - \mu}{\Delta} = \mu'_N(0),$$

and by similar reasoning in the common confound model $\mu'_{N'}(0) \geq \mu'_N(0)$.

Fix any population size $N+1$ and $\Delta > 0$. We embed the N-agent model in the $(N+1)$-agent model by considering it to consist of the first N agents' outputs. When vectors are written without subscripting, they will denote $(N+1)$-agent vectors. Let $\mathbf{a}^*(N+1)$ be the action vector with every entry equal to $a^*(N+1)$, and $\mathbf{a}^*(N)$ be action vector with every entry equal to $a^*(N)$.

Consider first the common type model. By definition

$$\begin{aligned}\mu_N(\Delta) = &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N) + \Delta, \mathbf{a}^*(N)_{2:N})) \\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S}_{1:N} = \mathbf{s}_{1:N}; \mathbf{a}_{1:N} = \mathbf{a}^*(N)_{1:N}) \right).\end{aligned}$$

Now note that because a shift in the action vector shifts up the output vector by the same amount for any realization of $\theta, \eta, \varepsilon$, we have

$$G^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}_{1:N} \mid \mathbf{a}_{1:N}) = G^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}_{1:N} + \delta\mathbf{a}_{1:N} \mid \mathbf{a}_{1:N} + \delta\mathbf{a}_{1:N})$$

and

$$F^T_\theta(\theta \mid \mathbf{S}_{1:N} = \mathbf{s}_{1:N}; \mathbf{a}_{1:N} = \mathbf{a}^*(N)_{1:N}) = F^T_\theta(\theta \mid \mathbf{S}_{1:N} = \mathbf{s}_{1:N} + \delta\mathbf{a}_{1:N}; \mathbf{a}_{1:N} + \delta\mathbf{a}_{1:N})$$

for any output realization $\mathbf{s}_{1:N}$ and any action vectors $\mathbf{a}_{1:N}$ and $\delta\mathbf{a}_{1:N}$. So let $\delta\mathbf{a}^*(N) \equiv \mathbf{a}^*(N+1) - \mathbf{a}^*(N)$. Then the previous representation of $\mu_N(\Delta)$ may be equivalently written

$$\begin{aligned}\mu_N(\Delta) = &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}_{1:N} + \delta\mathbf{a}^*_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N+1) + \Delta, \mathbf{a}^*(N+1)_{2:N})) \\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S}_{1:N} = \mathbf{s}_{1:N} + \delta\mathbf{a}^*_{1:N}; \mathbf{a}_{1:N} = \mathbf{a}^*(N+1)_{1:N}) \right).\end{aligned}$$

By a change of variables to the integrator $\mathbf{s}'_{1:N} = \mathbf{s}_{1:N} + \delta\mathbf{a}^*_{1:N}$, this may simply be written

$$\begin{aligned}\mu_N(\Delta) = &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}'_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N+1) + \Delta, \mathbf{a}^*(N+1)_{2:N})) \\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S}_{1:N} = \mathbf{s}'_{1:N}; \mathbf{a}_{1:N} = \mathbf{a}^*(N+1)_{1:N}) \right).\end{aligned}$$

Now, by the law of iterated expectations this expression may be expanded as

$$\begin{aligned}\mu_N(\Delta) = &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}'_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N}))\\ &\times \int dG^T_{N+1}(S_{N+1} = s'_{N+1} \mid \mathbf{S}_{1:N} = \mathbf{s}'_{1:N}; \mathbf{a} = \mathbf{a}^*(N+1))\\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a} = \mathbf{a}^*(N+1))\right).\end{aligned}$$

Due to the impact on the model of a shift in the action vector noted earlier, we have

$$\begin{aligned}&G^T_{N+1}(S_{N+1} = s'_{N+1} \mid \mathbf{S}_{1:N} = \mathbf{s}'_{1:N}; \mathbf{a} = \mathbf{a}^*(N+1))\\ &= G^T_{N+1}(S_{N+1} = s'_{N+1} \mid \mathbf{S}_{1:N} = (\mathbf{s}'_1 + \Delta, \mathbf{s}'_{2:N}); \mathbf{a} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N+1}))\end{aligned}$$

for every $\mathbf{s}'$. So the previous expression for $\mu_N(\Delta)$ may be equivalently written

$$\begin{aligned}\mu_N(\Delta) = &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}'_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N}))\\ &\times \int dG^T_{N+1}(S_{N+1} = s'_{N+1} \mid \mathbf{S}_{1:N} = (\mathbf{s}'_1 + \Delta, \mathbf{s}'_{2:N}); \mathbf{a} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N+1}))\\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a} = \mathbf{a}^*(N+1))\right).\end{aligned}$$

Now, by Lemma A.2 we know that $F^T_\theta(\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a})$ satisfies FOSD in $\mathbf{S}$ and thus $\int \theta \, dF^T_\theta(\theta \mid \mathbf{S}; \mathbf{a})$ is strictly increasing in S_{N+1}. The same lemma also tells us that $G^T_{N+1}(S_{N+1} \mid \mathbf{S}_{1:N}; \mathbf{a})$ satisfies FOSD in S_1. Hence

$$\begin{aligned}\mu_N(\Delta) > &\int dG^T_{1:N}(\mathbf{S}_{1:N} = \mathbf{s}'_{1:N} \mid \mathbf{a}_{1:N} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N}))\\ &\times \int dG^T_{N+1}(S_{N+1} = s'_{N+1} \mid \mathbf{S}_{1:N} = s'_{1:N}; \mathbf{a} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N+1}))\\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a} = \mathbf{a}^*(N+1))\right).\end{aligned}$$

By the law of iterated expectations, the right-hand size of this inequality may be written more simply as

$$\begin{aligned}\mu_N(\Delta) > &\int dG^T(\mathbf{S} = \mathbf{s}' \mid \mathbf{a} = (a^*(N+1)+\Delta, \mathbf{a}^*(N+1)_{2:N+1}))\\ &\times \left(\int \theta \, dF^T_\theta(\theta \mid \mathbf{S} = \mathbf{s}'; \mathbf{a} = \mathbf{a}^*(N+1))\right).\end{aligned}$$

But this final expression is exactly $\mu_{N+1}(\Delta)$, and so $\mu_N(\Delta) > \mu_{N+1}(\Delta)$. As this argument holds for arbitrary N, we conclude that $\mu_N(\Delta)$ is strictly decreasing in N, as desired.

The result for the common confound model follows by nearly identical reasoning, with one key difference: in this model Lemma A.2 tells us that $F^C_{\theta_1}(\theta_1 \mid \mathbf{S}; \mathbf{a})$ satisfies FOSD wrt $-\mathbf{S}_{-1}$ and thus that $\int \theta \, dF^C_{\theta_1}(\theta_1 \mid \mathbf{S}; \mathbf{a})$ is strictly decreasing in S_{N+1}. So the key inequality is reversed versus the common type model, implying that $\mu_N(\Delta) < \mu_{N+1}(\Delta)$ for all N.

B.2 The $N \to \infty$ limit

We now establish the limiting value of $MV(N) = \mu'_N(0)$ as $N \to \infty$ for each model.

Consider a limiting model in which the principal observes a countably infinite vector of outputs $\mathbf{S} = (S_1, S_2, ...)$. By the law of large numbers, in the common type model this means that the principal perfectly infers θ, while in the common confound model the principal perfectly infers η. Define $\mu_\infty(\Delta; \alpha)$ analogously to the finite-population case. In the common type model the principal's posterior inference is not affected by the value of a single output, and so each agent's expectation of the principal's forecast at any level of distortion Δ is $\mu_\infty(\Delta, \alpha) = \mu$. Then $a^*(\infty) = 0$ is the unique action level such that $\mu'_\infty(0, \mathbf{a}^*(\infty)) = C'(a^*(\infty))$. Meanwhile in the common confound model, Assumption 2 and reasoning very similar to the proof of Lemma A.4 imply that $\mu'_\infty(0, \alpha)$ exists, is independent of α, and lies in $[0, 1]$. So there exists a unique, finite $a^*(\infty)$ satisfying $\mu'(0; \mathbf{a}^*(\infty)) = C'(a^*(\infty))$. Define $\mu'_\infty(0) = \mu'(0; \mathbf{a}^*(\infty))$ in each model.

We will show that $\mu'_N(0) \to \mu'_\infty(0)$ in each model. Given that $\mu'_\infty(0) = 0$ in the common type model, this result implies in particular that that $MV(N) \to 0$ in that model as $N \to \infty$.

To prove the result, we will need the ability to change measure between the distribution of outputs at the equilibrium action profile, and one in which a single agent, without loss agent 1, deviates to a different action. For each model, define a reference probability space $(\Omega, \mathcal{F}, \mathcal{P}^{\mathbf{a}})$, containing all relevant random variables for arbitrary population sizes. For the common type model this space supports the latent types $\theta, \eta_1, \eta_2, ...$ as well as the outputs $S_1, S_2, ...$ Similarly, in the common confound model the space supports $\eta, \theta_1, \theta_2, ...$ and $S_1, S_2, ...$ In each model the probability measure $\mathcal{P}^{\mathbf{a}}$ depends on the vector of agent actions $\mathbf{a} = (a_1, a_2, ...)$, as the distributions of the outputs depend on the actions.

We will use $\mathcal{F}^\infty$ to denote the σ-algebra generated by the full vector of outputs $S_1, S_2, ...$ Note that by the LLN all latent types may be taken to be measurable wrt $\mathcal{F}^\infty$. Finally, for each population size N, we will let $\mathcal{P}^{*N}$ denote the restriction of the measure $\mathcal{P}^{\mathbf{a}^*(N)}$ to $(\Omega, \mathcal{F}^\infty)$, and similarly let $\mathcal{P}^{\Delta,N}$ denote the restriction of the measure $\mathcal{P}^{(a^*(N)+\Delta, \mathbf{a}^*(N))}$ to $(\Omega, \mathcal{F}^\infty)$. These measures represent the distributions over outputs induced when all agents take actions $a^*(N)$ and when agent 1 deviates to action $a^*(N) + \Delta$, respectively.

Lemma B.1. *The Radon-Nikodym derivative for the change of measure from* $(\Omega, \mathcal{F}^\infty, \mathcal{P}^{*N})$ *to* $(\Omega, \mathcal{F}^\infty, \mathcal{P}^{\Delta,N})$ *is*

$$\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} = \frac{g_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))}$$

in the common type model and

$$\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} = \frac{g_1^C(S_1 \mid \eta; a_1 = a^*(N) + \Delta)}{g_1^C(S_1 \mid \eta; a_1 = a^*(N))}$$

in the common confound model.

Proof. We derive the derivative for the common type model, with the expression for the common confound model following from nearly identical work. Fix any $\mathcal{F}^\infty$-measurable random variable X. Then there exists a measurable function $x : \mathbb{R}^\infty \to \mathbb{R}$ such that $X = x(\mathbf{S})$ a.s. Thus

$$\begin{aligned}
&\mathbb{E}[X \mid \mathbf{a} = (a^*(N) + \Delta, \mathbf{a}^*(N))] \\
&= \int dF_\theta(\theta)\, dG_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)\, dG_{-1}^T(\mathbf{S}_{-1} \mid \theta, S_1; \mathbf{a} = (a^*(N) + \Delta, \mathbf{a}^*(N))) \\
&\qquad \times x(\mathbf{S}).
\end{aligned}$$

As $\mathbf{S}_{-1}$ is independent of S_1 conditional on θ in the common type model, $G_{-1}^T(\mathbf{S}_{-1} \mid \theta, S_1; \mathbf{a} = (a^*(N) + \Delta, \mathbf{a}^*(N))) = G_{-1}^T(\mathbf{S}_{-1} \mid \theta; \mathbf{a}_{-1} = \mathbf{a}^*(N))$. So this expression may be equivalently written

$$\begin{aligned}
&\mathbb{E}[X \mid \mathbf{a} = (a^*(N) + \Delta, \mathbf{a}^*(N))] \\
&= \int dF_\theta(\theta)\, dG_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)\, dG_{-1}^T(\mathbf{S}_{-1} \mid \theta; \mathbf{a}_{-1} = \mathbf{a}^*(N)) \\
&\qquad \times x(\mathbf{S}) \\
&= \int dF_\theta(\theta)\, dG_1^T(S_1 \mid \theta; a_1 = a_N^*)\, dG_{-1}^T(\mathbf{S}_{-1} \mid \theta; \mathbf{a}_{-1} = \mathbf{a}^*(N)) \\
&\qquad \times \frac{g_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} x(\mathbf{S}) \\
&= \mathbb{E}\left[\frac{g_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} X \mid \mathbf{a} = \mathbf{a}^*(N)\right].
\end{aligned}$$

As this argument holds for arbitrary $\mathcal{F}^\infty$-measurable X, it must be that

$$\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} = \frac{g_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))}.$$

□

To establish the desired limiting result, we will prove that for any Δ and N,

$$|\mu_N(\Delta) - \mu_\infty(\Delta)| \leq \kappa_N(\Delta)\frac{\beta}{\sqrt{N}},$$

where

$$\kappa_N(\Delta) \equiv \left(\mathbb{E}\left[\left(\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} - 1\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right]\right)^{1/2}$$

and β is a finite constant independent of N and Δ whose value depends on the model. The following lemma establishes several important properties of κ_N.

Lemma B.2. *$\kappa_N(\Delta)$ is independent of N, $\kappa_N(0) = 0$, and $\overline{\kappa}'_{N,+}(0) = \limsup_{\Delta\downarrow 0} \kappa_N(\Delta)/\Delta < \infty$.*

Proof. We prove the theorem for the common type model, with nearly identical work establishing the result for the common confound model. Note that when $\Delta = 0$, $d\mathcal{P}^{\Delta,N}/d\mathcal{P}^{*N} = 1$, and so trivially $\kappa_N(0) = 0$. To see that $\kappa_N(\Delta)$ is independent of N, note that the distribution of each output satisfies the translation invariance property $g_i^T(S_i = s_i \mid \theta; a_i = \alpha) = g_i^T(S_i = s_i - \alpha \mid \theta; a_i = 0)$ for any s_i and α. So $\kappa_N(\Delta)$ may be written

$$\begin{aligned}
&\kappa_N(\Delta) \\
&= \int dF_\theta(\theta)\, ds_1\, g_1^T(S_1 = s_1 \mid \theta; a_1 = a^*(N)) \left(\frac{g_1^T(S_1 = s_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 = s_1 \mid \theta; a_1 = a^*(N))} - 1\right)^2 \\
&= \int dF_\theta(\theta)\, ds_1\, g_1^T(S_1 = s_1 - a^*(N) \mid \theta; a_1 = 0) \left(\frac{g_1^T(S_1 = s_1 - a^*(N) \mid \theta; a_1 = \Delta)}{g_1^T(S_1 = s_1 - a^*(N) \mid \theta; a_1 = 0)} - 1\right)^2
\end{aligned}$$

So perform a change of variables to $s_1' \equiv s_1 - a^*(N)$ to obtain the representation

$$\kappa_N(\Delta) = \int dF_\theta(\theta)\, ds_1'\, g_1^T(S_1 = s_1' \mid \theta; a_1 = 0) \left(\frac{g_1^T(S_1 = s_1' \mid \theta; a_1 = \Delta)}{g_1^T(S_1 = s_1' \mid \theta; a_1 = 0)} - 1\right)^2,$$

which is independent of N, as desired.

Next, note that by Bayes' rule

$$\begin{aligned}
&\mathbb{E}\left[\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} \,\middle|\, \mathbf{S}; \mathbf{a} = \mathbf{a}^*(N)\right] \\
&= \mathbb{E}\left[\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} \,\middle|\, S_1, \theta; a_1 = a^*(N)\right] \\
&= \int dF_{\eta_1}^T(\eta_1 \mid S_1, \theta; a_1 = a^*(N))\, \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} \\
&= \int dF_{\eta_1}(\eta_1 \mid \theta; a_1 = a^*(N)) \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))}.
\end{aligned}$$

Now, note that η_1 and θ are independent of each other and of agent 1's action, so $F_{\eta_1}(\eta_1 = t \mid \theta; a_1 = a^*(N)) = F_\eta(\eta_1 = t)$. Thus

$$\begin{aligned}
&\mathbb{E}\left[\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} \,\middle|\, \mathbf{S}; \mathbf{a} = \mathbf{a}^*(N)\right] \\
&= \int dF_\eta(\eta_1) \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} \\
&= \int dF_\eta(\eta_1) \frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} \\
&= \frac{g_1^T(S_1 \mid \theta; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta; a_1 = a^*(N))} \\
&= \frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}}.
\end{aligned}$$

Then $\kappa_N(\Delta)$ may be bounded using Jensen's inequality as

$$\begin{aligned}
\kappa_N(\Delta) &= \left(\mathbb{E}\left[\left(\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} - 1\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right]\right)^{1/2} \\
&= \left(\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} - 1\right) \,\middle|\, \mathbf{S}; \mathbf{a} = \mathbf{a}^*(N)\right]\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right]\right)^{1/2} \\
&\leq \left(\mathbb{E}\left[\left(\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} - 1\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right]\right)^{1/2}.
\end{aligned}$$

So define

$$\widetilde{\kappa}_N(\Delta) \equiv \left(\mathbb{E}\left[\left(\frac{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} - 1\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right]\right)^{1/2}.$$

Since $0 \leq \kappa_N(\Delta)/\Delta \leq \widetilde{\kappa}_N(\Delta)/\Delta$ for all $\Delta > 0$, it is enough to show that $\limsup_{\Delta \downarrow 0} \widetilde{\kappa}_N(\Delta)/\Delta$ exists and is finite.

Using the fact that θ_1, η_1, and a_1 each affect the output as an additive shift, $g_1^T(S_1 = s_1 \mid \theta = t, \eta_1 = t'; a_1) = f_\varepsilon(s_1 - t - t' - a_1)$ for any (s_1, t, t', a_1). Thus given that $S_1 = a^*(N) + \theta + \eta_1 + \varepsilon_1$ under $\mathcal{P}^{*,N}$, $g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N)) = f_\varepsilon(\varepsilon_1)$ and similarly $g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N) + \Delta) = f_\varepsilon(\varepsilon_1 - \Delta.)$ So

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{g_1^T(S_1 \mid \theta = t, \eta_1 = t'; a_1 = a^*(N) + \Delta)}{g_1^T(S_1 \mid \theta, \eta_1; a_1 = a^*(N))} - 1\right)^2 \,\middle|\, \mathbf{a} = \mathbf{a}^*(N)\right] \\
&= \int dF_\varepsilon(\varepsilon_1) \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)}\right)^2.
\end{aligned}$$

We must therefore show that the limit

$$\limsup_{\Delta \downarrow 0} \frac{1}{\Delta} \left(\int dF_\varepsilon(\varepsilon_1) \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)} \right)^2 \right)^{1/2}$$
$$= \left(\limsup_{\Delta \downarrow 0} \int dF_\varepsilon(\varepsilon_1) \frac{1}{\Delta^2} \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)} \right)^2 \right)^{1/2}$$

exists and is finite. By Assumption 3, for Δ sufficiently close to 0 there exists a non-negative, integrable function $J(\cdot)$ such that

$$\frac{1}{\Delta^2} \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)} \right)^2 \leq J(\varepsilon_1)$$

for all ε_1. Then by reverse Fatou's lemma,

$$\begin{aligned} \limsup_{\Delta \downarrow 0} \int dF_\varepsilon(\varepsilon_1) \frac{1}{\Delta^2} \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)} \right)^2 &\leq \int dF_\varepsilon(\varepsilon_1) \limsup_{\Delta \downarrow 0} \frac{1}{\Delta^2} \left(\frac{f_\varepsilon(\varepsilon_1 - \Delta) - f_\varepsilon(\varepsilon_1)}{f_\varepsilon(\varepsilon_1)} \right)^2 \\ &\leq \int dF_\varepsilon(\varepsilon_1)\, J(\varepsilon_1) < \infty, \end{aligned}$$

as desired. □

The bound on $|\mu_N(\Delta) - \mu_\infty(\Delta)|$ just claimed completes the proof because for $\Delta > 0$ it may be rewritten

$$|(\mu_N(\Delta) - \mu)/\Delta - (\mu_\infty(\Delta) - \mu)/\Delta| \leq \frac{\kappa_N(\Delta) - \kappa_N(0)}{\Delta} \frac{\beta}{\sqrt{N}},$$

and thus by taking $\Delta \downarrow 0$ the inequality

$$|\mu_N'(0) - \mu_\infty'(0)| \leq \overline{\kappa}_{N,+}'(0) \frac{\beta}{N}$$

must hold. Then as $\overline{\kappa}_{N,+}'(0)$ is finite and independent of N, $\mu_N'(0) \to \mu_\infty'(0)$ as $N \to \infty$, as desired.

We first derive the bound for the common type model. To streamline notation, we will write $\mathbb{E}^{*N}$ to represent expectations conditioning on $\mathbf{a} = \mathbf{a}^*(N)$, and $\mathbb{E}^{\Delta,N}$ to represent expectations conditioning on $a_1 = a^*(N) + \Delta$ and $\mathbf{a}_{-1} = \mathbf{a}^*(N)$. In the common type model, $\mu_\infty(\Delta) = \mu$ for any Δ, so

$$\begin{aligned} |\mu_N(\Delta) - \mu_\infty(\Delta)| &= |\mu_N(\Delta) - \mu| \\ &= |\mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}]] - \mu| \\ &= |\mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta]|, \end{aligned}$$

where we have used the fact that $\mathbb{E}^{\Delta,N}[\theta] = \mu$ for any Δ. Now, performing a change of measure,

$$\begin{aligned}
&\mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta] \\
&= \mathbb{E}^{*N}\left[\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}}\left(\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta\right)\right] \\
&= \mathbb{E}^{*N}\left[\left(\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} - 1\right)\left(\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta\right)\right] + \mathbb{E}^{*N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta] \\
&= \mathbb{E}^{*N}\left[\left(\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} - 1\right)\left(\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta\right)\right],
\end{aligned}$$

with the last equality following from the law of iterated expectations. The Cauchy-Schwarz inequality then implies the bound

$$\left|\mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta]\right| \leq \kappa_N(\Delta)\left(\mathbb{E}^{*N}\left[\left(\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta\right)^2\right]\right)^{1/2}.$$

Now, the posterior expectation minimizes mean-squared loss among all estimators of θ, so

$$\begin{aligned}
\mathbb{E}^{*N}\left[\left(\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta\right)^2\right] &\leq \mathbb{E}^{*N}\left[\left(\frac{1}{N}\sum_{i=1}^{N} S_i - \theta\right)^2\right] \\
&= \mathbb{E}^{*N}\left[\mathbb{E}^{*N}\left[\left(\frac{1}{N}\sum_{i=1}^{N} S_i - \theta\right)^2 \middle| \theta\right]\right] \\
&= \frac{\sigma_\eta^2 + \sigma_\varepsilon^2}{N},
\end{aligned}$$

with the last equality following from the fact that the S_i are independent conditional on θ and each has conditional mean θ and conditional variance $\sigma_\varepsilon^2 + \sigma_\eta^2$. Thus

$$\left|\mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta \mid \mathbf{S}_{1:N}] - \theta]\right| \leq \kappa_N(\Delta)\frac{\sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}}{\sqrt{N}},$$

which was the desired bound with $\beta = \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2}$.

We now derive the bound for the common confound model. Note first that the expected value of the principal's posterior estimate of θ_1 is a function only of the size of agent 1's distortion Δ, but *not* of the equilibrium action inference. Thus

$$\begin{aligned}
\mu_\infty(\Delta) &= \mathbb{E}[\mathbb{E}[\theta_1 \mid \mathbf{S}; \mathbf{a} = \mathbf{a}^*(\infty)] \mid \mathbf{a} = (a^*(\infty) + \Delta, \mathbf{a}^*(\infty))] \\
&= \mathbb{E}[\mathbb{E}[\theta_1 \mid \mathbf{S}; \mathbf{a} = \mathbf{a}^*(N)] \mid \mathbf{a} = (a^*(N) + \Delta, \mathbf{a}^*(N))] \\
&= \mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]].
\end{aligned}$$

So we may write

$$\mu_N(\Delta) - \mu_\infty(\Delta) = \mathbb{E}^{\Delta,N}[\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]].$$

Now, changing measure in a manner analogous to the common type case yields

$$\mu_N(\Delta) - \mu_\infty(\Delta) = \mathbb{E}^{*N}\left[\left(\frac{d\mathcal{P}^{\Delta,N}}{d\mathcal{P}^{*N}} - 1\right)\left(\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]\right)\right].$$

Then by an application of the Cauchy-Schwarz inequality,

$$|\mu_N(\Delta) - \mu_\infty(\Delta)| \le \kappa_N(\Delta)\left(\mathbb{E}^{*N}\left[\left(\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]\right)^2\right]\right)^{1/2}.$$

To bound the rhs, define the family of random variables $\widehat{\theta}_N(t) \equiv \mathbb{E}^{*N}[\theta_1 \mid S_1, \eta = t]$ for $t \in \mathbb{R}$. Note that $\widehat{\theta}_1(\eta) = \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]$, as $\mathbf{S}$ allows the principal to perfectly infer η, and the vector of outputs $\mathbf{S}_{-1}$ is independent of θ_1 conditional on η. Further, $\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] = \mathbb{E}^{*N}[\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}] \mid \mathbf{S}_{1:N}]$ is the mean-square minimizing estimator of $\widehat{\theta}_N(\eta)$ conditional on the performance vector $\mathbf{S}_{1:N}$. Another estimator of $\widehat{\theta}_N(\eta)$ is $\widehat{\theta}_N(\overline{\eta}_N)$, where $\overline{\eta}_N \equiv \frac{1}{N}\sum_{i=1}^N (S_i - \mu)$. So

$$\mathbb{E}^{*N}\left[\left(\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]\right)^2\right] \le \mathbb{E}^{*N}\left[\left(\widehat{\theta}_N(\overline{\eta}_N) - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]\right)^2\right].$$

Given that shifts in η affect the output additively, $\mathbb{E}^{*N}[\theta_1 \mid S_1 = s_1, \eta = t] = \mathbb{E}^{*N}[\theta_1 \mid S_1 = s_1 - t, \eta = 0]$ for every s_1 and t. Now recall that Assumption 2 implies that $\widehat{\theta}_N(t)$ is differentiable everywhere, while Lemma A.3 implies that $\widehat{\theta}'_N(t) \in [-1, 0]$. Thus by the fundamental theorem of calculus,

$$|\widehat{\theta}_N(\overline{\eta}_N) - \widehat{\theta}_N(\eta)| = \left|\int_\eta^{\overline{\eta}_N} \widehat{\theta}'_N(t)\, dt\right| \le \int_\eta^{\overline{\eta}_N} |\widehat{\theta}'_N(t)|\, dt \le |\overline{\eta}_N - \eta|.$$

Further note that

$$\overline{\eta}_N - \eta = \frac{1}{N}\sum_{i=1}^N (\theta_i - \mu + \varepsilon_i),$$

which has mean 0 and variance $(\sigma_\theta^2 + \sigma_\varepsilon^2)/N$ given that the θ_i and ε_i are mutually independent. So

$$\mathbb{E}^{*N}\left[\left(\mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}_{1:N}] - \mathbb{E}^{*N}[\theta_1 \mid \mathbf{S}]\right)^2\right] \le \frac{\sigma_\theta^2 + \sigma_\varepsilon^2}{N},$$

implying the desired bound with $\beta = \sqrt{\sigma_\theta^2 + \sigma_\varepsilon^2}$.

C Proofs for Section 4 (Main Results)

C.1 Proofs of Theorems 1 and 2

Opt-In Equilibrium. In any pure-strategy equilibrium in which all agents opt-in, the equilibrium effort level a^* must satisfy two conditions:

$$MV(N) = C'(a^*) \tag{C.1}$$

$$t_1 - C(a^*) \geq 0 \tag{C.2}$$

The expression in (C.1) guarantees that an agent who opts-in cannot strictly gain by deviating to a different effort choice. This is identical to the condition used in the exogenous entry model to solve for equilibrium. The expression in (C.2) guarantees that agents cannot profitably deviate to opting-out.

The marginal value $MV(N)$ is independent of a^*, and C' is strictly monotone. Thus (C.1) pins down a unique effort level $a^* = C'^{-1}(MV(N))$. Since C is everywhere increasing, the conditions in (C.1) and (C.2) can be simultaneously satisfied if and only if $0 \leq C'^{-1}[MV(N)] \leq a^{**} \equiv C^{-1}(t_1)$, or equivalently,

$$0 = C'(0) \leq MV(N) \leq C'(a^{**})$$

noting that C'^{-1} is everywhere increasing.

By Assumption 6, $t_1 > C(a^*(1))$. Since the cost function C has positive first and second derivatives, $t_1 > C(a^*(1))$ and $t_1 = C(a^{**})$ imply that $a^*(1) < a^{**}$, which further implies $C'(a^*(1)) < C'(a^{**})$. By Lemma 1, $MV(1) = MV_T(1) \geq MV_T(N)$. Thus

$$MV_T(N) \leq MV_T(1) = C'(a^*(1)) \leq C'(a^{**}),$$

and a symmetric all opt-in equilibrium exists in the common type model. In contrast, in the common confound model,

$$MV_C(N) \geq MV_C(1) = C'(a^*(1)) \tag{C.3}$$

so the inequality $MV_C(N) \leq C'(a^{**})$ is not guaranteed to hold. An opt-in equilibrium exists if and only if N is sufficiently small; specifically, $N \leq N^*$ where

$$N^* \equiv \sup\{N \ : \ MV_C(N) \leq C'(a^{**})\}.$$

(It is possible that N^* is infinite if $MV_C(N) \leq C'(a^{**})$ for all N.)

Finally, for the parameters $N \leq N^*$ where an opt-in equilibrium exists in both models, it is possible to rank equilibrium effort levels as follows: Define a_C^* and a_T^* to be the respective equilibrium effort levels. Then, since $MV_C(N) \geq MV(1) \geq MV_T(N)$ for all N,

$$a_C^* = C'^{-1}(MV_C(N)) \geq C^{-1}(MV_T(N)) = a_T^*$$

so equilibrium effort is higher in the common confound model.

Opt-Out Equilibrium. Under the imposed refinement on the principal's off-equilibrium belief about the agent's action, the optimal action conditional on entry is $a^*(1)$. Thus in an all opt-out equilibrium, the equilibrium action a^* must satisfy

$$t_1 - C(a^*(1)) < 0 \tag{C.4}$$

which violates Assumption 6. There are no pure-strategy equilibria in either model in which all agents choose to opt-out.

Mixed Equilibrium. For any probability $p \in [0, 1]$ and $M \in \{T, C\}$, let

$$MV_M(p, N) = \mathbb{E}[(MV_M(\widetilde{N} + 1) \mid \widetilde{N} \sim \text{Binomial}(N - 1, p)]$$

be the expected marginal impact for agent i of exerting additional effort beyond the principal's expectation, when agent i opts-in and all other agents opt-in with independent probability p. Note that because $MV_C(N)$ is increasing in N, and increasing p shifts up the distribution of $\widetilde{N}$ in the FOSD sense, $MV_C(p, N)$ is increasing in p. Further, because increasing p shifts $\Pr(\widetilde{N} \leq n)$ strictly downward for every $n < N - 1$, this monotonicity is strict whenever $MV_C(n)$ is not constant over the range $\{1, .., N\}$. For the same reasons, $MV_C(p, N)$ is increasing in N for fixed p, and strictly increasing whenever $p \in (0, 1)$ and $MV_C(n)$ is not constant over $\{1, ..., N\}$.

In a mixed equilibrium, the equilibrium effort level a^* and probability p assigned to opting-in must jointly satisfy

$$t_1 - C(a^*) = 0. \tag{C.5}$$

$$MV(p, N) = C'(a^*). \tag{C.6}$$

The expression in (C.5) pins down the equilibrium action, which is identical to the action defined as a^{**} above. Moreover, $C'(a)$ is independent of both the mixing probability p and also the fixed population size N. Therefore an equilibrium exists if and only if $MV(p, N) = C'(a^{**})$ for some $p \in [0, 1]$. But for all $p \in [0, 1]$,

$$MV_T(p, N) \leq \max_{1 \leq N' \leq N} MV_T(N') = MV_T(1) = C'(a^*(1)) < C'(a^{**})$$

using that MV_T is a decreasing function of N (Lemma 1). Thus the common type model does not admit a strictly mixed equilibrium.

Similarly if $MV_C(N) < C'(a^{**})$, then

$$MV_C(p, N) \leq \max_{1 \leq N' \leq N} MV_C(N') = MV_C(N) < C'(a^{**})$$

since MV_C is a strictly increasing function of N (Lemma 1). So there does not exist a strictly mixed equilibrium in the common confound model either. Indeed, this is exactly the range for N that supports the symmetric all opt-in equilibrium in the common confound model.

If however $MV(N) \geq C'(a^{**})$, then

$$MV_C(1) = MV_C(0, N) < C'(a^{**}) \leq MV_C(1, N) = MV_C(N).$$

This implies in particular that MV_C is not constant over the range $\{1, ..., N\}$, so that $MV_C(p, N)$ is strictly increasing in p. Since $MV_C(p, N)$ is also continuous in p, the intermediate value theorem yields existence of a unique $p^*(N) \in (0, 1]$ satisfying $MV_C(p^*(N), N) = C'(a^{**})$.

If $N \leq N^*$, i.e. $MV(N) = C'(a^{**})$, then it must be that $p^*(N) = 1$. Thus in particular the opt-in equilibrium is unique whenever it exists. Otherwise $p^*(N) < 1$, in which case the fact that $MV_C(p, N)$ is strictly increasing in N for fixed $p \in (0, 1)$ further implies that $p^*(N)$ must be strictly decreasing in N. Finally, the effort level a^{**} chosen in this equilibrium weakly exceeds the effort level a^*_C chosen in the symmetric opt-in equilibrium in the common confound model, since $t_1 \geq C(a^*_C)$ by (C.2), while $t_1 = C(a^{**})$ by (C.5).

C.2 Proof of Corollary 2

Suppose all agents in a population of size N enter and choose action a. Social welfare

$$W(1, a, N) = N \cdot (2\mu + t_1 + a - C(a))$$

is strictly increasing on $a \in [0, a_{FB})$. Thus the comparison $a^*_T(N) \leq a_{NS} < a_{FB}$ immediately implies that for all N, welfare is ranked

$$W_T(N) \leq W_{NS}(N) < W_{FB}$$

where the first inequality is strict for all $N > 1$.

For population sizes $N < N^*$, the equilibrium action in the common confound model satisfies $a^*_C(N) \in [a_{NS}, a_{FB})$ (Theorem 2), so similar arguments imply

$$W_T(N) \leq W_{NS}(N) \leq W^*_C(N) < W_{FB}$$

where again the inequalities are strict for $N > 1$. When the population size $N > N^*$,

$$W_C^*(N) = N \cdot [p(N) \cdot (2\mu + t_1 + a^{**} - C(a^{**}))] \equiv N \cdot f_C(N)$$

while

$$W_T^*(N) = N \cdot [2\mu + t_1 + a_T^*(N) - C(a_T^*(N))] \equiv N \cdot f_T(N)$$

Since $p(N) \to 0$ as $N \to \infty$ (Theorem 2), it follows that $f_C(N) \to 0$ while $f_T(N) \to 2\mu > 0$. So for N sufficiently large,

$$W_C^*(N) < W_T(N) < W_{NS}(N) < W_{FB}$$

Finally, since each of W_C, W_T, W_{NS} (extended to all of $\mathbb{R}$) are continuous functions of N, by the intermediate value theorem, there exists N such that

$$W_T(N) < W_C^*(N) < W_{NS}(N) < W_{FB}$$

This concludes the proof.

D Proofs for Section 5 (Gaussian)

D.1 Verification of Assumptions in 2.2

Here we verify that Gaussian uncertainty satisfies the stated assumptions. Assumptions 1, 2, and 4 are immediate. Assumption 5 is satisfied for any strictly convex cost function, since the second derivative of the posterior expectation in each signal realization is zero. Assumption 3 is verified in the following lemma:

Lemma D.1. *Suppose $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Then for any $\overline{\Delta} > 0$, the function*

$$J^*(\varepsilon) = \frac{1}{\overline{\Delta}^2}\left(\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right) + \exp\left(\frac{\overline{\Delta}|\varepsilon|}{\sigma^2}\right) - 2\right)^2$$

is $\mathcal{P}^0$-integrable and satisfies $|J(\varepsilon, \Delta)| \leq J^*(\varepsilon)$ for every $\varepsilon \in \mathbb{R}$ and $\Delta \in [-\overline{\Delta}, \overline{\Delta}]$.

Proof. Under the distributional assumption on ε, the density function f_ε has the form

$$f_\varepsilon(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right).$$

Therefore

$$\frac{1}{\Delta}\frac{f_\varepsilon(\varepsilon - \Delta) - f_\varepsilon(\varepsilon)}{f_\varepsilon(\varepsilon)} = \frac{\exp\left(\frac{1}{\sigma^2}\Delta(\varepsilon - \Delta/2)\right) - 1}{\Delta}.$$

Now, we may equivalently write

$$\begin{aligned}\frac{1}{\Delta}\frac{f_\varepsilon(\varepsilon-\Delta)-f_\varepsilon(\varepsilon)}{f_\varepsilon(\varepsilon)} &= \frac{1}{\sigma^2}\int_{\Delta/2}^{\varepsilon}\exp\left(\frac{1}{\sigma^2}\Delta(\widetilde{\varepsilon}-\Delta/2)\right)\,d\widetilde{\varepsilon} \\ &= \frac{\exp\left(-\frac{\Delta^2}{2\sigma^2}\right)}{\sigma^2}\int_{\Delta/2}^{\varepsilon}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon}.\end{aligned}$$

Hence

$$\begin{aligned}\left|\frac{1}{\Delta}\frac{f_\varepsilon(\varepsilon-\Delta)-f_\varepsilon(\varepsilon)}{f_\varepsilon(\varepsilon)}\right| &= \frac{\exp\left(-\frac{\Delta^2}{2\sigma^2}\right)}{\sigma^2}\int_{\min\{\Delta/2,\varepsilon\}}^{\max\{\Delta/2,\varepsilon\}}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon} \\ &\le \frac{1}{\sigma^2}\int_{\min\{\Delta/2,\varepsilon\}}^{\max\{\Delta/2,\varepsilon\}}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon}.\end{aligned}$$

Let

$$H(\varepsilon,\Delta)\equiv\frac{1}{\sigma^2}\int_{\min\{\Delta/2,\varepsilon\}}^{\max\{\Delta/2,\varepsilon\}}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon}.$$

We will show that $H(\varepsilon,\Delta)\le\sqrt{J^*(\varepsilon)}$ for all ε and $\Delta\in[-\overline{\Delta},\overline{\Delta}]$ in cases, depending on the signs of ε,Δ, and $\varepsilon-\Delta/2$.

Case 1: $\varepsilon\ge\Delta/2\ge0$. Then

$$\begin{aligned}H(\varepsilon,\Delta) &= \frac{1}{\sigma^2}\int_{\Delta/2}^{\varepsilon}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon} \\ &\le \frac{1}{\sigma^2}\int_{0}^{\varepsilon}\exp\left(\frac{\overline{\Delta}\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon} \\ &= \frac{1}{\overline{\Delta}}\left(\exp\left(\frac{\overline{\Delta}\varepsilon}{\sigma^2}\right)-1\right)\le\sqrt{J^*(\varepsilon)}.\end{aligned}$$

Case 2: $\varepsilon\ge0>\Delta/2$. Then

$$\begin{aligned}H(\varepsilon,\Delta) &= \frac{1}{\sigma^2}\int_{\Delta/2}^{\varepsilon}\exp\left(\frac{\Delta\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon} \\ &\le \frac{1}{\sigma^2}\left(\int_{0}^{\varepsilon}\exp\left(\frac{\overline{\Delta}\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon}+\int_{-\overline{\Delta}/2}^{0}\exp\left(-\frac{\overline{\Delta}\widetilde{\varepsilon}}{\sigma^2}\right)\,d\widetilde{\varepsilon}\right) \\ &= \frac{1}{\overline{\Delta}}\left(\exp\left(\frac{\overline{\Delta}\varepsilon}{\sigma^2}\right)+\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right)-2\right) \\ &= \sqrt{J^*(\varepsilon)}.\end{aligned}$$

Case 3: $\Delta/2 > \varepsilon \geq 0$. Then

$$\begin{aligned} H(\varepsilon, \Delta) &= \frac{1}{\sigma^2} \int_{\varepsilon}^{\Delta/2} \exp\left(\frac{\Delta \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &\leq \frac{1}{\sigma^2} \int_{0}^{\overline{\Delta}/2} \exp\left(\frac{\overline{\Delta} \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &= \frac{1}{\overline{\Delta}} \left(\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right) - 1 \right) \\ &\leq \sqrt{J^*(\varepsilon)}. \end{aligned}$$

Case 4: $\Delta/2 > 0 > \varepsilon$. Then

$$\begin{aligned} H(\varepsilon, \Delta) &= \frac{1}{\sigma^2} \int_{\varepsilon}^{\Delta/2} \exp\left(\frac{\Delta \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &\leq \frac{1}{\sigma^2} \left(\int_{0}^{\overline{\Delta}/2} \exp\left(\frac{\overline{\Delta} \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} + \int_{\varepsilon}^{0} \exp\left(-\frac{\overline{\Delta} \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \right) \\ &= \frac{1}{\overline{\Delta}} \left(\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right) + \exp\left(\frac{\overline{\Delta}|\varepsilon|}{\sigma^2}\right) - 2 \right) \\ &= \sqrt{J^*(\varepsilon)}. \end{aligned}$$

Case 5: $0 \geq \Delta/2 > \varepsilon$. Then

$$\begin{aligned} H(\varepsilon, \Delta) &= \frac{1}{\sigma^2} \int_{\varepsilon}^{\Delta/2} \exp\left(\frac{\Delta \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &\leq \frac{1}{\sigma^2} \int_{\varepsilon}^{0} \exp\left(-\frac{\overline{\Delta} \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &= \frac{1}{\overline{\Delta}} \left(\exp\left(\frac{\overline{\Delta}|\varepsilon|}{\sigma^2}\right) - 1 \right) \leq \sqrt{J^*(\varepsilon)}. \end{aligned}$$

Case 6: $0 > \varepsilon \geq \Delta/2$. Then

$$\begin{aligned} H(\varepsilon, \Delta) &= \frac{1}{\sigma^2} \int_{\Delta/2}^{\varepsilon} \exp\left(\frac{\Delta \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &\leq \frac{1}{\sigma^2} \int_{-\overline{\Delta}/2}^{0} \exp\left(-\frac{\overline{\Delta} \widetilde{\varepsilon}}{\sigma^2}\right) d\widetilde{\varepsilon} \\ &= \frac{1}{\overline{\Delta}} \left(\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right) - 1 \right) \leq \sqrt{J^*(\varepsilon)}. \end{aligned}$$

This establishes that $|J(\varepsilon, \Delta)| \leq H(\varepsilon, \Delta)^2 \leq J^*(\varepsilon)$ for every ε and $\Delta \in [-\overline{\Delta}, \overline{\Delta}]$, as desired. It remains only to show that J^* is $\mathcal{P}^0$-integrable. This follows because

$$\begin{aligned} J^*(\varepsilon) &\leq \frac{1}{\overline{\Delta}^2}\left(\exp\left(\frac{\overline{\Delta}^2}{2\sigma^2}\right) + \exp\left(\frac{\overline{\Delta}|\varepsilon|}{\sigma^2}\right)\right)^2 \\ &= \frac{1}{\overline{\Delta}^2}\left(\exp\left(\frac{\overline{\Delta}^2}{\sigma^2}\right) + 2\exp\left(\frac{\overline{\Delta}^2}{\sigma^2}\right)\exp\left(\frac{\overline{\Delta}|\varepsilon|}{\sigma^2}\right) + \exp\left(\frac{2\overline{\Delta}|\varepsilon|}{\sigma^2}\right)\right) \\ &= \frac{1}{\overline{\Delta}^2}\left(\exp\left(\frac{\overline{\Delta}^2}{\sigma^2}\right) + 2\exp\left(\frac{\overline{\Delta}^2}{\sigma^2}\right)\left(\exp\left(\frac{\overline{\Delta}\varepsilon}{\sigma^2}\right) + \exp\left(-\frac{\overline{\Delta}\varepsilon}{\sigma^2}\right)\right)\right. \\ &\qquad \left. + \exp\left(\frac{2\overline{\Delta}\varepsilon}{\sigma^2}\right) + \exp\left(-\frac{2\overline{\Delta}\varepsilon}{\sigma^2}\right)\right) \end{aligned}$$

The first term is a constant, while each of the remaining terms is proportional to a lognormal random variable. Thus each term has finite mean, and hence so does $J^*(\varepsilon)$. □

D.2 Proof of Corollary 3

Consider first the common type model, and suppose that agent i chooses effort $a_i = a^* + \Delta$ while all agents $j \neq i$ choose the equilibrium effort level a^*. Then from standard formulas for Bayesian updating to normal signals, the agent's expectation of the principal's posterior expectation of θ_i is

$$\begin{aligned} \mu_N(\Delta) =& \mathbb{E}\left[\frac{\sigma_\varepsilon^2 + \sigma_\eta^2}{N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2}\mu + \frac{N\sigma_\theta^2}{N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2}(\bar{S} + \Delta/N)\right] \\ &= \mu + \frac{\sigma_\theta^2}{N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2} \cdot \Delta \end{aligned}$$

where $\bar{S} + \Delta/N = (1/N)\sum_{j=1}^N (S_j - a^*)$. So the marginal value of effort is

$$\mu_N'(\Delta) = \frac{\sigma_\theta^2}{N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2}. \tag{D.1}$$

Note that this expression is constant in Δ. The equilibrium effort level a^* must satisfy implying $a^* = \sigma_\theta^2/(N\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)]$ as desired.

Consider now the common confound model with the same player actions described above. The principal's posterior belief about θ_i is independent of $\mathbf{S}_{-i}$ (the vector of outputs excluding S_i) conditional on η. Thus we can first update the principal's belief about the confound η to

$$\eta \mid \mathbf{S}_{-i} \sim \mathcal{N}\left(\frac{(N-1)\sigma_\eta^2}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2}\bar{S}_{-i}, \frac{\sigma_\eta^2(\sigma_\varepsilon^2 + \sigma_\theta^2)}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2}\right) \equiv \mathcal{N}(\tilde{\mu}_\eta, \tilde{\sigma}_\eta^2)$$

where $\bar{S}_{-i} = \frac{1}{N}\sum_{j\neq i}(S_j - a^*)$. Then, $(\theta_i, S_i - a_i)$ are jointly distributed

$$\begin{pmatrix} \theta_i \\ \theta_i + \eta + \varepsilon_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu + \tilde{\mu}_\eta \end{pmatrix}, \begin{pmatrix} \sigma_\theta^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \tilde{\sigma}_\eta^2 + \sigma_\varepsilon^2 \end{pmatrix} \right)$$

so the principal's posterior expectation of θ_i is

$$\begin{aligned} \mu_N(\Delta) &= \mathbb{E}\left[\left(\mu + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \tilde{\sigma}_\eta^2 + \sigma_\varepsilon^2}(S_i + \Delta - (\mu + \tilde{\mu}_\eta))\right)\right] \\ &= \frac{\sigma_\varepsilon^2 + \sigma_\theta^2}{(N-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} \cdot \mu + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \widetilde{\sigma}_\eta^2 + \sigma_\varepsilon^2} \cdot \Delta \end{aligned}$$

and $\mu_N'(\Delta) = \sigma_\theta^2/(\sigma_\theta^2 + \tilde{\sigma}_\eta^2 + \sigma_\varepsilon^2)$, which is again constant in Δ. The equilibrium effort level a^* must satisfy

$$C'(a_C^*) = \mu_N'(0) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \tilde{\sigma}_\eta^2 + \sigma_\varepsilon^2}$$

implying $a_C^* = \sigma_\theta^2/(\sigma_\theta^2 + \tilde{\sigma}_\eta^2 + \sigma_\varepsilon^2)$ as desired.

The threshold N^* must satisfy $\mu_N'(0) = C'(a^{**})$ where $a^{**} = C^{-1}(t_1) = \sqrt{2t_1}$. Thus we have

$$\sigma_\theta^2 / \left[\left(\sigma_\theta^2 + \frac{\sigma_\eta^2(\sigma_\varepsilon^2 + \sigma_\theta^2)}{(N^*-1)\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} + \sigma_\varepsilon^2\right)\right] = 2\sqrt{\mu}$$

which yields the expression given in (6).

D.3 Proof of Proposition 2

Parts (b) and (c) follow directly from concavity of $MV_C(N)$ (Lemma 2), since

$$MV(p(N)\cdot N) > \mathbb{E}[MV(\widetilde{N} \sim \text{Bin}(N, p(N)))] = MV(p^*(N)\cdot N)$$

and $MV(N)$ is increasing in its argument.

We now show Part (a), which says that the expected number of entrants $p(N)\cdot N$ is increasing in N. With N total agents, the number of entrants in the unique symmetric equilibrium is distributed as $\text{Bin}(N, p(N))$. Fix N, and let $q \equiv p(N)$. Consider increasing the number of agents to $N' > N$, and let $q' \equiv p(N)N/N'$, so that $\text{Bin}(N, q)$ and $\text{Bin}(N', q')$ have the same mean. By Lemma 2, $MV_C(N)$ is strictly concave in N. So let $\widetilde{N} \sim \text{Bin}(N, q)$ and $\widetilde{N}' \sim \text{Bin}(N', q')$. If $\widetilde{N}'$ is a mean-preserving spread of $\widetilde{N}$, then $\mathbb{E}[MV_C(\widetilde{N}')] < \mathbb{E}[MV_C(\widetilde{N})]$, meaning that

$$MV_C(p(N'), N') = C'(a^{**}) = \mathbb{E}[MV_C(\widetilde{N})] > \mathbb{E}[MV_C(\widetilde{N}')] = MV_C(q', N')$$

and therefore $p(N') > q'$ given that $MV_C(\cdot, N)$ is a strictly increasing function. Then since $q'N' = qN$, the mean number of entrants $p(N')N'$ with N' agents is strictly larger than the mean number $p(N)N$ with N agents. So the theorem is proven once we establish that $\widetilde{N}$ is a mean-preserving spread of N.

Because the mean-preserving spread property is transitive, it is sufficient to prove that for every positive integer N and $q \in (0,1)$, the random variable $\widetilde{X} \sim \text{Bin}(N+1, q')$ with $q' \equiv qN/(N+1)$ is a mean-preserving spread of the random variable $X \sim \text{Bin}(N, q)$. Define the CDFs F and $\widetilde{F}$ for X and $\widetilde{X}$, respectively, satisfying

$$F(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{N}{i} q^i (1-q)^{N-i}, & 0 \leq x \leq N \\ 0, & x < 0 \\ 1, & x > N \end{cases}$$

and

$$\widetilde{F}(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{N+1}{i} (q')^i (1-q')^{N+1-i}, & 0 \leq x \leq N+1 \\ 0, & x < 0 \\ 1, & x > N+1 \end{cases}$$

Also define

$$\Gamma(x) \equiv \int_{-\infty}^{t} (\widetilde{F}(t) - F(t))\, dt.$$

In order that $\widetilde{X}$ be a mean-preserving spread of X it must be that $\Gamma(x) \geq 0$ for all x. For $x \leq 0$ trivially $\Gamma(x) = 0$, and meanwhile for any distribution function G with mean μ and support contained in $(-\infty, \overline{x}]$,

$$\int_{-\infty}^{\overline{x}} G(t)\, dt = \int_{-\infty}^{\overline{x}} \int_{-\infty}^{t} dG(s)\, dt = \int_{-\infty}^{\overline{x}} dG(s) \int_{s}^{\overline{x}} dt = \int_{-\infty}^{\overline{x}} (\overline{x} - s)\, dG(s) = \overline{x} - \mu.$$

Hence $\Gamma(x) = 0$ for all $x \geq N+1$ as well. It therefore remains only to establish the result for $x \in (0, N+1)$.

Differentiating Γ yields

$$\Gamma'(x) = \widetilde{F}(x) - F(x).$$

Suppose that Γ' satisfies single-crossing on $[0, N+1]$; that is, there exists an $x_0 \in (0, N+1)$ such that $\Gamma'(x) > 0$ for $x < x_0$ and $\Gamma'(x) < 0$ for $x > x_0$. Then Γ is single-peaked on $[0, N+1]$, and given that $\Gamma(0) = \Gamma(N+1) = 0$, it must therefore be that $\Gamma(x) > 0$ for all $x \in (0, N+1)$. We complete the proof by establishing single-crossing.

First note that for $x \in [N, N+1)$, $\Gamma'(x) = \widetilde{F}(N) - 1 < 0$. So it is sufficient to establish single-crossing on $[0, N]$. As a preliminary step, note that

$$\Gamma'(0) = (1-q)^N - (1-q')^{N+1} = \xi(q)^{Nq} - \xi(q')^{(N+1)q'} = \xi(q)^{Nq} - \xi(q')^{Nq},$$

where $\xi(x) \equiv (1 + 1/x)^x$. This function is well-known to be strictly increasing in x, so $\Gamma'(0) > 0$. Then as $\Gamma'(N) < 0$, Γ' must cross zero an odd number of times. We complete the proof by showing that Γ' can cross zero at most twice, establishing that it must cross exactly once, as desired.

To show this final property, differentiate Γ' to obtain

$$\Gamma''(x) = \widetilde{f}(\lfloor x \rfloor) - f(\lfloor x \rfloor),$$

where

$$f(x) = \binom{N}{x} q^x (1-q)^{N-x}$$

and

$$\widetilde{f}(x) = \binom{N+1}{x} (q')^x (1-q')^{N+1-x}.$$

The sign of $\Gamma''(x)$ for integer x is the same as the sign of

$$\phi(x) \equiv \log \frac{\widetilde{f}(x)}{f(x)} = \log \left\{ \frac{N+1}{N+1-x} \left(\frac{N}{N+1} \right)^x \left(\frac{1 - q\frac{N}{N+1}}{1-q} \right)^{N-x} \left(1 - q\frac{N}{N+1} \right) \right\}.$$

Differentiating this function twice yields

$$\phi''(x) = \frac{1}{(N+1-x)^2} > 0,$$

so ϕ is a strictly convex function. Further,

$$\phi(0) = \log \frac{\widetilde{f}(0)}{f(0)} = \log \frac{\widetilde{F}(0)}{F(0)} > 0.$$

Thus ϕ is either always positive, downcrosses zero once, or downcrosses and then upcrosses zero once. The integer truncation in the definition of Γ'' cannot increase the number of crossings, so Γ'' must also exhibit one of these three behaviors. Given that $\Gamma'(0) > 0$, if Γ'' does not cross zero at all, then $\Gamma'(x) > 0$ for all $x \in (0, N)$. If Γ'' crosses zero exactly once as a downcrossing, then either $\Gamma'(x) > 0$ for all x, or else Γ' crosses zero once. Finally, if Γ'' crosses zero twice, then Γ' crosses zero either zero, one, or two times. This establishes the desired upper bound on the number of crossings.

E Proofs for Section 6 (Extensions)

E.1 Proof of Claims 2 and 3

The equilibrium condition $C'(a^*(\beta, N)) = MV(N)$ for an exogenous population of size N implies

$$a^*(\beta, N) = \frac{1}{\beta}(MV(N))^{1/(\gamma-1)} \tag{E.1}$$

which is strictly decreasing in β.

Define $N^*(\beta)$ to be the threshold population satisfying

$$C[a^*(\beta, N^*(\beta))] = t_1.$$

Using (E.1), this simplifies to

$$MV(N^*(\beta)) = \left(t_1\gamma\beta^{\gamma-1}\right)^{(\gamma-1)/\gamma}$$

where the RHS is strictly increasing in β. Since $MV_T(N)$ and $MV_C(N)$ are (respectively) decreasing and increasing in N (Lemma 1), the threshold in the common type model is decreasing in β, while the threshold in the common confound model is increasing in β.

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. 2019. "Too Much Data: Prices and Inefficiencies in Data Markets." Working Paper.

Acquisiti, Alessandro, Laura Brandimarte, and George Loewenstein. 2015. "Privacy and Human Behavior in the Age of Information." *Science*, 347: 509–514.

Agarwal, Anish, Munther Dahleh, and Tuhin Sarkar. 2019. "A Marketplace for Data: An Algorithmic Solution." *ACM Conference on Economics and Computation.*

Anderson, Simon, N. Larson, M. Sanchez, and A. Urbano. 2018. "Demand for privacy, selling consumer information, and consumer hiding vs. opt-out." Working Paper.

Auriol, Emmanuelle, Guido Friebel, and Lambros Pechlivanos. 2002. "Career Concerns in Teams." *Journal of Labor Economics.*

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin. 2018. "The Design and Price of Information." *American Economic Review.*

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan. 2019. "The Economics of Social Data." Working Paper.

Blackwell, David, and Lester Dubins. 1962. "Merging of Opinions with Increasing Information." *The Annals of Mathematical Statistics.*

Bonatti, Alessandro, and Gonzalo Cisternas. 2018. "Consumer Scores and Price Discrimination." Working Paper.

Chouldechova, Alexandra. 2017. "Fair prediction with disparat e impact: A study of bias in recidivism predic- tion instruments." *Big Data, Special issue on Social and Technical Trade-Offs.*

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures." *Review of Economic Studies.*

Dwork, Cynthia, and Aaron Roth. 2014. "The Algorithmic Foundations of Differential Privacy." *Foundations and Trends in Theoretical Computer Science*, 9.

Eliaz, Kfir, and Ran Spiegler. 2018. "Incentive-Compatible Estimators." Working Paper.

Fainmesser, Itay, Andrea Galeotti, and Ruslan Momot. 2019. "Digital Privacy."

Frankel, Alex, and Navin Kartik. 2019. "Improving Information from Manipulable Data." Working Paper.

Georgiadis, George, and Michael Powell. 2019. Optimal Incentives under Moral Hazard:From Theory to Practice.

Gomes, Renato, and Alessandro Pavan. 2018. "Price Customization and Targeting in Platform Markets." Working Paper.

Holmstrom, Bengt. 1982. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies.*

Hu, Lily, Nicole Immorlica, and Jennifer Wortman Vaughan. 2019. "The Disparate Effects of Strategic Manipulation."

Ichihashi, Shota. 2019. "Online Privacy and Information Disclosure by Consumers." *American Economic Review.*

Jin, Yizhou, and Shoshana Vasserman. 2019. "Buying Data from Consumers: The Impact of Monitoring in U.S. Auto Insurance." Working Paper.

Jullien, Bruno, Yassine Lefouili, and Michael Riordan. 2018. "Privacy Protection and Consumer Retention." Working Paper.

Kearns, Michael, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. "Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness." *ICML.*

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2017. "Inherent Trade-Offs in the Fair Determination of Risk Scores." *ITCS.*

Kostka, Genia. 2019. "China's social credit systems and public opinion: Explaining high levels of approval." *New Media and Society.*

Milgrom, Paul. 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics.*

Olea, Jose Luis Montiel, Pietro Ortoleva, Mallesh Pai, and Andrea Prat. 2018. "Competing Models." Working Paper.

Saumard, Adrien, and Jon A. Wellner. 2014. "Log-Concavity and Strong Log-Concavity: A Review." *Statistics Surveys.*