

# Incentivizing Information Acquisition<sup>\*</sup>

Fan Wu<sup>†</sup>

March 5, 2025

## Abstract

I study a principal-agent model in which a principal hires an agent to collect information about an unknown continuous state. The agent acquires a signal whose distribution is centered around the state, controlling the signal's precision at a cost. The principal observes neither the precision nor the signal, but rather, using transfers that can depend on the state, incentivizes the agent to choose high precision and report the signal truthfully. I identify a sufficient and necessary condition on the agent's information structure which ensures that there exists an optimal transfer with a simple cutoff structure: the agent receives a fixed prize when his prediction is close enough to the state and receives nothing otherwise. This condition is mild and applies to all signal distributions commonly used in the literature.

*Keywords.* Information Acquisition; Principal-Agent Model.

*JEL Codes.* C73, D82, D83, D86

---

<sup>\*</sup>I am indebted to my committee chair Luciano Pomatto and advisor Omer Tamuz for their generous encouragement, guidance, and support. This paper also benefits from insightful discussions with Alexander Bloedel, Peter Caradonna, Axel Niemeyer, and Kun Zhang. For helpful comments, I also thank S. Nageeb Ali, Pak Hung Au, Daghan Carlos Akkar, Simon Board, Simone Cerreia-Vioglio, Daniel Gottlieb, Simon Grant, Nima Haghpahan, Wei He, Jonathan Libgober, Shuo Liu, Yi Liu, Jay Lu, Yusufcan Masatlioglu, Doron Ravid, Fedor Sandomirskiy, Balazs Szentes, Xin Shan, Andrew J. Sinclair, Jakub Steiner, Rakesh Vohra, Mu Zhang, Mengxi Zhang, Weijie Zhong, seminar participants at Caltech, UCLA, HKU, HKUST, PKU, THU, and conference participants at 2024 AMES, 2024 Midwest Theory Meeting.

<sup>†</sup>Fan Wu: Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125 (email: fwu2@caltech.edu).

# 1 Introduction

In many situations, an agent is tasked with collecting information about a quantity of interest. Examples include a statistician collecting data for the Bureau of Statistics, a meteorologist forecasting weather for the National Weather Service, or a consultant assessing the profitability of a new market. A natural question is how to design a contract to incentivize the agent to collect information and report it accurately.

A difficulty in designing such contracts is that the agent’s effort and the information acquired are often unobservable to the principal. For example, a statistician might be able to fabricate part of their data. Similarly, in the case of a meteorologist analyzing weather patterns, the amount of time and effort spent on the analysis might not be easily observable. In the case of the consultant, predicting market profitability may rely on private connections or specialized skills, which are often proprietary. Therefore, the contract between the principal and the agent must take into account moral hazard.

In this paper, I study the design of an optimal contract for a principal (she) that incentivizes the agent (he) to gather information and maximize its precision. The agent conducts a location-scale experiment that is centered around the state and controls the precision (the inverse of the scale) at a cost. Eventually, the state is revealed and the contract can depend on the state and the agent’s report. I abstract from the principal’s broader decision problem—that is, the way in which the principal uses the information generated by the agent—and focus on the problem where the principal’s objective is only to maximize precision subject to some fixed budget constraint on transfers. My analysis and results shall be robust to whatever the broader decision problem may be. The budget reflects the maximum amount of transfer that the principal is able to use.

In practice, a budget that can only be used for a specific task is commonly observed. For example, government agencies typically operate within budget limits imposed by government funding and must return any unused funds to the government.<sup>1</sup> There are numerous reports of unspent government funding being returned to the government, such as [Mansouri \(2024\)](#), [Cuccia \(2023\)](#).<sup>2</sup> In firms, it is also common

---

<sup>1</sup>[Schick \(2008\)](#) provides an in-depth look at how the federal budget is structured and constrained by funding limits imposed by the government.

<sup>2</sup>Federal regulations also specify where unspent grant money is returned, as seen at [National Archives \(2024\)](#).

for agents to operate under a budget and to eventually return unused funds to the firm. [Anthony et al. \(2007\)](#) and [Horngren \(2009\)](#) provide comprehensive empirical evidence for this practice.<sup>3</sup> In the consulting example, the transfer could take the form of a promotion, a recommendation, or a grade. Sometimes, budgets are restricted by external funding designated for specific purposes, such as university grants.

In such settings, a simple contract is to allocate the entire budget to the agent if their report is close to the actual state, and withhold payment otherwise. I call these incentive schemes *cutoff transfers*. The cutoff transfer is the simplest transfer rule: It is straightforward to understand and easy to implement. The literature on principal-agent problems has long been interested in when simple contracts are optimal ([Carroll, 2015](#); [Herweg et al., 2010](#); [Oyer, 2000](#); [Gottlieb and Moreira, 2022](#)).

My main result is that cutoff transfers are optimal across a large variety of settings. In particular, I identify a sufficient and necessary condition on the agent’s information structure such that for all cost functions, there exists an optimal transfer that is a cutoff transfer.

Formally, I study a model in which the principal wants to incentivize the agent to acquire information on an underlying state  $\theta \in \mathbb{R}$ . For simplicity, I assume the state  $\theta$  admits an improper uniform common prior.<sup>4</sup> The agent can acquire a costly signal  $s \in \mathbb{R}$  about the state. The signal takes the form  $s = \theta + \frac{1}{\lambda}\varepsilon$  where  $\varepsilon$  is drawn from some symmetric and single-peaked probability density function, and  $\lambda$  is a measure of precision controlled by the agent. An important example is the case where  $\varepsilon$  admits a standard Gaussian distribution. The agent chooses precision  $\lambda$  at a cost  $c(\lambda)$ .

The principal observes neither the agent’s signal  $s$  nor the agent’s choice of precision  $\lambda$ . The agent makes a report  $a \in \mathbb{R}$  after observing the signal. The principal wants to incentivize the agent to maximize precision and truthfully report his signal. She does so by a transfer rule  $t$  that depends on the report  $a$  and the state  $\theta$ , which is eventually observed by both players.<sup>5</sup> I impose limited liability ( $t \geq 0$ ) and limited budget. The budget can only be used for this task. The agent’s payoff is his

---

<sup>3</sup>[Anthony et al. \(2007\)](#) offers an extensive analysis of organizational budget management, highlighting the frequent practice of returning unspent funds to a central or general fund in both private and public sectors. [Horngren \(2009\)](#) explores cost control and budget allocation, noting how unused resources are often returned to a central pool for future use.

<sup>4</sup>I generalize the model to the case where the state is multi-dimensional and the case with Gaussian prior.

<sup>5</sup>I study the case where the state is unobservable but the principal has a private signal about the state later.

expected utility on transfer minus the cost. I allow any increasing utility function, which accommodates arbitrary risk attitude of the agent.

I refer to a transfer rule as *optimal* if it induces the maximum precision among all transfers and elicits truthful reports. There are at least two difficulties in this problem. First, the transfer design problem is an infinite-dimensional optimization problem, as the transfer rule itself is a function. Second, there are two layers of incentive compatibility issues. The principal does not observe the precision chosen by the agent, which is a moral hazard problem, and she does not observe the signal either, leading to a communication problem.

I say that a transfer rule is a *cutoff* transfer if it pays the entire budget when the distance between the report and the state is below a cutoff  $d$  and pays 0 otherwise. For which signal distributions are cutoff transfers optimal?

My main result (Theorem 1) identifies a sufficient and necessary condition for cutoff transfers to be optimal. This condition is weaker than the monotone likelihood ratio property (Milgrom, 1981; Rogerson, 1985; Jewitt, 1988) and is satisfied by most common distributions, including Gaussian, Laplace, logistic, and the uniform distribution. Theorem 1 shows that if this condition holds, then for all cost functions cutoff transfers are optimal. Conversely, if the condition does not hold, then there is a cost function for which no cutoff transfer is optimal. Moreover, this cost function is not pathological and can be taken to be increasing and differentiable. Furthermore, when the condition holds, I characterize the optimal cutoff (Theorem 2).

An intuition for the optimality of cutoff rules is that they provide the strongest incentives among all contracts. Loosely speaking, since more precise signal structures generate signals that are “more concentrated” around the state, a precision-maximizing contract must pay the agent more for the report closer to the state. The cutoff transfers take this logic to the extreme: the principal exhausts her entire budget when the report is sufficiently close to the state and pays nothing otherwise. Importantly, however, this intuition is incomplete, because the notion of “more concentrated” signals is imprecise. My condition is exactly what is needed to make this logic tight.

All the results generalize to the  $n$ -dimensional case. The signal distribution is still symmetric, single-peaked, and centered around  $\theta$ . Here the symmetry means that the density of the distribution depends only on the Euclidean distance from the state  $\theta$ . A cutoff transfer pays 1 when the Euclidean distance between the report and the

state is less than a cutoff  $d$ . In this setting I show that an analogous result holds (Proposition 2).

I also generalize my results to the case of a proper Gaussian prior and Gaussian signals. I show that for all cost functions, cutoff transfers are optimal (Proposition 3).

In addition, I extend my results to the setting where the state is unobservable. Instead, the principal privately receives a signal about the state. When designing the transfer, she uses her signal instead of the state to discipline the agent. I show that given a Gaussian or uniform prior and Gaussian signal, cutoff transfers are optimal (Proposition 4).

Lastly, I apply my results to a classic principal-agent problem, offering new insights into the optimality of simple contracts. In the classic setting, a principal incentivizes the agent to produce an output. Unlike the traditional framework, in which the principal’s goal is to maximize expected payoff (of outputs) minus transfers, I assume the principal’s sole objective is to maximize output, constrained by a budget limit.<sup>6</sup> As a corollary of my main result, the monotone likelihood ratio property is sufficient to ensure the optimality of cutoff transfers (Corollary 3). Here, the cutoff transfer pays the entire budget if the output is above a cutoff. This corollary provides an alternative explanation for simple contracts: the optimality of cutoff transfers stems from the nature of budgets in many contracting scenarios. In contrast, closed-form solutions for the optimal transfers in the classic setting are generally not obtainable (see Bolton and Dewatripont, 2004, Chapter 4.5).

One point to note is that my proof relies on techniques from monotone comparative statics. This allows me to analyze the problem under weak assumptions on the cost function and signal distribution. In particular, I do not need to assume the validity of the first-order approach (Rogerson, 1985; Jewitt, 1988).

## 1.1 Related Literature

My paper is related to the literature on incentivizing information acquisition (see, e.g. Osband, 1989; Li and Libgober, 2023; Li et al., 2022; Neyman et al., 2021; Zermeno, 2011; Carroll, 2019; Chen et al., 2023; Chade and Kovrijnykh, 2016; Whitmeyer and

---

<sup>6</sup>This is a common practice in firms, see Anthony et al. (2007) and Horngren (2009).

Zhang, 2023; Sharma et al., 2023; Clark and Reggiani, 2021).<sup>7</sup>

In Li et al. (2022), an agent exerts a binary level of effort to refine a posterior from a prior. The paper studies optimizing proper scoring rules by maximizing the increase in the score with effort. In Li and Libgobber (2023), a principal hires an agent to learn about a binary state. The agent acquires information over time through a Poisson information arrival technology. The principal rewards the agent with a fixed-value prize as a function of the agent’s sequence of reports and the state. Li and Libgobber (2023) identify conditions that it is without loss to elicit a single report after all information has been acquired.

Osband (1989) studies a principal with a quadratic prediction cost who incentivizes an expert to collect information. The expert can increase the precision (the inverse of variance) of the prediction at a constant cost. The principal minimizes the sum of the expected error variance and the expected transfer. In this stylized setting, the optimal transfer consists of a quadratic report error term plus a linear term on a initial belief error, with three parameters.

In Whitmeyer and Zhang (2023) and Sharma et al. (2023), a rational inattentive agent can acquire information flexibly subject to a posterior separable cost. The principal wants to minimize the expected monetary cost of implementing a given information structure. Similarly, Clark and Reggiani (2021) study the Pareto optimal contract that maximizes social welfare. In all these papers, both the experiment and the signal are unobservable to the principal. In Rappoport and Somma (2017), a principal hires an agent to acquire costly information to influence a third party’s decision. This paper assumes that the realized piece of information is observable and contractible.

Zermeno (2011) and Papireddygar and Waggoner (2022) study menu design with information acquisition. In their setting, the principal first offers a menu of contracts. Then the agent privately acquire costly information. Next, the agent selects a contract from the menu. The selection therefore reveals some information the agent acquired.

In Carroll (2019), the principal is uncertain about the expert’s information acquisition technology and only knows some experiments that the agent can choose. The principal evaluates the incentive contract by a worse-case criterion.

---

<sup>7</sup>In Neyman et al. (2021), a rational expert aims to predict the probability of a biased coin flip. He acquires information by choosing the number of flip trials at a fixed cost per flip. Neyman et al. (2021) study the optimal scoring rule that incentivize precision.

Argenziano et al. (2016) and Kreutzkamp (2023) study costly information acquisition and transmission. In both papers, there is no transfer and the agent cares about the principal’s action. In Kreutzkamp (2023) setting, the sender publicly chooses an experiment. In Argenziano et al. (2016), the expert acquires information by choosing a number of binary trials to perform.

In the classic principal-agent model, several papers also show the optimality of the cutoff transfer but rely on different assumptions. Oyer (2000) assumes the validity of the first-order approach, the existence of the optimal contract, and the absence of the IR constraint. He shows that cutoff transfers are optimal among monotone contracts for a risk neutral agent. Herweg et al. (2010) show that cutoff transfers are optimal for expectation-based loss averse and risk neutral agent.

Notably, there is a rising literature adopting the same assumption that the principal cannot take money from the agent and can only reward the agent with a prize for which they have no other uses, see Li and Libgober (2023); Li et al. (2022); Deb et al. (2018, 2023); Dasgupta (2023); Hébert and Zhong (2022); Wong (2023).

## 2 The Model

I study a principal-agent model, where the principal (she) wants to incentivize the agent (he) to acquire information regarding an underlying state  $\theta \in \mathbb{R}$ . They share a common prior over  $\theta$ . It will be convenient to assume the prior to be an improper uniform prior on  $\mathbb{R}$ . I generalize my results to the case in which the state is multi-dimensional in Section 5.1, and I study the case of a proper Gaussian prior in Section 5.2.

The agent can acquire a costly signal  $s \in \mathbb{R}$  of the form

$$s = \theta + \frac{1}{\lambda}\varepsilon$$

where  $\varepsilon$  is drawn from a distribution with a symmetric and single-peaked probability density function (PDF)  $\phi$ . I assume  $\phi$  is continuously differentiable. The function  $\phi$  is supported on an interval, which can be bounded or unbounded. The parameter  $\lambda$  is a measure of the signal’s precision, a scale parameter that is inversely proportional to the standard deviation of  $s$ . The PDF of the signal, given  $\theta$  and  $\lambda$ , is denoted by

$\varphi(\cdot; \theta, \lambda)$ . It takes the form:

$$\varphi(x; \theta, \lambda) = \lambda \phi(\lambda(x - \theta)).$$

For example, if  $\phi$  is the PDF of the standard Gaussian distribution, then  $\varphi(\cdot; \theta, \lambda)$  is the PDF of a Gaussian distribution with mean  $\theta$  and standard deviation  $1/\lambda$ . Note that  $\phi(x) = \varphi(x; 0, 1)$ .

The agent chooses the precision  $\lambda$  at a cost  $c(\lambda)$ . For example, suppose the agent's signal is the aggregation of many small independent signals. Then as the number of small signals becomes large, the aggregate signal tends to a Gaussian distribution. If the agent incurs a cost that depends on the number of small signals he gathers, then the agent effectively controls the standard deviation of the aggregate signal at a cost. I assume that the cost function  $c$  is lower semicontinuous.<sup>8</sup> The pair  $(\phi, c)$  constitutes the primitive of the model.

The principal observes neither the agent's signal  $s$  nor the agent's choice of precision  $\lambda$ . Instead, after observing the signal, the agent sends a report  $a \in \mathbb{R}$  to the principal. Eventually both players observe the state  $\theta$ .<sup>9</sup> The principal wants to incentivize the agent to maximize precision and truthfully report his signal. She does so by means of a transfer. The transfer can depend on the state and the agent's report. The assumption that the realized state is contractible is familiar from the literature on belief elicitation via (proper) scoring rules and prediction markets, ubiquitous in the principal-expert literature, and well-suited to economic applications in which the state is publicly observable ex post (e.g., the outcome of an election being forecast by a pollster, or the conditions of a new market being analyzed by a consultant).

I assume that the transfer  $t$  is a function of the difference  $\theta - a$  and it vanishes at infinity. That is, it satisfies

$$\lim_{x \rightarrow -\infty} t(x) = \lim_{x \rightarrow \infty} t(x) = 0.$$

As I discuss below, the assumption that the transfer depends only on the difference between  $a - \theta$  is to ensure the agent's payoff is well-defined. I also assume limited

---

<sup>8</sup>It would be natural to also assume that the cost function is increasing—that is, the higher the precision, the higher the cost. Imposing this assumption or not does not affect my results.

<sup>9</sup>I relax this assumption in Section 5.3 where instead of eventually observing the state, they eventually observe a signal about the state.



liability, i.e., that the principal cannot take money from the agent, so that  $t \geq 0$ . In addition, the principal has a limited budget that does not depend on the state. She can use this budget only for the transfer. Without loss, I take the budget to be 1.

The timing of moves is as follows:

1. The principal commits to a transfer rule  $t: \mathbb{R} \rightarrow [0, 1]$ .
2. The agent chooses to accept or not. The game continues if the agent accepts and terminates otherwise.
3. The agent privately chooses a signal precision  $\lambda$ .
4. The agent privately observes a signal realization  $s$  drawn from the distribution  $\varphi(\cdot; \theta, \lambda)$ .
5. The agent makes a report  $a \in \mathbb{R}$ .
6. Finally, the state  $\theta$  is revealed and the agent receives the transfer  $t(\theta - a)$ .

The agent's payoff is given by the expected utility of the transfer minus the cost of acquiring information:

$$\mathbb{E}u(t) - c(\lambda)$$

where  $u: [0, 1] \rightarrow \mathbb{R}$  is strictly increasing and continuous, and normalized so that  $u(0) = 0$ , and  $u(1) = 1$ .<sup>10</sup> To ease the notation, it will be convenient to treat the transfer rule as paying in utils, rather than money. Under my assumptions on  $u$ , this is equivalent, and without loss of generality, to assuming  $u$  is the identity function. Finally, I assume the agent has an outside option which gives payoff 0 if he rejects the contract.

After choosing  $\lambda$ , the agent observes a signal. Conditional on a signal realization  $s$ , it follows from Bayes' rule (adapted to a uniform prior) that the agent's posterior has PDF  $\varphi(\cdot; s, \lambda)$ . The agent now chooses a report  $a$  to maximize the expected transfer, which is therefore given by:

$$E(\lambda; t) = \max_{a \in \mathbb{R}} \int_{\mathbb{R}} t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta'.$$

---

<sup>10</sup>Here I can drop the continuity assumption on  $u$  and only require  $u$  to be weakly increasing. All my results still hold.

Maximizing over report  $a$  shows that when choosing the precision, the agent anticipates how he optimally reports later. Thus, the model allows for double deviation. The assumption that  $t$  vanishes at infinity ensures that the maximum is well defined.<sup>11</sup>

Note that as  $t$  depends only on the difference  $\theta - a$ , the expected transfer computed above does not depend on  $s$ .<sup>12</sup> Consequently, the agent's (interim) expected transfer conditional on signal  $s$  does not depend on  $s$ , which must coincide with his unconditional expected transfer. If  $t$  depends arbitrarily on  $a$  and  $\theta$ , the agent's unconditional expected transfer might not be well-defined, since the distribution of  $s$  is improper. In Appendix A, I show that when the agent's expected transfer is well-defined, it is without loss to assume that the transfer depends only on the difference  $a - \theta$ . An interesting fact is that almost all popular key performance indicators in the forecast industry depend only on the prediction error  $a - \theta$  (Vandeput, 2021).

The principal's objective is to maximize  $\lambda$  under the constraint of inducing a truthful report from the agent. (As I show later, this constraint will not be binding.) This objective can be interpreted as capturing, in reduced-form, settings in which the principal uses the information conveyed by the agent's report to solve some (unmodeled) decision problem. Indeed, under broad conditions, it is optimal for the principal to maximize precision regardless of the decision problem that she faces. First, when the density  $\phi$  is strongly unimodal, for a large class of monotone decision problems (Karlin and Rubin, 1956), higher precision is always better for the principal (see Lehmann, 2011, Theorem 5.1 and 5.2).<sup>13</sup><sup>14</sup> Second, when the distribution  $\phi$  is *self-decomposable*, the signal with different precisions are ranked in the Blackwell order; thus the principal is better off with higher precision for all decision problems. The self-decomposable distributions include all stable distributions (such as Gaussian and Cauchy distributions) and some non-stable ones such as Laplace.<sup>15</sup>

---

<sup>11</sup>I prove this in Lemma 5 in the appendix.

<sup>12</sup>To see this, note that

$$\max_a \int_{\mathbb{R}} t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' = \max_a \int_{\mathbb{R}} t(\theta' - a + s) \varphi(\theta'; 0, \lambda) d\theta' = \max_a \int_{\mathbb{R}} t(\theta' - a) \varphi(\theta'; 0, \lambda) d\theta'$$

<sup>13</sup> $\phi$  is strongly unimodal if  $-\ln \phi$  is convex. This is slightly stronger than global increasing elasticity.

<sup>14</sup>The class of *monotone* decision problems is defined in terms of the action space and the permissible loss functions. For each  $\theta$ , there is a correct action  $A(\theta)$ . The function  $A(\theta)$  is real-valued and nondecreasing. The range of  $A(\theta)$  is the action space. The loss function is minimized at the correct action and is nondecreasing as the action moves away from the correct action on either side.

<sup>15</sup>See Goel and DeGroot (1979) and Lehmann (2011) for a detailed discussion.

### 3 Preliminary Analysis

First, I define a relaxed problem, in which the principal designs a transfer to maximize the agent's signal precision, without incentivizing truthful reporting.

$$\begin{aligned}
& \underset{\lambda, t}{\text{maximize}} && \lambda \\
& \text{subject to} && 0 \leq t \leq 1 \\
& && \lambda \in \arg \max E(\cdot; t) - c(\cdot) \quad \text{IC} \\
& && E(\lambda; t) - c(\lambda) \geq 0 \quad \text{IR}
\end{aligned} \tag{1}$$

The first constraint in the relaxed problem is limited liability and limited budget. The second constraint is an incentive compatibility constraint, since the precision  $\lambda$  is chosen by the agent and is unobserved by the principal and cannot be contracted upon.<sup>16</sup> The last constraint is a participation constraint, also known as individual rationality, and is implied by the outside option available to the agent. A transfer rule  $t$  is *optimal* if it solves problem (1) and induces the agent to report truthfully.

A quantity that will be key to the analysis is the elasticity of the standardized signal distribution  $\phi$ .

**Definition 1.** The *elasticity*  $\eta$  of signal distribution  $\phi$  at  $x > 0$  is defined as

$$\eta(x) = -\frac{d\phi(x)/\phi(x)}{dx/x}.$$

For  $x$  such that  $\phi(x) = 0$ , I let  $\eta(x) = +\infty$ .

Elasticity measures by how much a percentage change in  $x$  leads to a percentage change in the density function  $\phi(x)$ . For example, if  $\phi$  is the PDF of a standard Gaussian, then  $\eta(x) = x^2$ . If  $\phi$  is the PDF of a standard Laplace, then  $\eta(x) = x$ . Notice that both examples have weakly increasing elasticity. In fact, weakly increasing elasticity is satisfied by most common nonatomic distributions defined on an interval, including the uniform distribution, the triangular distribution, and the logistic distribution. The next lemma shows that, weakly increasing elasticity condition is equivalent to a monotone likelihood property: For any  $0 \leq x_1 \leq x_2$ , the

---

<sup>16</sup>The continuity of  $E(\cdot; t)$  and the lower semicontinuity of  $c$  ensures that the  $\arg \max$  in IC is well defined.

ratio  $\frac{\varphi(x_1;0,\lambda)}{\varphi(x_2;0,\lambda)}$  increases in  $\lambda$ . That is, a more accurate signal ( $x_1$ ) is more likely to appear than a less accurate signal ( $x_2$ ) as precision increases.

**Lemma 1.**

$$\frac{\partial[\ln \phi(\lambda x)]}{\partial \ln \lambda} = -\eta(\lambda x).$$

In particular, for any  $0 \leq x_1 \leq x_2$ ,

$$\frac{\varphi(x_1;0,\lambda)}{\varphi(x_2;0,\lambda)} \text{ increases in } \lambda \quad \Leftrightarrow \quad \eta(\lambda x_1) \leq \eta(\lambda x_2).$$

Next, I define a weaker condition.

**Definition 2.** The signal distribution satisfies *increasing elasticity above 1* if  $\eta$  single-crosses 1 from below and is weakly increasing after the cross, i.e.; if for every  $x > 0$

$$\eta(x) > 1 \quad \text{implies} \quad \eta(y) \geq \eta(x) \text{ for all } y > x.$$

I define a useful quantity.

$$\eta^{-1}(1) = \inf\{x \in \mathbb{R}_+ | \eta(x) > 1\}.$$

Note that  $\eta^{-1}(1)$  is well-defined, as the density function  $\phi$  is integrable.<sup>17</sup> Later on, when I generalize the state and the signal to be  $n$  dimensional,  $\eta^{-1}(n)$  is defined similarly.

## 4 Characterization of the Optimal Transfer

A simple transfer rule is the *cutoff transfer* that pays the agent 1 when the distance between the report and the state is less than a cutoff  $d$  and pays 0 otherwise (see Figure 1). As cutoff transfers are symmetric and single-peaked, they have the additional desirable property of inducing truthful reports by the agent.

As transfer rules are functions, the principal faces an infinite dimensional optimization problem. In my main result I show that when the standardized distribution  $\phi$  satisfies increasing elasticity above one—which holds for all commonly used signal

---

<sup>17</sup>Notice that function  $1/x$  is not integrable over any neighborhood around 0. Thus,  $\eta(0^+) < 1$  since  $\phi$  is integrable over any neighborhood around 0. Moreover, function  $1/x$  is not integrable over any neighborhood around  $+\infty$ . Thus,  $\eta(x) > 1$  must hold for some  $x \in \mathbb{R}_+$ .

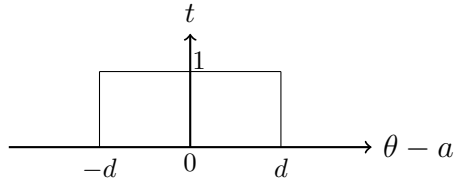


Figure 1: Cutoff Transfer

distributions—the principal chooses a cutoff transfer rule, reducing the problem to a one-dimensional one. Moreover, I show that this property of  $\phi$  is also necessary for cutoff transfer rules to always be optimal.

**Theorem 1.** *The following are equivalent.*

1. *For all cost functions, there exists an optimal transfer that is a cutoff transfer.*
2. *For all increasing and differentiable cost functions, there exists an optimal transfer that is a cutoff transfer.*
3. *Signal density function  $\phi$  satisfies increasing elasticity above 1.*

Note that Theorem 1 holds even if we assume that all cost functions are differentiable and increasing; that is, the result is not driven by considering ill behaved cost functions.

Once we verify the condition in statement (3), then we do not need to worry about the infinite dimensional optimization problem, as cutoff transfers shall be optimal. Moreover, the condition is very easy to check. We only need to compute the elasticity and check if it is monotone in some region. A direct implication of Theorem 1 is that for all cost functions, there exists an optimal transfer that is a cutoff transfer when the signal distribution is Gaussian, since the elasticity of a Gaussian distribution is increasing. The lower bound 1 in the statement is the dimension of the problem. So far both the state and the report are one-dimensional. In Section 5.1, where I generalize this result to  $n$ -dimensional signal and states, the dimension  $n$  will replace 1.

Note that there is a unique solution—which must be a cutoff transfer—if we additionally impose three mild assumptions.<sup>18</sup>

---

<sup>18</sup>I prove this uniqueness result at the end of the Proof of Theorem 1.

1. Signal density function  $\phi$  satisfies strictly increasing elasticity above<sup>19</sup> 1.
2. The cost function is continuously differentiable.
3. The optimal precision is an interior solution.

In the following, I shall first illustrate the sufficiency part (i.e., why (3) implies (1)). Second, I optimize over cutoff transfers and characterize the optimal cutoff. Third, I provide intuition for the necessity part (i.e., why (2) implies (3)).

I define the agent's expected payoff given transfer  $t$  as

$$\pi(t) = \max_{\lambda} E(\lambda; t) - c(\lambda)$$

and the agent's choice of precision given  $t$

$$\lambda(t) = \begin{cases} \max[\arg \max_{\lambda} E(\lambda; t) - c(\lambda)] & \text{if } \pi(t) \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

When analyzing cutoff transfers, with slight abuse of notations, I use  $E(\lambda; d)$  to represent the expected transfer at precision  $\lambda$  given the cutoff transfer with cutoff  $d$ , i.e.,  $E(\lambda; d) = E(\lambda; \mathbf{1}_{|\theta-a| \leq d})$ . Similarly, I denote by  $\pi(d)$  and  $\lambda(d)$  the expected payoff and choice of precision for the cutoff transfer  $d$ , respectively.

## 4.1 Cutoff Transfers Are Optimal

I explain why increasing elasticity above 1 implies that for all cost functions, cutoff transfers are optimal. Suppose that increasing elasticity above 1 holds. I show that for any transfer rule  $t$ , there exists a cutoff transfer that induces a weakly larger precision and truthful report.

Here, for ease of exposition, I assume that  $\eta$  is weakly increasing and consider a transfer rule  $t$  that is symmetric in  $\theta - a$  and weakly decreases in  $|\theta - a|$ . Given such a transfer rule, the agent reports truthfully and chooses the precision  $\lambda(t)$ . I construct a cutoff transfer by choosing a cutoff  $d$  such that

$$E(\lambda(t); d) = E(\lambda(t); t). \tag{2}$$

---

<sup>19</sup>Density function  $\phi$  satisfies *strictly increasing elasticity above 1* if  $\eta(\cdot)$  single-crosses 1 from below,  $\{x | \eta(x) = 1\}$  is a singleton, and is strictly increasing after the cross.

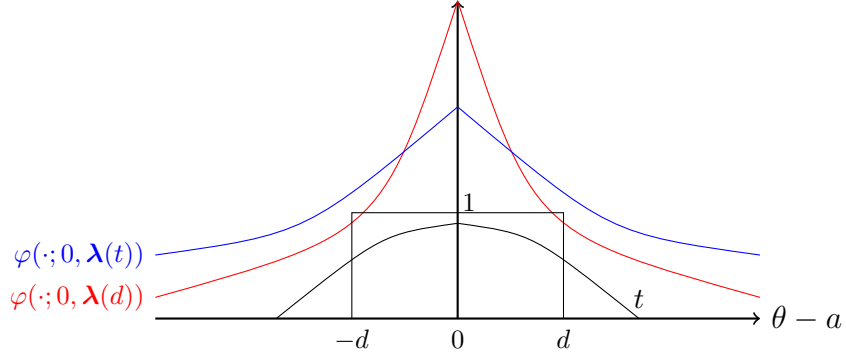


Figure 2: The Transfer  $t$  and The Cutoff Transfer  $d$ . The given transfer  $t$  and the matching cutoff transfer  $d$  are shown in black. The wider distribution shown in blue is the signal distribution chosen by the agent under  $t$ , while the narrower (more precise) distribution in red is the one chosen under  $d$ .

Such a  $d$  exists because  $E(\lambda(t); d)$  increases in  $d$  from 0 to 1 (as the budget is 1). See Figure 2 for an example.

Next, I argue that

$$\begin{aligned} E(\lambda; d) - E(\lambda; t) &\geq 0 \text{ for } \lambda > \lambda(t) \\ E(\lambda; d) - E(\lambda; t) &\leq 0 \text{ for } \lambda < \lambda(t). \end{aligned} \tag{3}$$

This is depicted in Figure 3.

To see this, note that as the cutoff transfer pays the entire budget 1 when  $|\theta - a| \leq d$ , it is larger than the transfer  $t$  within the cutoff region. In addition, Lemma 1 shows that, since the elasticity of  $\phi$  is weakly increasing, for any  $0 < x_1 < d < x_2$  the ratio

$$\frac{\varphi(x_1; 0, \lambda)}{\varphi(x_2; 0, \lambda)}$$

increases in  $\lambda$ . Since  $E(\lambda; t)$  is the integral of the transfer with respect to the signal distribution, (3) follows. As a result, we have  $\lambda(d) \geq \lambda(t)$  (see Figure 3).

In the argument above, I use a stronger condition, global increasing elasticity, to show that cutoff transfers are optimal. By Lemma 1, global increasing elasticity is equivalent to the monotone ratio likelihood property:

$$\text{For all } 0 < x_1 < x_2, \quad \frac{\varphi(x_1; 0, \lambda)}{\varphi(x_2; 0, \lambda)} \text{ is increasing in } \lambda.$$

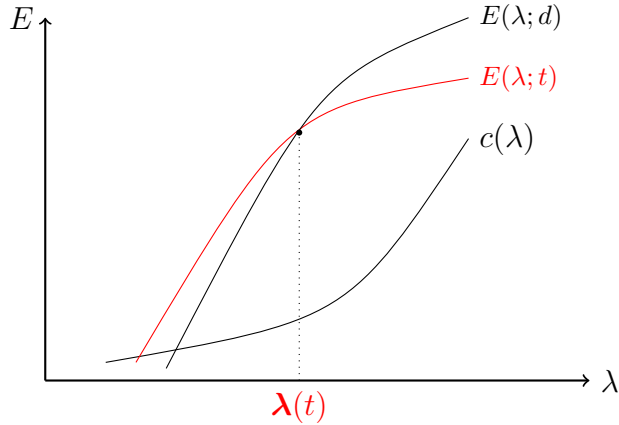


Figure 3: Expected Transfer

Thus, this property is a sufficient condition for cutoff transfers to be optimal.

**Corollary 1.** *If for all  $x_2 > x_1 > 0$ ,  $\frac{\varphi(x_1; 0, \lambda)}{\varphi(x_2; 0, \lambda)}$  increases in  $\lambda$ , there exists an optimal transfer that is a cutoff transfer.*

To complete the proof of sufficiency in the general setting, there are two more technical issues. First, for a general transfer  $t$  (which is not necessarily symmetric), the agent's report depends on both the signal and the precision. But in our previous example (Figure 3), the agent's report equals the signal and does not depend on the precision. Second, the proof sketch above assumed globally increasing elasticity, which is stronger than increasing elasticity above 1. Under this weaker assumption, we do not have for all  $x_2 > x_1 > 0$ ,  $\frac{\varphi(x_1; 0, \lambda)}{\varphi(x_2; 0, \lambda)}$  increases in  $\lambda$ . The proof in the appendix deal with both issues.

## 4.2 Optimization of Cutoff Transfers

When  $\phi$  satisfies increasing elasticity above 1, Theorem 1 reduces the infinite dimensional optimization problem to a one-dimensional problem of finding the value of the optimal cutoff. In this section, I will solve this one-dimensional problem and identify the optimal cutoff.

As the cutoff  $d$  increases, the agent's expected transfer  $E(\cdot; d)$  increases. Let  $\bar{d}$  denote the minimum cutoff such that IR constraint holds, i.e.,

$$\bar{d} = \min\{d \geq 0 \mid \pi(d) \geq 0\}.$$



Notice that  $\bar{d}$  depends on the cost function  $c$ . The continuity of  $E(\cdot; d)$  for all  $d$  and the lower semicontinuity of  $c$  ensures that this minimum is well defined.

**Theorem 2.** *If  $\phi$  satisfies increasing elasticity above 1, the cutoff transfer*

$$d^* = \min\{d \geq \bar{d} \mid \lambda(d)d \geq \eta^{-1}(1)\}$$

*is optimal.*

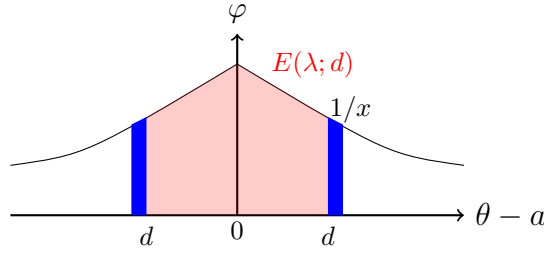


Figure 4: Increment of Expected Transfer When Increasing  $\lambda$

The agent slightly increases precision from  $\lambda$  to  $\lambda + \Delta\lambda$ . The area of the red region is the expected transfer  $E(\lambda; d)$ . The area of two blue regions is the increment of probability that the signal falls into the cutoff.

To obtain this result, the following lemma is crucial.

**Lemma 2** (Complements or Substitutes). *The expected transfer  $E(\lambda; d)$  satisfies:*

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \geq 0 \quad \Leftrightarrow \quad \eta(\lambda d) \leq 1.$$

This lemma characterizes whether the precision and the cutoff are complements or substitutes. When the cross derivative  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d}$  is positive, increasing the cutoff would increase the marginal return to the precision; thus they are complements. On the other hand, when the cross derivative is negative, the cutoff and the precision are substitutes.

I illustrate the intuition of the boundary case when  $\eta(\lambda d) = 1$  and show that

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} = 0.$$

I plot the agent's the expected transfer in Figure 4. Since the agent receives transfer 1 when the signal lies in the cutoff, the expected transfer  $E(\lambda; d)$  is the probability

that the signal lies in the cutoff, which is the area of the light red region. Suppose that the agent slightly increases precision from  $\lambda$  to  $\lambda + \Delta\lambda$ . The signal shall lie in the cutoff with a higher probability. The area of dark blue region is the incremental probability that the signal lies in the cutoff. Since  $\eta(\lambda d) = 1$ , in a neighborhood of  $d$ ,  $\varphi(\cdot; 0, \lambda)$  behaves like  $x^{-1}$ . We can compute the incremental expected transfer, which coincides with the incremental probability, by the area of the dark blue region

$$E(\lambda + \Delta\lambda; d) - E(\lambda; d) \approx 2\varphi(d; 0, \lambda) \left( d - d \frac{\lambda}{\lambda + \Delta\lambda} \right)$$

which is independent of  $d$  as  $\varphi(\cdot; 0, \lambda)$  decays at  $x^{-1}$ . As a result,  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} = 0$ . A similar argument shows that  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d}$  and  $\eta(\lambda d) - 1$  have the opposite sign. Here 1 appears as this problem is one dimensional, which is also why “increasing elasticity above 1” appears in Theorem 1 and  $\eta^{-1}(1)$  appears in Theorem 2. Later on when I generalize this problem to  $n$ -dimensional,  $n$  shows up in both characterizations.

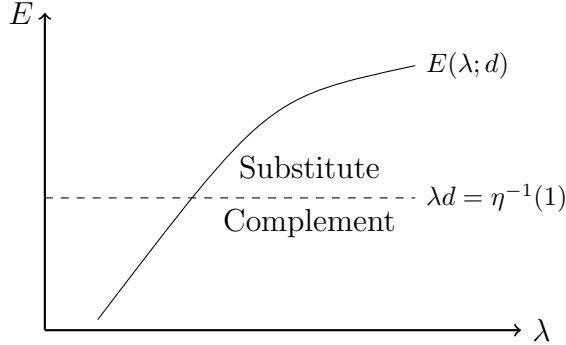


Figure 5: Expected Transfer

As the expected transfer  $E(\lambda; d)$  is the probability that a signal falls within the cutoff region, we have  $E(\lambda; d) = 2\Phi(\lambda d) - 1$  where  $\Phi$  is the CDF of the distribution  $\phi$ . Consequently,  $E(\lambda; d)$  is increasing in  $\lambda d$ . Since  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d}$  and  $\eta(\lambda d) - 1$  have the opposite sign, by the definition of  $\eta^{-1}(1)$ , we have

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \geq 0 \quad \text{if} \quad \lambda d \leq \eta^{-1}(1)$$

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0 \quad \text{if} \quad \lambda d \geq \eta^{-1}(1).$$

Thus, the graph of expected transfer is separated into two parts (see Figure 5). When

$\lambda d \geq \eta^{-1}(1)$ ,  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0$ , I refer to this part as the *substitute region*. When  $\lambda d < \eta^{-1}(1)$ ,  $\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} > 0$ , I refer to this part as the *complement region*.

Recall that  $\bar{d}$  denotes the minimum cutoff that the agent is willing to work. If  $(\bar{d}, \lambda(\bar{d}))$  lies in the substitute region, increasing  $d$  shall reduce  $\lambda(d)$ , which is not optimal for the principal. Consequently, she should set the cutoff at the minimum  $\bar{d}$ . In this case, the IR constraint is binding and the agent's surplus is zero. If  $(\bar{d}, \lambda(\bar{d}))$  lies in the complement region, increasing  $d$  shall first increase  $\lambda(d)$ , and  $\lambda(d)d$  shall increase until it hits the boundary  $\eta^{-1}(1)$ . After hitting the boundary, if we further increase  $d$ , then  $\lambda(d)$  shall decrease as  $(d, \lambda(d))$  enters the substitute region. Thus, the optimal cutoff is the  $d$  where  $\lambda(d)d$  first hits the boundary  $\eta^{-1}(1)$ . In this case, the IR constraint is relaxed and the agent enjoys some surplus.

Note that although  $\lambda^*$  is unique by design, the optimal cutoff transfer is not necessarily unique. To see this, recall that changing  $d$  causes the curve  $E(\cdot; d)$  to rotate. If the cost function  $c$  has a kink and  $\lambda(d)$  is stuck at this kink, rotating  $E(\cdot; d)$  slightly might not affect  $\lambda(d)$ , which gives rise to multiple optimal cutoff transfers. The cutoff  $d^*$  in Theorem 2 is the optimal cutoff transfer that provides the strongest local incentive around  $\lambda^*$ , i.e., it has the largest derivative  $\frac{\partial E(\lambda; d)}{\partial \lambda} \big|_{\lambda=\lambda^*}$  among all optimal cutoff transfers. But this example is not generic due to the kink in the cost function. This is also why a continuous differentiable cost function helps us to get a unique optimal transfer rule.

### 4.3 The Necessity of Increasing Elasticity above 1

In this section, I explain the intuition behind the necessity result—(2) implies (3)—in Theorem 1. That is, why increasing elasticity above 1 is necessary for a cutoff transfer to always be optimal.

I first provide an example where  $\phi$  does not satisfy increasing elasticity above 1. In this case, for some increasing and differentiable cost function, all cutoff transfers are suboptimal.

**Example 1.** The standardized signal distribution is truncated  $\exp(1/x)$ :

$$\phi(x) = \begin{cases} k \exp(1/\epsilon), & \text{if } x \in [0, \epsilon), \\ k \exp(1/x), & \text{if } x \in [\epsilon, 1], \\ 0, & \text{otherwise.} \end{cases}$$

where  $k$  is a normalizing factor (see the left panel of Figure 6). I plot the elasticity of  $\phi$  on the right panel of Figure 6. The signal distribution does not satisfy increasing elasticity above 1 as  $\eta(x)$  is decreasing for  $x \in [\epsilon, 1]$ . Fix a pair  $(\lambda^*, d^*)$  such that  $\epsilon < \lambda^* d^* < 1$ . Suppose an increasing cost function  $c \in C^1$  is tangent to  $E(\cdot; d^*)$  at  $\lambda = \lambda^*$  and  $c(\lambda)$  is strictly above  $E(\lambda; d^*)$  for all  $\lambda \neq \lambda^*$  (see the right panel of Figure 7).

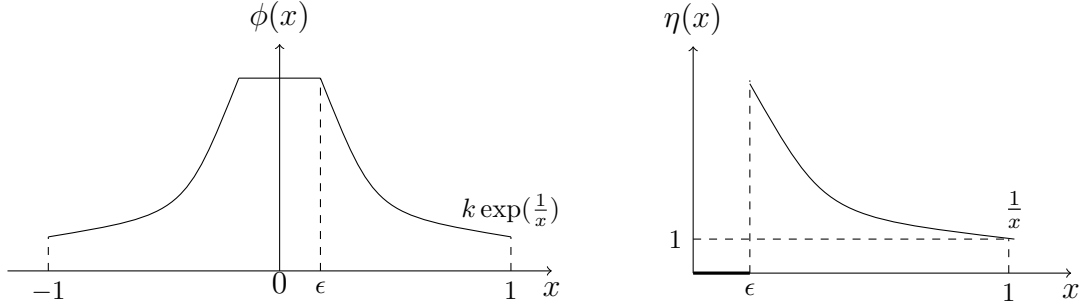


Figure 6: Signal Distribution and Elasticity

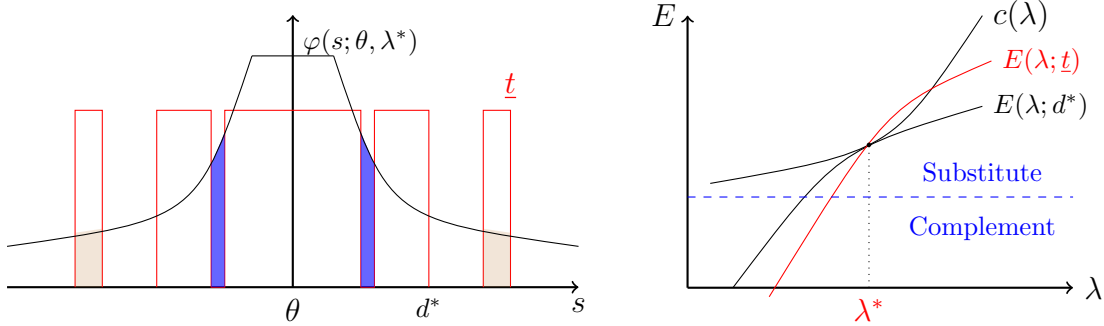


Figure 7: The New Transfer Rule and The Expected Transfer

First, note that  $d^*$  is the best cutoff among all cutoff transfers: By design,  $d^*$  is the minimum cutoff for the agent to work. By Lemma 2, as  $(\lambda^*, d^*)$  lies in the substitute region, any cutoff transfer with cutoff  $d > d^*$  induces a smaller precision. Thus,  $d^*$  is the best cutoff transfer.

Second, I construct a new transfer rule  $\underline{t}$  which induces a strictly larger precision. Starting with the best cutoff transfer rule  $d^*$ , I modify the transfer rule by setting one interior region (dark blue) within the cutoff to 0 and one exterior region (light brown) outside the cutoff to 1 (see the left panel of Figure 7). These two regions are chosen such that the area of the dark blue region and the light brown region are the

same. Note that the area of each region is the expected transfer contributed by this region at precision  $\lambda^*$ . Thus, the new transfer rule  $\underline{t}$  has the same expected transfer as cutoff transfer  $d^*$  at precision  $\lambda^*$ .

As the signal distribution features decreasing elasticity, the density function  $\varphi(s; \theta, \lambda^*)$  is steeper at the blue region. Once we increase precision, the blue area shall shrink more than the brown area. This implies that the new transfer rule  $\underline{t}$  has higher expected transfer than the cutoff transfer  $d^*$  for slightly higher precision. I translate this comparison to the right panel of Figure 7. For higher precision  $\lambda > \lambda^*$ ,  $E(\lambda; \underline{t}) > E(\lambda; d^*)$ . Thus,  $\underline{t}$  shall induces a higher precision.

This example is key to understand more generally the necessity of increasing elasticity above 1. I call a pair  $(\lambda^*, d^*)$  *exposed* if the cutoff transfer  $d^*$  is the optimal transfer for some increasing cost function  $c \in C^1$  and induces precision  $\lambda^*$ . A pair  $(\lambda^*, d^*)$  being exposed implies that  $\eta(x_1) \leq \eta(x_2)$  for all  $0 \leq x_1 \leq \lambda^* d^* \leq x_2$ . Otherwise, I can use the construction in Example 1 to construct a new transfer than induces a strictly larger precision. In Theorem 2, all  $(d^*, \lambda^*)$  pairs in the substitute region are exposed, as I can always find an increasing and continuously differentiable cost function that is tangent to  $E(\cdot; d^*)$  at  $\lambda^*$  and strictly above  $E(\cdot; d^*)$  everywhere else. Similarly, when for all increasing and differentiable cost functions there exists an optimal transfer that is a cutoff transfer, all pairs with  $\eta(d^* \lambda^*) \geq 1$  are exposed. This implies increasing elasticity above 1. However, as the construction in the example above only works in the substitute region, pairs in the complement regions are not exposed. Consequently, we do not have monotonicity for  $\eta$  when it is below 1.

## 4.4 Comparative Statics

Next I study how the optimal cutoff changes with the cost function. Without loss of generality, I assume that the cost function is weakly increasing.<sup>20</sup> I express dependence on the cost function. Let  $d^*(c)$  and  $\lambda^*(c)$  denote the optimal cutoff and precision given cost function  $c$ . Consider two cost functions  $c_1 \leq c_2$  where  $c_1$  is less costly. Intuitively, if the agent's cost is higher, the optimal cutoff increases. The next result formalizes this intuition.

---

<sup>20</sup>Given any cost function  $\tilde{c}$  that is not weakly increasing, I can define a new cost function  $c$  to be the largest weakly increasing function that is below  $\tilde{c}$ . The agent's precision choice problem under cost function  $c$  is the same as under cost  $\tilde{c}$ , because  $E(\lambda; d)$  is weakly increasing in  $\lambda$ .

**Proposition 1** (Comparative Statics). *Suppose that  $\phi$  satisfies increasing elasticity above 1 and  $c_1 \leq c_2$ . If  $c_2(\lambda) - c_1(\lambda)$  is weakly increasing in  $\lambda$ , then  $d^*(c_1) \leq d^*(c_2)$  and  $\lambda^*(c_2) \leq \lambda^*(c_1)$ . In particular, if  $c_2 = kc_1$  for some constant  $k > 1$ , then  $d^*(c_1) \leq d^*(c_2)$  and  $\lambda^*(c_2) \leq \lambda^*(c_1)$ .*

The cost difference  $c_2(\lambda) - c_1(\lambda)$  being weakly increasing occurs if  $c_2$  is more convex than  $c_1$ . Another interesting comparative statics is to change the budget from 1 to  $1/k$ . This is equivalent to keeping the budget at 1 and change the cost function from  $c_1$  to  $kc_1$ . Thus, lowering the budget would lead to a larger optimal cutoff.

As a direct corollary, I can study what happens if the agent's signal is more precise. Suppose that  $\varepsilon_2 = k\varepsilon_1$  with some constant  $k > 1$ . Recall that the agent's signal is  $s = \theta + \frac{1}{\lambda}\varepsilon$ . Thus, a smaller noise  $\varepsilon_1$  corresponds to a more precise signal, which leads to a smaller cutoff by the next result.

**Corollary 2.** *Suppose that the cost function is weakly convex. If  $\varepsilon_2 = k\varepsilon_1$  with some constant  $k > 1$ ,  $d_1^* \leq d_2^*$ .*

What if the cost difference  $c_2 - c_1$  is not increasing? Then the answer is more involved. Denote by  $\lambda(d; c)$  the induced precision given cutoff transfer  $d$  and the cost function  $c$ . Given the cost function  $c_1$ , the minimum cutoff  $\bar{d}(c_1)$  and the induced precision  $\lambda(\bar{d}(c_1); c_1)$  pair lies either in the substitute region  $\lambda(\bar{d}(c_1); c_1)\bar{d}(c_1) \geq \eta^{-1}(1)$  or not. If this pair lies in the substitute region, then the optimal cutoff coincides with the minimum cutoff,  $d^*(c_1) = \bar{d}(c_1)$ , by Theorem 2. Then given a larger cost function  $c_2$ , the minimum cutoff must be larger,  $\bar{d}(c_2) \geq \bar{d}(c_1)$ , which entails a larger optimal cutoff,  $d^*(c_2) \geq d^*(c_1)$ .

However, if the  $\lambda(\bar{d}(c_1); c_1), \bar{d}(c_1)$  pair lies in the complement region, the comparison of  $d^*$  can go both directions. If  $c_2 - c_1$  is decreasing, it may render higher precision relatively more attractive to the agent under cost  $c_2$  than  $c_1$ . This may induce a higher precision for each cutoff transfer. As the optimal precision and optimal cutoff are substitutes by Theorem 2, a higher induced precision leads to a lower optimal cutoff.

Readers may wonder what happens to the optimal precision  $\lambda^*$  given  $c_2 \geq c_1$ . It turns out that increasing the cost function can shift  $\lambda^*$  in both directions, depending on the shape of the cost difference  $c_2 - c_1$ .

## 5 Extensions

### 5.1 The Multi-Dimensional Case

My results can generalize to the case with a multi-dimensional state. In this section, suppose the state, the signal, and the report are  $n$ -dimensional, i.e.,  $\theta, s, a \in \mathbb{R}^n$ . The signal distribution is still symmetric, single-peaked, and centered around  $\theta$ . Here symmetry means that the density of the distribution depends only on the Euclidean distance from the state  $\theta$ . A cutoff transfer pays 1 when the Euclidean distance between the report and the state is less than a cutoff  $d$ . The next proposition shows that a statement analogous to that of Theorem 1 applies in this case too.

**Proposition 2.** *Suppose the state is  $n$ -dimensional. The following statements are equivalent.*

1. *For all cost functions, there exists an optimal transfer that is a cutoff transfer.*
2. *For all increasing and differentiable cost functions, there exists an optimal transfer that is a cutoff transfer.*
3. *Density function  $\phi$  satisfies increasing elasticity above  $n$ .*

Moreover, if  $\phi$  satisfies increasing elasticity above  $n$ , the cutoff transfer

$$d^* = \min\{d \geq \bar{d} \mid \lambda(d)d \geq \eta^{-1}(n)\}$$

is optimal.

Note that the increasing elasticity condition of Theorem 1 is generalized in this proposition to increasing elasticity above the dimension  $n$ , providing an explanation for why the number 1 appeared in Theorem 1.

### 5.2 Gaussian Prior and Gaussian Signal

In this section, I assume that both the prior and the signal admit Gaussian distributions. Since the prior is no longer uniform, a slight adjustment is that now the principal wants the agent to truthfully report the posterior mean rather than the signal.

Let the prior distribution be  $\mathcal{N}(0, 1/\lambda_0^2)$ . Conditional on a signal  $s$ , the agent's posterior is also Gaussian

$$\mathcal{N}\left(\frac{s\lambda^2}{\lambda_0^2 + \lambda^2}, \frac{1}{\lambda_0^2 + \lambda^2}\right)$$

where  $\lambda$  is the agent's choice of precision. I let  $\Lambda$  denote the precision of the posterior

$$\Lambda(\lambda) = \sqrt{\lambda_0^2 + \lambda^2}.$$

Then, the expected transfer becomes  $E(\Lambda(\lambda); t)$  instead of  $E(\lambda; t)$ . Therefore,  $\Lambda(\lambda)$  shall replace  $\lambda$  in the characterization. Other than this difference, the results remain the same as in previous sections, as the posterior is Gaussian whose elasticity is increasing.

**Proposition 3.** *Suppose the prior and the signal admit Gaussian distribution. There exists an optimal transfer that is a cutoff transfer. Moreover, the cutoff transfer*

$$d^* = \min\{d \geq \bar{d} | \Lambda(\lambda(d))d \geq \eta^{-1}(1)\}$$

*is optimal.*

### 5.3 Unobserved State

In this section, I extend my results to the setting where the state is unobservable. Instead, the principal has some private information about the state. When designing the transfer, she uses her private information instead of the state to discipline the agent.<sup>21</sup>

I assume that the prior is uniform or Gaussian  $\mathcal{N}(0, 1/\lambda_0^2)$ . The principal wants the agent to truthfully report his posterior mean and maximizes  $\lambda$ . The principal and the agent receive independent Gaussian signals:  $s_p \sim \mathcal{N}(\theta, 1/\lambda_p^2)$  and  $s \sim \mathcal{N}(\theta, 1/\lambda^2)$ . I show that it is without loss to focus on cutoff transfers that pay 1 when  $|s_p - a|$  is less than a cutoff and 0 otherwise.

**Proposition 4.** *Suppose that the state is unobservable and the prior admits uniform or Gaussian distribution. The principal and the agent receive independent Gaussian*

---

<sup>21</sup>Note that the principal does not have utility for their budget, and so have no incentive to lie about their private signals. Thus, it is without loss of generality to assume that the principal has commitment power.



signals. There exists an optimal transfer that is a cutoff transfer. Moreover, the cutoff transfer

$$d^* = \min\{d \geq \bar{d} | \Lambda(d)d \geq \eta^{-1}(1)\}$$

is optimal where  $\Lambda(d) = (\frac{1}{\lambda_p^2} + \frac{1}{\lambda^2(d)})^{-\frac{1}{2}}$  for the uniform prior and  $\Lambda(d) = (\frac{1}{\lambda_p^2} + \frac{1}{\lambda^2(d) + \lambda_0^2})^{-\frac{1}{2}}$  for the Gaussian prior.

## 6 Implications for Classic Principal-Agent Problem

In this section, I demonstrate how my results apply to a classic principal-agent problem, offering new insights into the optimality of simple contracts. In the classic setting, a principal incentivizes the agent to produce an output. Unlike the traditional framework, in which the principal's goal is to maximize expected payoff (of outputs) minus transfers, I assume the principal's sole objective is to maximize output, constrained by a budget limit. The budget with a specific usage is a common practice in firms. [Anthony et al. \(2007\)](#) and [Horngren \(2009\)](#) provide comprehensive empirical evidence for this approach. In firms, managers (the principals in my model) control the resources without being the ultimate owners, giving rise to agency costs, known as residual loss in finance ([Jensen and Meckling, 1976](#)). To combat this loss, a budget constraint with a specific usage is commonly deployed.

Suppose that an agent chooses an effort level  $e \geq 0$ , which can be discrete or continuous. A continuous output  $y \geq 0$  is random and distributed according to a probability density function  $g(y; e)$ . A principal's objective is to maximize the agent's output, subject to a budget constraint equals to 1. The transfer  $t(y)$  depends on the output  $y$ , such that  $0 \leq t(y) \leq 1$ . A cutoff transfer with parameter  $d$  is of the form  $t(y) = 1$  if  $y \geq d$  and  $t(y) = 0$  otherwise. The agent's cost of effort is given by a lower semi-continuous function  $c(e)$  and he maximizes the expected transfer minus the cost.

**Corollary 3.** *Suppose that  $g$  satisfies the monotone likelihood ratio condition: For all  $y_1 \leq y_2$ ,  $\frac{g(y_2; e)}{g(y_1; e)}$  is weakly increasing in  $e$ . Then, there exists an optimal transfer that is a cutoff transfer.<sup>22</sup>*

---

<sup>22</sup>[Banks and Sundaram \(1998\)](#) study how a long-lived principal interact with a series of short-

This is an immediate corollary of Corollary 1. My model can be “projected down” onto this classic setting. The argument involves making the signal  $s$  publicly observable, which removes the truthful report issue. Then set output  $y = 1/|s - \theta|$  and effort  $e = \lambda$ . I formalize a sense that my principal-expert setting is strictly richer than the classic principal-agent setting and derive new results for the classic setting. Economically, this corollary suggests that my insights apply not only to contracting for modern “knowledge/information economy” jobs but also for classic “production/manufacturing” jobs.

The literature on principal-agent problems has long been interested in when simple contracts are optimal (Carroll, 2015; Herweg et al., 2010; Oyer, 2000). This corollary provides an alternative explanation for simple contracts: the optimality of cutoff transfers stems from the nature of budgets in many contracting scenarios. The bounded budget that can only be used for this task can lead to the optimality of cutoff transfers, and the monotone likelihood ratio property is a sufficient condition to generate this sharp characterization.

Classic principal-agent models maximize the principal’s expected payoff (of outputs) minus transfers, subject to incentive compatibility (IC), individual rationality (IR), and sometimes limited liability constraints. Generally, closed-form solutions for the optimal transfers are not obtainable (see Bolton and Dewatripont, 2004, Chapter 4.5). In a special case with binary effort levels and a risk-neutral principal, a well-known result emerges: the monotone likelihood ratio property is necessary and sufficient for optimal transfers to increase with output (see Laffont and Martimort, 2009, Proposition 4.6). However, analyzing the general case with more than two effort levels is much more challenging. Grossman and Hart (1983) study this problem with continuous effort under the assumption of finite output levels. Their findings show that even the monotone likelihood ratio property is not sufficient to ensure optimal transfers to be increasing. Beyond some weak predictions, we do not have a characterization of optimal transfers.<sup>23</sup>

---

lived agent with moral hazard and adverse selection. Agents have different types (abilities) and can derive utility from working. Each agent can work for at most two periods. The principal designs a retention rule that maps from the first-period output to two outcomes: retain or fire. The principal aims to maximize output. Banks and Sundaram (1998) show that a cutoff retention rule can be optimal under some condition on the agent’s preference and MLRP. However, the principal is unable to induce effort with moral hazard alone. See their proposition 3.5 for a detailed discussion.

<sup>23</sup>For a risk-neutral principal, we only know that the optimal transfer cannot be decreasing everywhere, nor can it increase faster than the output everywhere.

From a methodological perspective, my results rely on techniques of monotone comparative statics, which allows me to impose relatively mild assumptions on the output distribution. Another popular method is the first-order approach ([Rogerson, 1985](#); [Jewitt, 1988](#)). To ensure the validity of the first-order approach, people either impose restrictive conditions on the distribution of outputs, like the convexity of output distribution ([Rogerson, 1985](#)), or impose strong assumptions on the agent's utility function ([Jewitt, 1988](#)). My results do not rely on the first-order approach.

## References

- Anthony, R. N., Govindarajan, V., Hartmann, F. G., Kraus, K., and Nilsson, G. (2007). *Management control systems*, volume 12. McGraw-Hill Boston.
- Argenziano, R., Severinov, S., and Squintani, F. (2016). Strategic information acquisition and transmission. *American Economic Journal: Microeconomics*, 8(3):119–155.
- Banks, J. S. and Sundaram, R. K. (1998). Optimal retention in agency problems. *Journal of Economic Theory*, 82(2):293–323.
- Bolton, P. and Dewatripont, M. (2004). *Contract theory*. MIT press.
- Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, 105(2):536–563.
- Carroll, G. (2019). Robust incentives for information acquisition. *Journal of Economic Theory*, 181:382–420.
- Chade, H. and Kovrijnykh, N. (2016). Delegated information acquisition with moral hazard. *Journal of Economic Theory*, 162:55–92.
- Chen, S., Wu, J., Wu, Y., and Yang, Z. (2023). Learning to incentivize information acquisition: Proper scoring rules meet principal-agent model. *arXiv preprint arXiv:2303.08613*.
- Clark, A. and Reggiani, G. (2021). Contracts for acquiring information. *arXiv preprint arXiv:2103.03911*.
- Cuccia, A. (2023). DC left millions unspent from federal grants to end homelessness. <https://thedcline.org/2023/06/08/dc-left-millions-unspent-from-federal-grants-to-end-homelessness/>.
- Dasgupta, S. (2023). Optimal test design for knowledge-based screening. *Available at SSRN 4403119*.
- Deb, R., Pai, M. M., and Said, M. (2018). Evaluating strategic forecasters. *American Economic Review*, 108(10):3057–3103.

- Deb, R., Pai, M. M., and Said, M. (2023). *Indirect Persuasion*. Centre for Economic Policy Research.
- Edlin, A. S. and Shannon, C. (1998). Strict monotonicity in comparative statics. *Journal of Economic Theory*, 81(1):201–219.
- Goel, P. K. and DeGroot, M. H. (1979). Comparison of experiments and information measures. *The Annals of Statistics*, 7(5):1066–1077.
- Gottlieb, D. and Moreira, H. (2022). Simple contracts with adverse selection and moral hazard. *Theoretical Economics*, 17(3):1357–1401.
- Grossman, S. J. and Hart, O. D. (1983). An analysis of the principal-agent problem. *Econometrica*, 51(1):7–45.
- Hébert, B. and Zhong, W. (2022). Engagement maximization. *arXiv preprint arXiv:2207.00685*.
- Herweg, F., Müller, D., and Weinschenk, P. (2010). Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, 100(5):2451–77.
- Horngren, C. T. (2009). *Cost accounting: a managerial emphasis*. Pearson Education India.
- Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.
- Jewitt, I. (1988). Justifying the first-order approach to principal-agent problems. *Econometrica: Journal of the Econometric Society*, pages 1177–1190.
- Karlin, S. and Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *The Annals of Mathematical Statistics*, pages 272–299.
- Kreutzkamp, S. (2023). Endogenous information acquisition in cheap-talk games. Technical report, Working paper.
- Laffont, J.-J. and Martimort, D. (2009). The theory of incentives: the principal-agent model. In *The theory of incentives*. Princeton university press.

- Lehmann, E. L. (2011). Comparing location experiments. In *Selected Works of EL Lehmann*, pages 779–791. Springer.
- Li, Y., Hartline, J. D., Shan, L., and Wu, Y. (2022). Optimization of scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 988–989.
- Li, Y. and Libgober, J. (2023). Optimalscorefordynamicinformation acquisition. *Working Paper*.
- Mansouri, K. (2024). Millions remain unspent in federal funds for homeless students — and time is running out. <https://www.lpm.org/news/2024-07-16/millions-remain-unspent-in-federal-funds-for-homeless-students-and-time-is-running>
- Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391.
- National Archives (2024). Code of Federal Regulations: 886.20 What happens to unused funds from my grant? <https://www.ecfr.gov/current/title-30/chapter-VII/subchapter-R/part-886/section-886.20>.
- Neyman, E., Noarov, G., and Weinberg, S. M. (2021). Binary scoring rules that incentivize precision. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 718–733.
- Osband, K. (1989). Optimal forecasting incentives. *Journal of Political Economy*, 97(5):1091–1112.
- Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of labor Economics*, 18(3):405–426.
- Papireddygar, M. and Waggoner, B. (2022). Contracts with information acquisition, via scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 703–704.
- Rappoport, D. and Somma, V. (2017). Incentivizing information design. *Available at SSRN 3001416*.

- Rogerson, W. P. (1985). The first-order approach to principal-agent problems. *Econometrica: Journal of the Econometric Society*, pages 1357–1367.
- Schick, A. (2008). *The federal budget: Politics, policy, process*. Brookings Institution Press.
- Sharma, S., Tsakas, E., and Voorneveld, M. (2023). Procuring unverifiable information. Technical report, Mimeo.
- Vandeput, N. (2021). *Data science for supply chain forecasting*. de Gruyter.
- Whitmeyer, M. and Zhang, K. (2023). Buying opinions. *arXiv preprint arXiv:2202.05249*.
- Wong, Y. F. (2023). Dynamic monitoring design. *Available at SSRN 4466562*.
- Zermeno, L. (2011). A principal-expert model and the value of menus. *Working Paper*.

## A The Assumption That Transfer Depends Only on the Difference Between the State and Report

In this section, I show that given the uniform prior, it is without loss to assume that the transfer depends only on the difference  $a - \theta$ . To make the expected transfer well-defined for the uniform prior, I take two approaches. In the first approach, I define the agent's expected transfer to be conditional on the event that the signal falls in some set of finite Lebesgue measure. Then I show that the conclusion holds for any such set. The conditioning on some set is merely an artifact to deal with the improperness of unbounded uniform prior. In the second approach, I study an alternative setting where the state space is a circle in  $\mathbb{R}^2$  and the prior is proper and uniform. This unit circle is similar to the one-point compactification of  $\mathbb{R}$ .

Throughout this section, I allow the transfer to depend on both the state and the report. To ensure that the agent's report is well-defined, I assume that the family of maps  $a \mapsto t(y + a, a)$  parametrized by  $y$  is equicontinuous. This is trivially satisfied when the transfer depends only on  $\theta - a$ , as  $t(y + a, a)$  stays constant as  $a$  varies. When the prior is improper uniform on  $\mathbb{R}$ , I additionally assume that  $t(\theta, \theta + x)$  vanishes uniformly (in  $\theta$ ) as  $x$  tends to infinity.

First, I show that the agent's report is well-defined.

**Lemma 3.** *The maximum*

$$\max_{a \in \mathbb{R}} \int_{\mathbb{R}} t(\theta', a) \varphi(\theta'; s, \lambda) d\theta'.$$

*is well-defined.*

*Proof of Lemma 3.* Fix a signal  $s$  and precision  $\lambda$ . Let  $T(a)$  be the agent's expected transfer after reporting  $a$ ,

$$T(a) = \int t(\theta', a) \varphi(\theta'; s, \lambda) d\theta'.$$

If  $T(a) = 0$  for all  $a \in \mathbb{R}$ , then the maximum is well-defined. Suppose that there exists  $a_1 \in \mathbb{R}$  such that  $T(a_1) > 0$ . As  $t(\theta, \theta + x)$  vanishes uniformly (in  $\theta$ ) as  $x$  tends to infinity, there exists a  $K_1 > 0$  such that for all  $|\theta - a| > K_1$ ,  $t(\theta, a) < \frac{T(a_1)}{2}$ . Moreover, there exists a  $K_2 > 0$  such that  $\Phi(\lambda K_2) > 1 - \frac{T(a_1)}{4}$ . Let  $K = \max(K_1, K_2)$ .



If  $|a - s| > 2K$ ,

$$\begin{aligned}
T(a) &= \int t(\theta', a) \varphi(\theta'; s, \lambda) d\theta' \\
&= \int_{\theta': |\theta' - s| < K} t(\theta', a) \varphi(\theta'; s, \lambda) d\theta' + \int_{\theta': |\theta' - s| \geq K} t(\theta', a) \varphi(\theta'; s, \lambda) d\theta' \\
&\leq \frac{T(a_1)}{2} \int_{\theta': |\theta' - s| < K} \varphi(\theta'; s, \lambda) d\theta' + 1 \cdot \int_{\theta': |\theta' - s| \geq K} \varphi(\theta'; s, \lambda) d\theta' \\
&\leq \frac{T(a_1)}{2} + 2[1 - \Phi(\lambda K)] \\
&= T(a_1).
\end{aligned}$$

Thus,

$$\max_{a \in \mathbb{R}} T(a) = \max_{|a - s| \leq 2K} T(a).$$

Next, I show that  $T(\cdot)$  is continuous.

$$\begin{aligned}
T(a + \epsilon) - T(a) &= \int t(\theta', a + \epsilon) \varphi(\theta'; s, \lambda) d\theta' - \int t(\theta', a) \varphi(\theta'; s, \lambda) d\theta' \\
&= \int [t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)] \varphi(\theta'; s, \lambda) d\theta' + \int [t(\theta' - \epsilon, a) - t(\theta', a)] \varphi(\theta'; s, \lambda) d\theta' \\
&= \int [t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)] \varphi(\theta'; s, \lambda) d\theta' + \int t(\theta', a) [\varphi(\theta' + \epsilon; s, \lambda) - \varphi(\theta'; s, \lambda)] d\theta'
\end{aligned}$$

$$\begin{aligned}
|T(a + \epsilon) - T(a)| &\leq \sup_{\theta'} |t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)| + \left| \int t(\theta', a) [\varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda)] d\theta' \right| \\
&\leq \sup_{\theta'} |t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)| + \int |\varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda)| d\theta' \\
&\leq \sup_{\theta'} |t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)| + 2 \int_{\theta' \leq s - \frac{\epsilon}{2}} \varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda) d\theta' \\
&\leq \sup_{\theta'} |t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)| + 2[\Phi(\frac{\epsilon}{2}\lambda) - \Phi(-\frac{\epsilon}{2}\lambda)]
\end{aligned}$$

Note that  $\sup_{\theta'} |t(\theta', a + \epsilon) - t(\theta' - \epsilon, a)|$  tends to 0 as  $\epsilon$  tends to 0 by the equicontinuity of  $t$ . Moreover, the CDF  $\Phi$  is continuous. Thus,  $T$  is continuous. A continuous function achieves a maximum on a compact set. So the maximum is well-defined.  $\square$

Next, I show that it is without loss to assume that the transfer depends only on the difference  $\theta - a$ . For a transfer rule  $t(\theta - a)$  that depends only on the difference

$\theta - a$ , let  $\hat{E}(\lambda; t)$  denote the agent's expected transfer when he has to report truthfully

$$\hat{E}(\lambda; t) = \int_{\mathbb{R}} t(\theta') \varphi(\theta'; 0, \lambda) d\theta'.$$

Let  $\hat{\lambda}(t)$  denote the corresponding induced precision

$$\hat{\lambda}(t) = \begin{cases} \max[\arg \max \hat{E}(\cdot; t) - c(\cdot)], & \text{if } \max \hat{E}(\cdot; t) - c(\cdot) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

For a transfer rule  $t(\theta, a)$  that depends on both the state and the report, we need to integrate the signal when computing the expected transfer. Let  $g$  denote the PDF of the signal, which is also uniform. Fix any measureable set  $A$  of finite Lebegue measure. I define the agent's expected transfer conditional on the signal falls in set  $A$ , i.e.,  $s \in A$

$$E(\lambda; t) = \int_{s \in A} g(s) ds \left( \max_a \int t(\theta', a) \varphi(\theta'; s, \lambda) d\theta' \right)$$

$$\lambda(t) = \begin{cases} \max[\arg \max E(\cdot; t) - c(\cdot)], & \text{if } \max E(\cdot; t) - c(\cdot) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

**Lemma 4.** *For a transfer rule  $t(\theta, a)$ , there exists a transfer rule  $t^*(\theta - a)$  that depends only on the difference  $\theta - a$ ; given transfer rule  $t^*$ , the agent chooses precision  $\lambda(t)$  when he is forced to reveal the signal. That is,*

$$\lambda(t) = \hat{\lambda}(t^*)$$

*Proof.* Let  $\mathbf{a}(s; \lambda)$  denote the agent's report after observing signal  $s$  given precision  $\lambda$ .

$$\mathbf{a}(s; \lambda) = \arg \max_a \int_{\mathbb{R}} t(\theta', a) \varphi(\theta'; s, \lambda) d\theta'.$$

$$E(\lambda; t) = \int_{s \in A} g(s) ds \int t(\theta', \mathbf{a}(s; \lambda)) \varphi(\theta'; s, \lambda) d\theta'$$

Note that this integral is well-defined since  $\mathbf{a}(s; \lambda)$  is well-defined and the set  $A$  has

finite Lebegue measure. Let  $\theta'' = \theta' - s$ ,

$$E(\lambda; t) = \int_{s \in A} g(s) ds \int t(\theta'' + s, \mathbf{a}(s; \lambda)) \varphi(\theta''; 0, \lambda) d\theta''.$$

Let  $t^*(\theta'') = \int_{s \in A} g(s) t(\theta'' + s, \mathbf{a}(s; \boldsymbol{\lambda}(t))) ds$ . By Fubini-Tonelli Theorem,

$$\begin{aligned} E(\boldsymbol{\lambda}(t); t) &= \int_{s \in A} g(s) ds \int t(\theta'' + s, \mathbf{a}(s; \boldsymbol{\lambda}(t))) \varphi(\theta''; 0, \boldsymbol{\lambda}(t)) d\theta'' \\ &= \int_{\mathbb{R}} t^*(\theta'') \varphi(\theta''; 0, \boldsymbol{\lambda}(t)) d\theta'' \\ &= \hat{E}(\boldsymbol{\lambda}(t); t^*) \end{aligned}$$

Moreover, at  $\lambda \neq \boldsymbol{\lambda}(t)$ ,

$$\begin{aligned} E(\lambda; t) &= \int_{s \in A} g(s) ds \int_{\mathbb{R}} t(\theta', \mathbf{a}(s; \lambda)) \varphi(\theta'; s, \lambda) d\theta' \\ &\geq \int_{s \in A} g(s) ds \int_{\mathbb{R}} t(\theta', \mathbf{a}(s; \boldsymbol{\lambda}(t))) \varphi(\theta'; s, \lambda) d\theta' \\ &= \int_{s \in A} g(s) ds \int t(\theta'' + s, \mathbf{a}(s; \boldsymbol{\lambda}(t))) \varphi(\theta''; 0, \lambda) d\theta'' \\ &= \int_{\mathbb{R}} t^*(\theta'') \varphi(\theta''; 0, \lambda) d\theta'' \\ &= \hat{E}(\lambda; t^*). \end{aligned}$$

where the inequality follows by the definition of  $\mathbf{a}(s; \lambda)$ . By the definition of  $\boldsymbol{\lambda}(t)$ ,  $\hat{\boldsymbol{\lambda}}(t^*) = \boldsymbol{\lambda}(t)$ .  $\square$

Given this Lemma, we can input the constructed  $t^*$  into the proof of Theorem 1, where I show that there exists a cutoff transfer that induces a higher precision.

In the second approach, the state space is a circle in  $\mathbb{R}^2$  with a circumference of 1, and the prior distribution is uniform. This setup corresponds to cases where the state is periodic, such as the unit vector in  $\mathbb{R}^2$  which corresponds a direction in  $\mathbb{R}^2$  (In Machina triangle, each direction corresponds to an indifference curve.), the time of day (e.g., 15:20) or the day of the year (e.g., October 19). The model remains almost the same. The only exception is that  $\phi$  admits a compact support  $[-M, M]$  and the precision  $\lambda \in [2M, +\infty]$ . The proof that it is without loss to assume that transfer

depends only on the difference  $\theta - a$  is almost the same as above. The exception is that the integral of  $s$  in Lemma 4 can be taken over the entire circle.

## B Omitted Proofs

*Proof of Lemma 1.*

$$\frac{\partial \ln \phi(\lambda x)}{\partial \ln \lambda} = \frac{\phi'(\lambda x)}{\phi(\lambda x)} \lambda x = -\eta(\lambda x).$$

Thus, we have

$$\frac{\partial [\ln \frac{\phi(\lambda x_1)}{\phi(\lambda x_2)}]}{\partial \ln \lambda} = \eta(\lambda x_2) - \eta(\lambda x_1).$$

The second part follows by

$$\frac{\partial [\frac{\varphi(x_1; 0, \lambda)}{\varphi(x_2; 0, \lambda)}]}{\partial \lambda} = \frac{\partial [\frac{\phi(\lambda x_1)}{\phi(\lambda x_2)}]}{\partial \lambda}$$

□

*Proof of Lemma 2.* Let  $\Phi$  denote the anti-derivative of  $\phi$ , i.e.,  $\Phi(x) = \int_{-\infty}^x \phi(y) dy$ . We compute the expected transfer given the cutoff transfer  $d$

$$E(\lambda; d) = 2\Phi(\lambda d) - 1.$$

The derivative of the expected transfer with respect to  $\lambda$  is

$$\frac{\partial E(\lambda; d)}{\partial \lambda} = 2d\phi(\lambda d).$$

Let us see how the derivative changes when varying  $d$ ,

$$\frac{\partial \frac{\partial E(\lambda; d)}{\partial \lambda}}{\partial d} = 2\phi(\lambda d)[1 - \eta(\lambda d)].$$

□

**Lemma 5.** *The agent's expected transfer*

$$E(\lambda; t) = \max_{a \in \mathbb{R}} \int_{\mathbb{R}} t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta'.$$

is well-defined.

*Proof of Lemma 5.* I show that the maximum is well-defined, given that the transfer rule  $t$  vanished at infinity. Fix a signal  $s$  and precision  $\lambda$ . Let  $T(a)$  be the agent's expected transfer after reporting  $a$ ,

$$T(a) = \int t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta'.$$

If  $T(a) = 0$  for all  $a \in \mathbb{R}$ , the  $\arg \max_a T(a)$  is well-defined. Suppose that there exists  $a_1 \in \mathbb{R}$  such that  $T(a_1) > 0$ . As  $t$  vanished at infinity, there exists a  $K_1 > 0$  such that for all  $|x| > K_1$ ,  $t(x) < \frac{T(a_1)}{2}$ . Moreover, there exists a  $K_2 > 0$  such that  $\Phi(\lambda K_2) > 1 - \frac{T(a_1)}{4}$ . Let  $K = \max(K_1, K_2)$ . If  $|a - s| > 2K$ ,

$$\begin{aligned} T(a) &= \int t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' \\ &= \int_{\theta': |\theta' - s| < K} t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' + \int_{\theta': |\theta' - s| \geq K} t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' \\ &\leq \frac{T(a_1)}{2} \int_{\theta': |\theta' - s| < K} \varphi(\theta'; s, \lambda) d\theta' + 1 \cdot \int_{\theta': |\theta' - s| \geq K} \varphi(\theta'; s, \lambda) d\theta' \\ &\leq \frac{T(a_1)}{2} + 2[1 - \Phi(\lambda K)] \\ &= T(a_1). \end{aligned}$$

Thus,

$$\arg \max_{a \in \mathbb{R}} T(a) = \arg \max_{|a - s| \leq 2K} T(a).$$

Next, I show that  $T(\cdot)$  is continuous.

$$\begin{aligned}
|T(a + \epsilon) - T(a)| &= \left| \int t(\theta' - a - \epsilon) \varphi(\theta'; s, \lambda) d\theta' - \int t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' \right| \\
&= \left| \int t(\theta' - a) \varphi(\theta' + \epsilon; s, \lambda) d\theta' - \int t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' \right| \\
&= \left| \int t(\theta' - a) \varphi(\theta'; s - \epsilon, \lambda) d\theta' - \int t(\theta' - a) \varphi(\theta'; s, \lambda) d\theta' \right| \\
&= \left| \int t(\theta' - a) [\varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda)] d\theta' \right| \\
&\leq \int \left| \varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda) \right| d\theta' \\
&= 2 \int_{\theta' \leq s - \frac{\epsilon}{2}} \varphi(\theta'; s - \epsilon, \lambda) - \varphi(\theta'; s, \lambda) d\theta' \\
&= 2 \left[ \Phi\left(\frac{\epsilon}{2}\lambda\right) - \Phi\left(-\frac{\epsilon}{2}\lambda\right) \right]
\end{aligned}$$

Thus,  $T$  is continuous as the CDF  $\Phi$  is continuous. A continuous function achieves a maximum on a compact set. So the maximum in  $E(\lambda; t)$  is well-defined.  $\square$

Let  $\hat{E}(\lambda; t)$  denote the agent's expected transfer when he has to report truthfully

$$\hat{E}(\lambda; t) = \int_{\mathbb{R}} t(\theta') \varphi(\theta'; 0, \lambda) d\theta'.$$

Let  $\hat{\lambda}(t)$  denote the corresponding induced precision

$$\hat{\lambda}(t) = \begin{cases} \max[\arg \max \hat{E}(\cdot; t) - c(\cdot)], & \text{if } \max \hat{E}(\cdot; t) - c(\cdot) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

**Lemma 6.** *Suppose  $f$  satisfies increasing elasticity above 1. Suppose the agent has to report truthfully. Given a symmetric transfer rule  $\tilde{t}$  and induced  $\lambda^* = \hat{\lambda}(\tilde{t})$ , the new transfer*

$$\tilde{t}'(x) = \begin{cases} 1, & \text{if } \lambda^* |x| < \eta^{-1}(1), \\ \tilde{t}(x), & \text{otherwise} \end{cases}$$

*can induce a weakly higher precision  $\hat{\lambda}(\tilde{t}') \geq \hat{\lambda}(\tilde{t})$ .*

*Proof of Lemma 6.* Let  $\Delta t = \tilde{t}' - \tilde{t}$ . I shall show that

$$\hat{E}(\lambda^*; \Delta t) \geq \hat{E}(\lambda; \Delta t) \quad \text{for all } \lambda \leq \lambda^*.$$

This implies that  $\hat{\lambda}(\tilde{t}') \geq \hat{\lambda}(\tilde{t})$  since  $\hat{\lambda}(t) = \max[\arg \max \hat{E}(\cdot; t) - c(\cdot)]$ .

Consider two cutoff transfers  $0 < d_1 < d_2 \leq \eta^{-1}(1)/\lambda^*$ . For all  $\lambda \leq \lambda^*$ , we have  $\eta(\lambda d_1) \leq 1$ ,  $\eta(\lambda d_2) \leq 1$ . By Lemma 2,

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \geq 0 \quad \Leftrightarrow \quad \eta(\lambda d) \leq 1.$$

This implies  $\frac{\partial E(\lambda; d)}{\partial \lambda}$  increases in  $d \in [d_1, d_2]$  for all  $\lambda \leq \lambda^*$ , which implies  $E(\lambda; d_2) - E(\lambda; d_1)$  increases in  $\lambda \in [0, \lambda^*]$ . As  $\Delta t$  is an integral of such difference of cutoff transfers and  $\hat{E}(\lambda; t)$  is linear in  $t$ ,  $\hat{E}(\lambda; \Delta t)$  increases in  $\lambda \in [0, \lambda^*]$ . □

**Lemma 7.** Suppose  $f$  satisfies increasing elasticity above 1. Let  $\eta^{-1}(1) < x_1 < x_2$ . Then

$$\frac{\phi(\lambda' x_2)}{\phi(\lambda' x_1)} \geq \frac{\phi(x_2)}{\phi(x_1)} \quad \text{for all } \lambda' < 1.$$

*Proof of Lemma 7.* By Lemma 1,

$$\ln \frac{\phi(\lambda' x)}{\phi(x)} = \int_{\lambda'}^1 \eta(\lambda x) d \ln \lambda.$$

Then,

$$\frac{\phi(\lambda' x_1)}{\phi(x_1)} \leq \frac{\phi(\lambda' x_2)}{\phi(x_2)} \quad \text{for all } \lambda' < 1$$

is equivalent to

$$\int_{\lambda'}^1 \eta(\lambda x_1) d \ln \lambda \leq \int_{\lambda'}^1 \eta(\lambda x_2) d \ln \lambda \quad \text{for all } \lambda' < 1.$$

Let  $y = \ln \lambda$ , then it suffices to show

$$\int_{\ln \lambda'}^0 \eta(x_1 \exp(y)) dy \leq \int_{\ln \lambda'}^0 \eta(x_2 \exp(y)) dy \quad \text{for all } \lambda' < 1.$$

Let  $\Delta y = \ln(x_2/x_1) > 0$ . Then it suffices to show

$$\int_{\ln \lambda'}^0 \eta(x_2 \exp(y - \Delta y)) dy \leq \int_{\ln \lambda'}^0 \eta(x_2 \exp(y)) dy \quad \text{for all } \lambda' < 1.$$

Let  $g(y) = \eta(x_2 \exp(y))$ . Then it suffices to show

$$\int_{\ln \lambda'}^0 g(y - \Delta y) dy \leq \int_{\ln \lambda'}^0 g(y) dy \quad \text{for all } \lambda' < 1. \quad (4)$$

Note that function  $g(y)$  is increasing in  $y \in [\ln \frac{\eta^{-1}(1)}{x_2}, 0]$  and single-crosses 1 from below at  $\ln \frac{\eta^{-1}(1)}{x_2}$ .

If  $\ln \lambda' \geq -\Delta y$ , then for all  $y \in [\ln \lambda', 0]$ ,

$$g(y) \geq g(\ln \lambda') \geq g(-\Delta y) \geq g(y - \Delta y).$$

Suppose  $\ln \lambda' < -\Delta y$ . Equation (4) is

$$\int_{\ln \lambda' - \Delta y}^{-\Delta y} g(y) dy \leq \int_{\ln \lambda'}^0 g(y) dy \quad \text{for all } \lambda' < 1.$$

I can subtract  $\int_{\ln \lambda'}^{-\Delta y} g(y) dy$  from both sides. Then the inequality becomes

$$\int_{\ln \lambda' - \Delta y}^{\ln \lambda'} g(y) dy \leq \int_{-\Delta y}^0 g(y) dy \quad \text{for all } \lambda' < 1.$$

It holds since for all  $y \in [-\Delta y, 0]$ ,  $y' \in [\ln \lambda' - \Delta y, \ln \lambda']$ ,

$$g(y) \geq g(-\Delta y) \geq g(\ln \lambda') \geq g(y').$$

□

*Proof of Theorem 1 (3) implies (1) and Proof of Theorem 2.* Suppose that  $f$  satisfies increasing elasticity above 1. If there is no  $t$  such that  $\lambda(t) > 0$ , then the problem is trivial.<sup>24</sup> Now suppose there exists a  $t$  such that  $\lambda(t) > 0$ . In step 1, I shall show that for all transfer  $t$  with  $\lambda(t) > 0$ , I can construct a new cutoff transfer  $d$  such that  $\lambda(d) \geq \lambda(t)$ . That is, cutoff transfer  $d$  weakly improves over  $t$ . In step 2, I shall

---

<sup>24</sup>In this case, we adopt the convention that every transfer is optimal.



optimize over cutoff transfers and prove Theorem 2. In step 3, I show uniqueness under stronger conditions.

### Step 1: Optimality of Cutoff Transfers

Fix a transfer rule  $t$ . First, I show that there exists a transfer rule  $t^*(\theta - a)$  such that given  $t^*$ , the agent chooses precision  $\lambda(t)$  when he has to report truthfully. That is,

$$\hat{\lambda}(t^*) = \lambda(t).$$

For the transfer rule  $t$  that depends on the state and the report, Lemma 4 gives us the desired  $t^*$ . Now assume that  $t$  depends only on the difference between the state and the report,  $\theta - a$ . The construction of  $t^*$  is simpler and is as follows. Let  $\mathbf{a}(s; \lambda(t))$  denote the agent's report after observing signal  $s$  at the precision  $\lambda(t)$ .

$$\mathbf{a}(s; \lambda(t)) = \arg \max_a \int_{\mathbb{R}} t(\theta' - a) \varphi(\theta'; s, \lambda(t)) d\theta'.$$

As both  $t$  and  $\varphi$  are translation invariant,  $\mathbf{a}(s; \lambda(t)) - s$  is a constant. I define a new transfer rule  $t^*$  by

$$t^*(x) = t(x + s - \mathbf{a}(s; \lambda(t))).$$

Given transfer  $t^*$ , the agent reports truthfully at  $\lambda(t)$ . Note that replacing  $t$  by  $t^*$  only changes the agent's report by a constant. Two transfers  $t$  and  $t^*$  provide the same incentive for the agent to choose  $\lambda$ , i.e.,

$$E(\cdot; t) = E(\cdot; t^*), \quad \lambda(t^*) = \lambda(t).$$

I define an auxiliary problem: the agent has to reveal his signal. The agent's expected transfer is

$$\hat{E}(\lambda; t^*) = \int_{\mathbb{R}} t^*(\theta') \varphi(\theta'; 0, \lambda) d\theta'.$$

It must be true that for all  $\lambda \in \mathbb{R}_+$ ,

$$\hat{E}(\lambda; t^*) \leq E(\lambda; t^*) \tag{5}$$

and

$$\hat{E}(\lambda(t); t^*) = E(\lambda(t); t^*). \tag{6}$$

The inequality holds as the agent loses the flexibility to misreport in the auxiliary

problem. The equality holds as the agent reports truthfully at  $\lambda = \boldsymbol{\lambda}(t)$ . By Equation (5) and (6),

$$\hat{\boldsymbol{\lambda}}(t^*) = \boldsymbol{\lambda}(t^*) = \boldsymbol{\lambda}(t).$$

Then, I symmetrify the transfer rule  $t^*$  to obtain

$$\tilde{t}(x) = \frac{1}{2}[t^*(x) + t^*(-x)].$$

For all  $\lambda \in \mathbb{R}_+$ , we have

$$\hat{E}(\lambda; \tilde{t}) = \hat{E}(\lambda; t^*).$$

$$\hat{\boldsymbol{\lambda}}(\tilde{t}) = \hat{\boldsymbol{\lambda}}(t^*) = \boldsymbol{\lambda}(t)$$

By Lemma 6, I can augment transfer  $\tilde{t}$  to a new transfer  $\tilde{t}'$

$$\tilde{t}'(x) = \begin{cases} 1, & \text{if } \boldsymbol{\lambda}(t)|x| < \eta^{-1}(1), \\ \tilde{t}(x), & \text{otherwise} \end{cases}$$

and induces higher precision

$$\hat{\boldsymbol{\lambda}}(\tilde{t}') \geq \hat{\boldsymbol{\lambda}}(\tilde{t}) = \boldsymbol{\lambda}(t)$$

Then I can construct a cutoff transfer  $d$ . I pin down the cutoff by requiring that cutoff transfer  $d$  and  $\tilde{t}'$  offers the same expected transfer at  $\boldsymbol{\lambda}(t)$ , i.e.,

$$E(\boldsymbol{\lambda}(t); d) = \hat{E}(\boldsymbol{\lambda}(t); \tilde{t}'). \quad (7)$$

Let  $\Delta t(x) = 1_{|x| \leq d} - \tilde{t}'(x)$  denote the transfer difference. The transfer difference must take the form

$$\Delta t(x) \begin{cases} = 0, & \text{if } |x| \leq \eta^{-1}(1)/\boldsymbol{\lambda}(t), \\ \geq 0, & \text{if } \eta^{-1}(1)/\boldsymbol{\lambda}(t) < |x| \leq d, \\ \leq 0, & \text{otherwise.} \end{cases}$$

Compare the expected transfer given cutoff transfer  $d$  and the expected transfer  $\tilde{t}'$  in the auxiliary problem.

$$E(\lambda; d) - \hat{E}(\lambda; \tilde{t}') = \int_{\mathbb{R}} \Delta t(\theta') \varphi(\theta'; 0, \lambda) d\theta'$$

Consider  $y_1, y_2$  such that  $\eta^{-1}(1)/\lambda(t) < y_1 < y_2$ . Let  $x_1 = \lambda(t)y_1$ ,  $x_2 = \lambda(t)y_2$ . Then  $\eta^{-1}(1) < x_1 < x_2$ . By Lemma 7, for all  $\lambda < \lambda(t)$ , we have

$$\begin{aligned}\frac{\phi(x_2\lambda/\lambda(t))}{\phi(x_1\lambda/\lambda(t))} &\geq \frac{\phi(x_2)}{\phi(x_1)} \\ \frac{\phi(\lambda y_2)}{\phi(\lambda y_1)} &\geq \frac{\phi(\lambda(t)y_2)}{\phi(\lambda(t)y_1)} \\ \frac{\varphi(y_2; 0, \lambda)}{\varphi(y_1; 0, \lambda)} &\geq \frac{\varphi(y_2; 0, \lambda(t))}{\varphi(y_1; 0, \lambda(t))}.\end{aligned}$$

This implies

$$E(\lambda; d) - \hat{E}(\lambda; \tilde{t}') \leq 0, \text{ if } \lambda < \lambda(t).$$

As  $\hat{\lambda}(\tilde{t}') \geq \lambda(t)$  and the definition  $\lambda(t) = \max[\arg \max E(\cdot; t) - c(\cdot)]$ , we have  $\lambda(d) \geq \lambda(t)$  (see Figure 8).

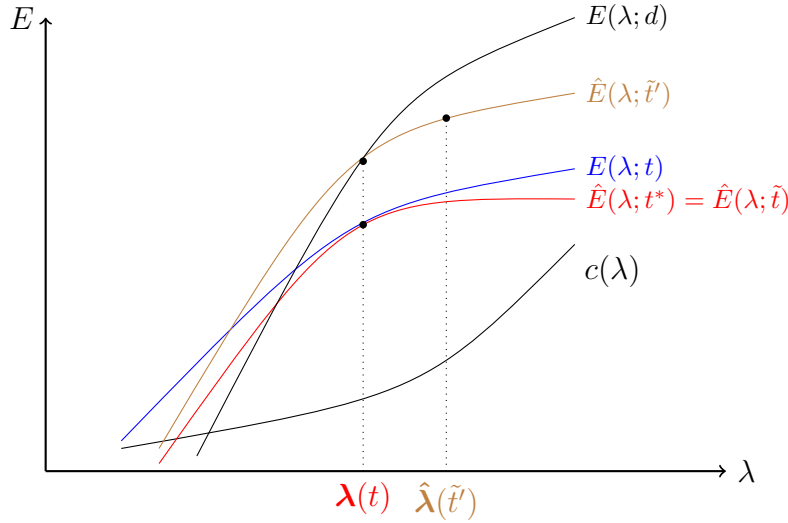


Figure 8: Expected Transfer

### Step 2: Proof of Theorem 2

Next, I optimize over cutoff transfers to prove Theorem 2 and show the existence of the optimal cutoff transfer. If we set  $d < \bar{d}$ , the agent never chooses to work. So consider  $d \geq \bar{d}$ .

**Case 1:**  $\lambda(\bar{d})\bar{d} \geq \eta^{-1}(1)$ .

Let  $\tilde{\lambda} = \eta^{-1}(1)/\bar{d} \leq \lambda(\bar{d})$ . For any  $d > \bar{d}$ , if  $\lambda(d) \leq \tilde{\lambda}$ , then  $\lambda(d) \leq \lambda(\bar{d})$ . Consider

$\lambda(d) > \tilde{\lambda}$ . For all  $d \geq \bar{d}$  and  $\lambda \geq \tilde{\lambda}$ ,

$$\eta(\lambda d) \geq \eta(\bar{d}\tilde{\lambda}) \geq 1 \quad \Rightarrow \quad \frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0.$$

Since  $\lambda(d) = \max[\arg \max E(\cdot; d) - c(\cdot)]$ . By the Topkis' monotone comparative statics theorem,  $\lambda(d) \leq \lambda(\bar{d})$  for all  $d \geq \bar{d}$ . Thus,  $\bar{d}$  is the optimal cutoff.

**Case 2:**  $\lambda(\bar{d})\bar{d} < \eta^{-1}(1)$ . For all  $\bar{d} \leq d \leq d^*$  and  $\lambda \leq \frac{\eta^{-1}(1)}{d}$ ,

$$\eta(\lambda d) \leq 1 \quad \Rightarrow \quad \frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \geq 0.$$

This implies  $\lambda(d)$  is increasing in  $d$  for  $\bar{d} \leq d \leq d^*$ . Now suppose  $d > d^*$ . If  $\lambda(d)d \leq \eta^{-1}(1)$ , then

$$\lambda(d) \leq \eta^{-1}(1)/d \leq \eta^{-1}(1)/d^* = \lambda^*.$$

If  $\lambda(d)d > \eta^{-1}(1)$ , for all  $d \geq d^*$  and  $\lambda \geq \eta^{-1}(1)/d$ ,

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0 \quad \Rightarrow \quad \lambda(d) \leq \lambda(d^*)$$

by the Topkis' monotone comparative statics theorem.

Now, I prove existence. In case 1,  $\bar{d}$  is the optimal cutoff transfer. In case 2, I can increase  $d$  starting from  $\bar{d}$ . Before hitting  $d^*$ ,  $\lambda(d)$  increases in  $d$ . As  $\eta^{-1}(1)$  is well-defined and finite given that  $f$  satisfies increasing elasticity above 1. Eventually, increasing  $\lambda(d)d$  can hit  $\eta^{-1}(1)$ . Thus,  $d^*$  is well defined.

### Step 3: Uniqueness

Finally, suppose that  $\phi$  satisfies strictly increasing elasticity above 1, the cost function is continuously differentiable, and the optimal precision is an interior solution (not 0 or  $+\infty$ ).<sup>25</sup> I show that the optimal transfer rule is unique. First, I show that for any transfer rule  $t$  that is not a cutoff transfer, the cutoff transfer constructed above induces a strictly larger precision:  $\lambda(d) > \lambda(t)$ . Since  $t$  is not a cutoff transfer,

---

<sup>25</sup>Distribution  $\phi$  satisfies *strictly increasing elasticity above 1* if  $\eta(\cdot)$  single-crosses 1 from below,  $\{x | \eta(x) = 1\}$  is a singleton, and is strictly increasing after the cross.

at least one of  $\tilde{t}' - \tilde{t}$  and  $\Delta t$  is a non-zero function. If  $\tilde{t}' - \tilde{t}$  is non-zero,

$$\left. \frac{\partial[\hat{E}(\lambda; \tilde{t}') - \hat{E}(\lambda; \tilde{t})]}{\partial \lambda} \right|_{\lambda=\lambda^*} > 0. \quad (8)$$

This follows by a strict version of Lemma 6 in which strictly increasing elasticity above 1 implies

$$\eta(\lambda d) < 1 \quad \Rightarrow \quad \frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} > 0.$$

As  $\hat{E}(\cdot; \tilde{t}) - c(\cdot)$  is continuously differentiable,<sup>26</sup>  $\hat{\lambda}(\tilde{t})$  is interior, Equation (8) implies  $\hat{\lambda}(\tilde{t}') > \hat{\lambda}(\tilde{t})$  by Edlin and Shannon (1998). This implies  $\lambda(d) > \lambda(t)$ .

If  $\Delta t$  is non-zero, next I show

$$\left. \frac{\partial[E(\lambda; d) - \hat{E}(\lambda; \tilde{t}')] }{\partial \lambda} \right|_{\lambda=\lambda(t)} > 0. \quad (9)$$

By the construction of  $\Delta t$ ,

$$\int \Delta t(x) \varphi(x; 0, \lambda(t)) dx = 0 \quad (10)$$

$$\int \Delta t(x) \phi(\lambda(t)x) dx = 0$$

Take some small  $\epsilon > 0$ . Let  $\lambda' = \lambda(t) \exp(\epsilon)$ . Consider

$$E(\lambda'; d) - \hat{E}(\lambda'; \tilde{t}') = \int \Delta t(x) \lambda' \phi(\lambda' x) dx$$

By  $\ln(\lambda') - \ln(\lambda(t)) = \epsilon$  and Lemma 1,

$$\ln \phi(\lambda' x) - \ln \phi(\lambda(t)x) = -\eta(\lambda(t)|x|)\epsilon + \mathcal{O}(\epsilon)$$

---

<sup>26</sup>  $\hat{E}(\cdot; \tilde{t})$  is continuously differentiable as  $\phi$  is continuously differentiable.

Thus,

$$\begin{aligned}
E(\lambda'; d) - \hat{E}(\lambda'; \tilde{t}') &= \int \Delta t(x) \lambda' \phi(\lambda' x) dx \\
&= \lambda' \int \Delta t(x) \phi(\boldsymbol{\lambda}(t)x) \exp(-\epsilon \eta(\boldsymbol{\lambda}(t)|x|)) dx + o(\epsilon) \\
&= -\lambda' \int \Delta t(x) \phi(\boldsymbol{\lambda}(t)x) \epsilon \eta(\boldsymbol{\lambda}(t)|x|) dx + o(\epsilon)
\end{aligned}$$

where the last equality follows by Equation (10).

$$\begin{aligned}
\lim_{\epsilon \rightarrow 0} \frac{E(\lambda'; d) - \hat{E}(\lambda'; \tilde{t}')}{\epsilon} &= -\boldsymbol{\lambda}(t) \int \Delta t(x) \phi(\boldsymbol{\lambda}(t)x) \eta(\boldsymbol{\lambda}(t)|x|) dx \\
&= - \int \Delta t(x) \varphi(x; 0, \boldsymbol{\lambda}(t)) \eta(\boldsymbol{\lambda}(t)|x|) dx
\end{aligned}$$

$$\left. \frac{\partial [E(\lambda; d) - \hat{E}(\lambda; \tilde{t}')] }{\partial \lambda} \right|_{\lambda=\boldsymbol{\lambda}(t)} = -\frac{1}{\boldsymbol{\lambda}(t)} \int \Delta t(x) \varphi(x; 0, \boldsymbol{\lambda}(t)) \eta(\boldsymbol{\lambda}(t)|x|) dx$$

which is strictly positive by Equation (10),  $\eta(y)$  being strictly increasing for  $y > \eta^{-1}(1)$ , and  $\Delta t(x)$  being supported on  $|x| \geq \frac{\eta^{-1}(1)}{\boldsymbol{\lambda}(t)}$ . Again by  $\hat{E}(\cdot; \tilde{t}') - c(\cdot)$  being continuously differentiable,  $\hat{\boldsymbol{\lambda}}(\tilde{t}')$  is interior, Equation (9) implies  $\boldsymbol{\lambda}(d) > \hat{\boldsymbol{\lambda}}(\tilde{t}') \geq \boldsymbol{\lambda}(t)$  by Edlin and Shannon (1998).

I have shown that for any transfer rule  $t$  that is not a cutoff transfer,  $\boldsymbol{\lambda}(d) > \boldsymbol{\lambda}(t)$ . Thus, any optimal transfer must be a cutoff transfer. Given strictly increasing elasticity above 1, cost function being continuously differentiable, optimal precision being interior, all comparative statics analyses in the proof of Theorem 2 are strict. Thus, there exists a unique optimal transfer rule, which is a cutoff transfer.  $\square$

**Lemma 8.** *Suppose that for all cost functions, there exists an optimal transfer that is a cutoff transfer. For a cost function  $c \in C^1$ , let  $d^*$  be the optimal cutoff and  $\lambda^*$  be the induced precision with  $\eta(\lambda^* d^*) \geq 1$ . Then for all  $x_1 \in [0, \lambda^* d^*)$  and  $x_2 \in [\lambda^* d^*, +\infty)$ ,*

$$\eta(x_1) \leq \eta(x_2).$$

*Proof of Lemma 8.* Without loss I can assume  $c(\lambda^*) = E(\lambda^*; d^*)$  since I can increase  $c$  by a constant without affecting  $d^*$  and  $\lambda^*$ . Similarly, without loss of generality, I

assume that  $\lambda^*$  uniquely maximizes  $E(\lambda; d^*) - c(\lambda)$ .<sup>27</sup>

Towards a contradiction, suppose  $\exists x_1 \in [0, \lambda^* d^*)$ ,  $x_2 \in [\lambda^* d^*, +\infty)$  such that  $\eta(x_1) > \eta(x_2)$ . As  $\eta(\cdot)$  is continuous, I can pick  $\delta_1 > 0$ ,  $\delta_2 > 0$  small enough such that for all  $y_1 \in [\frac{x_1}{\lambda^*} - \delta_1, \frac{x_1}{\lambda^*} + \delta_1]$  and  $y_2 \in [\frac{x_2}{\lambda^*} - \delta_2, \frac{x_2}{\lambda^*} + \delta_2]$ ,

$$\eta(y_1 \lambda^*) > \eta(y_2 \lambda^*) \quad (11)$$

and

$$\int_{\frac{x_1}{\lambda^*} - \delta_1}^{\frac{x_1}{\lambda^*} + \delta_1} \varphi(x; 0, \lambda^*) dx = \int_{\frac{x_2}{\lambda^*} - \delta_2}^{\frac{x_2}{\lambda^*} + \delta_2} \varphi(x; 0, \lambda^*) dx.$$

I define

$$\Delta t(x) = -1_{|x| \in [\frac{x_1}{\lambda^*} - \delta_1, \frac{x_1}{\lambda^*} + \delta_1]} + 1_{|x| \in [\frac{x_2}{\lambda^*} - \delta_2, \frac{x_2}{\lambda^*} + \delta_2]}$$

and a new transfer  $\underline{t}(x) = 1_{|x| \leq d^*} + \Delta t(x)$ .

I can set  $\delta_1$  and  $\delta_2$  to be small enough such that the agent reports truthfully given the transfer  $\underline{t}$ , since  $\varphi'(d^*; 0, \lambda^*) < 0$  is bounded from above due to  $\eta(\lambda^* d^*) \geq 1$ . To see this, it suffices to consider the case  $\lambda^* = 1$ . If the agent reports truthfully, the expected transfer is  $2\Phi(d^*) - 1$ . Now suppose the agent misreports by  $\epsilon' > 0$ . Then his expected transfer is

$$\begin{aligned} & \Phi(d^* - \epsilon') + \Phi(d^* + \epsilon') - 1 \\ & + \Phi(x_2 - \epsilon' + \delta_2) - \Phi(x_2 - \epsilon' - \delta_2) + \Phi(x_2 + \epsilon' + \delta_2) - \Phi(x_2 + \epsilon' - \delta_2) \\ & - [\Phi(x_1 - \epsilon' + \delta_1) - \Phi(x_1 - \epsilon' - \delta_1)] - [\Phi(x_1 + \epsilon' + \delta_1) - \Phi(x_1 + \epsilon' - \delta_1)] \end{aligned}$$

where the second and the third line can be made arbitrarily close to 0 as  $\delta_1$  and  $\delta_2$  tend to 0, since  $F$  is continuous. Moreover, the difference

$$\Phi(d^* - \epsilon') + \Phi(d^* + \epsilon') - 2\Phi(d^*)$$

is strictly negative and is strictly decreasing in  $\epsilon'$ , due to  $f$  being single-peaked and  $\eta(d^*) = -\frac{f'(d^*)}{\phi(d^*)} d^* \geq 1$ . Therefore, the expected transfer under misreport is lower than  $2\Phi(d^*) - 1$  when  $\delta_1$  and  $\delta_2$  are small enough.

Then we have

$$E(\lambda^*; \underline{t}) = E(\lambda^*; d^*).$$

---

<sup>27</sup>Otherwise, I can always increase  $c(\lambda)$  for  $\lambda \neq \lambda^*$ , without affecting  $d^*$  and  $\lambda^*$ .

By Lemma 1 and Equation (11),

$$\left. \frac{\partial[E(\lambda; \underline{t}) - E(\lambda; d^*)]}{\partial \lambda} \right|_{\lambda=\lambda^*} > 0.$$

Since previously  $\lambda^*$  uniquely maximizes  $E(\lambda; d^*) - c(\lambda)$ , we can pick  $\delta_1$  and  $\delta_2$  small enough such that  $\lambda(\underline{t}) \in (\lambda^* - \epsilon, \lambda^* + \epsilon)$ , for some small  $\epsilon > 0$ . As  $c$  is continuously differentiable,  $E(\lambda^*; \underline{t}) = E(\lambda^*; d^*)$ ,

$$\left. \frac{\partial[E(\lambda; \underline{t}) - E(\lambda; d^*)]}{\partial \lambda} \right|_{\lambda=\lambda^*} > 0,$$

by Edlin and Shannon (1998) we have  $\lambda(\underline{t}) > \lambda^*$ , contradicting that  $(d^*, \lambda^*)$  is optimal (Figure 9).

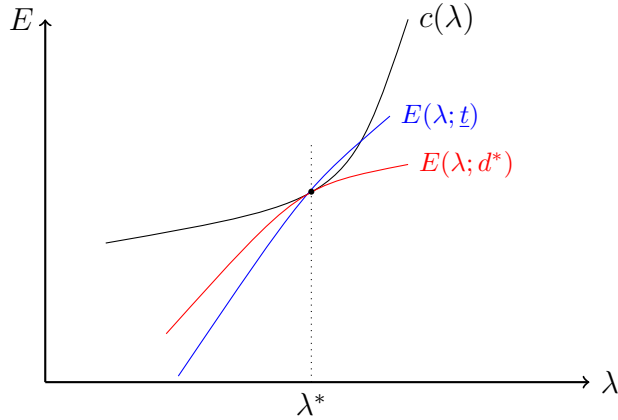


Figure 9: Expected Transfer

□

*Proof of Theorem 1 (2) implies (3).* Suppose that for all increasing and continuously differentiable cost functions, there exists an optimal transfer that is a cutoff transfer. Take an increasing cost function  $c_0 \in C^1$ , with  $d_0$  and  $\lambda_0$  being the corresponding optimal cutoff and induced precision. I can pick  $c_0$  such that  $\lambda_0$  is the unique maximizer of<sup>28</sup>  $E(\cdot; d_0) - c_0(\cdot)$ . As  $E(\cdot; d_0) - c(\cdot)$  is continuously differentiable, I can use the

<sup>28</sup>Given any increasing cost function  $c_0 \in C^1$ , I can always construct a new increasing and continuously differentiable cost function by increasing  $c_0(\lambda)$  for all  $\lambda \neq \lambda_0$  while keeping  $c_0(\lambda_0)$  and  $\frac{dc_0}{d\lambda}|_{\lambda=\lambda_0}$  unchanged.



first-order approach. As

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0 \quad \Leftrightarrow \quad \eta(\lambda d) \geq 1,$$

we have  $\eta(\lambda_0 d_0) \geq 1$ . If not, I can slightly increase  $d_0$ , which leads to a larger

$$\left. \frac{\partial E(\lambda; d)}{\partial \lambda} \right|_{\lambda=\lambda_0}$$

This induces a larger  $\lambda > \lambda_0$ , a contradiction.<sup>29</sup>

Since the cutoff transfer  $d_0$  is the optimal transfer and  $c_0 \in C^1$ , by Lemma 8 and  $\eta(\cdot)$  being continuous,

$$\eta(x) \leq \eta(\lambda_0 d_0) \quad \text{if } x \leq \lambda_0 d_0$$

$$\eta(x) \geq \eta(\lambda_0 d_0) \quad \text{if } x \geq \lambda_0 d_0.$$

Note that  $\eta(0^+) < 0$  (as  $f$  is integrable around 0),  $\eta(\lambda_0 d_0) \geq 1$ ,  $\eta(\cdot)$  is continuous, I can find the largest  $x$  where  $\eta(\cdot)$  crosses 1 from below

$$x_1 = \min\{x | \eta(x') \geq 1, \text{ if } x' \geq x\} \leq \lambda_0 d_0.$$

$\eta(x) \geq 1$  when  $x \geq x_1$ .

Now pick  $d_1$  and  $\lambda_1$  such that  $\lambda_1 d_1 = x_1$ . Construct an increasing cost function  $c_1 \in C^1$  such that

$$c_1(\lambda) > E(\lambda; d_1) \quad \text{if } \lambda \neq \lambda_1$$

$$c_1(\lambda_1) = E(\lambda_1; d_1)$$

$$\left. \frac{\partial E(\lambda; d_1)}{\partial \lambda} \right|_{\lambda=\lambda_1} = \left. \frac{dc_1(\lambda)}{d\lambda} \right|_{\lambda=\lambda_1}.$$

As

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0 \quad \Leftrightarrow \quad \eta(\lambda d) \geq 1$$

and  $\eta(x) \geq 1$  for all  $x \geq x_1$ , the cutoff transfer  $d_1$  is the best among all cutoff transfers for cost function  $c_1$ . Since there exists an optimal transfer that is a cutoff transfer, cutoff transfer  $d_1$  is the optimal transfer, and  $\lambda_1$  is the maximum precision.

---

<sup>29</sup>As  $\lambda_0$  is the unique maximizer of  $E(\cdot; d_0) - c_0(\cdot)$ , I can always make the increment on  $d_0$  small enough such that the new maximizer never falls in  $[0, \lambda_0)$ .

By Lemma 8,

$$\eta(x) \leq 1 \quad \text{if } x \leq x_1$$

$$\eta(x) \geq 1 \quad \text{if } x \geq x_1.$$

Thus,  $\eta(\cdot)$  single-crosses 1 at  $x_1$ .

Similarly, for all  $x_2 > x_1$ , I can pick  $d_2$  and  $\lambda_2$  such that  $\lambda_2 d_2 = x_2$ . Pick a increasing cost function  $c_2 \in C^1$  such that

$$c_2(\lambda) > E(\lambda; d_2) \quad \text{if } \lambda \neq \lambda_2$$

$$\begin{aligned} c_2(\lambda_2) &= E(\lambda_2; d_2) \\ \left. \frac{\partial E(\lambda; d_2)}{\partial \lambda} \right|_{\lambda=\lambda_2} &= \left. \frac{dc_2(\lambda)}{d\lambda} \right|_{\lambda=\lambda_2} \end{aligned}$$

By the same argument, we have

$$\eta(x) \leq \eta(x_2) \quad \text{if } x \leq x_2$$

$$\eta(x) \geq \eta(x_2) \quad \text{if } x \geq x_2.$$

Thus,  $\eta(\cdot)$  is increasing once it goes above 1 at  $x_1$ .

□

*Proof of Proposition 1.* First suppose for all  $\lambda$ ,  $c_2(\lambda) \geq E(\lambda; d^*(c_1))$ . Then  $\bar{d}(c_2) \geq d^*(c_1)$ , implying  $d^*(c_2) \geq d^*(c_1)$ . Moreover, since

$$\text{for all } \lambda \geq \lambda^*(c_1) \text{ and } d \geq d^*(c_1), \quad \frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \leq 0$$

and  $c_2 - c_1$  is increasing,

$$\max[\arg \max_{\lambda} E(\lambda; d^*(c_2)) - c_2(\lambda)] \leq \max[\arg \max_{\lambda} E(\lambda; d^*(c_1)) - c_1(\lambda)].$$

Thus,  $\lambda^*(c_2) \leq \lambda^*(c_1)$ . Now suppose for some  $\lambda$ ,  $c_2(\lambda) \leq E(\lambda; d^*(c_1))$ . Note that

$$\begin{aligned} & \max[\arg \max_{\lambda} E(\lambda; d^*(c_1)) - c_2(\lambda)] \\ &= \max[\arg \max_{\lambda} E(\lambda; d^*(c_1)) - c_1(\lambda) - (c_2(\lambda) - c_1(\lambda))] \\ &\leq \max[\arg \max_{\lambda} E(\lambda; d^*(c_1)) - c_1(\lambda)] = \lambda^*(c_1) \end{aligned}$$

where the inequality follows by  $c_2 - c_1$  being increasing. Thus,

$$\lambda(d^*(c_1); c_2) d^*(c_1) \leq \lambda^*(c_1) d^*(c_1) = \eta^{-1}(1).$$

By the proof of Theorem 2,  $d^*(c_2) \geq d^*(c_1)$  and  $\lambda^*(c_2) \leq \lambda^*(c_1)$ . □

*Proof of Corollary 2.* Changing the noise from  $\varepsilon_1$  to  $k\varepsilon_1$  is equivalent to keeping the noise at  $\varepsilon_1$  and changing the cost function from  $c_1(\lambda)$  to  $c_2(\lambda) = c_1(k\lambda)$ . Notice that

$$c_2(\lambda) - c_1(\lambda) = c_1(k\lambda) - c_1(\lambda)$$

which is increasing in  $\lambda$  due to the convexity of  $c_1$ . The conclusion follows by Proposition 1. □

*Proof of Proposition 2.* The proof of the first part remains the same as Theorem 1. For proving the second part, the only difference is to note that the expected transfer is

$$E(\lambda; d) = \int_0^{\lambda d} \phi(r) \frac{\pi^{n/2}}{\Gamma(n/2 + 1)} dr^n$$

where  $\Gamma$  is the gamma function and  $\frac{\pi^{n/2}}{\Gamma(n/2+1)} r^n$  is the volume of the  $n$ -dimensional ball.

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} = \frac{n\pi^{n/2}}{\Gamma(n/2 + 1)} \phi(\lambda d) (\lambda d)^{n-1} [n - \eta(\lambda d)].$$

Thus,

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} \geq 0 \quad \Leftrightarrow \quad \eta(\lambda d) \leq n.$$

The rest of the argument remains the same. □

*Proof of Proposition 3.* The optimality of cutoff transfers follows by the Proof of Theorem 1 (3) implies (1), as Gaussian distributions satisfy increasing elasticity.

For proving the counterpart of Theorem 2, it suffices to notice that

$$\frac{\partial^2 E(\lambda; d)}{\partial \lambda \partial d} = \frac{n\pi^{n/2}}{\Gamma(n/2 + 1)} \phi(\Lambda(\lambda)d) (\Lambda(\lambda)d)^{n-1} \frac{d\Lambda(\lambda)}{d\lambda} [n - \eta(\Lambda(\lambda)d)]$$

where  $\Lambda(\lambda) = (\frac{1}{\lambda_p^2} + \frac{1}{\lambda^2})^{-\frac{1}{2}}$  for uniform prior and  $\Lambda(\lambda) = (\frac{1}{\lambda_p^2} + \frac{1}{\lambda^2 + \lambda_0^2})^{-\frac{1}{2}}$  for Gaussian prior.  $\square$

*Proof of Proposition 4.* I first prove the case of uniform prior. Given the agent's signal  $s$ , his posterior belief about the state is  $\mathcal{N}(s, 1/\lambda^2)$ . Since the principal's signal is Gaussian centered at  $\theta$  with variance  $\frac{1}{\lambda_p^2}$ , the agent's posterior belief about  $s_p$  is

$$\mathcal{N}(s, \frac{1}{\lambda^2} + \frac{1}{\lambda_p^2}).$$

Now the agent's payoff maximization problem is same as in Section 4 except that  $s_p$  replaces  $\theta$  and his precision is  $(\frac{1}{\lambda^2} + \frac{1}{\lambda_p^2})^{-\frac{1}{2}}$ . Thus, Proposition 4 follows from Theorem 1 and 2.

The proof for the Gaussian prior is similar. The only difference is that conditional on a signal  $s$ , the agent's posterior about  $s_p$  is

$$\mathcal{N}(\frac{s\lambda^2}{\lambda_0^2 + \lambda^2}, \frac{1}{\lambda_0^2 + \lambda^2} + \frac{1}{s_p^2}).$$

Thus, the precision is  $\Lambda = [\frac{1}{\lambda_0^2 + \lambda^2} + \frac{1}{s_p^2}]^{-\frac{1}{2}}$   $\square$