Online Appendix for
“False Positives and Transparency”
By Jonathan Libgober

A. PROOFS

I prove the following Lemma (which allows for more than two possible signals), and note that Lemma 2 follows immediately when pure strategy equilibrium is imposed (noting that the requirement in the Lemma holds immediately whenever $a = \tilde{a}$).

Lemma 5. *Suppose the receiver infers the sender's choice as $\tilde{a}$. Then if the sender chooses experiment a such that $\mathbb{P}[y \mid \tilde{a}, \theta] = 0$ implies $\mathbb{P}[y \mid a, \theta] = 0$, sender's payoffs can be written:*

$$\mathbb{E}_{y \sim \tilde{a}}\left[\mathbb{E}_{\theta \sim \hat{p}_{\tilde{a}}(y)}\left[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta)\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]}\right]\right] - c_S(a).$$

Proof of Lemma 5. The sender's payoffs can be expressed as:

$$\mathbb{E}_{y,\theta}[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta)] - c_S(a),$$

where the expectation is taken over realizations of θ and y, and $e_{\tilde{a}}$ is the receiver's equilibrium effort as a function of y. Note that the distribution over y is a function of the true sender choice (i.e., a), and not the inferred sender choice (i.e., $\tilde{a}$). Further noting that we can restrict to y which occur under a with positive probability, and that $\mathbb{P}[y \mid a, \theta] > 0$ implies $\mathbb{P}[y \mid \tilde{a}, \theta] > 0$. We thus rewrite the expectation as over realizations of y and θ:

$$\begin{aligned}
\mathbb{E}_{y,\theta}[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta)] &= \sum_{\theta}\left(\sum_{y} U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta)\mathbb{P}[y \mid a, \theta]\right)\mathbb{P}[\theta] \\
&= \sum_{\theta}\left(\sum_{y} U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]}\right) \cdot \mathbb{P}[y \mid \tilde{a}, \theta]\right)\mathbb{P}[\theta] \\
&= \sum_{\theta}\left(\sum_{y} U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]}\right) \cdot \frac{\mathbb{P}[y \mid \tilde{a}, \theta]\mathbb{P}[\theta]}{\sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}]} \cdot \sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right) \\
&= \sum_{y}\left(\sum_{\theta} U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]}\right) \cdot \frac{\mathbb{P}[y \mid \tilde{a}, \theta]\mathbb{P}[\theta]}{\sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}]} \cdot \sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right)
\end{aligned}$$

$$= \sum_y \left(\sum_\theta U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]} \right) \cdot \mathbb{P}[\theta \mid \tilde{a}, y] \cdot \sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}] \right)$$

$$= \sum_y \left(\mathbb{E}_{\theta \sim \hat{p}_{\tilde{a}}(y)} \left[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]} \right) \right] \sum_{\tilde{\theta}} \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}] \right)$$

$$= \sum_{\tilde{\theta}, y} \mathbb{E}_{\theta \sim \hat{p}_{\tilde{a}}(y)} \left[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \cdot \left(\frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]} \right) \right] \mathbb{P}[y \mid \tilde{a}, \tilde{\theta}]\mathbb{P}[\tilde{\theta}]$$

$$= \mathbb{E}_{y \sim \tilde{a}} \left[\mathbb{E}_{\theta \sim \hat{p}_{\tilde{a}}(y)} \left[U^S(e_{\tilde{a}}(y), \hat{p}_{\tilde{a}}(y), \theta) \frac{\mathbb{P}[y \mid a, \theta]}{\mathbb{P}[y \mid \tilde{a}, \theta]} \right] \right].$$

Essentially, the argument follows from noting that under the full support condition, it is possible to divide and multiply every term in the sum by $\mathbb{P}[y \mid \tilde{a}, \theta]$ and $\mathbb{P}_{\tilde{a}}[y]$, in order to move between the objective distribution over states and the distribution perceived by the receiver after observing y. The fifth line follows from an application of Bayes rule, noting that this term is equal to the posterior belief that the state is θ when the chosen experiment is $\tilde{a}$. Finally, in any pure strategy equilibrium, we have both that the sender chooses a and that the receiver infers that the sender chose experiment a. Hence $\mathbb{P}[y \mid a, \theta] = \mathbb{P}[y \mid \tilde{a}, \theta]$, giving the second expression. □

An alternative way of expressing sender's payoffs in the case where $\tilde{a} \neq a$ is:

$$\mathbb{E}_{y \sim \tilde{a}} \left[\mathbb{E}_{\theta \sim \hat{p}_{\tilde{a}}(y)} \left[U^S(e_{\tilde{a}}(y), \hat{p}_a(y), \theta) \frac{\hat{p}_a(y)[\theta]}{\hat{p}_{\tilde{a}}(y)[\theta]} \cdot \frac{\mathbb{P}[y \mid a]}{\mathbb{P}[y \mid \tilde{a}]} \right] \right] - c_S(a),$$

which follows from noting that $\hat{p}_a(y)[\theta] \cdot \mathbb{P}[y \mid a] = \mathbb{P}[\theta, y \mid a] = \mathbb{P}[y \mid a, \theta]\mathbb{P}[\theta]$. In other words, to use the belief ratio to switch between "actual" and the "inferred." In the case where the sender and receiver disagree over the *prior* (as opposed to the experiment), Alonso and Câmara (2016) use belief ratios to rewrite the receiver's induced belief as a function of the sender's (their Proposition 1), and subsequently apply the belief based approach to the sender's value function. Note that their Proposition 1 is true regardless whether the experiment is sender-optimal, although their primary applications of this result relate to this case.

The next proof is stated assuming no restrictions on the number of indices. In this case, Assumptions 2 and 3 should be understood as applying to the vector of indices which may potentially be unobservable to the receiver.

Proof of Lemma 1. Suppose M is the index set of observable indices, and partition the sender's action into $a = (a_M, a_{-M})$. This proof shows that there is some a^*_{-M} such that when the receiver conjectures that a^*_{-M} are the unobserved actions of the sender, the sender's best response is to follow action a^*_{-M}. Since p_0 is interior and, for any choice of experiment, some signal occurs with positive probability in some state, the receiver always puts non-negative probability on observing any $y \in \{0, 1\}$, for any conjecture regarding the sender's behavior. Therefore, there is a unique belief profile $(\hat{p}_a(y))_{y \in Y}$ formed after observing any signal, for any equilibrium strategy of the sender. In fact, since A is compact, we have that $\mathbb{P}[y]$ is bounded away from 0 for all y. This implies that beliefs are a continuous function of actions, and well-defined given any conjecture. With these preliminaries in mind, the proof applies Kakutani's theorem to the sender's best reply correspondence. Define the function $\phi(a)$ as follows:

$$\phi(a) = \underset{\tilde{a} \in A_{-M}}{\arg\max} \overbrace{\sum_{\theta} \left(\sum_{y} U^S(e_a(y), \hat{p}_a(y), \theta) \mathbb{P}[y \mid \theta, a_M, \tilde{a}] \right) \mathbb{P}[\theta]}^{(\dagger\dagger)} - c(a_M, \tilde{a}).$$

Note that $\phi(a)$ gives the payoff maximizing response, assuming (observable) actions a_M are chosen and a conjecture of a. We first show this is function is upper hemicontinuous in a_{-M}, noting that the variables inside U^S do not respond to $\tilde{a}$. Take $a_n \to a$, and $b_n \in \phi(a_n)$ with $b_n \to b$. Write $a_n = (a_M, a^n_{-M})$.

Recall that beliefs are continuous in the sender's action choice, since $\mathbb{P}[y \mid a, \theta]$ is bounded away from 0 on a compact set. We now show that if $\hat{p}_n \to \hat{p}^*$, then $e(\hat{p}_n) \to e(\hat{p}^*)$ (that is, $e(\hat{p})$ is continuous in $\hat{p}$); since effort is chosen from a compact set and the receiver's best response is unique, we can ensure $e(\hat{p}_n) \to e^*$, passing to a subsequence if necessary by compactness of the receiver's action set. If e^* does not maximize $\mathbb{E}_{\theta \sim \hat{p}^*}[U^R(e, \theta)]$, then there is some e^{**} where the receiver does strictly better when the induced belief is $\hat{p}^*$. But continuity of the receiver payoff function implies that $\mathbb{E}_{\theta \sim \hat{p}_n}[U^R(e(\hat{p}_n), \theta)] \to \mathbb{E}_{\theta \sim \hat{p}^*}[U^R(e^*, \theta)]$, which implies that e^{**} would be a preferred action choice to $e(\hat{p}_n)$ for some n sufficiently large, contradicting the definition of $e(\hat{p}_n)$.

From this, we conclude that $(\dagger\dagger)$ is simply the sum and product of terms that are continuous in a, and so:

$$\sum_{\theta} \left(\sum_{y} U^S(e_{a_n}(y), \hat{p}_{a_n}(y), \theta) \mathbb{P}[y \mid (a_M, \tilde{a}_{-M}), \theta] \right) \mathbb{P}[\theta] \to^n$$
$$\sum_{\theta} \left(\sum_{y} U^S(e_{(a_M, a)}(y), \hat{p}_{(a_M, a)}(y), \theta) \mathbb{P}[y \mid (a_M, \tilde{a}), \theta] \right) \mathbb{P}[\theta].$$

If $b \notin \phi(a)$, then there exists some value δ such that a deviation to δ would result in a higher objective than b, namely we would have:

$$\sum_\theta \left(\sum_y U^S(e_a(y), \hat{p}_a(y), \theta)(\mathbb{P}[y \mid \theta, (a_M, \delta)] - \mathbb{P}[y \mid \theta, (a_M, b_{-M})]) \right) \mathbb{P}[\theta] > c(a_M, \delta) - c(a_M, b_{-M}).$$

But since $a^n_{-M} \to a_{-M}$ and $b_n \to b$, by continuity we would be able to find some find some n sufficiently large such that this inequality would also be satisfied replacing b by b_n and a with a_n—that is, sufficiently close to the limit—which would contradict our assumption that b_n is a maximizer of $\phi(a_n)$. Hence the map ϕ is upper-hemicontinuous.

Furthermore, $\phi(a)$ is nonempty and closed. To see this, note that A_{-M} is compact, and as we have argued, the objective is continuous in $\tilde{a}$. Since the set of maximizers of a continuous function on a compact set is itself a compact nonempty set, we have that $\phi(a)$ is compact.

To show that $\phi(a)$ is convex, we show that the objective is concave in $\tilde{a}$. Note that we can write each term inside the sum over θ in $(\dagger\dagger)$ as:

$$\left(U^S(e_a(0), \hat{p}_a(0), \theta) + (U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta))\mathbb{P}[1 \mid \theta, a_M, \tilde{a}] \right) \mathbb{P}[\theta].$$

Note that $U^S(e_a(0), \hat{p}_a(0), \theta)$ is a constant in $\tilde{a}$, and by Assumption 3, since $U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta) > 0$ and $\mathbb{P}[1 \mid \theta, a_M, \tilde{a}]$ is concave, we have this expression is concave in $\tilde{a}$ (after summing over states θ). Furthermore, by Assumption 2, c_S is convex in $\tilde{a}$, meaning that the objective in the definition of ϕ is concave in $\tilde{a}$.

So suppose that a', a'' are both in $\phi(a)$ (noting that these must differ only in the coordinates A_{-M}; so in other words, supposing both a' and a'' are maximizing choices of $\tilde{a}$). Since $\phi(a)$ is concave $\tilde{a}$, it follows that if this expression is maximized at a' and a'', it must also be maximized at every $a''' = \alpha a' + (1-\alpha)a''$, as desired. Having demonstrated that the conditions for Kakutani's fixed point theorem are satisfied, an equilibrium exists when a_M is observed, for any choice of a_M.

To conclude that a PBE exists, note that the above result shows that we can write the inferred choice of a_{-M} as a function of a_M, i.e. $a_{-M}(a_M)$, given some selection. (When there are two dimensions, this selection is written $a_2(a_1)$) Since A_M (or A_1) is finite, it follows that the image of A_1 under the sender's utility is finite as well. Its maximizer is achieved at an element of A_1, and so sender maximizes payoff by choosing this element of a_1, yielding a PBE.

I now show the claim on mixed strategy equilibria. If there were multiple equilibria, then the first order condition must hold for two values of $\tilde{a}$, say $\tilde{a}^1 < \tilde{a}^2$. On the other hand, the receiver's beliefs do not depend on the choice of $\tilde{a}$. Using expression for the sender's benefit in terms of

$U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)$, we have:

$$\nabla_{\tilde{a}} c(a_M, \tilde{a}^i) = \sum_\theta \left(U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)\right) \mathbb{P}[\theta] \nabla_{\tilde{a}} \mathbb{P}[y = 1 \mid (a_M, \tilde{a}^i), \theta],$$

and hence subtracting the equation for $i = 1$ from the equation for $i = 2$, and taking the dot product for some arbitrary α with $||\alpha|| = 1$:

$$\alpha \cdot \nabla_{\tilde{a}}(c(a_M, \tilde{a}^2) - c(a_M, \tilde{a}^1)) =$$
$$\sum_\theta (U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)) \mathbb{P}[\theta] (\alpha \cdot \nabla_{\tilde{a}}(\mathbb{P}[y \mid (a_M, \tilde{a}^2), \theta] - \mathbb{P}[y \mid (a_M, \tilde{a}^1), \theta])).$$

By the mean value theorem, applied to $\mathbb{P}[y \mid a, \theta]$ and c in order to obtain choices a_θ and a_c which are all convex combinations of $\tilde{a}^1$ and $\tilde{a}^2$ such that:

$$\alpha \cdot (\nabla^2_{\tilde{a}} c(a_M, a_c) \cdot (\tilde{a}^2 - \tilde{a}^1)) =$$
$$\sum_\theta (U^S(e_{(a_M,a)}(1), \hat{p}_{(a_M,a)}(1), \theta) - U^S(e_{(a_M,a)}(0), \hat{p}_{(a_M,a)}(0), \theta)) \mathbb{P}[\theta] \alpha \cdot \nabla^2_{\tilde{a}} \mathbb{P}[y = 1 \mid (a_M, a_\theta), \theta] \cdot (\tilde{a}^2 - \tilde{a}^1).$$

But by the strictness of concavity or convexity, either the left hand side is strictly positive or the right hand side is strictly negative, with both being at least weakly so, a contradiction. Hence in equilibrium, there can only be pure strategies. □

Proof of Proposition 1. I provide a proof showing that (1) being positive implies higher choices of a_2; showing that lower choices are implied when this expression is negative follows identical reasoning. First, note that the sender's payoff (in equilibrium) can be written:

$$\sum_y \sum_\theta U^S(\hat{p}_{\tilde{a}}(y), \theta) \mathbb{P}[y \mid a, \theta] \mathbb{P}[\theta] - c_S(a).$$

The proof considers the first-order condition of this expression. Note that the first-order conditions (as an equality for interior a_2 and an inequality for $a_2 \in \{\min A_2, \max A_2\}$) must hold both under full transparency, as well as under partial transparency (as per the proof of Lemma 1). The difference in these expression is that when a_2 is unobservable, $\hat{p}_{\tilde{a}}$ does not not change as the sender changes a_2.

First consider limited transparency. Given a pure strategy PBE, the first order condition following a correct inference by the receiver is:

$$\frac{dc_S(a)}{da_2} = \sum_y \sum_\theta U^S(\hat{p}_a(y), \theta) \frac{d\mathbb{P}[y \mid a, \theta]}{da_2} \mathbb{P}[\theta].$$

When a_2 is observable, a term is added to the right hand side, which corresponds to the change in the receiver's belief about the state. The action is higher whenever this term is positive. Using that the beliefs are differentiable as a function of the action, chain rule gives us that the added term is:

$$\sum_y \sum_\theta \nabla_{\hat{p}} U^S(\hat{p}_a(y), \theta) \cdot \frac{d\hat{p}_a(y)[\cdot]}{da_2} \mathbb{P}[y \mid a, \theta]\mathbb{P}[\theta],$$

which is (1) (noting that the brackets reflect that $\hat{p}$ is a belief over many states).

Now, suppose the first order condition holds at an interior value of a_2, say a^*_{obs}. If (1) is positive, then:

$$\sum_y \sum_\theta U^S(\hat{p}_a(y), \theta) \frac{d\mathbb{P}[y \mid a, \theta]}{da_2} \mathbb{P}[\theta] - \left.\frac{dc_S(a)}{da_2}\right|_{a=a^*_{obs}} < 0,$$

since this holds with equality when (1) is added. Note that the objective, as a function of a, is concave in every coordinate under Assumptions 1-3, meaning that the left hand side is decreasing in a_2. It follows that for the first order conditions to hold, the resulting a_2 must be lower. Hence if (1) is positive, then keeping a_2 hidden lowers it, so that the choice of a_2 is higher under observability, as claimed. Similar reasoning reveals the same conclusion holds at boundary cases, although the change need not be strict since the first order conditions only hold as inequalities in these cases (with the direction of the inequality depending on which boundary is considered). □

Proof of Lemma 3. Denote by $a_2^*(a_1)$ the equilibrium response of a_2, fixing the choice of a_1. By Lemma 1, this is characterized by the first order condition:

$$\frac{\partial}{\partial a_2} c_S(a_1, a_2^*(a_1)) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_2} \mathbb{P}[y \mid (a_1, a_2^*(a_1)), \theta]\mathbb{P}[\theta], \tag{6}$$

with equality holding whenever a_2 is interior. Using that Y is binary and evaluating at $a_1 = \tilde{a}_1$, we rewrite this as:

$$\frac{\partial}{\partial a_2} c_S(\tilde{a}_1, a_2^*(\tilde{a}_1)) \leq \sum_\theta (U^S(\hat{p}(1), \theta) - U^S(\hat{p}(0), \theta)) \frac{\partial}{\partial a_2} \mathbb{P}[y = 1 \mid (\tilde{a}_1, a_2^*(\tilde{a}_1)), \theta]\mathbb{P}[\theta]. \tag{7}$$

First suppose that the first order condition defining $a_2^*(\tilde{a}_1)$ holds with equality. Then adding $\frac{\partial c_S(a_1^*, a_2)}{\partial a_2} - \frac{\partial c_S(\tilde{a}_1, a_2)}{\partial a_2} \leq M_{a_1^*}(a_2) - M_{\tilde{a}_1}(a_2)$ to both sides of (7) when it holds with equality yields:

$$\frac{\partial}{\partial a_2} c_S(a_1^*, a_2^*(a_1^*)) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_2} \mathbb{P}[y \mid (a_1^*, a_2^*(a_1^*)), \theta] \mathbb{P}[\theta], \tag{8}$$

Using the fact that $c(a_1, a_2)$ is convex in a_2, and that $\mathbb{P}[y = 1 \mid (a_1, a_2), \theta]$ is concave in a_2, it follows that given an inferred choice of a_2 must be higher when a_1^* is chosen, and strictly so when a_2 is interior and the inequality is strict. On the other hand, if the first order condition holds as a strict inequality, then a_2 is chosen as an edge case, and the same reasoning implies that a_2 could only increase as well. □

Proof of Proposition 2. The key observation is that, even though transparency may increase a_2 uniformly over all choices of a_1, the losses from this increase in a_2 is small relative to the benefit (to the receiver) from inducing a higher a_1. Therefore, I first argue that it suffices to show the following, given the conditions of the Proposition:

- The losses (to the sender) from keeping $a_1 = a_1^{obs}$ are large, and
- The losses (to both the sender and the receiver) due to higher a_2 are small when $a_1 > a_1^{obs}$.

To see that this suffices, first note that the receiver's payoff function is continuous in a_2, which follows immediately from continuity assumption on the sender's experiment choice. More precisely, since the proof of Lemma 1 shows that effort is continuous in posterior beliefs, as well as that beliefs are continuous in a_2, it follows that continuity of receiver's payoffs are maintained, given any (realized or conjectured) a_2.

Since the receiver's payoff increases in a_1, holding *fixed* the choice of a_2, there exists a discrete increase in the receiver's payoffs, say α, when the sender chooses a higher a_1. Therefore, by continuity of the receiver's payoff function, we can find some ε such that an increase from a_1 to $\alpha_i > a_1^{obs}$ also delivers a higher payoff, whenever the increase in a_2 is no more than ε (since as $\varepsilon \to 0$ corresponds to the case where a_2 does not increase).

Using these arguments, the proposition follows from the following observations: First, when (1) is sufficiently large, the loss to the sender is large as well. And second, the change in a_2 is minimal when a_1 increases, provided $M_{\alpha_i}(a_2) - M_{a_1^{obs}}(a_2) - \left(\frac{\partial c_S(\alpha_i, a_2)}{\partial a_2} - \frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}\right)$ is large relative to (1).

Both of these claims follow immediately from the same arguments as in Proposition 1 and Lemma 3. Denoting the equilibrium response of a_2 given an (observable) choice of a_1 by $a_2^*(a_1)$, given (1) sufficiently positive relative to $\frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}$, Proposition 1 shows that $a_2^*(a_1^{obs})$ approaches $\max A_2$. By assumption, this lowers sender's payoff relative to any other action, provided a_2 is chosen sufficiently close to a_2^{obs}. Indeed, Lemma 3 shows that a_2^* decreases when a_1 increases. But in fact, inspecting the first order conditions, we see that if the second bulletpoint holds, then the first order condition will be satisfied at a value of a_2 no more than ε larger than a_2^{obs}.

Hence as discussed above, the conditions ensure that the change in payoffs due to the (potentially) higher a_2 is small relative to the increase in payoffs due to higher a_1. This proves the proposition. The reasoning for the converse case is identical and hence omitted. $\square$

Proof of Lemma 4. I first use Lemma 2 to write the sender's payoff from $(*)$ as a function of the receiver's ex-post beliefs. Applying this Lemma shows that the Sender's payoff is proportional to $\hat{p}e(\hat{p})$. Thus, to prove Lemma 4, it suffices to show that $\hat{p}e(\hat{p})$ is convex. This is immediate in the case of polynomial effort costs, since then $e(\hat{p})$ is proportional to $\hat{p}^{1/(n-1)}$, so that $\hat{p} \cdot \hat{p}^{1/(n-1)}$ is convex. More generally, take the second derivative of $pe(p)$ and observe that it is equal to:

$$2e'(p) + pe''(p).$$

Since $c_R(e)$ is strictly convex, $e'(p) > 0$, which follows from the receiver's first order condition:

$$b_R p = c_R'(e(p)),$$

differentiating with respect to p to obtain:

$$b_R = c_R''(e(p))e'(p).$$

Differentiating again gives:

$$0 = c_R'''(e)(e'(p))^2 + c_R''(e)e''(p).$$

Since $e(p)$ is strictly increasing, the assumptions on $c_R'''(e)$ ensure that $e''(p) \geq 0$, and hence the objective is convex. $\square$

In general, convexity of receiver effort by itself is not a strong enough assumption in order to ensure that $pe(p)$ is convex. To see this, suppose that:

$$c_R(e) = 1 - \sqrt{1-e} \Rightarrow c_R'(e) = \frac{1}{2\sqrt{1-e}} > 0 \Rightarrow c_R''(e) = \frac{1}{4(1-e)^{3/2}} > 0.$$

In that case:

$$e(p) = \max\{0, -\frac{1}{4b_R^2 p^2} + 1\},$$

and observe that $pe(p)$ is concave whenever $e(p) > 0$.

B. MISCELLANEOUS

B.1. Counterexample to Lemma 1 when Assumptions are Violated

I briefly demonstrate, by example, on the possibility of a failure of pure strategy equilibrium existence when results cannot be classified into positives and negatives. The failure arises due to a failure of concavity in the objective stated in Lemma 1. This can be avoided to a certain extent by taking a transformation of the index; hence the point of this example is to show that the real technical issue arises when the "positive" and "negative" label depends on the experiment choice, which is not fixed by taking a monotone transformation of a_2. Note that in this example, the feasible experiment set is convex.

Let $|\Theta| = |Y| = 2$, with $\Theta = \{-1, 1\}$ and $\mathbb{P}[\theta = 1] = 1/2$. Consider the following sender preferences:

$$U^S(\hat{p}, \theta) = -\hat{p}[\theta = 1] \cdot \theta$$

And let:

$$\mathbb{P}[Y = 1 \mid \theta = 1] = \mathbb{P}[Y = 0 \mid \theta = -1] = 1 - \mathbb{P}[Y = 1 \mid \theta = -1] = 1 - \mathbb{P}[Y = 0 \mid \theta = 1] = a^2.$$

with $c(a) = a/4$. In this example, in state $\theta = 1$, the event $Y = 1$ is a positive result when $a < \frac{1}{\sqrt{2}}$ is inferred (in which case it is evidence for the state $\theta = -1$), and a negative result otherwise. In state $\theta = -1$, this is flipped. So when $a > \frac{1}{\sqrt{2}}$, the concavity assumption is satisfied in state $\theta = 1$ but violated in state $\theta = -1$, and the opposite is true when $a < \frac{1}{\sqrt{2}}$.

Write the payoff to the sender from an experiment a when it is inferred as $\tilde{a}$ (noting that $a = \tilde{a}$ in equilibrium). By symmetry, the probability of a positive result $1/2$ ex-ante, for any choice of experiment. Hence the payoff is:

$$-\frac{1}{4}a + \frac{1}{2}\left(-\tilde{a}^2 \cdot \left(a^2\right) + \tilde{a}^2\left(1 - a^2\right) - (1 - \tilde{a}^2)(1 - a^2) + (1 - \tilde{a}^2) \cdot a^2\right).$$

which reduces to:

$$-\frac{1}{4}a + \frac{1}{2}\left(-4\tilde{a}^2 a^2 + 2\tilde{a}^2 + 2a^2 - 1\right).$$

Given a conjecture of $\tilde{a}$, sender chooses a to maximize the objective $-\frac{a}{4} + a^2(1 - 2\tilde{a}^2)$. Note that if $\tilde{a}^2 \geq \frac{1}{2}$, this is maximized at $a = 0$, since the objective is negative for all other values of a. Thus, there is no pure strategy equilibrium where the sender chooses $\tilde{a} \geq \frac{1}{\sqrt{2}}$. On the other hand, if $\tilde{a}^2 < \frac{1}{2}$, then the second derivative of the objective is positive. Since this is a quadratic function, the optimum is on the boundary of the choice set, i.e., either 0 or 1, for any choice of $\tilde{a}$. Hence the only choice of a less than $\frac{1}{\sqrt{2}}$ that could possibly be part of a pure strategy equilibrium would $a = 0$. However, $a = 1$ is a best response to $\tilde{a} = 0$, showing that there is no pure strategy equilibrium.

B.2. Preferences over y

I comment on a modification to the model where I allow for the sender to have preferences over y itself. For simplicity, I consider the case where the payoffs are separable, and the sender obtains an added benefit of $\lambda_y \cdot y$. from a positive result. In principle, this model still is amenable to the belief-based approach, noting that any positive result leads to a higher belief and any negative result leads to a lower belief. Hence this setting is as if there were a jump in the sender's payoff function at the prior (as commented on in Footnote 2). That said, it is simplest to comment on this case simply by inspection. In this case, it is immediate that the sender is incentivized to maximize the biasing action in this case (whether higher informative actions will be taken depends on the prior):

Proposition 3. *As $\lambda_y \to \infty$, the sender's choice of experiment converges to the one which maximizes $\mathbb{P}[y = 1]$.*

However, I also comment that transparency does not interact with the experiment choice when all that matters is whether the result is positive or not (and would similarly expect a limited impact if this consideration itself was overwhelmingly dominant):

Proposition 4. *Suppose the sender's payoffs $U^S(e, \hat{p}, y, \theta)$ is constant in e and $\hat{p}$. Then the sender's experiment choice does not differ depending on transparency regime.*

This is immediate since the (ex-ante) probability of an outcome y, conditional on the experiment, does not depend on transparency regime, only on the realized experiment choice. Hence neither do payoffs if all that matters for the scientist is the probability of a positive result.

While this is theoretically immediate, it may seem surprising in the context of the application—the preference for positive results appears so widespread that it is tempting to think it is intrinsic. This paper takes the view that the benefit from a "positive" or "negative" result is endogenous and depends on the belief movement. If a preference for positive results emerges entirely because negative results are harder to publish, then this would suggest an *interaction* between having a positive result and the other payoff terms. While this would add a non-convexity in the sender's payoff as a function of the receiver's belief (due to the jump at the prior), provided it is small, the main conclusions of the paper should not change drastically (albeit with some additional notation). On the other hand, it would make it more unwieldy to characterize the preference for information in certain places (e.g., Section 2), without changing intuition. It may also be that positive results that are obtained "cheaply" (via bias) are less meaningful, but those that are achieved "scrupulously" (via informativeness) are more meaningful. This would suggest greater interdependence between the cost function and the benefit than what I have here. While these observations may call for more empirical commentary to identify the source and nature of any intrinsic preference for positive results, this is left for future work.

B.3. Example Illustrating the Role of Off-path beliefs

This section presents an example illustrating the role of the "correct inference assumption." This example is a modified version of Section 2. Specifically, consider payoffs exactly as in Section 2, but suppose instead the signal distribution is as follows:

$$\begin{aligned}
&\mathbb{P}[y=1 \mid \theta=T, a_1=0, a_2=0]=2/5, && \mathbb{P}[y=1 \mid \theta=F, a_1=0, a_2=0]=0,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=0, a_2=1]=1/2, && \mathbb{P}[y=1 \mid \theta=F, a_1=0, a_2=1]=1/6,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=1, a_2=0]=3/4, && \mathbb{P}[y=1 \mid \theta=F, a_1=1, a_2=0]=0,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=1, a_2=1]=7/8, && \mathbb{P}[y=1 \mid \theta=F, a_1=1, a_2=1]=2/3.
\end{aligned}$$

Intuitively, relative to the previous, we now let the $a_1 = 1$ experiment be (severely) susceptible to bias. Take $c(a_1, a_2) = c \cdot a_1 + k \cdot a_2$. Note that $\pi_R(a_1, a_2)$ is exactly as before for $a \neq (1, 1)$. Recall that in Section 2, if $c \in (\pi_R(1, 0) - \pi_R(0, 0), \pi_R(1, 0) - \pi_R(0, 1))$, then the sender would choose $a = (0, 0)$ under full transparency, but would choose $a = (1, 0)$ in order to credibly show that $a_2 \neq 1$.

However, whereas in Section 2 the $a_1 = 1$ experiment is *resistant* to bias, here the $a_1 = 1$ experiment is *highly susceptible* to bias. This suggests complementarity goes in the opposite direction, and thus that limited transparency would favor the $a_1 = 0$ experiment instead.[27] Now, $c > \pi_R(1, 0) - \pi_R(0, 0)$ implies the sender would rather choose $a_1 = 0$ if a_2 were inferred equal to 0. Furthermore, if k is very large, then indeed $a_2(1) = a_2(0) = 0$, since the marginal benefit to a higher probability of $y = 1$ is fixed and finite. I conclude that the limited transparency and full transparency experiments coincide if $c > \pi_R(1, 0) - \pi_R(0, 0)$ and k is sufficiently large, under the belief refinement in Lemma 1.

[27] To see this, first note that $\pi_R(1, 1) < \pi_R(0, 1)$, and $\pi_R(1, 0) - \pi_R(1, 1) > \pi_R(0, 0) - \pi_R(0, 1)$. In particular, one can calculate that $\pi_R(1, 1) = \frac{29}{414} \approx .07 < \frac{1}{12}$. We have $\pi_R(1, 0) - \pi_R(1, 1) = \frac{5}{26} - \frac{29}{414} \approx 0.12$, and $\pi_R(0, 0) - \pi_R(0, 1) = \frac{1}{8} - \frac{1}{12} \approx 0.04$. As a result, for $k = 0$ and $c \in (\pi_R(1, 1) - \pi_R(0, 1), \pi_R(1, 0) - \pi_R(0, 0))$, the sender chooses $a_1 = 1$ under full transparency but $a_1 = 0$ under partial transparency.

Now take k large and $c \in (\pi_R(1,0) - \pi_R(0,0), \pi_R(1,0) - \pi_R(0,1))$. With different off-path beliefs, limited transparency can favor $a_1 = 1$, despite the opposite complementarity. Consider the following profile under limited transparency:

- Sender chooses $(1,0)$
- Following an observation of $a_1 = 1$, the receiver infers $a_2 = 0$. Following an observation of $a_1 = 0$, the receiver infers $a_2 = 1$.

If beliefs are not restricted off-path, then the sender would not have any incentive to deviate from $(1,0)$ given the receiver's inference following an observation of $a_1 = 0$. And while this inference is inconsistent with an action profile where $a_1 = 0$ is exogenously imposed on the sender, the stated belief profile is valid if off-path beliefs are not restricted. I thus conclude that, despite the opposite complementarity from Section 2, again the more informative experiment is chosen in this equilibrium under limited transparency.