

Optimal Information Disclosure*

Luis Rayo

Booth School of Business, University of Chicago, Chicago, IL 60637

luis.rayo@uchicago.edu

Ilya Segal

Department of Economics, Stanford University, Stanford, CA 94305

ilya.segal@stanford.edu

May 1, 2009

WORK IN PROGRESS

Abstract

A “Sender” (internet advertising platform, seller, rating agency, or school) has a probability distribution over “prospects” (internet ads, products, bonds, or students). Each prospect is characterized by its profitability to the Sender and its relevance to a “Receiver” (internet user, consumer, investor, or employer). The Sender privately observes the profitability and relevance of the prospect, whereas the receiver observes only a signal provided by the Sender (the prospect’s “rating”). The Receiver accepts a given prospect only if his Bayesian inference about its relevance exceeds a private opportunity cost that is uniformly drawn from $[0,1]$. We characterize the Sender’s optimal information disclosure rule assuming commitment power on her behalf. While the Receiver’s welfare is maximized by full disclosure, the Sender’s profits are typically maximized by partial disclosure, in which the Receiver is induced to accepting less relevant but more profitable prospects (“switches”) by pooling them with more relevant but less profitable ones (“baits”). Extensions of the model include maximizing a weighted sum of Sender profits and Receiver welfare, and allowing the Sender to subsidize or tax the Receiver.

*Acknowledgements to be added.

1 Introduction

An Internet advertising platform can provide some information to users about the relevance of its ads. This information can be signaled by such features as the ad's position on the web page, its font size, color, flashing, etc. Suppose that users have rational expectations and are sophisticated enough to interpret these signals. Then user welfare would be maximized by communicating the ads' relevance to them, allowing them to make fully informed decisions about which ads to click.

The platform, however, may care not just about user welfare, but about its own profits. Suppose that each potential ad is characterized by its value to consumers and its per-click profits to the platform, and the two are not always aligned. Then the platform would increase its profits by inducing users to click on more profitable ads.

While the platform would not be able to fool rational users systematically to induce them to click more on less relevant ads, a similar effect could be achieved by withholding some information from them, pooling the less relevant but more profitable ads with those that are more relevant and less profitable.

Similar information disclosure problems arise in other economic settings:

1. A seller chooses which information to disclose to consumers about its products, which vary both in their profitability to the seller and their value to consumers.
2. A bond rating agency chooses what information to disclose to investors about bond issuers, who also make payments to the agency for the rating.
3. A school chooses what information to disclose to prospective employers about the ability of its students, who also pay tuition to the school.

In each of these cases, the profit-maximizing information disclosure rule may be partially but not fully revealing.

This paper characterizes the optimal information disclosure rule in such settings. Our basic model has two agents - the "Sender" and the "Receiver." The Sender (who can be alternatively interpreted as an advertising platform, seller, rating agency, or school) has a probability distribution over "prospects" (ads, products, bonds, or students, respectively). Each prospect is characterized by its profitability to the Sender and its value to the Receiver (user, consumer, investor, or employer), which are not observed by the Receiver. First, the Sender chooses an information disclosure rule about the prospects. Then a prospect is

drawn at random and a signal about it is shown to the Receiver according to the rule. The Receiver then makes a rational inference about the prospect's value from the disclosed signal and chooses whether to accept the prospect (click on the ad, buy the product, invest in the bond, hire the student) or to reject it.

The problem of designing the optimal information disclosure rule turns out to be amenable to elegant analysis under the special assumption that the Receiver's private reservation value (equivalently, opportunity cost of accepting a prospect) is drawn from a uniform distribution, with support normalized to the interval $[0,1]$. Then the probability of a prospect's acceptance simply equals the Receiver's expectation of its value. For convenience, we also assume that the distribution from which prospects come is finite-valued, and that the Sender can randomize in sending signals.¹ Under these assumptions, we characterize the optimal disclosure rule. In particular, we establish that this rule must have the following properties:

- It is optimal to pool two prospects (i.e., to send the same signal for each of them with a positive probability) when they are “non-ordered” (i.e. one has a higher and lower value than the other). When two prospects “ordered” (i.e. one dominates the other in both value and profit), it is never optimal to pool them.
- When we describe each potential signal shown to the Receiver by the prospect's expected profit and expected value conditional on the signal, the set of such signals must be ordered, i.e., for any two signals, one must dominate the other in both value and profit.
- Any set of prospects that are pooled with each other (i.e., result in the same signal) with a positive probability must lie on a straight line in the (profit,value) space. For the “generic” case in which no three prospects are on the same line, this implies that any signal can pool at most two prospects.
- Two intervals connecting pooled prospects cannot intersect in the (profit,value) space.
- When one prospect is higher than another in both value and profit, it can only be pooled into a higher signal than the other.

¹We believe that such randomization would become unnecessary with a continuous convex-support distribution of prospects, but the full analysis of such a case is technically more challenging and we have not attempted it yet

- In the “generic” case, the set of prospects can be partitioned into three subsets: “profit” prospects, “value” prospects, and “isolated” prospects, so that any possible pooling involves one “profit” prospect and one “value” prospect, with the “profit” prospect having a higher profit and a lower value than the “value” prospect it is pooled with. Each “profit” or “value” prospect is pooled with some other prospect with probability 1, whereas each “isolated” prospect is never pooled.

While these results tell us a great deal about the optimal disclosure mechanism, they do not fully describe it: They still leave many ways to choose the pooling graph, and to choose probabilities with which a given prospect is pooled with its potential pooling partners. Fortunately, the Sender’s expected profit-maximization problem with these probabilities turns out to have a concave objective function and linear constraints (that the probabilities add up to 1). Its solution can then be characterized by the first-order conditions, which we derive. A complication arises due to the fact that the objective function is not differentiable in the probabilities of pooling with a given signal when this signal has probability zero. This matters because typically only a subset of signals can have positive probabilities at a solution. One way to overcome this problem is by trying different subset of signals (pooling pairs), using first-order conditions to find optimal probabilities of pooling into these signals, and then choose the subset with the maximal expected profits. Another way is by first solving the perturbed maximization problem subject to the additional constraint that each prospect is pooled with each potential partner with probability at least epsilon, and then taking epsilon to zero to approach the solution to the unconstrained problem.

In the general analysis we take the profitability of each prospect to the Sender as given. Yet we can apply this analysis to applications in which the Sender is an intermediary between the Receiver and an independent Advertiser who owns the prospects. The Sender’s mechanism design problem then includes the design of payments that the Advertiser is charged for the signal about his prospect that is shown to the Receiver. For example, an online advertising platform charges advertisers different payments for different signals (ad placement etc.). In the extreme case where the Sender has full information about the Advertiser’s profits, he can charge him payments that extract these profits fully, in which case the disclosure rule design problem becomes the same as if she owned the prospects. But we also consider the more interesting case in which the Advertiser has private information about the prospects’ profitability to them. For example, online advertisers may have private information about their per-click profits, and so any mechanism designed by the platform will leave advertis-

ers with some information rents. By subtracting these rents from the total profits, we can calculate the profits collected by the platform as the Advertiser’s “virtual profits,” which is the part of his profits that can be appropriated by the platform.

We consider an application with this structure, where the Advertiser’s private information is his per-click profit θ . In addition, there is a signal ρ of the Advertiser’s relevance that is observed both by the Sender and the Advertiser. The prospect’s value to consumers is given by a function $v(\theta, \rho)$, which allows for the Advertiser’s private information to affect the prospect’s value to the Receiver. The Sender (e.g., an advertising platform) offers a mechanism to the Advertiser, which without loss can be a direct revelation mechanism: the Advertiser reports his profits θ (e.g., through his bid per click), which together with the relevance parameter ρ determines the probability distribution over the signals revealed to the Receiver about the prospect, as well as the Advertiser’s payment to the Sender. We then argue through a simple example that this model may help account for some stylized features of internet advertising.

Finally, we consider a few extensions of the model. First, if instead of a monopolistic platform there are several platforms competing for users, we may expect a different Pareto optimal information disclosure rule to emerge, which maximizes a weighted sum of expected profits and consumer welfare. The problem of maximizing this weighted sum is mathematically equivalent to the original problem, upon a linear change of coordinates. As the relative weight on consumer welfare goes up, the optimal rule becomes more revealing (in the limit, it becomes fully revealing).

The second perturbation is to allow the platform to offer monetary subsidies or taxes per click. We find that given the optimal choice of subsidies/taxes, it becomes optimal to have a fully revealing disclosure rule.

Finally, we have assumed that Receiver has a uniformly distributed private reservation utility. This is a very special distributional assumption. (Although similar assumptions have proven necessary to obtain tractable results in other communication models, such as Crawford-Sobel, 1982, and Athey-Ellison, 2008). For general distributions, the mathematical problem becomes a lot more challenging, but we show that several of our results continue to hold.

2 Related Literature

There exists a large literature on communicating information in Sender-Receiver games, using costly signals such as education (Spence 1977) or advertising (Nelson, 1974, Kihlstrom and Riordan, 1984), disclosure of verifiable information (e.g., Milgrom’s 2008 survey), or cheap talk (Crawford and Sobel, 1982). Our approach is distinct from this literature in two key respects: (1) our Sender is able to commit to a disclosure rule (thus, formally, we consider the Stackelberg equilibrium rather than the Nash equilibrium of the game), and (2) our Sender has two-dimensional rather than one-dimensional private information. These differences fundamentally alter information disclosure outcomes.

We believe that commitment to an information disclosure rule is a sensible assumption in the applications discussed in the introduction. We can view the Sender as a “long-run” player facing a sequence of “short-run” Receivers. In such a repeated game, a patient long-run player will be able to develop a reputation to play its Stackelberg strategy, provided that enough information is revealed concerning history of play (Fudenberg and Levine 1989). While an internet advertising platform may be tempted in the short run to fool users into clicking more profitable ads by overstating their relevance, pursuing this strategy would be detrimental to the platform’s long-run profits.

Several papers have considered commitment to optimal disclosure policy by an auctioneer using a given auction format (Milgrom and Weber 1982) and by a monopolist designing an optimal price-discrimination mechanism (Ottaviani and Prat 2001). The literature has focused on providing sufficient conditions for full information disclosure to be optimal (e.g., using the “linkage principle”). In our main model, the Sender does not have a pricing choice, and full information disclosure is generally not optimal.

Another literature related to this paper is that on certification intermediaries, starting with Lizzeri (1999). In Lizzeri’s basic model, the certification intermediary is able to capture the whole surplus by revealing either no information, or just enough information for consumers to make efficient choices. The ability to extract consumer surplus is due to the assumption that consumers have no private information (the demand curve is perfectly elastic), as well as the ability to vary the price to consumers (which is not present in our main analysis). The ability to extract producer surplus is due to the producers having no informational advantage over the intermediary (in contrast to our application in which prospects’ profitability is their private information). The key feature distinguishing our model from this literature is the two-dimensional space of prospects: they differ not just in their value to

consumers but in their profitability (equivalently, costs) to sellers. The key new conclusion in this two-dimensional space is that we have partial information disclosure and partial pooling in specific directions. Adding some price flexibility to our model (such as per-click subsidies or taxes considered in Subsection 8.2) may make it more appropriate for some applications.

Our model is also closely related to Rayo (2005), who examines the optimal mechanism for selling conspicuous goods (such as luxury watches, pens, jewelry, or cars) whose only purpose is assumed to be signaling of wealth. This is parallel to our model once we interpret the seller as the Sender, consumers as prospects, and conspicuous goods as signals. (The Receiver is present in Rayo's model in reduced form only, whose beliefs enter the consumer's utility.) The main difference from our model is again in the dimension of the type space: in Rayo's model type is one-dimensional and prospects/consumers who have a higher type (i.e., higher value) are also the ones for whom signaling a higher type is more profitable.

Athey and Ellison (2008) examine the inferences of rational users about ad relevance in sponsored-search position auctions, and the design of auctions that take these inferences into account. While they are concerned with similar issues, their model is different in a number of respects. First, they do not consider general information disclosure mechanisms, instead focusing on auctions where the only signal of an ad's relevance is its ranking on the screen. Second, they focus on one-dimensional advertiser type with the sorting condition in the right direction, so that the more profitable ads are also more valuable to users. On the other hand, they focus on some aspects of sponsored-search auctions that we abstract from, in particular on users' short-run learning about the relevance of a given ad panel by clicking on other ads, and the externalities among ads due to substitution among their products (a consumer who finds a match on one advertised website does not click on any more ads).

In recent work, Kamenica and Gentzkow (2009) and Ostrovsky and Schwarz (2008) consider games in which a Sender with commitment power influences a rational Receiver through his choice of information disclosure. Kamenica and Gentzkow find conditions under which such influence is desirable for the Sender, while Ostrovsky and Schwarz study the impact of disclosure over unraveling in matching markets. In contrast to our work, these papers consider the case in which the Sender alone has private information, and their main focus is not the case of two-dimensional private information for the Sender.

3 Setup

We begin with two players: the Sender and the Receiver. The Sender is endowed with a prospect, which is randomly drawn from a finite set $P = \{1, \dots, N\}$. The probability of prospect i being realized is denoted by $p_i > 0$, with $\sum_{i \in P} p_i = 1$. Each prospect $i \in P$ is characterized by its *payoffs* $(\pi_i, v_i) \in \mathbb{R}^2$, where π_i is the prospect's profitability to the Sender and v_i is its value to the Receiver.

The realized prospect is not directly observed by the Receiver. Instead, the Receiver is shown a signal about this prospect, according to an information disclosure rule:

Definition 1 A “disclosure rule” $\langle \sigma, S \rangle$ consists of a finite set S of signals² and a mapping $\sigma : P \rightarrow \Delta(S)$ that assigns to each prospect i a probability distribution $\sigma(i) \in \Delta(S)$ over signals.

For example, at one extreme, the “full separation” rule is implemented by taking the signal space $S = P$ and the disclosure rule $\sigma_s(i) = 1$ if $s = i$ and $\sigma_s(i) = 0$ otherwise. At the other extreme, the “full pooling” rule is implemented by letting S be a singleton.

After observing the signal s , the Receiver decides whether to “accept” ($a = 1$) or “not accept” ($a = 0$) the prospect. Whenever the Receiver accepts the prospect, he forgoes an outside option worth $r \in \mathbb{R}$. Thus, the Sender and Receiver obtain payoffs equal respectively to $a\pi$ and $a(v - r)$.

We assume the Sender commits to a disclosure rule before the prospect is realized. Thus, the timing is as follows:

1. The Sender chooses a disclosure rule $\langle \sigma, S \rangle$.
2. A prospect $i \in P$ is drawn.
3. A signal $s \in S$ is drawn from distribution $\sigma(i)$ and shown to the Receiver.
4. The Receiver accepts or rejects the prospect.

Example 1 A search engine (the Sender) shows a consumer (the Receiver) an online advertisement with a link. Based on the characteristics of this advertisement (s), and his own opportunity cost (r), the consumer decides whether or not to click on the link. The online

²The restriction to a finite set of signals is without loss of generality in this setting.

advertisement, for instance, may describe a product sold by a separate firm, in which case the search engine's payoff (π) may correspond to a fee paid by such firm. We consider this possibility in greater detail in Section 7.

In principle, the Sender may be able to “exclude” a prospect (e.g., by not showing it to the Receiver at all), thus enforcing acceptance decision $a = 0$. For expositional simplicity for now we do not consider this possibility. Our analysis will thus apply conditional on the probability distribution of the prospects that are not excluded. (And when all prospects have nonnegative profits, the Sender will indeed find it optimal not to exclude any of them.) We explicitly introduce optimal exclusion decisions below in Section 7.

The Receiver knows his outside option r but must make an inference about the prospect's value v from the signal s shown to him. Accordingly, the Receiver optimally sets $a = 1$ if and only if his expected value conditional on the signal, $\mathbb{E}[v|s]$, exceeds r . In what follows, we assume that r is drawn from a uniform distribution with support normalized to $[0, 1]$. For simplicity, we also restrict v to be in $[0, 1]$. Under these assumptions, conditional on receiving a given signal s , the probability that $a = 1$ (the Receiver's “acceptance rate”) is equal to $\mathbb{E}[v|s]$. The resulting expected surplus obtained by the Receiver is given by

$$\int_0^1 \max \{ \mathbb{E}[v|s] - r, 0 \} dz = \frac{1}{2} \mathbb{E}[v|s]^2.$$

As for the Sender, her expected profit from signal s being accepted is $\mathbb{E}[\pi|s]$, hence her expected profit from sending the signal is $\mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s]$. Taking the ex ante expectation over the signals, the Receiver and Sender's expected payoffs for a given disclosure rule can be written, respectively, as

$$U_R = \mathbb{E} \left[\frac{1}{2} \mathbb{E}[v|s]^2 \right], \tag{1}$$

$$U_S = \mathbb{E} \left[\mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s] \right]. \tag{2}$$

Observe that for the purposes of computing the parties' payoffs, an information disclosure rule $\langle \sigma, S \rangle$ is characterized by the total probabilities $q_s = \sum_{i \in P} p_i \sigma_s(i)$ that signals $s \in S$ are sent as well as the parties' posterior expected payoffs conditional on these signals:

$$\mathbb{E}[v|s] = \frac{1}{q_s} \sum_{i \in P} p_i \sigma_s(i) v_i, \quad \mathbb{E}[\pi|s] = \frac{1}{q_s} \sum_{i \in P} p_i \sigma_s(i) \pi_i.$$

Thus, showing the Sender a signal s is equivalent to showing him a single disclosed prospect with payoffs $(\mathbb{E}[\pi|s], \mathbb{E}[v|s])$. This observation will prove useful in analyzing optimal disclosure rules.

Note, in particular, that if we have two different signals with the same expected payoffs $(\mathbb{E}[\pi|s], \mathbb{E}[v|s])$, they can be merged into one signal with their combined probability. Thus, we can restrict attention without loss to disclosure rules that are *non-redundant*, i.e., where different signals have different expected payoffs $(\mathbb{E}[\pi|s], \mathbb{E}[v|s])$, and all signals are sent with positive probabilities. We will also view different disclosure rules that coincide up to a relabeling of signals as equivalent. For example, we can then say that there is a unique (non-redundant) full-separation rule and a unique (non-redundant) full-pooling rule.

Now we consider the effect of information disclosure on the two parties' payoffs. As far as the Receiver is concerned, it is clear that the more information is disclosed to him, the higher is his expected payoff. Thus, the Receiver's expected payoff is maximized by full separation rule, which gives him a payoff of $\mathbb{E}[\frac{1}{2}v^2]$. One way to see this is using Jensen's inequality. Namely, for any disclosure rule,

$$\mathbb{E} \left(\frac{1}{2} \mathbb{E}[v|s]^2 \right) \leq \mathbb{E} \left(\frac{1}{2} \mathbb{E}[v^2|s] \right) = \mathbb{E} \left[\frac{1}{2} v^2 \right].$$

At the other extreme, under full pooling, the Receiver's expected payoff is only $\frac{1}{2}\mathbb{E}[v]^2$. Again by Jensen's inequality, this is the smallest possible payoff among all disclosure rules:

$$\frac{1}{2}\mathbb{E}[v]^2 = \frac{1}{2} [\mathbb{E}(\mathbb{E}[v|s])]^2 \leq \frac{1}{2}\mathbb{E} (\mathbb{E}[v|s]^2).$$

We now turn to the problem of choosing the disclosure rule to maximize the *Sender's* expected payoff, which proves to be substantially more complicated and which in general is not solved by either full separation or full pooling.

4 Characterizing Profit-Maximizing Disclosure

The goal is to find a disclosure rule that maximizes the expected product of the two coordinates $\mathbb{E}[\pi|s]$ and $\mathbb{E}[v|s]$:

$$\mathbb{E} (\mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s]). \tag{3}$$

We begin with a simple exercise that will then be used as a key building block for the

analysis. The Sender's expected gain from pooling two prospects i and j into one signal \hat{s} (while disclosing information about the other prospects as before) is given by:

$$\begin{aligned}
& (p_i + p_j) \mathbb{E}[\pi_k | k \in \{i, j\}] \cdot \mathbb{E}[v_k | k \in \{i, j\}] - p_i \pi_i v_i - p_j \pi_j v_j \\
&= (p_i + p_j) \cdot \frac{p_i \pi_i + p_j \pi_j}{p_i + p_j} \cdot \frac{p_i v_i + p_j v_j}{p_i + p_j} - p_i \pi_i v_i - p_j \pi_j v_j \\
&= -\frac{p_i p_j}{p_i + p_j} (\pi_i - \pi_j)(v_i - v_j).
\end{aligned} \tag{4}$$

Thus, we see that the profitability of pooling two prospects depends on how their payoffs are ordered:

Definition 2 *Two prospects' payoffs $(\pi_i, v_i), (\pi_j, v_j) \in \mathbb{R}^2$ are ordered if either $(\pi_i, v_i) \leq (\pi_j, v_j)$ or $(\pi_j, v_j) \leq (\pi_i, v_i)$. The two prospects are unordered if $(\pi_i, -v_i) \leq (\pi_j, -v_j)$ or $(\pi_j, -v_j) \leq (\pi_i, -v_i)$. The two prospects are strictly ordered if they are ordered and not unordered; they are strictly unordered if they are unordered and not ordered.*

Examination of (4) immediately yields:

Lemma 1 *Pooling two prospects yields (strictly) higher profits for the Sender than separating them if the prospects are (strictly) unordered, and yields (strictly) lower profits if the prospects are (strictly) ordered.*

A simple intuition for this result is that pooling two prospects preserves the expected acceptance rate but shifts it from the more valuable to the less valuable prospect. When the more valuable prospect is also more profitable (the "ordered" case), this shift reduces the Sender's expected profits. When instead the more valuable prospect is less profitable (the "unordered" case), this shift raises the Sender's expected profits.

A more formal explanation for the same result comes from examining the curvature of the product function in different directions $(\Delta\pi, \Delta v)$:

$$\frac{d^2}{dt^2} [(\pi + t\Delta\pi)(v + t\Delta v)] = 2\Delta\pi\Delta v,$$

and so the function is convex when $\text{sign}(\Delta\pi) = \text{sign}(\Delta v)$, but concave when $\text{sign}(\Delta\pi) \neq \text{sign}(\Delta v)$. Thus, by Jensen's inequality, full separation is optimal in the directions of convexity while full pooling is optimal in the directions of concavity.

This simple observation has far-reaching implications for the optimal disclosure rule with any number of prospects. The simplest one is

Lemma 2 *In a profit-maximizing disclosure rule, the set of the signals' payoffs*

$$\{(\mathbb{E}[\pi|s], \mathbb{E}[v|s]) : s \in S\}$$

is ordered (i.e., any two of its elements are ordered).

Proof. If there were two signals $s_1, s_2 \in S$ sent with positive probabilities such that $(\mathbb{E}[\pi|s_1], \mathbb{E}[v|s_1])$ and $(\mathbb{E}[\pi|s_2], \mathbb{E}[v|s_2])$ are not ordered, then by Lemma 1 the expected profits would be increased by pooling these two signals into one. ■

For example, if we just have $N = 2$ prospects, and the prospects are strictly unordered, optimal information disclosure must involve full pooling, for otherwise there would be two strictly unordered signals. If we instead have $N = 2$ prospects that are strictly ordered, optimal information disclosure must involve full separation, for otherwise by Lemma 1 we would improve by breaking any signal into separation of the prospects.

For cases with more than two prospects, characterization of optimal disclosure requires more work, as it is typically neither full separation nor full pooling.

Definition 3 *The pool of a signal $s \in S$ is the set of prospects for which signal s is sent with positive probability, i.e.,*

$$P_s = \{i \in P : \sigma_s(i) > 0\}.$$

Lemma 3 *In a profit-maximizing disclosure rule, for any given signal $s \in S$, the payoffs of the prospects in the pool of s , $\{(\pi_i, v_i) : i \in P_s\}$, lie on a straight line with a nonpositive slope.³*

Proof. (See Figure 1 for reference.) First, suppose in negation that the payoffs do not lie on a straight line. Then the convex hull of $\{(\pi_i, v_i) : i \in P_s\}$, which we denote by H , has a

³Note that it is important for this lemma, unlike the previous results, that randomized disclosure rules be allowed. By virtue of this lemma, allowing for randomization actually simplifies the characterization of optimal information disclosure, contrary to what one might expect a priori. We expect that randomization becomes superfluous when the prospects are drawn from a continuous distribution on a convex set; however, analysis of such a case requires different techniques and is not undertaken here.

nonempty interior, which contains $(\mathbb{E}[\pi|s], \mathbb{E}[v|s])$. Then H also contains $(\mathbb{E}[\pi|s], \mathbb{E}[v|s]) - (\delta, \delta)$ for small enough $\delta > 0$, i.e., there exists $\lambda \in \Delta(P_s)$ such that

$$(\mathbb{E}[\pi|s], \mathbb{E}[v|s]) - (\delta, \delta) = \sum_{i \in P_s} \lambda_i \cdot (\pi_i, v_i).$$

Now replace the original signal s with two new signals s_1, s_2 and consider the new disclosure rule $\hat{\sigma}$ that for each $i \in P_s$ has $\hat{\sigma}_{s_1}(i) = \varepsilon \lambda_i$ and $\hat{\sigma}_{s_2}(i) = \sigma_s(i) - \varepsilon \lambda_i$, where $\varepsilon > 0$ is chosen small enough so that $\hat{\sigma}_{s_2}(i) \geq 0$ for all $i \in P_s$. (Let $\hat{\sigma}_t(i) = \sigma_t(i)$ for all $t \in S \setminus \{s\}$, $i \in P \setminus P_s$.) By construction, we obtain

$$\begin{aligned} (\mathbb{E}[\pi|s_1], \mathbb{E}[v|s_1]) &= (\mathbb{E}[\pi|s], \mathbb{E}[v|s]) - (\delta, \delta) \text{ and} \\ \frac{\varepsilon}{q_s} \cdot (\mathbb{E}[\pi|s_1], \mathbb{E}[v|s_1]) + \frac{q_s - \varepsilon}{q_s} \cdot (\mathbb{E}[\pi|s_2], \mathbb{E}[v|s_2]) &= (\mathbb{E}[\pi|s], \mathbb{E}[v|s]). \end{aligned}$$

This in turn implies

$$(\mathbb{E}[\pi|s_2], \mathbb{E}[v|s_2]) = (\mathbb{E}[\pi|s], \mathbb{E}[v|s]) + \frac{\varepsilon}{q_s - \varepsilon} (\delta, \delta).$$

Thus, the points $(\mathbb{E}[\pi|s_1], \mathbb{E}[v|s_1])$ and $(\mathbb{E}[\pi|s_2], \mathbb{E}[v|s_2])$ are strictly ordered, and by Lemma 1 the expected profit from separating signals s_1 and s_2 is strictly higher than the expected profit from pooling them into one signal s . This contradicts the optimality of the original disclosure rule.

Finally, if the straight line containing P_s were strictly upward sloping, then by Lemma 1 the expected profits would be increased by breaking up signal s into full separation. ■

Lemma 4 *In a profit-maximizing disclosure rule σ , suppose we have prospects a_1, a_2, b_1, b_2 and signals s_1, s_2 such that: $a_1, b_1 \in P_{s_1}$, $a_2, b_2 \in P_{s_2}$, and the payoffs of these prospects do not lie on the same line. Then, the intervals⁴*

$$[(\pi_{a_1}, v_{a_1}), (\pi_{b_1}, v_{b_1})] \text{ and } [(\pi_{a_2}, v_{a_2}), (\pi_{b_2}, v_{b_2})]$$

can only intersect if they share an end point.

Proof. See Figure 2 for intuition. Suppose, in negation, that the above intervals intersect at point z , and this point lies in the interior of at least one of the intervals. For $j = 1, 2$, let

⁴Where $[x, y] = \{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}$.

$\lambda_j \in [0, 1]$ be such that

$$\lambda_j (\pi_{a_j}, v_{a_j}) + (1 - \lambda_j) (\pi_{b_j}, v_{b_j}) = z.$$

Since $\lambda_j \in (0, 1)$ for some j , we may assume without loss that $\lambda_1 \in (0, 1)$.

Now consider a new disclosure rule $\hat{\sigma}$ that is identical to σ with the following exception: for all $j \neq k = 1, 2$,

$$\begin{aligned} \hat{\sigma}_{s_j}(a_j) &= \sigma_{s_j}(a_j) - \varepsilon \lambda_j, & \hat{\sigma}_{s_j}(b_j) &= \sigma_{s_j}(b_j) - \varepsilon(1 - \lambda_j), \\ \hat{\sigma}_{s_j}(a_k) &= \sigma_{s_j}(a_k) + \varepsilon \lambda_k, & \hat{\sigma}_{s_j}(b_k) &= \sigma_{s_j}(b_k) + \varepsilon(1 - \lambda_k), \end{aligned}$$

where $\varepsilon > 0$ is chosen small enough so that $\hat{\sigma}_{s_j}(a_j)$ and $\hat{\sigma}_{s_j}(b_j)$ are positive.

By construction, $\hat{\sigma}$ and σ place the same total probability on every signal. In addition, the posterior payoffs for the affected signals s_j are identical under both rules:

$$\begin{aligned} \mathbb{E}_{\hat{\sigma}}[\pi, v | s_j] &= \frac{1}{q_{s_j}} \sum_{i \in P} \hat{\sigma}_{s_j}(i) (\pi_i, v_i) \\ &= \frac{1}{q_{s_j}} \sum_{i \in P} \sigma_{s_j}(i) (\pi_i, v_i) - \frac{\varepsilon}{q_{s_j}} \{ \lambda_j (\pi_{a_j}, v_{a_j}) + (1 - \lambda_j) (\pi_{b_j}, v_{b_j}) \} \\ &\quad + \frac{\varepsilon}{q_{s_j}} \{ \lambda_k (\pi_{a_k}, v_{a_k}) + (1 - \lambda_k) (\pi_{b_k}, v_{b_k}) \} = \mathbb{E}_{\sigma}[\pi, v | s_j], \end{aligned}$$

where $q_{s_j} = \sum_{i \in P} \sigma_{s_j}(i) = \sum_{i \in P} \hat{\sigma}_{s_j}(i)$ and the last equality above follows from the fact that both expressions in braces are equal to z .

As a result, $\hat{\sigma}$ delivers the same payoff for the Sender as σ , and is therefore optimal. Nevertheless, since $\lambda_1 \in (0, 1)$, the pool of signal s_2 now contains all four prospects, which is a contradiction to lemma 3. ■

Lemma 5 *In any optimal disclosure rule, for any two signals $s, s' \in S$ and any two prospects $i \in P_s, i' \in P_{s'}$, whenever $(\pi_{i'}, v_{i'}) \geq (\pi_i, v_i)$ and the two points do not coincide, either $E[(\pi, v) | s'] \geq E[(\pi, v) | s]$, or it is optimal to pool the two signals.*

Proof. By Lemma 2, we must either have $E[(\pi, v) | s'] \geq E[(\pi, v) | s]$ or $E[(\pi, v) | s'] \leq E[(\pi, v) | s]$. Suppose in negation that (a) the former inequality does not hold, hence the latter one holds, and (b) it is not optimal to pool the two signals, hence $E[(\pi, v) | s'] \neq E[(\pi, v) | s]$.

Let

$$\begin{aligned} x &= (\pi_i, v_i), x' = (\pi_{i'}, v_{i'}), y = E[(\pi, v) | s], y' = E[(\pi, v) | s'], \\ L &= \{\lambda x + (1 - \lambda)y | \lambda \in \mathbb{R}\}, L' = \{\lambda x' + (1 - \lambda)y' | \lambda \in \mathbb{R}\}. \end{aligned}$$

By Lemma 3, both L and L' must have a non-positive slope. If $L = L'$, then by Lemma 1 it is optimal to pool signals s and s' . Thus we can focus on the case where $L \neq L'$, hence the two lines have at most one intersection.

Since by assumption x' is strictly above L and y' is below L , there exists $C \in [x', y']$ such that $\{C\} = L \cap L'$. By the symmetric argument, we also have $C \in [x, y]$.

But then we can find $j \in P_s, j' \in P_{s'}$ such that the interiors of intervals $[(\pi_i, v_i), (\pi_j, v_j)]$ and $[(\pi_{i'}, v_{i'}), (\pi_{j'}, v_{j'})]$ intersect at C , and since we also know that the lines L, L' on which they lie do not coincide, by Lemma 4 this contradicts optimality of the disclosure rule. ■

We can further narrow down the structure of optimal pooling when we focus on the “generic” case:

Definition 4 *The problem is “generic” if no three prospects lie on the same straight line.*

In this case, Lemma 3 tells us that no more than two types can share a given signal s .⁵ Thus, any given signal s either fully reveals a specific prospect i , or, alternatively, it pools exactly two different prospects $\{i, j\}$. Then the disclosure rule induces a “pooling graph” on P , in which two prospects are linked if and only if they are pooled into one signal. (Note that by Lemma 2 it cannot be optimal to have two distinct signals that both pool two strictly unordered prospects, since then the two signals would themselves be strictly unordered.) In the generic case, we can say even more about the pooling graph:

Definition 5 *For two prospects $i, j \in P$, if $\pi_i \geq \pi_j$ and $v_i \leq v_j$ then we say that i is “to the SE” of j , and that j is “to the NW” of i .*

Lemma 6 *In the generic case, an optimal information disclosure rule partitions P into three subsets: the set V of “value prospects,” the set Π of “profit prospects,” and the set I of “isolated prospects,” so that for any signal s , the pool P_s consists either of a single prospect $i \in I$ or of two prospects $\{i, j\}$ with $i \in V$ and $j \in \Pi$, with i being to the NW of j .*

⁵Clearly, if we instead considered a continuous distribution of prospects, then this notion of genericity would not be appropriate, and we would typically expect many prospects to be pooled into one signal.

Proof. Observe that a given prospect i cannot be optimally pooled with a prospect i_{SE} to the SE of it and also with another prospect i_{NW} to the NW of it. Indeed, were this to happen, letting s_{SE} and s_{NW} represent the two respective signals, the respective posteriors $E[\pi, v|s_{SE}]$, $E[\pi, v|s_{SW}]$ would be strictly unordered (here also using genericity), and so by Lemma 2 this could not be an optimal disclosure rule.

Thus, for any given prospect i , there are just three possibilities: (i) it does not participate in any pools, in which case we assign i to I , (ii) all of its pooling partners are to the SE of i , in which case we assign it to V , and (iii) all of its pooling partners are to the NW of i , in which case we assign it to Π . ■

Intuitively, Lemma 6 allows us to interpret the prospects from set Π as “profit” prospects and those from set V as “value” prospects. A value prospect is always used as a “bait” to attract consumers, while a profit prospect is always used as a “switch” to exploit the attracted consumers. (Of course, since consumers are rational, they take the probability of being “switched” into account.) The substantive contribution of the lemma is in showing that the role of a pooled prospect in the optimal disclosure rule cannot change: it is either always used as the “bait” or always used as the “switch.”

Example 2 (Taxonomy of optimal pooling with 4 prospects) *Focusing on the case where all prospects are pooled ($I = \emptyset$), these are all the possibilities (see Figure 3):*

a) $|V| = 1, |\Pi| = 3$ or $|V| = 3, |\Pi| = 1$: 3 signals (“Fan”)

b) $|V| = |\Pi| = 2$:

b.1) 2 signals, 1-to-1 pooling between V and Π (“Two lines”)

b.2) 3 signals (“Zigzag”)

b.3) 4 signals (“Cycle”)

Furthermore, it turns out that cycles are “fragile:” they can only be optimal for non-generic parameter combinations, and even for such combinations there exists another optimal pooling graph that does not contain cycles.

5 Solving for Optimal Disclosure

The lemmas in the previous section tell us a great deal about the optimal disclosure rule, but do not fully nail it down. In this section we discuss how to solve for the optimal rule. For

simplicity we restrict attention to the generic case, in which, by Lemma 3, we can restrict attention to signals that either pool a pair of prospects or separate a prospect. Thus, we can take $S = \{s \subset P : |s| = 1 \text{ or } |s| = 2\}$, where a single-element signal $\{i\}$ separates prospect i while a two-element signal $\{i, j\}$ is a pool of prospects i and j .

One way to describe such a disclosure rule is by defining, for any two-element signal $\{i, j\} \subset P$, the weight $\beta_{ij} = p_i \sigma_{\{i,j\}}(i)$ – i.e., the mass of point i that is pooled into signal $\{i, j\}$. Given these weights, we can calculate the Sender’s expected payoff (3) as follows. For each signal $\{i, j\}$ that is sent with a positive probability (i.e., $\beta_{ij} + \beta_{ji} > 0$), the expected payoff from using this signal relative to that from breaking it up into separation can be obtained using formula (4), substituting into it $p_i = \beta_{ij}$ and $p_j = \beta_{ji}$. Thus, the seller’s expected payoff can be written as

$$F(\beta) = \sum_{i \in P} p_i \pi_i v_i - \sum_{\{i,j\} \subset S} g(\beta_{ij}, \beta_{ji}) Z_{ij}, \quad (5)$$

where $g(b_1, b_2) = \begin{cases} b_1 b_2 / (b_1 + b_2) & \text{if } b_1 + b_2 > 0, \\ 0 & \text{otherwise,} \end{cases}$

and $Z_{ij} = (v_j - v_i)(\pi_j - \pi_i)$ for all $i, j \in P$.

The Sender will choose nonnegative weights to maximize this function subject to the constraints

$$\sum_{\{j,j\} \subset P} \beta_{ij} \leq p_i \text{ for all } i \in P,$$

$$\beta_{ij} \geq 0 \text{ for all } \{i, j\} \subset P.$$

(When the first constraint holds with strict equality for some prospect i this means that with the remaining probability the prospect is separated.)

Furthermore, note that the seller strictly prefers not to use any signals $\{i, j\}$ for which $Z_{ij} > 0$, i.e., that are strictly ordered. (This is also clear from Lemma 1.) Thus, we can restrict attention to pools from the set

$$U = \{\{i, j\} \subset P : Z_{ij} \leq 0\}.$$

Thus, the Sender's program can be written as

$$\max_{\beta \in \mathbb{R}^U} \sum_{i \in P} p_i \pi_i v_i - \sum_{\{i,j\} \in U} g(\beta_{ij}, \beta_{ji}) Z_{ij}, \text{ s.t.} \quad (6)$$

$$\sum_{\{j\} \leq U} \beta_{ij} \leq p_i \text{ for all } i \in P, \quad (7)$$

$$\beta_{ij} \geq 0 \text{ for all } (i, j) \in U. \quad (8)$$

Lemma 7 *The objective function in (6) is continuous and concave on \mathbb{R}_+^U .*

Proof. For continuity, it suffices to show that the function $g(b_1, b_2)$ is continuous on $(b_1, b_2) \in \mathbb{R}_+^2$. Continuity at any point $(b_1, b_2) \neq (0, 0)$ follows from the fact that it is a composition of continuous functions. To see continuity at $(0, 0)$, note that

$$0 \leq g(b_1, b_2) \leq b_1, b_2, \text{ hence}$$

$$\lim_{b_1, b_2 \rightarrow +0} g(b_1, b_2) = 0 = g(0, 0)$$

For concavity, since $Z_{ij} \leq 0$ for all $\{i, j\} \in U$, it suffices to show that $g(b_1, b_2)$ is a concave function on \mathbb{R}_+^2 . We first show that it is concave on $\mathbb{R}_+^2 \setminus \{(0, 0)\}$ by expressing its Hessian at any $(b_1, b_2) \neq (0, 0)$ as

$$D^2 g(b_1, b_2) = \frac{2}{(b_1 + b_2)^3} \begin{pmatrix} -b_2^2 & b_1 b_2 \\ b_1 b_2 & -b_1^2 \end{pmatrix},$$

and noting that it is negative semidefinite. Moreover, since g is continuous at $(0, 0)$, its concavity is preserved when adding this point to the set. ■

The Lemma implies that the set of solutions to the above program is continuous and convex. We proceed to write first-order conditions for the this program. However, before doing so, a word of caution is in order: The function $F(\beta)$ proves non-differentiable in (β_{ij}, β_{ji}) at points where $\beta_{ij} = \beta_{ji} = 0$. Indeed, on the one hand, the partial derivative of F with respect to either β_{ij} or β_{ji} is zero at any such point. This is simply because raising one of the weights while holding the other at zero has no effect on the information disclosure rule. However, the directional derivative of F in any direction in which β_{ij} and β_{ji} are raised at once is not zero: in particular, it is positive when i and j are strictly unordered.

We can still make use of first-order conditions for program (6) in the variables β_{ij}, β_{ji} for signals $\{i, j\}$ such that $(\beta_{ij}, \beta_{ji}) \neq (0, 0)$, holding the set of such signals fixed at some $\hat{S} \subset U$. Letting λ_i denote the Lagrange multipliers with adding-up constraints (7), the first-order conditions can be written as:

$$\frac{\beta_{ji}^2}{(\beta_{ij} + \beta_{ji})^2} |Z_{ij}| \leq \lambda_i, \text{ with equality if } \beta_{ij} > 0. \quad (9)$$

In particular, for signals $\{i, j\}$ and $\{i, k\}$ to both be sent with a positive probability, we must have

$$\frac{1 + \beta_{ij}/\beta_{ji}}{1 + \beta_{ik}/\beta_{ki}} = \sqrt{\frac{Z_{ij}}{Z_{ik}}}.$$

Thus, one way to solve for an optimal disclosure rule is by trying different sets of signals $\hat{S} \subset U$, writing interior first-order conditions for all signals from \hat{S} to be sent with a positive probability, solving for the optimal disclosure rule given \hat{S} , and calculating the resulting expected profit of the Sender. Then we can choose the set \hat{S} that maximizes the Sender's expected profits. We can use Lemma 6 to narrow down the set of possible signal combinations that could be optimal. Still, when the set P of prospects is large, this procedure may be infeasible, since the set of possible signal combinations \hat{S} can grow exponentially with the number of prospects. For such cases, we propose an alternative approach: choose $\varepsilon > 0$ and add the constraints $\beta_{ij} + \beta_{ji} \geq \varepsilon$ for each $(i, j) \in U$. Within the constrained set, the objective function is totally differentiable, hence the solutions can be characterized by first-order conditions to the problem. Then, by taking ε to zero, we will approach a solution to the unconstrained program.

Finally, while so far we have not allowed the Sender to exclude prospects, it is easy to introduce this possibility, by letting the Sender choose any $p_i \leq \bar{p}_i$, where \bar{p}_i is the true probability of prospect i , and the prospect is therefore excluded with probability $\bar{p}_i - p_i$. Note that the Sender will never exclude a prospect with a positive profit, since it can always be separated from the others. But the Sender may choose not to exclude even some prospects with a negative profit, since if this prospect has a high value the Sender may benefit by pooling it with another, profitable prospect.

6 Cycles and Generic Uniqueness

We begin by noting that there is a trivial reason why problem (6) may have multiple solutions. Suppose there is an optimum β^* such that, for some prospect i , we have $\beta_{ji}^* = 0$ for all $j \in P(i)$. In this case, regardless of the values of β_{ij}^* , prospect i is isolated. For this reason, the values of β_{ij}^* become immaterial both in terms of the payoff for the Sender and the information transmitted to the Receiver, and can therefore be chosen arbitrarily (indeed, the function $g(\beta_{ij}, \beta_{ji})$ is zero whenever either one of its arguments is zero). In order to abstract from this specific source of multiplicity, we restrict attention to the subset of optima such that

$$\beta_{ij} = 0 \Leftrightarrow \beta_{ji} = 0 \text{ for all } i, j \in U. \quad (10)$$

Denote this subset of optima $\widehat{B} = \{\beta \in B^* : (10) \text{ holds}\}$. The following results establish properties of these optima. All proofs are collected in the Appendix (Section 9). We assume throughout that $Z_{ij} \neq 0$ for all $i, j \in U$ (a generic property).

Lemma 8 *The set of optima \widehat{B} is convex and compact. Thus, by the Krein-Milman theorem, it is the convex hull of its vertices.*

\widehat{B} may in principle include two types of optima: cyclic and acyclic. Figure 3 illustrates examples of each.⁶

Lemma 9 *An optimum $\beta \in \widehat{B}$ is acyclic if and only if it is a vertex of \widehat{B} .*

Under an additional generic property for the prospects' payoffs, the optimum is guaranteed to be unique, and therefore acyclic.

Condition 1 *For every subset of prospects $\{i_1, i_2, \dots, i_K\}$ with more than two elements we have*

$$\sum_{k=1}^K (-1)^k \sqrt{Z_{i_k, i_{k \bmod K+1}}} \neq 0.$$

Proposition 1 *There exists an acyclic optimum. Moreover, under Condition 1, \widehat{B} has a single element.*

⁶Formally, we say that β is cyclic if its pooling graph contains a cycle, namely, there exists a set of prospects $\{i_1, i_2, \dots, i_K\}$, with more than two elements, such that both $\beta_{i_k, i_{k \bmod K+1}}$ and $\beta_{i_{k \bmod K+1}, i_k}$ are strictly positive for every $k = 1, 2, \dots, K$.

7 An Independent Advertiser

Here we assume that the prospect is owned by a new player, called the Advertiser, rather than the Sender. This prospect is characterized by a parameter vector $y = (\theta, \rho)$ that is randomly drawn from a finite set $Y \subset \mathbb{R}^2$. The first component θ represents the profit obtained by the Advertiser if the prospect is accepted ($a = 1$). The second component ρ is a “relevance” parameter that, in combination with θ , determines the benefit obtained by the Receiver conditional on $a = 1$. This benefit, in particular, is given by a function $f(\theta, \rho) \in [0, 1]$, which with slight abuse of notation we denote $v(\theta, \rho)$.

The prospect’s profit parameter θ is privately observed by the Advertiser and its relevance parameter ρ is jointly observed by the Advertiser and the Sender. In this way, the Sender enjoys at least partial knowledge of v . (The Receiver observes neither θ nor ρ .) Let $h(\theta | \rho)$ denote the probability of θ conditional on ρ , with cumulative function $H(\theta | \rho)$.

The Sender sells a signal lottery to the Advertiser using a direct-revelation mechanism. For each value of ρ , this mechanism requests a report $\hat{\theta}$ of the Advertiser’s profitability θ and, based on this report, determines: (1) a lottery $\sigma(\hat{\theta}, \rho) \in \Delta(S)$ and (2) a monetary transfer $t(\hat{\theta}, \rho) \in \mathbb{R}$ from the Advertiser to the Sender. The goal of the Sender is to maximize expected revenues $\mathbb{E}[t(\theta, \rho)]$ subject to the relevant participation and incentive constraints.

The timing is as follows:

1. The Sender chooses a mechanism consisting of a disclosure rule $\sigma : Y \rightarrow \Delta(S)$ and a transfer rule $t : Y \rightarrow \mathbb{R}$.
2. The Advertiser draws prospect parameters $(\theta, \rho) \in Y$.
3. The Advertiser reports $\hat{\theta}$ and transfers $t(\hat{\theta}, \rho)$ to the Sender.
4. A signal $s \in S$ is drawn from distribution $\sigma(\hat{\theta}, \rho)$
5. The Receiver observes s and his reservation value r .
6. The Receiver accepts or rejects the prospect.

We assume that the Receiver has knowledge of the mechanism chosen by the Sender as well as the prior distribution of (θ, ρ) . Accordingly, for any given s , the Receiver’s acceptance rate is given by $\mathbb{E}[v | s] = \mathbb{E}[v(\theta, \rho) | s]$, where the expectation is taken over (θ, ρ) .

On the other hand, for any given mechanism, the net expected profit obtained by an Advertiser who is endowed with parameters (θ, ρ) , and who reports type $\hat{\theta}$, is given by

$$\theta \cdot \mathbb{E} \left[\mathbb{E} [v | s] | \sigma(\hat{\theta}, \rho) \right] - t(\hat{\theta}, \rho),$$

where the first expectation is taken over s according to the lottery $\sigma(\hat{\theta}, \rho)$. The participation and incentive constraints indicate, respectively, that this payoff must be non-negative and maximized at $\hat{\theta} = \theta$.

For any given ρ , the highest transfers that the Sender can obtain are determined by a binding participation constraint for the Advertiser with the lowest value of θ , and a binding downward-adjacent incentive constraint for all other Advertisers. Accordingly, the Sender's objective becomes

$$\mathbb{E} [t(\theta, \rho)] = \mathbb{E} (\mathbb{E} [\pi(\theta, \rho) | s] \cdot \mathbb{E} [v(\theta, \rho) | s]), \quad (11)$$

where $\pi(\theta, \rho)$ denotes the “virtual profit” that the Sender obtains from an Advertiser with parameters (θ, ρ) . This virtual profit is given by

$$\pi(\theta, \rho) = \theta - (\theta' - \theta) \frac{1 - H(\theta | \rho)}{h(\theta | \rho)}, \quad (12)$$

where θ' denotes the type immediately above θ (provided such a type exists) and $\frac{1 - H(\theta | \rho)}{h(\theta | \rho)}$ is the inverse hazard rate for θ .

In addition, the incentive constraints indicate that the Sender must restrict to disclosure rules $\langle \sigma, S \rangle$ that result in a monotonic allocation. Namely, for any given ρ , the expected probability that $a = 1$ must be a nondecreasing function of the Advertiser's profit parameter θ :

$$\mathbb{E} [\mathbb{E} [v | s] | \sigma(\theta, \rho)] \text{ is non-decreasing in } \theta \text{ for all } \rho. \quad (M)$$

Notice that, other than the monotonicity constraint, the Sender's problem of maximizing (11) is identical to the original problem of maximizing (3), where π and v are now simply indexed by $y = (\theta, \rho)$. Consequently, whenever the monotonicity constraint is slacked, all results derived in Section 5 apply. The following conditions guarantee that this constraint is in fact slacked:

Condition 2 $\pi(\theta, \rho)$ is increasing in θ for all ρ .

Condition 2 is automatically met when θ takes only two values, and is satisfied in general

when the distribution h has an increasing hazard rate and adjacent types θ, θ' are evenly spaced.

Condition 3 $v(\theta, \rho)$ is nondecreasing in θ for all ρ .

Condition 3 indicates that more profitable advertisers also deliver higher consumer surplus.⁷

Lemma 10 Under conditions 2 and 3 the monotonicity constraint (M) does not bind.

Proof. Consider a disclosure rule $\langle \sigma^*, S \rangle$ that maximizes (11) and is such that the posterior payoffs of all signals are strictly ordered (which is without loss for the Sender due to Lemma 2 and the fact that any pair of signals with posterior payoffs that are both ordered and unordered can be pooled without changing her objective). We show that such disclosure rule satisfies (M).

Suppose not. Then, for some ρ , there must exist a pair θ_1, θ_2 , with $\theta_1 < \theta_2$, such that

$$\mathbb{E}[\mathbb{E}[v | s] | \sigma^*(\theta_1, \rho)] > \mathbb{E}[\mathbb{E}[v | s] | \sigma^*(\theta_2, \rho)].$$

This inequality in turn implies that there exist two signals s_1, s_2 , with $\sigma_{s_1}^*(\theta_1, \rho), \sigma_{s_2}^*(\theta_2, \rho) > 0$, such that

$$\mathbb{E}[v | s_1] > \mathbb{E}[v | s_2]. \quad (13)$$

When combined with the fact that the posterior payoffs of all signals are strictly ordered, this inequality implies that

$$\mathbb{E}[\pi | s_1] > \mathbb{E}[\pi | s_2]. \quad (14)$$

On the other and, since v and π are, respectively, nondecreasing and increasing in θ , we have $v(\theta_1, \rho) \leq v(\theta_2, \rho)$ and $\pi(\theta_1, \rho) < \pi(\theta_2, \rho)$. But when combined with (13) and (14), these inequalities contradict Lemma 5. ■

⁷For instance, a more profitable advertiser may have a higher-quality product and therefore charge a higher price than his competitors. But this higher price may only partially capture the consumer's higher willingness to pay for the advertiser's product, therefore leaving more surplus in the consumer's hands. To see this more formally, consider a simple example, gracefully suggested by Michael Schwarz. Suppose each advertiser has a underlying private type t . When a consumer clicks on an ad, he draws a gross private value z for the advertiser's product, with z uniformly distributed over $[0, 1]$, and purchases this product when z exceeds its price. Assuming zero marginal costs, the optimal price for the advertiser is $t/2$. This price delivers expected profits $t/4$ for the advertiser (which correspond to θ in our model) and expected surplus $t/8$ for the consumer (which corresponds to v). As a result, profits and consumer value are positively related. (This type of example can also be extended to include a role for the relevance parameter ρ .)

7.1 A Stylized Application

In practice, online search engines typically display links to their search results in three broad categories: left-hand-side sponsored links, left-hand-side organic links (displayed immediately below the sponsored links), and right-hand-side sponsored links. The engine receives direct revenues from all sponsored links (which are auctioned off), but not from the organic ones (which are chosen based on a measure of consumer value). The links on the left normally enjoy a significantly higher acceptance rate (or clickthrough) than those on the right. Moreover, it is reasonable to assume that many consumers do not draw a sharp distinction between the top organic links and the sponsored links on the left.⁸ Indeed, search engines normally offer only mild visual distinction between the two, such as slight background shading. Thus, we can roughly interpret this scenario as the engine showing two types of signals: a high quality “left-hand-side” signal that is shared by top sponsored and organic links, and a low quality “right-hand-side” signal that includes only lower-revenue sponsored links.⁹

The following simple example illustrates how the model can provide a stylized rationale for the above practice:

Example 3 *Suppose the Sender (search engine) has three prospects. The first two prospects (1 and 2) represent advertisers that share the same value of ρ , but have different profit levels θ , with $\theta_1 > \theta_2$, and therefore, from eq. (12), deliver different virtual profits π for the Sender, with $\pi_1 > \pi_2$ (i.e., Condition 2 is met). Assume also that $\pi_2 > 0$. Moreover, suppose the associated consumer values v satisfy $v_1 > v_2$ (i.e., Condition 3 is also met). Finally, suppose the third prospect represents an organic website that does not purchase an ad from the Sender, so that $\pi_3 = 0$, but delivers high value for the consumer, with $v_3 > v_1, v_2$. Notice that the payoffs of prospects 1 and 2 are ordered, $(\pi_1, v_1) \gg (\pi_2, v_2)$, while the payoffs of prospect 3 lie to the NW of the first two, $\pi_3 < \pi_1, \pi_2$ and $v_3 > v_1, v_2$.*

Since Conditions 2 and 3 are met, Lemma 10 indicates that the monotonicity constraint is slacked and therefore the optimal disclosure policy solves problem (6).

Lemma 11 *The optimal disclosure rule for Example 3 involves two signals s_1 and s_2 . Advertiser $i = 1, 2$ is assigned signal s_i with probability one. The organic prospect, in contrast, serves as Bait and is randomly assigned one of the two signals (with possibly degenerate*

⁸We are grateful to Glenn Ellison for making this observation.

⁹Under this rough interpretation, we are ignoring any potential substitutability across links and we are also abstracting from the fact that the specific position in which a sponsored link is displayed (within a given side of the page) also has an important effect on clickthrough.

probabilities). Other things equal, the Bait is assigned the signal of advertiser i with a higher probability if: (1) this advertiser has a larger mass p_i , and (2) the payoffs of this advertiser are more unordered vis-a-vis the payoffs of the Bait (i.e., $|Z_{i3}|$ is large).

Proof. The optimal disclosure rule solves the Sender's problem (6) with $P = \{1, 2, 3\}$, $U = \{\{1, 3\}, \{2, 3\}\}$, and the payoffs (π_i, v_i) described in the example, so that Z_{13} and Z_{23} are strictly negative. Since prospects 1 and 2 have only one potential pooling partner each (prospect 3), we can set, without loss, $\beta_{13} = p_1$ and $\beta_{23} = p_2$, so that the full masses of prospects 1 and 2 are pooled, respectively, into the signals $\{1, 3\}$ and $\{2, 3\}$. Denote these signals s_1 and s_2 .

It remains only to find optimal weights β_{31} and β_{32} (with $\beta_{31} + \beta_{32} = p_3$), which indicate how the total mass of prospect 3 (the Bait) is distributed between s_1 and s_2 . The corresponding first-order conditions (9) for these two weights are:

$$\begin{aligned} \left(\frac{p_1}{\beta_{31} + p_1}\right)^2 \cdot |Z_{13}| &\leq \lambda_3, \text{ with equality if } \beta_{31} > 0, \text{ and} \\ \left(\frac{p_2}{\beta_{32} + p_2}\right)^2 \cdot |Z_{23}| &\leq \lambda_3, \text{ with equality if } \beta_{32} > 0. \end{aligned}$$

Depending on the parameter values, we have three possible types of solutions. First, if $\left(\frac{p_1}{p_3+p_1}\right)^2 \cdot |Z_{13}| \geq |Z_{23}|$, we obtain a corner solution in which $\beta_{31} = p_3$ and $\beta_{32} = 0$. In this case, the Bait is exclusively pooled with advertiser 1 into signal s_1 , and advertiser 2 receives his own signal s_2 . Second, if $|Z_{13}| \leq \left(\frac{p_2}{p_3+p_2}\right)^2 \cdot |Z_{23}|$, we obtain the opposite corner solution in which $\beta_{32} = p_3$ and $\beta_{31} = 0$. In this case, the Bait is exclusively pooled with advertiser 2 into signal s_2 , and advertiser 1 receives his own signal s_1 . Third, in all other cases, we obtain an interior solution with $\beta_{31}, \beta_{32} > 0$, so that the Bait shares part of his mass with each advertiser. In this case, both first-order conditions above hold with equality and we obtain:

$$\frac{1 + \beta_{31}/p_1}{1 + \beta_{32}/p_2} = \sqrt{\frac{|Z_{13}|}{|Z_{23}|}}.$$

Inspection of this expression delivers the last statement in the Lemma. ■

The two signals s_1 and s_2 in the Lemma can be interpreted as the “left-hand-side” and “right-hand-side” signals that are used in practice. Moreover, while the organic prospect in the example can in principle serve as Bait for both advertisers, it will be pooled exclusively with the high-profit advertiser whenever $|Z_{13}|$ is large relative to $|Z_{23}|$. This would occur,

for instance, when the difference in profitability between the two advertisers is sufficiently large.¹⁰ This pooling case can be interpreted as the “bundling” we observe in practice between the top organic links and the left-hand-side sponsored links.

While stylized, this example helps explain why, in practice, not all sponsored links are grouped together (for example, on the right) and also why search engines do not introduce a sharper visual distinction between the top organic and sponsored links on the left (for example, by introducing a third column for the high-revenue sponsored links). Indeed, the example tells us that if all sponsored links were grouped, advertisers that are likely to be ordered would be bundled, therefore reducing profits. And it also tells us that introducing a sharper distinction between organic and sponsored links on the left would make the organic links a less effective bait for these high-revenue sponsored links.

8 Extensions

8.1 Pareto Optimal Allocations

Here we consider the more general problem of maximizing a weighted average of expected Receiver surplus and expected Sender profit, rather than focusing on expected profit alone. The objective becomes

$$\lambda \mathbb{E} \left(\frac{1}{2} \mathbb{E} [v | s]^2 \right) + (1 - \lambda) \mathbb{E} (\mathbb{E} [\pi | s] \cdot \mathbb{E} [v | s]), \quad (15)$$

where $\lambda \in [0, 1]$ represents an arbitrary Pareto weight on the Receiver. For example, when facing competitive pressure, a platform (Sender) may wish to increase the welfare of each user (Receiver) in order to increase the total number of users that patronize this platform. As before, we can interpret the Sender as either the direct owner of each prospect or simply as an intermediary between an Advertiser and the Receiver.

From linearity of the expectation operator, the above objective can be expressed as

$$\begin{aligned} \mathbb{E} \left(\left[\frac{\lambda}{2} \mathbb{E} [v | s] + (1 - \lambda) \mathbb{E} [\pi | s] \right] \cdot \mathbb{E} [v | s] \right) = \\ \mathbb{E} \left(\mathbb{E} \left[\frac{\lambda}{2} v + (1 - \lambda) \pi \mid s \right] \cdot \mathbb{E} [v | s] \right). \end{aligned}$$

¹⁰Indeed, $|z_{i3}| = |v_i - v_3| \pi_i$, which is increasing in π_i .

It follows that the problem of maximizing (15) is mathematically equivalent to the original problem after a linear transformation of the prospect's payoffs (π, v) into the new payoffs $(\widehat{\pi}(\lambda), v)$, with $\widehat{\pi}(\lambda) = \frac{\lambda}{2}v + (1-\lambda)\pi$.¹¹ Graphically, as shown in Figure 4, we can think of this transformation as a horizontal shift of all payoffs toward a ray from the origin with slope 2, where the new payoffs correspond to a weighted average between the original payoffs (π, v) and the point $(\frac{1}{2}v, v)$.

The difference with the basic case in which $\lambda = 0$ is that the optimal rule becomes progressively more revealing as λ increases. For instance, in the extreme when $\lambda = 1$ the Sender cares exclusively about Receiver surplus and, therefore, full separation becomes optimal (i.e., all new payoffs lie on the ray with positive slope and therefore are strictly ordered).

For intermediate levels of λ , it may still be optimal to pool some pairs of prospects but not others. Let

$$\begin{aligned} Z_{ij}(\lambda) &= (\widehat{\pi}_j(\lambda) - \widehat{\pi}_i(\lambda))(v_j - v_i) \\ &= \frac{\lambda}{2}(v_j - v_i)^2 + (1-\lambda)(\pi_j - \pi_i)(v_j - v_i), \end{aligned}$$

so that the payoffs of any two prospects i and j are ordered if and only if $Z_{ij}(\lambda) \geq 0$. For instance, if the original payoffs of these prospects $((\pi_i, v_i)$ and $(\pi_j, v_j))$ were strictly ordered, it follows that the new payoffs are strictly ordered as well and therefore these prospects are never pooled with each other.

On the other hand, if the original payoffs of the prospects were unordered, then the new payoffs remain weakly unordered as long as $\lambda \in [0, \widehat{\lambda}_{ij}]$, where

$$\widehat{\lambda}_{ij} = \frac{(\pi_j - \pi_i)}{(v_i - v_j)} \left[1 + \frac{(\pi_j - \pi_i)}{(v_i - v_j)} \right]^{-1},$$

which is inversely proportional to the slope of the line connecting the original payoffs (π_i, v_i) and (π_j, v_j) , measured in absolute value. Beyond this critical value for λ , the payoffs of i and j become strictly ordered and therefore these prospects no longer constitute potential pooling partners for each other.

¹¹For the case in which the sender acts as an intermediary (section 7), the assumption that π is increasing and v is nondecreasing in θ remains sufficient for the monotonicity constraint to have slack. Indeed, when $\lambda < 1$, this assumption implies that $\widehat{\pi}(\lambda)$ is also increasing in θ (as required by Lemma 10), and when $\lambda = 1$ we obtain full separation, in which case the monotonicity constraint is automatically met.

Note that for the generic case we have $\widehat{\lambda}_{ij} \neq \widehat{\lambda}_{ik}$ for all $i \neq j \neq k$. Thus, as λ increases, each prospect i progressively loses its potential pooling partners j , one at a time, in inverse order of the slopes $\frac{(\pi_j - \pi_i)}{(v_i - v_j)}$.

8.2 Receiver Incentives

Here we return to the original problem of maximizing expected profits but we consider the case in which the Sender offers the consumer a subsidy (or tax) conditional on accepting the prospect. We allow the Sender to use a potentially different subsidy for each signal s . For expositional clarity, we begin with the case in which both π and v lie in $[0, 1]$, and then consider the case in which π can also be negative or larger than 1. For this section, it is also convenient to break any indifference in favor of separating two prospects.

For any given s , with posterior payoffs $\mathbb{E}[\pi, v \mid s]$, the optimal subsidy, denoted $\alpha(s)$, solves:

$$\max_{\alpha} \mathbb{E}[\pi - \alpha \mid s] \cdot \mathbb{E}[v + \alpha \mid s],$$

where the subsidy is added to consumer value (resulting in a higher acceptance rate) but is also subtracted from the Sender's profits. The solution to this problem is uniquely given by

$$\alpha(s) = \frac{1}{2} \mathbb{E}[(\pi - v) \mid s],$$

where a negative subsidy $\alpha(s) < 0$ corresponds to a tax. Accordingly, the net expected profit and acceptance rate are both equated to $\frac{1}{2} \mathbb{E}[(\pi + v) \mid s]$.

Substituting this solution in the objective, the optimized payoff for the Sender (conditional on s) becomes

$$U_S = \frac{1}{4} \mathbb{E}[(\pi + v) \mid s]^2.$$

This expression has a simple structure, as it is convex in π and v (and strictly convex in all directions $\Delta\pi, \Delta v$ except those in which $\Delta\pi + \Delta v = 0$). Thus, from Jensen's inequality we conclude that full separation is optimal. In particular, for any disclosure rule $\langle \sigma, S \rangle$,

$$\mathbb{E} \left(\frac{1}{4} \mathbb{E}[(\pi + v) \mid s]^2 \right) \leq \mathbb{E} \left(\frac{1}{4} \mathbb{E}[(\pi + v)^2 \mid s] \right) \leq \mathbb{E}[(\pi + v)^2],$$

where $\mathbb{E}[(\pi + v)^2]$ corresponds to expected profits under full separation. Consider, moreover, the "generic" case in which the sum $(\pi_i + v_i)$ is never equal for two different projects. In

this case, it is strictly optimal to fully separate every prospect because, along the interval connecting the payoffs of any two prospects, the payoff function is strictly convex.

Recall that the original motivation for pooling was to increase the acceptance rate of high-profit prospects by pooling these “switch” prospects with “bait” prospects. But once subsidies are allowed, the Sender effectively replaces this strategy with (more efficient) direct monetary incentives. Of course, offering such subsidies may prove impractical because the Receiver can potentially game the contract (e.g., there may exist a mass of strategic internet users with very low clicking costs that are not interested in the Advertiser’s product per se, but nevertheless click on the ad in order to exploit the subsidy), or it may prove infeasible if the Sender cannot directly contract with the Receiver (e.g., a university may not be capable of offering payments to future employers of its students).

We now extend the above results to the case in which π is allowed to lie anywhere along the real line. In this case, we must explicitly restrict the Receiver’s acceptance a to lie in $[0, 1]$, which, from the Sender’s standpoint, is equivalent to restricting $\mathbb{E}[v | s] + \alpha$, for each s , to lie in this interval. (Notice that when both π and v belonged to $[0, 1]$ this restriction was automatically met since the optimized acceptance rate was $v + \alpha(s) = \frac{1}{2}(\mathbb{E}[\pi | s] + \mathbb{E}[v | s])$.) The resulting constrained problem is

$$\begin{aligned} \max_{\alpha} \quad & \mathbb{E}[\pi - \alpha | s] \cdot \mathbb{E}[v + \alpha | s] \\ \text{s.t.} \quad & \\ & 0 \leq \mathbb{E}[v | s] + \alpha \leq 1. \end{aligned}$$

Since the unconstrained optimal subsidy is $\hat{\alpha}(s) = \frac{1}{2}\mathbb{E}[(\pi - v) | s]$, the constrained solution becomes

$$\alpha(s) = \begin{cases} -\mathbb{E}[v | s] & \text{if } \hat{\alpha}(s) \leq 0, \\ \hat{\alpha}(s) & \text{if } 0 < \hat{\alpha}(s) < 1, \\ 1 - \mathbb{E}[v | s] & \text{if } \hat{\alpha}(s) > 1. \end{cases}$$

The first case corresponds to the corner solution in which $a = 0$ (representing exclusion), the second case corresponds to an unconstrained interior solution, and the last case is the opposite corner solution in which $a = 1$ (representing 100% acceptance rate).

The optimized payoff for the Sender (conditional on s) is therefore

$$U_S = \begin{cases} 0 & \text{if } \hat{\alpha}(s) < 0, \\ \frac{1}{4}\mathbb{E}[(\pi + v) | s]^2 & \text{if } 0 \leq \hat{\alpha}(s) \leq 1, \\ \pi + v - 1 & \text{if } \hat{\alpha}(s) > 1. \end{cases}$$

This function is continuous and, since each of its segments is either linear or convex, it remains weakly convex. As a result, the full separation rule remains weakly optimal. What is new relative to the case in which $\pi \in [0, 1]$ is that any prospect with a negative combined payoff $(\pi_i + v_i) < 0$ is optimally excluded ($a = 0$), and any prospect with an average payoff $\frac{1}{2}(\pi_i + v_i)$ greater than 1 receives 100% acceptance rate ($a = 1$). Notice, finally, that when two prospects receive 100% acceptance rate they become weakly ordered and therefore the Sender is indifferent between separating and pooling these prospects.

That full disclosure is optimal when transfers are allowed is consistent with the findings of Ottaviani and Prat (2001), who show that a monopolist designing a price-discrimination mechanism finds it optimal to commit to publicly reveal information affiliated to the consumer's valuation.

8.3 Non-Uniform Acceptance Rate

Here we discuss the case in which the Receiver's reservation value r is drawn from a general distribution G over $[0, 1]$. Conditional on receiving a given signal s , the Receiver's acceptance rate becomes $\text{prob}\{r \leq \mathbb{E}[v | s]\} = G(\mathbb{E}[v | s])$. Thus, the Sender's expected profit from sending signal s is $\mathbb{E}(\mathbb{E}[\pi | s] \cdot G(\mathbb{E}[v | s]))$. Taking the ex-ante expectation over signals, the Sender's expected profit for a given disclosure σ is now

$$\mathbb{E}_\sigma(\mathbb{E}[\pi | s] \cdot G(\mathbb{E}[v | s])). \quad (16)$$

As before, we begin by computing the Sender's expected gain from pooling two prospects i and j into one signal \hat{s} (while disclosing information about the other prospects as before). This gain is given by

$$(p_i + p_j) \mathbb{E}[\pi | \hat{s}] \cdot G(\mathbb{E}[v | \hat{s}]) - p_i \pi_i G(v_i) - p_j \pi_j G(v_j)$$

$$\begin{aligned}
&= -\frac{p_i p_j}{p_i + p_j} (\pi_i - \pi_j) (G(v_i) - G(v_j)) \\
&\quad + (p_i + p_j) \mathbb{E}[\pi | \hat{s}] \cdot \{G(\mathbb{E}[v | \hat{s}]) - \mathbb{E}[G(v) | \hat{s}]\}.
\end{aligned} \tag{17}$$

When both prospects have the same acceptance rate G , pooling has no impact. In contrast, when $G(v_i) \neq G(v_j)$, pooling has two effects. First, as before, it shifts acceptance rate from the more valuable prospect (with a higher rate G) to the less valuable prospect. This effect is captured by the first term in (17), which indicates that the shift in acceptance rate raises the Sender's expected profits when the more valuable prospect is also less profitable (the unordered case), and vice versa.

Second, depending on the curvature of G , pooling may also change the overall acceptance rate. This effect is captured by the second term in (17). For example, when G is strictly concave, pooling increases the overall acceptance rate (by Jensen's inequality the expression in braces is positive), therefore raising profits. The opposite occurs when G is strictly convex.

Once both effects are combined we obtain:

Lemma 12 *Pooling two prospects with different acceptance rates yields (strictly) higher profits for the Sender than separating them if the prospects are (strictly) unordered and G is (strictly) concave, and yields (strictly) lower profits if the prospects are (strictly) ordered and G is (strictly) convex. The remaining cases are ambiguous.¹²*

To further understand this result, it is useful to examine the curvature of the Sender's profit function $\pi \cdot G(v)$ in different directions $(\Delta\pi, \Delta v)$:

$$\frac{d^2}{dt^2} [(\pi + t\Delta\pi) \cdot G(v + t\Delta v)] = 2G'(v)\Delta\pi\Delta v + G''(v)\pi\Delta v^2, \tag{18}$$

and so the function is concave (making pooling optimal for the Sender) when both $\text{sign}(\Delta\pi) \neq \text{sign}(\Delta v)$ and $G''(v) \leq 0$, and convex when both $\text{sign}(\Delta\pi) = \text{sign}(\Delta v)$ and $G''(v) \geq 0$.

The difference vis-a-vis the case of a uniform acceptance rate is that two ordered prospects will be optimally pooled when G is sufficiently concave (in order to increase the overall accep-

¹²When G is twice differentiable, we can obtain further intuition for this lemma by examining the curvature of the function $\pi \cdot G(v)$ in different directions $(\Delta\pi, \Delta v)$:

$$\frac{d^2}{dt^2} [(\pi + t\Delta\pi) \cdot G(v + t\Delta v)] = 2G'(v)\Delta\pi\Delta v + G''(v)\pi\Delta v^2,$$

and so the function is concave (making pooling optimal for the Sender) when both $\text{sign}(\Delta\pi) \neq \text{sign}(\Delta v)$ and $G''(v) \leq 0$, and convex when both $\text{sign}(\Delta\pi) = \text{sign}(\Delta v)$ and $G''(v) \geq 0$.

tance rate), and two unordered prospects will be optimally separated when G is sufficiently convex (in order to avoid a reduction in the acceptance rate). Nevertheless, several properties of the optimal disclosure rule derived above continue to hold. Most notably, prospects that are pooled together must still lie on a straight line:

Lemma 13 *Assume G is differentiable and strictly increasing. In a profit-maximizing disclosure rule σ , for any given signal $s \in S$, the payoffs of the prospects in the pool of s , $\{(\pi_i, v_i) : i \in P_s\}$, lie on a straight line.*

Proof. Suppose not. Then the convex hull of $\{(\pi_i, v_i) : i \in P_s\}$, which we denote by H , has a nonempty interior that contains $(\mathbb{E}[\pi|s], \mathbb{E}[v|s])$. In addition, H contains $(\mathbb{E}[\pi|s], \mathbb{E}[v|s]) - (\delta_1, \delta_2)$ for small $\delta_1, \delta_2 > 0$. Let $\lambda \in \Delta(P_s)$ be such that

$$(\mathbb{E}[\pi|s], \mathbb{E}[v|s]) - (\delta_1, \delta_2) = \sum_{i \in P_s} \lambda_i \cdot (\pi_i, v_i).$$

Now consider a new disclosure rule $\hat{\sigma}$ that replaces the original signal s with two new signals s_1, s_2 , and for each $i \in P_s$ has $\hat{\sigma}_{s_1}(i) = \varepsilon \lambda_i$ and $\hat{\sigma}_{s_2}(i) = \sigma_s(i) - \varepsilon \lambda_i$, where $\varepsilon > 0$ is chosen small enough so that $\hat{\sigma}_{s_2}(i) > 0$ for all $i \in P_s$. (Also set $\hat{\sigma}_t(i) = \sigma_t(i)$ for all $t \in S \setminus \{s\}$, $i \in P \setminus P_s$.) Let $(\bar{\pi}, \bar{v}) = (\mathbb{E}[\pi|s], \mathbb{E}[v|s])$ and $(\bar{\pi}_k, \bar{v}_k) = (\mathbb{E}[\pi|s_k], \mathbb{E}[v|s_k])$ for $k = 1, 2$. By construction, we obtain

$$(\bar{\pi}_1, \bar{v}_1) = (\bar{\pi}, \bar{v}) - (\delta_1, \delta_2) \text{ and} \tag{19}$$

$$\frac{\varepsilon}{q_s}(\bar{\pi}_1, \bar{v}_1) + \frac{q_s - \varepsilon}{q_s}(\bar{\pi}_2, \bar{v}_2) = (\bar{\pi}, \bar{v}), \text{ where } q_s = \sum_{i \in P_s} \sigma_s(i).$$

These equations in turn imply

$$(\bar{\pi}_2, \bar{v}_2) = (\bar{\pi}, \bar{v}) + \alpha \cdot (\delta_1, \delta_2), \text{ where } \alpha = \frac{\varepsilon}{q_s - \varepsilon}. \tag{20}$$

The Sender's gain from adopting $\hat{\sigma}$ relative to σ is

$$\begin{aligned} & \varepsilon \cdot \bar{\pi}_1 G(\bar{v}_1) + (q_s - \varepsilon) \cdot \bar{\pi}_2 G(\bar{v}_2) - q_s \cdot \bar{\pi} G(\bar{v}) \\ &= \frac{\varepsilon(q_s - \varepsilon)}{q_s} (\bar{\pi}_2 - \bar{\pi}_1) (G(\bar{v}_2) - G(\bar{v}_1)) \\ & - q_s \bar{\pi} \left(G(\bar{v}) - \frac{\varepsilon}{q_s} \cdot G(\bar{v}_1) - \frac{q_s - \varepsilon}{q_s} \cdot G(\bar{v}_2) \right). \end{aligned}$$

From (19) and (20) this gain is equal to

$$\frac{\varepsilon(q_s - \varepsilon)}{q_s}(1+\alpha)\delta_1(G(\bar{v} + \alpha\delta_2) - G(\bar{v} - \delta_2)) - q_s\bar{\pi} \left(G(\bar{v}) - \frac{\varepsilon}{q_s} \cdot G(\bar{v} - \delta_2) - \frac{q_s - \varepsilon}{q_s} \cdot G(\bar{v} + \alpha\delta_2) \right),$$

which we denote by $\Phi(\varepsilon, \delta_1, \delta_2)$. Now fix small $\varepsilon, \delta_1 > 0$. Notice that $\Phi(\varepsilon, \delta_1, 0)$ is zero and the partial derivative $\frac{\partial}{\partial \delta_2} \Phi(\varepsilon, \delta_1, 0)$ is strictly positive:

$$\frac{\partial}{\partial \delta_2} \Phi(\varepsilon, \delta_1, 0) = \frac{\varepsilon(q_s - \varepsilon)}{q_s}(1+\alpha)^2\delta_1 G'(\bar{v}) > 0.$$

It follows that $\Phi(\varepsilon, \delta_1, \delta_2)$ is strictly positive for any small $\delta_2 > 0$, which contradicts the optimality of the original disclosure rule σ . ■

To understand this result, it is useful to re-examine the curvature of the Sender's profit function $\pi \cdot G(v)$ in (18). Note that the first term in this equation is proportional to Δv , whereas the second term is proportional to Δv^2 . Thus, starting from any arbitrary point (π, v) , provided $G'(v) > 0$ there always exists an ordered direction $(\Delta\pi, \Delta v)$, with sufficiently small $\Delta v \neq 0$, along which the first term in (18) is larger than the second term, and therefore $\pi \cdot G(v)$ is strictly convex.¹³ Consequently, if a given signal pools prospects that do not lie on the same straight line, this signal can always be spread out in a direction of convexity (as in Figure 1, but now spread out along a line with sufficiently small slope), therefore increasing expected profits.

Thanks to this lemma, we can also show that pooling intervals cannot intersect at an interior point:

Lemma 14 *Assume G is differentiable and strictly increasing. In a profit-maximizing disclosure rule σ , suppose we have prospects a_1, a_2, b_1, b_2 and signals s_1, s_2 such that: $a_1, b_1 \in P_{s_1}$, $a_2, b_2 \in P_{s_2}$, and the payoffs of these prospects do not lie on the same line. Then, the intervals¹⁴*

$$[(\pi_{a_1}, v_{a_1}), (\pi_{b_1}, v_{b_1})] \text{ and } [(\pi_{a_2}, v_{a_2}), (\pi_{b_2}, v_{b_2})]$$

can only intersect if they share an end point.

¹³Indeed, for small Δv the acceptance rate G is approximately linear, and therefore the curvature of the Sender's objective is essentially determined by the sign of $\Delta\pi\Delta v$.

¹⁴Where $[x, y] = \{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}$.

Proof. Identical to the proof of lemma 4, but with lemma 13 replacing lemma 3 in the last line of the proof. ■

The remaining results in section 4 are dependant on prospects being pooled or not based exclusively on their ordering. Indeed, since the curvature of G can have a decisive effect on the desirability to pool any given pair of prospects, regardless of their payoffs, little can be said in general about the optimal pooling graph. In contrast, when the curvature of G is mild, the findings of section 3 continue to be relevant. For example, if G is everywhere concave and its curvature is sufficiently mild so that strictly ordered prospects remain separated, then all lemmas derived under the uniform assumption remain valid.

9 Appendix

Before proving Lemmas 8 and 9, and Proposition 1, we derive some preliminary results.

Lemma 15 *Suppose $\langle \sigma, S \rangle$ are $\langle \sigma', S' \rangle$ are optimal disclosure rules. Then, for any pair of signals $s \in S$ and $s' \in S'$, the posterior payoffs $\mathbb{E}[(\pi, v)|s]$ and $\mathbb{E}[(\pi, v)|s']$ are ordered.*

Proof. Suppose without loss that the sets S and S' have no signal in common (which is always possible through a relabeling of signals). Now consider a new disclosure rule $\langle \sigma'', S'' \rangle$ that results from randomizing between the two original disclosure rules $\langle \sigma, S \rangle$ and $\langle \sigma', S' \rangle$ with equal probability assigned to each. Namely, $S'' = S \cup S'$ and $\sigma''_s(i) = \frac{1}{2} \{ \sigma_s(i) + \sigma'_s(i) \}$ for every $i \in P$ and $s \in S''$.

Since S and S' do not intersect, for any given $s \in S$ and $s' \in S'$, the posterior payoffs $\mathbb{E}[(\pi, v)|s]$ and $\mathbb{E}[(\pi, v)|s']$ are equal under the original and new disclosure rules. As a result, the expected payoff delivered by $\langle \sigma'', S'' \rangle$ is

$$\sum_{s \in S''} \sum_{i \in P} \sigma''_s(i) \mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s] = \frac{1}{2} \left\{ \sum_{s \in S} \sum_{i \in P} \sigma_s(i) \mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s] + \sum_{s \in S'} \sum_{i \in P} \sigma'_s(i) \mathbb{E}[\pi|s] \cdot \mathbb{E}[v|s] \right\},$$

where the two terms in braces represent, respectively, the payoffs delivered by $\langle \sigma, S \rangle$ and $\langle \sigma', S' \rangle$. It follows that $\langle \sigma'', S'' \rangle$ is also optimal. Consequently, from Lemma 2, the set of posterior payoffs under $\langle \sigma'', S'' \rangle$, which includes all posterior payoffs from the original disclosure rules, must be ordered. ■

Corollary 1 *Suppose β and β' are solutions to problem (6). If a given pair of prospects $i, j \in U$ is pooled under both β and β' , then the posterior payoffs $\mathbb{E}[(\pi, v)|\{i, j\}]$ conditional on signal $\{i, j\}$ must be equal for both solutions. As a result,*

$$\frac{\beta_{ij}}{\beta_{ji}} = \frac{\beta'_{ij}}{\beta'_{ji}}.$$

Proof. For each solution, the posterior payoffs $\mathbb{E}[(\pi, v)|\{i, j\}]$ lie on the straight line with negative slope connecting (π_i, v_i) and (π_j, v_j) . Consequently, if these posterior payoffs differed across solutions, they would be strictly unordered, a contradiction to Lemma 15. ■

Proof of Lemma 8.

That \widehat{B} is convex follows from the fact that the objective in (6) is concave (Lemma 7) and the set of vectors $\beta \in \mathbb{R}^U$ that satisfy constraints (7), (8), and (10) is convex. For compactness, it suffices to show that \widehat{B} contains its boundary. Suppose β' belongs to the boundary of \widehat{B} and let $\beta^n \in \widehat{B}$ ($n = 1, 2, \dots$) be a sequence converging to β' . Constraint (10) and Corollary 1 imply that there exists a constant $C > 0$ such that, for every n and every $i, j \in U$,

$$\beta_{ij}^n = C\beta_{ji}^n.$$

Taking the limit as $n \rightarrow \infty$, this equality implies that β' satisfies constraint (10). In addition, since the objective $F(\beta)$ is continuous, β' must also be an optimum. It follows that $\beta' \in \widehat{B}$. *Q.E.D.*

Proof of Lemma 9.

We begin with necessity (\Rightarrow). Suppose that $\beta \in \widehat{B}$ is not a vertex of \widehat{B} . Since \widehat{B} is convex, there must exist an optimum $\beta' \in \widehat{B}$ that is arbitrarily close to β and yet $\beta' \neq \beta$. Indeed, we can select β' such that, for every $i, j \in U$, $\beta'_{ij} > 0 \Leftrightarrow \beta_{ij} > 0$. Let $V \subset U$ denote the subset of pairs $i, j \in U$ such that $\beta_{ij} > 0$. Since β is not a vertex of \widehat{B} , V is nonempty. Moreover, from Corollary 1,

$$\frac{\beta_{ij}}{\beta_{ji}} = \frac{\beta'_{ij}}{\beta'_{ji}} \text{ for all } i, j \in V. \tag{21}$$

Now let $\Delta\beta_{ij} = \beta_{ij} - \beta'_{ij}$. Constraint (2), which binds under both β and β' , implies

$$\sum_{j \in P(i)} \Delta\beta_{ij} = 0 \text{ for all } i \in P. \quad (22)$$

Since $\beta' \neq \beta$, there must exist a pair $i, j \in V$ such that $\Delta\beta_{ij} \neq 0$. Moreover, whenever $\Delta\beta_{ij} \neq 0$, (21) implies that $\Delta\beta_{ji} \neq 0$ (with $\text{sign}(\Delta\beta_{ji}) = \text{sign}(\Delta\beta_{ij})$), and equation (22) in turn implies that there exists a prospect $k \in P(j)$, with $k \neq i$, such that $\Delta\beta_{jk} \neq 0$ (with $\text{sign}(\Delta\beta_{jk}) \neq \text{sign}(\Delta\beta_{ji})$). It follows that we can select an infinite sequence of prospects i_1, i_2, \dots (with repeated elements) such that, for all $k = 1, 2, \dots$, we have: $i_k \neq i_{k+1}$ and $\Delta\beta_{i_k i_{k+1}} \neq 0$. Moreover, since $\Delta\beta_{i_k i_{k+1}} \neq 0$ requires by construction that $\beta_{i_k i_{k+1}} > 0$, and the set of prospects P is finite, β must contain a cycle.

We now turn to sufficiency (\Leftarrow). Suppose $\beta \in \widehat{B}$ contains a cycle among prospects $\{i_1, i_2, \dots, i_K\}$. Without loss, denote these prospects $\{1, 2, \dots, K\}$. For every k in this set, let $\gamma_k = \frac{\beta_{k, k \bmod K+1}}{\beta_{k, k \bmod K+1} + \beta_{k \bmod K+1, k}}$ (i.e., the share of k in signal $\{k, k \bmod K+1\}$) and let $A_k = \sqrt{Z_{k, k \bmod K+1}}$. The first-order conditions (9) for weights $\beta_{k, k \bmod K+1}$ and $\beta_{k, k \bmod K-1}$ (which are both positive) are

$$(1 - \gamma_k) \cdot A_k = \gamma_{k \bmod K-1} \cdot A_{k \bmod K-1} = \sqrt{\lambda_k}. \quad (23)$$

Multiplying these first-order conditions across k , and rearranging terms, we obtain

$$\prod_{k=1}^K \frac{1 - \gamma_k}{\gamma_k} = 1. \quad (24)$$

We now show that there exist two optima $\beta', \beta'' \in \widehat{B}$, both different from β , such that $\beta = \frac{1}{2}(\beta' + \beta'')$, which in turn implies that β is not a vertex of \widehat{B} . Select a small $\varepsilon > 0$ and, for all $k = 1, 2, \dots, K$, let

$$\begin{aligned} \beta'_{k, k \bmod K+1} &= \beta_{k, k \bmod K+1} + \Delta_k, & \beta'_{k, k \bmod K-1} &= \beta_{k, k \bmod K-1} - \Delta_k, \\ \beta''_{k, k \bmod K+1} &= \beta_{k, k \bmod K+1} - \Delta_k, & \beta''_{k, k \bmod K-1} &= \beta_{k, k \bmod K-1} + \Delta_k, \end{aligned} \quad (25)$$

where the values of Δ_k satisfy $\Delta_1 = \varepsilon$ and

$$\Delta_{k \bmod K+1} = -\frac{1 - \gamma_k}{\gamma_k} \cdot \Delta_k. \quad (26)$$

(That the above equation can be satisfied for all k follows from equation (24)). For all other pairs $i, j \in U$, let $\beta'_{ij} = \beta''_{ij} = \beta_{ij}$. Notice that, provided ε is small, β' and β'' satisfy (10) and $\beta = \frac{1}{2}(\beta' + \beta'')$. Moreover, combining equations (25) and (26) we obtain

$$\frac{\beta'_{k, k \bmod K+1}}{\beta'_{k \bmod K+1, k}} = \frac{\beta''_{k, k \bmod K+1}}{\beta''_{k \bmod K+1, k}} = \frac{\beta_{k, k \bmod K+1}}{\beta_{k \bmod K+1, k}}.$$

As a result, β' and β'' lead to the same values of γ_k as the original optimum β , and therefore they also meet the first-order conditions (23). It follows that $\beta', \beta'' \in \widehat{B}$. *Q.E.D.*

Proof of Proposition 1.

The first part of the Corollary follows directly from Lemma 9. For the second part, we show that, under Condition 1, no $\beta \in \widehat{B}$ can be cyclic. As a result, from Lemma 9, \widehat{B} must be a singleton. Suppose in negation that $\beta \in \widehat{B}$ contains a cycle among a subset of prospects denoted $\{1, 2, \dots, K\}$. From the proof of Lemma 9, for every k in this subset, the first-order conditions (23) must be met. Combining these first-order conditions to solve for the value of γ_1 we obtain

$$\gamma_1 = \gamma_1 + \frac{1}{A_1} \sum_{k=1}^K (-1)^k A_k,$$

where $A_k = \sqrt{Z_{k, k \bmod K+1}}$. But this equation can only hold when $\sum_{k=1}^K (-1)^k A_k = 0$, which is ruled out by Condition 1. *Q.E.D.*

10 References

1. Athey, S., and G. Ellison (2008), “Position Auctions with Consumer Search,” working paper
2. Crawford, V., and J. Sobel, “Strategic Information Transmission,” *Econometrica* 50
3. Kamenica, E., and M. Gentzkow (2009), “Bayesian Persuasion,” working paper
4. Kihlstrom, R., and M. Riordan (1984), “Advertising as a Signal,” *Journal of Political Economy* 92(3)
5. Lizzeri, A. (1999), “Information Revelation and Certification Intermediaries,” *RAND Journal of Economics* 30(2)
6. Milgrom, P. (2008), “What the Seller Won’t Tell You: Persuasion and Disclosure in Markets,” *Journal of Economic Perspectives* 22(2)
7. Milgrom, P., and R. Weber (1982), “A Theory of Auctions and Competitive Bidding,” *Econometrica* 49
8. Nelson, P. (1974), “Advertising as Information,” *Journal of Political Economy* 82
9. Ostrovsky, M., and M. Schwarz (2008), “Information Disclosure and Unraveling in Matching Markets,” working paper
10. Ottaviani, M., and A. Prat (2001), “The Value of Public Information in Monopoly,” *Econometrica* 69(6)
11. Rayo, L. (2005), “Monopolistic Signal Provision,” *B.E. Journals* (accepted)
12. Spence, M. (1973), “Job Market Signaling,” *Quarterly Journal of Economics* 87(3)

Figure 1: Pooling along Straight Lines

Black balls: prospects

White ball: signal that pools all prospects

Grey balls: new signals

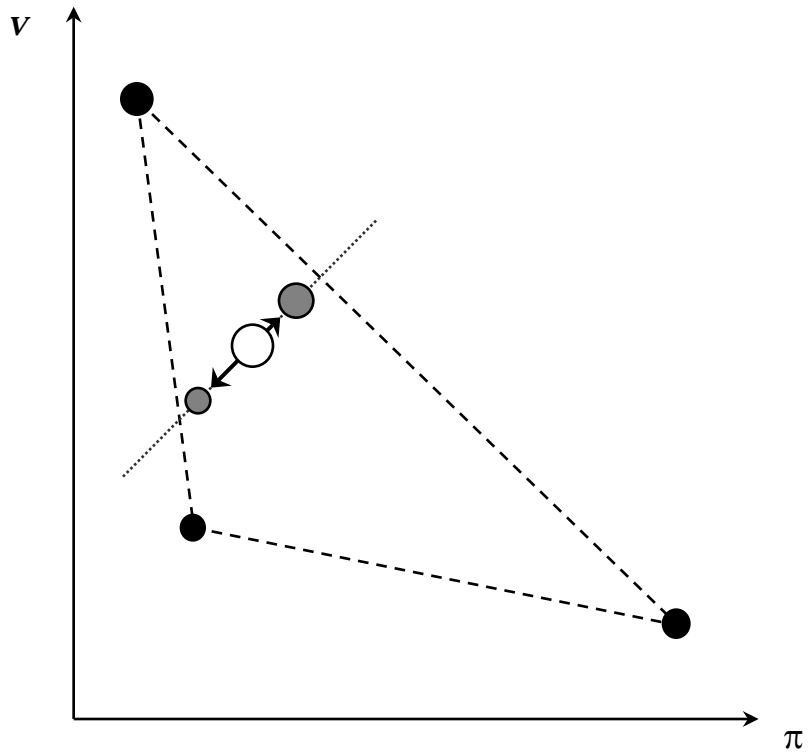


Figure 2: Pooling Intervals do not Intersect

Proof of Lemma 4. If there is an intersection point z , the masses of signals s_1 and s_2 can be recombined so that: (i) the position and total mass of each signal does not change, and (ii) the pool of each signal now contains all four prospects – a contradiction to Lemma 3.

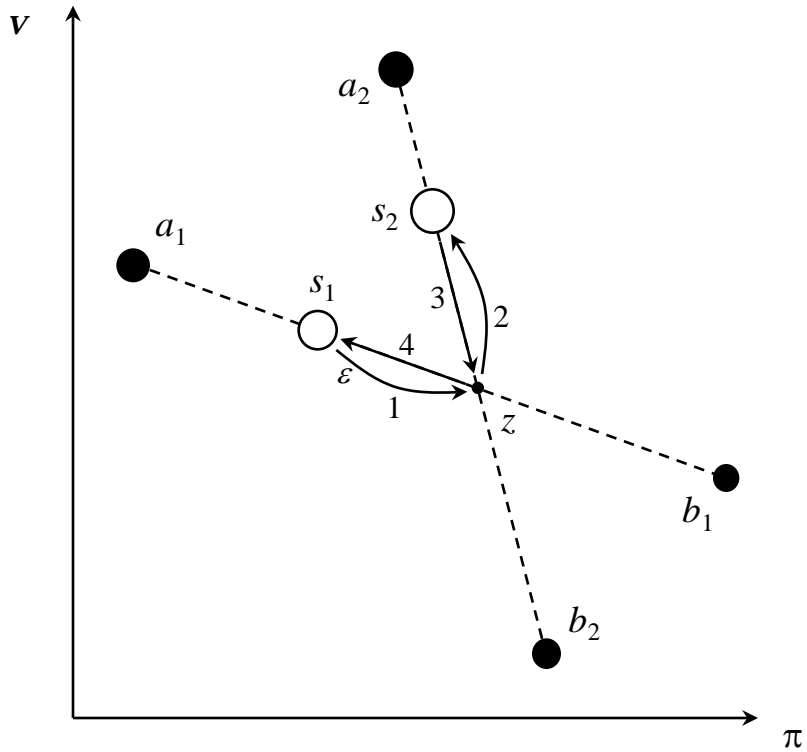


Figure 3: Taxonomy with 4 Prospects

Panel A (top left): Fan. **Panel B** (top right): Two Lines.

Panel C (bottom left): Zigzag. **Panel D** (bottom right): Cycle.

Black discs represent individual prospects. White discs represent pools.

The size of each disc is proportional to its mass.

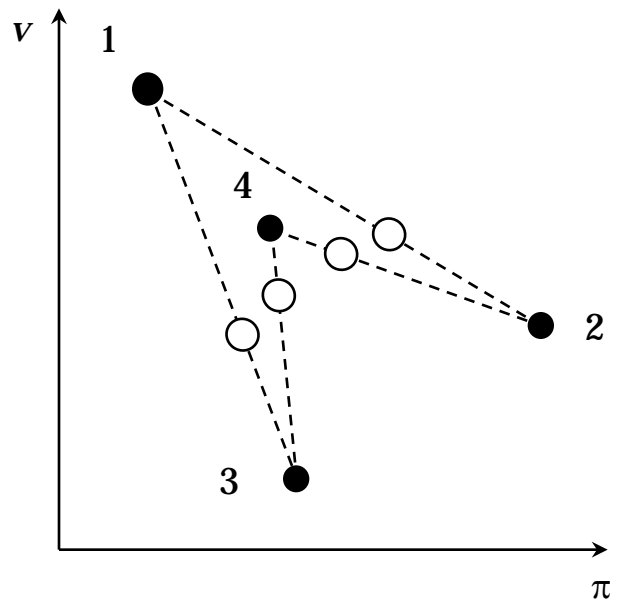
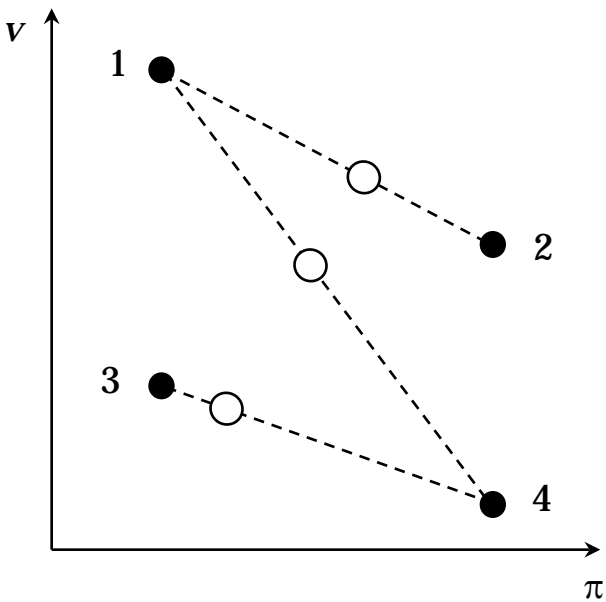
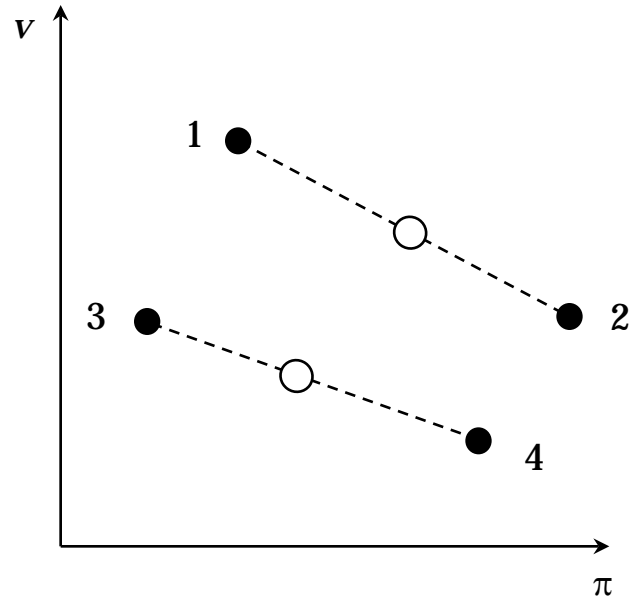
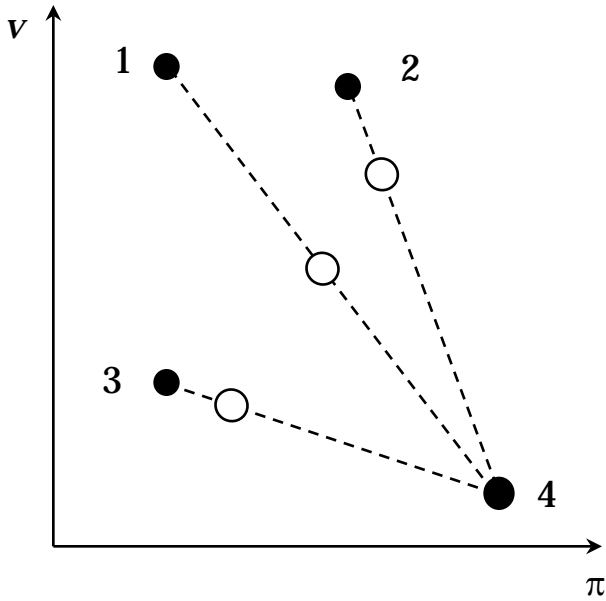


Figure 4: Pareto-Weighted Payoffs

$\lambda = 0$ (black balls): pairs $\{1,2\}$ and $\{1,3\}$ are strictly unordered

$\lambda = \frac{1}{2}$ (grey balls): only pair $\{1,2\}$ is strictly unordered

$\lambda = 1$ (white balls): all pairs are strictly ordered

