Improving Information from Manipulable Data*

Alex Frankel† Navin Kartik‡

September 23, 2019

Abstract

Data-based decision-making must account for data manipulation, or gaming, by agents who are aware of how decisions are being made. We study a framework in which this manipulation makes data less informative when decisions depend more strongly on data. We formalize why and how a decision-maker should commit to under-utilizing data in order to attenuate information loss.

JEL Classification: C72; D40; D82

Keywords: Gaming; Goodhart's Law; Strategic Classification

*We thank Ralph Boleslavsky, Max Farrell, Pepe Montiel Olea, and conference audiences for helpful comments. Bruno Furtado and Suneil Parimoo provided excellent research assistance.

†University of Chicago Booth School of Business; afrankel@chicagobooth.edu.

‡Columbia University, Department of Economics; nkartik@columbia.edu.

1. Introduction

In various situations an agent receives an allocation based on some prediction about her characteristics—a prediction that relies on data generated by the agent's own behavior. Firms use a consumer's web browsing history for price discrimination or ad targeting; a prospective borrower's loan decision and interest rate depend on her credit score; and web search rankings take as input a web site's own text and metadata. In all these settings, agents who understand the prediction algorithm can alter their behavior to receive a more desirable allocation. Consumers can adjust browsing behavior to mimic those with low willingness to pay; borrowers can open or close various accounts to improve their credit score; and web sites can engage in search engine optimization to improve their rankings. How should a designer account for data manipulation when setting the allocation rule?

First consider a naive designer, one who is unaware of the potential for manipulation. Before implementing an allocation rule, the designer gathers data about agents and fits a model to estimate agents' types (i.e., the relevant characteristics) from observables. The *naive allocation rule* then assigns each agent the allocation that is optimal according to this estimate of her type. But after the rule is implemented, agents' behavior changes: if agents with "higher observables" x receive a "higher allocation" y under the allocation rule $Y(x)$, and if agents prefer higher allocations, then some agents will find ways to game the rule by increasing their x. In line with Goodhart's Law, the original estimation is no longer accurate.

A more sophisticated designer would realize that agent behavior had changed, and could gather new data and then re-estimate the relationship between observables and type. After the designer updates the allocation rule based on the new prediction, agent behavior would of course change once again. The designer could iterate until finding a *fixed point*: an allocation rule that is a best response to the data that is generated under this very rule. But the resulting allocation need not match the desired agent characteristics very well.

The question of this paper is how a designer with *commitment* power—a Stackelberg leader—should adjust a fixed-point allocation rule in order to improve the accuracy of the allocation. We find that a designer should commit to making the allocation rule less sensitive to manipulable data than under the fixed point. In other words, the designer should "flatten" the allocation rule. Flattening the allocation rule results in "ex-post" suboptimality in the sense that an agent's allocation will generally not be optimal given the information the designer obtains about her type. By contrast, fixed-point allocations are ex-post optimal. However, a flatter allocation rule reduces agent incentives for manipulation, which makes

the data more informative about the agent's type. Allocation accuracy improves on balance. We develop and explore this logic in what we believe is a compelling and canonical model of information loss due to manipulation.

By way of background, note that in some environments there are fixed-point rules that actually deliver the designer's full-information outcome. We can think of a fixed-point rule as corresponding to the designer's equilibrium strategy in a signaling game in which the designer and agent best respond to each other. Under a standard single-crossing condition à la Spence (1973)—the designer wants to give more desirable allocations to agents with higher types, and higher types have lower marginal costs of taking higher observable actions—this signaling game has a fully separating equilibrium, i.e., one in which the designer perfectly matches the agent's allocation to her type.

We are interested in settings in which the agent's manipulation leads to information loss. We build on the general framework of "muddled information" (Frankel and Kartik, 2019), specifically the linear-quadratic model introduced in Fischer and Verrecchia (2000) and Bénabou and Tirole (2006); see also Ali and Bénabou (2019), Gesche (2019), Ball (2019), and, outside the linear-quadratic class, Hu, Immorlica, and Vaughan (2019). In this framework, single-crossing fails because agents have a two-dimensional type. The dimension of interest to the designer is the agent's *natural action* $\eta \in \mathbb{R}$, which determines her observable $x \in \mathbb{R}$ prior to any manipulation. Agents are also heterogeneous in their *gaming ability*, $\gamma \in \mathbb{R}$, which summarizes how much they adjust their action in response to incentives: gaming ability may represent marginal costs of altering one's observable, or marginal benefits of improving one's allocation. In this framework, we consider a designer who seeks to minimize quadratic loss between an agent's allocation $y \in \mathbb{R}$ and the natural action η. We restrict attention to linear allocation rules or policies $Y(x) = \beta x + \beta_0$, and we posit that agents adjust their observable x in proportion to $\gamma\beta$, i.e., their gaming ability times the sensitivity of allocations to observables.[1]

Our main result establishes that in this environment the optimal policy is less sensitive to observables than is the fixed-point policy. Mathematically, for allocation rules $Y(x) = \beta x + \beta_0$, the designer flattens the fixed-point rule by attenuating the coefficient β towards zero. For instance, suppose the sensitivity of the naive policy is $\beta = 1$: when the designer does not condition the allocation on observables, the linear regression coefficient of type η on observable x is 1, and the naive designer responds by matching her allocation rule's

[1] Subsection 2.1 points out that such behavior for the agents can be microfounded. Subsection 4.1 discusses the optimality of linear allocation rules.

sensitivity to this regression coefficient. The fixed-point policy may have $\beta = 0.7$. That is, when the designer sets $\beta = 0.7$ and runs a linear regression of η on x (using data generated by the agent in response to $\beta = 0.7$), the regression coefficient is the same 0.7. Our result is that the optimal policy has $\beta \in (0, 0.7)$, say $\beta = 0.6$. Note that the designer recognizes and commits to ex-post misallocations under this optimal policy: after the designer sets $\beta = 0.6$, the corresponding linear regression coefficient could be $\simeq 0.75$. We emphasize that our argument for shrinking regression coefficients is driven by the informational benefit from reduced manipulation, and in turn, the resulting improvement in allocations. It is orthogonal to concerns about model overfitting.

In comparing our commitment solution with the fixed-point benchmark, it is helpful to keep in mind two distinct interpretations of the fixed point. The first concerns a designer who has market power in the sense that agents adjust their manipulation behavior in response to this designer's policies. Think of web sites engaging in search engine optimization to specifically improve their Google rankings; third party sellers paying for fake reviews on the Amazon platform; or citizens trying to game an eligibility rule for a targeted government policy. In these cases the designer may settle on a fixed point by iterating policies until reaching an ex post optimum. Our paper highlights that this fixed point may yet be suboptimal ex ante, and offers the prescriptive advice of flattening the allocation rule.

A second perspective is that the fixed-point policy represents the outcome of a competitive market. With many banks, any one bank that uses credit information in an ex-post suboptimal manner will simply be putting itself at a disadvantage to its competitors. So the fixed point becomes a descriptive prediction of the market outcome, i.e., the equilibrium of a signaling game. In that case, the optimal policy we derive may suggest a proposal for a government intervention to improve market allocations, or it may suggest the direction that firms would move in if they could collude.

Related Literature. At a very broad level, our flattening result is reminiscent of the "downward distortion" of allocations in screening problems following Mussa and Rosen (1978). That said, our framework, analysis, and emphasis—on manipulation and information loss, allocation accuracy, contrasting commitment with fixed points—are not readily comparable with most of that literature.

One paper on screening we would highlight is Bonatti and Cisternas (2019), although their model is still quite different from ours. They study a price discrimination problem in

which a designer with access to a long-lived consumer's purchase history chooses what to reveal to a sequence of short-lived firms. They find that firms get better information about consumer types, and hence higher steady-state profits, if the designer reveals a statistic that underweights recent consumer behavior. Suitable underweighting dampens consumer incentives to manipulate demand.

There is a literature in finance concerning the difficulty of using financial activity to learn fundamentals when market participants have incentives to manipulate such learning. Although these models are again very different from ours, some papers highlight, as we do, the benefits of commitments to "underutilizing information".[2] See, for example, Bond and Goldstein (2015) and Boleslavsky, Kelly, and Taylor (2017). These authors study models of trading in financial markets when there is a policymaker who, upon observing prices and/or order flows, can intervene in the market. A common theme is that the anticipation of intervention can affect market participants' behavior in a manner that makes the financial market less informative about a fundamental to which the policymaker would like to tailor her intervention. Both papers establish that the policymaker may benefit from a commitment that, in some sense, entails ex-post underutilization of information in order to improve the market's informativeness. In particular, Bond and Goldstein (2015, Proposition 2) highlight a local first-order information benefit vs. second-order allocation loss akin to our Lemma 1. Unlike us, they do not study global optimality.

The motivation underlying our work also relates to that of other economists who have studied the design of testing regimes and other instruments to improve information extraction. Harbaugh and Rasmusen (2018) show how a test that pools together certain realizations may yield more information than a fully informative test if agents who expected bad outcomes on the fully informative test would choose not to participate. Perez-Richet and Skreta (2018) study how a principal may benefit from a noisy test when the agent can manipulate the test and the principal's allocation best responds to the agent's manipulation. Martinez-Gorricho and Oyarzun (2019) find conditions under which a designer concerned with manipulation should commit to a "conservative" (or "confirmatory") threshold for overturning an agent's default allocation.[3] Jann and Schottmüller (2018), Ali and Bénabou (2019), and Frankel and Kartik (2019) analyze strategic environments in which hiding information about agents'

[2] Less directly related, Duffie and Dworczak (2018) design financial benchmarks to be robust to the incentives of traders to distort these benchmarks; Zhang's (2019) related work explores the susceptibility of financial derivatives to price manipulation.

[3] Conservatism has also been advocated to mitigate distortions in other contexts (e.g., Li, 2001).

actions—increasing their privacy—can improve an observer's information about the agents' characteristics.[4] In this vein, Ball (2019) considers an environment with multidimensional actions and proposes that an intermediary should add more noise to the more manipulable dimensions of agents' actions.

Finally, our paper connects to a recent computer science literature studying how a designer with commitment power should set up a classification algorithm in the presence of strategic manipulation, and the welfare effects of the designer's policy. See, among others, Hardt, Megiddo, Papadimitriou, and Wootters (2016), Hu et al. (2019), Milli, Miller, Dragan, and Hardt (2018), and Kleinberg and Raghavan (2019). Note that, in contrast to the latter paper, we are only concerned with "gaming" by agents; we do not model an agent's effort as either producing desirable output or intrinsically affecting the agent's optimal allocation. Moreover, our designer's objective only values the accuracy of the allocations, and not (directly) agents' costs of manipulation. In a binary strategic classification problem, Braverman and Garg (2019) discuss the role of random allocations in not only improving allocation accuracy but also reducing manipulation costs.[5]

2. Model

The agent has a type $(\eta, \gamma) \in \mathbb{R}^2$ drawn from some joint distribution F. It may be helpful to remember the mnemonics η for *n*atural action, and γ for *g*aming ability; see Subsection 2.1. Assume the variances $\mathrm{Var}(\eta) = \sigma_\eta^2$ and $\mathrm{Var}(\gamma) = \sigma_\gamma^2$ are positive and finite.[6] Denote the means of η and γ by μ_η, μ_γ, and assume their correlation is $\rho \in (-1, 1)$, with $\rho = \mathrm{Cov}(\eta, \gamma)/(\sigma_\eta \sigma_\gamma)$.

The designer seeks to match an allocation $y \in \mathbb{R}$ to η, with a quadratic loss of $(y - \eta)^2$. The designer chooses $y = Y(x)$ as a function of an observed action $x \in \mathbb{R}$ that is chosen by the agent. Thus, the designer's welfare loss is

$$\text{Welfare Loss} \equiv \mathbb{E}[(Y(x) - \eta)^2]. \tag{1}$$

The agent chooses x as a function of her type (η, γ) after observing the designer's allocation

[4] Eliaz and Spiegler (2019) explore the distinct question of whether an agent has incentives to reveal her own data to a "non-Bayesian statistician" who is making predictions about her.

[5] In the economics literature, Ederer, Holden, and Meyer (2018) study randomized rewards scheme to reduce gaming in a multi-tasking environment. Their focus is on improving effort rather than information.

[6] Throughout, we use 'positive' without qualification to mean 'strictly positive', and similarly for 'negative', 'larger', and 'smaller'.

rule Y. In a manner detailed later, the agent will have an incentive to choose a higher x to obtain a higher y. Given a strategy of the agent, the designer can compute the distribution of x and the value of $\mathbb{E}[\eta|x]$ for any x the agent may choose. A standard decomposition[7] is

$$\text{Welfare Loss} = \underbrace{\mathbb{E}[(\mathbb{E}[\eta|x] - \eta)^2]}_{\text{Info loss of estimating } \eta \text{ from } x} + \underbrace{\mathbb{E}[(Y(x) - \mathbb{E}[\eta|x])^2]}_{\text{Misallocation loss given estimation}} . \tag{2}$$

Plainly, holding fixed the agent's strategy, it is "ex-post optimal" for the designer to set $Y(x) = \mathbb{E}[\eta|x]$. However, the agent's strategy responds to Y. So it is possible that the designer may prefer to use an ex-post suboptimal allocation rule because that improves her ability to estimate η from x, as seen in the first term of (2). That is, the designer may benefit from the power to commit to her allocation rule.

2.1. Linearity Assumptions

Assume the designer restricts attention to linear allocation rules: the designer chooses policy parameters $(\beta, \beta_0) \in \mathbb{R}^2$ such that

$$Y(x) = \beta x + \beta_0. \tag{3}$$

Also assume that, given the designer's policy (β, β_0), the agent chooses x according to a linear strategy $X_\beta(\eta, \gamma)$ that takes the form

$$X_\beta(\eta, \gamma) = \eta + m\beta\gamma \tag{4}$$

for some given parameter $m > 0$. Thus η is the agent's "natural action": the action that would be taken when the designer's policy does not depend on x (i.e., $\beta = 0$). The variable γ represents idiosyncratic responsiveness to the designer's policy: agents with higher γ increase their action from their natural level by more for any $\beta > 0$. The parameter m captures a common component of responsiveness across all agents.

[7] The right-hand sides of (1) and (2) are equal if

$$\mathbb{E}[(Y(x))^2 - 2\eta Y(x) + \eta^2] = \mathbb{E}\big[\eta^2 - 2\eta\mathbb{E}[\eta|x] + (\mathbb{E}[\eta|x])^2 + (Y(x))^2 - 2Y(x)\mathbb{E}[\eta|x] + (\mathbb{E}[\eta|x])^2\big].$$

Canceling out like terms and rearranging, it suffices to show that

$$2\mathbb{E}\big[(\mathbb{E}[\eta|x] - \eta)Y(x)\big] = 2\mathbb{E}\big[(\mathbb{E}[\eta|x] - \eta)\mathbb{E}[\eta|x]\big].$$

This equality holds by the orthogonality condition $\mathbb{E}[(\mathbb{E}[\eta|x] - \eta)g(x)] = 0$ for all functions $g(x)$.

The strategy in Equation 4 can be motivated as the best response for an agent who maximizes a utility of

$$m\gamma y - (x-\eta)^2/2.$$

Here m captures the "stakes" that agents face to obtain higher y, and γ is an idiosyncratic marginal benefit. Alternatively, the strategy is also optimal for an agent with $\gamma > 0$ who maximizes

$$y - (x-\eta)^2/(2m\gamma).$$

Here m parameterizes the "manipulability" of the action x, and γ is an agent's idiosyncratic "gaming ability".

2.2. The Designer's Problem

We study a setting in which the designer commits to her policy (β, β_0), which the agent observes and responds to according to (4). Plugging the rule (3) and the strategy (4) into the welfare loss function (1) yields

$$\text{Welfare Loss} = \mathbb{E}[(\beta(\eta + m\beta\gamma) + \beta_0 - \eta)^2].$$

The designer's problem is therefore to choose (β, β_0) to minimize the above loss function, which is quartic in β.[8] We denote the solution as (β^*, β_0^*).

2.3. Discussion

Given the asymmetry between the characteristics η and γ in the agent's strategy (4), it is crucial for our results that the designer seeks to match her allocation with η rather than γ. The reason is that when the designer's policy puts more weight on the data—i.e., when β increases—the agent's action x becomes less informative about η but more informative about γ; Remark 1 below makes this point precise.

It is, on the other hand, straightforward to generalize our analysis to the allocation matching some other characteristic of the agent, τ, that is correlated with η. The assumption we would require is that $\mathbb{E}[\tau|\eta]$ is independent of γ and linear in η. The welfare loss $\mathbb{E}[(Y(x)-\tau)^2]$

[8] Using standard mean-variance decompositions, some algebra shows that

$$\text{Welfare Loss} = (1-\beta)^2\sigma_\eta^2 + m^2\beta^4\sigma_\gamma^2 - 2(1-\beta)m\beta^2\rho\sigma_\eta\sigma_\gamma + (\beta_0 - (1-\beta)\mu_\eta + m\beta^2\mu_\gamma)^2.$$

could be decomposed as $\mathbb{E}[(Y(x)-\mathbb{E}[\tau|\eta])^2]+\mathbb{E}[(\mathbb{E}[\tau|\eta]-\tau)^2]$. As the second term here—the information loss in predicting τ from η—is independent of the allocation rule $Y(x)$, it would not affect the designer's choice of $Y(x)$. The designer would effectively be trying to match the allocation to (a linear function of) η.

2.4. Preliminaries

2.4.1. Linear regression of type η on action x

When the designer uses a policy $(\tilde{\beta}, \tilde{\beta}_0)$, the agent responds with the strategy $X_{\tilde{\beta}}(\eta, \gamma) = \eta + m\tilde{\beta}\gamma$. Suppose the designer were to gather data under this agent behavior and then estimate the relationship between the type dimension of interest η and the action x. Specifically, let $\hat{\eta}_{\tilde{\beta}}(x)$ denote the best linear estimator of η from x under a quadratic loss objective:

$$\hat{\eta}_{\tilde{\beta}}(x) \equiv \hat{\beta}(\tilde{\beta})x + \hat{\beta}_0(\tilde{\beta}),$$

with $\hat{\beta}$ and $\hat{\beta}_0$ the coefficients of an ordinary least squares (OLS) regression of η on x.

Following standard results for OLS,

$$\hat{\beta}(\tilde{\beta}) = \frac{\sigma_\eta^2 + m\rho\sigma_\eta\sigma_\gamma\tilde{\beta}}{\sigma_\eta^2 + m^2\sigma_\gamma^2\tilde{\beta}^2 + 2m\rho\sigma_\eta\sigma_\gamma\tilde{\beta}}, \tag{5}$$

where the right-hand side's numerator is the covariance of x and η given the strategy $X_{\tilde{\beta}}$, and its denominator is the variance of x.[9] Correspondingly,

$$\hat{\beta}_0(\tilde{\beta}) = \mu_\eta - \hat{\beta}(\tilde{\beta})[\mu_\eta + m\tilde{\beta}\mu_\gamma].$$

It is useful to further rewrite the welfare loss (2) as follows, for any policy (β, β_0) defining the linear allocation rule $Y(x) = \beta x + \beta_0$:[10]

$$\text{Welfare Loss} = \underbrace{\mathbb{E}[(\hat{\eta}_\beta(x) - \eta)^2]}_{\text{Info loss of linearly estimating } \eta \text{ from } x} + \underbrace{\mathbb{E}[(Y(x) - \hat{\eta}_\beta(x))^2]}_{\text{Misallocation loss given linear estimation}}. \tag{6}$$

Some readers may find it helpful to note that information loss (the first term in (6)) is the

[9] Our maintained assumption of $\rho \in (-1, 1)$ ensures the denominator is non-zero.

[10] This derivation is identical to that in fn. 7, only replacing $\mathbb{E}[\eta|x]$ by $\hat{\eta}_\beta(x)$ and applying the orthogonality condition $\mathbb{E}[(\hat{\eta}_\beta(x) - \eta)g(x)] = 0$ for all affine functions $g(x)$.

variance of the residuals in an OLS regression of η on x; put differently, $\mathbb{E}[(\hat{\eta}_\beta(x) - \eta)^2] = \sigma_\eta^2(1 - R_{x\eta}^2)$, where $R_{x\eta}^2$ is the coefficient of determination in that regression. It bears emphasis that Equation 6 is simply a welfare decomposition for any linear allocation rule, which will aid our interpretation and intuition; the appearance of OLS here does not impose any additional restriction on the designer.

Remark 1. For $\rho \geq 0$, $\hat{\beta}(\tilde{\beta})$ is decreasing on $\tilde{\beta} \geq 0$. To see why, notice that when $\tilde{\beta}$ increases the agent's action x depends more on the variable γ. This increases $\text{Var}(x)$ and, when $\rho \geq 0$, also provides the designer with less information about the variable η that she is trying to estimate from x.[11] Both effects lead to a lower $\hat{\beta}$. By contrast, if the designer were trying to estimate γ rather than η (i.e., minimizing $\mathbb{E}[(y - \gamma)^2]$ rather than $\mathbb{E}[(y - \eta)^2]$), then for $\rho \geq 0$, the analogous regression coefficient of γ on x would be increasing on $\tilde{\beta} \geq 0$.

2.4.2. Benchmark policies

Constant. A rule that does not condition the allocation on the observable corresponds to a constant policy (β, β_0) with $\beta = 0$. A constant policy gives rise to a welfare loss of $\sigma_\eta^2 + (\beta_0 - \mu_\eta)^2$. In the decomposition of Equation 6, the entire welfare loss is due to misallocation; the information loss is zero because the agent's behavior $x = \eta$ fully reveals the natural action η. Under the constant policy the linear estimator $\hat{\eta}_0$ has coefficients $\hat{\beta}(0) = 1$ and $\hat{\beta}_0(0) = 0$.

Naive. Consider a designer who gathers data on the relationship between η and x produced by agents responding to a constant policy. Suppose further that the designer fails to account for manipulation, expecting agents to maintain the strategy $X_0(\eta, \gamma) = \eta$ regardless of the policy (β, β_0). Then the designer would (incorrectly) perceive her optimal policy to be $(\beta^{\text{n}}, \beta_0^{\text{n}}) \equiv (\hat{\beta}(0), \hat{\beta}_0(0)) = (1, 0)$.

Designer's best response. More generally, suppose the designer expects the agent to use the strategy $X_{\tilde{\beta}}(\eta, \gamma) = \eta + m\tilde{\beta}\gamma$ regardless of the policy (β, β_0). The designer would find it optimal in response to set an allocation rule $Y(x)$ equal to the best linear estimator of η from x, i.e., a policy $(\hat{\beta}(\tilde{\beta}), \hat{\beta}_0(\tilde{\beta}))$ yielding $Y(x) = \hat{\eta}_{\tilde{\beta}}(x)$.

[11] Less information is not generally in the Blackwell (1951) sense unless the prior on (η, γ) is bivariate normal. Rather, it is in the sense of a higher information loss of linearly estimating η from x: $\mathbb{E}[(\hat{\eta}_{\tilde{\beta}}(x) - \eta)^2]$ is increasing in $\tilde{\beta}$.

Fixed point. We say that a policy $(\beta^{\text{fp}}, \beta_0^{\text{fp}})$ is a *fixed point* if

$$\beta^{\text{fp}} = \hat{\beta}(\beta^{\text{fp}}) \quad \text{and} \quad \beta_0^{\text{fp}} = \hat{\beta}_0(\beta^{\text{fp}}).$$

A fixed point corresponds to a Nash equilibrium of a game in which the designer's policy is set simultaneously with the agent's strategy. That is, instead of the designer committing to a policy (the Stackelberg solution), the policy is a best response to the agent's strategy that the policy induces. In the decomposition of Equation 6, a fixed-point policy may have a positive information loss, but it has zero misallocation loss—the designer is choosing the optimal policy given the information that she has.

As elaborated in Subsection 4.1, an elliptical prior distribution F (subsuming normal distributions) ensures that fixed points also correspond to equilibria of a signaling game in which the agent first chooses her observable action x and the designer then chooses the allocation y.

Figure 1 illustrates some designer best response functions and fixed points.

3. Analysis

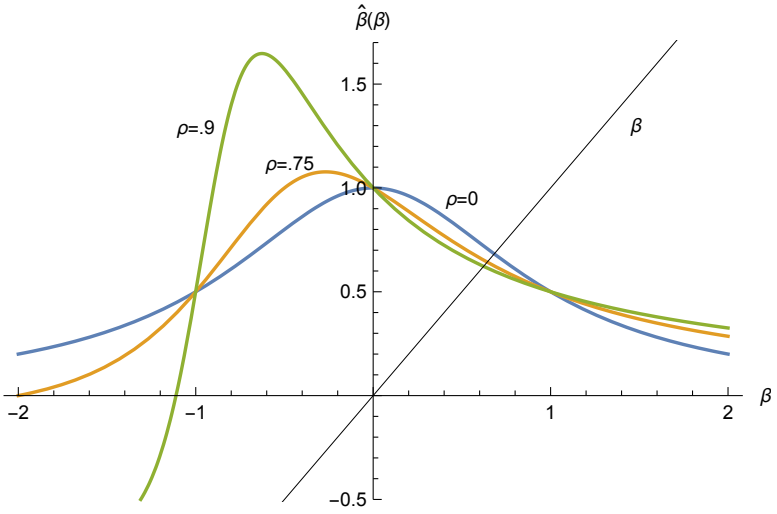
3.1. Main Result

We seek to compare the designer's optimal policy (β^*, β_0^*) with the fixed points $(\beta^{\text{fp}}, \beta_0^{\text{fp}})$. There can, in general, be multiple fixed points, but there is always at least one with a positive sensitivity or weight on the agent's action, i.e., $\beta^{\text{fp}} > 0$. Moreover, when there is nonnegative correlation in the agent's characteristics (i.e., $\rho \geq 0$), there is only one nonnegative fixed point, and it satisfies $\beta^{\text{fp}} \in (0, 1)$. See Proposition A.1 in the Appendix.

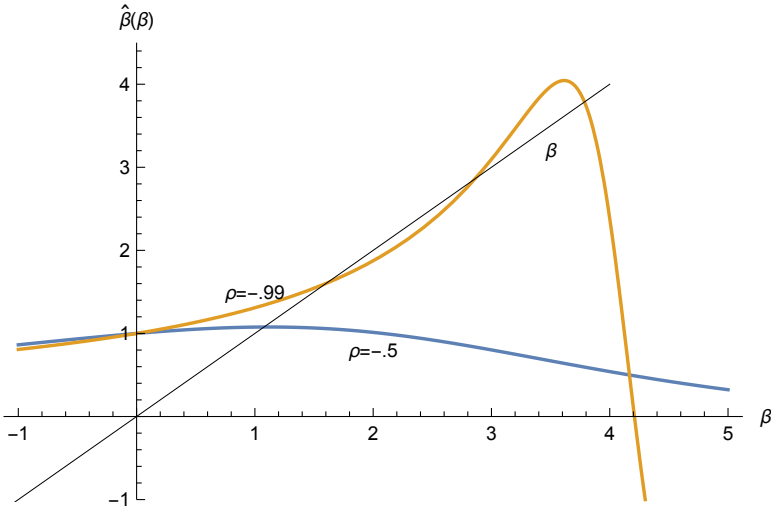
Take any fixed-point sensitivity $\beta^{\text{fp}} > 0$. Our main result is that the optimal policy puts less weight on the agent's action than does the fixed point:

Proposition 1. *There is a unique optimum,* (β^*, β_0^*). *It has* $\beta^* > 0$ *and* $\beta^* < \beta^{\text{fp}}$ *for any fixed point* $\beta^{\text{fp}} > 0$.

Here is the intuition, as illustrated graphically in Figure 2. Consider the designer choosing $\beta = \beta^{\text{fp}} > 0$. At this point the designer's policy is ex-post optimal in the sense that the loss from misallocation (the second term in the welfare decomposition Equation 6) is minimized at zero. Adjusting the sensitivity β in either direction from β^{fp} leads to an increase in



(a) Parameters: $\sigma_\eta = \sigma_\gamma = 1$ and $m = 1$.



(b) Parameters: $\sigma_\eta = \sigma_\gamma = 1$ and $m = 0.24$.

Figure 1 – The best response function $\hat{\beta}$. As shown in Figure 1a, $\hat{\beta}$ is decreasing on $[0, \infty)$ when $\rho \geq 0$. Figure 1b illustrates that this need not be true when $\rho < 0$. In all cases, intersections of $\hat{\beta}$ with β correspond to fixed points β^{fp}.

misallocation loss, but this harm is second order because we are starting from a minimum. On the other hand, at β^{fp} there is a positive information loss (the first term in (6)) because x does not reveal η. As suggested by Remark 1, lowering β reduces information loss, which yields a first-order benefit. While Remark 1 was restricted to $\rho \geq 0$, it turns out that the first-order improvement intuition is general. Hence, on net, there is a first-order welfare benefit of lowering β from β^{fp}. Of course, the designer wouldn't lower β all the way down to 0, since making some use of the information from the data is better than not using it at all.[12] To complete the proof of Proposition 1, we establish uniqueness of the global optimum, rule out that it is negative, show that it is less than every fixed point $\beta^{\text{fp}} > 0$.

To formalize the key step of the proof, which establishes a first-order benefit in reducing β from any β^{fp}, let $\mathcal{L}(\beta)$ be the welfare loss from choosing policy β, with derivative $\mathcal{L}'(\beta)$.[13]

Lemma 1. *For any β^{fp}, it holds that $\mathcal{L}'(\beta^{\text{fp}}) > 0$.*

Note that Lemma 1 also applies to negative values of β^{fp} when those exist.

Remark 2. The welfare gains from commitment can be substantial. For suitable parameters, the welfare in the unique fixed point is arbitrarily close to that of the constant policy $Y(x) = \mu_\eta$, while the welfare under commitment is arbitrarily close to that of the first best.[14]

3.2. Comparative Statics

We provide a few comparative statics below. In taking comparative statics, it is helpful to observe that the designer's best response $\hat{\beta}(\beta)$ defined in Equation 5 depends on parameters m, σ_η, and σ_γ only through the statistic $k \equiv m\sigma_\gamma/\sigma_\eta$, as does the welfare loss $\mathcal{L}(\beta)$ divided by σ_η^2 (see Equation A.2 in Appendix A.2). Therefore, the optimal and fixed-point values β^* and β^{fp} also only depend on these parameters through k. The parameter k summarizes the susceptibility of the allocation problem to manipulation: higher k (arising from higher stakes or manipulability m of the mechanism, greater variance in gaming ability σ_γ^2, or lower variance in natural actions σ_η^2) means that under any fixed policy, agents as a whole

[12] Indeed, any fixed-point policy itself does better than the best constant policy $(\beta, \beta_0) = (0, \mu_\eta)$. Note, however, that this constant policy can be better than the naive policy $(\beta^{\text{n}}, \beta_0^{\text{n}}) = (1, 0)$.

[13] We write $\mathcal{L}(\beta)$ rather than $\mathcal{L}(\beta, \beta_0)$ because for any β there is uniquely optimal β_0 that can be substituted in; see the proof of Proposition 1, which also confirms that $\mathcal{L}(\cdot)$ is differentiable.

[14] The parameters are such that $m\sigma_\gamma/\sigma_\eta \to 1/4^+$ and correlation $\rho \to -1$. Note that both the first-best welfare and that under the constant policy are independent of ρ; the former is 0 (by normalization) while the latter is $-\sigma_\eta^2$, which can be arbitrarily low.

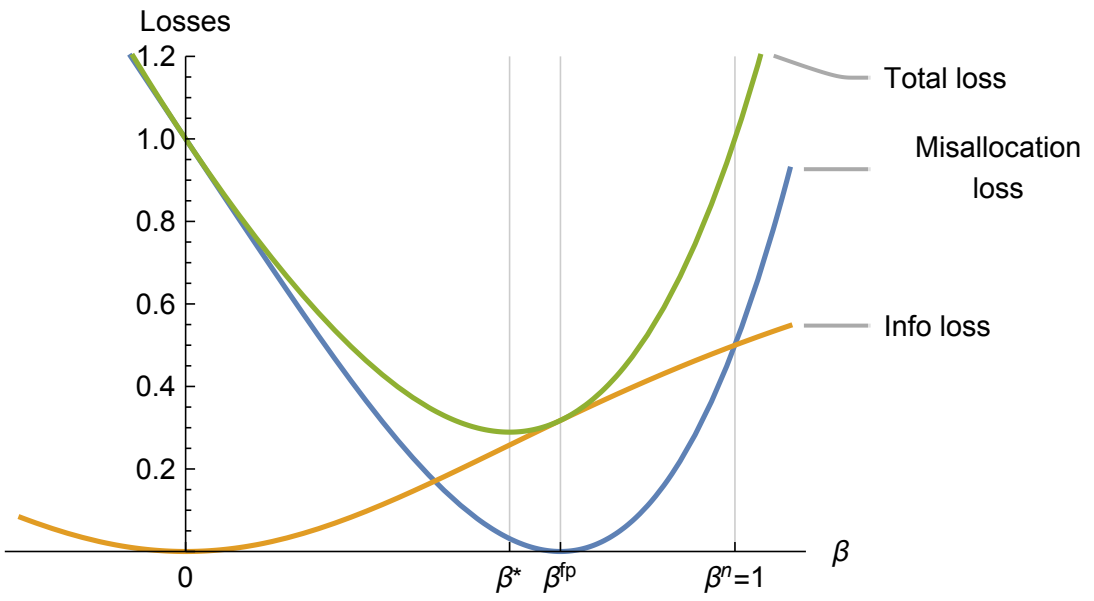


Figure 2 – The welfare loss decomposition from Equation 6 for policy (β_0, β), with the optimal β_0 plugged in for each β on the horizontal axis. Parameters: $\sigma_\eta = \sigma_\gamma = \rho = 1$ and $m = 1$. Numerical solutions: $\beta^* = 0.590$ and $\beta^{\text{fp}} = 0.682$.

adjust their observable action x further from their natural action η, relative to the spread of observables prior to manipulation. Hence, for comparative statics over model primitives, it is sufficient to consider only the statistic k and the correlation parameter ρ.

Proposition 2. *For $k \equiv m\sigma_\gamma/\sigma_\eta$, the following comparative statics hold.*

1. *As $k \to \infty$, $\beta^* \to 0$; as $k \to 0$, $\beta^* \to 1$. If $\rho \geq 0$, then β^* is strictly decreasing in k; if $\rho < 0$, then β^* is strictly quasi-concave in k, attaining a maximum at some point.*

2. *β^* is strictly increasing in ρ when $k > 3/4$, strictly decreasing in ρ when $k < 3/4$, and independent of ρ when $k = 3/4$.*

3. *When $\rho = 0$, $\beta^*/\beta^{\text{fp}}$ is strictly decreasing in k, approaching $\sqrt[3]{1/2} \approx 0.79$ as $k \to \infty$ and 1 as $k \to 0$.*

Part 1 of the proposition implies that when agents' characteristics are nonnegatively correlated, a designer faced with a more manipulable environment should put less weight on the agents' observable action. While such monotonicity is intuitive, it does not hold when there is negative correlation. Similarly, one might expect greater positive correlation to increase the optimum β^*; indeed, Frankel and Kartik (2019, Proposition 4) establish that

it does have this effect on the (unique) positive fixed point $\beta^{\text{fp}} > 0$. But we see in part 2 of Proposition 2 that this holds for β^* only when the susceptibility-to-manipulation statistic k is large enough. Finally, part 3 implies that when the characteristics are uncorrelated, the ratio $\beta^{\text{fp}}/\beta^*$ decreases as the statistic k increases. As $k \to 0$, the fixed point fully reveals an agent's natural action ($\beta^{\text{fp}} \to 1$) and so the designer does not benefit from commitment power: the fixed point is optimal as it provides the minimum possible welfare loss. As $k \to \infty$, both β^* and β^{fp} tend to zero yet the ratio $\beta^*/\beta^{\text{fp}}$ stays bounded.

4. Discussion

4.1. Nonlinear Policies

As explained in Subsection 2.1, there are reasonable objective functions under which it is optimal for the agent to use a linear strategy of the form (4) when the designer uses a linear allocation rule of the form (3). That the designer uses linear allocation rules is, however, generally restrictive. Fischer and Verrecchia (2000), Bénabou and Tirole (2006), Gesche (2019), and Frankel and Kartik (2019) have shown that fixing any linear strategy for the agent, the designer's best response is also linear if the agent's type distribution is bivariate elliptical (Gómez, Gómez-Villegas, and Marín, 2003), subsuming bivariate normal. Hence, under these joint distributions, the linear fixed-point policies of the current paper correspond to equilibrium response functions in a signaling game. Ball (2019) extends these results to a multidimensional action space. A plausible conjecture is that elliptical distributions also ensure global optimality of linear allocation rules when the designer has commitment power.

4.2. Alternative Models of Information loss

The fundamental logic underlying our main result is simply that "flattening" the allocation rule from any fixed point yields a first-order improvement in information while only a second-order loss from misallocation. We have developed this point in what we believe is a canonical model of information loss from manipulation, one used in a number of aforementioned papers. But we think the point applies more broadly, including in other models of information loss. For instance, even a model with a one-dimensional type (such as the model in this paper with no heterogeneity on the gaming ability γ) can lead to information loss when there is a bounded action space and strong manipulation incentives. The reason is pooling "at the top". We establish in Appendix B a version of our result for a simple model in this vein.

4.3. General Allocation Problems

We conclude by sketching a proposal for attenuating the impact of manipulable data in more general allocation problems. We have in mind an abstract environment in which a designer estimates an agent's characteristic η from the observation of some x, and then assigns an allocation y based on both x and the estimate of η. None of these variables need be scalar; in particular, some components of x may be manipulable and some not. As such, the functional form of the allocation rule need not have any easily interpreted coefficient measuring how "flat" or "steep" it is with respect to x.

To formalize our proposal—estimation with noise—let a *data set* be a joint distribution over (x, η). Let ML be an *estimation procedure* (e.g., a machine learning algorithm) that takes as input an observable x and a data set d, and then outputs an allocation y. We interpret $ML(x; d)$ as first estimating η from x after being fit to the training data d, and then outputting the designer's preferred allocation given x and the estimate of η.

Estimation with noise. Recall the classical econometric result that measurement error on an independent variable leads to attenuation bias, i.e., to an estimated coefficient in a linear regression that is biased towards zero. Applying this concept, here is one approach for generating the optimal policy of Sections 2–3, in which a one unit increase in x leads to a β^* increase in y. First gather training data set $\tilde{d}$ from some linear policy $\tilde{Y}(x) = \beta_0 + \beta x$, where we take the coefficient β such that we expect the best response $\hat{\beta}(\beta)$ to be above β^*. For instance, we might start from a fixed-point policy $\beta = \beta^{\text{fp}}$ or (if $\rho \geq 0$) from a constant policy with $\beta = 0$. Then add noise to the measurements of x in the data set $\tilde{d}$ to generate a new data set d'. For instance, replace each data point (x_i, η_i) in $\tilde{d}$ with data point (x'_i, η_i) in d', where the new regressor x' is defined as $x'_i = x_i + c + \varepsilon_i$ for $c \in \mathbb{R}$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2_\varepsilon)$. When we linearly regress η on x' in the data set d', attenuation bias establishes that we find a smaller coefficient than $\hat{\beta}(\beta)$: increasing the variance of the noise σ^2_ε from 0 to infinity reduces the estimated coefficient of η on x' from $\hat{\beta}(\beta)$ to 0. For an appropriate level of noise, we hit the optimal coefficient β^*. Finally the constant c, added to or subtracted from all points x', can be adjusted so that the average allocation is equal to μ_η and thus the constant term in the regression is optimal.

We can generalize this estimation with noise to arbitrary estimation procedures on arbitrary data sets. Start with the training data set $\tilde{d}$ induced by some original policy $\tilde{Y}$. To generate the new data set d', add noise—perhaps with nonzero mean—to any manipulable

components of x to get x', while keeping η unchanged.[15] Now define the estimation with noise policy Y^{ews} as

$$Y^{\text{ews}}(x) = ML(x; d').$$

Crucially, when determining the allocation for an agent with observable x, we do not add noise to this agent's x. The noise is only added to the data set on which the algorithm is trained.[16] In other words, Y^{ews} sets each agent's allocation based on an estimate of η, where η is estimated using artificially noised up data. The logic of attenuation bias suggests that Y^{ews} is in some sense "flatter" with respect to the manipulable components of x, or "puts less weight" on those components, relative to the best response policy that does not add noise.

We hope future research will explore this proposal systematically and study its benefits in improving information from manipulable data in complex environments.

References

Ali, S. N. and R. Bénabou (2019): "Image Versus Information: Changing Societal Norms and Optimal Privacy," Unpublished.

Ball, I. (2019): "Incentive-Compatible Prediction," ArXiv:1909.01888v1 [econ.TH].

Bénabou, R. and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, vol. 1, 93–102.

Boleslavsky, R., D. L. Kelly, and C. R. Taylor (2017): "Selloffs, Bailouts, and Feedback: Can Asset Markets Inform Policy?" *Journal of Economic Theory*, 169, 294–343.

[15] Of course, if the manipulable components of x lie in some bounded set such as $\{0, 1\}$, any noise would need to be nonzero mean (at some values of x).

[16] Note that adding noise to the data here does not necessarily mean that the policy function Y^{ews} will be stochastic; indeed, by estimating on resampled data points with independent noise draws, the function can be made essentially deterministic conditional on the true data. In contrast, mechanisms designed to keep agent characteristics hidden from an observer may require stochastic output conditional on the underlying data. See the literature on differential privacy, surveyed in Dwork (2011).

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

Bond, P. and I. Goldstein (2015): "Government Intervention and Information Aggregation by Prices," *Journal of Finance*, 70, 2777–2812.

Braverman, M. and S. Garg (2019): "The Role of Randomness and Noise in Strategic Classification," ACM EC 2019 Workshop on Learning in Presence of Strategic Behavior.

Duffie, D. and P. Dworczak (2018): "Robust Benchmark Design," NBER working paper 20540.

Dwork, C. (2011): "Differential Privacy," *Encyclopedia of Cryptography and Security*, 338–340.

Ederer, F., R. Holden, and M. Meyer (2018): "Gaming and Strategic Opacity in Incentive Provision," *RAND Journal of Economics*, 49, 819–854.

Eliaz, K. and R. Spiegler (2019): "The Model Selection Curse," *American Economic Review: Insights*, 1, 127–40.

Fischer, P. E. and R. E. Verrecchia (2000): "Reporting Bias," *The Accounting Review*, 75, 229–245.

Frankel, A. and N. Kartik (2019): "Muddled Information," *Journal of Political Economy*, 129, 1739–1776.

Gesche, T. (2019): "De-biasing Strategic Communication," Unpublished.

Gómez, E., M. A. Gómez-Villegas, and J. M. Marín (2003): "A Survey on Continuous Elliptical Vector Distributions," *Revista matemática complutense*, 16, 345–361.

Harbaugh, R. and E. Rasmusen (2018): "Coarse grades: Informing the Public by Withholding Information," *American Economic Journal: Microeconomics*, 10, 210–35.

Hardt, M., N. Megiddo, C. Papadimitriou, and M. Wootters (2016): "Strategic Classification," in *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, ACM, 111–122.

Hu, L., N. Immorlica, and J. W. Vaughan (2019): "The Disparate Effects of Strategic Manipulation," in *ACM Conference on Fairness, Accountability, and Transparency*, Atlanta, Georgia.

Jann, O. and C. Schottmüller (2018): "An Informational Theory of Privacy," Unpublished.

Kleinberg, J. and M. Raghavan (2019): "How Do Classifiers Induce Agents to Invest Effort Strategically?" in *Proceedings of the 2019 ACM Conference on Economics and Computation*, ACM, 825–844.

Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

Li, H. (2001): "A Theory of Conservatism," *Journal of Political Economy*, 109, 617–636.

Martinez-Gorricho, S. and C. Oyarzun (2019): "Hypothesis Testing with Endogenous Information," Unpublished.

Milli, S., J. Miller, A. D. Dragan, and M. Hardt (2018): "The Social Cost of Strategic Classification," ArXiv:1808.08460v2 [cs.LG].

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Perez-Richet, E. and V. Skreta (2018): "Test Design Under Falsification," Unpublished.

Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

Zhang, A. (2019): "Competition and Manipulation in Derivative Contract Markets," Unpublished.

Appendices

A. Proofs

A.1. Preliminary Results

Proposition A.1. *There exists $\beta^{\text{fp}} > 0$ satisfying $\hat{\beta}(\beta^{\text{fp}}) = \beta^{\text{fp}}$. If $\rho \geq 0$ there is only one $\beta^{\text{fp}} \geq 0$, and it satisfies $\beta^{\text{fp}} \in (0, 1)$.*

That there is only one positive fixed point under nonnegative correlation has been noted in different form in Frankel and Kartik (2019, Proposition 4).

Proof. For $\beta \geq 0$, Equation 5 can be rewritten as the cubic equation

$$m^2\sigma_\gamma^2\beta^3 + 2m\rho\sigma_\eta\sigma_\gamma\beta^2 + (\sigma_\eta^2 - m\rho\sigma_\eta\sigma_\gamma)\beta - \sigma_\eta^2 = 0. \tag{A.1}$$

The left-hand side of (A.1) is continuous, negative at $\beta = 0$ and tends to ∞ as $\beta \to \infty$. There is a positive solution to (A.1) by the intermediate value theorem.

For the second statement of the proposition, differentiate $\hat{\beta}(\cdot)$ from Equation 5 to obtain

$$\hat{\beta}'(\beta) = -\frac{m\sigma_\eta\sigma_\gamma(2\beta m\sigma_\eta\sigma_\gamma + \rho\sigma_\eta^2 + \rho\beta^2 m^2\sigma_\gamma^2)}{(\sigma_\eta^2 + 2\beta m\rho\sigma_\eta\sigma_\gamma + \beta^2 m^2\sigma_\gamma^2)^2}.$$

When $\rho \geq 0$, this derivative is negative for all $\beta > 0$. The result follows from the fact that $\hat{\beta}(0) = 1$ and, when $\rho \geq 0$, $\hat{\beta}(1) < 1$. *Q.E.D.*

In subsequent proofs, we will appeal to the following standard fact concerning monotone comparative statics.

Fact 1. Let $T \subseteq \mathbb{R}$, $Z \subseteq \mathbb{R}$ be open, and $f : Z \times T \to \mathbb{R}$ be continuously differentiable in z with for all $t \in T$, $\arg\min_{z \in Z} f(z, t) \neq \emptyset$. Define $M(t) \equiv \arg\min_{z \in Z} f(z, t)$. For any $\overline{t} \in T$ and $\underline{t} \in T$ with $\overline{t} > \underline{t}$, it holds that:

1. If $f_z(z, \overline{t}) > f_z(z, \underline{t})$ for all $z \in Z$, then for any $\overline{m} \in M(\overline{t})$ and any $\underline{m} \in M(\underline{t})$ it holds that $\overline{m} < \underline{m}$.

   Proof: For any $\hat{z} > \underline{m}$,

$$f(\hat{z}, \overline{t}) - f(\underline{m}, \overline{t}) = \int_{\underline{m}}^{\hat{z}} f_z(z, \overline{t})\mathrm{d}z > \int_{\underline{m}}^{\hat{z}} f_z(z, \underline{t})\mathrm{d}z = f(\hat{z}, \underline{t}) - f(\underline{m}, \underline{t}) \geq 0.$$

Hence $\overline{m} \leq \underline{m}$. The inequality must be strict because otherwise the first-order conditions yield $0 = f_z(\overline{m}, \overline{t}) = f_z(\underline{m}, \overline{t}) > f_z(\underline{m}, \underline{t}) = 0$.

2. If $f_z(z, \overline{t}) < f_z(z, \underline{t})$ for all $z \in Z$, then for any $\overline{m} \in M(\overline{t})$ and any $\underline{m} \in M(\underline{t})$ it holds that $\overline{m} > \underline{m}$. (We omit a proof, as it is analogous to that above.)

A.2. Proof of Proposition 1

Recall from Subsection 2.2 that (β^*, β_0^*) solves

$$\min_{(\beta,\beta_0)\in\mathbb{R}^2} \mathbb{E}[(m\beta^2\gamma + \beta_0 - (1-\beta)\eta)^2].$$

The first-order condition with respect to β_0 implies

$$\beta_0^* = \mathbb{E}[(1-\beta)\eta - m\beta^2\gamma] = (1-\beta)\mu_\eta - m\beta^2\mu_\gamma.$$

Substituting β_0^* back into the objective, the designer chooses β to minimize

$$\begin{aligned}
&\mathbb{E}[(m\beta^2(\gamma-\mu_\gamma) - (1-\beta)(\eta-\mu_\eta))^2] \\
=&(1-\beta)^2\sigma_\eta^2 + m^2\beta^4\sigma_\gamma^2 - 2(1-\beta)m\beta^2\mathrm{Cov}(\eta,\gamma) \\
=&(1-\beta)^2\sigma_\eta^2 + m^2\beta^4\sigma_\gamma^2 - 2(1-\beta)m\beta^2\rho\sigma_\eta\sigma_\gamma \\
=&\sigma_\eta^2\left[\left((1-\beta) - k\beta^2\right)^2 + 2(1-\rho)\beta^2(1-\beta)k\right],
\end{aligned}$$

where

$$k \equiv m\sigma_\gamma/\sigma_\eta > 0.$$

Equivalently, for any parameters $k > 0$ and $\rho \in (-1, 1)$, the designer chooses β to minimize

$$L(\beta, k, \rho) \equiv \left(k\beta^2 + \beta - 1\right)^2 + 2(1-\rho)\beta^2(1-\beta)k. \tag{A.2}$$

Differentiating,

$$L_\beta(\beta, k, \rho) = -2(1-\beta) + 4k^2\beta^3 + 2\rho k\beta(3\beta - 2). \tag{A.3}$$

Note that $L_\beta(0, k, \rho) < 0$. This means the designer gains a first-order benefit from putting at least some weight on the agent's action, i.e., increasing β from zero. Furthermore, $L_\beta(\beta, k, \rho) \to \infty$ as $\beta \to \infty$.

Proposition 1 is implied by Lemma 1 and the following result. We will abuse notation hereafter and drop the arguments k and ρ from $L(\cdot)$ when those parameters are being held fixed. So, for example, $L(\beta)$ means that both k and ρ are fixed.

Lemma A.1. *There exists $\beta^* \in (0,2)$ such that:*

1. *The loss function $L(\beta)$ from (A.2) is uniquely minimized over $\beta \in \mathbb{R}$ at β^*.*
2. *$\beta^* = \min_{\beta \geq 0}\{\beta : L'(\beta) \geq 0\}$.*
3. *$L''(\beta^*) > 0$.*

Proof. The proof has a few steps below. Steps 1–3 are building blocks to Step 4, which establishes that all minimizers of $L(\beta)$ are in $(0,2)$. Step 5 then establishes there is in fact a unique minimizer, and it has the requisite properties. It is useful in this proof to extend the domain of the function L defined in (A.2) to include $\rho = -1$ and $\rho = 1$.

Step 1: We begin by establishing some properties of $L(\beta, \rho = 1)$. Simplifying (A.2),

$$L(\beta, \rho = 1) = \left(k\beta^2 + \beta - 1\right)^2$$

is the square of a quadratic function. The quadratic function $k\beta^2 + \beta - 1$ is minimized at

$$\beta = \beta^m \equiv -1/(2k) < 0, \tag{A.4}$$

and the function has two roots, one of which is negative and the other is

$$\beta = \overline{\beta} \equiv \frac{-1 + \sqrt{1 + 4k}}{2k} \in (0,2).$$

So $L(\cdot, \rho = 1)$ is minimized at $\overline{\beta}$, and there is no other nonnegative minimizer. Moreover, $L(\cdot, \rho = 1)$ is strictly quasiconvex on $(-\infty, \beta^m]$, strictly decreasing on $[\beta^m, \overline{\beta}]$, and strictly increasing on $[\overline{\beta}, \infty)$. Still further, $L(\cdot, \rho = 1)$ is symmetric around β^m: for any $x \in \mathbb{R}$, $L(\beta^m + x, \rho = 1) = L(\beta^m - x, \rho = 1)$.

Step 2: We claim that for any $\beta < 0$ and $\rho < 1$, there is $\tilde{\beta} \geq 0$ such that $L(\tilde{\beta}) < L(\beta)$. Since $L'(0) < 0$—and hence $L(\beta)$ is not minimized at $\beta = 0$—it follows that for $\rho < 1$, $\arg\min L(\beta, \rho) \subset \mathbb{R}_{++}$.

To prove the claim, we first establish that for any $x > 0$ and $\beta = \beta^m - x$ (where β^m is defined in (A.4)), the symmetric point $\beta^m + x$ has a lower loss when $\rho < 1$; note that $\beta^m + x$ may also be negative. The argument is as follows:

$$\begin{aligned} L(\beta^m - x, \rho) - L(\beta^m + x, \rho) &= L(\beta^m - x, \rho = 1) + 2(1-\rho)(\beta^m - x)^2(1 - \beta^m + x)k \\ &\quad - \left[L(\beta^m + x, \rho = 1) + 2(1-\rho)(\beta^m + x)^2(1 - \beta^m - x)k\right] \\ &= 2(1-\rho)k\left[(\beta^m - x)^2(1 - \beta^m + x) - (\beta^m + x)^2(1 - \beta^m - x)\right] \\ &= 4(1-\rho)kx\left(\beta^m(3\beta^m - 2) + x^2\right) \\ &\geq 0, \end{aligned}$$

where the first equality is from the definition of $L(\cdot)$ in Equation A.2, the second is because Step 1 established that $L(\beta^m + x, \rho = 1) = L(\beta^m - x, \rho = 1)$, the third equality is algebraic simplification, and the inequality is because $\beta^m < 0$, $x > 0$, and $\rho < 1$.

It now suffices to establish that $L(0, \rho) < L(\beta, \rho)$ for all $\beta \in [\beta^m, 0)$. Differentiating (A.3) yields $L_{\beta\rho}(\beta, \rho) = 2k\beta(3\beta - 2) > 0$ when $\beta < 0$. Hence for $\beta \in [\beta^m, 0)$, $L(0, \rho) - L(\beta, \rho) \leq L(0, \rho = 1) - L(\beta, \rho = 1) < 0$, where the strict inequality is from Step 1.

Step 3: $\arg\min_\beta L(\beta, \rho = -1) \cap (0, 2] \neq \emptyset$.

To prove this, begin by simplifying (A.2) to get

$$L(\beta, \rho = -1) = \left(k\beta^2 - \beta + 1\right)^2.$$

The quadratic function $k\beta^2 - \beta + 1$ is strictly convex in β and is minimized at $\beta = 1/(2k)$; moreover, if $k \geq 1/4$ then the function is nonnegative for all β, and otherwise it is equal to zero at $\beta = \frac{1 \pm \sqrt{1-4k}}{2k}$. It follows that if $k \geq 1/4$, $\arg\min L(\beta, \rho = -1) = \{1/(2k)\}$, and hence the unique minimizer is in $(0, 2]$. If $k < 1/4$, $\arg\min L(\beta, \rho = -1) = \{\frac{1-\sqrt{1-4k}}{2k}, \frac{1+\sqrt{1-4k}}{2k}\}$, and routine algebra verifies that the smaller minimizer, $\frac{1-\sqrt{1-4k}}{2k}$, is in $(0, 2)$.

Step 4: For $\rho \in (-1, 1)$, $\arg\min_\beta L(\beta, \rho) \subset (0, 2)$.

This follows from a standard monotone comparative statics argument (see Fact 1): since $L_{\beta\rho}(\beta, \rho) = 2k\beta(3\beta - 2) > 0$ when $\beta > 2/3$, on the domain $(2/3, \infty)$ every minimizer of $L(\cdot, \rho)$ when $\rho > -1$ is smaller than every minimizer of $L(\cdot, \rho = -1)$. Step 3 then implies that all minimizers for $\rho > -1$ are less than 2; Step 2 established that when $\rho < 1$, all minimizers are larger than 0.

Step 5: Finally, we claim that for $\rho \in (-1, 1)$, $L'(\beta)$ has only one root in $(0, 2)$; moreover,

$L''(\beta) > 0$ at that root. The lemma then follows because $L'(\beta)$ is continuous and $L'(0) < 0$.

To prove the claim, we begin by observing from Equation A.3 that $L'(\beta)$ is cubic function that is initially strictly concave and then strictly convex, with an inflection point at $\beta = -\rho/(2k)$. For the rest of the proof, arguments of L' or L'' refer to values of β.

1. If $\rho \geq 0$, then the inflection point is negative, and thus L' is strictly convex on $\beta > 0$. Since $L'(0) < 0$, L' has only one positive root, and $L'' > 0$ at that root.

2. Now consider $\rho \in (-1, 0)$. L'' is minimized at the inflection point of L'. Differentiating Equation A.3, it holds at the inflection point that

$$L''\left(\frac{-\rho}{2k}\right) = 2 + 12k^2\left(\frac{-\rho}{2k}\right)^2 + 4\rho k\left(3\left(\frac{-\rho}{2k}\right) - 1\right) = 2 - 3\rho^2 - 4k\rho.$$

   If this expression is positive, then $L''(\beta) > 0$ for all β, i.e., L' is strictly increasing and hence has a unique root.

   So suppose instead that $2 - 3\rho^2 - 4k\rho \leq 0$. Equivalently, since $\rho < 0$, suppose that

$$k \leq \frac{2 - 3\rho^2}{4\rho}.$$

   The right-hand side of this inequality is less than $-\rho/4$ because $\rho \in (-1, 0)$, and hence $k < -\rho/4$. Consequently, the inflection point, $\beta = -\rho/(2k)$, is larger than 2, and therefore $L'(\beta)$ is concave over $\beta \in (0, 2)$. Moreover, recall that $L'(0) < 0$, and also observe that $L'(2) = 32k^2 + 16k\rho + 2 > 0$.[17] It follows that L' has only one root on $(0, 2)$, and $L'' > 0$ at that root. *Q.E.D.*

A.3. Proof of Lemma 1

It holds that

$$\mathbb{E}[(\hat{\eta}_\beta(x) - \eta)^2] = \sigma_\eta^2\left(1 - R_{\eta x}^2\right) = \sigma_\eta^2 - (\hat{\beta}(\beta))^2\mathrm{Var}(x),$$

where the first equality was noted in the main text after Equation 6, and the second equality holds because $R_{\eta x}^2 = (\mathrm{Cov}(x, \eta))^2/(\mathrm{Var}(\eta)\mathrm{Var}(x))$ and $\hat{\beta}(\beta) = \mathrm{Cov}(x, \eta)/\mathrm{Var}(x)$. Recall that $\mathrm{Var}(x) > 0$ because of our maintained assumption that $\rho \in (-1, 1)$.

[17] To see that $L'(2) > 0$, observe that the quadratic expression $32k^2 + 16k\rho + 2$ is minimized over choice of k at $k = -\rho/4$, at which point its value is $-2\rho^2 + 2$. Since $k < -\rho/4$, we have $L'(2) > -2\rho^2 + 2$, and this right-hand side is larger than 0 because $\rho \in (-1, 0)$.

We also have

$$\begin{aligned}\mathbb{E}[(Y(x)-\hat{\eta}_\beta(x))^2] &= \mathbb{E}[(\beta x+\beta_0-\hat{\beta}(\beta)x-\hat{\beta}_0(\beta))^2] \qquad \text{from definitions}\\ &= \mathbb{E}\left[\Big((\beta-\hat{\beta}(\beta))(x-\mathbb{E}[x])\Big)^2\right]\\ &= (\beta-\hat{\beta}(\beta))^2\mathrm{Var}(x),\end{aligned}$$

where the second line is because the choice of β_0 and $\hat{\beta}(\beta)$ are such that $\beta\mathbb{E}[x]+\beta_0=\mu_\eta=\hat{\beta}(\beta)\mathbb{E}[x]+\hat{\beta}_0(\beta)$ (the second equality here is standard; for the first, see the beginning of the proof of Proposition 1) and hence $\beta_0-\hat{\beta}_0(\beta)=(\hat{\beta}(\beta)-\beta)\mathbb{E}[x]$.

Substituting these formulae into Equation 6 yields

$$\mathcal{L}(\beta)=\underbrace{\sigma_\eta^2-(\hat{\beta}(\beta))^2\mathrm{Var}(x)}_{\text{Info loss}}+\underbrace{(\beta-\hat{\beta}(\beta))^2\mathrm{Var}(x)}_{\text{Misallocation loss}}.$$

Differentiating,

$$\mathcal{L}'(\beta)=\overbrace{\left(-2\hat{\beta}(\beta)\hat{\beta}'(\beta)\mathrm{Var}(x)-(\hat{\beta}(\beta))^2\frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Var}(x)\right)}^{\text{Marginal change in info loss}}+\underbrace{\left(-2(\beta-\hat{\beta}(\beta))\hat{\beta}'(\beta)\mathrm{Var}(x)+(\beta-\hat{\beta}(\beta))^2\frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Var}(x)\right)}_{\text{Marginal change in misallocation loss}}.$$

When $\beta=\beta^{\mathrm{fp}}=\hat{\beta}(\beta^{\mathrm{fp}})$, the marginal change in misallocation loss is evidently zero (intuitively because the misallocation loss is minimized at $\beta=\beta^{\mathrm{fp}}$). Thus,

$$\mathcal{L}'(\beta^{\mathrm{fp}})=-2\beta^{\mathrm{fp}}\hat{\beta}'(\beta^{\mathrm{fp}})\mathrm{Var}(x)-(\beta^{\mathrm{fp}})^2\frac{\mathrm{d}}{\mathrm{d}\beta}\mathrm{Var}(x).$$

Using $\mathrm{Var}(x)=\mathrm{Cov}(x,\beta)/\hat{\beta}(\beta)$, $\mathrm{Cov}(x,\beta)=\sigma_\eta^2+m\rho\sigma_\eta\sigma_\gamma\beta$, $\mathrm{Var}(x)=\mathrm{Cov}(x,\beta)+m\rho\sigma_\eta\sigma_\gamma\beta+m^2\sigma_\gamma^2\beta^2$, and $\beta^{\mathrm{fp}}=\hat{\beta}(\beta^{\mathrm{fp}})$, some algebra then yields[18]

$$\mathcal{L}'(\beta^{\mathrm{fp}})=\frac{2m^2}{\mathrm{Var}(x)}(\beta^{\mathrm{fp}})^2\sigma_\eta^2\sigma_\gamma^2(1-\rho^2).$$

Since $\beta^{\mathrm{fp}}\neq 0$ (as $\hat{\beta}(0)=1$ from Equation 5) and $\rho\in(-1,1)$, it follows that $\mathcal{L}'(\beta^{\mathrm{fp}})>0$.

[18] Letting C and V be shorthand for $\mathrm{Cov}(x,\beta)$ and $\mathrm{Var}(x)$ respectively, a prime denote the derivative

A.4. Proof of Proposition 2

The proof is via the following claims. Applying Lemma A.1, we without loss restrict attention to $\beta \in (0, 2)$ throughout this subsection.

Claim A.1. *β^* is continuously differentiable in ρ and k.*

Proof. Lemma A.1 established that $\text{sign}[L''(\beta^*)] > 0$. Thus, the implicit function theorem guarantees the existence of $\frac{\mathrm{d}\beta^*}{\mathrm{d}k} = -\frac{L_{\beta k}}{L_{\beta\beta}}$ and $\frac{\mathrm{d}\beta^*}{\mathrm{d}\rho} = -\frac{L_{\beta\rho}}{L_{\beta\beta}}$. *Q.E.D.*

Claim A.2. *If $k > 3/4$ then $\beta^* < 2/3$ and is strictly increasing in ρ. If $k < 3/4$ then $\beta^* > 2/3$ and is strictly decreasing in ρ. If $k = 3/4$ then $\beta^* = 2/3$ independent of ρ.*

Proof. From Equation A.3 compute the cross partial

$$L_{\beta\rho} = 2k\beta(3\beta - 2).$$

Hence $L_{\beta\rho} < 0$ when $\beta < 2/3$, while $L_{\beta\rho} > 0$ when $\beta > 2/3$. Moreover, it follows from Equation A.3 that when $\beta = 2/3$, $\text{sign}[L_\beta] = \text{sign}[k - 3/4]$ independent of ρ.

1. Consider $k = 3/4$. Routine algebra verifies that L_β is strictly increasing in β, and hence $L_\beta = 0 \implies \beta = 2/3$, i.e., $\beta^* = 2/3$ independent of ρ.

2. Consider $k > 3/4$. Since $L_\beta > 0$ when $\beta = 2/3$, it follows that $\beta^* < 2/3$. (Recall $L_\beta < 0$ when $\beta = 0$, and Lemma A.1 implies that $\beta^* = \min\{\beta > 0 : L_\beta = 0\}$.) Since $L_{\beta\rho} < 0$ on the domain $\beta < 2/3$, monotone comparative statics imply β^* is strictly increasing in ρ.

3. Consider $k < 3/4$. For $\rho = 0$, we have $L_{\beta k} = 8k\beta^3 > 0$ and hence $\beta^* > 2/3$ using $\beta^* = 2/3$ when $k = 3/4$ and monotone comparative statics. It follows that $\beta^* > 2/3$ for all ρ because β^* is continuous in ρ and $L_\beta < 0$ when $\beta = 2/3$ whereas $L_\beta = 0$ when

---

with respect to β, suppressing arguments, evaluating all functions at β^{fp}, and using the properties noted:

$$\begin{aligned}
\mathcal{L}' &= -2\beta^{\text{fp}}\hat{\beta}'V - (C/V)^2V' = -2C\hat{\beta}' - (C/V)^2V' = (-2CVC' + C^2V')/V^2 \\
&= (C/V^2)\left[-2CC'(C + m\rho\sigma_\eta\sigma_\gamma\beta^{\text{fp}} + m^2\sigma_\gamma^2(\beta^{\text{fp}})^2) + C^2(2C' + 2m^2\sigma_\gamma^2\beta^{\text{fp}})\right] \\
&= (2\beta^{\text{fp}}C/V^2)\left[-C'(m\rho\sigma_\eta\sigma_\gamma + m^2\sigma_\gamma^2\beta^{\text{fp}}) + Cm^2\sigma_\gamma^2\right] \\
&= (2\beta^{\text{fp}}C/V^2)\left[-(m\rho\sigma_\eta\sigma_\gamma)^2 - (m\rho\sigma_\eta\sigma_\gamma)m^2\sigma_\gamma^2\beta^{\text{fp}}) + (\sigma_\eta^2 + m\rho\sigma_\eta\sigma_\gamma\beta^{\text{fp}})m^2\sigma_\gamma^2\right] \\
&= (2\beta^{\text{fp}}C/V^2)m^2(\sigma_\eta\sigma_\gamma)^2(1-\rho^2) = 2(\beta^{\text{fp}})^2(1/V)m^2\sigma_\eta^2\sigma_\gamma^2(1-\rho^2).
\end{aligned}$$

$\beta = \beta^*$. Since $L_{\beta\rho} > 0$ on the domain $\beta > 2/3$, monotone comparative statics imply β^* is strictly decreasing in ρ. *Q.E.D.*

Claim A.3. *As $k \to \infty$, $\beta^* \to 0$; as $k \to 0$, $\beta^* \to 1$. If $\rho \geq 0$ then β^* is strictly decreasing in k. If $\rho < 0$ then β^* is strictly quasi-concave in k, attaining a maximum at some point.*

Proof. The first statement about limits is evident from inspecting Equation A.3. For the comparative statics, compute the cross partials

$$L_{\beta k} = 8k\beta^3 + 2\rho\beta(3\beta - 2) \quad \text{and} \quad L_{\beta kk} = 8\beta^3 > 0.$$

Since $\frac{\mathrm{d}\beta^*}{\mathrm{d}k} = -\frac{L_{\beta k}}{L_{\beta\beta}}$ and, from Lemma A.1, $L_{\beta\beta} > 0$ at $\beta = \beta^*$, the sign of $\frac{\mathrm{d}\beta^*}{\mathrm{d}k}$ is the sign of $-L_{\beta k}$. Using $\beta^* \to 1$ as $k \to 0$, we see that for small k and at $\beta = \beta^*$, $L_{\beta k}$ is larger than but arbitrarily close to 2ρ.

1. It follows that $L_{\beta k} > 0$ for all k and $\beta = \beta^*$ when $\rho \geq 0$. That is, $\frac{\mathrm{d}\beta^*}{\mathrm{d}k} < 0$ when $\rho \geq 0$.

2. Consider $\rho < 0$. Plainly $L_{\beta k} < 0$ for small k and $\beta = \beta^*$, while for some k it becomes positive (since $\beta^* \to 0$ as $k \to \infty$). Since $L_{\beta k}$ is strictly increasing in k, it follows that $\frac{\mathrm{d}\beta^*}{\mathrm{d}k}$ is strictly decreasing in k, initially positive and eventually negative. *Q.E.D.*

Claim A.4. *Assume $\rho = 0$. There is a unique β^{fp}, which is positive. Both β^{fp} and $\beta^*/\beta^{\mathrm{fp}}$ are strictly decreasing in k. Moreover, $\beta^*/\beta^{\mathrm{fp}} \to 1$ as $k \to \infty$ and $\beta^*/\beta^{\mathrm{fp}} \to \sqrt[3]{1/2}$ as $k \to 0$.*

Proof. Assume $\rho = 0$. Equation A.1 simplifies to

$$k^2(\beta^{\mathrm{fp}})^3 + \beta^{\mathrm{fp}} - 1 = 0, \tag{A.5}$$

which has a unique solution, with $\beta^{\mathrm{fp}} \in (0, 1)$ strictly decreasing in k with range $(0, 1)$.

The first order condition for β^* simplifies to

$$2k^2(\beta^*)^3 + \beta^* - 1 = 0, \tag{A.6}$$

which has a unique solution, also in $(0, 1)$ and strictly decreasing in k with range $(0, 1)$.

Hence, $\beta^*/\beta^{\mathrm{fp}} \to 1$ as $k \to 0$. Moroever, Equation A.5 and Equation A.6 imply that as $k \to \infty$, $k^2(\beta^{\mathrm{fp}})^3 \to 1$ and $2k^2(\beta^*)^3 \to 1$, and hence $(\beta^*/\beta^{\mathrm{fp}}) \to \sqrt[3]{1/2}$.

It remains to prove that $\beta^*/\beta^{\text{fp}}$ is strictly decreasing in k. Applying the implicit function theorem to Equation A.5 and Equation A.6 (which is indeed valid) and doing some algebra,

$$\frac{\mathrm{d}\beta^*}{\mathrm{d}k} = -\frac{4k(\beta^*)^3}{6k^2(\beta^*)^2+1},$$
$$\frac{\mathrm{d}\beta^{\text{fp}}}{\mathrm{d}k} = -\frac{2k(\beta^{\text{fp}})^3}{3k^2(\beta^{\text{fp}})^2+1}.$$

$\beta^*/\beta^{\text{fp}}$ is strictly decreasing in k if and only if $\beta^{\text{fp}}\frac{\mathrm{d}\beta^*}{\mathrm{d}k} - \beta^*\frac{\mathrm{d}\beta^{\text{fp}}}{\mathrm{d}k} < 0$. Substituting in the formulae above, this inequality is equivalent to

$$\frac{2k(\beta^{\text{fp}})^3\beta^*}{3k^2(\beta^{\text{fp}})^2+1} < \frac{4k(\beta^*)^3\beta^{\text{fp}}}{6k^2(\beta^*)^2+1}$$
$$\iff \left(6k^2(\beta^*)^2+1\right)(\beta^{\text{fp}})^2 < \left(3k^2(\beta^{\text{fp}})^2+1\right)2(\beta^*)^2$$
$$\iff \beta^{\text{fp}} < \beta^*\sqrt{2}.$$

Plainly, the last inequality holds as $k \to 0$ because both $\beta^{\text{fp}} \to 1$ and $\beta^* \to 1$ as $k \to 0$. By continuity, we are done if there is no k at which $\beta^{\text{fp}} = \beta^*\sqrt{2}$. Indeed there is not because then Equation A.5 would become equivalent to

$$2k^2(\beta^*)^3 + \beta^* - 1/\sqrt{2} = 0,$$

contradicting Equation A.6. *Q.E.D.*

B. Alternative Model of Information Loss

Let the agent take action $x \in \{0,1\}$ with natural action $\eta \in \{0,1\}$. The agent's type η is her private information, drawn with ex-ante probability $\pi \in (0,1)$ that $\eta = 1$. After observing x, the designer chooses allocation $y \in \mathbb{R}$ with payoff $-(y-\eta)^2$. We assume, for simplicity, that the agent of type $\eta = 1$ must choose $x = 1$.[19] The payoff for type $\eta = 0$ is $y - cx$, where $c > 0$ is a commonly known parameter. To streamline the analysis, we assume $c \in (0, \pi)$.

A pure allocation rule or policy is $Y : \{0,1\} \to \mathbb{R}$. Due to the designer's quadratic loss

[19] Our main point goes through so long as action $x = 1$ is no more costly than $x = 0$ for type $\eta = 1$, as this will ensure it is optimal for type $\eta = 1$ to choose $x = 1$.

payoff, it is without loss to focus on pure policies. Given a policy Y, let $\Delta \equiv Y(1) - Y(0)$ be the difference in allocations across the two actions of the agent. We focus, without loss, on policies with $\Delta \geq 0$. A policy with a smaller Δ is a "flatter" policy, i.e., it is less sensitive to the agent's action. The naive policy Y^{n} sets $Y^{\mathrm{n}}(1) = 1$ and $Y^{\mathrm{n}}(0) = 0$, corresponding to a naive allocation difference of $\Delta^n = 1$. Let Δ^{fp} and Δ^* denote the corresponding differences from fixed point and commitment policies.

B.1. Naive Policy

Take any policy with $\Delta = 1$. Since we assume $c < \pi < 1$, even the agent with $\eta = 0$ will then choose $x = 1$. So welfare—the designer's ex-ante expected payoff—from the naive policy is

$$-\pi(0-0)^2 - (1-\pi)(1-0)^2 = -(1-\pi).$$

B.2. Fixed Point

At a Bayesian Nash equilibrium (of either the simultaneous move game, or when the agent moves first), $Y(x) = \mathbb{E}[\eta|x]$ for any x on the equilibrium path. If $x = 0$ is on the equilibrium path, $Y(0) = 0$ because type $\eta = 1$ does not play $x = 0$.

There is a fully-pooling equilibrium with both types playing $x = 1$: the designer plays $Y(1) = \pi$ and $Y(0) = 0$, and it is optimal for type $\eta = 0$ to play $x = 1$ because $c < \pi$. The corresponding welfare is

$$-\pi(\pi-1)^2 - (1-\pi)(\pi-0)^2 = -\pi(1-\pi).$$

There is no equilibrium in which the agent of type $\eta = 0$ puts positive probability on action $x = 0$, because that would imply $Y(1) > \pi$ and $Y(1) = 0$, against which the agent's unique best response is to play $x = 1$.

Therefore, we have identified the (essentially unique, up to the off-path allocation following $x = 0$) fixed point policy: $Y^{\mathrm{fp}}(1) = \pi$, $Y^{\mathrm{fp}}(0) = 0$, and therefore $\Delta^{\mathrm{fp}} = \pi$. The agent pools on $x = 1$, and welfare is $-\pi(1-\pi)$.[20] This welfare is larger than that of the naive policy.

[20] The choice of $Y^{\mathrm{fp}}(0) = 0$ can be justified from the perspective of the agent "trembling". In particular, in the signaling game where the agent moves before the designer, any sequential equilibrium (Kreps and Wilson, 1982) has $Y(0) = 0$, as only type $\eta = 0$ can play $x = 0$. But note that no matter how $Y(0)$ is specified, it must hold in a fixed point that $\Delta \leq c$; otherwise the agent will not pool at $x = 1$.

B.3. Commitment

Now suppose the designer's commits to a policy before the agent moves. From the earlier analysis, if $\Delta > c$ the agent will pool at $x = 1$ and so an optimal such policy is the fixed point policy Y^{fp}. For any $\Delta < c$, there is full separation: the agent's best response is $x = \eta$. Indeed, full separation is also a best response for the agent when $\Delta = c$. Given that the designer wants to match the agent's type, it follows that the optimal way to induce full separation is to set $\Delta = c$ (or $\Delta = c^-$), i.e., have $Y^*(1) = Y^*(0) + c$.

At such an optimum, quadratic loss utility implies that the designer sets an average action of $(1 - \pi)Y^*(0) + \pi Y^*(1)$ equal to $\mathbb{E}[\eta] = \pi$. Plugging in $Y^*(1) = Y^*(0) + c$ yields

$$(1 - \pi)Y^*(0) + \pi(Y^*(0) + c) = \pi,$$

and hence the solution

$$Y^*(0) = \pi(1 - c), \quad Y^*(1) = \pi(1 - c) + c.$$

The corresponding welfare is

$$-(1 - \pi)(\pi(1 - c) - 0)^2 - \pi(\pi(1 - c) + c - 1)^2 = -(1 - c)^2(1 - \pi)\pi.$$

This welfare is larger than that under the fixed point. Moreover, the optimal policy has $\Delta^* = c$ while the fixed point has $\Delta^{\text{fp}} = \pi$ and the naive policy has $\Delta^{\text{n}} = 1$. Thus the optimal policy is flatter than the fixed point, which in turn is flatter than the naive policy:

$$\Delta^* < \Delta^{\text{fp}} < \Delta^{\text{n}}.$$

Note that the designer obtains no benefit from reducing Δ from $\Delta^{\text{fp}} = \pi$ until reaching $\Delta^* = c$; this is an artifact of the assumption that there is no heterogeneity in the manipulation cost c. In a model with such heterogeneity, there would be a more continuous benefit of reducing Δ from the fixed point.