A case for transparency in principal-agent relationships

Emiliano Catonini[*] and Sergey Stepanov[†]

July 9, 2022

Abstract

When is transparency optimal for the principal in principal-agent relationships? We consider the following setting. The principal has private information that affects the agent's incentives to exert effort. Higher effort leads to higher material utility for both parties but the agent bears the cost of effort. The principal can share her information with the agent and can commit to any information structure. We obtain interpretable and easily verifiable sufficient conditions for the optimality of full disclosure. With this, we show that full disclosure is optimal under some modeling assumptions commonly used in applied principal-agent papers.

JEL classification: D82, D83

Keywords: information design, Bayesian persuasion, transparency

1 Introduction

In many principal-agent relationships, the principal has private information about some exogenous factor that influences the agent's incentives to exert effort. In this paper we address a simple question: when is it ex-ante optimal for the principal to implement the policy of full disclosure of such information? The answer is non-obvious, because disclosing states that motivate the agent to exert more effort entails discouraging the agent in other states.

[*]NYU Shanghai. Email: ec1838@nyu.edu.

[†]HSE University, International College of Economics and Finance and Faculty of Economic Sciences, Russia. Postal address: Office S641, Pokrovsky Boulevard 11, 109028 Moscow, Russia. Email: sstepanov@hse.ru

Following the Bayesian persuasion literature pioneered by Rayo and Segal (2010) and Kamenica and Gentzkow (2011), we consider a principal who has commitment power over the information she reveals to the agent. Without setting any restrictions on possible persuasion strategies, we search for conditions that make full disclosure optimal. Differently from other complicated information design schemes, just disclosing the truth seems to be a realistic goal in many scenarios — e.g., with transparency policies in organizations.

As the state space in our model can be a continuum, the concavification approach of Kamenica and Gentzkow (2011) cannot be applied. Despite that, we manage to find a rather mild and easy to check condition that ensures the optimality of transparency. Differently from several recent papers (Dworczak and Kolotilin (2019), Dworczak and Martini (2019), Gentzkow and Kamenica (2016), Kolotilin (2018), Kolotilin et al. (2021), Arieli et al. (2020)), our condition speaks directly to the underlying incentives of the parties, as opposed to the indirect utility function of the principal, and we do not assume that the sender's (principal's) payoff is a function of the expected state (or any moments of the posterior distribution).

To see why the effect of transparency may be non-trivial, consider a simple example. The agent generates an output which he and the principal split in a fixed proportion. The output is increasing in the agent's effort, and the agent bears the cost of effort. The underlying state of nature determines the productivity of effort, with a higher state resulting in higher productivity. At first sight, the principal would always want to commit to revealing the state to the agent, as both parties *seem* to benefit from effort more when the state is higher. Here is a simple argument why it does not have to be the case. Suppose the agent is sufficiently risk averse. Then, good news about the productivity may actually depress effort. This is because a higher productivity implies that the agent reaches a higher income, hence a lower marginal utility, at lower levels of effort. If the principal is risk neutral, then disclosure discourages the agent precisely when the principal benefits more from effort (and incentivizes the agent when the principal gains less from effort). In such a case, transparency is unlikely to be optimal. Alternatively, the agent may increase effort under the good news and reduce it under the bad news, as the principal wants, but the increase may be smaller than the decrease, to the point that the overall effect on the principal's utility is negative.

We look for conditions on the incentives of the parties that guarantee optimality of full disclosure in the following general setting. Both the principal and the agent derive some material utility from the agent's effort. Higher effort yields higher material utility for both parties, but the agent also suffers a cost of effort. The underlying state of nature determines the effect of effort on the utility of the principal and the agent. The principal can commit to any mapping from the state space to messages for the agent. We are after a condition under which any message that pools (or partially pools) more than one state, can be "split" so as to improve the principal's ex-ante payoff.

To understand our condition, consider the simple case of a message that pools two ex-ante equally likely states — we provide an example of this kind in the next section. The principal contemplates splitting this message into two messages that reveal the state. Then, if this makes any difference, the agent will increase effort under one state and decrease it under the other state. Two forces determine whether the principal gains from the split or not: the changes in the agent's effort and the changes in the principal's utility per unit of effort. Intuitively, the variations of effort are driven by the speed at which the agent's marginal utility of effort changes compared to its marginal cost. For normalization, let us call the cost of effort simply "effort". Then we show that what matters is the ratio between the derivative of the principal's utility with respect to effort and the absolute value of the second derivative of the agent's utility with respect to effort. This ratio, in fact, measures how much the utility of the principal increases/decreases as the agent increases/reduces effort, per unit of the agent's marginal utility of effort that is lost/gained in the process. We show that the principal benefits from the split if the ratio is larger under the effort-increasing state than under the effort-decreasing state, for any pair of effort levels between the old one and the two new ones.

Our main result generalizes this argument to all possible messages. Specifically, in Section 3.1, we show that *any* message with a non-singleton support can be split so as to improve the principal's welfare if the ratio introduced above increases under any change in the effort and state that involves an increase in both the effort and the agent's marginal utility of effort. This condition, hence, ensures the optimality of full disclosure. Under the assumptions of continuity of the state space and differentiability of the derivatives of the parties' utilities with respect to the state, we also provide an equivalent "derivative" condition, which may be easier to check in some economic applications.

In addition to our sufficient condition for transparency, we also derive a sufficient condition for suboptimality of transparency. Although there remains a "gap" between the two conditions (one is not a negation of the other), the latter is still helpful to establish when transparency is definitely not optimal, as we show in an example.

Then we consider a special case in which the principal's utility is a convex transformation of the agent's material utility. This feature naturally arises in settings when the effort is exerted to generate some monetary output (profit), the agent is more risk averse than the principal and receives a compensation that is not too convex (e.g., linear) in the output. Under the additional assumption that the agent's material utility is supermodular in effort and state, we provide a sufficient condition for the optimality of full disclosure that involves only the agent's preferences.

Finally, we discuss several examples demonstrating that our condition is easy to check, it is not very strong, and is often satisfied in economic applications. The examples also shed light on the role of risk aversion in the optimality/suboptimality of transparency.

The first example follows the setting outlined in the fourth paragraph of the introduction. In this example (Section 4.1), we assume that both parties exhibit CRRA and look more closely at the effects of the parties' risk aversion. Transparency turns out to be optimal, when the agent is more risk averse than the principal (a typical textbook situation) but not too risk averse (with the coefficient of relative risk aversion below one). In this case, the preferences of the parties are sufficiently aligned. For both parties, state and effort remain complements. Disclosing the states raises the effort in expectation, and the principal benefits more from effort in higher states. Instead, when the agent becomes too risk averse (while the principal remains moderately risk averse), transparency ceases to be optimal. As we have discussed in the beginning in this section, under high risk aversion, good news about productivity *depress* effort, that is, effort and state become substitutes for the agent while remaining complements for the principal.

Another interesting, though less realistic, case discussed in Section 4.1 is when the agent is sufficiently risk averse, and the principal is at least as risk averse as the agent. In that case, the average effort falls but the principal nevertheless gains from transparency. This happens because for the principal effort and state are even more substitutes than for the agent. Bad news about productivity encourages effort, and the principal benefits even more from effort in lower states than the agent does.

In Section 4.2 we simplify preferences by assuming risk neutrality for both parties and focus instead on the properties of the production function that ensure the optimality of transparency.

Finally, in Section 5 we consider a particular setting that allows deriving a necessary and sufficient condition for transparency using Jensen's inequality. We then provide examples when our sufficient condition becomes necessary and sufficient, and when, on the contrary, transparency can be optimal even when our condition is not satisfied.

Two recent papers, Dworczak and Martini (2019) and Kolotilin (2018), have provided conditions for the optimality of full disclosure in persuasion. Using a "duality approach" and linear programming techniques, they provide a tractable solution method when the indirect utility of the sender depends only on the expected state.[1] As mentioned, differently from us, they give conditions in terms of the sender's indirect utility function, abstracting away from the underlying economic problem.[2] We focus instead on a (broad) class of economic problems, but without any restriction on how the state affects utilities. More importantly, we directly look for conditions in terms of the utility functions of both the principal and the agent, in order to gain economic intuition and policy implications.

In a framework similar to ours, under rather general assumptions on the payoffs, Kolotilin and Wolitzky (2020) provide a necessary and sufficient condition for the optimality of full disclosure. However, also their condition is expressed in terms of the indirect sender's utility function and has a very general form that does not provide a tractable recipe for checking whether full disclosure is optimal for given utility functions of the parties.

Using a concept analogous to the concept of "virtual value" in the mechanism design literature, Mensch (2021) offers conditions for transparency jointly on the receiver's utility function and on a transformation of the sender's utility function that takes into account the incentive compatibility constraint of the receiver ("virtual utility"). His focus is on the importance of complementarities between states and actions, and whether these complementarities "point in the same direction" for the sender and the receiver. Our and Mensch's papers can be viewed as complementary to each other. While Mensch's condition for transparency (Theorem 5) is intuitive, it is rather abstract and not

[1] See also Dworczak and Kolotilin (2019), Dizdar and Kováč (2020), Galperti and Perego (2018).

[2] On a completely different note, Jehiel (2004) shows that transparency is generically suboptimal. Roughly speaking, this is because the conditions for transparency can only be expressed as conditions at each and every point of the domain of the parameters space, so they are generically violated somewhere.

straightforward to apply, as it requires a derivation of the "virtual utility". Instead, our conditions are directly on primitives of the model, that is, the shapes of the parties' utility functions. We provide an additional discussion of the relationship between our results and Mensch (2012) in Section 4.

The paper is organized as follows. Section 2 provides a simple example that aims to deliver intuition for the main results. Section 3 derives sufficient conditions for the optimality as well as suboptimality of full disclosure in a general framework. Section 4 discusses examples, the role of risk aversion and complementarity/substitutability between effort and state. In Section 5, we compare our condition with a necessary and sufficient condition in a simple example.

2 Example with two equally likely states

Consider the following principal-agent problem. The principal wants the agent to exert effort on a project. The cost of effort for the agent is linear: $c(e) = e$. There is a binary state of the world $\omega \in \{\omega_1, \omega_2\}$ ($\omega_1 < \omega_2$), with common prior $\Pr(\omega = \omega_1) = 1/2$. The principal's utility from the project is $g(\omega, e)$, increasing and differentiable in e. The agent's material utility from the project is $f(\omega, e)$, and his total utility is $f(\omega, e) - e$, where $f(\omega, e)$ is twice continuous and differentiable, and strictly increasing and concave in e. Assume also that $f_{\omega e}(\omega, e) > 0$ for all ω and e, and, for each $\omega \in \{\omega_1, \omega_2\}$, $f_e(\omega, e) = 1$ has a (finite) solution $e(\omega) > 0$.

When does the policy of full disclosure of the state benefit the principal, compared to the policy of no disclosure? Although this question ignores other persuasion policies, answering it will provide intuition for our general result in the next section. Let us have a look at the figure below, where $\overline{f(\omega, e)}$ denotes the average $f(\omega, e)$.

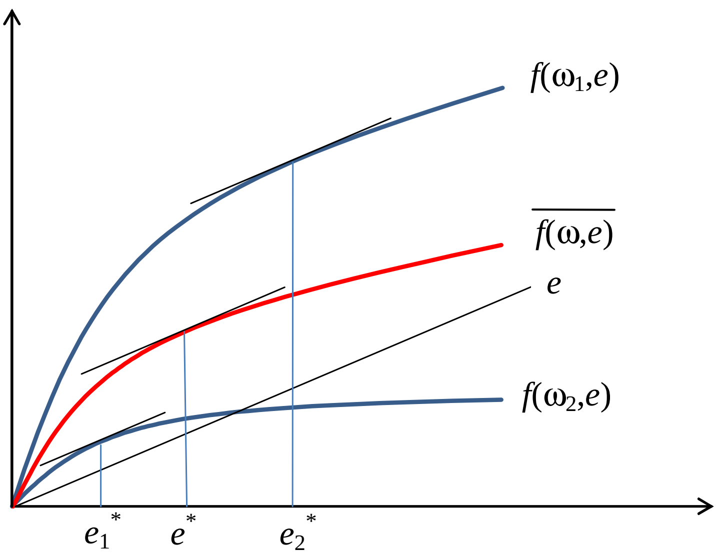

Figure 1.

Under pooling (no disclosure), the agent chooses $e = e^*$ so that

$$\frac{1}{2}f_e(\omega_1, e^*) + \frac{1}{2}f_e(\omega_2, e^*) = 1. \tag{1}$$

Under full disclosure, the choice of effort, e_i^*, in state ω_i solves $f_e(\omega_i, e_i^*) = 1$. Consider a switch from pooling to full disclosure. The effort rises from e^* to e_2^* in state ω_2 and falls from e^* to e_1^* in state ω_1. So, given that the states are equally likely, if the increase in the principal's utility in the high state exceeds its decrease in the low state, the principal benefits from transparency.

Now, since $f(\omega, e)$ is strictly concave in e, $f_e(\omega, e)$ is monotonic in e. Therefore, instead of looking at the change of e from e^* to e_i^* in state ω_i, we can equivalently consider the change of $f_e(\omega_i, e)$ from $f_e(\omega_i, e^*)$ to $f_e(\omega_i, e_i^*)$. Notice that $f_e(\omega_1, e_1^*) = f_e(\omega_2, e_2^*) = 1$. Furthermore, from equation (1) it follows that $f_e(\omega_1, e^*) = 1 - a$ and $f_e(\omega_2, e^*) = 1 + a$ for some constant a. Then, the increase in the principal's utility in state ω_2 can be represented as

$$\int_1^{1+a} -\frac{dg(\omega_2, e)}{df_e(\omega_2, e)} df_e(\omega_2, e),$$

and its fall in state ω_1 as

$$\int_{1-a}^{1} -\frac{dg(\omega_1, e)}{df_e(\omega_1, e)} df_e(\omega_1, e),$$

Since the length of the interval of integration is the same in the two integrals, a sufficient condition for the increase to exceed the fall is that

$$-\frac{dg(\omega_2, e_2)}{df_e(\omega_2, e_2)} > -\frac{dg(\omega_1, e_1)}{df_e(\omega_1, e_1)} \text{ for each } e_1 \in [e_1^*, e^*),\ e_2 \in [e^*, e_2^*),$$

which can also be written as

$$-\frac{g_e(\omega_2, e_2)}{f_{ee}(\omega_2, e_2)} > -\frac{g_e(\omega_1, e_1)}{f_{ee}(\omega_1, e_1)} \text{ for each } e_1 \in [e_1^*, e^*),\ e_2 \in [e^*, e_2^*)$$

(we show the equivalence of the two formulations in Section 3.1).

Intuitively, the gain from transparency is determined by two things. First, how large is the increase in the agent's effort in the high state relative to its decrease in the low state, once we move from pooling to transparency? This is determined by the speed at which the agent's marginal utility changes in the low and high states, when the agent re-adjusts his effort in response to learning the state. Hence, we have $f_{ee}(\omega_1, e_1)$ for $e_1 \in [e_1^*, e^*)$ and $f_{ee}(\omega_2, e_2)$ for $e_2 \in [e^*, e_2^*)$ in the formula. Second, what also matters is the change in the principal's utility per unit of effort in the each state. Hence, we have $g_e(\omega_1, e_1)$ and $g_e(\omega_2, e_2)$.

The next section extends the logic of this example to an arbitrary message with a non-singleton support in a more general model.

3 General model

There are a principal (she) and an agent (he). The agent takes a non-contractible action $a \in \mathbb{R}^+$, which generates a disutility $e = e(a)$ for himself, with function $e(\cdot)$ being strictly increasing. As there is one-to-one correspondence between a and e, we will work with e and call it "effort" (having in mind "disutility of effort").

There is a state of the world $\omega \in \Omega = [0, 1]$ with common prior $p \in \Delta(\Omega)$. Effort and state jointly determine the agent's material utility $f(\omega, e)$. For brevity, we will call $f(\omega, e)$ simply "agent's utility". His total utility is $f(\omega, e) - e$ — we are assuming separability in the two components.

We assume that $f(\omega, e)$ is twice continuous and differentiable, and strictly concave in e. We also assume that, for each ω, $f_e(\omega, e) = 1$ has a (finite) solution $e(\omega)$ in the set of admissible values for e.

The principal's utility is $g(\omega, e)$, and we assume it to be differentiable in e. Until Section 4, we abstract from where $f(\omega, e)$ and $g(\omega, e)$ originate from.

Before learning the state, the principal can commit to an information structure, whereby the agent receives some information about the state before choosing the effort. Formally, following the standard Bayesian persuasion framework, the principal commits to a mapping from the set of states Ω to distributions over messages $m \in \mathbb{R}$ that are sent to the agent. The information structure chosen by the principal is common knowledge. The goal of the principal is to select an information structure that maximizes her expected utility.

After hearing message m, the agent solves

$$\max_e \{E(f(\omega, e)|m) - e\}$$

Due to our assumptions on $f(\omega, e)$, the optimal effort chosen by the agent is unique under every posterior belief about the state, and it changes continuously in the posterior belief. With this, the persuasion problem of the principal is well-defined and has a solution.

We are going to find a sufficient condition for the optimality of full disclosure. Commitment to full disclosure is typically easier to achieve than to "garbling" information structures; for instance, it may be achieved with transparency policies. Note also that the state may just represent whatever information is available in possession of the principal/organization, about a "more primitive" state that influences the output.

We call "full disclosure" the information structure where the agent learns the state. We are going to prove optimality of full disclosure by showing that any alternative information structure is suboptimal: a "vague" message about the state can be improved upon by being fractioned into smaller messages.

3.1 Sufficient conditions for full disclosure

Consider the following condition on f and g:

$$\text{For all } e_1,\ e_2,\ \omega_1,\ \omega_2,\ \begin{cases} e_1 < e_2 \\ f_e(\omega_1, e_1) < f_e(\omega_2, e_2) \end{cases} \Rightarrow \frac{g_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{g_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)}\ . \quad (2)$$

The ratio $-g_e(\omega_1, e_1)/f_{ee}(\omega_1, e_1)$ measures how much the utility of the principal increases, as the agent increases effort, per unit of agent's marginal utility of effort that is lost in the process. To see this, let $y = f_e(\omega, e)$ and compute the derivative of g with respect to changes of the agent's marginal utility, holding ω fixed:

$$\frac{\partial g(\omega, f_e^{-1}(\omega, y))}{\partial y} = g_e(\omega, f_e^{-1}(\omega, y))\frac{\partial f_e^{-1}(\omega, y)}{\partial y} = \frac{g_e(\omega, f_e^{-1}(\omega, y))}{f_{ee}(\omega, f_e^{-1}(\omega, y))} = \frac{g_e(\omega, e)}{f_{ee}(\omega, e)}.$$

Condition (2) essentially requires that this measure of the principal's return from the agent's effort improves as we move towards a state that generates stronger incentives for the agent. More precisely, the marginal increase of the principal's utility induced by a marginal decrease of the agent's marginal utility (which entails an increase of effort) must be higher under the state-effort pairs that entail higher effort and yet higher marginal utility of effort for the agent. This guarantees that revealing "good news" and "bad news" is better than pooling them: under the news that correspond to a higher marginal utility of effort for the agent, as he increases effort to readjust his marginal utility, the utility gain for the principal in the process exceeds the utility loss of the symmetric process under the alternative news. So, condition (2) is sufficient for optimality of full disclosure.

Theorem 1 *Under condition (2), full disclosure maximizes the principal's expected utility.*

Proof of Theorem 1 We show that transparency is optimal for the principal by showing that any message m^* that generates a posterior with support Ω^* is either equivalent to revealing the states in the support, or suboptimal; then transparency is optimal because, as already argued, an optimal solution exists.

For each $\omega \in \Omega$ and $x \in \mathbb{R}^+$, let $f_e^{-1}(\omega, x)$ denote the effort e such that $f_e(\omega, e) = x$, if any. Let $\widetilde{f}(m, e)$ denote the agent's expected material utility under message m and effort e, $E(f(\omega, e)|m)$. Let e^* denote the agent's optimal effort upon receiving m^*. It is obtained by solving the first-order condition $\widetilde{f}_e(m^*, e) = 1$.

If it is not true that revealing all states in Ω^* induces effort e^* for each of the states, then there exist Borel subsets Ω_1^*, Ω_2^* of Ω^* such that $p(\Omega_1^*)p(\Omega_2^*) > 0$ and

$$\overline{x}_1 := \sup_{\omega_1 \in \Omega_1^*} f_e(\omega_1, e^*) < 1 < \inf_{\omega_2 \in \Omega_2^*} f(\omega_2, e^*) =: \underline{x}_2, \tag{3}$$

and additionally, for all $\alpha \in [0,1]$, $(\omega_1, \omega_2) \in \Omega_1^* \times \Omega_2^*$, and $(\omega_1', \omega_2') \in \Omega_1^* \times \Omega_2^*$[3]

$$f_e(\omega_2, f_e^{-1}(\omega_2', 1 + \alpha\,[\underline{x}_2 - 1])) - f_e(\omega_1, f_e^{-1}(\omega_1', \overline{x}_1 + \alpha\,[1 - \overline{x}_1])) > 0. \tag{4}$$

To see that such Ω_1^*, Ω_2^* exist, construct them as follows. Start with $\Omega_k^* = \{\omega_k^*\}$ $(k = 1,2)$ for some $\omega_1^*, \omega_2^* \in \Omega^*$ that satisfy (3). Then, (4) becomes

$$[1 + \alpha\,(\underline{x}_2 - 1)] - [\overline{x}_1 + \alpha\,(1 - \overline{x}_1)] > 0.$$

This inequality always holds as the left-hand side is equal to $1 - \overline{x}_1 + \alpha(\underline{x}_2 - \overline{x}_1) > 0$. Hence, (4) is preserved by continuity when Ω_1^*, Ω_2^* are sufficiently small neighbourhoods of ω_1^*, ω_2^*. Such neighbourhoods have positive prior probability because they have positive posterior probability by definition of support.

Now we decompose m^* into three messages: m_1, m_2 that induce posteriors with supports Ω_1^*, Ω_2^*, and a complementary message $\widetilde{m}$ that induces effort e^* like m^*. That is, $\Pr(m_1|m^*) + \Pr(m_2|m^*) + \Pr(\widetilde{m}|m^*) = 1$, $\widetilde{f}_e(\widetilde{m}, e^*) = 1$, $\widetilde{f}_e(m_1, e^*) < 1$, $\widetilde{f}_e(m_2, e^*) > 1$. Such messages can be constructed because starting from m^* and taking away probability from Ω_1^* and Ω_2^* has opposite effects on the marginal expected utility by (3).[4]

To simplify notation, let us drop the conditioning notation and denote the above conditional probabilities simply by $\Pr(m_1)$, $\Pr(m_2)$, $\Pr(\widetilde{m})$. Since m^* induces effort e^*, we must have

$$\Pr(m_1) \cdot \widetilde{f}_e(m_1, e^*) + \Pr(m_2) \cdot \widetilde{f}_e(m_2, e^*) + \Pr(\widetilde{m}) \cdot \widetilde{f}_e(\widetilde{m}, e^*) = 1,$$

Since message $\widetilde{m}$ induces effort e^* as well, $\widetilde{f}_e(\widetilde{m}, e^*) = 1$, hence

$$\frac{\Pr(m_1) \cdot \widetilde{f}_e(m_1, e^*) + \Pr(m_2) \cdot \widetilde{f}_e(m_2, e^*)}{\Pr(m_1) + \Pr(m_2)} = 1,$$

[3] $f_e^{-1}(\omega_2, 1 + \alpha\,[\underline{x}_2 - 1])$ is well-defined because, by the existence of an optimal effort, by (3), and by continuity of f_e, $f_e(\omega_2, e)$ takes all the values between 1 and $1 + \alpha\,[\underline{x}_2 - 1]$; likewise for $f_e^{-1}(\omega_1, \overline{x}_1 + \alpha\,[1 - \overline{x}_1])$.

[4] Notice that we assume that the three messages are drawn after that m^* has been drawn. This is not important; we could equally assume that the messages are independent of m^* conditional on the state, we would only need to make sure that $\Pr(m_1|\omega) + \Pr(m_2|\omega) + \Pr(\widetilde{m}|\omega) = \Pr(m^*|\omega)$ for any ω.

which can be rewritten as

$$\Pr(m_1)\left[1-\widetilde{f}_e(m_1,e^*)\right] = \Pr(m_2)\left[\widetilde{f}_e(m_2,e^*)-1\right]. \tag{5}$$

Call e_1^*, e_2^* the agent's optimal efforts under m_1, m_2, i.e, $\widetilde{f}_e(m_1,e_1^*) = \widetilde{f}_e(m_2,e_2^*) = 1$. Notice that $e_1^* < e^* < e_2^*$, as $\widetilde{f}_e(m_1,e^*) < 1 < \widetilde{f}_e(m_2,e^*)$ and f is strictly concave in e. We show that the principal's expected utility increases after the decomposition, that is

$$\begin{aligned} & \Pr(m_1)\widetilde{g}(m_1,e^*) + \Pr(m_2)\widetilde{g}(m_2,e^*) + \Pr(\widetilde{m})\widetilde{g}(\widetilde{m},e^*) \\ < \ & \Pr(m_1)\widetilde{g}(m_1,e_1^*) + \Pr(m_2)\widetilde{g}(m_2,e_2^*) + \Pr(\widetilde{m})\widetilde{g}(\widetilde{m},e^*). \end{aligned}$$

Rewrite this as

$$\Pr(m_1)\left[\widetilde{g}(m_1,e^*) - \widetilde{g}(m_1,e_1^*)\right] < \Pr(m_2)\left[\widetilde{g}(m_2,e_2^*) - \widetilde{g}(m_2,e^*)\right], \tag{6}$$

and then as

$$\Pr(m_1)\int_{e_1^*}^{e^*} \widetilde{g}_e(m_1,e)de < \Pr(m_2)\int_{e^*}^{e_2^*} \widetilde{g}_e(m_2,e)de.$$

For each $k=1,2$, let $x_k = \widetilde{f}_e(m_k,e)$, so that $dx_k = \widetilde{f}_{ee}(m_k,e)de$ and $e = \widetilde{f}_e^{-1}(m_k,x_k)$,[5] and operate a change of variable (noticing that $\widetilde{f}_e(m_1,e_1^*) = \widetilde{f}_e(m_2,e_2^*) = 1$):

$$\begin{aligned} & -\Pr(m_1)\int_{\widetilde{f}_e(m_1,e^*)}^{1} \widetilde{g}_e(m_1,\widetilde{f}_e^{-1}(m_1,x_1)) \cdot \frac{1}{\widetilde{f}_{ee}\left(m_1,\widetilde{f}_e^{-1}(m_1,x_1)\right)}dx_1 \\ < \ & -\Pr(m_2)\int_{1}^{\widetilde{f}_e(m_2,e^*)} \widetilde{g}_e(m_2,\widetilde{f}_e^{-1}(m_2,x_2)) \cdot \frac{1}{\widetilde{f}_{ee}\left(m_2,\widetilde{f}_e^{-1}(m_2,x_2)\right)}dx_2. \end{aligned} \tag{7}$$

Letting $x_1 = \widetilde{f}_e(m_1,e^*) + \alpha\left[1-\widetilde{f}_e(m_1,e^*)\right]$ and $x_2 = 1+\alpha\left[\widetilde{f}_e(m_2,e^*)-1\right]$, inequality (7) can be rewritten as

$$\begin{aligned} & \int_0^1 \frac{\widetilde{g}_e(m_1,\widetilde{f}_e^{-1}(m_1,x_1))}{-\widetilde{f}_{ee}\left(m_1,\widetilde{f}_e^{-1}(m_1,x_1)\right)}\Pr(m_1)\left[1-\widetilde{f}_e(m_1,e^*)\right]d\alpha \\ < \ & \int_0^1 \frac{\widetilde{g}_e(m_2,\widetilde{f}_e^{-1}(m_2,x_2))}{-\widetilde{f}_{ee}\left(m_2,\widetilde{f}_e^{-1}(m_2,x_2)\right)}\Pr(m_2)\left[\widetilde{f}_e(m_2,e^*)-1\right]d\alpha. \end{aligned}$$

[5] $\widetilde{f}_e^{-1}(m_k,x_k)$ is defined as the value of e such that $\widetilde{f}_e(m_k,e) = x_k$.

Thus, by (5), a sufficient condition for (7) to hold is that, for each $\alpha \in [0,1]$,

$$\frac{\widetilde{g}_e(m_1, \widetilde{f}_e^{-1}(m_1, x_1))}{-\widetilde{f}_{ee}\left(m_1, \widetilde{f}_e^{-1}(m_1, x_1)\right)} < \frac{\widetilde{g}_e(m_2, \widetilde{f}_e^{-1}(m_2, x_2))}{-\widetilde{f}_{ee}\left(m_2, \widetilde{f}_e^{-1}(m_2, x_2)\right)}. \tag{8}$$

Call $e_1 = \widetilde{f}_e^{-1}(m_1, x_1)$ and $e_2 = \widetilde{f}_e^{-1}(m_2, x_2)$. Note that, by the definition of $\overline{x}_1$ and $\underline{x}_2$ and the fact that the supports of m_1 and m_2 are Ω_1^* and Ω_2^*, we have $\widetilde{f}_e(m_1, e^*) \leq \overline{x}_1$ and $\widetilde{f}_e(m_2, e^*) \geq \underline{x}_2$. Therefore $x_1 \leq \overline{x}_1 + \alpha\,[1 - \overline{x}_1]$ and $x_2 \geq 1 + \alpha\,[\underline{x}_2 - 1]$, which imply

$$e_1 \geq \widetilde{f}_e^{-1}(m_1, \overline{x}_1 + \alpha\,[1 - \overline{x}_1]), \tag{9}$$

$$e_2 \leq \widetilde{f}_e^{-1}(m_2, 1 + \alpha\,[\underline{x}_2 - 1]). \tag{10}$$

Inequality (8) can be rewritten as

$$\frac{\widetilde{f}_{ee}\,(m_2, e_2)}{\widetilde{f}_{ee}\,(m_1, e_1)} < \frac{\widetilde{g}_e(m_2, e_2)}{\widetilde{g}_e(m_1, e_1)}. \tag{11}$$

For each $\omega_1 \in \Omega_1^*$ and $\omega_2 \in \Omega_2^*$, we have

$$\begin{gathered} f_e(\omega_1, e_1) \leq f_e(\omega_1, \widetilde{f}_e^{-1}(m_1, \overline{x}_1 + \alpha\,[1 - \overline{x}_1])) < \\ f_e(\omega_2, \widetilde{f}_e^{-1}(m_2, 1 + \alpha\,[\underline{x}_2 - 1])) \leq f_e(\omega_2, e_2), \end{gathered}$$

where the first and third inequalities are by (9) and (10), and the second inequality is by (4). Notice also that $e_1 < e_2$, as even when x_1 takes the minimum value, $\widetilde{f}_e(m_1, e^*)$, and x_2 takes the maximum value, $\widetilde{f}_e(m_2, e^*)$, $e_1 = \widetilde{f}_e^{-1}(m_1, \widetilde{f}_e(m_1, e^*)) = \widetilde{f}_e^{-1}(m_2, \widetilde{f}_e(m_2, e^*)) = e_2$. However, it is impossible that $x_1 = \widetilde{f}_e(m_1, e^*)$ and $x_2 = \widetilde{f}_e(m_2, e^*)$ at the same time. Hence, if (2) is satisfied, then

$$\frac{g_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{g_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)},$$

or

$$\frac{f_{ee}(\omega_2, e_2)}{f_{ee}(\omega_1, e_1)} < \frac{g_e(\omega_2, e_2)}{g_e(\omega_1, e_1)}.$$

Since this is true for all $\omega_1 \in \Omega_1^*$ and $\omega_2 \in \Omega_2^*$, (11) holds. $\blacksquare$

Condition (2) can be expressed in terms of just derivatives of f and g. To see this, notice that it is equivalent to stating that, at each (ω, e), $-g_e(\omega, e)/f_{ee}(\omega, e)$ is increasing in all directions in which both e and $f_e(\omega, e)$ increase. So, by applying directional derivatives, one can show the lemma below. Namely, consider the following conditions:

$$\text{For each } (\omega, e) \text{ s.t. } f_{e\omega} > 0, \quad \begin{cases} f_{ee\omega} g_e \geq g_{e\omega} f_{ee} \\ g_e(f_{eee} f_{e\omega} - f_{ee\omega} f_{ee}) \geq f_{ee}(g_{ee} f_{e\omega} - g_{e\omega} f_{ee}) \end{cases}, \tag{12}$$
$$\text{with at least one inequality being strict}$$

and

$$\text{For each } (\omega, e) \text{ s.t. } f_{e\omega} < 0, \quad \begin{cases} f_{ee\omega} g_e \leq g_{e\omega} f_{ee} \\ g_e(f_{eee} f_{e\omega} - f_{ee\omega} f_{ee}) \leq f_{ee}(g_{ee} f_{e\omega} - g_{e\omega} f_{ee}) \end{cases}, \tag{13}$$
$$\text{with at least one inequality being strict}$$

Lemma 1 *Condition (2) is equivalent to (12) and (13).*

Proof. See the Appendix. ■

Notice that (12) and (13) do not cover the case $f_{e\omega} = 0$. This is because, when $f_{e\omega} = 0$, there is simply no direction in which both e and f_e increase.

3.2 Sufficient condition for suboptimality of full disclosure

The previous subsection delivered a sufficient condition for the optimality of transparency. It turns out that we can apply almost the same scheme of reasoning as above to derive a sufficient condition for *sub*optimality of transparency. The difference is that instead of the existence of a welfare-improving split for any message with a non-sigleton support, the suboptimality of transparency requires the existence of at least one pair of states that can be pooled (or partially pooled) so as to improve the principal's welfare.

Namely, fix a pair of states ω_1, ω_2 and consider the following condition

$$\text{For all } e_1,\ e_2, \begin{cases} e_1 < e_2 \\ f_e(\omega_1, e_1) < f_e(\omega_2, e_2) \end{cases} \Rightarrow \frac{g_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} > \frac{g_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)} \tag{14}$$

This condition resembles (2) except that it is formulated for given ω_1 and ω_2 and the sign of the inequality between the ratios flips.

Theorem 2 *Suppose there exist* $e_1,\ e_2,\ \omega_1,\ \omega_2$ *such that* $e_1 < e_2$ *and* $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$. *Then, if there exists a pair of states* $\omega_1,\ \omega_2$, *such that (14) holds, full disclosure is suboptimal.*

Proof. See the Appendix. ■

Notice that the condition of Theorem 2 is not the negation of the condition of Theorem 1. That is, Theorem 2 does not imply that Theorem 1 delivers a necessary and sufficient condition for the optimality of transparency. There may be a pair of states under which the relation between the ratios does not keep the same sign for all $e_1,\ e_2$ such that $e_1 < e_2$ and $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$. Also, of course, it may be the case that for no pair of states there exist e_1 and e_2 such that $e_1 < e_2$ and $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$. The simplest example of such a situation is when $f_e(\omega, e)$ does not depend on ω, as the concavity of f with respect to e implies that $f_e(\omega, e)$ is decreasing in e.

To understand Theorem 2, one can recall the example of Section 2. There, a separation of ω_1 and ω_2 was optimal if

$$-\frac{g_e(\omega_1, e_1)}{f_{ee}(\omega_1, e_1)} < -\frac{g_e(\omega_2, e_2)}{f_{ee}(\omega_2, e_2)} \text{ for each } e_1 \in [e_1^*, e^*),\ e_2 \in [e^*, e_2^*)$$

Clearly, a pooling of ω_1 and ω_2 would be optimal if the relation between the ratios were just flipped

$$-\frac{g_e(\omega_1, e_1)}{f_{ee}(\omega_1, e_1)} > -\frac{g_e(\omega_2, e_2)}{f_{ee}(\omega_2, e_2)} \text{ for each } e_1 \in [e_1^*, e^*),\ e_2 \in [e^*, e_2^*)$$

For each $e_1 \in [e_1^*, e^*),\ e_2 \in [e^*, e_2^*)$, $e_1 < e_2$, and also $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$ (as $f_e(\omega_1, e_1^*) = f_e(\omega_2, e_2^*) = 1$). Hence, if (14) holds, pooling ω_1 and ω_2 is optimal.

3.3 Special case

Assume the principal's utility is an increasing and weakly convex transformation of the agent's: $g(\omega, e) = h(f(\omega, e))$, where h is a weakly convex function. That means that the agent's utility is weakly concave in that of the principal. Such a feature naturally arises in settings when the effort is exerted to generate some monetary output (profit), the agent is more risk averse than the principal and receives a compensation that is not too convex (e.g., linear). Assume also that $f_{e\omega} > 0$ and $f(\omega, 0)$ is weakly increasing in ω:

that is, effort and state are complements for the agent and $f(\omega, e)$ is (weakly) increasing in ω for any given e.

Condition

$$\frac{g_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{g_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)} \tag{15}$$

from (2) can be written as

$$\frac{h'(f(\omega_1, e_1)) \cdot f_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{h'(f(\omega_2, e_2)) \cdot f_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)}$$

Due to the assumptions of this section, $e_1 < e_2$ and $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$ together imply $f(\omega_1, e_1) < f(\omega_2, e_2)$. Consequently, $h'(f(\omega_1, e_1)) \leq h'(f(\omega_2, e_2))$, and

$$\frac{f_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{f_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)} \tag{16}$$

implies (15). Hence, we can formulate the following sufficient condition for the optimality of transparency

$$\text{For all } e_1,\ e_2,\ \omega_1,\ \omega_2,\ \left\{ \begin{array}{c} e_1 < e_2 \\ f_e(\omega_1, e_1) < f_e(\omega_2, e_2) \end{array} \right. \Rightarrow \frac{f_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} < \frac{f_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)}\ , \tag{17}$$

and state the following result:

Theorem 3 *Suppose the principal's utility is a weakly convex transformation of the agent's material utility, the agent's utility is (weakly) increasing in ω for a given e, and e and ω are (weak) complements for the agent. Then, under condition (17), full disclosure maximizes the principal's expected utility.*

Under the assumptions of this section (summarized in the theorem), (16) is stronger than (15). Hence, our new condition is stronger than (2). Its advantage is that one does not need to know how exactly the principal's utility looks like. It is sufficient to know that it is a convex transformation of the agent's utility, that effort and state are (weak) complements for the agent, and that $f(\omega, e)$ is (weakly) increasing in ω for any given e.

Condition (17) can also be expressed through the equivalent "derivative conditions":

Lemma 2 *Under the assumption that $f_{e\omega} > 0$, condition (17) is equivalent to the fol-*

lowing condition:

$$\textit{For each } (\omega, e), \quad \begin{cases} f_{ee\omega} f_e \geq f_{e\omega} f_{ee} \\ f_{eee} f_{e\omega} \geq f_{ee} f_{ee\omega} \end{cases}, \tag{18}$$

where at least one of the inequalities has to be strict.

Proof. See the Appendix. ∎

4 Discussion and Examples

Mensch (2021) points at the role of complementarities between state and action. He argues that it is important that complementarities for the sender and the receiver "point in the same direction" for transparency to be optimal.[6] Indirectly, the interaction between effort and state plays a role in our condition as well, because it affects whether $g_e(\omega, e)$ comoves with $f_e(\omega, e)$ when both e and $f_e(\omega, e)$ increase.

Specifically, when effort and state are complementary for the agent, higher $f_e(\omega, e)$ together with higher e imply higher ω, meaning that $\omega_2 > \omega_1$ in (2). Then, if effort and state are complementary for the principal as well, higher ω pushes $g_e(\omega, e)$ upward for given e, thereby relaxing (2). In contrast, if effort and state are substitutes for the principal, higher ω pushes $g_e(\omega, e)$ downward for given e, thereby tightening (2).[7] By similar logic, if effort and state are substitutes for the agent, (2) is more (less) likely to be satisfied when they are substitutes (complements) for the principal.

Examples below are aimed to shed more light on the role of the interaction between effort and state as well as on the effects of the agent's and principal's risk-aversion. The examples will also help us to grasp whether we shall expect (2) to be satisfied in typical economic applications.

In all examples of this section we consider the following classical setting: An agent produces an output which he splits with a principal. Namely, there is output $y(\omega, e)$, the agent's and the principal's utilities of money are (weakly) concave functions $u(\cdot)$ and $v(\cdot)$ respectively, the agent receives wage $w(y)$, and the principal receives $y - w(y)$. For the

[6] Instead of "complements" versus "substitutes", the author speaks about "direction of complementarity". Essentially, complemetarities pointing in "different directions" in Mensch's terminology means that action and state are complements for one party and substitutes for the other.

[7] To avoid confusion, complemetarity for both parties does *not* guarantee the optimality of transparency. In a specific example of subsection 5.2, though state and effort are complements for both parties, condition (23) must be satisfied for transparency to be optimal.

sake of simplicity, we assume that the wage is linear, that is, the agent receives a fixed share δ of the output. Whereas we take the compensation scheme for the agent as given, the conclusions about the optimality of transparency will not depend on δ, as we will see. However, allowing for a non-linear wage schedule and solving for the optimal wage schedule and disclosure policy jointly could be an interesting avenue for future research.

4.1 Effects of risk aversion in a simple setting

Consider the following setting:

$$\begin{aligned} y(\omega, e) &= \omega\sqrt{e},\ w(y) = \delta y \\ u(x) &= \frac{x^{1-\gamma}}{1-\gamma},\ v(x) = \frac{x^{1-\rho}}{1-\rho} \end{aligned}$$

That is, both the agent and the principal exhibit CRRA with coefficients γ and ρ respectively. We can compute:

$$\begin{aligned} f(\omega, e) &= \frac{1}{1-\gamma}(\delta\omega)^{1-\gamma}e^{\frac{1}{2}(1-\gamma)},\ f_e(\omega, e) = \frac{1}{2}(\delta\omega)^{1-\gamma}e^{-\frac{1}{2}-\frac{1}{2}\gamma},\ f_{ee}(\omega, e) = -(1+\gamma)\frac{1}{4}(\delta\omega)^{1-\gamma}e^{-\frac{3}{2}-\frac{1}{2}\gamma} \\ g(\omega, e) &= \frac{1}{1-\rho}((1-\delta)\omega)^{1-\rho}e^{\frac{1}{2}(1-\rho)},\ g_e(\omega, e) = \frac{1}{2}((1-\delta)\omega)^{1-\rho}e^{-\frac{1}{2}-\frac{1}{2}\rho} \end{aligned}$$

With some algebra, one can then derive

$$\frac{g_e(\omega, e)}{-f_{ee}(\omega, e)} = const \cdot (f_e(\omega, e))^{\frac{\gamma-\rho}{1-\gamma}} \cdot e^{\frac{1-\rho}{1-\gamma}},$$

where *const* is a positive constant.

Consider first the case when $\gamma < 1$. Notice, first of all, that there exists a direction in which both f_e and e go up: For any increase in e, ω can be increased so that f_e rises as well.

If $\rho < \gamma$, then the ratio is increasing in both f_e and e. Hence, (2) holds, and transparency is optimal.

If $\rho \in (\gamma, 1)$, then the ratio is decreasing in f_e and increasing in e. Neither the sufficient condition for transparency nor the sufficient condition for suboptimality of transparency is satisfied. Hence, our analysis is inconclusive in this case.

If $\rho > 1$, then the ratio is decreasing in both f_e and e. According to Theorem 2, transparency is thus suboptimal.

Consider now the case when $\gamma > 1$. Here again there exists a direction in which both f_e and e go up (now it necessitates a sufficient decrease in ω).

If $\rho > \gamma$, then the ratio is increasing in both f_e and e. Hence, (2) holds, and transparency is optimal.

If $\rho \in (1, \gamma)$, then the ratio is decreasing in f_e and increasing in e. Neither the sufficient condition for transparency nor the sufficient condition for suboptimality of transparency is satisfied. Hence, our analysis is inconclusive in this case.

If $\rho < 1$, then the ratio is decreasing in both f_e and e. According to Theorem 2, transparency is thus suboptimal.

We can notice that transparency fails to be optimal when ρ and γ are on the opposite sides from 1. This is related to the fact that, in this case, state end effort are complements for one party and substitutes for the other, which can be seen by examining the expressions for $f_e(\omega, e)$ and $g_e(\omega, e)$. In contrast, when ρ and γ are both smaller or both greater than 1, the direction of interaction between state and effort is the same for both parties, and, thus, transparency gets a chance.

For example, consider a typical textbook situation with a risk neutral principal ($\rho = 0$) and a risk averse agent. If the agent is not too risk averse ($\gamma < 1$), transparency is optimal. Since state and effort are complements for both parties, the principal benefits more from effort exactly when the agent has higher incentives to exert effort.

Instead, when the agent becomes too risk averse ($\gamma > 1$), while the principal remains risk neutral, state and effort turn substitutes for the agent. As a result, good news about productivity *depress* effort, while the principal still benefits more from effort in higher states. As a result, transparency ceases to be optimal.

When the principal is highly risk averse ($\rho > 1$) the story is reversed: now *insufficient* risk aversion of the agent ($\gamma < 1$) becomes detrimental to the optimality of transparency. This is because now the principal benefits more from effort under *lower* states, while for the agent state and effort are complements. One needs to make the agent sufficiently risk averse ($\gamma > 1$) to align the interaction of effort and state between the two parties, so that transparency can be optimal.

What is interesting about the case of a highly risk averse principal is that transparency can be optimal despite lowering the expected effort and can be harmful despite raising the expected effort. Indeed, one can easily derive that disclosure of states in the support of any given message increases the expected effort under $\gamma < 1$ and lowers it under $\gamma > 1$. This observation demonstrates that an increase (decrease) in the average effort due to disclosure is not sufficient to make transparency optimal (suboptimal), as the direction and strength of the interaction between state and effort in the principal's payoff matters too.

4.2 Risk neutral agent and principal, "multiplicatively-additively" separable production function

In the previous example we assumed a simple production function and played with risk aversion of the parties. Let us now simplify the preferences of both parties by assuming that their material utilities are linear in output and focus instead on the properties of the production function. Linearity in output for both parties would arise, for example, in a setting where both parties are risk neutral and the wage is linear in output. Risk neutrality is often assumed in applications, so it is interesting too see what it takes to make transparency optimal in such a simple framework.

The material utilities of the agent and the principal under these assumptions are (up to affine transformations) δy and $(1-\delta)y$, respectively. Then, (2) becomes simply

$$\text{For all } e_1,\ e_2,\ \omega_1,\ \omega_2,\ \left\{\begin{array}{c} e_1 < e_2 \\ y_e(\omega_1, e_1) < y_e(\omega_2, e_2) \end{array}\right. \Rightarrow \frac{y_e(\omega_1, e_1)}{-y_{ee}(\omega_1, e_1)} < \frac{y_e(\omega_2, e_2)}{-y_{ee}(\omega_2, e_2)}. \quad (19)$$

Consider output functions of the following form:

$$y(\omega, e) = \alpha(\omega)\varphi(e) + \beta(\omega) + \xi(e), \quad (20)$$

with $\alpha(\cdot) > 0,\ \alpha'(\cdot) > 0,\ \varphi(\cdot) > 0,\ \varphi'(\cdot) > 0,\ \varphi''(\cdot) < 0,\ \xi'(\cdot) \geq 0,\ \xi''(\cdot) \leq 0$. This output function can be called "multiplicatively-additively" separable in state and effort; we will call it just "separable", for simplicity. Special cases of this form (such as $\omega\varphi(e)$ employed in the previous subsection) are commonly used in the literature.

Notice that (19) is the same as (17) but with y instead of f. Furthermore, due to our assumption on $\alpha'(\cdot)$ and $\varphi'(\cdot)$, state and effort are complements. Thus, we can apply

Lemma 2 to argue that (19) is equivalent to (18) with y instead of f. Then, given (20), one can easily derive that the corresponding derivative conditions become:

$$\begin{cases} \varphi''(e)\xi'(e) \geq \varphi'(e)\xi''(e) \\ \alpha(\omega)\left[\varphi'''(e)\varphi'(e) - (\varphi''(e))^2\right] \geq \xi''(e)\varphi''(e) - \xi'''(e)\varphi'(e) \end{cases}, \tag{21}$$

where at least one inequality has to be strict.

Notice first that $\beta(\omega)$ does not enter the conditions in any way. This is intuitive: As $\beta(\omega)$ has no effect on effort, it is irrelevant for the disclosure policy.

Let us now examine some commonly used functional forms. To begin with, assume that $\xi(\cdot)$ is a constant. Then, the first condition holds as equality, and the second one becomes $\varphi'''(e)\varphi'(e) \geq (\varphi''(e))^2$. Then, if $\varphi(\cdot)$ is a concave power function, $\varphi(e) = ae^s$ with $a > 0,\ s \in (0,1)$, it is straightforward to check that the second condition is satisfied.

Assume now both $\varphi(\cdot)$ and $\xi(\cdot)$ are concave power functions: $\varphi(e) = ae^s,\ \xi(e) = be^t$, with $a > 0,\ b > 0,\ s \in (0,1),\ t \in (0,1)$. It is straightforward to derive that the first condition boils down to $s \geq t$, and the second one in (21) always holds as a strict inequality whenever $s \geq t$. Thus, $s \geq t$ becomes a sufficient condition for the optimality of transparency.

As another example, let us assume that $\varphi(e) = s \cdot \ln e$ and $\xi(e) = t \cdot \ln e$. Then the first inequality holds as equality, and it can be easily derived that the second one is always satisfied as a strict inequality.

5 Necessary and sufficient conditions under a separable output and linear utilities

In the previous sections we provided sufficient conditions for the optimality of transparency. Can we hope that they are also necessary? In this section, we give examples when they are and when they are not. We stick to the specific setting of subsection 4.2. Hence, (21), with at least one inequality being strict, is a sufficient condition for the optimality of transparency.

It turns out that, in this setting, the principal's utility turn out to depend only on the expected $\alpha(\omega)$, and, hence, its convexity in α becomes the necessary and sufficient condition for the optimality of transparency (likewise, its concavity will be necessary and sufficient for the optimality of complete pooling).[8]

5.1 Jensen's inequality application

Consider a decomposition of message m^* into m_1 and m_2 (i.e., $\Pr(m_1|m^*)+\Pr(m_2|m^*) = 1$). Let

$$\begin{aligned} \alpha^* &: = E(\alpha(\omega)|m^*),\ \alpha_1 := E(\alpha(\omega)|m_1),\ \alpha_2 := E(\alpha(\omega)|m_2), \\ \beta^* &: = E(\beta(\omega)|m^*),\ \beta_1 := E(\beta(\omega)|m_1),\ \beta_2 := E(\beta(\omega)|m_2). \end{aligned}$$

Let A be the set of all values of $\alpha(\omega)$ as ω runs from 0 to 1.

The choice of effort will depend only on the expected $\alpha(\omega)$, given a message. Namely, for given m, $e^*(\alpha)$ solves

$$\frac{\partial E(\delta y(\omega, e)|m)}{\partial e} = 1,$$

or

$$\alpha\varphi'(e^*) + \xi'(e^*) = 1/\delta, \text{ where } \alpha \in \{\alpha^*, \alpha_1, \alpha_2\}.$$

Then, under the separation of m^* into m_1 and m_2, the principal's payoff, divided by $(1-\delta)$, conditionally on m^* is

$$\begin{aligned} & \Pr(m_1|m^*)\left[\alpha_1\varphi(e^*(\alpha_1)) + \beta_1 + \xi(e^*(\alpha_1))\right] \\ & + \Pr(m_2|m^*)\left[\alpha_2\varphi(e^*(\alpha_2)) + \beta_2 + \xi(e^*(\alpha_2))\right] \\ \equiv\ & \beta^* + \Pr(m_1|m^*)\left[\alpha_1\varphi(e^*(\alpha_1)) + \xi(e^*(\alpha_1))\right] \\ & + \Pr(m_2|m^*)\left[\alpha_2\varphi(e^*(\alpha_2)) + \xi(e^*(\alpha_2))\right] \end{aligned}$$

Under pooling, it is

$$\begin{aligned} & \Pr(m_1|m^*)\left[\alpha_1\varphi(e^*(\alpha^*)) + \beta_1 + \xi(e^*(\alpha^*))\right] \\ & + \Pr(m_2|m^*)\left[\alpha_1\varphi(e^*(\alpha^*)) + \beta_2 + \xi(e^*(\alpha^*))\right] \\ \equiv\ & \beta^* + \alpha^*\varphi(e^*(\alpha^*)) + \xi(e^*(\alpha^*)) \end{aligned}$$

Hence, given that $\alpha^* = \Pr(m_1|m^*)\alpha_1 + \Pr(m_2|m^*)\alpha_2$, the comparison of the two

[8] This result is a simple extension of the results derived in Dworczak and Martini (2019) obtained for a setting when the sender's utility depends only on the expected state.

payoffs boils down to checking if $\gamma(\alpha) \equiv \alpha\varphi(e(\alpha)) + \xi(e(\alpha))$ is convex or concave, where $e(\alpha)$ solves $\alpha\varphi'(e) + \xi'(e) = 1/\delta$, and applying Jensen's inequality. Namely, we can state the following lemma.

Lemma 3 *When the output has form (20) and the principal's and the agent's payoffs are linear in the output, full disclosure is optimal for the principal if and only if function* $\gamma(\alpha) \equiv \alpha\varphi(e(\alpha)) + \xi(e(\alpha))$, *where* $e(\alpha)$ *solves* $\alpha\varphi'(e) + \xi'(e) = 1/\delta$, *is convex for all* $\alpha \in A$.

Proof. The lemma immediately follows from Jensen's inequality: If the payoff is convex everywhere, any split of any message with a non-singleton support is beneficial. If it is concave somewhere, full transparency cannot be optimal, because pooling messages in the concavity region improves the expected payoff. ■

Notice that the lemma can obviously be extended to claiming that complete *pooling* is optimal if and only if $\gamma(\alpha)$ is concave. Notice also that the condition of Lemma 3 only depends on the functions of effort, and not of the state (i.e., on $\varphi(\cdot)$ and $\xi(\cdot)$, but not on $\alpha(\cdot)$ or $\beta(\cdot)$). At the same time, the second inequality in the sufficient condition (21) contains $\alpha(\omega)$. This hints that the (21) is generally only a sufficient condition but not necessary. The following example provides more detail on this matter.

5.2 The case when the separate term depending on effort is linear

For convenience, let $t := 1/\delta$. Using the expression for $\gamma(\alpha)$ and the relationship $\alpha\varphi'(e) + \xi'(e) = t$, we can derive

$$\begin{aligned}\gamma'(\alpha) &= \varphi(e) + \alpha\varphi'(e)e'(\alpha) + \xi'(e)e'(\alpha) = \varphi(e) + te'(\alpha) \\ &= \varphi(e) - t\frac{\varphi'(e)}{\alpha\varphi''(e) + \xi''(e)}\end{aligned}$$

Suppose $\xi(\cdot)$ is linear, with $\xi'(\cdot) = s$. Then $\xi''(e)$ in the above expression disappears, and, using $\alpha\varphi'(e) + \xi'(e) = t$, we can rewrite

$$\gamma'(\alpha) = \varphi(e) - t\frac{\varphi'(e)}{\frac{t-\xi'(e)}{\varphi'(e)}\varphi''(e)} = \varphi(e) - t\frac{(\varphi'(e))^2}{(t-s)\varphi''(e)}$$

Hence,

$$\gamma''(\alpha) = \left[\varphi'(e) - \frac{t}{t-s}\frac{2\varphi'(e)(\varphi''(e))^2 - \varphi'''(e)(\varphi'(e))^2}{(\varphi''(e))^2}\right] e'(\alpha)$$

Since $e'(\alpha) > 0$ the sign of $\gamma''(\alpha)$ is determined by the expression in brackets. Straightforward algebra yields that this expression is positive if and only if

$$t\varphi'''(e)\varphi'(e) > (t+s)(\varphi''(e))^2 \tag{22}$$

When $\xi(e)$ is a constant, the condition becomes:[9]

$$\varphi'''(e)\varphi'(e) > (\varphi''(e))^2 \tag{23}$$

Due to Lemma 3, condition (22) is necessary and sufficient for full transparency to be optimal.

Now, the sufficient condition (21) in the case of linear $\xi(\cdot)$ writes:

$$\begin{cases} \varphi''(e)s \geq 0 \\ \varphi'''(e)\varphi'(e) \geq (\varphi''(e))^2 \end{cases}, \tag{24}$$

where at least one of the inequalities has to be strict.

For $s > 0$, (24) is so strong that it never holds, as the first condition is never satisfied. For $s < 0$, the first condition in (24) always holds, but the second one is obviously stronger than (22). For $s = 0$, the first condition holds as an equality, hence the second condition must hold as a strict inequality, which coincides with (23). Hence, for $s = 0$ (that is, when $\xi(\cdot)$ is a constant), our sufficient condition is also necessary.

6 Conclusion

In this paper we have addressed the following question: When is commitment to a full disclosure of payoff-relevant information to the agent benefits the principal? In a rather general setting, we have obtained an interpretable and easily verifiable sufficient condition for the optimality of transparency. We have also provided a sufficient condition under

[9] In applications, people sometimes assume a linear output but a convex cost of effort. Obviously, there exists a tansformation $a := z(e)$ such that $\eta(a) \equiv \varphi(e)$ is a linear function of a and $k(a) \equiv e$ is a convex function of a. Then the condition is equivalent to $k'''(a)k'(a) < 2(k''(a))^2$.

which transparency is suboptimal. There are several avenues for further research. One interesting question is: When is complete *non*-transparency optimal? Another issue is the interaction between explicit compensation schemes for the agent and disclosure policy, and joint determination of the optimal compensation and disclosure.

7 Appendix

Proof of Lemma 1. Let $h(\omega, e) := \dfrac{g_e(\omega, e)}{-f_{ee}(\omega, e)}$. Condition (2) is equivalent to the statement that $h(\omega, e)$ increases in all directions in which e and $f_e(\omega, e)$ jointly increase. So, let us define a direction through a function $\omega(e)$ and take the full derivative of $h(\omega, e)$ wrt e:

$$\begin{aligned}
\frac{dh}{de} &= \frac{-\dfrac{dg_e}{de} f_{ee} + \dfrac{df_{ee}}{de} g_e}{(f_{ee})^2} > 0 \\
&\Leftrightarrow \frac{df_{ee}}{de} g_e - \frac{dg_e}{de} f_{ee} > 0
\end{aligned}$$

Taking into account that

$$\begin{aligned}
\frac{dg_e}{de} &= g_{ee} + g_{e\omega} \frac{d\omega}{de} \\
\frac{df_{ee}}{de} &= f_{eee} + f_{ee\omega} \frac{d\omega}{de}
\end{aligned}$$

the inequality becomes

$$\begin{aligned}
& f_{eee} g_e + f_{ee\omega} g_e \frac{d\omega}{de} - \left(g_{ee} f_{ee} + g_{e\omega} f_{ee} \frac{d\omega}{de} \right) \\
\equiv \; & f_{eee} g_e - g_{ee} f_{ee} + (f_{ee\omega} g_e - g_{e\omega} f_{ee}) \frac{d\omega}{de} > 0
\end{aligned} \tag{25}$$

We need that it holds for all $\omega(e)$ such that $\dfrac{df_e}{de} > 0$ (i.e., all directions in which f_e increases as well). As $\dfrac{df_e}{de} = f_{ee} + f_{e\omega} \dfrac{d\omega}{de}$, we have that $\dfrac{df_e}{de} > 0$ is equivalent to

$$\begin{cases}
\dfrac{d\omega}{de} > -\dfrac{f_{ee}}{f_{e\omega}} \text{ if } f_{e\omega} > 0 \\
\dfrac{d\omega}{de} < -\dfrac{f_{ee}}{f_{e\omega}} \text{ if } f_{e\omega} < 0
\end{cases}$$

Consider first the case when $f_{e\omega} > 0$. Then, the necessary and sufficient conditions for (25) to hold for all $\omega(e)$ such that $\frac{df_e}{de} > 0$, given that $\frac{d\omega}{de}$ can take any value above $-\frac{f_{ee}}{f_{e\omega}}$, are the following:

$$\begin{cases} f_{ee\omega}g_e - g_{e\omega}f_{ee} \geq 0 \\ f_{eee}g_e - g_{ee}f_{ee} - (f_{ee\omega}g_e - g_{e\omega}f_{ee})\frac{f_{ee}}{f_{e\omega}} \geq 0 \end{cases},$$

which becomes

$$\begin{cases} f_{ee\omega}g_e \geq g_{e\omega}f_{ee} \\ g_e(f_{eee}f_{e\omega} - f_{ee\omega}f_{ee}) \geq f_{ee}(g_{ee}f_{e\omega} - g_{e\omega}f_{ee}) \end{cases}, \tag{26}$$

where at least one of the inequalities has to be strict.

Consider now the case when $f_{e\omega} < 0$. Then, following the same steps we get

$$\begin{cases} f_{ee\omega}g_e \leq g_{e\omega}f_{ee} \\ g_e(f_{eee}f_{e\omega} - f_{ee\omega}f_{ee}) \leq f_{ee}(g_{ee}f_{e\omega} - g_{e\omega}f_{ee}) \end{cases},$$

where at least one of the inequalities has to be strict. ■

Proof of Lemma 2. To prove the lemma we just need to replace g with f in the proof of Lemma 1 and consider the case when $f_{e\omega} > 0$. That is, the sufficient condition is (26) with f instead of g:

$$\begin{cases} f_{ee\omega}f_e \geq f_{e\omega}f_{ee} \\ f_{eee}f_{e\omega} \geq f_{ee}f_{ee\omega} \end{cases},$$

where at least one of the inequalities has to be strict. ■

Proof of Theorem 2. To prove the theorem we borrow the construction and notation we used in the proof of Theorem 1 with a few simplifications and the "opposite" ultimate goal: deriving a condition when *pooling* two states benefits the principal.

Consider two states, ω_1 and ω_2, such that they generate different effort levels, if revealed. Consider pooling of ω_1 and ω_2 into one message m^*. Denote its support $\{\omega_1, \omega_2\}$ by Ω^*. Let e^* denote the agent's optimal effort upon receiving m^*. Clearly $f_e(\omega_i, e^*)$ must be above 1 in one state and below 1 in the other one, so WLOG let $f_e(\omega_1, e^*) < 1$.

Sets Ω_1^* and Ω_2^* from the proof of Theorem 1 are simply ω_1 and ω_2 now. Respectively, formula (3) simplifies to

$$\overline{x}_1 := f_e(\omega_1, e^*) < 1 < f(\omega_2, e^*) =: \underline{x}_2$$

Now, with ω_1 and ω_2 replacing m_1 and m_2 and dropping message $\widetilde{m}$ (we do not have it now), we obtain the counterpart of (5):

$$\Pr(\omega_1)\left[1 - f_e(\omega_1, e^*)\right] = \Pr(\omega_2)\left[f_e(\omega_2, e^*) - 1\right],$$

where $\Pr(\omega_i)$ denotes $\Pr(\omega_i|\Omega^*)$, for simplicity. Then, calling $e_1^*,\ e_2^*$ the agent's optimal efforts under ω_1, ω_2, pooling improves the principal's welfare whenever

$$\Pr(\omega_1)g(\omega_1, e^*) + \Pr(\omega_2)g(\omega_2, e^*) > \Pr(\omega_1)g(\omega_1, e_1^*) + \Pr(\omega_2)g(\omega_2, e_2^*),$$

or

$$\Pr(\omega_1)\left[g(\omega_1, e^*) - g(\omega_1, e_1^*)\right] > \Pr(\omega_2)\left[g(\omega_2, e_2^*) - g(\omega_2, e^*)\right],$$

which is a counterpart of (6) but with the opposite sign.

Now, repeating the same steps as in the proof of Theorem 1 with ω_1 and ω_2 rather than m_1 and m_2 and with tildes removed we obtain the counterpart of (11) but with the opposite sign:

$$\frac{g_e(\omega_1, e_1)}{-f_{ee}(\omega_1, e_1)} > \frac{g_e(\omega_2, e_2)}{-f_{ee}(\omega_2, e_2)} \tag{27}$$

Obviously the relationships $e_1 < e_2$ and $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$ hold, as they are based on the properties of $\Omega^*,\ \Omega_1^*,\ \Omega_2^*$, which remain the same as in the proof of Theorem 1.

Hence, if there exist ω_1, ω_2 such that $e_1 < e_2$ and $f_e(\omega_1, e_1) < f_e(\omega_2, e_2)$ imply (27), transparency cannot be optimal. ■

References

[1] Arieli I., Y. Babichenko, R. Smorodinsky, and T. Yamashita, "Optimal persuasion via bi-pooling," working paper, available at https://ssrn.com/abstract=3511516.

[2] Dizdar D., and E. Kováč, 2020, "A simple proof of strong duality in the linear persuasion problem," *Games and Economic Behavior*, 122: 407-12.

[3] Dworczak, P., and A. Kolotilin, 2019, "The persuasion duality," working paper, available at https://ssrn.com/abstract=3474376.

[4] Dworczak P, and G. Martini, 2019, "The simple economics of optimal persuasion," *Journal of Political Economy*, 127(5): 1993-2048.

[5] Galperti S, and J. Perego, 2018, "A dual perspective on information design," working paper, available at https://ssrn.com/abstract=3297406.

[6] Gentzkow M, and E. Kamenica, 2016, "A Rothschild-Stiglitz approach to Bayesian persuasion," *American Economic Review*, 106(5): 597-601.

[7] Jehiel P., 2015, "On transparency in organizations," *The Review of Economic Studies*, 82(2): 736-61.

[8] Kamenica E, and M. Gentzkow, "Bayesian persuasion," *American Economic Review*, 101(6): 2590-615.

[9] Kolotilin, A., 2018, "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 13(2) :607-35.

[10] Kolotilin, A. and A. Wolitzky, 2020, "Assortative information disclosure," working paper, available at https://ssrn.com/abstract=3616781

[11] Kolotilin, A., T. Mylovanov, and A. Zapechelnyuk, 2021, "Censorship as optimal persuasion," working paper.

[12] Mensch J., 2021, "Monotone persuasion," *Games and Economic Behavior*, 130: 521-42.

[13] Rayo L, and I. Segal, 2010, "Optimal information disclosure," *Journal of Political Economy*, 118(5): 949-87.