Data and Incentives*

Annie Liang[†] Erik Madsen[‡]

January 5, 2021

Abstract

Markets for lending and insurance incentivize good behavior by forecasting risk on the basis of past outcomes. As "big data" expands the set of covariates used to predict risk, how will these incentives change? We show that "attribute" data which is informative about consumer quality tends to decrease effort, while "circumstance" data which predicts idiosyncratic shocks to outcomes tends to increase it. When covariates are independent, this effect is uniform across all consumers. Under more general forms of correlation, this effect continues to hold on average, but observation of a new covariate may lead to disparate impact—increasing effort for some consumer groups and decreasing it for others. A regulator can improve social welfare by restricting the use of either attribute or circumstance data, and by limiting the use of covariates with substantial disparate impact.

1 Introduction

Lenders and insurers have long personalized the terms offered to individual customers based on an estimate of the customer's risk level. Historically, these estimates have taken into account some personal demographic data, such as age and gender, but have also placed

*We are grateful to Eduardo Azevedo, Cuimin Ba, Dirk Bergemann, Alessandro Bonatti, Sylvain Chassang, Yash Deshpande, Ben Golub, Matt Jackson, Yizhou Jin, Navin Kartik, Rishabh Kirpalani, Alessandro Lizzeri, Steven Matthews, Xiaosheng Mu, Larry Samuelson, Andrzej Skrzypacz, Juuso Toikka, and Weijie Zhong for useful conversations, and to National Science Foundation Grant SES-1851629 for financial support. We thank Changhwa Lee for valuable research assistance on this project.

[†]Department of Economics and Department of Computer Science, Northwestern University

[‡]Department of Economics, New York University

significant weight on the customer's past record (e.g. his history of loan repayments or insurance claims). When a customer's record shapes the market's perception of his risk, he is incentivized to exert effort to improve that record and establish a good reputation, for instance by driving carefully to avoid auto accidents or budgeting expenditures to repay loans. This reputational incentive for effort plays an important role in mitigating moral hazard.

Increasingly, big data is reshaping risk forecasting. For example, financial technology startups evaluate creditworthiness based on personal data such as the frequency with which a potential borrower changes residence[1]; health insurers collect data on insurees' hobbies to forecast healthcare costs[2]; and automobile insurers condition rates on detailed data about traffic, weather, and environmental factors near the driver's address.[3] In other domains, such as college admissions or workplace promotion decisions, "big data" is not yet widely used, but may very well be in the future.

We propose a model of how increasing access to fine-grained consumer data changes incentives for effort. The core of our model is the well-known Holmström (1999) framework: An agent has a private type (e.g. driving ability) which is unknown to both him and the market. This type, along with a random shock (encountering icy roads) and a choice of effort (how carefully to drive), jointly determine a publicly observed outcome (claims rate). The market updates its beliefs about the agent's type based on this outcome. Under suitable regularity assumptions on type and shock distributions, higher effort levels lead the market to raise its expectations of the agent's type. As a result, an agent who benefits from a good type reputation will be incentivized to exert costly effort.

Our contribution is to model the agent's type and shock as predictable from observable covariates, some of which are *attributes* correlated with the agent's type, and others of which are *circumstances* predictive of the agent's shock. For example, an agent's driving ability may be predicted from their attention span and hand-eye coordination, while the shock to

[1] See for example `https://wapo.st/35bWgDZ`.

[2] "[Aetna] had obtained personal information from a data broker... The data contained each person's habits and hobbies, like whether they owned a gun, and if so, what type... The goal was to see how people's personal interests and hobbies might relate to their health care costs." `https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates`

[3] The data analytics firm Verisk includes these datasets in the "environmental module" of its auto risk analyzer: `https://www.verisk.com/insurance/products/iso-risk-analyzer-suite/iso-risk-analyzer-personal-auto/`. At least one major US auto insurance company has announced its use of the Verisk environmental module: `https://www.nbcnews.com/id/wbna41990026`

their accident probability on a given day may be correlated with current weather conditions near their home. The market has access to some, but not all, of these covariates, which it uses to group consumers into categories. We model "big data" as an expansion of the set of covariates available to the market, leading to finer partitioning of consumers.

Our main results characterize how access to new covariates impacts consumer incentive for effort. We show that new attributes tend to decrease effort, while new circumstances tend to increase it. We first establish this result for a class of "mean-shifter" covariates, encompassing models with statistically independent covariates and correlated Gaussian settings, among others. Within this class, any new attribute reduces effort for all individuals (regardless of the individual's value for the newly acquired attribute), while any new circumstance increases effort for all agents.

Outside of the class of mean-shifters, we show by example that access to a new covariate can lead to *disparate impact*, increasing effort for some consumer groups and decreasing it for others. Even so, we prove that if covariates satisfy a positive dependence condition, then conditioning forecasts on new attributes reduces effort on average across each consumer category, while conditioning forecasts on new circumstances increases effort on average. Thus, attributes and circumstances differ qualitatively in their impact on a population of consumers, although the direction of their impact on specific consumers is ambiguous.

Besides their implications for the use of new data in lending and insurance, our results make an important theoretical contribution to the career concerns literature (Holmström, 1999), by characterizing how incentives for effort vary with the amount of prior information about the agent. In particular, we are able to analyze this force in general informational environments, in which the "amount of information" cannot be reduced to the precision of beliefs about the agent's type and shock. These results complement the analyses of Dewatripont et al. (1999) and Rodina (2018), which consider how effort incentives vary when a forecaster observes an additional non-output signal *after* the agent has exerted effort. (We discuss on this relationship further in Section 3.2.)

We apply our results to examine how a social planner should regulate access to new covariates. We show that if existing reputational incentives for effort are inefficiently weak, then a regulator optimally bans access to new attribute data, but may improve welfare by allowing access to new circumstance data. The opposite conclusion holds when reputational incentives are inefficiently strong. These results indicate that it is important to distinguish between attribute and circumstance data when regulating data usage, a dichotomy that is obscured in popular discussions of "big data."

Additionally, we show that all else equal, the regulator optimally exhibits a bias against access to covariates which have a heterogeneous impact across agents. This implies that the regulator should penalize covariates which identify some consumers as belonging to well-understood categories while pooling others into a poorly-understood group exhibiting substantial heterogeneity.

The remainder of this paper proceeds as follows. Section 1.1 discusses related literature; Section 2 describes our model; Section 3 establishes our main results about the impact of covariates on effort; Section 4 describes the implications of our findings for data regulation; Section 5 discusses extensions; and Section 6 concludes. Supporting discussion and all proofs are collected in the Appendix.

1.1 Related Literature

Our paper contributes to an emerging literature regarding how consumer behavior responds to the use of big data for forecasting. A central concern is the possibility that consumers may be able to distort the data being used, opening the door to strategic interaction between consumers and forecasting mechanisms.

One set of papers in this literature examine incentives for "gaming" forecasts, by misreporting private information or exerting effort to distort an observed outcome; see, e.g., Eliaz and Spiegler (2019), Hu et al. (2019), Ball (2020), Eliaz and Spiegler (2020), Frankel and Kartik (2020), and Hennessy and Goodhart (2020). These papers treat gaming as intrinsically inefficient, either because it reduces the precision of forecasts, or because effort is costly and signal distortion generates no social value. In another set of papers (e.g. Frankel and Kartik (2019) and Haghtalab et al. (2020)), effort improves the agent's covariates, or signals a type which the market wishes to forecast. Effort is therefore not intrinsically wasteful, though outcomes may still be inefficient if too little or too much effort is incentivized in equilibrium.

Both sets of papers treat the data environment as fixed, and focus on equilibrium outcomes or design of an optimal forecasting mechanism. Our paper instead considers how outcomes change as the forecaster gains access to richer data. We model applications in which effort is productive, as in the second set of papers, and characterize the effect of observing additional covariates. Tirole (2020) similarly considers the consequences of varying the information available to forecasters in an environment with productive effort. While we focus on how information impacts effort on a task of interest to the market, Tirole (2020) considers whether a designer might wish to muddle a signal in one domain to induce effort in another.

More broadly, our work joins a set of recent papers studying the economic consequences of big data. Data usage has been recognized as having important implications for price discrimination (Bergemann et al., 2020; Ichihashi, 2019; Bonatti and Cisternas, 2019; Yang, 2020; Hidir and Vellodi, 2020; Elliott et al., 2020); targeted advertising (Jullien et al., 2020; Gomes and Pavan, 2019); adverse selection in insurance markets (Brunnermeier et al., 2020; Braverman and Chassang, 2020); and privacy (Acemoglu et al., 2020; Acquisti et al., 2015; Dwork and Roth, 2014; Fainmesser et al., 2020; Eilat et al., 2020). In addition, the widespread collection and sale of big data has motivated the study of optimal data bundling (Bergemann et al., 2018; Segura-Rodriguez, 2020) and data market design (Agarwal et al., 2019).

Finally, a leading application of our model is to insurance markets. Jin and Vasserman (2020) show empirically that in the auto insurance market, a short-term monitoring program which collects data on driving behavior incentivizes drivers to substantially increase driving effort. Notably, this behavior change occurs even though the insurer does not directly reward better driving, and instead collects data only in order to forecast future accident risk and adjust insurance premiums. Their study indicates that incentives for effort deriving from reputational concerns are of significant practical relevance in this market.

2 Model

In this section we describe our model of data and reputational incentives for effort. Our model takes as a starting point the classic Holmström (1999) framework, which we review in Section 2.1. In Section 2.2, we build on that framework by modeling the dependence of the agent's type and shock on a set of underlying covariates, some of which will be observable to the market. In Section 2.3 we provide specific examples of what those covariates might be across different applications, and in Section 2.4 we discuss key modeling assumptions.

2.1 Review of Holmström (1999)

An agent participates in a market across two periods $t = 1, 2$.[4] He possesses a quality type $\theta \sim F_\theta$, which is persistent across time and unknown to himself and the market. In **period 1**, the agent generates an observable outcome

$$Y = e + \theta + \varepsilon,$$

which is determined by the agent's quality θ, a transient shock $\varepsilon \sim F_\varepsilon$ independent of the agent's type, and an effort level $e \in \mathbb{R}_+$ privately chosen by the agent.

The agent incurs a cost to exert effort, which we take to be $C(e) = \frac{1}{2}e^2$ for expositional simplicity. (We extend our results to general cost functions in Section 5.3.) We suppose that any payment to the agent to enter the market is sunk,[5] so the agent's period-1 payoff from participation is just his total cost of effort:

$$U_1 = -C(e) = -\frac{1}{2}e^2$$

In **period 2**, the agent receives a reputational payoff standing in for returns from future participation in the market. This payoff is equal to the market's expectation of his quality conditional on the outcome variable Y.[6] Since the agent's effort choice is private, the market's forecast is based on a conjectured level of effort $\hat{e}$. Letting $Y^{\hat{e}} \equiv \hat{e} + \theta + \varepsilon$ be output supposing the market's effort conjecture is correct, the agent's second-period payoff conditional on realized output $Y = y$ is

$$U_2 = \mathbb{E}^{\hat{e}}\left[\theta \mid Y = y\right] \equiv \int \theta dF_\theta(\theta \mid Y^{\hat{e}} = y) \tag{1}$$

where $\mathbb{E}^{\hat{e}}[\theta \mid Y = y]$ denotes the market's (potentially misspecified) expectation of θ, updat-

[4]In general, strong parametric assumptions are needed to make a many-period version of this model tractable. We focus on a 2-period model to permit study of more general information structures, but our main results continue to hold in a many-period version of the model with Gaussian uncertainty (see Section 5.4).

[5]This assumption does not imply that the agent would prefer not to participate in the period-1 market. It entails only that the agent's utility from participation is independent of his period-1 outcome, as is true, for instance, for a driver who is fully insured against accidents in that period.

[6]None of our results would be impacted if the agent's reputational payoff were instead the market's expectation of any strictly increasing function of θ. Our model therefore accommodates a variety of interpretations for the source of reputational returns from effort.

ing to realized output assuming that $Y = Y^{\hat{e}}$.

The agent's ex post payoff from participating in the market in both periods is a discounted sum of period payoffs:

$$U = U_1 + \beta \cdot U_2,$$

where $\beta > 0$ denotes the agent's discount factor. (We allow β to be greater than 1 to reflect the relative importance of the agent's future reputation over a long time horizon; see Section 4 for further discussion). The agent's expected payoff under effort level e is therefore

$$\mathbb{E}^e[U] = \beta \cdot \mathbb{E}^e[\mathbb{E}^{\hat{e}}[\theta \mid Y]] - \frac{1}{2}e^2$$

where $\mathbb{E}^e$ denotes the agent's expectation over output given the true effort level e.

In equilibrium, the agent must have no incentive to deviate from the market's conjectured level of effort. That is, the marginal value of effort at the equilibrium level e^* must equal its marginal cost:

$$\beta \cdot \left. \frac{\partial \mathbb{E}^e[\mathbb{E}^{e^*}[\theta \mid Y]]}{\partial e} \right|_{e=e^*} = e^* \tag{2}$$

The left-hand side of this equation is independent of e^* due to the additive impact of effort on output, so there is a unique solution to the first-order condition. Throughout this paper, we will assume that the first-order approach is valid, so that this solution constitutes the unique equilibrium effort level.

2.2 Model of Covariates and Beliefs

We view the agent's type θ and shock ε as predictable from data about the agent's characteristics, some of which may be available to the market. In this section we propose a model of this dependence. Data about the agent is summarized by a set of real-valued *covariates*, which are initially unknown and so are formally random variables, drawn according to a joint distribution to be specified in more detail later. These covariates are divided into two categories: *attributes* $\boldsymbol{a} = (a_1, a_2, \ldots, a_J)$, and *circumstances* $\boldsymbol{c} = (c_1, c_2, \ldots, c_K)$. Attributes are predictive of the agent's quality type θ, while circumstances are predictive of his shock ε. To maintain independence of the agent's type and shock, we assume that $\boldsymbol{a} \perp\!\!\!\perp \boldsymbol{c}$.

In the main text we assume that covariates impact types and outcomes additively. (We extend our results to a general nonlinear model in Section 5.1.) The impact of each covariate is captured by the deterministic *effect size functions* Ψ^j and Λ^k. Specifically, the variables θ and ε are decomposable as

$$\begin{aligned} \theta &= \mu + \sum_{j=1}^{J} \Psi^j(a_j) + \theta^{\perp} \\ \varepsilon &= \quad\;\; \sum_{k=1}^{K} \Lambda^k(c_k) + \varepsilon^{\perp}, \end{aligned}$$

where μ is a deterministic scalar. The random variables $\theta^{\perp}$ and $\varepsilon^{\perp}$ are independent of each other and all attributes and circumstances, and represent the unlearnable components of the agent's type and shock. Without loss we assume that each of the random components contributing to type and shock has mean zero, i.e. $\mathbb{E}[\Psi^j(a_j)] = \mathbb{E}[\Lambda^k(c_k)] = \mathbb{E}[\theta^{\perp}] = \mathbb{E}[\varepsilon^{\perp}] = 0$ for every j and k. Thus $\Psi^j(a_j)$ and $\Lambda^k(c_k)$ represent the de-meaned impacts of the covariates on outcomes, while the scalar μ captures the mean outcome.

We assume that the agent and market commonly observe the distributions of qualities and shocks within certain *subpopulations* of agents, as identified by a set of observed covariates.

Definition 1. *A* subpopulation $\mathcal{S}$ *is a quadruple* $(\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, *where:*

- $\mathcal{J} \subset \{1, \ldots, J\}$ *is a set of observed attributes,* $\mathcal{K} \subset \{1, \ldots, K\}$ *is a set of observed circumstances, and*
- $\boldsymbol{\alpha} \in \mathbb{R}^J$ *and* $\boldsymbol{\gamma} \in \mathbb{R}^K$ *are realizations of the covariate vectors* $\boldsymbol{a}_{\mathcal{J}}$ *and* $\boldsymbol{c}_{\mathcal{K}}$.

Given a set of covariates $(\mathcal{J}, \mathcal{K})$, *a* $(\mathcal{J}, \mathcal{K})$-subpopulation *is any subpopulation whose observed covariates are* $(\mathcal{J}, \mathcal{K})$.

We will subsequently suppose that the agent's covariates in a set $(\mathcal{J}, \mathcal{K})$ are known to both the agent and the market, and that the distributions of θ and ε within the agent's $(\mathcal{J}, \mathcal{K})$-subpopulation are common knowledge. When $\mathcal{J} = \mathcal{K} = \emptyset$, this assumption amounts to the standard one that the distributions of θ and ε are commonly known. For more refined subpopulations, we interpret this assumption as reflecting "big data" not only in the sense of many covariates, but also in the sense of past outcomes for a large number of agents. These records allow insurers and lenders to infer the aggregate distribution of qualities and shocks within observed subpopulations.[7] It is not necessary for our subsequent results that actors in the model possess any additional information about the joint distributions of $\boldsymbol{a}$ and $\boldsymbol{c}$ or the shapes of the effect size functions $(\Psi^j)_{j=1}^J$ and $(\Lambda^k)_{k=1}^K$. Given knowledge of the conditional distributions of θ and ε in a subpopulation, forecasts of θ do not change with additional information about these parameters.

We impose the following weak technical conditions on the distributions of latent variables, which are maintained throughout the paper. The first ensures that the agent's utility function is differentiable, while the second ensures that types and shocks have full support on $\mathbb{R}$.[8]

Assumption 1 (Differentiability). *For every set of observed covariates $(\mathcal{J}, \mathcal{K})$, each $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, and every effort level e, $\frac{\partial}{\partial y}\mathbb{E}^e\left[\theta \mid Y = y, \mathcal{S}\right]$ exists and is uniformly bounded across all realizations y of Y.*

Assumption 2 (Residual Noise). *The random variables $\theta^\perp$ and $\varepsilon^\perp$ have full support on $\mathbb{R}$.*

Interactions proceed according to the model described in Section 2.1: In period $t = 1$, the agent chooses effort $e \in \mathbb{R}_+$ and incurs the cost of effort. In period $t = 2$, the agent's outcome Y is realized, and the agent receives the market's forecast of his type. Our main results describe how an agent's effort depends on which sets of covariates $(\mathcal{J}, \mathcal{K})$ are observed, and on the realized values of those covariates.

2.3 Motivating Examples

Below we describe leading applications of our model, with accompanying interpretations of model parameters and possible covariates:

Automobile Insurance. The agent is a fully insured driver.[9] Effort e corresponds to more attention to careful driving, and the outcome $-Y$ is the driver's insurance claim rate. His

[7] This interpretation may be less realistic in the short-run, if the population becomes finely partitioned into groups with small memberships. We abstract away from estimation considerations in this paper (see Braverman and Chassang (2020) for related work that tackles this question more explicitly), but note that in practice firms can choose the number of covariates used to segment agents, and may do so in a way that preserves subpopulations of sufficient size.

[8] We expect that our results continue to hold in a more general setting without full support. We impose this assumption to simplify proofs which establish affiliation of sets of random variables, by allowing us to rely on characterizations of affiliation for probability distributions with strictly positive densities.

[9] If the driver has partial insurance, so that he is exposed to some of the downside of any accident, our subsequent results extend straightforwardly whenever the agent must pay a fixed fraction of any damages incurred. We expect the forces we identify here to also play a role in a richer model of nonlinear insurance contracting, but do not pursue that more complex analysis here.

type θ is his driving ability, which may be predicted from attributes such as distractability, aggression, or hand-eye coordination. The shock ε may be predicted from circumstances such as traffic incidents, abnormal road conditions (e.g. construction), or weather events. The driver's period-2 payoff is his insurance premium in the next claims cycle.

Health Insurance. The agent is a fully insured patient. Effort e corresponds to exercising regularly and eating well, and the outcome $-Y$ is the patient's insurance claim rate. His type θ is his health, which may be predicted from attributes such as his social media activity, online purchases, or hobbies (e.g. gun ownership). The shock ε captures a transient health shock, and may be predicted from circumstances such as food contamination outbreaks, COVID incidence, or air quality. The patient's period-2 payoff is his insurance premium in the next coverage cycle.

Lending. The agent is a borrower with a credit card. Effort e corresponds to actions taken to help pay off his credit card balance (such as budgeting expenses or earning additional income), and the outcome Y is the borrower's repayment rate. His type θ is his creditworthiness, and may be predicted from attributes such as the agent's social network size, financial literacy, and political affiliation. The shock ε captures transient financial shocks, and may be predicted from circumstances such as one-time expenses (e.g. a wedding), windfall gains (such as a lawsuit settlement), and work furloughs. The borrower's period-2 payoff is the rate of interest charged on future credit card balances.

Admissions/internship. The agent is a student applying to college or an internship. Effort e corresponds to how hard he studies, and the outcome Y is his GPA. His type θ is his intrinsic ability, and may be predicted from attributes such as parental background, reading habits, and interview scores. The shock ε is an idiosyncratic shock to GPA, and may be predicted from circumstances such as illness, injury, financial hardship, and school disruptions. The student's period-2 payoff is the quality of the university he attends/internship he obtains.

Labor markets. The agent is a worker. Effort e corresponds to how hard he works, and the outcome Y is the worker's output. His type θ is his intrinsic ability, and may be predicted from attributes such as personality metrics or professional certifications. The shock to output ε may be predicted from circumstances such as the industry growth rate and employer restructurings. The worker's period-2 payoff is his future wage.

2.4 Discussion of Modeling Choices

We now briefly discuss and interpret important features of the model.

Symmetric uncertainty. In our model, agents do not possess private information about their type beyond what the market knows. This assumption does not require that agents are unaware of covariate values unobserved by the market, but does require that they cannot forecast how those covariates impact outcomes. While agents may be able to learn how a firm's current statistical model classifies their subpopulation, they are much less likely to know how that model would change when augmented with additional covariates.[10] Nonetheless, Section 5.2 demonstrates a related extension to model uncertainty on one side, and we conjecture that our main results would extend also given asymmetric information advantaging either side. We leave exploration of this question to future work.

The agent's reputational payoff. As in Holmström (1999), we suppose that the agent's second-period payoff is the market's expectation of his type conditional on his period-1 outcome. This specification may be directly microfounded by assuming that the agent participates in the market twice; the agent's type is persistent across periods, while his shock is drawn anew; and competition between firms to serve the agent results in his being paid his expected output in each period. We formalize this interpretation in Appendix A, and show that it reduces to the model described above.

The reputational payoff can also be understood as a reduced-form stand-in for future benefits accrued by the agent (in this and other markets) by being perceived as high-quality. From this perspective, the discount factor β can be interpreted as the weight of the discounted reputational rewards across multiple time periods or markets, accounting for changes in the scale of service over time.

Interpretation of effort, outcome, and attributes. Our model contrasts attributes—which are fixed characteristics of the agent—and the outcome Y, which is susceptible to manipulation by effort. While not all covariates are truly fixed,[11] we view this separation as

[10] Brunnermeier et al. (2020) emphasize the possibility that big data may actually give insurers an informational advantage over insurees regarding how covariates map to outcomes. We abstract from such considerations of "inverse selection" in our analysis.

[11] See Haghtalab et al. (2020) for a complementary analysis in which the agent cannot directly influence Y, but can manipulate individual attributes.

a useful one for the following reasons: First, our view is that in general θ and ε are determined by the aggregation of many covariates, each of which individually plays only a small role. Our exercise focuses on the agent's incentives to influence the relatively informative output signal Y, while abstracting from any costly distortion of individual less-informative covariates. Second, to compute the value of manipulating a covariate, the agent must know something about the shape of the effect size function describing how different covariate values impact the market's beliefs. Such knowledge requirements are substantially more demanding than the ones we have imposed.

3 Main Results

Our main results characterize how equilibrium effort changes as the market gains access to new covariates. Section 3.1 establishes necessary notation and definitions. Section 3.2 considers a benchmark setting with independent covariates, and establishes the main result that attributes deterministically reduce effort while circumstances deterministically increase it. Section 3.3 generalizes this result to a class of "mean-shifter" covariates. Finally, Section 3.4 demonstrates that in a larger class of affiliated covariates, attributes reduce effort and circumstances increase effort *on average*.

3.1 Preliminary Definitions and Notation

Fix a baseline set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, which we will use as the reference point for all results in this section. Henceforth, we will say that a property holds "on $\mathcal{S}$" as shorthand for "conditional on the realizations $a_j = \alpha_j$ for all $j \in \mathcal{J}$ and $c_k = \gamma_k$ for all $k \in \mathcal{K}$."

To streamline exposition, we define *type components* and *shock components* to capture the impact of a given covariate an agent's type or shock. We also define *residual type components* and *residual shock components* to capture the part of the agent's type or shock that is unexplained by a given set of covariates.

Definition 2. *For each attribute j, define $\theta_j \equiv \Psi^j(a_j)$ to be the corresponding* type component. *For each set of attributes* $\mathcal{J}$, *define*

$$\theta^{-\mathcal{J}} \equiv \sum_{j \notin \mathcal{J}} \theta_j + \theta^{\perp}$$

to be the corresponding residual type component. *Similarly, for each circumstance* k, *define* $\varepsilon_k \equiv \Lambda^k(c_k)$ *to be the corresponding* shock component. *For each set of circumstances* $\mathcal{K}$, *define*

$$\varepsilon^{-\mathcal{K}} \equiv \sum_{k \notin \mathcal{K}} \varepsilon_k + \varepsilon^{\perp}$$

to be the corresponding residual shock component.

When the observed covariates are $(\mathcal{J}, \mathcal{K})$, the agent's type θ and the shock ε to his outcome can be decomposed as

$$\theta = \mu + \sum_{j \in \mathcal{J}} \theta_j + \theta^{-\mathcal{J}} \qquad\qquad \varepsilon = \sum_{k \in \mathcal{K}} \varepsilon_k + \varepsilon^{-\mathcal{K}}$$

where $\mu + \sum_{j \in \mathcal{J}} \theta_j$ and $\sum_{k \in \mathcal{K}} \varepsilon_k$ are known to the agent and market while $\theta^{-\mathcal{J}}$ and $\varepsilon^{-\mathcal{K}}$ are not.

In our main results, we impose regularity log-concavity conditions on the distributions of type and shock components. Log-concavity of latent variables in an additive model is a standard assumption ensuring that higher outputs lead to larger inferred types and shocks, so that higher effort leads to better perceptions of quality. (See Appendix E.1 for a list of examples of covariate distributions and effect size functions which generate log-concave type and shock components.)

Definition 3. *A random variable* X *is* log-concave *if the distribution function of* X *admits a log-concave density function. A conditional random variable* $X|Y$ *is* log-concave *if conditional on any realization of* Y, *the distribution function of* X *admits a log-concave density function.*

Definition 4 (Regularity)**.** *Fix a set of covariates* $(\mathcal{J}, \mathcal{K})$ *and a* $(\mathcal{J}, \mathcal{K})$*-subpopulation* $\mathcal{S}$. *Say that* $\mathcal{S}$ *is* regular *if* $\theta^{-\mathcal{J}}$ *and* $\varepsilon^{-\mathcal{K}}$ *are log-concave on* $\mathcal{S}$.

Definition 5 ($\mathcal{S}$-Regularity)**.** *Fix a set of covariates* $(\mathcal{J}, \mathcal{K})$ *and a* $(\mathcal{J}, \mathcal{K})$*-subpopulation* $\mathcal{S}$. *An attribute* $j' \notin \mathcal{J}$ *is* $\mathcal{S}$-regular *if* $\theta^{-\mathcal{J} \cup \{j'\}} \mid a_{j'}$ *is log-concave on* $\mathcal{S}$. *Similarly, a circumstance* $k' \notin \mathcal{K}$ *is* $\mathcal{S}$-regular *if* $\varepsilon^{-\mathcal{K} \cup \{k'\}} \mid c_{k'}$ *is log-concave on* $\mathcal{S}$.

Regularity requires that in the baseline setting, higher realizations of the outcome lead to higher inferences about the unobserved components of type and shock. $\mathcal{S}$-regularity imposes the same condition in the setting where an additional attribute or circumstance is observed. For all of the results of this section, we maintain the following assumption on the baseline subpopulation:

Assumption 3. *The subpopulation $\mathcal{S}$ is regular.*

3.2 Independent Covariates

Our first result applies to environments in which all covariates are independent. In this case, observing an additional attribute leads deterministically to a decrease in equilibrium effort, while observing an additional circumstance leads deterministically to an increase in effort.

Proposition 1. *Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. If all covariates are mutually independent, then:*

(a) Observing the additional attribute j' reduces the agent's effort.

(b) Observing the additional circumstance k' increases the agent's effort.

Further, the magnitude of the effort change is independent of the observed value of the additional covariate.

Our result is stated from a single-agent perspective. There is an alternative populational interpretation: If the market gains access to a new attribute, then effort decreases *for all* agents in subpopulation $\mathcal{S}$; if the market gains access to a new circumstance, then effort increases *for all* agents in $\mathcal{S}$.

The key to this result is that—under the assumption of independence—observing an additional covariate reduces the market's uncertainty about the corresponding component of output, regardless of the realized value of the covariate. Therefore if a new attribute is acquired, the market's uncertainty about θ drops, and so there is less to learn about θ from the realization of the agent's outcome Y. This effect reduces the marginal value of exerting effort to improve the realization of Y. By contrast, acquisition of an additional circumstance reduces uncertainty about the shock ε, making the outcome Y a more informative signal of the agent's type. This effect increases the marginal value of improving the realization of Y.

By the first-order condition (2), these changes in the marginal value of effort are directly inherited by the level of equilibrium effort.

This result is closely related to Proposition 5.1 in Dewatripont et al. (1999), with the key difference that our result concerns *ex-post* effort in an environment where the agent knows his subpopulation before choosing effort, while Dewatripont et al. (1999) analyzes *ex-ante* effort when the agent does not. A key difference is that under the ex post perspective, agents may react differently to acquisition of the same covariate, depending on their value of that covariate. As Proposition 1 shows, the ex post and ex ante outcomes coincide when covariates are mutually independent. That is, the ex post effect is equal to the ex ante effect realization by realization. More generally, the two outcomes can diverge.

3.3 Mean-Shifter Covariates

We now examine the extent to which our findings for the independent-covariate baseline generalize when covariates are correlated. In this section, we define a class of correlated-covariate models with the property that, when a new covariate is observed, posterior uncertainty does not depend on the realization of the new covariate. We show that Proposition 1 can be extended to cover this class.

Definition 6. *Given a set of attributes $\mathcal{J}$, the* de-meaned residual type component $\tilde{\theta}^{-\mathcal{J}}$ *is the part of the residual type component $\theta^{-\mathcal{J}}$ not explained by the observed attributes $\boldsymbol{a}_{\mathcal{J}}$:*

$$\tilde{\theta}^{-\mathcal{J}} \equiv \theta^{-\mathcal{J}} - \mathbb{E}[\theta^{-\mathcal{J}} \mid \boldsymbol{a}_{\mathcal{J}}].$$

Similarly, given any set of circumstances $\mathcal{K}$, the de-meaned residual shock component $\tilde{\varepsilon}^{-\mathcal{K}}$ *is the part of the residual shock component $\varepsilon^{-\mathcal{J}}$ not explained by the observed circumstances $\boldsymbol{c}_{\mathcal{K}}$:*

$$\tilde{\varepsilon}^{-\mathcal{K}} \equiv \varepsilon^{-\mathcal{K}} - \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid \boldsymbol{c}_{\mathcal{K}}].$$

Definition 7 (Mean shifters)**.** *Fix any attribute $j' \notin \mathcal{J}$ and circumstance $k' \notin \mathcal{K}$.*

*(a) Attribute j' is an $\mathcal{S}$-*mean shifter *if the de-meaned residual type component $\tilde{\theta}^{-\mathcal{J}'}$ is independent of the random variable $a_{j'}$ on $\mathcal{S}$, where $\mathcal{J}' = \mathcal{J} \cup \{j'\}$.*

*(b) Circumstance k' is an $\mathcal{S}$-*mean shifter *if the de-meaned residual shock component $\tilde{\varepsilon}^{-\mathcal{K}'}$ is independent of the random variable $c_{k'}$ on $\mathcal{S}$, where $\mathcal{K}' = \mathcal{K} \cup \{k'\}$.*

The assumption that a particular covariate is a mean shifter is a substantive restriction on permitted correlation structures. The de-meaned type and shock components $\tilde{\theta}^{-\mathcal{J}'}$ and $\tilde{\varepsilon}^{-\mathcal{K}'}$ must, by construction, have mean zero conditional on any realization of the observed covariates $\boldsymbol{a}_{\mathcal{J}'} = (\boldsymbol{a}_{\mathcal{J}}, a_{j'})$ and $\boldsymbol{c}_{\mathcal{K}'} = (\boldsymbol{c}_{\mathcal{K}}, c_{k'})$. However, the higher moments of $\tilde{\theta}^{-\mathcal{J}'}$ and $\tilde{\varepsilon}^{-\mathcal{K}'}$ may depend on the realizations of $a_{j'}$ and $c_{k'}$. The substance of the mean shifter property is to assume away such variation. While it is certainly restrictive, this class encompasses a diverse range of examples:[12]

Example 1. Suppose the vectors of type and shock components each follow multivariate normal distributions.[13] Then the mean-shifter property is satisfied *globally*—that is, for any set of observed covariates $(\mathcal{J}, \mathcal{K})$ and each $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, every attribute $j' \notin \mathcal{J}$ is an $\mathcal{S}$-mean shifter and every circumstance $k' \notin \mathcal{K}$ is an $\mathcal{S}$-mean shifter. (See Appendix E.2 for details.)

Example 2. Suppose attribute a_1 is uniformly distributed on $[a, b]$ with $\Psi^1(a_1) = \sqrt{a_1}$. Define $a_2 \equiv a_1 \cdot X$, where X is a random variable that follows an exponential distribution, and let $\Psi^2(a_2) = \log(a_2)$. Then a_1 is a mean shifter when no covariates are observed in the baseline.

Example 3. Suppose Ψ^1 and Ψ^2 are both affine functions, and a_1 is a log-concave random variable. Let $a_2 \equiv a_1 + X$ for any log-concave random variable X that is independent of a_1. Then a_1 is a mean-shifter when no covariates are observed in the baseline.

The following theorem extends Proposition 1 to settings in which the markets observes an additional mean shifter.

Theorem 1. *Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

(a) If j' is an $\mathcal{S}$-mean shifter, observing the additional attribute j' reduces the agent's effort.

(b) If k' is an $\mathcal{S}$-mean shifter, observing the additional circumstance k' increases the agent's effort.

Further, the magnitude of the effort change is independent of the observed value of the additional covariate.

[12] These examples also satisfy all of our standing technical assumptions, including regularity.

[13] This special case of our model closely corresponds to models studied in Meyer and Vickers (1997), Acemoglu et al. (2020), and Bergemann et al. (2020).

We prove this result by "de-meaning" the model's covariates, which transforms the setting back into one of independent covariates. We briefly sketch our approach for the case of an additional attribute, with the case of an additional circumstance following from similar arguments. The key idea is to decompose the residual component $\theta^{-\mathcal{J}}$ into the sum of two parts: the expectation of $\theta^{-\mathcal{J}}$ conditional on the new covariate $a_{j'}$, and the de-meaned residual component $\tilde{\theta}^{-\mathcal{J}'}$ which captures all remaining variation. In the baseline, both are random variables from the perspective of the market. When the new attribute $a_{j'}$ is observed, it perfectly reveals the former variable but is independent of the latter. Thus we return a setting that resembles that of the previous section, where Y is decomposed into the sum of independent parts, and the new attribute $a_{j'}$ reveals one of these. The main technical challenge is that we are not guaranteed that regularity holds in this transformed setting: That is, the assumption that j' is $\mathcal{S}$-regular implies that higher realizations of Y lead to higher inferences of $\theta^{-\mathcal{J}'}$, but does not immediately imply that higher realizations of Y leads to higher inferences about the de-meaned residual $\tilde{\theta}^{-\mathcal{J}'}$. To complete the proof, we establish regularity for this transformed environment.

3.4 General Covariates

Beyond the class of mean-shifter covariates, new forces emerge. First, we demonstrate that uniformity of impact can break in a strong sense: Not only does effort generally vary with the realization of the new covariate, but even the *directional* effect may differ, so that the new covariate raises effort for some agents while decreasing it for others.

Example 4 (Disparate Impact). Suppose $-Y$ is a driver's automobile claims amount, which is modeled as

$$Y = e + \Psi^1(a_1) + \Psi^2(a_2) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1)$$

where $a_1 \in [0, 10]$ is the age of the car, while $a_2 \in [0, 1]$ is the driver's education quantile. The effect size functions are $\Psi^1(a_1) = -a_1$ and $\Psi^2(a_2) = -(1 - a_2)/10$, so risk levels (i.e. outcomes are lower) are higher for drivers of older cars, and drivers with lower levels of education.

The driver's education level and car age are correlated. The education quantile a_2 is uniformly distributed on $[0, 1]$, while the distribution of the age of the driver's car a_1 conditional on a_2 is:

$$a_1|a_2 \sim \begin{cases} U([0,1]) & \forall a_2 \geq 0.05 \\ U([0,10]) & \forall a_2 < 0.05 \end{cases}$$

Under this distribution, individuals in the bottom 5% of educational attainment—i.e. those without a high-school diploma—drive cars whose ages are both higher on average as well as significantly more variable.

If the market observes neither attribute, then all individuals exert the same amount of effort $e^* \approx 0.16$.[14] If educational attainment is observed, equilibrium effort for individuals with a high-school diploma declines to $e^{**} \approx 0.077$, while equilibrium effort for individuals without a high-school diploma rises to $\tilde{e}^{**} \approx 0.82$. The collection of data on educational attainment thus affects individuals in the population unequally.

In this example, access to a new attribute does not uniformly decrease effort across the population. It therefore demonstrates that the monotonicity result of Theorem 1 may fail in the presence of correlated covariates. However, the example does exhibit a weaker form of monotonicity: When educational attainment is observed, average effort across the population falls to approximately 0.11, compared to the baseline of 0.16. Thus, the finding of Theorem 1 does generalize in the sense that access to the new attribute reduces effort *on average*. This property turns out to hold across a large class of correlated covariates, which we now define.

Definition 8 (Affiliation)**.** *An attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-*affiliated *if $\Psi^{j'}$ is one-to-one and $(\theta_{j'}, \theta^{-\mathcal{J}\cup\{j'\}})$ are affiliated on $\mathcal{S}$. Similarly, a circumstance $k' \notin \mathcal{K}$ is $\mathcal{S}$-*affiliated *if $\Lambda^{k'}$ is one-to-one and $(\varepsilon_{k'}, \varepsilon^{-\mathcal{K}\cup\{k'\}})$ are affiliated on $\mathcal{S}$.*

Affiliation is a well-known form of positive dependence between random variables.[15] If a covariate satisfies $\mathcal{S}$-affiliation, then "good news" about that covariate's contribution to the outcome is also good news about the contribution of all unobserved covariates. For example, suppose in the context of auto insurance that the total shock to a driver's insurance claim

[14]The marginal value of effort in this example does not have an analytical closed form, but may be calculated numerically. The code used to perform the calculation is available upon request.

[15]An n-vector of random variables $\boldsymbol{Z}$ with joint density function $f(\boldsymbol{z})$ is *affiliated* if for every $\boldsymbol{z}, \boldsymbol{z'} \in \mathbb{R}^n$, $f(\boldsymbol{z} \wedge \boldsymbol{z'})f(\boldsymbol{z} \vee \boldsymbol{z'}) \geq f(\boldsymbol{z})f(\boldsymbol{z'})$, where $\boldsymbol{z} \wedge \boldsymbol{z'}$ and $\boldsymbol{z} \vee \boldsymbol{z'}$ are the componentwise minimum and maximum of $\boldsymbol{z}$ and $\boldsymbol{z'}$. When f is strictly positive and twice-differentiable everywhere, this condition is equivalent to $\partial^2 \log f / \partial z_i z_j \geq 0$ everywhere for every pair of components i and $j \neq i$.

may be decomposed as $\varepsilon = \Lambda^1(c_1) + \Lambda^2(c_2) + \Lambda^3(c_3) + \varepsilon^\perp$, where c_1 is precipitation, c_2 is the amount of ice on the ground, and c_3 is the stress level of the driver, with each Λ^k an increasing function. The precipitation covariate satisfies affiliation if a higher realization of $\Lambda^1(c_1)$ is associated with a higher belief about the sum $\Lambda^2(c_2) + \Lambda^3(c_3) + \varepsilon^\perp$. The additional requirement that the associated effect size functions be one-to-one simplifies statement of the property, by ensuring that type and shock components θ_j and ε_k are a sufficient statistic for the underlying covariate values a_j and c_k.

One setting satisfying affiliation is Example 4, where the attribute a_2, representing educational attainment, satisfies affiliation when no covariates are observed in the baseline. The following example illustrates another environment with affiliated covariates.[16]

Example 5. Suppose that all effect size functions Ψ^j are one-to-one, and type components are independent exponentially distributed random variables conditional on a common rate parameter: $\theta_j|\lambda \sim_{iid} \text{Exp}(\lambda)$. Suppose further that the rate parameter λ is distributed as $\lambda \sim \text{Gamma}(\alpha, \beta)$ with $\alpha \geq 1$. Then the affiliation property is satisfied globally—for any set of covariates $(\mathcal{J}, \mathcal{K})$ and $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$, each attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-affiliated. (See Appendix E.3 for a proof.) An analogous statement holds for circumstance variables exhibiting this correlation structure.

We now establish that our characterization of the directional effect of new covariates extends *on average* to settings with affiliated covariates.

Theorem 2. *Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

(a) If j' is $\mathcal{S}$-affiliated, observing the additional attribute j' reduces the agent's effort on average.

(b) If k' is $\mathcal{S}$-affiliated, observing the additional circumstance circumstance k' increases the agent's effort on average.

We defer the details of the proof to the appendix, but provide a brief intuition here for why affiliation has implications for effort. Consider observation of a new attribute j'. The desired result holds if can we show that the posterior forecast of the residual unknown $\theta^{-\mathcal{J}} = \theta_{j'} + \theta^{-\mathcal{J}'}$ is (in expectation) less responsive to the realization of the outcome Y

[16] Any positively correlated multivariate normal model also satisfies affiliation, although this case is already covered by the stronger result of Theorem 1.

when conditioned on $\theta_{j'}$. Whether or not attribute j' is observed, improving the realization of Y directly improves inferences about $\theta^{-\mathcal{J}'}$. In the baseline where $\theta_{j'}$ is not observed, an increase in Y has a further direct effect of improving inferences about $\theta_{j'}$, as well as an *indirect* effect: a larger inferred $\theta_{j'}$ shifts up the posterior distribution of $\theta^{-\mathcal{J}'}$ (by affiliation of $\theta_{j'}$ and $\theta^{-\mathcal{J}'}$). When $\theta_{j'}$ is observed, these latter two effects are eliminated, since the distribution of $\theta_{j'}$ is unresponsive to manipulation of Y. As a result, the agent faces weaker incentives to manipulate Y when j' is observed.

4 Welfare and Regulation of Data

We now apply our main results to study the welfare implications of access to new covariates. Our welfare criterion is the expected total social surplus generated by the agent's effort. As in Holmström (1999), we assume that the outcome variable Y directly measures social value generated by the agent in period 1, while the effort costs incurred by the agent are socially dissipative. Therefore, the contribution to welfare by an agent exerting equilibrium effort e is

$$W(e) \equiv \mathbb{E}^e[Y] - C(e) = \mu + e - C(e).$$

(We do not include the agent's equilibrium reputational payoff in the welfare calculation because, as noted in Section 2.1, that payoff is fixed at $\mathbb{E}^e[\mathbb{E}^e[\theta \mid Y]] = \mu$, independent of the available covariates and the equilibrium effort level.) This welfare function is strictly concave and maximized at the first-best effort level e^{FB} satisfying $C'(e^{FB}) = 1$. It is therefore possible for equilibrium effort to be too high or too low, depending on how the discounted marginal value of effort compares to 1.

Our main finding is that when the weight the agent places on future reputation is low (small β), observing additional attributes decreases welfare, while observing new circumstances increases it. These statements are reversed when the agent's concerns for future reputation are significant (large β). A regulator can therefore improve welfare by prohibiting use of one kind of data while encouraging the other.

Formally, fix a baseline subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. Suppose a regulator can choose whether to permit access by the market to a novel covariate for forecasting.[17] If the new

[17] Similar regulations already exist at the level of individual covariates (e.g. banks are not permitted to use race to set credit limits), although the motivations for existing regulations are different from the ones

covariate is not permitted, agents in subpopulation $\mathcal{S}$ exert some (common) effort level e^*. Suppose first that the conditions of Theorem 1 apply and the new covariate is a mean-shifter. Then following observation of the covariate, effort shifts to the new (common) level e^{**}, where $e^{**} \leq e^*$ if the new covariate is an attribute, and the inequality is reversed if the new covariate is a circumstance. We therefore obtain the following result as a straightforward corollary of Theorem 1:

Corollary 1. *Suppose the new covariate is an $\mathcal{S}$-mean shifter. Then:*

(a) If the covariate is an attribute, there exists a threshold discount factor β_* such that it is (weakly) optimal for the regulator to permit access if and only if $\beta \geq \beta_*$.

(b) If the covariate is a circumstance, there exists a threshold discount factor β^* such that it is (weakly) optimal for the regulator to permit access if and only if $\beta \leq \beta^*$.

That is, the regulator should prohibit use of circumstance data when the weight on the future is (sufficiently) high, and should prohibit use of attribute data when the weight on the future is (sufficiently) low. The intuition for this result is as follows: The welfare-maximizing effort level e^{FB} is characterized by $C'(e^{FB}) = 1$. Equilibrium effort, in contrast, is characterized by the first-order condition

$$C'(e) = \beta \cdot MV(\mathcal{S})$$

where $MV(\mathcal{S})$ denotes the marginal value of effort in subpopulation $\mathcal{S}$. Thus if $\beta \cdot MV(\mathcal{S}) < 1$, then effort is lower than the first-best level. Basing the forecast on an additional attribute decreases effort, and so makes equilibrium effort even more inefficient. In contrast, access to a new circumstance increases effort, which potentially moves effort closer to the first-best. An analogous line of reasoning applies when $\beta \cdot MV(\mathcal{S}) > 1$, reversing the roles of attributes and circumstances.

For any subpopulation $\mathcal{S}$, the marginal value of effort $MV(\mathcal{S})$ is bounded above by 1 under our regularity conditions, reflecting the fact that a 1-unit increase in output must lead to no more than a 1-unit inferred increase in θ. As a result, for any $\mathcal{S}$ and any new covariate, the bounds β_* and β^* appearing in Corollary 1 are both greater than 1. Given that our discount factor β is a reduced-form stand-in for any factors that may impact the value of

we consider here.

future reputation, we view both β larger and smaller than 1 as relevant in applications. In particular, β may be very large if the agent expects to enjoy reputational benefits over many future periods of service, or if he expects the scale of transactions to increase in the future. Conversely, β may be small if the agent participates in the market infrequently, or is unsure whether he will continue requiring service.

Suppose now that the new covariate is not a mean shifter, but satisfies the conditions of Theorem 2. Introduction of such a covariate moves the average effort level, but also introduces dispersion in effort across the subpopulation. This latter effect always lowers welfare due to the concavity of $W(e)$. On the other hand, the first effect can improve welfare, depending on whether the average effort level moves towards first best. When it does, the overall welfare change depends on which of these two opposing forces wins out, and it is not possible to systematically sign the effect of the new covariate.

When the average effect moves away from first-best, then introduction of the new covariate is always welfare-reducing. As a result, a regulator should disallow use of a new attribute when β is too low and disallow use of a new circumstance when β is high:

Corollary 2. *Suppose the new covariate is $\mathcal{S}$-affiliated. Then:*

(a) If the covariate is an attribute, there exists a threshold β_* such that it is (weakly) optimal for the regulator to forbid access when $\beta \leq \beta_*$.

(b) If the covariate is a circumstance, there exists a threshold β^* such that it is (weakly) optimal for the regulator to forbid access when $\beta \geq \beta^*$.

We conclude with two final remarks. First, holding fixed a change in average effort, covariates that induce very heterogeneous changes in effort across individuals are worse for welfare than ones that induce similar effort changes.[18] Second, although the analysis of this section has focused on a regulator's treatment of a single new covariate in isolation, in some cases the regulator may have the opportunity to jointly evaluate a set of new covariates. Attributes could then potentially be paired with circumstances to calibrate the overall effect of the approved covariates on effort. In particular, there may exist covariates which, on their own, reduce welfare, but which move effort closer to its first-best level when used jointly.[19]

[18] This result is complementary to findings of Frankel and Kartik (2020) and Ball (2020) that increased effort dispersion may reduce forecasting accuracy. In both their setting and ours, uncertainty about effort is bad for welfare.

[19] One risk of such a strategy is that the market may not actually use all approved covariates, so the

5 Extensions

5.1 Nonlinear Models

Our results so far have been developed in the context of an additive model, but they have natural analogues in a more general model that we now briefly outline.

Suppose θ and ε decompose as

$$\begin{aligned}\theta &= \mu + \Psi(\boldsymbol{a}) + \theta^{\perp} \\ \varepsilon &= \Lambda(\boldsymbol{c}) + \varepsilon^{\perp}.\end{aligned}$$

where $\Psi : \mathbb{R}^J \to \mathbb{R}$ and $\Lambda : \mathbb{R}^K \to \mathbb{R}$ are general (not necessarily additive) effect size functions. Without loss, again suppose $\mathbb{E}[\Psi(\boldsymbol{a})] = \mathbb{E}[\Lambda(\boldsymbol{c})] = 0$, taking $\Psi(\boldsymbol{a})$ and $\Lambda(\boldsymbol{c})$ to represent the de-meaned impact of the covariates on outcomes, and $\mu \in \mathbb{R}$ to represent the mean outcome. We generalize the notion of type and shock components as follows:

Definition 9. *Given a set of observed attributes $\mathcal{J}$, define*

$$\theta^{\mathcal{J}} \equiv \mathbb{E}[\Psi(\boldsymbol{a}) \mid \boldsymbol{a}_{\mathcal{J}}], \quad \theta^{-\mathcal{J}} \equiv \Psi(\boldsymbol{a}) - \theta^{\mathcal{J}} + \theta^{\perp}$$

to be the associated observed type component *and* residual type component*. Similarly, given a set of observed circumstances $\mathcal{K}$, define*

$$\varepsilon^{\mathcal{K}} \equiv \mathbb{E}[\Lambda(\boldsymbol{c}) \mid \boldsymbol{c}_{\mathcal{K}}], \quad \varepsilon^{-\mathcal{K}} \equiv \Lambda(\boldsymbol{c}) - \varepsilon^{\mathcal{K}} + \varepsilon^{\perp}$$

to be the associated observed shock component *and* residual shock component*.*

Given a set of observed covariates $(\mathcal{J}, \mathcal{K})$, the agent's type and the shock to his outcome can be decomposed as

$$\begin{aligned}\theta &= \mu + \theta^{\mathcal{J}} + \theta^{-\mathcal{J}} \\ \varepsilon &= \varepsilon^{\mathcal{K}} + \varepsilon^{-\mathcal{K}},\end{aligned}$$

where $\mu + \theta^{\mathcal{J}}$ and $\varepsilon^{\mathcal{K}}$ are observed while $\theta^{-\mathcal{J}}$ and $\varepsilon^{-\mathcal{K}}$ are not. Note that in this general

covariates that are incorporated into market forecasts may not improve welfare. By contrast, a case-by-case evaluation of new covariates avoids negative outcomes even if approved covariates are not actually used.

nonlinear framework, residual type and shock components are automatically "de-meaned", in a manner similar to the model with mean shifters.

For any new covariates $j' \notin \mathcal{J}$ and $k' \notin \mathcal{K}$, define

$$\theta_{j'}^{\mathcal{J}} \equiv \theta^{\mathcal{J}\cup\{j'\}} - \theta^{\mathcal{J}}$$

and

$$\varepsilon_{k'}^{\mathcal{K}} \equiv \varepsilon^{\mathcal{K}\cup\{k'\}} - \varepsilon^{\mathcal{K}}$$

to be the market's updates to the agent's expected type and shock when the additional attribute j' and circumstance k' are observed. Then θ and ε may be further decomposed as

$$\begin{aligned}
\theta &= \mu + \theta^{\mathcal{J}} + \theta_{j'}^{\mathcal{J}} + \theta^{-\mathcal{J}\cup\{j'\}} \\
\varepsilon &= \varepsilon^{\mathcal{K}} + \varepsilon_{k'}^{\mathcal{K}} + \varepsilon^{-\mathcal{K}\cup\{k'\}}.
\end{aligned}$$

Fix a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. In the general nonlinear model, we may define notions of regularity and $\mathcal{S}$-regularity exactly as in the additive environment, with type and shock components in those definitions interpreted according to the definitions just given. We may further define an attribute $j' \notin \mathcal{J}$ to be $\mathcal{S}$-*affiliated* if $\theta_{j'}^{\mathcal{J}}$ is a one-to-one-function of $a_{j'}$ (holding $\boldsymbol{a}_{\mathcal{J}}$ fixed) and $(\theta_{j'}^{\mathcal{J}}, \theta^{-\mathcal{J}\cup\{j'\}})$ are affiliated on $\mathcal{S}$. The notion of $\mathcal{S}$-affiliation for a circumstance is defined analogously. With these definitions in hand, the following result holds:

Theorem 3. *Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

- *If j' is $\mathcal{S}$-affiliated, further observing the attribute j' reduces the agent's effort on average.*
- *If k' is $\mathcal{S}$-affiliated, further observing the circumstance k' increases the agent's effort on average.*

The proof of this theorem follows exactly the same lines as the proof of Theorem 2. Despite its similarity to that theorem, however, Theorem 3 is in fact a generalization of Theorem 1, because the affiliation condition is applied to residual type and shock components which are de-meaned. This parallel can be sharpened by considering a related result under a stronger condition on covariates: say that an attribute $j' \notin \mathcal{J}$ is $\mathcal{S}$-*independent* if $\theta^{-\mathcal{J}\cup\{j'\}}$ is independent of $a_{j'}$ on $\mathcal{S}$, with an analogous definition of $\mathcal{S}$-independence for circumstances. Then the following result holds:

Theorem 4. *Fix any $\mathcal{S}$-regular attribute $j' \notin \mathcal{J}$ and $\mathcal{S}$-regular circumstance $k' \notin \mathcal{K}$. Then:*

- *If j' is $\mathcal{S}$-independent, observing the additional attribute j' reduces the agent's effort deterministically.*
- *If k' is $\mathcal{S}$-independent, observing the additional circumstance k' increases the agent's effort deterministically.*

Under $\mathcal{S}$-independence, mononotonicity holds not just on average but uniformly across realizations of the additional covariate, for the same reasons as in Theorem 1. In the special case where Ψ and Λ are additively separable, Theorems 1 and 4 exactly coincide.

5.2 Model Uncertainty and Misspecification

Our main results are established for an environment in which the agent is aware both of the covariates used by the market for forecasting, as well as the conditional distributions of types and shocks for agents with his covariate values. Effectively, the agent is perfectly aware of the market's forecasting model. Our results are robust to relaxation of this assumption.

In particular, suppose that the agent is subjectively uncertain about the market's statistical model of his subpopulation, where a model is comprised of a joint distribution over output and types for agents in the subpopulation. This uncertainty may be due to uncertainty over which covariates the market observes, as well as over other aspects of the market's perception of the relationship between covariates and outcomes. (It is not important that the agent's beliefs be correctly specified, in the sense that the market's true model need not be contained in the support of the agent's beliefs.) We will continue to maintain the assumption that the agent is not asymmetrically informed about his type, and so his own subjective belief about the distribution of his type and outcome is just the expectation of his belief about the market's model of his subpopulation.[20]

In this setting, all of our results extend in the following sense: If the agent becomes convinced that the market's statistical model has become "better-informed" about the agent's

[20]Thus, the agent does not attempt to use any private information about his covariates to refine his own beliefs about the distribution of output, consistent with a view that the agent is less knowledgeable about the forecasting model than the market.

type or shock, his effort will move in the direction predicted by our results, so long as the corresponding statistical assumptions hold for each model in the support of the agent's beliefs. More precisely, an agent believes the market has become "better-informed" if he thinks that, regardless of what statistical model it is in fact using, the market has gained access to additional attributes or additional covariates and refined its model accordingly. The robustness of our result in this environment follows directly from the fact that the marginal value of effort moves in the same direction conditional on any model in the support of the agent's beliefs, and therefore the expected marginal value of effort moves in this direction as well. Under model uncertainty, the agent's effort is determined by the discounted expected marginal value of effort, and so shifts in this quantity lead to effort shifts in the same direction.

5.3 General Convex Cost Functions

In the main text, we maintained the assumption that effort costs were quadratic: $C(e) = \frac{1}{2}e^2$. Under this cost structure, equilibrium effort is identical to the marginal value of effort, allowing us to characterize the former by analyzing the latter. More generally, when C is a strictly convex cost function, equilibrium effort is a uniquely determined, strictly increasing function of the marginal value of effort:

$$e^* = (C')^{-1}(MV),$$

where MV is the marginal value of effort (which is independent of e^*). As a result, for any strictly convex cost function, a deterministic shift in the marginal value of effort implies a change in effort in the same direction. This implies in particular that the results of sections 3.2 (for independent covariates) and 3.3 (for mean-shifters) extend immediately.

Our results for more general correlation structures become slightly more complex under non-quadratic cost functions. The new force which arises is that average effort may respond to mean-preserving spreads of the marginal value of effort. To illustrate the idea, consider any cost function $C(e) \propto e^k$, where $k > 2$. Under such a cost function the marginal cost of effort is convex, so that equilibrium effort is a concave function of the marginal value of effort. Hence any mean-preserving spread of the marginal value of effort reduces average effort. Conversely, if $2 > k > 1$, effort is a convex function of the marginal value of effort, and a mean-preserving spread of the marginal value of effort increases average effort.

In the case of general correlation (under the conditions of Theorem 2), observing an additional attribute has two effects: it lowers the *average* marginal value of effort, and additionally introduces a *spread* in the distribution of marginal values (relative to its baseline value in the subpopulation). If the marginal cost of effort is convex, these two forces work together to lower average effort; on the other hand, if the marginal cost of effort is sufficiently concave, average effort could increase. Analogous results hold when an additional circumstance is observed.

We view our results for the quadratic-cost case as a natural baseline for analyzing the impact of novel covariates which are "small" contributors to an agent's overall type or shock, in the sense that they don't change the marginal value of effort too much regardless of their realization. In that limit, the marginal cost of effort may be approximated to first order by a linear function, and the directional effects of adding a covariate identified by our main results will hold.

5.4 Dynamic Model

We have so far developed our formal results in the context of a 2-period model, in which the agent exerts effort once in order to influence a single reputational reward. This setup maps directly onto applications involving a one-shot interaction, such as college admissions. On the other hand, settings such as auto insurance or lending are more likely to consist of multiple interactions over time, with the agent receiving service and choosing effort to impact an outcome in each period.

Our model can be extended to accommodate multiple periods of service by assuming that, in each period $t = 1, ..., T$, the agent receives the market's period-t expectation of his outcome $\mathbb{E}_t[Y_t]$ as an up-front payment,[21] and then exerts effort to impact that period's outcome. Each Y_t is determined via

$$Y_t = e_t + \theta + \varepsilon_t,$$

where e_t is the agent's period-t effort and the shocks $\varepsilon_1, \varepsilon_2, ..., \varepsilon_T$ are identically distributed and mutually independent, corresponding to new circumstance realizations each period.

[21] This payment can be microfounded as arising from competition between service providers. See Appendix A for details.

Each attribute informs about θ as in the baseline model, while access to a circumstance means that this covariate is observed each period, and thus informs about each shock ε_t. (For example, access to weather data means that weather is observed for each time t.)[22] If $T = 2$, this model reduces to our baseline setting.

With more than 2 periods, the agent's effort in a given period t continues to be shaped by his desire to impact the payment $\mathbb{E}_{t+1}[Y_{t+1}] = e^*_{t+1} + \mathbb{E}_{t+1}[\theta]$, where e^*_{t+1} is the equilibrium effort level in period $t + 1$. In addition, however, effort in period t impacts payments in periods beyond $t + 1$, as a higher forecasted type yields dividends in all future periods. To formally extend our monotonicity results, it is sufficient to show that the sensitivity of forecasts at later dates responds to new covariates in the same way as the sensitivity of next period's forecast.

When type and shock components follow multivariate normal distributions, this result can be established by direct computation. In this setting, the type and shock in any subpopulation are ex ante normally distributed. Thus, as demonstrated in Holmström (1999),[23] the marginal value of effort in any period t is

$$MV_t = \beta\alpha_1^{(T-t)} + \beta^2\alpha_2^{(T-t)} + ... + \beta^{T-t}\alpha_{T-t}^{(T-t)},$$

where $\alpha_s^{(T-t)}$ are a series of positive coefficients capturing the marginal value of effort on the forecast in each period $t + s$. Holmström (1999) shows also that each $\alpha_s^{(T-t)}$ is decreasing in the precision of the time-t belief about θ and increasing in the precision of the belief about each shock. Since observing an additional attribute raises the precision of the belief about θ, while observing an additional circumstance raises the precision of the belief about each ε_t, our monotonicity results generalize for arbitrarily many periods.

Outside of the multivariate normal setting, it is technically challenging to determine the marginal impact of effort on type forecasts multiple periods ahead. We conjecture that the forces we have identified will continue to ensure monotonicity in a many-period setting, but leave a full exploration to future work.

[22] Our results would also extend if the market were able to observe the agent's circumstances in only a single period, using arguments similar to the ones outlined here.

[23] One subtlety of the current setting is that the agent may change subpopulations over time as his circumstances change. However, when error components are multivariate normal, every subpopulation has an error term which is identically distributed up to a mean shift, and so this detail does not change the value of effort.

6 Conclusion

As firms and governments move towards collecting large datasets of consumer transactions and behavior as inputs to decision-making, the question of whether and how to regulate the usage of consumer data has emerged as an important policy question. Recent regulations, such as the European Union's General Data Protection Regulation (GDPR), have focused on protecting consumer privacy and improving transparency regarding what kind of data is being collected. An important complementary consideration is how data impacts social and economic behaviors. In the present paper, we analyze one such factor—the effect that a market's access to novel covariates has on consumer incentives for effort. We show that "attribute" covariates, which inform about persistent quality, and "circumstance" covariates, which inform about an idiosyncratic shock, have opposing implications for effort. Thus it is important to distinguish between these two classes of covariates when regulating data usage.

We conclude by noting that while we have taken the set of covariates to be exogenously given, it would be interesting to understand which sets of covariates consumers and firms might strategically choose to reveal or acquire. For instance, gains from trade in insurance markets may be increased by limiting the number of attributes acquired by insurers, since this policy enables consumers to be insured against uncertainty about their long-run risk. But in a competitive market, firms may feel obliged to acquire all available attributes so as not to face an adversely selected pool if their competitors acquire such data and use it for screening purposes (a consideration not present in our framework). In such a situation, it becomes all the more important for a regulator to place restrictions on use of certain attributes for forecasting—not because these attributes are intrinsically invasive or unethical, but because they lead to poor economic outcomes. We leave the formalization of interesting issues such as these to future work.

Appendix

A Microfoundation

In this appendix we show that the model in the main text arises as the reduced form of a model in which firms compete to serve an agent.

The agent obtains service from a competitive market consisting of $J \geq 2$ homogeneous firms across two periods $t = 1, 2$. Service in each period generates a surplus equal to Y_t, which is collected by whichever firm serves the agent.[24] Each period's surplus is determined according to

$$Y_t = e_t + \mu + \theta + \varepsilon_t,$$

where e_t is the agent's effort choice, θ is distributed as in the model in the main text, and $\varepsilon_1, \varepsilon_2$ are drawn iid and distributed as in the model in the main text.

Firms compete to attract the agent in each period by offering up-front monetary rewards. The timeline of service in each period is as follows:

I. Firms simultaneously offer rewards.

II. The agent chooses a firm, receives the reward offered by their chosen firm, and exerts effort.

III. The outcome is realized and collected as profit by the firm providing service.

Rewards cannot be made contingent on the outcome of service, so firms cannot write incentive contracts. However, realized outcomes are public, and so firms may condition their offered rewards in period 2 on the realized outcome in period 1.

The agent's payoff in a given period is the reward he receives from his chosen firm, minus any effort costs he incurs to impact the outcome of service. That is, the agent's period-t payoff from selecting firm j and exerting effort e is

$$U_t = R_t^j - C(e),$$

[24] Nothing would change if the agent instead enjoyed a fixed surplus S from service, and the firm incurred a cost D_t to serve the agent, with $S - D_t = Y_t$. The key assumption is that all variation in outcomes arising from the agent's type and shock impacts the firm's payoff but not the agent's.

where R_t^j is firm j's offered reward to the agent in period t. The agent's total ex post payoff across both periods is a discounted sum of period payoffs:

$$U = U_1 + \beta \cdot U_2.$$

Each firm j receives a payoff in period t equal to zero if it does not serve the agent, and $Y_t - R_t^j$ otherwise. Firm j's total payoff across both periods is the discounted sum

$$\Pi^j = \Pi_1^j + \beta \cdot \Pi_2^j.$$

(Nothing would change if firms discounted the future at any other strictly positive rate.)

Equilibrium outcomes are determined under a mild restriction on off-path beliefs.

Definition A.1. *An* equilibrium *is a pure-strategy perfect Bayesian equilibrium in which firms believe the agent exerts the same period-1 effort regardless of his period-1 firm choice.*

This refinement is necessary because the agent chooses a firm and effort level simultaneously, and so sequential rationality imposes no restriction on firm beliefs about agent actions following an unexpected period-1 firm choice. As a result, there exist perfect Bayesian equilibria in the agent is "locked-in" to a firm offering an unfavorable reward due to very high conjectured effort following an off-path firm choice. The refinement of Definition A.1 eliminates such equilibria.

The following lemma, whose proof can be found in the online appendix, establishes that in this model, the agent's equilibrium period-2 payoff is exactly the market's posterior expectation of the agent's type. The agent also receives a payment in period 1 equal to his ex ante expected type plus his equilibrium effort level. Because that payment is independent of the agent's effort choice and outcome in period 1, incentives are unchanged if it is left out of the agent's utility function. With that normalization, the model just described reduces to the model in the main text.

Lemma A.1. *In equilibrium:*

- *The agent exerts effort e^* in period 1 and no effort in period 2,*
- *The highest rewards offered to the agent in each period are $R_1^* = \mu + e^*$ and $R_2^* = \mathbb{E}^{e^*}[\theta \mid Y_1]$.*

B Characterization of MV

Given a subpopulation $\mathcal{S}$, let e^* be equilibrium effort, and define

$$MV(\mathcal{S}) \equiv \frac{d}{de}\mathbb{E}^e\left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*}$$

to be the agent's equilibrium marginal value of effort, where $\mathbb{E}^{\hat{e}}[\cdot]$ denotes expectations under the measure in which $Y = \hat{e} + \theta + \varepsilon$ almost surely.

Lemma B.1. *Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and any $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S}$. Then the equilibrium marginal value of effort on $\mathcal{S}$ is*

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \mid \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J}\theta_j + \theta^{\perp} + \sum_{k=1}^{K}\varepsilon_k + \varepsilon^{\perp}.$$

Proof. Fix a subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. Let $h(y \mid t, e)$ be the conditional density of $Y \mid \theta^{-\mathcal{J}}, e$ on $\mathcal{S}$ and $h(y \mid e)$ be the conditional density of $Y \mid e$ on $\mathcal{S}$. Because effort affects output as an additive shift, $h(y \mid t, e) = h(y - e \mid t, 0)$ and $h(y \mid e) = h(y - e \mid 0)$ for every (y, t, e). So let $f(t \mid y, e)$ be the conditional distribution of $\theta^{-\mathcal{J}} \mid Y, e$ on $\mathcal{S}$, and let $f^0(t)$ be the conditional distribution of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$. Then by Bayes' rule,

$$f(t \mid y, e) = \frac{h(y \mid t, e) f^0(t)}{h(y \mid e)} = \frac{h(y - e \mid t, 0) f^0(t)}{h(y - e \mid 0)} = f(t \mid y - e, 0).$$

Hence

$$\begin{aligned}
\mathbb{E}^{e^*}[\theta \mid Y = y, \mathcal{S}] &= \sum_{j \in \mathcal{J}} \alpha_j + \mathbb{E}^{e^*}[\theta^{-\mathcal{J}} \mid Y = y, \mathcal{S}] \\
&= \sum_{j \in \mathcal{J}} \alpha_j + \int t\, h(t \mid y, e^*)\, dt \\
&= \sum_{j \in \mathcal{J}} \alpha_j + \int t\, h(t \mid y - e^*, 0)\, dt \\
&= \sum_{j \in \mathcal{J}} \alpha_j + \mathbb{E}^0[\theta^{-\mathcal{J}} \mid Y = y - e^*, \mathcal{S}].
\end{aligned}$$

Under the measure corresponding to $e = 0$, the variable Y is equal to Y^0 almost surely. So

$$\mathbb{E}^0[\theta^{-\mathcal{J}} \mid Y = y - e^*, \mathcal{S}] = \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}].$$

Now,

$$\begin{aligned}
\mathbb{E}^e \left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right] &= \int dy\, h(y \mid e)\mathbb{E}^{e^*}[\theta \mid Y = y, \mathcal{S}] \\
&= \int dy\, h(y \mid e) \left(\sum_{j\in\mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}]\right) \\
&= \int dy\, h(y - e \mid 0) \left(\sum_{j\in\mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y - e^*, \mathcal{S}]\right),
\end{aligned}$$

and so by making the variable substitution $y' = y - e$ we may write

$$\mathbb{E}^e \left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right] = \int dy'\, h(y' \mid 0) \left(\sum_{j\in\mathcal{J}} \alpha_j + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y' - e^* + e, \mathcal{S}]\right).$$

Differentiating wrt e and invoking Assumption 1 so that the dominated convergence theorem may be applied yields

$$\frac{\partial}{\partial e}\mathbb{E}^e \left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*} = \int dy'\, h(y' \mid 0)\, \frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}]\Big|_{Y^0=y'}.$$

Recall that $Y \overset{d}{=} Y^0$ conditional on $e = 0$, so $h(y' \mid 0)$ is the conditional density of Y^0 on $\mathcal{S}$. The rhs of the previous expression may therefore be written

$$\frac{\partial}{\partial e}\mathbb{E}^e \left[\mathbb{E}^{e^*}[\theta \mid Y, \mathcal{S}] \mid \mathcal{S}\right]\Big|_{e=e^*} = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \,\Big|\, \mathcal{S}\right],$$

as desired. □

C Proof of Theorem 2

We prove part (a) here. The proof of part (b) follows along very similar lines, and is relegated to Online Appendix F. In Section C.1 we use the assumption that $\Psi^{j'}$ is one-to-one to replace observation of attribute $a_{j'}$ with observation of the corresponding type component $\theta_{j'}$. The desired result then holds if can we show that the conditional forecast of the residual unknown $\theta^{-\mathcal{J}}$ is (in expectation) less responsive to the realization of the outcome Y when conditioned on $\theta_{j'}$.

In Section C.2, we prove an important intermediate lemma: $(\theta^{-\mathcal{J}}, \theta_{j'}, Y)$ are affiliated, so that holding fixed any value of Y, higher realizations of $\theta_{j'}$ lead to higher conditional inferences about $\theta^{-\mathcal{J}}$. This result follows from our maintained regularity and affiliation assumptions. In Section C.3, we consider the effect of effort under the baseline covariates (where $\theta_{j'}$ is not observed) and separate the impact of increasing Y into two effects: a shift up in the conditional distribution of $\theta_{j'}$ and, for each realization of $\theta_{j'}$, a shift up in the conditional distribution of $\theta^{-\mathcal{J}}$. The affiliation lemma tells us that these effects amplify one another, leading to a higher marginal impact of manipulation of Y. In contrast, when $\theta_{j'}$ is observed, it must follow its true distribution irrespective of the manipulation of Y. Larger Y still shifts up the conditional distribution of $\theta^{-\mathcal{J}}$, but we show that this alone is a smaller effect.

C.1 Preliminaries

Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$.

As established in Lemma B.1, the marginal value of effort in subpopulation $\mathcal{S}$ is

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \;\middle|\; \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J} \theta_j + \theta^{\perp} + \sum_{k=1}^{K} \varepsilon_k + \varepsilon^{\perp}$$

is the baseline value of output after subtracting out the agent's effort.

Now suppose the market additionally observes the additional attribute $j' \notin \mathcal{J}$, and let $\mathcal{J}' = \mathcal{J} \cup \{j'\}$. Under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] \;\middle|\; a_{j'}, \mathcal{S}\right],$$

where $\mathbb{E}[\cdot \mid a_{j'} = \alpha_{j'}, \mathcal{S}]$ denotes expectations with respect to the distributions of θ and ε in the subpopulation of $\mathcal{S}$ whose value of covariate j' is $\alpha_{j'}$. Note that on $\mathcal{S}$, $MV(\mathcal{S}, a_{j'})$ is a random variable whose value is a function of the realization of $a_{j'}$.

Because $\Psi^{j'}$ is a one-to-one mapping, the subpopulation of $\mathcal{S}$ whose value of covariate j' is $\alpha_{j'}$ is identical to the subpopulation whose type component $\theta_{j'}$ has value $\Psi(\alpha_{j'})$. So we may equivalently write the agent's marginal value of effort under the expanded set of covariates as

$$MV(\mathcal{S}, \theta_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \theta_{j'}, \mathcal{S}\right],$$

where $\mathbb{E}[\cdot \mid \theta_{j'}, \mathcal{S}]$ is interpreted analogously to $\mathbb{E}[\cdot \mid a_{j'}, \mathcal{S}]$, and $MV(\mathcal{S}, \theta_{j'})$ is a random variable which depends on $a_{j'}$ only via the type component $\theta_{j'}$.

C.2 Affiliation Lemma

Lemma C.1. *$(\theta^{-\mathcal{J}}, \theta_{j'}, Y^0)$ are jointly affiliated on $\mathcal{S}$.*

Proof. Let $f(u, t, y)$ be the conditional joint density of $(\theta^{-\mathcal{J}}, \theta_{j'}, Y^0)$ on $\mathcal{S}$. We will show that f is log-supermodular.

Let $f_{j'}(t)$ be the conditional density of $\theta_{j'}$ on $\mathcal{S}$, $f_{-\mathcal{J}|j'}(u \mid t)$ be the conditional density of $\theta^{-\mathcal{J}} \mid \theta_{j'}$ on $\mathcal{S}$, and $h_{Y|-\mathcal{J}}(y \mid u)$ be the conditional density of $Y^0 \mid \theta^{-\mathcal{J}}$ on $\mathcal{S}$. Note that Y^0 is independent of $\theta_{j'}$ conditional on $\theta^{-\mathcal{J}}$ on $\mathcal{S}$, and so

$$f(u, t, y) = f_{j'}(t) f_{-\mathcal{J}|j'}(u \mid t) h_{Y|-\mathcal{J}}(y \mid u).$$

It is therefore sufficient to show that $h_{Y|-\mathcal{J}}$ and $f_{-\mathcal{J}|j'}$ are log-supermodular.

Consider $h_{Y|-\mathcal{J}}$. Let $\mu_{\mathcal{S}} \equiv \sum_{j\in\mathcal{J}} \alpha_j + \sum_{k\in\mathcal{K}} \gamma_k + \theta^{-\mathcal{J}}$. On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Let $g_{-\mathcal{K}}(z)$ be the conditional density of $\varepsilon^{-\mathcal{K}}$ on $\mathcal{S}$. Then

$$h_{Y|-\mathcal{J}}(y \mid u) = g_{-\mathcal{K}}(y - \mu_{\mathcal{S}} - u).$$

The assumption that $\mathcal{S}$ is a regular subpopulation implies that $g_{-\mathcal{K}}$ is log-concave, therefore $h_{Y|-\mathcal{J}}$ is log-supermodular.

As for $f_{-\mathcal{J}|j'}$, let $f_{-\mathcal{J}'|j'}(w \mid t)$ be the conditional density of $\theta^{-\mathcal{J}'} \mid \theta_{j'}$ on $\mathcal{S}$. As $\theta^{-\mathcal{J}} =$

$\theta_{j'} + \theta^{-\mathcal{J}'}$, it follows that

$$f_{-\mathcal{J}|j'}(u \mid t) = f_{-\mathcal{J}'|j'}(u - t \mid t).$$

Hence by the chain rule,

$$\frac{\partial^2}{\partial u \partial t} \log f_{-\mathcal{J}|j'}(u \mid t) = \left[\frac{\partial^2}{\partial w \partial t} \log f_{-\mathcal{J}'|j'}(w \mid t) - \frac{\partial^2}{\partial w^2} \log f_{-\mathcal{J}'|j'}(w \mid t) \right]_{w=u-t}.$$

As attribute j' is $\mathcal{S}$-regular, the second term is non-negative. Meanwhile since j' is $\mathcal{S}$-affiliated, $(\theta^{-\mathcal{J}'}, \theta_{j'})$ are affiliated on $\mathcal{S}$ and so the first term is also non-negative. Hence

$$\frac{\partial^2}{\partial u \partial t} \log f_{-\mathcal{J}|j'}(u \mid t) \geq 0,$$

establishing the desired log-supermodularity. □

C.3 Comparison of MV

$MV(\mathcal{S}, \theta_{j'})$ can be compared to $MV(\mathcal{S})$ as follows. Define

$$F(t \mid y) \equiv \Pr(\theta_{j'} \leq t \mid Y^0 = y, \mathcal{S})$$

to be the conditional distribution function of $\theta_{j'}$ given the outcome Y^0, and

$$\phi(y, t) \equiv \mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0 = y, \theta_{j'} = t, \mathcal{S}]$$

to be the conditional expectation of $\theta^{-\mathcal{J}'}$ given Y^0 and $\theta_{j'}$.

By the law of total probability

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] = \int dF(t \mid y) \, (t + \phi(y, t))$$

and so the change in the conditional expectation of the unobserved $\theta^{-\mathcal{J}}$ as Y^0 moves from y to $y' > y$ is

$$\begin{aligned} &\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y', \mathcal{S}] - \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] \\ &= \int dF(t \mid y') \, (t + \phi(y', t)) - \int dF(t \mid y) \, (t + \phi(y, t)) \end{aligned} \tag{C.1}$$

This difference can be signed using Lemma C.1: Since $(\theta^{-\mathcal{J}}, \theta_{j'})$ are affiliated, the expression $t + \phi(y,t)$ is increasing in t. Since $(\theta_{j'}, Y^0)$ are affiliated, $\mathbb{E}[\pi(\theta_{j'}) \mid Y^0, \mathcal{S}]$ is increasing in Y^0 for any increasing function π. Thus

$$\int dF(t \mid y')\,(t + \phi(y,t))$$

is increasing in y', and so the expression in (C.1) can be bounded below by

$$\int dF(t \mid y)\,(\phi(y',t) - \phi(y,t))\,.$$

It follows that

$$\frac{\partial}{\partial y}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \mathcal{S}] \geq \int dF(t \mid y)\frac{\partial \phi}{\partial y}(y,t).$$

The lhs is the marginal improvement in the posterior expectation of $\theta^{-\mathcal{J}}$ when the realization of Y^0 is increased. The rhs is the expected marginal improvement in the posterior expectation of $\theta^{-\mathcal{J}'}$ when it is conditioned on the manipulated realization of Y^0 and the *un-manipulated* realization of $\theta_{j'}$.

To complete the proof, rewrite this inequality as:

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right].$$

Taking the expectation of each side conditional on $\mathcal{S}$ yields

$$MV(\mathcal{S}) \geq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right].$$

By the law of iterated expectations, the rhs may be expanded as

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \mathcal{S}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \theta_{j'}, \mathcal{S}] \;\middle|\; \theta_{j'}, \mathcal{S}\right] \;\middle|\; \mathcal{S}\right] \\
&= \mathbb{E}\left[MV(\mathcal{S}, \theta_{j'}) \mid \mathcal{S}\right].
\end{aligned}$$

Therefore

$$MV(\mathcal{S}) \geq \mathbb{E}[MV(\mathcal{S}, \theta_{j'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is higher than the expected marginal value once attribute j' is additionally observed.

D Proof of Theorem 1

We again prove part (a), leaving the proof of part (b) to Online Appendix F.

The structure of the proof is very similar to that of Theorem 2, but $(\theta^{-\mathcal{J}}, \theta_{j'}, Y)$ is no longer guaranteed to satisfy affiliation, so Lemma C.2 may fail.[25] This precludes an argument along identical lines. Instead, in Section D.1, we define a decomposition of the residual component $\theta^{-\mathcal{J}}$ into the sum of two random variables $\tilde{\theta}_{j'}$ and $\tilde{\theta}^{-\mathcal{J}'}$, where the new attribute $a_{j'}$ perfectly reveals the former but is independent of the latter. Rather than studying the impact of observing the type component $\theta_{j'}$, we characterize the equivalent impact of observing $\tilde{\theta}_{j'}$. We then show that $(\theta^{-\mathcal{J}}, \tilde{\theta}_{j'}, Y)$ are affiliated. At this point arguments used in the proof of Theorem 2 can be applied without modification, and we further strengthen these to show deterministic impact.

D.1 Decomposition

Fix a set of observed covariates $(\mathcal{J}, \mathcal{K})$ and a $(\mathcal{J}, \mathcal{K})$-subpopulation $\mathcal{S} = (\mathcal{J}, \mathcal{K}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$. As established in Lemma B.1, the marginal value of effort in subpopulation $\mathcal{S}$ is

$$MV(\mathcal{S}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \;\middle|\; \mathcal{S}\right],$$

where

$$Y^0 \equiv \mu + \sum_{j=1}^{J} \theta_j + \theta^{\perp} + \sum_{k=1}^{K} \varepsilon_k + \varepsilon^{\perp}$$

is the baseline value of output after subtracting out the agent's effort.

Now suppose the market additional observes the additional attribute $j' \notin \mathcal{J}$, and let $\mathcal{J}' = \mathcal{J} \cup \{j'\}$. Under the expanded set of observed covariates, the marginal value of effort

[25] For a simple example, set $\theta_{j'} = -\theta^{-\mathcal{J}'}$. Then $\theta_{j'}$ is an $\mathcal{S}$-mean shifter but it is not $\mathcal{S}$-affiliated.

becomes

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] \,\middle|\, a_{j'}, \mathcal{S}\right],$$

where, on $\mathcal{S}$, $MV(\mathcal{S}, a_{j'})$ is a random variable whose value is a function of the realization of $a_{j'}$.

On $\mathcal{S}$, the outcome Y^0 may be decomposed as

$$Y^0 = \mu_{\mathcal{S}} + \Psi^{j'}(a_{j'}) + \theta^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}, \tag{D.1}$$

where $\mu_{\mathcal{S}} \equiv \sum_{j\in\mathcal{J}} a_j + \sum_{k\in\mathcal{K}} \gamma_k$ is a known constant while $\Psi^{j'}(a_{j'})$, $\theta^{-\mathcal{J}'}$ and $\varepsilon^{-\mathcal{K}}$ are random variables. We will now de-mean the unknown type residual $\theta^{-\mathcal{J}'}$.

For each possible realization $\alpha_{j'}$ of the attribute $a_{j'}$, define

$$\xi(\alpha_{j'}) \equiv \mathbb{E}[\theta^{-\mathcal{J}'} \mid a_{j'} = \alpha_{j'}, \mathcal{S}]$$

to be the conditional expectation of the residual $\theta^{-\mathcal{J}'}$ when $a_{j'}$ takes value $\alpha_{j'}$. Let $\tilde{\theta}_{j'} \equiv \Psi^{j'}(a_{j'}) + \xi(a_{j'})$ be the sum of the j-th type component and this posterior mean. Then we can rewrite (D.1) as

$$Y^0 = \mu_{\mathcal{S}} + \tilde{\theta}_{j'} + \tilde{\theta}^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}$$

where (by assumption that j' is a $\mathcal{S}$-mean shifter) the new residual $\tilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ on $\mathcal{S}$, while $\mu_{\mathcal{S}} + \tilde{\theta}_{j'}$ is a known constant given $a_{j'}$ (and hence independent of Y^0). So the expectation of the type residual $\theta^{-\mathcal{J}'}$ given $a_{j'}$ is

$$\begin{aligned}
\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] &= \tilde{\theta}_{j'} + \mathbb{E}[\tilde{\theta}^{-\mathcal{J}'} \mid Y^0, a_{j'}, \mathcal{S}] \\
&= \tilde{\theta}_{j'} + \mathbb{E}[\tilde{\theta}^{-\mathcal{J}'} \mid Y^0, \tilde{\theta}_{j'}, \mathcal{S}]
\end{aligned}$$

and $MV(\mathcal{S}, a_{j'})$ may be written

$$MV(\mathcal{S}, a_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\tilde{\theta}^{-\mathcal{J}'} \mid Y^0, \tilde{\theta}_{j'}, \mathcal{S}] \,\middle|\, a_{j'}, \mathcal{S}\right].$$

Note that the random variable inside the outer expectation depends on $a_{j'}$ only through $\tilde{\theta}_{j'}$. Thus the marginal value of effort after observing j' depends on the realization of $a_{j'}$ only through $\tilde{\theta}_{j'}$, and so we may denote this marginal value of effort

$$MV(\mathcal{S}, \tilde{\theta}_{j'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\tilde{\theta}^{-\mathcal{J}'} \mid Y^0, \tilde{\theta}_{j'}, \mathcal{S}] \;\middle|\; \tilde{\theta}_{j'}, \mathcal{S}\right].$$

The desired result holds if can we show that the conditional expectation of $\tilde{\theta}^{-\mathcal{J}'}$ is less responsive to the realization of the outcome Y than the conditional expectation of the original residual $\theta^{-\mathcal{J}}$. Note that $\theta^{-\mathcal{J}}$ is the sum of the (conditionally) independent variables $\tilde{\theta}_{j'}$ and $\tilde{\theta}^{-\mathcal{J}'}$, so uncertainty about $\tilde{\theta}^{-\mathcal{J}'}$ is mechanically lower than uncertainty about $\theta^{-\mathcal{J}}$. But this does not directly translate into a statement that the posterior expectation of $\tilde{\theta}^{-\mathcal{J}'}$ is less sensitive to the realization of Y. In general, we are not even guaranteed that higher realizations of Y lead to higher inferences about $\tilde{\theta}^{-\mathcal{J}}$ once we have conditioned on the realization of $\tilde{\theta}_{j'}$.[26] In the next section, we prove a key technical lemma, which will imply an analogue of regularity for our transformed environment.

D.2 Affiliation Lemma

Lemma D.1. $(\theta^{-\mathcal{J}}, \tilde{\theta}_{j'}, Y^0)$ *are jointly affiliated on* $\mathcal{S}$.

Proof. Let $\tilde{f}(u, t, y)$ be the conditional joint density of $(\theta^{-\mathcal{J}}, \tilde{\theta}_{j'}, Y^0)$ on $\mathcal{S}$. We will show that $\tilde{f}$ is log-supermodular.

Use $\tilde{f}_{j'}(t)$ to denote the conditional density of $\tilde{\theta}_{j'}$ on $\mathcal{S}$, $\tilde{f}_{-\mathcal{J}|j'}(u \mid t)$ to denote the conditional density of $\theta^{-\mathcal{J}} \mid \tilde{\theta}_{j'}$ on $\mathcal{S}$, and $h_{Y|-\mathcal{J}}(y \mid u)$ to denote the conditional density of $Y^0 \mid \theta^{-\mathcal{J}}$ on $\mathcal{S}$. Note that Y^0 is independent of $\tilde{\theta}_{j'}$ conditional on $\theta^{-\mathcal{J}}$ on $\mathcal{S}$. So $\tilde{f}$ may be decomposed as

$$\tilde{f}(u, t, y) = \tilde{f}_{j'}(t)\tilde{f}_{-\mathcal{J}|j'}(u \mid t)h_{Y|-\mathcal{J}}(y \mid u).$$

It is therefore sufficient to show that $h_{Y|-\mathcal{J}}$ and $\tilde{f}_{-\mathcal{J}|j'}$ are log-supermodular.

First consider $h_{Y|-\mathcal{J}}$. Recall that on $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

[26] Recall that our regularity assumptions are imposed on the original type component $\theta_{j'}$, and not on the constructed $\tilde{\theta}_{j'}$.

Let $g_{-\mathcal{K}}(z)$ be the conditional density of $\varepsilon^{-\mathcal{K}}$ on $\mathcal{S}$. Then

$$h_{Y|-\mathcal{J}}(y \mid u) = g_{-\mathcal{K}}(y - \mu_{\mathcal{S}} - u).$$

The assumption that $\mathcal{S}$ is a regular subpopulation implies that $g_{-\mathcal{K}}$ is log-concave, therefore $h_{Y|-\mathcal{J}}$ is log-supermodular.

As for $\tilde{f}_{-\mathcal{J}|j'}$, let $\tilde{f}_{-\mathcal{J}'}(w)$ be the conditional density of $\tilde{\theta}^{-\mathcal{J}'}$ on $\mathcal{S}$. Decompose $\theta^{-\mathcal{J}}$ as $\theta^{-\mathcal{J}} = \tilde{\theta}_{j'} + \tilde{\theta}^{-\mathcal{J}'}$, and recall that if j' is an $\mathcal{S}$-mean shifter then $\tilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ and hence $\tilde{\theta}_{j'}$ on $\mathcal{S}$. It follows that

$$\tilde{f}_{-\mathcal{J}|j'}(u \mid t) = \tilde{f}_{-\mathcal{J}'}(u - t),$$

and hence

$$\frac{\partial^2}{\partial u \partial t} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) = -\left.\frac{\partial^2}{\partial w^2} \log \tilde{f}_{-\mathcal{J}'}(w)\right|_{w=u-t} = -\frac{\partial^2}{\partial u^2} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t).$$

Now, let $f^0_{-\mathcal{J}|j'}(u \mid \alpha_{j'})$ denote the conditional density of $\theta^{-\mathcal{J}} \mid a_{j'}$ on $\mathcal{S}$. Define

$$\zeta(\alpha_{j'}) = \Psi^{j'}(\alpha_{j'}) + \xi(\alpha_{j'})$$

so that

$$\theta^{-\mathcal{J}} = \zeta(a_{j'}) + \tilde{\theta}^{-\mathcal{J}'},$$

If j' is an $\mathcal{S}$-mean shifter,

$$f^0_{-\mathcal{J}|j'}(u \mid \alpha_{j'}) = \tilde{f}_{-\mathcal{J}'}(u - \zeta(\alpha_{j'})) = \tilde{f}_{-\mathcal{J}|j'}(u \mid \zeta(\alpha_{j'})).$$

Let $\tilde{\theta} \equiv \{t \ : \ \zeta(\alpha_{j'}) = t \text{ for some } \alpha_{j'} \in A_{j'}\}$ denote the support of $\tilde{\theta}_{j'}$. Fix any $t \in \tilde{\theta}$, and let $\alpha_{j'} \in A_{j'}$ be such that $\zeta(\alpha_{j'}) = t$. Then for all u,

$$\frac{\partial^2}{\partial u^2} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) = \frac{\partial^2}{\partial u^2} \log f^0_{-\mathcal{J}|j'}(u \mid \alpha_{j'}).$$

As attribute j' is $\mathcal{S}$-regular, this final expression is non-positive, meaning

$$\frac{\partial^2}{\partial u \partial t} \log \tilde{f}_{-\mathcal{J}|j'}(u \mid t) \geq 0$$

for every u and $t \in \tilde{\theta}$. Hence $\tilde{f}_{-\mathcal{J}|j'}$ is log-supermodular, as desired. $\square$

D.3 Comparison of MV

Following arguments identical to those used in Section C.3 for the proof of Theorem 2 (with $\tilde{\theta}^{-\mathcal{J}'}$ and $\tilde{\theta}_{j'}$ everywhere replacing $\theta^{-\mathcal{J}'}$ and $\theta_{j'}$), Lemma D.1 implies

$$MV(\mathcal{S}) \geq \mathbb{E}[MV(\mathcal{S}, \tilde{\theta}_{j'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is higher than the expected marginal value once attribute j' is additionally observed.

To complete the proof, we must establish that monotonicity holds uniformly across realizations of the additional covariate, and not just on average. This follows immediately once we establish that $MV(\mathcal{S}, a_{j'})$ is independent of the realization of $a_{j'}$. On $\mathcal{S}$, when attribute j' is additionally observed and is found to have value $a_{j'} = \alpha_{j'}$, Y may be decomposed as

$$Y = e + \mu_{\mathcal{S}} + \Psi^{j'}(\alpha_{j'}) + \xi(\alpha_{j'}) + \tilde{\theta}^{-\mathcal{J}'} + \varepsilon^{-\mathcal{K}}.$$

Because $\tilde{\theta}^{-\mathcal{J}'}$ is independent of $a_{j'}$ on $\mathcal{S}$, $\alpha_{j'}$ enters the market's inference problem as a known additive shift to the agent's type distribution, and therefore its value does not impact incentives for effort. So incentives for effort must be independent of $\alpha_{j'}$, as claimed.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2020): "Too Much Data: Prices and Inefficiencies in Data Markets," Working Paper.

Acquisti, A., L. Brandimarte, and G. Loewenstein (2015): "Privacy and Human Behavior in the Age of Information," *Science*, 347, 509–514.

Agarwal, A., M. Dahleh, and T. Sarkar (2019): "A Marketplace for Data: An Algorithmic Solution," in *Proceedings of the 2019 ACM Conference on Economics and Computation*, 701–726.

Ball, I. (2020): "Scoring Strategic Agents," Working Paper.

Bergemann, D., A. Bonatti, and T. Gan (2020): "The Economics of Social Data," Working Paper.

Bergemann, D., A. Bonatti, and A. Smolin (2018): "The Design and Price of Information," *American Economic Review*, 108, 1–48.

Bonatti, A. and G. Cisternas (2019): "Consumer Scores and Price Discrimination," *Review of Economic Studies*, forthcoming.

Braverman, M. and S. Chassang (2020): "Data-Driven Incentive Alignment in Capitation Schemes," Working Paper.

Brunnermeier, M., R. Lamba, and C. Segura-Rodriguez (2020): "Inverse Selection," Working Paper.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 66, 183–198.

Dwork, C. and A. Roth (2014): "The Algorithmic Foundations of Differential Privacy," *Found. Trends Theor. Comput. Sci.*, 9, 211–407.

Eilat, R., K. Eliaz, and X. Mu (2020): "Bayesian Privacy," *Theoretical Economics*, Forthcoming.

Eliaz, K. and R. Spiegler (2019): "The Model Selection Curse," *American Economic Review: Insights*, 1, 127–140.

——— (2020): "On Incentive-Compatible Estimators," Working Paper.

Elliott, M., A. Galeotti, and A. Koh (2020): "Market segmentation through information," Working Paper.

Fainmesser, I. P., A. Galeotti, and R. Momot (2020): "Digital Privacy," Working Paper.

Frankel, A. and N. Kartik (2019): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776.

——— (2020): "Improving Information from Manipulable Data," Working Paper.

Gomes, R. and A. Pavan (2019): "Price Customization and Targeting in Matching Markets," Working Paper.

Haghtalab, N., N. Immorlica, B. Lucier, and J. Wang (2020): "Maximizing Welfare with Incentive-Aware Evaluation Mechanisms," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 160–166.

Hennessy, C. A. and C. A. E. Goodhart (2020): "Goodhart's Law and Machine Learning," Working Paper.

Hidir, S. and N. Vellodi (2020): "Privacy, Personalization and Price Discrimination," *Journal of the European Economic Association*, Forthcoming.

Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182.

Hu, L., N. Immorlica, and J. W. Vaughan (2019): "The Disparate Effects of Strategic Manipulation," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 259–268.

Ichihashi, S. (2019): "Online Privacy and Information Disclosure by Consumers," *American Economic Review*, 110, 569–595.

Jin, Y. and S. Vasserman (2020): "Buying Data from Consumers: The Impact of Monitoring Programs in U.S. Auto Insurance," Working Paper.

Jullien, B., Y. Lefouili, and M. H. Riordan (2020): "Privacy Protection, Security, and Consumer Retention," Working Paper.

Meyer, M. A. and J. Vickers (1997): "Performance Comparisons and Dynamic Incentives," *Journal of Political Economy*, 105, 547–581.

Rodina, D. (2018): "Information Design and Career Concerns," Working Paper.

Segura-Rodriguez, C. (2020): "Selling Data," Working Paper.

Tirole, J. (2020): "Digital Dystopia," *American Economic Review*, Forthcoming.

Yang, K. H. (2020): "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and its Consequences," Working Paper.

Online Appendix

E Additional Material

E.1 Supplementary Material for Section 3.1

We provide some examples below of individual type or shock components that are log-concave. This list is not meant to be exhaustive.

Example 6. The attribute a is height, it is normally distributed, and $\Psi(a)$ is affine.

Example 7. The attribute a is a one-dimensional location variable which is uniformly distributed on an interval $[c, d]$, while $\Psi(a) = a - x$ is (signed) distance from a fixed point $x \in [c, d]$.

Example 8. The attribute a is the expected number of friends that one can borrow money from, and it is exponentially distributed, while $\Psi(a) = \sqrt{a}$.

Example 9. The attribute a is days between social media posts, and it has a gamma distribution, while $\Psi(a) = \log(a)$.

Example 10. c is the number of inches of precipitation last month, and it has an exponential distribution, while $\Lambda(c) = -\log(c)$.

E.2 Supplementary Material for Example 1

We show below that when type components and shock components are jointly normal, then the mean-shifter property is satisfied globally. First consider the two-attribute model

$$Y = e + \theta_1 + \theta_2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

with

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$

Using standard formulas for Bayesian updating to Gaussian signals, the conditional distribution of θ_1 given θ_2 is

$$\theta_1 \mid \theta_2 \sim \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\theta_2 - \mu_2) + \mathcal{N}\left(0, \sigma_1^2(1-\rho^2)\right), \tag{E.1}$$

which depends on θ_2 in mean only. Since the family of normal variables is closed under conditioning and summation, the above model is without loss: that is, for any $\mathcal{J}$ and $j' \notin \mathcal{J}$, we may set θ_1 equal to $\theta^{-\mathcal{J}'}$ and θ_2 equal to $\theta_{j'}$, so that (E.1) implies that attribute j' is an $\mathcal{S}$-mean shifter. (The argument applies identically for circumstance variables.)

E.3 Supplementary Material for Example 5

A standard result about sums of iid exponential random variables is that they follow a Gamma distribution. In particular, if there are J total attributes, then $\theta^{-\mathcal{J}'}|\lambda \sim \text{Gamma}(J - |\mathcal{J}'|, \lambda)$, and $\theta^{-\mathcal{J}'}$ and $\theta_{j'}$ are independent conditional on λ.

Define $k \equiv \lambda^{-1}$. The joint density of $(\theta_{j'}, \theta^{-\mathcal{J}'})$ conditional on $\boldsymbol{\theta}_{\mathcal{J}}$ may be written

$$\eta(\theta_{j'}, \theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_{\mathcal{J}}) = \int dk\, \eta(k \mid \boldsymbol{\theta}_{\mathcal{J}})\eta(\theta_{j'} \mid k)\eta(\theta^{-\mathcal{J}'} \mid k).$$

As marginalization preserves affiliation, $(\theta_{j'}, \theta^{-\mathcal{J}'})$ are affiliated conditional on $\boldsymbol{\theta}_{\mathcal{J}}$ if $(\theta_{j'}, k)$ and $(\theta^{-\mathcal{J}'}, k)$ are each affiliated. Note that

$$\log \eta(\theta_{j'} \mid k) = -\log k - \frac{\theta_{j'}}{k},$$

so that

$$\frac{\partial^2}{\partial \theta_{j'} \partial k} \log \eta(\theta_{j'} \mid k) = \frac{1}{k^2} > 0,$$

while

$$\log \eta(\theta^{-\mathcal{J}'} \mid k) = -N \log k - \log \Gamma(N) + (N-1) \log \theta^{-\mathcal{J}'} - \frac{\theta^{-\mathcal{J}'}}{k}$$

for $N = J - |\mathcal{J}'|$, so that similarly

$$\frac{\partial^2}{\partial \theta^{-\mathcal{J}'} \partial k} \log \eta(\theta^{-\mathcal{J}'} \mid k) = \frac{1}{k^2} > 0.$$

Hence $(\theta_{j'}, k)$ and $(\theta^{-\mathcal{J}'}, k)$ are affiliated, as desired.

Meanwhile, the conditional density of $\theta^{-\mathcal{J}'}$ may be written

$$\eta(\theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_{\mathcal{J}'}) = \int d\lambda\, \eta(\lambda \mid \boldsymbol{\theta}_{\mathcal{J}'})\eta(\theta^{-\mathcal{J}'} \mid \lambda).$$

Log-concavity is also preserved by marginalization, so $\eta(\theta^{-\mathcal{J}'} \mid \boldsymbol{\theta}_{\mathcal{J}'})$ is log-concave wrt $\theta^{-\mathcal{J}'}$

if $\eta(\theta^{-\mathcal{J}'} \mid \lambda)$ is log-concave wrt $(\theta^{-\mathcal{J}'}, \lambda)$ and $\eta(\lambda \mid \theta_{\mathcal{J}'})$ is log-concave wrt λ. We have

$$\log \eta(\theta^{-\mathcal{J}'} \mid \lambda) = N \log \lambda - \log \Gamma(N) + (N-1) \log \theta^{-\mathcal{J}'} - \lambda \theta^{-\mathcal{J}'},$$

which is a sum of concave functions of $(\theta^{-\mathcal{J}'}, \lambda)$, hence itself concave. Meanwhile, note that the Gamma function is a conjugate prior for the exponential likelihood function, and so conditional on $\boldsymbol{\theta}_{\mathcal{J}'}$, $\lambda \sim \text{Gamma}\left(\alpha + |\mathcal{J}'|, \beta + \sum_{j \in \mathcal{J}'} \frac{1}{\theta_j}\right)$. The Gamma distribution is log-concave whenever its shape parameter is at least 1, and as $\alpha \geq 1$ it follows that $\eta(\lambda \mid \theta_{\mathcal{J}'})$ is log-concave wrt λ.

This work establishes that this system of attributes satisfies the conditions of Theorem 2 for the attribute case whenever the conditional distribution of $\varepsilon^{-\mathcal{K}}$ is log-concave.

F Completing the Proofs of Theorems 1 and 2

F.1 Proof of Theorem 1 Part (b)

Suppose the market observes the additional circumstance $k' \notin \mathcal{K}$. Under the expanded set of observed covariates, the marginal value of effort is

$$MV(\mathcal{S}, c_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0} \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, c_{k'}, \mathcal{S}] \;\middle|\; c_{k'}, \mathcal{S}\right].$$

Let $\mathcal{K}' = \mathcal{K} \cup \{k'\}$, and define the function $\eta(\gamma_{k'})$ by

$$\eta(\gamma_{k'}) \equiv \mathbb{E}[\varepsilon^{-\mathcal{K}'} \mid c_{k'} = \gamma_{k'}, \mathcal{S}]$$

Let $\tilde{\varepsilon}_{k'} \equiv \Lambda^{k'}(c_{k'}) + \eta(c_{k'})$. On $\mathcal{S}$, Y^0 may be decomposed as

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \tilde{\varepsilon}_{k'} + \tilde{\varepsilon}^{-\mathcal{K}'}.$$

If k' is a $\mathcal{S}$-mean shifter, then $\tilde{\varepsilon}^{-\mathcal{K}'}$ is independent of $c_{k'}$ on $\mathcal{S}$, and so the distribution of Y^0 depends on $c_{k'}$ only through $\tilde{\varepsilon}_{k'}$. Thus

$$\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, c_{k'}, \mathcal{S}] = \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S}].$$

Therefore, in a manner analogous to the attribute case, the marginal value of effort after observing j' depends on $c_{k'}$ only through $\tilde{\varepsilon}_{k'}$ and may be written

$$MV(\mathcal{S}, \tilde{\varepsilon}_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S}] \;\middle|\; \tilde{\varepsilon}_{k'}, \mathcal{S}\right].$$

Lemma F.1. $(\tilde{\varepsilon}_{k'}, \tilde{\varepsilon}^{-\mathcal{K}}, Y^0)$ *are affiliated on* $\mathcal{S}$.

Proof. This proof follows along very similar lines to the proof of Lemma D.1, so we omit the details. Let $f_{-\mathcal{J}}(u)$ be the conditional density of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$ and $g^0_{-\mathcal{K}'|k'}(x \mid z)$ be the conditional density of $\varepsilon^{-\mathcal{K}'} \mid c_{k'}$ on $\mathcal{S}$. The conditions required for the steps of that proof to go through are that $f_{-\mathcal{J}}(u)$ is log-concave, $g^0_{-\mathcal{K}'|k'}(x \mid z)$ is log-concave in x for all z, and $\tilde{\varepsilon}^{-\mathcal{K}'}$ is independent of $\tilde{\varepsilon}_{k'}$ on $\mathcal{S}$. The first two properties hold by $\mathcal{S}$-regularity of circumstance k', while the final property holds because k' is an $\mathcal{S}$-mean shifter. $\square$

We compare $MV(\mathcal{S}, \tilde{\varepsilon}_{k'})$ and $MV(\mathcal{S})$ in a manner very similar to the attribute case. Define

$$\widetilde{G}(z \mid y) \equiv \Pr(\tilde{\varepsilon}_{k'} \leq z \mid Y^0 = y, \mathcal{S})$$

to be the conditional CDF of $\tilde{\varepsilon}_{k'}$ given the outcome Y^0.

On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Taking expectations of each side conditional on $(Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S})$ yields

$$Y^0 = \mu_{\mathcal{S}} + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S}] + \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S}].$$

Hence

$$\begin{aligned} &\int d\widetilde{G}(z \mid y')\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \tilde{\varepsilon}_{k'} = z, \mathcal{S}] \\ &= y - \mu_{\mathcal{S}} - \int d\widetilde{G}(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \tilde{\varepsilon}_{k'} = z, \mathcal{S}]. \end{aligned} \tag{F.1}$$

Lemma F.1 directly implies that

$$\int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}]$$

is increasing in y', so (F.1) is decreasing in y'.

Following the same logic as in the attributes case, monotonicity of (F.1) implies that

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \leq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \tilde{\varepsilon}_{k'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right],$$

and it follows that

$$MV(\mathcal{S}) \leq \mathbb{E}[MV(\mathcal{S}, \tilde{\varepsilon}_{k'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is lower than the expected marginal value of effort when the circumstance k' is additionally observed.

The final step in the proof is again establishing that monotonicity holds uniformly across realizations of the additional circumstance, and not just on average. This follows from nearly identical work to the argument for the attributes case.

F.2 Proof of Theorem 2 Part (b)

Suppose that the market observes the additional circumstance $k' \notin \mathcal{K}$. Under the expanded set of observed covariates, the marginal value of effort becomes

$$MV(\mathcal{S}, c_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, c_{k'}, \mathcal{S}] \;\middle|\; c_{k'}, \mathcal{S}\right].$$

Because $\Lambda^{k'}$ is a one-to-one-mapping, $\varepsilon_{k'}$ is a sufficient statistic for the dependence of the distribution of $\varepsilon^{-\mathcal{K}'}$ on $c_{k'}$. We may therefore equivalently write the agent's marginal value of effort under the expanded set of covariates as

$$MV(\mathcal{S}, \varepsilon_{k'}) = \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}'} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] \;\middle|\; \varepsilon_{k'}, \mathcal{S}\right].$$

Lemma F.2. *$(\varepsilon_{k'}, \varepsilon^{-\mathcal{K}}, Y^0)$ are affiliated on $\mathcal{S}$.*

Proof. This is established along nearly identical lines to the proof of Lemma C.1, so we omit the details. Let $f_{-\mathcal{J}}(u)$ be the conditional density of $\theta^{-\mathcal{J}}$ on $\mathcal{S}$ and $g_{-\mathcal{K}'|k'}(x \mid z)$ be the conditional density of $\varepsilon^{-\mathcal{K}'} \mid \varepsilon_{k'}$ on $\mathcal{S}$. The conditions required for the steps of that proof to go through are that $f_{-\mathcal{J}}(u)$ is log-concave, $g_{-\mathcal{K}'|k'}(x \mid z)$ is log-concave in x for all z, and $(\varepsilon^{-\mathcal{K}'}, \varepsilon_{k'})$ are affiliated on $\mathcal{S}$. The first two properties hold by $\mathcal{S}$-regularity of circumstance k', while the final property holds by $\mathcal{S}$-affiliation of circumstance k'. □

We compare $MV(\mathcal{S}, \varepsilon_{k'})$ with $MV(\mathcal{S})$ in a manner very similar to the case of an additional attribute. Let $\mathcal{K}' = \mathcal{K} \cup \{k'\}$, and define

$$G(z \mid y) \equiv \Pr(\varepsilon_{k'} \leq z \mid Y^0 = y, \mathcal{S})$$

to be the conditional CDF of $\varepsilon_{k'}$ given the outcome Y^0. On $\mathcal{S}$, Y^0 may be written

$$Y^0 = \mu_{\mathcal{S}} + \theta^{-\mathcal{J}} + \varepsilon^{-\mathcal{K}}.$$

Taking expectations of each side conditional on $(Y^0, \varepsilon_{k'}, \mathcal{S})$ yields

$$Y^0 = \mu_{\mathcal{S}} + \mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] + \mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}].$$

Hence

$$\begin{aligned} &\int dG(z \mid y')\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}] \\ &= y - \mu_{\mathcal{S}} - \int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}]. \end{aligned} \tag{F.2}$$

Lemma F.2 directly implies that

$$\int dG(z \mid y')\mathbb{E}[\varepsilon^{-\mathcal{K}} \mid Y^0 = y, \varepsilon_{k'} = z, \mathcal{S}]$$

is increasing in y', so (F.2) is decreasing in y'.

Following the same logic as in the attributes case, monotonicity of (F.2) implies that

$$\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \mathcal{S}] \leq \mathbb{E}\left[\frac{\partial}{\partial Y^0}\mathbb{E}[\theta^{-\mathcal{J}} \mid Y^0, \varepsilon_{k'}, \mathcal{S}] \;\middle|\; Y^0, \mathcal{S}\right],$$

and it follows that

$$MV(\mathcal{S}) \leq \mathbb{E}[MV(\mathcal{S}, \varepsilon_{k'}) \mid \mathcal{S}].$$

Thus the marginal value of effort in subpopulation $\mathcal{S}$ is lower than the expected marginal value of effort when the circumstance k' is additionally observed.

G Proof of Lemma A.1

Fix an equilibrium, and consider the agent's choice of firm and effort in period 2 following any history h which firms have posted offers $R_2^j(h)$ (not necessarily on the equilibrium path). The agent's payoff from choosing firm j and effort e_2 is just $R_2^j(h) - C(e_2)$, which is strictly decreasing in e_2 for any choice of j. Thus the agent's unique optimal effort choice is $e_2 = 0$. Further, the agent optimally chooses a firm j such that $R_2^j(h) \geq R_2^{j'}(h)$ for every $j' \neq j$. Thus in any equilibrium, following any history the agent must exert zero effort in period 2 and choose a firm offering the highest reward.

Now consider each firm's reward offers to the agent in period 2 following some period-1 history h of reward offers, firm choice, and outcome. Let $e_1^*(h)$ be the equilibrium effort level in period 1, with $R_2^{*,j}(h)$ each firm's equilibrium period-2 reward offer. (In principle the equilibrium effort level may depend on the set of period-1 reward offers, and the period-2 rewards may depend on period-1 offers, the agent's firm choice, and outcomes.) Let $\overline{\theta}(h) \equiv \mathbb{E}^{e_1^*(h)}[\theta \mid Y]$. Given that the agent exerts no effort in period 2 no matter what rewards are posted, the off-path belief refinement imposed in Definition A.1 implies that, regardless of which firm the agent chose in period 1, a firm winning some the agent in period 2 after posting a reward R believes it will make expected profits $R - \overline{\theta}$ from the agent.

Suppose first that $\max_j R_2^{*,j}(h) < \overline{\theta}(h)$. Let firm j be a firm who does not win the agent in equilibrium, and consider a deviation to $R_j^2(h) \in (\max_j R_2^{*,j}(h), \overline{\theta}(h))$. This deviation ensures that firm j wins the agent and makes strictly positive profits from him, strictly improving on the profits it would have made in equilibrium. So it must be that $\max_j R_2^{*,j}(h) \geq \overline{\theta}(h)$.

Suppose instead that $\max_j R_2^{*,j}(h) > \overline{\theta}(h)$. Let firm j be the firm which wins the agent in equilibrium, and consider a deviation to $R_j^2(h) < \max_{j' \neq j} R_2^{*,j'}(h)$. This deviation ensures that firm j does not win the agent, strictly improving on the profits it would have made in equilibrium. So it must be that $\max_j R_2^{*,j}(h) = \overline{\theta}(h)$. A corollary is that all firms must make zero profits in period 2 in equilibrium following any period-1 history.

We now consider period-1 strategies. The work so far has established that, following any period-1 history h, in period 2 the agent receives a reward equal to $R_2^*(h) = \overline{\theta}(h)$ in equilibrium. Given any set of reward offers and any choice of firm j in period 1, the agent's payoff from exerting effort e is therefore

$$R_1^j - C(e) + \beta \cdot \mathbb{E}^e[\mathbb{E}^{e_1^*(h)}[\theta \mid Y]].$$

This function is maximized at $e_1^*(h)$ iff $e_1^*(h) = e^*$, the equilibrium effort level in the model of the main text.

It follows that regardless of the posted rewards, a firm who wins the agent by offering a reward R believes it will earn an expected profit of $\mu + e^* - R$ in this period from doing so. Further, in equilibrium its period-2 profits are independent of reward offers and firm choices in the period-1 market. Logic very similar to that used to characterize period-2 rewards then implies that the agent must receive a reward of $R_1^* = \mu + e^*$ in period 1 in equilibrium.