The Scale of Data

Information Spillovers and Investment Allocation in the AI Era*

Sylvia Chu

Advised by Professor Laura L. Veldkamp

November 25, 2026

Unpublished manuscript

Abstract. Economic activity produces data, but data guide investment only when they can be matched to the opportunities an investor faces. I study complementary technologies whose joint return differs across observable production types, with the assignment of returns to types unknown. A report that pools outcomes reveals the distribution of returns but not which type earns which; a type-linked report reveals both. When one firm's deployment generates such a report for separately owned firms, the deployer ignores the recipients' gain, so private deployment is never more frequent than the planner's. The main result is that paying for data fixes whether data are produced, not what is recorded. A fee that fully captures the recipients' gain aligns deployment, yet the source can still choose the less informative pilot, and when it chooses how precisely to record it records weakly less than the planner. In a Gaussian class where the signs follow from primitives, more payment lowers the precision the source records and a large enough payment ends recording, while the planner's precision grows with the number of recipients; no fee contingent on deployment alone restores it, and if the report can leak the gap grows without bound in the number of recipients. If recipients can verify the precision with some probability, recording resumes only when that probability exceeds a threshold, which in large markets equals the share of final precision that the later report supplies.

Keywords: data economy; information externalities; experimentation; complementary technologies; investment allocation. **JEL:** D83, O31, O33.

*I thank Laura Veldkamp for early discussions that helped shape the research question. All errors are mine.

1 Introduction

A vendor sells an AI inspection system to manufacturers and is paid each time a plant adopts it. Before it sells, it pilots the system at one plant. How finely the pilot logs what happens, by line, by batch and by operating condition, is costly and is not something buyers can check. A vendor paid per adoption has little reason to log finely. It will sell the system to every plant whatever the pilot shows, so learning from the pilot before selling has no value to it. A fee tied to adoption therefore pays the vendor to deploy and does not pay it to leave a record that other plants could use. The plants get the system without the evidence on where it works. This paper studies that problem: when one firm's activity generates data that others can use, payment for the activity can fix whether the data are produced and still fail to fix what is recorded.

Why the question matters. Models of the data economy treat data mainly as a quantity that accumulates as a by-product of activity, with its value depending on how much there is (Veldkamp and Chung, 2024; Farboodi and Veldkamp, 2026). The quality and form of the record are taken as given, and in these models a producer who is paid for data is also paid for good data. But recorded precision is a costly choice and is rarely what a market pays for: contracts pay for deployment, usage, or access. The paper separates the two margins and finds that they respond to payment in opposite ways. Three implications follow, each conditional on the model. First, data markets and platforms that pay per use may need to pay for verified quality instead, and Proposition 3 says how good the verification must be: in large markets it must exceed the share of final precision that a later report supplies, and below that threshold adding verification changes nothing. Second, data-sharing rules and logging standards matter more, not less, when payment is tied to deployment, because the producer's private incentive to learn before deploying is then weakest. Third, the market-size result says that the shortfall grows with the number of users, which is where data are supposed to be most valuable.

A second ingredient: the form of the record. A record also differs in whether outcomes remain attached to the observable type that a reader can act on. Suppose a plant that installs inspection together with batch traceability reports either that the combination cut defects sharply on some lines and not at all on others, or the same outcomes line by line: large gains on short-run, high-mix lines and none on continuous lines. A second firm with a high-mix line gains only from the second. The two write-ups contain the same outcomes. They differ in whether the outcomes remain linked to types. The vignette is stylized and does not describe a particular firm. Sections 4 to 8 build a discrete model around this distinction, in which a pilot may record outcomes with or without their correspondence to types, and the source's choice of pilot is the choice of which data to collect. The choice of which data to collect, in this sense and in the precision choice of Section 3, is what determines how much the spillover to other firms is worth.

Results. The core model (Section 3) is Gaussian. A source chooses the precision x of an initial

public record at cost $c(x)$; the precision is not contractible. It then decides whether to deploy, which yields a later report of precision q that n recipients can use. A recipient who acts on the same unknown with quadratic loss values information of precision p at $-B/p$, so the later report is worth $H(x) = Bq/((p_0+x)(p_0+x+q))$ to each recipient, which falls as the earlier record becomes more precise. A fee for deployment raises the source's return to deploying whatever the record shows, and so removes the value of learning before deploying (Figure 1). The results are as follows.

- *Payment fixes deployment, not recording* (Proposition 1). The source's chosen precision is nonincreasing in the fee and is zero for a large enough fee, while the planner's precision grows like $\sqrt{n}$. Figure 2 shows this in one example.
- *No fee on deployment alone repairs both margins* (Proposition 2). At the fee that makes deployment efficient, precision is too low; at a lower fee deployment is too rare.
- *Verification restores recording only above a threshold* (Proposition 3). If recipients can verify precision with probability φ, the source's precision rises with φ and is below the planner's unless $\varphi = 1$. In large markets, with linear cost, recording is zero for φ below $q/(p_0+q)$ and positive above it. This is the model's account of why precision is hidden: it is a matter of degree, and the degree matters.
- *Leakage widens the gap* (Proposition 4). If a report can be passed on at no cost, so that the source can collect no more than one recipient's gain, the gap between planner and source is unbounded in n.

The discrete model supplies the other results. Type correspondence has investment value even when the distribution of returns is known (Proposition 5 and Corollary 1). Privately chosen deployment is never more frequent than the planner's, and can be too rare (Proposition 6), and this holds for any recipient decision problem (Proposition 14). Greater access to operating reports has no universal effect on the value of an informative pilot, because a later report can fill the gap a less informative pilot leaves (Proposition 7 and Corollary 3). A fee that captures the recipient's gain on the deployment margin aligns deployment but not the choice of pilot (Proposition 8), and when the fee is the outcome of a pricing problem the source captures only part of the gain and excludes some recipients (Proposition 9 and Corollary 4). The shortfall in the source's incentive to run the informative pilot grows in proportion to the number of recipients (Propositions 10 and 11), and its sign follows the curvature of the recipient's value of information (Proposition 12). Appendix G shows that the discrete results survive many types, general recipients, a known correspondence, and a discrete precision grid (Propositions 13 to 17).

Relation to the literature. The closest papers are of two kinds. Activity-generated data and spillovers are studied in earlier work (Farboodi and Veldkamp, 2026; Farboodi et al., 2026; Teng,

2026); in those papers the record is a stock or an input and its quality is not chosen against payment. Papers on a seller's choice of precision are nearer the recording margin: Ozerturk (2004), Liu et al. (2025), and Colombo et al. (2026) show that the payment scheme shapes the precision a producer chooses. I did not find a paper that obtains the deployment-fee result with a public record, and I may have missed one; the search is described in Section 2. The tools are standard, namely Bayesian value of information and dynamic optimization, and the contribution is the specification of the problem and its analytical implications.

What the paper does not do. The core model has one source, one public record, and recipients that do not interact. Competition among recipients for a scarce slot, and a full resale game, are not modeled; Proposition 4 and the remark after it bracket the effect of resale without solving it. Precision is hidden by assumption in Proposition 1 and by the probability φ in Proposition 3, not derived from a contracting problem. The paper is theoretical and contains no empirical results. The numerical examples illustrate the results and are not estimates. It is not a growth model, and it does not generate multiple equilibria or a Big Push. Section 11 lists what remains.

Roadmap. Section 2 reviews related work. Section 3 presents the Gaussian core model, Propositions 1 to 4, and a welfare decomposition. Sections 4 to 8 develop the discrete model: Section 4 sets out the environment, Section 5 studies type correspondence, Section 6 deployment, Section 7 initial experimentation with endogenous continuation, and Section 8 contracts and pricing, with Section 8.5 on what the results imply for those who collect, buy, or regulate data. Section 9 shows how the discrete results scale with the number of recipients and what the two models share. Section 10 discusses testable implications and Section 11 concludes. Appendix E gives a Gaussian covariance benchmark, Appendix F the proofs, and Appendix G the supporting extensions.
Related literature

Data as information for decisions. Veldkamp and Chung (2024) organize the channels through which data enter economic models, including information for decisions, innovation, and the accumulation of productive assets. I use the information channel directly. A record changes a conditional belief about an investment return. It does not enter a production function as another factor, and the model holds technological performance fixed when it compares linked and pooled versions of a report. Farboodi and Veldkamp (2026), *A Model of the Data Economy*, is the natural starting point for activity-generated information: operating decisions produce data that affect later decisions. I keep that feedback but use a finite-horizon, heterogeneous-return environment to focus on where the information can be applied. There is no aggregate data stock in the main model. The relevant state is a posterior over the assignment of interaction returns to observable site types. My comparison uses the accessible author draft and does not presume that every feature is unchanged in the published version.

Cong et al. (2021) study data in knowledge accumulation, privacy, and growth. Those channels

matter for any broader development of deployment-generated data, but I do not add them to obtain a growth interpretation. In my model a report improves a recipient's choice among existing configurations. It neither creates a technology nor changes the growth rate of knowledge. The narrower interpretation lets the information mechanism be identified before innovation or aggregate accumulation is introduced.

Closest work. I compare my model with Farboodi et al. (2026) on the full June 2026 draft of *Data-Driven Automation*. Their data are a task-specific stock, accumulated as a by-product of production and aggregated into the effective data available for each task through a fixed spillover kernel. Atomless producers share data freely and do not internalize its value for future production. Four differences can be checked against that text.

(D1) *No unknown correspondence.* Their model has no uncertainty and no belief state; a search of the draft finds no reference to Bayesian updating, posteriors, or beliefs. Data raise productivity rather than resolve uncertainty, and the spillover kernel is known. There is therefore no counterpart to the pooled-versus-linked comparison of Propositions 5 and 13, which requires an unknown assignment of returns to types. Proposition 16 shows that when the correspondence is known, the channel in this paper is inactive.

(D2) *Sign of the wedge.* Their Proposition 6 states that the direction in which the planner tilts capital allocation relative to the equilibrium depends on the elasticity of substitution relative to $1/\eta$, and that, in their coarsened economy, the equilibrium is efficient only on a knife edge. Here the sign of the deployment wedge is fixed (Proposition 14): private deployment is never more frequent than the planner's. The objects differ, a capital tilt against a deployment decision, so this is a difference in structure, not a contradiction.

(D3) *What the choice is about.* In their model the planner can tilt capital toward tasks to accumulate data there, so the choice concerns where production happens. Data are a stock indexed by task. The choice in this paper concerns how outcomes are recorded relative to observable types, at a cost in immediate proceeds, so the pilot-choice results (Corollary 3 and Proposition 15) have no counterpart in a model where data are a stock.

(D4) *Different externalities.* Theirs is an accumulation externality in a general-equilibrium economy. Mine is an information externality from a single deployment. Neither implies the other, and a model containing both is a natural extension.

Teng's (2026) work on learning complementarity through experiment is especially close to my technological setting, but I compare it at the level of its public abstract, which describes firms that experiment on combinations of ingredients, uncertain ingredient qualities and complementarities, Bayesian updating, knowledge diffusion through patent disclosure, numerical solution with graph neural networks, and a result that the gains from disclosure policy depend on how disclosed ingredients complement a firm's portfolio. That last result is close in spirit to Proposition 13. The abstract

does not say whether disclosed outcomes retain their correspondence to the recipient's portfolio, whether disclosure is endogenous, or whether the efficiency properties include a sign result like Proposition 14. I therefore claim no difference from Teng's full model and no priority for learning complementarities through experiment. The differences I state are limited to what the abstract makes visible: an analytical rather than numerical solution, a separately owned recipient with an observable type, deployment that produces the report, and the pooled-versus-linked comparison.

Costly precision and the sale of information. Several papers study a seller who chooses the quality of what it sells, and they are closer to the recording margin than the data-economy papers above. Ozerturk (2004) models a seller who chooses costly, unobservable signal precision and sells to investors who also learn from the price. With compensation tied to quadratic forecast error the chosen precision rises with the volume sold, while a fixed fee leaves it unchanged. This is consistent with Proposition 3: payment that depends on an ex post measure of accuracy is a form of verification, and payment attached to use is not. I have not formalized that link. Liu et al. (2025) study a seller who chooses data quality at a convex cost and cannot commit against future sales, so that later sales dilute the value to earlier buyers and weaken the incentive to produce quality; subscriptions restore it, and their buyers cannot resell. Their friction is dilution through the seller's own future sales, mine is that payment is attached to use rather than to quality, and the two are complementary rather than substitutes. Colombo et al. (2026) show that a constant subsidy that aligns investment at a fixed information structure can fail to induce efficient information acquisition, and that transfers depending on the state can align both even when information is unverifiable. The logic is close to mine, since a payment on the action margin misprices the information margin, but their instrument is richer than a fee on deployment alone, which is the restriction in Proposition 2. Bimpikis et al. (2019) study a seller who chooses the precision of signals sold to competing downstream firms; their signal is endowed, not produced at a cost, and they show that strategic substitutes lead the seller to degrade information or exclude buyers. Kastl et al. (2018) and Drakopoulos and Makhdoumi (2023), whom I know only from abstracts, study imprecise information sold to competitive firms and the value of free samples. The closest margin to mine in this group is Ozerturk's; I did not find a paper that has a public record, a fee on deployment, and recipients who use the record. The search was limited to sources I could reach.

What I have read. Farboodi, Koh, and Xia (2026), Ichihashi (2021), Acemoglu et al. (2022), Bergemann et al. (2018), and Bergemann, Bonatti, and Gan (2022, author version) were read in full, including appendices, for the characterizations above; Ozerturk (2004) and Liu, Ma, and Veldkamp (2025, working-paper version) were read in the model and introduction sections, and Colombo et al. (2026) and Bimpikis et al. (2019) in the sections named. Teng (2026), Kastl et al. (2018), Drakopoulos and Makhdoumi (2023) and Admati and Pfleiderer (1986) were not read in full. The remaining citations are used for background and were checked against abstracts or parts of the text. No peer in the field has read this manuscript.

Information value and experimentation. Chen and Waggoner (2017) develop general notions of informational substitutes and of the value of signals in decision problems. My nonnegative value-of-information result follows the familiar logic: posterior beliefs are martingales, and optimized expected payoffs are convex in beliefs. The comparison between pilots is not an immediate application of a generic substitutes theorem, because the later report exists only when deployment is selected and a change in access reoptimizes that deployment. I derive the relevant dynamic comparison explicitly.

Markets for information and data. Bergemann et al. (2018) study a seller who designs and prices experiments for a buyer facing a decision problem, and show that the nature of information products widens the scope for price discrimination. My pricing result is a single posted price for a given report, so it lacks their design margin. The one place where version choice enters is Corollary 4, where the source chooses between a linked and a garbled version at a cost. Bergemann et al. (2022) study an intermediary that buys signals from consumers and resells information to a producer, where one consumer's data are predictive of others' and so can lower the compensation the intermediary must pay; externalities can also be positive, and the intermediary chooses whether acquired data are anonymized. Their consumers' initial precision is given, though an appendix lets the intermediary add noise at collection. The externality there runs through the supply of data; mine runs through the use of one firm's report by another. Acemoglu et al. (2022) show, with users whose signals have fixed noise and who choose only whether to share, that externalities among users depress the price of data and can lead to too much sharing. The sign of my wedge is the opposite, too little production, because the spillover here is a benefit to recipients and not a loss of privacy. The two concern different margins and do not contradict each other. Jones and Tonetti (2020) emphasize that data are nonrival, which creates a case for broad use, and that firms may hoard data. Nonrivalry enters my model through the number of recipients of a report (Proposition 10). Admati and Pfleiderer (1986) show that a monopolist seller of information may prefer to sell noisier versions of it, and personalized noisy versions to different buyers, in a noisy rational-expectations market (I rely on the author's abstract, lecture notes and a survey, not on the original text); Bergemann and Bonatti (2019) survey this literature. The choice between a linked and a garbled report in Corollary 4 is in the same spirit, but here the source does not choose the noise to maximize revenue: it chooses how much to link at a cost, and the recipients are investors choosing among installations. Three further papers are close to the recording margin. Ichihashi (2021) studies a single firm that buys information from consumers whose data are correlated, so that one consumer's sale changes what the firm learns from the others. When the firm can design the experiments it buys without restriction, its profit equals total surplus less the consumers' worst feasible outside options, and full revelation is optimal when consumers' preferences are individually monotone. Suppliers there choose whether to sell, not how precisely to record. There the collector is the buyer of data and the externality runs among suppliers of data. Here the collector is the

producer of the record, the recipients are third parties who would use it, and the question is whether the producer records enough in a form they can use; the two results concern different agents and are not in tension. Galperti et al. (2024) characterize the value of an individual data record and identify a pooling externality between records, which arises when an intermediary pools records to withhold the information they contain. Pooling in my model is a feature of the report, and what is priced is the correspondence between outcomes and types, not the value of a single record to an intermediary. Dominitz and Manski (2017) study the allocation of a sampling budget between processes that yield more data and processes that yield better data. Their decision maker is the only user of the data; the new element here is that a second party uses the record and does not pay for its form. Akcigit and Liu (2016) study competing experimenters who withhold dead-end findings and show that a dynamic disclosure mechanism can restore efficiency. Their margin is whether a finding is disclosed; mine is how the finding is recorded, which determines whether a disclosed report can be used at all.

Learning from others' experiments. Bolton and Harris (1999) show that when agents facing the same experimentation problem observe one another, information is a public good and free-riding arises. Their agents share one problem; here the source's experiment informs recipients whose problem differs by type, and the object studied is whether outcomes are recorded in a form the recipient can use. Foster and Rosenzweig (1995) and Conley and Udry (2010) provide evidence that farmers learn from others' experience when adopting new technologies. I do not use their evidence or estimates. They are cited as background on why such spillovers are economically relevant, which is a claim about those papers and not a result of this one.

Coordination. Murphy et al.'s (1989) Big Push is a coordination model in which an individual investment's profitability depends on other economic activity. A positive information externality is not sufficient to reproduce that equilibrium structure. In my model the source's privately optimal installation and the planner's installation are two optimization problems, not two private equilibria of the same economy. Establishing a Big Push would require an additional model of mutually dependent participation or repeated investment, which this finite sequence does not provide.

2 A core model: payment and the precision of the record

This section is the paper's core model. A source chooses how precisely to record, recipients later use a report that the source's deployment produces, and the sign of every effect of payment on the recording margin is proved from primitives. With Gaussian signals and quadratic loss the recipients' loss depends only on the precision of what they know, so the type-correspondence mechanism of Section 5 has no counterpart here: what remains is the incentive to choose how precisely to record. Sections 4 to 8 then add the second ingredient, the form of the record, in a discrete model, and

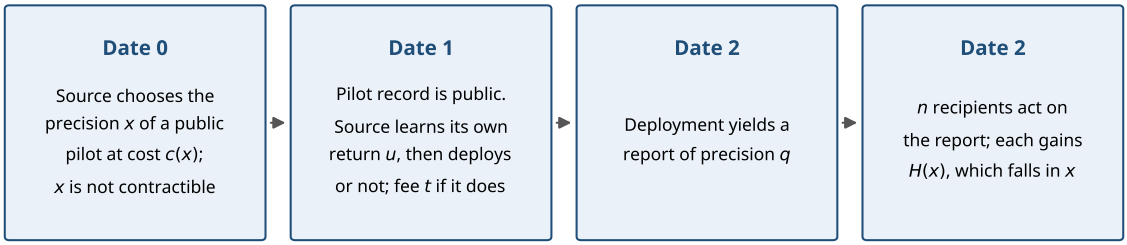

Figure 1: Timing in the core model.

Notes. Figure 1 shows the order of choices in this section. Recipients do not observe the precision the source chooses at date 0 when they contract, and the fee t is paid only if the source deploys.

Section 9.2 states what the two have in common.

Environment. A common state $\theta \sim N(0, 1/p_0)$ is relevant to the source's own deployment return and to n recipients. The source chooses the precision $x \geq 0$ of a public pilot at cost $c(x)$, where c is C^1, $c(0) = 0$, and $c'(x) \geq k > 0$. The pilot leaves the posterior mean m with distribution $N(0, v(x))$, $v(x) = x/(p_0(p_0+x))$. The source then decides whether to deploy, which yields the return $u = s_0 + s_1 m$ and a public report of precision q on θ. A recipient who must act on θ with quadratic loss has expected loss $B/(\text{precision of its information})$, so its value after the pilot is $w(x) = -B/(p_0+x)$ and its gain from the report is

$$H(x) = w(x+q) - w(x) = \frac{Bq}{(p_0+x)(p_0+x+q)},$$

which is positive and strictly decreasing in x: the pilot and the report are substitutes for the recipients. With β the discount factor and fee t (total capture, $0 \leq t \leq n$) paid on deployment, the source's value and the planner's value are

$$V_S(x;t) = -c(x) + \beta\, \mathbb{E}\max\{0,\, u + t\beta H(x)\}, \qquad V_P(x;n) = V_S(x;n) + \beta^2 n\, w(x).$$

Write $\sigma(x) = s_1\sqrt{v(x)}$, $a(x,t) = s_0 + t\beta H(x)$, and $P_x(D_t) = \Phi(a/\sigma)$ for the probability that the source deploys, with $P_0(D_t) = 1$ when $a > 0$.

Proposition 1 (Payment and recording precision in the Gaussian class)**.** *Let $s_0 \geq 0$. (i)* Envelope. *$\partial V_S/\partial t = \beta^2 H(x) P_x(D_t)$. (ii)* Payment lowers precision. *For $t' > t$, $V_S(x;t') - V_S(x;t)$ is nonincreasing in x. Hence the largest and the smallest maximizers of $V_S(\cdot\,;t)$ are nonincreasing in t. In particular, at full capture the source's precision is nonincreasing in the number of recipients n. (iii)* Large payments end learning. *There is $\bar{t} < \infty$ such that for every $t > \bar{t}$ the unique maximizer is $x = 0$: the source deploys with probability one and records nothing. (iv)* The planner records more. *For every $x > 0$, $\partial V_P/\partial x \;\geq\; -c'(x) + n\beta^2 B/(p_0+x+q)^2$. Hence every maximizer of $V_P(\cdot\,;n)$*

is at least $\hat{x}(n)$, the root of $n\beta^2 B/(p_0+\hat{x}+q)^2 = c'(\hat{x})$ (or 0 if there is none). With $c(x)=kx$, $\hat{x}(n) = \sqrt{n\beta^2 B/k} - p_0 - q$ whenever this is positive, so the planner's precision is unbounded in n while the source's is zero for $n > \bar{t}$. (v) Comparison. *The smallest maximizer of $V_P(\cdot\,;n)$ is strictly above the largest maximizer of $V_S(\cdot\,;n)$ unless both are zero.*

Appendix F.1 gives the proof. Part (ii) is the central sign: payment lowers recording precision because the pilot and the fee are substitutes in the source's deployment decision. A larger fee makes deployment more likely whatever the pilot shows, which removes the option value of learning before deploying. Part (iii) is the extreme form: a buyer who is sure to pay removes the reason to experiment. Part (iv) says the planner moves the other way: the more recipients, the more the public pilot is worth to them, and the marginal benefit of precision to the recipients' baseline rises one for one with n.

The sign in (ii) needs deployment to be profitable on average. For $s_0 < 0$ the argument still works on the fee range where it is: let $\tau_c = -s_0/(\beta H(X_0))$. For $\tau \geq \tau_c$ we have $a(x,\tau) \geq 0$ for every $x \in [0, X_0]$, so the same argument shows that $V_S(x;t') - V_S(x;t)$ is nonincreasing in x for $t' > t \geq \tau_c$, and the source's maximizers are nonincreasing in t on $[\tau_c, \infty)$. Below τ_c the probability of deployment can rise with precision and the sign can reverse. In the sweep below, with $s_0 \in \{-0.1, -0.3\}$ (972 configurations, ten fees from 0 to 64), 80 configurations show a rising source precision somewhere, and none of them does so at a fee above τ_c; with $s_0 \in \{0, 0.05, 0.3\}$ (1,458 configurations, source precision along $n \in \{0,1,2,4,8,16,32\}$) none does. This is the Gaussian counterpart of the sign switch in Proposition 11: the effect of payment on the pilot is positive where deployment is unlikely and negative once it is likely.

Numerical check and example. The sweep covers $p_0, B \in \{0.5, 1, 2\}$, $q \in \{0.5, 1, 3\}$, $\beta \in \{0.8, 0.95\}$, $s_1 \in \{0.5, 1, 2\}$, and $k \in \{0.005, 0.02, 0.1\}$, on a grid in x and in t. The envelope formula matches finite differences to 4×10^{-7}. In all 2,430 configurations the planner's maximizer is at least the source's at full capture, and in all 1,458 with $s_0 \geq 0$ the source records nothing at $n = 32$. Table 1 shows one parameter set and Figure 4(a) plots it: the source's precision is 2.36 with no recipients and 1.45 at $n = 0.5$, and it is zero from $n = 1$ on, while the planner's grows roughly with $\sqrt{n}$ and stays just above $\hat{x}(n)$.

Table 1. Pilot precision at full capture, Gaussian class.

n	Source, $t = n$	Planner	Planner lower bound $\hat{x}(n)$	Source deploys with prob.	Planner deploys with prob.
0	2.36	2.36	0.00	0.52	0.52
0.5	1.45	4.44	2.75	0.55	0.53
1	0.00	6.05	4.72	1.00	0.53

n	Source, $t = n$	Planner	Planner lower bound $\hat{x}(n)$	Source deploys with prob.	Planner deploys with prob.
2	0.00	8.57	7.50	1.00	0.53
5	0.00	13.85	13.02	1.00	0.53
8	0.00	17.75	17.00	1.00	0.53
20	0.00	28.70	28.04	1.00	0.53
40	0.00	41.08	40.48	1.00	0.53

Notes. $p_0 = B = q = 1$, $\beta = 0.95$, $s_0 = 0.05$, $s_1 = 1$, $c(x) = 0.02x$. Maximization is over a fine grid in x (step 0.025 near the optimum) by the script `verify_v16.py`. At $n = 0$ there are no recipients and the two coincide.

Proposition 2 (A fee on deployment cannot implement the planner's choice)**.** *Suppose x is chosen by the source and is not contractible, so that the only payments available are fees $t \in [0, n]$ contingent on deployment, and suppose the planner's precision is positive (which holds if $n\beta^2 B/(p_0+q)^2 > c'(0)$). Then no such fee induces the planner's precision and deployment rule. At $t = n$ deployment is efficient given the source's precision, but the source's precision is strictly too low. At any $t < n$ the source either records at positive precision and deploys too rarely on a set of positive probability, or records nothing.*

Appendix F.2 gives the proof. The statement does not say that a better instrument does not exist. If x were verifiable, recipients could pay for it directly, as in the contracting benchmark of Section 8; the proposition says that fees tied to the deployment decision, which is what an observable installation allows, cannot do both jobs. The friction is natural when precision is a by-product of how operations are logged (sensor resolution, sampling frequency, labelling effort) and can be checked by recipients only after use. A fee contingent on the realized accuracy of the report would be a different instrument, which Proposition 2 does not rule out. In the example of Table 1 at $n = 8$ the cost of this failure is large. The source stops recording at a fee of about 0.63, under a tenth of the recipients' total gain, and social value is -3.56 at $t = n$, against -1.60 at $t = 0$ and -0.34 for the planner (Figure 4(b)). Only differences in these values are meaningful. On the fee grid I computed, welfare is highest at no payment: the fee corrects deployment, but destroys the learning that the recipients would have used. This is one parameter set; the proposition itself does not say that a positive fee is always harmful.

Table 2. Welfare decomposition for $n = 8$ recipients, Gaussian class.

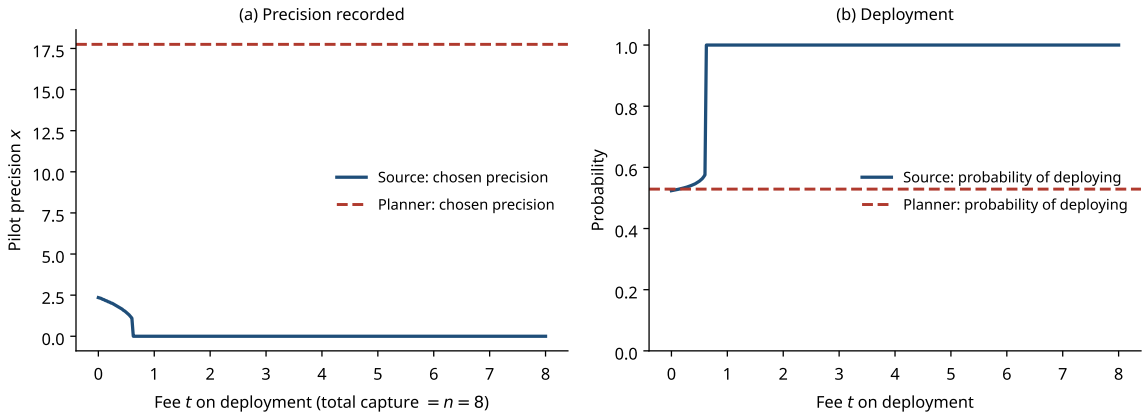

Figure 2: Payment fixes deployment but not recording: the source's choices as the fee rises, $n = 8$.

Notes. The figure uses the parameters of Table 1 with $n = 8$ recipients. Panel (a) shows the precision the source chooses as the fee rises, against the planner's precision of 17.75. Panel (b) shows the probability that the source deploys, against the planner's 0.53. The source stops recording at a fee of about 0.63, under a tenth of the recipients' total gain of 8, and from then on deploys with probability one. At that point deployment is efficient given what the source has recorded; the loss is in what it has recorded. Table 2 gives the welfare consequences. One parameter set.

Arrangement	Source's precision x	Probability of deployment	Recipients' baseline value	Social value
No payment ($t = 0$)	2.35	0.52	-2.16	-1.60
Fee $t = 0.65$	0.00	1.00	-7.22	-3.56
Full capture ($t = n$)	0.00	1.00	-7.22	-3.56
Planner	17.75	0.53	-0.39	-0.34

Notes. Parameters as in Table 1. The recipients' baseline value is $\beta^2 n\, w(x)$, which rises with the precision recorded. Social value is the source's own return plus the recipients' gain from the report plus their baseline value, net of the cost of recording; the fee is a transfer and does not enter. Only differences are meaningful. The values are computed by `make_figs_v21.py` and match `verify_v16.py`. One parameter set.

Why precision is hidden, and how much verification would restore it. Proposition 2 assumes that recipients can neither observe nor contract on x. That is the extreme case of a more general friction, and the friction can be put into the model. Suppose that with probability $\varphi \in [0,1]$ the recipients can verify the precision the source recorded (an audit of the log, a validation sample, a certified sensor), and that when they can, they pay the source what that precision is worth to them, $\beta^2 n\,[w(x) - w(0)]$, in addition to the fee $t = n$ on deployment. With probability $1 - \varphi$ the precision

is not verified and nothing is paid for it. The source's value is

$$V^{\varphi}(x) = V_S(x;n) + \varphi\,\beta^2 n\,[w(x) - w(0)],$$

so that $\varphi = 0$ is Proposition 1 at full capture, and $\varphi = 1$ is the planner's problem up to a constant: with full verification the recipients can contract on precision directly, as in the financing benchmark of Section 8.1.

Proposition 3 (Partial verification)**.** *Let $s_0 \geq 0$. (i)* Monotonicity. *The largest and the smallest maximizers of V^{φ} are nondecreasing in φ. (ii)* Never at the planner's choice unless verification is complete. *If $\varphi < 1$ and the planner's largest maximizer x_P is positive, every maximizer of V^{φ} is strictly below x_P. (iii)* A threshold in large markets. *Let $c(x) = kx$ and $\varphi^* = q/(p_0+q)$. If $\varphi < \varphi^*$, there is $\bar{n}$ such that for all $n \geq \bar{n}$ the unique maximizer of V^{φ} is $x = 0$. If $\varphi > \varphi^*$, there is $\bar{n}$ such that for all $n \geq \bar{n}$ every maximizer is positive.*

Appendix F.3 gives the proof. Parts (i) and (ii) say that hidden precision is not an on-off assumption: the source's precision rises continuously in the chance that it is verified and reaches the planner's only when verification is certain. Part (iii) says that verification must be good enough and not merely present. The fee on deployment already pays the source for the whole value of the report, so a payment for verified precision has to outweigh the option value of learning that the fee destroys. In a large market it does so only if φ exceeds $q/(p_0+q)$, the share of the final precision that the report supplies: the more the later report adds relative to what the earlier record already provides, the closer to certain verification must be. Below the threshold, adding some verification changes nothing.

In the example of Table 1 at $n = 8$ the threshold is $\varphi^* = 0.5$. The source records nothing for every $\varphi \leq 0.5$ on the grid I computed and social value stays at -3.56. At $\varphi = 0.6$ it records 13.2, about three quarters of the planner's 17.75, and social value is -0.37 against -0.34 for the planner; at $\varphi = 0.8$ and 0.9 it records 15.7 and 16.8 (Figure 3). For large n the first-order condition suggests that the source's precision is about $\sqrt{\varphi}$ times the planner's, because the marginal return to precision is about $\varphi B/x^2$ per recipient at large x against B/x^2 for the planner. Numerically the ratio is 0.7437, 0.7735 and 0.7746 at $\varphi = 0.6$ for $n = 8$, 10^4 and 10^8, against $\sqrt{0.6} = 0.7746$ (`verify_v22b_sqrtphi.py`). This is a heuristic and a numerical check, not a theorem. The threshold is a statement about large markets: at $n = 2$ in the same example recording resumes at $\varphi \approx 0.45$. Over 4,374 parameter configurations the source's maximizer is nondecreasing in φ in every one and is never above the planner's at $\varphi = 0.75$ (`verify_v22.py`). One parameter set is shown in Figure 3.

Proposition 4 (Leakage)**.** *Suppose the recipients are identical and that once one of them has paid, the report can be passed to the others at no cost, so that the source can collect no more than*

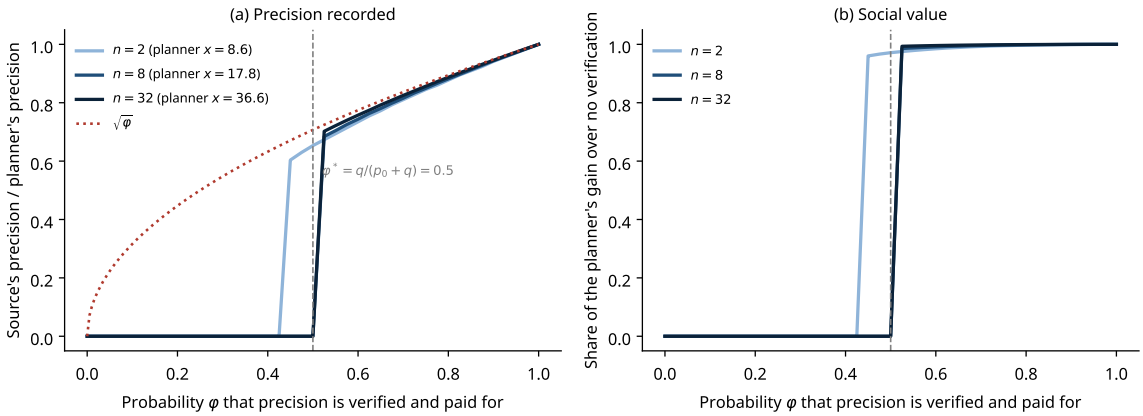

Figure 3: Partial verification: the source's precision and social value as the probability of verification rises.

Notes. The figure uses the parameters of Table 1. Panel (a) plots the source's precision as a share of the planner's, for $n = 2$, 8 and 32, with $\sqrt{\varphi}$ for reference; panel (b) plots the gain in social value over no verification as a share of the planner's gain. The dashed line is $\varphi^* = q/(p_0+q) = 0.5$. Below it, payment for deployment alone leaves the record at zero in the larger markets. One parameter set; the threshold in Proposition 3 is proved for large n.

one recipient's gain: the total capture is $t \in [0,1]$ whatever n. Let $s_0 \geq 0$. Then every precision the source chooses lies in $[0, x_S(0)]$, a set that does not depend on n, where $x_S(0)$ is the largest maximizer with no payment, while every planner maximizer is at least $\hat{x}(n)$. The planner's precision exceeds the source's by at least $\hat{x}(n) - x_S(0)$, which is unbounded in n when $c(x) = kx$.

Appendix F.4 gives the proof. This is a bound on capture, not an equilibrium analysis of resale: I do not model the resale game, which with heterogeneous recipients has multiple equilibria. What it shows is that the distortion does not shrink as the market grows, and in this benchmark it widens with the market. In the parameter set of Table 1 the source's precision is at most 2.36 and is zero at a fee of 1, while the planner's is 17.75 for $n = 8$ and 65.72 for $n = 100$ (the lower bound $\hat{x}(n) - x_S(0)$ is 14.64 and 62.82).

Resale and commitment. The capture bound is an assumption about what resale does, and it is worth saying which way each case cuts. If the source can commit to supply a single copy, and the holder can resell it at no cost, the holder is a monopolist reseller and the source can in principle collect up to n recipients' gain from that one sale: this is the case $t = n$ of Proposition 1, in which payment lowers precision. If the source cannot commit and sells copies to several recipients, any two holders compete in resale and drive its price toward zero, so each non-holder expects to obtain the report free and the source collects little beyond what the first buyers will pay; the bound $t \leq 1$ is the limiting case of that argument. The two cases bracket the outcome of a resale game without solving it, and the paper's conclusion holds in both: payment lowers precision when capture is large, and the gap is unbounded when capture is small. I do not claim that every equilibrium of a resale

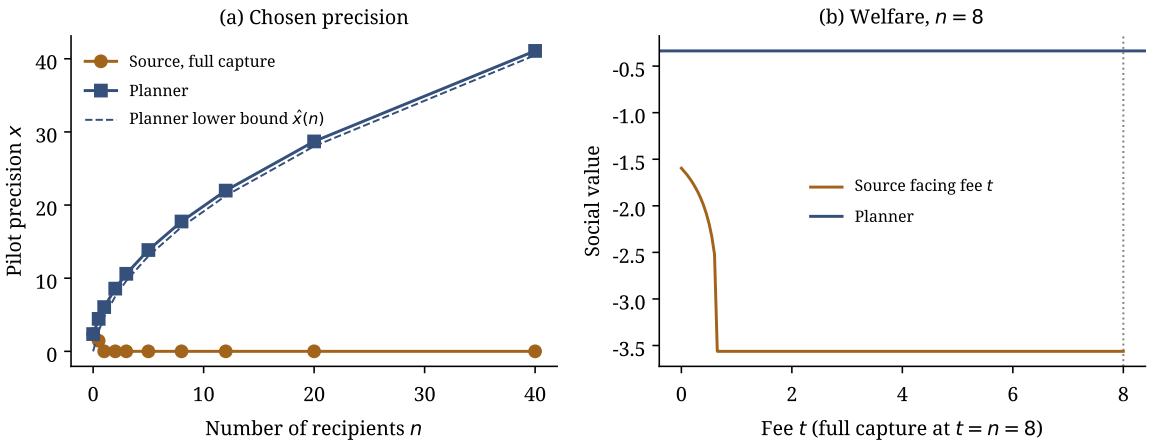

Figure 4: Gaussian class: chosen precision as the number of recipients varies, and social value as the fee varies.

Notes. The figure plots the entries of Table 1 in panel (a) and the welfare of the source facing a fee t, for $n = 8$ and the same parameters, against the planner's value in panel (b). Both panels are one parameter set. The fall in panel (b) at a fee of about 0.65 is the point at which the source stops recording.

game falls in one of the two cases, and I do not model recipients that compete with one another for a scarce installation slot, where the report may also carry a negative externality.

3 Environment, pilot design, and information rights

3.1 Technology and the uncertain assignment

A site has an observable type $t \in \{1,2\}$ and chooses a configuration $(a,k) \in \{0,1\}^2$. Here a indicates inspection and k indicates traceability. Gross operating surplus relative to the incumbent is

$$g_t(a,k) = \alpha a + \gamma k + \theta_t ak.$$

Net surplus subtracts the finite installation cost c_{ak}, with $c_{00} = 0$. Throughout the main analytical model, set the known standalone traceability return to $\gamma = 0$ and give configuration 01 a strictly positive cost. That configuration is then dominated. This simplifies the choice problem without implying that traceability has no standalone value. Each site operator bears its own configuration costs and controls its own complete configuration.

The interaction vector takes one of two values,

$$(\theta_1, \theta_2) \in \{(h,\ell),(\ell,h)\}, \qquad \Delta = h - \ell > 0.$$

The initial probability that type 1 has the high interaction is $p_0 = 1/2$. The standalone return α is independently uncertain, with $\mathbb{E}|\alpha| < \infty$. Both pilots reveal it before installation decisions. The finite interaction levels h and ℓ are known; the assignment is not. Thus even the complete distribution of interaction returns, not just its mean, can be known before the informative pilot is run. The interaction need not be positive at every type; the numerical examples have $h > 0 > \ell$.

The fixed sum $\theta_1 + \theta_2 = h + \ell$ imposes complete negative dependence between the two type returns. It is deliberately strong: it isolates assignment uncertainty and implies that evidence favorable to type 1 is unfavorable to type 2. Appendix G shows that the information results do not need it. The dynamic results of Sections 6 to 8 do use it.

3.2 The two pilots

A source operator has a budget for eight configuration-cell observations. Pilot A observes two repetitions of 00 and 10 at each type. Pilot B observes 00, 10, 01, and 11 once at each type. Their observation counts are equal. They are assumed to have the same common real experiment cost C.

For the analytical benchmarks, the 00 and 10 contrasts reveal α without error. Under B, the 01 outcome identifies the known γ, and each 11 contrast adds an independent measurement error to its interaction return. One explicit measurement system is

$$M_{t,00} = 0, \quad M_{t,10} = \alpha, \quad M_{t,01} = \gamma,$$

$$M_{t,11} = \alpha + \gamma + \theta_t + \varepsilon_t.$$

Then $Z_t = M_{t,11} - M_{t,10} - M_{t,01} + M_{t,00} = \theta_t + \varepsilon_t$. This system makes common learning of α explicit. Equal counts alone would not guarantee equal precision or equal real costs in a general measurement model. In particular, eight observations with ordinary noise in every cell would not perfectly reveal α.

A linked B report gives (Z_1, Z_2), with the correspondence to types retained. Its pooled version gives only $\{Z_1, Z_2\}$. The pooled report retains the outcomes but not their match to the types known at investment. A leaves the assignment prior unchanged. These pilots are an admissible menu, not a solution to an unrestricted optimal experimental-design problem. The pilot comparison is conditional on a pilot being commissioned; laboratory entry and the total volume of experimentation are not modeled.

The source receives immediate real net pilot proceeds b_q, $q \in \{A, B\}$, before subtraction of the common cost C. They can represent operating output or other within-source benefits of conducting a particular pilot. They are specified reduced-form surpluses, not estimated values and not transfers

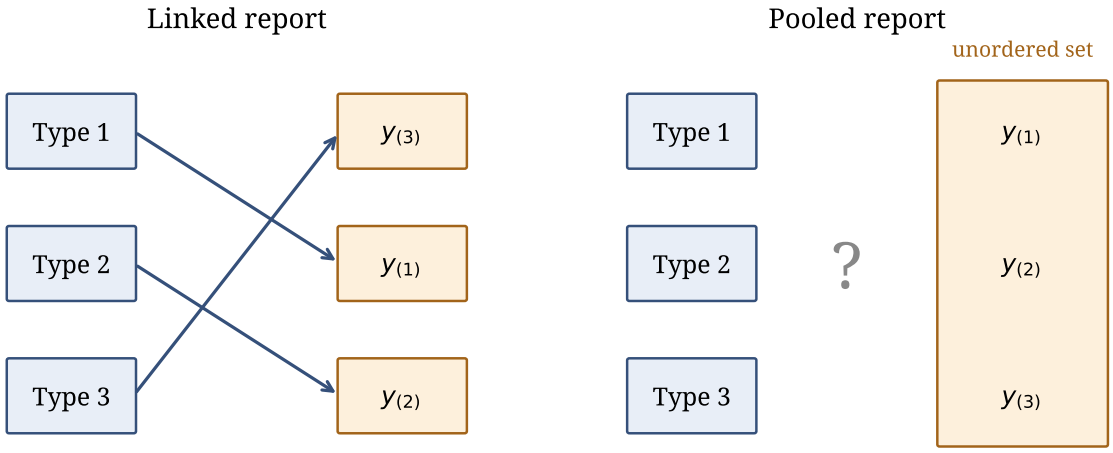

Figure 5: The same outcomes, recorded with and without their correspondence to types.

Notes. The figure is a schematic with no numerical content. The linked report keeps the match between each type and its outcome; the pooled report keeps the outcomes but not the match. This is the sense in which the recorded form, and not only the amount, of the data matters.

from the recipient. Let $d = b_A - b_B$. A difference in b_q can make a less informative pilot attractive to the source. In Sections 4 to 7 I take that difference as given rather than deriving it from the measurement count. Appendix G.4 and Section 3 derive it from a cost of recording precision, and the Gaussian class of Section 3 has no exogenous difference at all: the source chooses a continuous precision and bears its cost.

3.3 Ownership, access, and timing

The source owns the pilot and the type-1 installation opportunity. A distinct recipient owns the type-2 installation opportunity. Both are risk neutral. A social planner sums their real net payoffs. The benchmark fixes a dissemination arrangement with no payments from the recipient to the source. A report may be accessible, but the source has no use-based claim on the recipient's return. This is an institutional restriction, not a consequence deduced from data nonrivalry. Section 8 relaxes it.

At date 0 the source chooses a pilot, pays C, and receives b_q. Both operators observe the resulting α and public posterior p_q. At date 1 the source chooses its configuration. A joint installation generates a validated report Y about θ_1. At date 2 the recipient chooses its configuration. Conditional on production, the operating report reaches the recipient independently with probability $\lambda \in [0,1]$. The discount factor is $\beta \in (0,1]$.

There is a separate installation opportunity at each date and site. Installing at type 1 does not mechanically use up type 2's opportunity. This is the principal sequential model. Section 5 also

studies a recipient with only one installation across two types as a separate allocation benchmark, using the same return and information objects but a different feasibility constraint.

The source acts using the public pilot posterior, with no additional private signal before installation. Its decision rule and costs are public. Observing deployment therefore conveys no additional information conditional on the pilot posterior. A report that is absent because no joint deployment occurred is not an independent bad signal. Hidden source information would require an additional action-signaling model.

4 Type correspondence and the value of an installation opportunity

This section is a motivating benchmark. Section 5.1 shows that a pooled report cannot identify the match between outcomes and types. Section 5.2 then uses a scarce installation slot, so that a recipient's choice includes which type to equip, in order to show the value of the correspondence as simply as possible. Sections 6 to 8 return to separate opportunities for the source and the recipient, which is the structure on which the rest of the paper depends.

4.1 Identification from a pooled report

Proposition 5 (Type-match identification)**.** *Suppose contrast errors have a state-independent exchangeable joint distribution. The pooled report* $\{Z_1, Z_2\}$ *has the same conditional distribution under the two assignments of* (h, ℓ)*. Its posterior assignment probability equals its prior. With independent normal errors of variance* $\sigma^2 > 0$*, the linked posterior satisfies*

$$\operatorname{logit} p(Z_1, Z_2) = \operatorname{logit} p_0 + \frac{\Delta(Z_1 - Z_2)}{\sigma^2}.$$

Under the symmetric prior, a pooled report leaves $p = 1/2$. A linked report need not. The result depends on joint exchangeability and absence of other information that reveals the correspondence. Equality of the two error marginals alone is not sufficient. Appendix A provides the proof.

Intuition. The pooled report contains the same two numbers under either assignment. Only their pairing says which number belongs to which type, and an investor can act only on types it can observe.

The proposition concerns a specific loss of conditional information. If the published report already contains comparable type-specific four-configuration means, its type-specific interaction can be computed without individual record IDs. Unequal sampling weights, other observable outcomes,

or asymmetric measurement errors might also reveal the assignment. A report with those features is not the pooled experiment defined here.

4.2 A scarce-slot valuation benchmark

For this subsection only, let a recipient have at most one installation across the two types. It may choose 00, 10, or 11 and, for 11, choose which type receives the installation. Conditional on α and posterior p, its value is

$$V(p,\alpha) = \max\{0, \alpha - c_{10}, \alpha + \ell + \Delta \max\{p, 1-p\} - c_{11}\}.$$

Corollary 1 (Investment value of type-linked data)**.** *Conditional on common learning of α, a linked report weakly raises expected optimized installation value relative to its pooled version. The inequality is strict if no single feasible installation maximizes conditional expected payoff almost surely across the linked posterior realizations.*

The linked posterior is a martingale and V is the maximum of the expected payoffs of a finite set of actions, hence convex in the assignment probability. The corollary is a standard information-value comparison. Its economic content is the allocation of an installation opportunity when the distribution of interaction returns is held fixed.

For a transparent numerical example, let $\alpha \in \{-0.3, 0.3\}$ with equal probabilities, $h = 0.8$, $\ell = -0.4$, $c_{10} = 0.1$, and $c_{11} = 0.3$. Under the pooled report, the expected return to 11 at either type is $\alpha - 0.1$, the same as 10. The recipient installs only in the high-α state, giving ex ante value 0.1. A perfect linked report locates the high-interaction type. Joint installation there gives 0.8 when $\alpha = 0.3$ and 0.2 when $\alpha = -0.3$, giving ex ante value 0.5.

Set $b_A = 0.15$ and $b_B = 0$. Before the common pilot cost, A produces total value 0.25 and exact linked B produces 0.50. The source prefers A, but the planner prefers linked B. More generally, the perfect-report reversal occurs for $0 < d < 0.4$. It is a conditional result: if the assignment is already known, or if the interaction does not vary across types, there is no additional matching value.

The comparison also depends on precision. With normal contrast noise, linked B gives recipient values 0.309092 at $\sigma = 1$ and 0.236397 at $\sigma = 1.5$. At the latter precision, B's total value falls below A's 0.25. The model does not say that a report with a type label is always worth its opportunity cost. All numbers here and below are constructed theoretical values, not data estimates.

5 Deployment generates another investment signal

5.1 Public beliefs and the recipient's option value

Return to the principal sequential model. Under A, $p_A = 1/2$. Under linked B, let $Z_{0t} = \theta_t + \varepsilon_{0t}$, where the errors are independent normal with variance σ_0^2. Write $D_0 = Z_{01} - Z_{02}$. The symmetric-prior posterior is

$$\operatorname{logit} p_B = \frac{\Delta D_0}{\sigma_0^2}.$$

The source's established option includes not installing:

$$s(\alpha) = \max\{0, \alpha - c_{10}\}.$$

The expected return to its joint installation and the incremental private return are

$$J_1(p,\alpha) = \alpha + \ell + \Delta p - c_{11}, \qquad r_1(p,\alpha) = J_1(p,\alpha) - s(\alpha).$$

Only joint deployment generates $Y = \theta_1 + \eta$, where $\eta \sim N(0, \sigma_1^2)$ is independent of the pilot errors and assignment. If the report is received,

$$\operatorname{logit} p_Y = \operatorname{logit} p + \frac{\Delta}{\sigma_1^2}\left(Y - \frac{h+\ell}{2}\right).$$

Type 2 has the high interaction when type 1 has the low interaction. The recipient's value without the operating report is therefore

$$w(p,\alpha) = \max\{s(\alpha), \alpha + h - \Delta p - c_{11}\}.$$

Define the conditional information benefit, before the access probability is applied, by

$$H(p,\alpha) = \mathbb{E}[w(p_Y,\alpha) \mid p,\alpha] - w(p,\alpha).$$

This definition keeps technical surplus separate from information. The report does not improve the recipient's machine. It changes the arrangement the recipient chooses given that machine's return distribution.

5.2 The deployment wedge

Proposition 6 (Deployment and the information wedge)**.** *The report value is nonnegative. It is strictly positive when* $0 < p < 1$, $0 < \sigma_1 < \infty$, *and*

$$\alpha + h - c_{11} > s(\alpha) > \alpha + \ell - c_{11}.$$

Under the no-transfer dissemination arrangement, the source chooses joint deployment if $r_1 \geq 0$. *The planner chooses it if*

$$r_1 + \beta\lambda H \geq 0.$$

If $\beta\lambda > 0$, *the private cutoff lies strictly inside* $(0,1)$, *and* H *is strictly positive at that cutoff, the two rules differ on an open interval of posterior beliefs.*

To see the wedge, the source compares the real date-1 returns s and J_1. The planner also includes the date-2 recipient's optimized return. Joint deployment changes that recipient value from w to $w + \lambda H$. Discounting back to date 1 gives the additional term $\beta\lambda H$. Strict information value follows from a report with positive probability of crossing the recipient's action threshold. Appendix B gives a full proof and a closed-form expression.

Intuition. The source is rewarded for its own operating return but not for what its installation teaches the recipient. When the lesson is valuable enough, the planner installs where the source would not.

The access condition matters. At $\lambda = 0$ the report cannot affect the recipient, so the source's and planner's deployment thresholds are identical. An interior private threshold alone does not imply a strict wedge. At a degenerate prior, or when the same recipient arrangement is optimal in both states, $H = 0$ as well.

An open disagreement interval is not automatically an event of positive probability after every pilot. A's posterior is an atom at $1/2$, which can miss the interval. Finite-noise B has full posterior support on $(0,1)$, so the interval has positive conditional probability when the relevant α is in the support.

Because returns are negatively related in this benchmark, a favorable source report can reduce type 2's expected joint return. A disappointing source report can increase it. The information benefit is nonnegative because the recipient can reoptimize; it does not require every signal to increase investment or every recipient to adopt. Choosing an established arrangement after learning can be an efficient response.

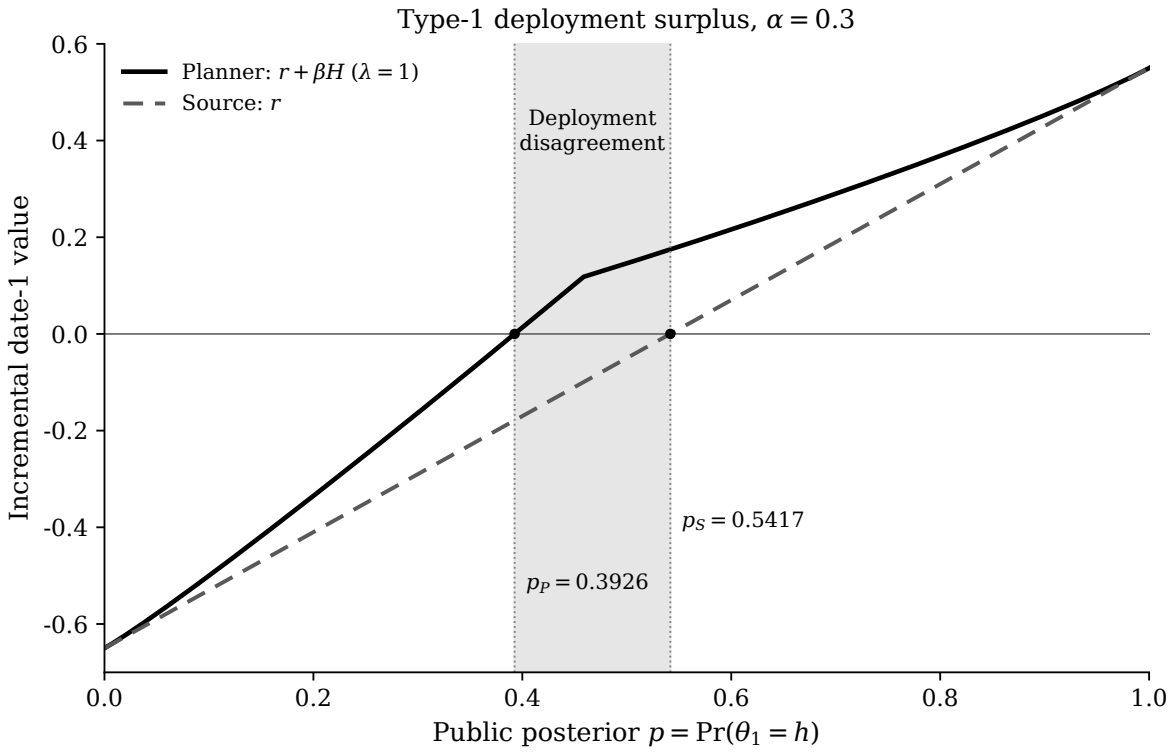

Figure 6: Private and social incremental deployment returns as functions of the public posterior.

Notes. The figure uses the dynamic numerical parameters below, with $\alpha = 0.3$ and full operating-report access. The social curve includes the recipient's information value. Shading identifies beliefs at which the planner installs jointly while the source does not. This is a mathematical illustration, not an estimated adoption relationship.

5.3 Standalone capability need not move joint adoption

Corollary 2 (Standalone improvements and joint adoption)**.** *For $\alpha \geq c_{10}$, if an increase in α raises the return to 10 and 11 equally while interaction returns, installation costs, and report precision remain fixed, both the private and social joint-deployment rules are independent of α.*

In this range,

$$r_1 = \ell + \Delta p - (c_{11} - c_{10}),$$

$$w(p, \alpha) = \alpha - c_{10} + \max\{0, h - \Delta p - (c_{11} - c_{10})\}.$$

The common α term cancels from the information benefit H. Thus a stronger standalone capability can increase both operators' profits without changing whether adding traceability is worthwhile. Evidence about the interaction, or a change in the incremental installation cost, can move that decision.

This result is conditional on the additive-return specification and the profitable standalone outside option. It does not apply if better inspection changes the interaction θ_t, if α falls below the standalone threshold, or if capability improvements change measurement quality. It is a prediction about a particular adoption margin, not a general explanation of observed AI adoption.

6 Initial experimentation with endogenous continuation

6.1 Full date-0 objectives

The source and planner must be allowed the same physical deployment opportunities after either pilot. In particular, the planner can induce a joint installation and obtain an operating report after A as well as B. Let $P(x) = \max\{x, 0\}$ and let $\mathbb{E}_q$ average over α and the public posterior induced by pilot q. The source's date-0 value is

$$U_q = b_q - C + \beta \mathbb{E}_q[s(\alpha) + P(r_1(p_q, \alpha))].$$

The optimized social value is given below. Arguments are suppressed: $s = s(\alpha)$, while r_1, H, and w are evaluated at (p_q, α).

$$W_q(\lambda) = b_q - C + \beta \mathbb{E}_q[s + P(r_1 + \beta\lambda H)] + \beta^2 \mathbb{E}_q[w].$$

The last term remains even if the source does not deploy jointly or the operating report is inaccessible. The recipient still knows the initial pilot posterior. Setting $\lambda = 0$ removes the additional operating report, not the information already delivered by B.

Define source and planner continuation gains from B relative to A before the difference in immediate proceeds:

$$\Delta_S = (U_B - b_B) - (U_A - b_A),$$

$$\Delta_P(\lambda) = (W_B(\lambda) - b_B) - (W_A(\lambda) - b_A).$$

The source prefers A while the planner prefers B exactly when

$$\Delta_S < d < \Delta_P(\lambda).$$

The interval need not exist. The planner comparison combines the initial signal's direct benefit

to the recipient, its benefit to the source's deployment choice, and the operating reports induced by reoptimization.

6.2 Access and the relative value of the pilot

Proposition 7 (Report access and experiment choice)**.** *For each pilot, $W_q(\lambda)$ is nondecreasing and convex in λ. At access probabilities with zero probability of a deployment tie,*

$$\frac{dW_q}{d\lambda} = \beta^2 \mathbb{E}_q \left[H(p_q, \alpha) \mathbf{1}\{r_1(p_q, \alpha) + \beta\lambda H(p_q, \alpha) > 0\} \right].$$

Write the expected information value at selected deployments as

$$K_q(\lambda) = \mathbb{E}_q[H\mathbf{1}\{r_1 + \beta\lambda H > 0\}].$$

Consequently, wherever both derivatives exist,

$$\frac{d\Delta_P}{d\lambda} = \beta^2[K_B(\lambda) - K_A(\lambda)].$$

There is no general sign for this difference. At a tie, the left derivative uses strict selection and the right derivative includes tied deployments with $H > 0$; derivatives exist almost everywhere.

For a fixed pilot realization, the only access-dependent component of value is the maximum of an affine function of λ and zero. Its slope is nonnegative because $H \geq 0$. This establishes monotonicity and convexity. The envelope derivative counts the information value only at deployments that are actually selected. Appendix C formalizes the result, including the atom in A's posterior distribution.

The comparative static distinguishes two forces. A more informative initial pilot can put the source closer to deployment and thereby create more opportunities for operating reports. But it can also leave less uncertainty for those reports to resolve. Access strengthens the joint pilot's relative advantage only when its selected deployments generate a larger expected report benefit than selected deployments after A. It weakens that advantage under the reverse inequality. In words, a later report substitutes for an informative initial pilot to the extent that it is most valuable after the uninformative one.

Corollary 3 (Sign of the access effect)**.** *Let $K_q(\lambda)$ be as defined in Proposition 7.*

(a) Substitution. *If linked B reveals the assignment, so that $p_B \in \{0, 1\}$ almost surely, then $K_B(\lambda) = 0$ and*

$$\frac{d\Delta_P}{d\lambda} = -\beta^2 K_A(\lambda) \leq 0.$$

The inequality is strict whenever the planner selects, with positive probability after A, a deployment whose report has positive value.

(b) Complementarity. *If $K_A(\lambda) = 0$, so that after A the planner selects no deployment whose report has positive value, and $K_B(\lambda) > 0$, then*

$$\frac{d\Delta_P}{d\lambda} = \beta^2 K_B(\lambda) > 0.$$

Appendix F gives the proof. Part (a) says that later reports substitute for an initial pilot that already resolves the uncertainty they would resolve: a report about an assignment already known has no value. Part (b) says that access can raise the relative value of the informative pilot when, without it, reports would not be produced at all. Between these cases the sign is that of $K_B - K_A$, which depends on primitives. In the example of Section 7.3, $K_A = 0$ for $\lambda < \lambda_c \approx 0.2551$ and $K_A = 0.1031$ above it, while K_B lies between 0.0056 and 0.0087 at the values tested (Appendix F.5). The slope of Δ_P is therefore about $+0.005$ below λ_c and between -0.087 and -0.085 above it, in line with the rise and fall in Figure 6.

Convexity of social value in access has an extensive-margin interpretation. A small chance of transmitting a useful report may not justify a privately costly installation. A larger chance may induce deployment and create the report. This optimization effect is not a statement that exogenously supplied signals are universally complementary, and it is not an equilibrium-multiplicity result.

6.3 A fully specified dynamic example

For this example take $h = 0.8$, $\ell = -0.4$, $\alpha \in \{-0.3, 0.3\}$ with equal probabilities, $c_{10} = 0.1$, $c_{11} = 0.35$, $\beta = 0.95$, $\sigma_0 = 0.6$, $\sigma_1 = 0.5$, and $b_A = 0.16$, $b_B = 0$. The joint cost differs from the static benchmark, and the two sites have separate installation opportunities. Report noise parameters are standard deviations.

At $\lambda = 1$, the source continuation values before immediate proceeds and common cost are 0.095000 after A and 0.226176 after B. Adding immediate proceeds gives private totals before C of 0.255000 and 0.226176. The planner continuation values are 0.254591 and 0.447541; its totals before C are 0.414591 and 0.447541. Thus

$$\Delta_S = 0.131176 < d = 0.16 < \Delta_P(1) = 0.192950.$$

A local deployment disagreement occurs at $\alpha = 0.3$ and $p = 0.45$. The source's incremental return is $r_1 = -0.11$. The conditional recipient information value is $H = 0.224047$. At date 1 the planner's incremental value is $-0.11 + 0.95H = 0.102845 > 0$. The private loss is outweighed by

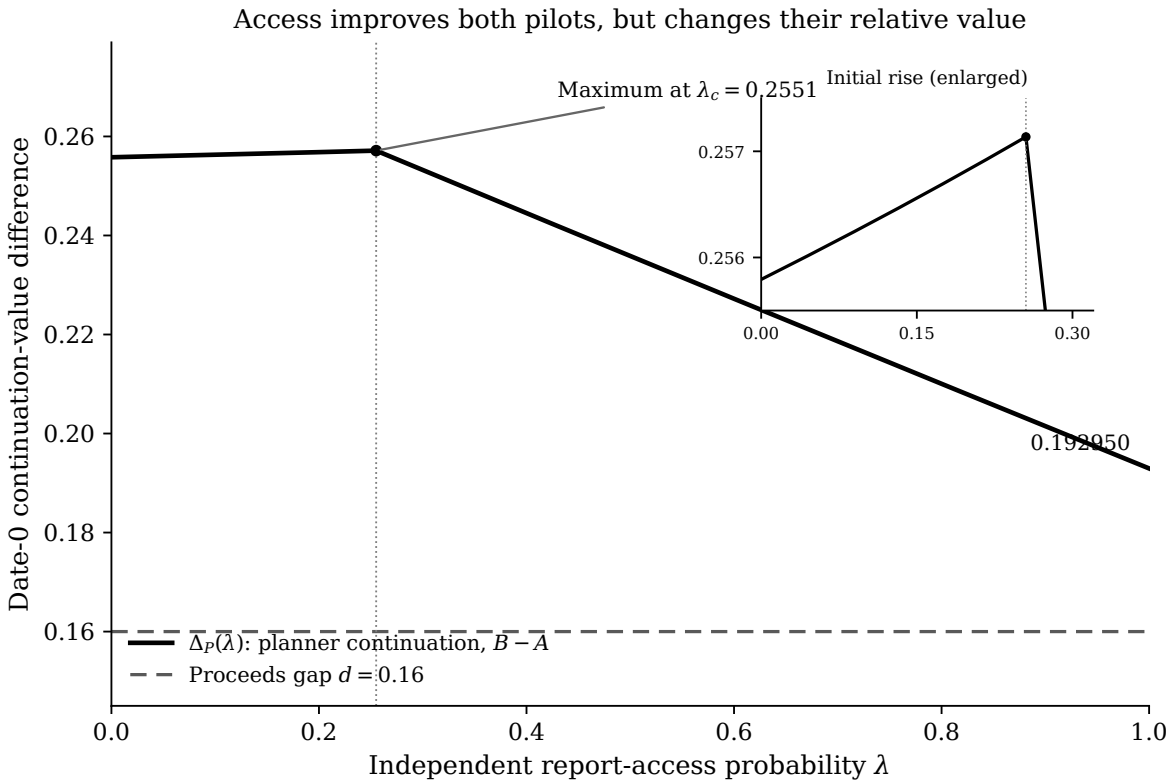

Figure 7: The joint pilot's social continuation advantage as operating-report access varies.

Notes. The figure plots $\Delta_P(\lambda)$ for the stated parameters and the immediate-proceeds difference $d = 0.16$. The same later deployment menu is available after both pilots. The figure is computed from the model's optimized values, not from an exogenous sequence of always-produced reports.

the discounted information benefit.

With no operating-report access, planner continuation values are 0.185250 after A and 0.441043 after B, yielding $\Delta_P(0) = 0.255793$. Opening access raises both absolute continuation values but reduces the relative B advantage to 0.192950. At $d = 0.16$, the planner chooses B at both endpoints. Therefore the later report is not necessary for the pilot-ranking reversal in this example, even though it can change deployment.

The complete curve is not even monotonically declining in this example. Initially B's advantage rises slightly. After A in the high-α state, the source's private joint increment is -0.05 and $H(1/2, 0.3) = 0.206295778$. At $\lambda_c = 0.255126787$, the planner starts inducing joint deployment after A. The relative advantage reaches 0.257136017 at that activation kink and then falls. The minimum on the plotted access range is the full-access endpoint, still above $d = 0.16$; the planner selects B throughout this parameter set.

The endpoint comparison is a counterexample to necessary amplification, while the activation kink illustrates why endogenous report production must be retained. Neither feature is a general

substitutes characterization. Any claim about the sign over an access interval must use the selected-deployment comparison in Proposition 7.

7 Contracts, appropriation, and the welfare boundary

7.1 Financing the static linked experiment

In the scarce-slot example, the recipient gains 0.4 from an exact linked B report relative to A, while the source gives up 0.15 of own proceeds. An enforceable procurement contract that requires B and its linked report can transfer an amount T satisfying $0.15 < T < 0.4$. Both parties then prefer the contract to A. A commitment to the experiment and report quality is important; an unverifiable promise to supply useful data is not the same contract.

This payment need not depend on every later use being technologically excludable. Even if the delivered report can be copied, an identified beneficiary may finance its production in advance when the source can commit to the contracted pilot. Nonrivalry alone therefore does not prove that a linked experiment must be publicly financed.

7.2 A deployment-contingent capture benchmark

As a counterfactual to free dissemination, suppose the source can commit to report-access rights, the public posterior and report-production event are contractible, and there are no financing or contracting costs. A recipient that does not contract has the same initial pilot information but does not receive the new operating report. This changes the benchmark's access entitlement; it is not a fee charged for evidence the recipient would already obtain freely. The schedule is committed before pilot selection. Conditional on joint deployment, the recipient pays a date-1 fee

$$F_\rho(p,\alpha) = \rho\beta\lambda H(p,\alpha), \qquad 0 \le \rho \le 1,$$

before the operating signal is realized. This is a stipulated fee schedule capturing a fraction of the recipient's expected information gain, not a bargaining solution. The recipient's expected date-1 benefit from the report is $\beta\lambda H$, so the schedule satisfies ex ante participation at that margin.

Proposition 8 (Partial and full internalization)**.** *Under this contract, the source deploys jointly when*

$$r_1 + \rho\beta\lambda H \ge 0.$$

The induced deployment set expands as ρ rises. At $\rho = 1$, the source's deployment rule equals

the planner's. This equality does not generally align date-0 pilot choice. Under common ownership and control of both sites, the date-0 objective equals W_q and both margins are internalized.

The fee is a transfer: it enters the source's payoff positively and the recipient's negatively, and cancels from their total. At full capture, the source values the incremental report benefit from installing, but still need not value the initial pilot's direct effect on the recipient's baseline information. Its objective contains the source's optimized continuation term, not the additional term $\beta^2 \mathbb{E}_q[w]$ in total welfare. Paying for deployment evidence alone is therefore different from owning or contracting over the entire continuation.

In the same dynamic example with full access and full capture, source continuation values are 0.164341 after A and 0.232674 after B. Its continuation gain from B is only 0.068333, below $d = 0.16$, so the source still chooses A even though its deployment rule is efficient. The planner's continuation gain remains 0.192950. Here, pricing only the later report rewards the path that leaves more information for that report to resolve; it does not price the recipient's direct gain from the initial linked pilot.

Participation is assessed before the report and access event are realized. At $\rho = 1$ it is weak, not strict; acceptance or a consistent tie convention is required. This does not establish ex post participation after every realized signal or nonreceipt. If the report would remain freely available without a contract, the no-report outside option is invalid and the fee argument does not apply.

Common ownership internalizes the full real continuation return by definition. An enforceable agreement that gives equivalent control and incentives can reproduce that comparison, subject to participation and verifiability. I do not claim that an arbitrary upfront payment or a simple data license implements every state-contingent decision. Appendix D proves the threshold and clarifies the pilot distinction.

7.3 Endogenous pricing and disclosure

Section 8.2 stipulated a fee. This subsection derives one from a pricing problem. Suppose a unit mass of potential recipients, indexed by j, and let recipient j's information value from the report be $\xi_j H(\pi)$. Here H is the aggregate information value of Appendix G.2 and $\xi_j \geq 0$ is a recipient-specific scale with distribution G and mean one. The scale represents differences in how much the report can improve a recipient's decision, such as the size of the installation at stake. The report is excludable: after deploying, the source can withhold it until paid. Deployment and the public belief π are observable. After deploying, the source posts a uniform price f. Recipient j buys if $\beta \xi_j H(\pi) \geq f$ and then receives the report with probability one. Payment precedes the realization of the report, so the participation constraint that matters is the one at the time of purchase, and it holds by construction.

Let $D(x) = \Pr(\xi \geq x)$ and

$$\rho^* = \max_{x \geq 0} xD(x).$$

Proposition 9 (Endogenous pricing and partial capture)**.** *(i) The optimal price is $f^*(\pi) = x^*\beta H(\pi)$, where x^* attains ρ^*, and the resulting revenue is $\rho^*\beta H(\pi)$. This is the subgame-perfect price: no commitment before deployment is needed.*

(ii) $0 < \rho^* \leq 1$, and $\rho^* = 1$ if and only if ξ takes at most one strictly positive value almost surely.

(iii) The source deploys if and only if $r(\pi) + \rho^*\beta H(\pi) \geq 0$. Its deployment set contains the no-payment set and is contained in the planner's. If $\rho^* < 1$ and $H > 0$, the containment in the planner's set is strict: for every belief with $-\beta H \leq r < -\rho^*\beta H$ the planner deploys and the source does not.

(iv) Recipients with $\xi < x^*$ are excluded even though supplying the report costs nothing. The value they forgo is $\beta H\,\mathbb{E}[\xi \mathbf{1}\{\xi < x^*\}]$, which is positive whenever $\Pr(0 < \xi < x^*) > 0$.

Appendix F.6 gives the proof. Two features matter for the paper's argument. First, the partial capture rate ρ^* does not depend on the belief, so the pricing problem is exactly the fee schedule of Section 8.2 with $\rho = \rho^*$, and Proposition 15 applies to it. Second, exclusion means that pricing is not a free way to restore efficiency: it trades the deployment wedge against an access wedge. The fee of Section 8.2 corresponds to the case $\rho^* = 1$, which requires the seller to know each recipient's valuation.

Example. Let ξ be uniform on $[0,2]$. Then $D(x) = 1 - x/2$, $x^* = 1$, and $\rho^* = 0.5$. The source charges βH and sells to the half of recipients with $\xi \geq 1$. It captures 50 percent of the aggregate information benefit. Buyers keep 25 percent, and the remaining 25 percent is lost to exclusion. In the dynamic example of Section 7.3 with full access, the deployment threshold in the public belief at $\alpha = 0.3$ is $p = 0.5417$ with no payment, 0.4524 with this price, and 0.3926 for the planner. The source's continuation gain from the informative pilot is 0.1312 with no payment, 0.1112 with this price, and 0.0683 at full capture, always below $d = 0.16$, while the planner's is 0.1930. The source therefore chooses the less informative pilot at every capture rate, and the planner does not.

Corollary 4 (Disclosure and linking)**.** *Suppose that, conditional on deployment, the source can release the report in a linked version L or in a version P that is any garbling of L, such as the same outcomes with the type labels removed. Linking costs the source $k > 0$, for example for data cleaning or privacy protection. Let $H^L \geq H^P$ be the corresponding information values. Under no payment the source never links. Under the pricing of Proposition 9, with the same heterogeneity for both versions, the source links if and only if $\rho^*\beta(H^L - H^P) \geq k$, while the planner links if and only if $\beta(H^L - H^P) \geq k$. For $\rho^* < 1$ and $H^L > H^P$ there is a nonempty interval of costs, $(\rho^*\beta(H^L - H^P), \beta(H^L - H^P)]$, at which the planner links and the source does not.*

The proof of Corollary 4 is in Appendix F.6. The inequality $H^L \geq H^P$ is the value-of-information result applied to a garbling. The corollary is the disclosure analogue of Proposition 6. It says that the choice of which data to release, as well as the choice of whether to deploy, is distorted toward too little matching.

7.4 What a policy comparison would require

The no-transfer benchmark is informative about a dissemination institution but does not identify the best one. A policy analysis would need a specified impediment to appropriation, such as noncontractible report quality, costly bargaining with dispersed beneficiaries, or inability to commit before an otherwise public report is generated, and it would have to include cheaper private experiments that supply the same payoff-relevant posterior. A larger λ need not maximize net welfare when dissemination has real costs. The paper's welfare statement is therefore conditional: within this environment, the source omits the recipient's information benefit. Choosing a reporting policy is a separate problem.

7.5 What the results say to those who collect, buy, or regulate data

The propositions have direct readings for three decision makers. They are conditional implications of the model, not recommendations established by evidence.

A firm deciding what to record. Recording outcomes together with the observable conditions under which they arose, at some cost, is what makes a report usable by others (Propositions 5 and 13). A firm that sells or shares reports should expect to under-invest in this linking when it captures only part of the recipients' gain (Corollary 4), and a plain fee will not correct that when the choice is made at the pilot stage (Proposition 8).

A firm deciding whether to pay for data. Paying for the later report of a deployment can secure the deployment, but it does not secure the experiment that precedes it. A buyer who wants the informative pilot needs a contract that covers the pilot itself, which requires that its quality be verifiable (Section 8.1), or common ownership of the continuation. Verification that is only partial does not restore the planner's precision, and in a large market it does nothing at all below a threshold (Proposition 3).

A regulator or platform designing access. Greater access raises social value but has no universal effect on the value of an informative pilot (Corollary 3). A rule that mandates broad sharing of operating reports can reduce the incentive to run an informative pilot when later reports would have resolved the same uncertainty. Whether open access helps therefore depends on where later reports are most valuable, which the model locates but does not measure. # Scale, and what the two models

share This section does two things. It shows how the discrete model's pilot-choice results scale with the number of recipients, and it states the object that the discrete model and the Gaussian core model have in common: the sign of the effect of payment on the informative pilot is the sign of one change in the recipient's value of information. ## How the discrete results scale with the number of recipients Section 8.3 and Appendix G.2 treat a population of recipients with total mass one. Let the mass be $n > 0$, with the same distribution of scales ξ. Each recipient's information value $\xi_j H$ and recipient value $\xi_j w$ are unchanged, so the aggregate values are nH and nw. Recipients do not interact: each faces its own decision problem and none competes for the same slot. This is a limitation that I return to below.

Proposition 10 (Scale of the recipient population)**.** *(i) The optimal capture rate ρ^* of Proposition 9 does not depend on n or on the belief, and revenue after deployment is $\rho^* n\beta H(\pi)$.*

(ii) The source deploys if and only if $r(\pi) + \rho^* n\beta H(\pi) \geq 0$ and the planner deploys if and only if $r(\pi) + n\beta H(\pi) \geq 0$. The surplus that the source does not capture, $(1-\rho^*)n\beta H(\pi)$, is proportional to n.

(iii) Let $\Delta_S(\rho n)$ denote the source's continuation gain from the informative pilot when the fee is $\rho n\beta\lambda H$, as in Section 8.2. The source's gain at price f^* is $\Delta_S(\rho^* n)$, and the planner's is $\Delta_S(n) + \beta^2 n\,(\mathbb{E}_B[w] - \mathbb{E}_A[w])$.

Appendix F.7 gives the proof. The proposition says that a larger population of recipients does not by itself cure the wedges, because both the captured and the uncaptured parts of the gain scale with n. Whether the pilot-choice wedge widens or narrows depends on how Δ_S varies with the fee, and the model does not sign it in general.

Example. Take the parameter set of Section 7.3 with full access, $\alpha = \pm 0.3$ averaged as there, $d = 0.16$, and $\rho^* = 0.5$. The source's gain from the informative pilot at the monopoly price is 0.1112 for $n = 1$ and 0.0683, 0.0259, -0.0162, -0.0997 and -0.1826 for $n = 2, 3, 4, 6$ and 8. It is below d for every $n \geq 1$, and it becomes negative: the source then prefers the less informative pilot even if the pilots cost the same. The planner's gain is 0.1930 for $n = 1$ and rises to 0.2331, 0.2742, 0.3159, 0.4003 and 0.4858, and exceeds d for every n above about 0.22. Here the revenue from the later report is worth more when the report resolves more uncertainty, so the source's reward from keeping uncertainty grows with the number of buyers. This is a numerical statement for one parameter set. I have not proved that the source's gain is monotone in n in general.

Proposition 11 (Capture, scale, and the pilot-choice wedge)**.** *Let $t = \rho n$ be the total capture, so that the fee is $t\beta\lambda H$, and let $\Delta_S(t)$ be the source's continuation gain from the informative pilot B over A, as in Proposition 10(iii). Let $D(t) = \{r + t\beta\lambda H \geq 0\}$ be the deployment set. (i) Δ_S is absolutely*

continuous, and for almost every t

$$\Delta_S'(t) = \beta^2 \lambda \Big(\mathbb{E}_B\big[H\,\mathbf{1}_{D(t)}\big] - \mathbb{E}_A\big[H\,\mathbf{1}_{D(t)}\big] \Big).$$

(ii) At full capture, $t = n$, the planner's gain exceeds the source's by exactly $\beta^2 n\,(\mathbb{E}_B[w] - \mathbb{E}_A[w])$, which is proportional to n. At the monopoly price, $t = \rho^* n$, the excess is

$$\beta^2 n\,(\mathbb{E}_B[w] - \mathbb{E}_A[w]) + \beta^2 \lambda \int_{\rho^* n}^{n} \Big(\mathbb{E}_B[H\mathbf{1}_{D(t)}] - \mathbb{E}_A[H\mathbf{1}_{D(t)}] \Big) dt.$$

(iii) If after the uninformative pilot A the report is not deployed at t, so that $\mathbb{E}_A[H\mathbf{1}_{D(t)}] = 0$, then $\Delta_S'(t) \geq 0$. If $\mathbb{E}_B[H\mathbf{1}_{D(t)}] \leq \mathbb{E}_A[H\mathbf{1}_{D(t)}]$ on an interval of t, then Δ_S is nonincreasing on it: a larger capture, or a larger population of paying recipients, then lowers the source's incentive to run the informative pilot.

Appendix F.8 gives the proof. The mechanism is the one in Corollary 3 seen from the seller's side. A fee on the later report rewards the pilot after which the later report is more often deployed and more informative, and an informative pilot resolves some of the uncertainty that the later report would otherwise resolve. Part (i) is Proposition 7 applied to the source's objective: the fee t plays the role of the access probability λ, and the term $\beta^2 n \mathbb{E}_q[w]$ of Proposition 15 is the only part of the planner's objective that the source never values. The new content is part (ii), which prices the disagreement exactly. The set of pilot-proceeds differences d at which the source runs the less informative pilot and the planner runs the informative one is, at full capture, the interval $\big(\Delta_S(n), \Delta_S(n) + \beta^2 n(\mathbb{E}_B[w] - \mathbb{E}_A[w])\big)$. Its length is proportional to n. Whether it contains a given d also depends on where $\Delta_S(n)$ lies, and in this parameter set $\Delta_S(n)$ falls as n rises while the upper end rises, so the interval contains $d = 0.16$ for every n tested. For the monopoly price, the identification $t = \rho^* n$ uses $\lambda = 1$, as in Proposition 9, where a buyer who has paid receives the report with certainty. The proposition states the sign condition exactly instead of assuming it. I do not obtain it from concavity of H and Jensen's inequality, because H is not concave in the belief for this parameter set.

Numerical check. In the parameter set of Section 7.3, with $\lambda = 1$, the formula in (i) agrees with finite differences of the computed Δ_S to within 2×10^{-7}. The slope is $+0.0049$ at $t = 0$ and $+0.0056$ at $t = 0.25$, where A is not deployed. It is -0.0864 at $t = 0.5$ and $-0.0852, -0.0839, -0.0827, -0.0817$ at $t = 1, 2, 4, 8$. The sign changes between $t = 0.25$ and $t = 0.5$, at the capture level where the deployment set after A becomes nonempty, which is the threshold of Corollary 3. The sign of the slope is therefore not universal, and the statement that the source's incentive falls as recipients are added holds in this parameter set above that threshold, not in general.

What the extension does not do is model strategic interaction among recipients. If recipients

compete for the same installation slot, the report's value to each falls with the number of recipients who act on it, and the aggregate values are no longer linear in n. If a recipient can resell or share the report after buying it, a posted price is no longer the relevant instrument and the capture rate ρ^* may fall. Both are real limitations. They are the natural next step, and I do not claim results for them here.

Table 3. Continuation gain from the linked pilot B relative to A, by number of recipients.

n	Source, priced report	Source, full capture	Planner	Source chooses B?	Planner chooses B?
1	0.1112	0.0683	0.1929	no	yes
2	0.0683	-0.0162	0.2331	no	yes
3	0.0259	-0.0997	0.2742	no	yes
4	-0.0162	-0.1826	0.3159	no	yes
6	-0.0997	-0.3474	0.4003	no	yes
8	-0.1826	-0.5112	0.4858	no	yes

Notes. The parameter set is that of Section 7.3 with full access and $d = b_A - b_B = 0.16$. A pilot is chosen if its continuation gain exceeds d; the "source chooses B" column refers to the priced report. Values are computed numerically by the script `verify_v09.py` in the replication package and are illustrations, not estimates. The decline of the source's gain in n is observed for this parameter set only.

7.6 What the discrete and Gaussian models share

The two models are different, but the proofs of Propositions 11 and 1 use the same object. In both, the source's value has the form $-c + \beta\,\mathbb{E}[\max\{0, u + t\beta H\}]$, so the derivative of its value with respect to the fee is $\beta^2\,\mathbb{E}_\pi[H\,\mathbf{1}_{D(t)}]$, where π is the pilot and $D(t)$ the deployment set (Propositions 11(i) and 1(i)). The effect of payment on how informative a pilot the source chooses has the sign of the change in $\mathbb{E}_\pi[H\,\mathbf{1}_{D(t)}]$ when the pilot becomes more informative. A more informative pilot resolves uncertainty the report would otherwise resolve, so it lowers H, and it widens the spread of the source's own return, which lowers the probability of deployment when deployment is already likely. Both effects push the same way, and the fee therefore lowers the incentive to run an informative pilot, once deployment is likely. When deployment is unlikely, a wider spread raises the probability of deployment instead, and the sign can reverse. This is the sign pattern of Proposition 11(iii) in the discrete model (in the two-type example the slope is $+0.0049$ and $+0.0056$ before the report is deployed after the uninformative pilot and -0.0864 at $t = 0.5$; the three-type example of Appendix G.5 gives $+0.0040$ and -0.0445) and of the fee range below τ_c in the Gaussian class. In

both models the sign is the sign of one object: whether the pilot and the report are informational substitutes for the recipient, that is, whether the recipient's value of information has diminishing returns in precision. In the Gaussian class with quadratic loss the value is $-B/(p_0+x)$, which is concave, so they are always substitutes. In the discrete model the recipient makes a binary choice, and the value of information is not concave in general. Proposition 12 states exactly where it is and is not, in the case where joint installation pays at the worst assignment. certain.

Proposition 12 (Substitutes and complements in the discrete model)**.** *Fix α and write $m(\alpha) = \alpha + h - c_{11} - s(\alpha)$, so that the recipient installs the joint technology iff $p \le p^* = m(\alpha)/\Delta$. Suppose $0 < p^* < 1$, let $\ell^* = \log \frac{p^*}{1-p^*}$, and set $S_B = 2\Delta^2/\sigma_0^2$ and $S_Y = \Delta^2/\sigma_1^2$, the precisions of the linked pilot and of the report. Let $G(S) = \mathbb{E}\max\{0, p^* - p_S\}$ where p_S is the posterior from a signal of precision S starting at $1/2$, so that the recipient's value of precision S is $s(\alpha) + \Delta G(S)$. (i) $G'(S) = \dfrac{\sqrt{p^*(1-p^*)}}{4\sqrt{2\pi S}} \exp\Big(-\dfrac{\ell^{*2}}{2S} - \dfrac{S}{8}\Big)$. (ii) G is convex on $[0, S_c]$ and concave on $[S_c, \infty)$, where $S_c = 2\big(\sqrt{1+\ell^{*2}} - 1\big)$; $S_c = 0$ iff $p^* = 1/2$. (iii) If joint installation pays at the worst assignment, $r_1(0,\alpha) \ge 0$, then for every fee $t \ge 0$ the contribution of α to $\Delta_S'(t)$ in Proposition 11(i) is $\beta^2 \lambda \Delta \kappa$, with $\kappa = \big[G(S_B + S_Y) - G(S_B)\big] - \big[G(S_Y) - G(0)\big]$. (iv) $\kappa < 0$ if $p^* = 1/2$: the pilot and the report are substitutes and a larger fee lowers the source's incentive to run the informative pilot. $\kappa > 0$ if $S_B + S_Y \le S_c$: they are complements and a larger fee raises that incentive.*

Appendix F.9 gives the proof. Part (iv) is a pair of sufficient conditions, not a full characterization: for p^* away from $1/2$ and larger precisions the sign of κ depends on the parameters, and it is computed by the integral in (iii). The result is the Radner–Stiglitz nonconcavity of the value of information, here in closed form. It explains why the earlier attempt to sign Proposition 11 by concavity and Jensen's inequality fails: H is a difference of values of information, and it is not concave in the belief. It also says that the paper's headline has a boundary. When the recipient's indifference belief is far from the prior and the signals are weak, the report is worth more after the pilot than before it, and a fee on the report encourages the informative pilot instead of discouraging it. The numerical check finds substitutes failing in 12 of 48 combinations of $p^* \in \{0.5, 0.6, 0.7, 0.8\}$, $S_B \in \{1, 2, 4\}$, $S_Y \in \{0.5, 1, 2, 4\}$, all at $p^* \ge 0.7$; the closed form matches numerical differentiation to 10^{-7}; and the complements and substitutes cases of (iv) are confirmed on 66 and 25 grid points (`verify_v17.py`, `verify_v17b.py`). # Testable implications and measurement

Information versus capability. A direct test would hold the underlying operating evidence and technology fixed while varying whether investors receive its type correspondence. Eligible recipients would see either the validated type-linked report or a truthful pooled version with the same unordered outcomes, and their site types would be recorded before treatment. A false mapping or misleading relabeling is not the pooled experiment. The first outcome is the revision in expected joint returns by recipient type. Under assignment uncertainty, linked access should move beliefs differently across types even when the pooled distribution is unchanged. In the two-state benchmark,

a report favorable to type 1 lowers the inferred joint return at type 2. Beliefs, and not only the number of installations, need to be measured.

The second outcome is investment conditional on the opportunity available. With a scarce slot, the model predicts reallocation toward the type with the better conditional return. With separate opportunities, information can induce or discourage joint installation, and a decision not to adopt after unfavorable evidence is consistent with learning. A distinct capability comparison would raise the standalone inspection payoff without changing the interaction, the incremental cost, or report precision. Corollary 2 predicts no movement in the joint-adoption threshold within the profitable-standalone range. Actual AI exposure measures may combine several of these channels and cannot be mapped to α without additional evidence.

Report production versus access. The later report is selected by the source's deployment. A comparison of firms with and without observed reports therefore mixes information with the activity that produced it. A clean access comparison would randomize the release of otherwise validated reports to eligible recipients and record the whole assignment and delivery process. It would identify the effect of access to evidence conditional on its existence, not the private-social incentive to produce that evidence. To investigate Proposition 7, initial information and later access would have to be separately observed or varied, together with the later deployment choice. The comparison is between reoptimized paths after different initial reports. If the pilot posterior is not public or the source has private information, deployment itself may reveal information; the benchmark rules this channel out.

Required records. The relevant data would link a pre-existing site-type registry to dated configuration installations, comparable outcomes for 00, 10, 01, and 11, validated report versions, the eligible-recipient population, report assignment and actual access, recipient forecasts, and later investment. Common identifiers and timestamps must establish what a recipient could know before committing capital, and prices, demand, and implementation support must be measured well enough to separate informational responses from changes in physical returns.

The World Bank Firm Adoption of Technology survey is a real source of establishment-level technology-use information, covering fifteen countries and more than 21,000 establishments in the 2019–2023 survey collection (World Bank, 2026). Such surveys can motivate which configurations are economically relevant, but without linked records and identification they do not establish the information effect in this model. Common adoption of inspection and supplier-management systems is not itself evidence of a causal interaction return, and any proxy for the technologies must be validated rather than chosen for cell coverage or a favorable sign. This paper presents no survey tabulations, estimates of type-specific interaction returns, or causal investment responses.

8 Conclusion

The investment value of operating data depends on what its outcomes can be matched to. In a two-type benchmark, and more generally whenever types are exchangeable ex ante, an unordered report preserves the distribution of interaction returns but cannot identify which observable type benefits, while a type-linked report can change investment allocation. When installation produces a later report, the source's private return omits the information value for a separately owned recipient, so private deployment is never more frequent than the planner's and can be less frequent.

Greater access to such reports raises optimized social value within the stated dissemination regime, but its effect on the relative value of the initial joint pilot has no universal sign. Access substitutes for a pilot that reveals the assignment and can complement a pilot when later reports would otherwise go unproduced. Full payment for the later report would close the deployment wedge without necessarily closing the pilot-choice wedge. When the source prices the report to heterogeneous recipients, it captures only part of the gain, excludes some recipients, and releases reports that are too weakly linked to types. Common ownership internalizes both margins. In a tractable Gaussian benchmark, covariance determines how informative an activity is for other investment opportunities.

Taken together, the results say that making data tradable fixes the margin on which data are produced, but not the margin on which they are designed. In the numerical illustration the source's reluctance to run the informative pilot grows with the number of paying recipients (Propositions 10 and 11), and when the source chooses how precisely to record, it records less than the planner even at full capture (Proposition 17). In the Gaussian class of Section 3 this is proved from primitives: payment lowers the precision the source records, a large enough payment ends recording altogether, and a fee on deployment alone cannot restore the planner's choice (Propositions 1 and 2), and if the report can leak, the gap widens with the number of recipients (Proposition 4). Hidden precision is a matter of degree: recording resumes only when recipients can verify it with probability above $q/(p_0+q)$ in large markets (Proposition 3). In the discrete model the sign is that of the curvature of the recipient's value of information, and it can reverse (Proposition 12).

Limitations and next steps. The results describe a finite-horizon information-allocation problem. They do not establish aggregate growth, multiple equilibria, or a Big Push. Five extensions remain. First, the main numerical example is two-type, and the three-type example of Appendix G.5 is one parameter set with Monte Carlo error. A systematic many-type exploration would show how large the wedges can be across parameters. Second, pricing here is a single seller facing recipients who do not interact, and the scale result of Section 9.1 is one parameter set, proved only for the precision margin of the Gaussian class of Section 3, which is an analogue of the discrete model and not a special case; competition for a scarce slot is not modeled, and resale of the report enters only as a cap on capture (Proposition 4), not as a game. A game among several sources and recipients,

with competition between recipients or mutually dependent participation, is the natural place to look for Big Push effects. Third, the literature comparison is incomplete: I have not compared the mechanism with Teng's full model, which I could not obtain, nor read the original text of Admati and Pfleiderer; the full texts I did read are listed in Section 2, and no one in the field has read this manuscript. Fourth, precision is hidden by assumption or through the verification probability φ of Proposition 3, and is not derived from a contracting problem with moral hazard; and a resale game, in which commitment determines whether the source captures the whole gain or one recipient's gain, is not solved. Fifth, the measurement strategy in Section 10 has not been tried on data.

References

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4), 2022.

Anat R. Admati and Paul Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

Ufuk Akcigit and Qingmin Liu. The role of information in innovation and competition. *Journal of the European Economic Association*, 14(4), 2016.

Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11:85–107, 2019.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American Economic Review*, 108(1):1–48, 2018.

Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. *RAND Journal of Economics*, 53(2):263–296, 2022.

Kostas Bimpikis, Davide Crapis, and Alireza Tahbaz-Salehi. Information sale and competition. *Management Science*, 65(6):2646–2664, 2019.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

Yiling Chen and Bo Waggoner. Informational substitutes. Full author manuscript of the FOCS 2016 paper, arXiv:1703.08636, 2017.

Luca V. A. Colombo, Gianluca Femminis, and Alessandro Pavan. Investment subsidies with spillovers and endogenous private information: Why Pigou got it all right. *Journal of Economic Theory*, 233:106160, 2026.

Lin William Cong, Danxia Xie, and Longtian Zhang. Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10):6480–6492, 2021. doi: 10.1287/mnsc.2021.3986. Model comparison uses the September 21, 2021 author deposit.

Timothy G. Conley and Christopher R. Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, 2010.

Jeff Dominitz and Charles F. Manski. More data or better data? a statistical decision problem. *Review of Economic Studies*, 84(4):1583–1605, 2017.

Kimon Drakopoulos and Ali Makhdoumi. Providing data samples for free. *Management Science*, 69(6):3536–3560, 2023.

Maryam Farboodi and Laura Veldkamp. A model of the data economy. *The Review of Eco-*

nomic Studies, 2026. Advance article rdag074. doi:10.1093/restud/rdag074. Substantive model comparison uses the July 9, 2024 author draft, not the unread published full model.

Maryam Farboodi, Andrew Koh, and Anchi Xia. Data-driven automation. Author draft dated June 10; NBER Working Paper 35320. `https://arxiv.org/pdf/2606.10127`, 2026.

Andrew D. Foster and Mark R. Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209, 1995.

Simone Galperti, Aleksandr Levkun, and Jacopo Perego. The value of data records. *Review of Economic Studies*, 91(2):1007–1038, 2024.

Shota Ichihashi. The economics of data externalities. *Journal of Economic Theory*, 196:105316, 2021.

Charles I. Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020.

Jakub Kastl, Marco Pagnozzi, and Salvatore Piccolo. Selling information to competitive firms. *RAND Journal of Economics*, 49(1):254–282, 2018.

Ernest Liu, Song Ma, and Laura Veldkamp. Data sales and data dilution. *Journal of Financial Economics*, 169:104053, 2025.

Kevin M. Murphy, Andrei Shleifer, and Robert W. Vishny. Industrialization and the big push. *Journal of Political Economy*, 97(5):1003–1026, 1989. doi: 10.1086/261641. Detailed framework comparison uses the September 1988 NBER Working Paper 2708.

Saltuk Ozerturk. Direct sale of information when precision is unobservable. *Canadian Journal of Economics*, 37(2):269–293, 2004.

Songyuan Teng. Learning complementarity via experiment: A Bayesian theory of combinatorial innovation. Work-in-progress abstract; the full model is not used for this paper's comparison. `https://songyuan-teng.github.io/files/Abstract_Learning.pdf`, 2026.

Laura Veldkamp and Cindy Chung. Data and the aggregate economy. *Journal of Economic Literature*, 62(2):458–484, 2024. doi: 10.1257/jel.20221580. Author-hosted published-format text consulted.

World Bank. Firm adoption of technology survey 2019–2023, 2026. Microdata catalog WLD_2019-2023_FAT_v01_M. doi:10.48529/assd-3j65. `https://microdata.worldbank.org/index.php/catalog/8209`.

Appendix A. Identification and the static allocation benchmark

A.1 Proof of Proposition 5

Let $f(u,v)$ be the state-independent joint density of the contrast errors, with $f(u,v) = f(v,u)$. For a distinct observed unordered pair $\{z_1, z_2\}$, its density under the state (h,ℓ) is proportional to

$$f(z_1 - h, z_2 - \ell) + f(z_2 - h, z_1 - \ell).$$

Under (ℓ,h) it is proportional to

$$f(z_1 - \ell, z_2 - h) + f(z_2 - \ell, z_1 - h).$$

Exchangeability makes the two sums equal. Ties do not change the conclusion. The corresponding argument uses probabilities rather than densities for a discrete error distribution. Bayes' rule therefore leaves the assignment odds unchanged for the pooled report.

For independent normal errors, the linked log-likelihood ratio is

$$-\frac{(z_1-h)^2 + (z_2-\ell)^2 - (z_1-\ell)^2 - (z_2-h)^2}{2\sigma^2} = \frac{\Delta(z_1 - z_2)}{\sigma^2}.$$

Adding prior log odds gives the stated posterior. The proof requires the errors' joint exchangeability, not just matching marginal distributions.

A.2 Proof of Corollary 1 and numerical integration

For fixed α, the expected payoff of each feasible installation is affine in p. Optimizing over those installations makes V convex. The linked posterior has conditional mean p_0. Jensen's inequality gives $\mathbb{E}[V(p_L,\alpha)] \geq V(p_0,\alpha)$. By Proposition 5, the pooled report gives the latter value. Strictness follows when posterior variation crosses distinct optimal-action regions in a way that rules out a common maximizing affine payoff.

In the static numerical case, write $D = Z_1 - Z_2$ and $p(D) = (1 + \exp\{-\Delta D/\sigma^2\})^{-1}$. The recipient value is

$$V(p(D),\alpha) = \max\{0, \alpha - 0.1, \alpha - 0.7 + 1.2\max[p(D), 1 - p(D)]\}.$$

Conditional on type 1 being high, $D \sim N(1.2, 2\sigma^2)$. The value is symmetric in D, so integration over that conditional distribution and averaging over $\alpha = \pm 0.3$ suffices. This gives 0.309092373 at

$\sigma = 1$ and 0.236397278 at $\sigma = 1.5$. The no-report value is 0.1 and the exact linked value is 0.5.

Appendix B. Report value and deployment

B.1 Nonnegative and strictly positive information value

The posterior p_Y is a martingale conditional on (p, α). The function $w(p, \alpha)$ is the maximum of two affine functions of p, so Jensen's inequality implies $H \geq 0$. Let $j_h = \alpha + h - c_{11}$ and $j_\ell = \alpha + \ell - c_{11}$. If $j_h > s > j_\ell$, the recipient's action threshold is strictly inside the posterior interval. For $0 < p < 1$ and finite positive normal noise, the likelihood ratio has support $(0, \infty)$; hence p_Y has positive probability on both sides of the threshold. No single affine branch maximizes everywhere, giving strict information value.

At $p = 0$ or $p = 1$, posterior variation is absent. If the same arrangement is optimal in both states, optimizing does not gain from the signal and $H = 0$. These cases must be handled separately rather than substituted into an interior logarithmic formula.

B.2 Closed-form expected recipient value

Suppose $j_h > s > j_\ell$. The recipient installs jointly after a sufficiently low source report. Its report threshold is

$$y_c = \frac{h + \ell}{2} + \frac{\sigma_1^2}{\Delta} \log \left(\frac{(1-p)(j_h - s)}{p(s - j_\ell)} \right).$$

Writing Φ for the standard normal distribution function, expected optimized recipient value after receipt is

$$F(p, \alpha) = s + (1-p)(j_h - s)\Phi \left(\frac{y_c - \ell}{\sigma_1} \right) - p(s - j_\ell)\Phi \left(\frac{y_c - h}{\sigma_1} \right).$$

Then $H = F - w$. The first incremental term applies in the state where the source is low and type 2 is high. The negative term reflects choosing joint installation after a low report when type 2 is actually low. Integrating state-contingent payoffs below y_c gives the expression.

B.3 Proof of Proposition 6 and Corollary 2

Let $x \in \{0, 1\}$ indicate source joint deployment rather than its established option. Date-1 total discounted continuation is

$$s + xr_1 + \beta[w + x\lambda H].$$

Only $r_1 + \beta\lambda H$ depends on the deployment choice. This gives the planner threshold, while the source's no-transfer threshold is $r_1 \geq 0$. At an interior private cutoff p_S, $r_1(p_S, \alpha) = 0$. If $\beta\lambda H(p_S, \alpha) > 0$, continuity makes the planner increment positive immediately below p_S. Since r_1 is strictly increasing in p, the source increment there is negative. This establishes an open disagreement interval.

For Corollary 2, $\alpha \geq c_{10}$ implies $s = \alpha - c_{10}$. Substitution gives the two expressions in Section 6.3. The common additive term cancels from $\mathbb{E}[w(p_Y, \alpha)] - w(p, \alpha)$, making H, r_1, and the two deployment rules independent of α within the stated range.

Appendix C. Date-0 value, access, and numerical methods

C.1 Derivation of the continuation objectives

The source receives date-1 payoff $s + P(r_1)$. Discounting it once and adding date-0 net pilot proceeds gives U_q. For the planner, optimizing x in the date-1 expression in Appendix B gives

$$s + P(r_1 + \beta\lambda H) + \beta w.$$

Discounting this expression to date 0 gives the objective W_q. The operating information benefit is discounted twice in total date-0 welfare. The baseline recipient value w remains after either pilot even at zero operating-report access.

C.2 Proof of Proposition 7

For a fixed (p, α), define $L(\lambda) = P(r_1 + \beta\lambda H)$. Since $H \geq 0$, L is nondecreasing and convex. Expectation preserves both properties. The remaining terms in W_q do not depend on λ, establishing the first claim.

Away from deployment ties, $L'(\lambda) = \beta H \mathbf{1}\{r_1 + \beta\lambda H > 0\}$. The finite-state return levels bound H, so differentiation under expectation is valid whenever the tie event has zero probability. The left derivative uses the strict selection set, and the right derivative includes the tie set with $H > 0$. Convexity ensures almost-everywhere differentiability. This is relevant for A, whose posterior is an atom at $1/2$.

Subtracting the two derivatives gives the stated expression for $d\Delta_P/d\lambda$. No general inequality

orders the two selected information-value expectations. In particular, the initial signal changes both residual uncertainty and the set of deployments that generate a later signal.

C.3 Integration and numerical checks

Under B, D_0 has the equal mixture of $N(\Delta, 2\sigma_0^2)$ and $N(-\Delta, 2\sigma_0^2)$. I apply the logistic posterior transformation and average the optimized continuation over that mixture and the two α values. The closed-form F from Appendix B reduces the calculation to one-dimensional deterministic quadrature. Integration is split at decision thresholds and uses absolute and relative tolerances of 2×10^{-12}, with tails truncated at least twelve normal standard deviations.

The analytic expression for F was separately checked against direct integration of the posterior-optimal recipient payoff. At $p = 0.45$ and $\alpha = 0.3$, $F = 0.434047145$, $w = 0.21$, and $H = 0.224047145$. The two integrations agree to numerical precision. The reported pilot values and the Gaussian example also reproduce their six-decimal rounding. These are numerical checks of specified theoretical examples, not formal interval bounds or simulated empirical observations.

Appendix D. Contracting boundary

For the static example, the source's increase from switching A to exact linked B under payment is $T - d$, and the recipient's increase is $0.4 - T$. Both are strictly positive for $d < T < 0.4$, which is feasible when $0 < d < 0.4$. The common experiment cost cancels. The contract must make the switch and delivered information enforceable.

For the deployment-contingent schedule, the source's date-1 incremental payoff is $r_1 + F_\rho = r_1 + \rho\beta\lambda H$. The recipient's expected report gain after the fee is $(1-\rho)\beta\lambda H \geq 0$. Transfers cancel from total payoff. Increasing ρ weakly expands the set of beliefs at which the source deploys, and full capture reproduces the planner's deployment coefficient.

The source's date-0 payoff under this fee is

$$U_q^\rho = b_q - C + \beta\mathbb{E}_q[s + P(r_1 + \rho\beta\lambda H)].$$

At $\rho = 1$, this still differs from W_q by $\beta^2\mathbb{E}_q[w]$. If that term varies with the pilot, aligning the deployment rule does not align the experiment objective. Under common ownership and control of both sites, both real continuation payoffs enter the owner's objective; it is exactly W_q. This proves Proposition 8 without imposing an arbitrary private appropriation rule on welfare.

Appendix E. Gaussian cross-site information value

Statement and discussion. The two-type model makes information relevance especially transparent: the two returns are perfectly negatively related. To distinguish the direction of a response from the value of information more generally, consider a separate Gaussian benchmark. It is a continuous-adjustment specialization, not a many-site theorem for the binary sequential model.

Conditional on the common information available before deployment, let $\theta \sim N(m,\Sigma)$. An ex ante source at site i can generate $Y_i = \theta_i + \varepsilon_i$, with independent error variance τ_i^2. A receiving site $j \neq i$ chooses a signed local adjustment $x_j \in \mathbb{R}$ around an existing operating baseline, with payoff

$$\mathbb{E}[\theta_j \mid \mathscr{I}]x_j - \frac{b_j}{2}x_j^2, \qquad b_j > 0.$$

The curvature b_j is distinct from the earlier pilot proceeds b_q. The adjustment is not a literal negative installation. Receiver decisions are separable and unconstrained by a common capacity budget. Gaussian conditioning yields

$$\mathbb{E}[\theta_j \mid Y_i] = m_j + \frac{\Sigma_{ji}}{\Sigma_{ii}+\tau_i^2}(Y_i - m_i).$$

Gaussian benchmark (Cross-site report value)**.** *If $\Sigma_{ii}+\tau_i^2 > 0$, the expected increase in recipients' optimized returns is*

$$H_i = \sum_{j\neq i} \frac{\Sigma_{ji}^2}{2b_j(\Sigma_{ii}+\tau_i^2)}.$$

It is strictly positive exactly when at least one recipient has nonzero covariance with the source return. A negative covariance contributes positively to information value while reversing the sign of that recipient's response to the report. Appendix E derives the expression from posterior-mean variation and quadratic optimization.

Under a no-transfer arrangement and independent report access λ_i, a source with real incremental return r_i deploys when $r_i \geq 0$, while the planner includes $\beta\lambda_i H_i$. Ranking alternative report sources by $r_i + \beta\lambda_i H_i$ additionally requires a common no-report welfare benchmark, or that candidate-specific opportunity costs have been included in r_i. The formula for H_i alone does not eliminate those differences.

For an illustrative common-baseline example, set $m = 0$, $\Sigma = vv^{\mathsf{T}} + 0.2I$, $v = (1.2, 0.8, 0.1)$, $b_j = 1$, $\tau_i^2 = 0.25$, $\beta = 0.8$, and $\lambda_i = 1$. The report values are $(0.247619, 0.425688, 0.022609)$. With $r = (0.10, 0.05, 0.02)$, the private source choice is site 1. The planner values are $(0.298095, 0.390550, 0.038087)$, selecting site 2.

This is a standard Gaussian-quadratic information calculation. Its use here is economic: an activity's own return, own uncertainty, and usefulness to other investors are different objects. Report relevance depends on covariance with those investors' uncertain incremental returns, precision, and adjustment costs. A covariance network based on aggregate revenues is not automatically an estimate of Σ, and proximity in that network does not demonstrate that a recipient received or used a report.

Proof and numerical example. Conditional on Y_i, the optimal receiver adjustment is $x_j = \mu_j / b_j$, where $\mu_j = \mathbb{E}[\theta_j \mid Y_i]$. Optimized expected payoff is $\mu_j^2/(2b_j)$. The posterior mean has ex ante mean m_j and variance

$$\mathrm{Var}(\mu_j) = \frac{\Sigma_{ji}^2}{\Sigma_{ii} + \tau_i^2}.$$

Subtracting the no-report optimized payoff $m_j^2/(2b_j)$ gives the recipient information value. Summing over $j \neq i$ proves the formula above. The sign of covariance affects the direction of the conditional adjustment; its square determines ex ante value. Independent access and discounting multiply the value by λ_i and β.

For the numerical example, covariance diagonals are $(1.64, 0.84, 0.21)$ and off-diagonals are $(\Sigma_{12}, \Sigma_{13}, \Sigma_{23}) = (0.96, 0.12, 0.08)$. The matrix is positive definite because it equals $vv^{\mathsf{T}} + 0.2I$. Using noise variance 0.25, not standard deviation 0.25, gives the reported values. Setting $m = 0$ makes the no-report optimized adjustment payoff zero for every candidate recipient set, so the illustrative source ranking does not omit a candidate-specific baseline payoff. # Further proofs

E.1 Proof of Proposition 1

Let $G(a,\sigma) = \mathbb{E}\max\{0, a + \sigma Z\} = a\Phi(a/\sigma) + \sigma\varphi(a/\sigma)$ for $Z \sim N(0,1)$, with $G(a,0) = \max\{0,a\}$. Then $G_a = \Phi(a/\sigma)$ and $G_\sigma = \varphi(a/\sigma)$ for $\sigma > 0$. Since $u = s_0 + \sigma(x)Z$ and $V_S = -c + \beta G(a(x,t), \sigma(x))$, (i) follows from $\partial a/\partial t = \beta H$. At $x = 0$, $s_0 \geq 0$ gives $V_S(0;t) = \beta(s_0 + t\beta H(0))$ and the same derivative, with $P_0 = 1$.

(ii). Take $t' > t$ and integrate (i): $V_S(x;t') - V_S(x;t) = \beta^2 \int_t^{t'} H(x)\Phi(a(x,\tau)/\sigma(x))\, d\tau$. For $\tau > 0$ and $s_0 \geq 0$, $a(x,\tau) > 0$ is strictly decreasing in x because H is, and $\sigma(x) = s_1\sqrt{x/(p_0(p_0+x))}$ is increasing. So a/σ is decreasing and $\Phi(a/\sigma)$ is nonincreasing in x. H is positive and decreasing, and the product of two nonnegative nonincreasing functions is nonincreasing, so the integrand is nonincreasing in x and so is the integral. We next show that all maximizers lie in a compact interval $[0, X_0]$ independent of t. For $x > 0$,

$$\partial_x V_S = -c'(x) + \beta\left[t\beta H'(x)\Phi(a/\sigma) + \varphi(a/\sigma)\sigma'(x)\right],$$

and $H' < 0$, so the first bracketed term is nonpositive. Now $\sigma'(x) = s_1\sqrt{p_0}/\left(2\sqrt{x}(p_0+x)^{3/2}\right)$ is strictly decreasing and tends to zero, so there is X_0 with $\beta\varphi(0)\sigma'(x) < k$ for $x \geq X_0$. For such x, $\partial_x V_S < -c' + k \leq 0$ for every $t \geq 0$, so no maximizer exceeds X_0. On the lattice $[0, X_0]$, V_S is continuous and has decreasing differences in (x,t), so by Topkis's monotone comparative statics theorem the largest and smallest maximizers are nonincreasing in t. Setting $t = n$ gives the last sentence.

(iii). On $(0, X_0]$, $a \geq t\beta H(X_0)$ since $s_0 \geq 0$ and H is decreasing. Also $\sigma' = s_1^2 v'/(2\sigma)$ with $v' = (p_0+x)^{-2} \leq p_0^{-2}$, so $\varphi(a/\sigma)\sigma' = \frac{s_1^2 v'}{2a} \cdot \frac{a}{\sigma}\varphi(\frac{a}{\sigma}) \leq \frac{s_1^2\varphi(1)}{2p_0^2 t\beta H(X_0)}$, using $u\varphi(u) \leq \varphi(1)$. Therefore $\partial_x V_S \leq -k + \dfrac{s_1^2\varphi(1)}{2p_0^2 tH(X_0)}$, which is negative for $t > \bar{t} := s_1^2\varphi(1)/(2p_0^2 kH(X_0))$. With the bound for $x \geq X_0$, V_S is continuous and strictly decreasing on $[0,\infty)$, so $x = 0$ is the unique maximizer, and $P_0 = 1$.

(iv). $\partial_x V_P = \partial_x V_S(x;n) + n\beta^2 w'(x)$ with $w'(x) = B/(p_0+x)^2$. In $\partial_x V_S(x;n)$, $H' \leq 0$ and $\Phi \leq 1$ give $n\beta^2 H'\Phi \geq n\beta^2 H'$, and $\varphi\sigma' \geq 0$. Thus $\partial_x V_P \geq -c' + n\beta^2(w' + H') = -c'(x) + n\beta^2 w'(x+q)$, because $w' + H' = w'(x+q)$. The function $x \mapsto n\beta^2 w'(x+q) - c'(x)$ is strictly decreasing since w' is, and c' is nondecreasing by convexity, so it is positive on $[0,\hat{x})$ and V_P is strictly increasing there: no maximizer lies below $\hat{x}$. For $c = kx$, $\hat{x}$ solves $n\beta^2 B/(p_0+\hat{x}+q)^2 = k$. The last claim follows from (iii).

(v). Let $\bar{x}$ be the largest maximizer of $V_S(\cdot\,;n)$. Because w is strictly increasing, for $x < \bar{x}$ we have $V_S(x;n) \leq V_S(\bar{x};n)$ and $w(x) < w(\bar{x})$, so $V_P(x) < V_P(\bar{x})$ and no planner maximizer lies below $\bar{x}$. If $\bar{x} > 0$, it is interior (by (ii) it is at most X_0 and the derivative is negative above), so $\partial_x V_S(\bar{x};n) = 0$ and $\partial_x V_P(\bar{x}) = n\beta^2 w'(\bar{x}) > 0$, so $\bar{x}$ is not a planner maximizer. If $\bar{x} = 0$, the smallest planner maximizer is weakly above zero, with equality only if it is zero. □ The numerical values are computed by `verify_v16.py`.

E.2 Proof of Proposition 2

Suppose a fee $t \in [0,n]$ induces the planner's precision $x_P > 0$ and the planner's deployment rule, which is to deploy iff $u + n\beta H(x_P) \geq 0$. The source deploys iff $u + t\beta H(x_P) \geq 0$. Since $x_P > 0$, u has full support on the real line, and $H > 0$, the two rules agree with probability one only if $t = n$. At $t = n$, x_P must maximize $V_S(\cdot\,;n)$. It is interior, by the argument for (v) above, so $\partial_x V_S(x_P;n) = 0$, and then $\partial_x V_P(x_P) = n\beta^2 w'(x_P) > 0$, contradicting the optimality of x_P for the planner. Hence the planner's allocation is not implementable, and for $t = n$ the source's largest maximizer is strictly below the planner's smallest by Proposition 1(v) since $x_P > 0$. For $t < n$ and a source maximizer $x > 0$, the set $\{u + n\beta H(x) \geq 0 > u + t\beta H(x)\}$ has positive probability, so deployment is too rare on that set. If the source's maximizer is 0, it differs from $x_P > 0$. The sufficient condition for $x_P > 0$

follows from (iv), since $\hat{x} > 0$ iff $n\beta^2 B/(p_0+q)^2 > c'(0)$. □

E.3 Proof of Proposition 3

Write $V^\varphi(x) = V_P(x;n) - (1-\varphi)\beta^2 n\, w(x) - \varphi\beta^2 n\, w(0)$, which holds because $V_P = V_S(\cdot\,;n) + \beta^2 nw$. *(i).* $\partial V^\varphi(x)/\partial\varphi = \beta^2 n\,[w(x) - w(0)]$ is increasing in x, so V^φ has increasing differences in (x,φ). All maximizers lie in a compact interval independent of φ: for large x, $\partial_x V_S(x;n) < 0$ by the argument in the proof of Proposition 1(ii), and $\varphi\beta^2 nw'(x) = \varphi\beta^2 nB/(p_0+x)^2 \to 0$ while $c' \geq k$. Topkis's theorem gives (i). *(ii).* Let a maximize V^φ. Suppose $a > x_P$. Since x_P is the largest planner maximizer, $V_P(a) < V_P(x_P)$, and $w(a) > w(x_P)$ gives $-(1-\varphi)\beta^2 n\, w(a) < -(1-\varphi)\beta^2 n\, w(x_P)$ when $\varphi < 1$. Hence $V^\varphi(a) < V^\varphi(x_P)$, a contradiction. Suppose $a = x_P > 0$. Then x_P is interior for V_P, so $V_P'(x_P) = 0$, and $(V^\varphi)'(x_P) = -(1-\varphi)\beta^2 nw'(x_P) < 0$, so x_P is not a maximizer of V^φ. *(iii).* Let $a(x) = s_0 + n\beta H(x) > 0$ and $G(a,\sigma) = a + R$ with $R = \mathbb{E}[(-a-\sigma Z)^+]$, so that $0 \leq R \leq \sigma\varphi(a/\sigma)$ for $a \geq 0$. Since $s_0 \geq 0$, $V_S(0;n) = \beta a(0)$ and

$$V^\varphi(x) - V^\varphi(0) = -kx + n\beta^2 h(x) + \beta R(x), \qquad h(x) = H(x) - H(0) + \varphi\,[w(x) - w(0)].$$

Here $h(0) = 0$, $h'(x) = w'(x+q) - (1-\varphi)w'(x) = B\big[(p_0+x+q)^{-2} - (1-\varphi)(p_0+x)^{-2}\big]$, and $h(\infty) = \frac{B}{p_0}(\varphi - \varphi^*)$ because $H(\infty) = 0$, $H(0) = Bq/(p_0(p_0+q))$ and $w(\infty) - w(0) = B/p_0$. *Case* $\varphi < \varphi^*$. Then $1-\varphi > p_0/(p_0+q) > (p_0/(p_0+q))^2$, so $h'(0) < 0$. The sign of h' changes once, from negative to positive, because $(p_0+x)/(p_0+x+q)$ increases to one. Together with $h(\infty) < 0$ this gives $h < 0$ on $(0,\infty)$, $h(x) \leq -\gamma x$ on some $[0,\varepsilon_0]$ with $\gamma > 0$, and $\eta := -\sup_{x\geq\varepsilon_0} h > 0$. For $x \geq \varepsilon_0$, $R \leq \sigma\varphi(0) \leq C := s_1/\sqrt{2\pi p_0}$, so $V^\varphi(x) - V^\varphi(0) \leq \beta C - n\beta^2\eta < 0$ for large n. For $0 < x < \varepsilon_0$, $a \geq A_n := n\beta H(\varepsilon_0)$, and $\varphi(z) \leq 0.294\, z^{-2}$ gives $R \leq 0.294\,\sigma^3/A_n^2 \leq Dx/A_n^2$ with $D = 0.294\, s_1^3 \varepsilon_0^{1/2} p_0^{-3}$ (using $\sigma^2 \leq s_1^2 x/p_0^2$), so $V^\varphi(x) - V^\varphi(0) \leq x[-k - n\beta^2\gamma + \beta D/A_n^2] < 0$ for large n. Hence $x = 0$ is the unique maximizer. *Case* $\varphi > \varphi^*$. Take $x_n = \sqrt{n}$. Since $R \geq 0$, $V^\varphi(x_n) - V^\varphi(0) \geq -k\sqrt{n} + n\beta^2 h(\sqrt{n})$ and $h(\sqrt{n}) \to h(\infty) > 0$, so the right-hand side is positive for large n and 0 is not a maximizer. □ The numbers are computed by `verify_v22.py` and `make_figs_v22.py`. ## Proof of Proposition 4 With capture $t \in [0,1]$ and $s_0 \geq 0$, Proposition 1(ii) gives that the source's maximizers are nonincreasing in t, and all lie in $[0, X_0]$. So every precision the source can choose at any $t \in [0,1]$ is at most the largest maximizer at $t = 0$, $x_S(0)$, which depends on $(p_0, B, q, \beta, s_0, s_1, c)$ but not on n. By Proposition 1(iv) every planner maximizer is at least $\hat{x}(n)$, and with $c(x) = kx$, $\hat{x}(n) = \sqrt{n\beta^2 B/k} - p_0 - q \to \infty$. The difference is at least $\hat{x}(n) - x_S(0)$. □ The numbers are computed by `verify_v16c.py`.

E.4 Proof of Corollary 3

Proposition 7 gives $d\Delta_P/d\lambda = \beta^2[K_B(\lambda) - K_A(\lambda)]$ wherever both derivatives exist, with $K_q \geq 0$ because $H \geq 0$.

Part (a). If $p_B \in \{0,1\}$ almost surely, the belief is degenerate. At a degenerate belief the posterior after any report equals the prior, so $\mathbb{E}[w(p_Y)] = w(p)$ and $H = 0$. Hence $K_B = 0$ and the derivative equals $-\beta^2 K_A \leq 0$, which is strictly negative if $K_A > 0$.

Part (b). If $K_A = 0$ and $K_B > 0$, the derivative equals $\beta^2 K_B > 0$. □

Numerical values in Section 7.3. In the dynamic example the script `verify_v07.py` computes K_A and K_B with the verified v0.4 routines. At $\lambda = 0.05, 0.15, 0.25$, $K_A = 0$ and $K_B = 0.00558, 0.00590, 0.00626$. At $\lambda = 0.30, 0.50, 0.80, 1.00$, $K_A = 0.10315$ and $K_B = 0.00646, 0.00737, 0.00824, 0.00869$. The formula $\beta^2(K_B - K_A)$ matches a finite-difference derivative of Δ_P to within 10^{-7} at every tested λ. With $\sigma_0 = 0.03$, so that B nearly reveals the assignment, K_B is below 10^{-100} and the slope equals $-\beta^2 K_A = -0.0931$ at $\lambda = 0.5$ and $\lambda = 1$, as part (a) requires.

E.5 Proof of Proposition 9 and Corollary 4

Proposition 9(i). If $H(\pi) = 0$ revenue is zero. Otherwise recipient j buys at price f if and only if $\xi_j \geq f/(\beta H)$. Writing $x = f/(\beta H)$, revenue is $fD(x) = \beta H\, xD(x)$. The function $xD(x)$ is upper semicontinuous because D is nonincreasing and left-continuous, and $xD(x) \leq \mathbb{E}[\xi \mathbf{1}\{\xi \geq x\}] \to 0$ as $x \to \infty$. The maximum ρ^* is therefore attained at some x^*. After deployment the report exists and the source's only remaining choice is the price. This static problem has the same solution whether or not a price was announced earlier, so the solution is subgame perfect without commitment.

(ii). Since $\mathbb{E}\xi = 1$, $\Pr(\xi > 0) > 0$, so $xD(x) > 0$ for small x and $\rho^* > 0$. For any x, $xD(x) = \mathbb{E}[x\mathbf{1}\{\xi \geq x\}] \leq \mathbb{E}[\xi \mathbf{1}\{\xi \geq x\}] \leq \mathbb{E}\xi = 1$. Equality in both steps requires $\xi = x$ almost surely on $\{\xi \geq x\}$ and $\xi = 0$ almost surely on $\{\xi < x\}$, that is, $\xi \in \{0, x\}$ almost surely.

(iii). After the price is paid the source's total payoff from deploying is $r + \rho^* \beta H$. The remaining claims follow from $0 < \rho^* \leq 1$, $H \geq 0$, and the planner's rule $r + \beta H \geq 0$. If $-\beta H \leq r < -\rho^* \beta H$, then $r + \beta H \geq 0 > r + \rho^* \beta H$.

(iv). Recipients with $\xi < x^*$ do not buy. Their information value is $\beta H \xi$, and supplying the report to them is free, so the forgone value is $\beta H\, \mathbb{E}[\xi \mathbf{1}\{\xi < x^*\}]$.

Example. For $\xi \sim U[0,2]$, $xD(x) = x - x^2/2$ is maximized at $x^* = 1$ with value 0.5. The excluded mass is $G(1) = 0.5$ and the forgone value is $\int_0^1 (\xi/2)\, d\xi = 0.25$ of the benefit βH. Buyers' surplus is $\int_1^2 (\xi - 1)\, d\xi/2 = 0.25$. The three shares sum to one. The numbers for the dynamic example come from `verify_v08.py`: thresholds 0.541667, 0.452382, 0.392628, and source gains

0.131176, 0.111231, 0.068333.

Corollary 4. Suppose version P is a garbling of L, so $\pi_{Y^P} = \mathbb{E}[\pi_{Y^L} \mid Y^P]$. Because each w_j is convex, Jensen's inequality gives $\mathbb{E}[w_j(\pi_{Y^P})] \le \mathbb{E}[w_j(\pi_{Y^L})]$, so $H^P \le H^L$. Under the pricing of Proposition 9 with the same G for both versions, revenue from version v is $\rho^*\beta H^v$. The source's net gain from linking is $\rho^*\beta(H^L - H^P) - k$ and the planner's is $\beta(H^L - H^P) - k$. Under no payment revenue is zero and the source's gain is $-k < 0$. The stated interval follows. □

E.6 Proof of Proposition 10

(i) At a posted price f, a recipient with scale ξ buys if $\beta\xi H \ge f$. Write $x = f/(\beta H)$. Revenue from a mass n of recipients is $nf\Pr(\xi \ge x) = n\beta H\, xD(x)$. For $H > 0$ the maximizer of $xD(x)$ does not depend on n or H, so $f^* = x^*\beta H$ and revenue is $\rho^* n\beta H$. (ii) The source's payoff from deploying is $r + \rho^* n\beta H$ and its payoff from not deploying is zero, which gives the rule; the planner's rule is the same with $\rho = 1$. The difference in the left-hand sides is $(1-\rho^*)n\beta H$. (iii) With mass n the aggregate information value is nH and the aggregate recipient value is nw, because each recipient's values scale by ξ_j and the recipients do not interact. The source's objective is then the objective of Section 8.2 with fee $\rho n\beta\lambda H$, and Proposition 15(i) applied to the aggregate recipient gives the planner's objective as the source's at full capture plus $\beta^2 n\mathbb{E}_q[w]$. Differencing over the two pilots gives the stated expressions. □ The numerical values in the example are computed by `verify_v09.py`.

E.7 Proof of Proposition 11

Fix a pilot q. By the decomposition of Proposition 15, the source's continuation value with total capture t is $b_q - C + \beta\mathbb{E}_q[s + P(r + t\beta\lambda H)]$ with $P(x) = \max\{x, 0\}$. The map $t \mapsto P(r + t\beta\lambda H)$ is convex and piecewise linear for each realization, with right derivative $\beta\lambda H\,\mathbf{1}\{r + t\beta\lambda H \ge 0\}$ and $0 \le H$ bounded. By dominated convergence the expectation is convex with right derivative $\beta\lambda\mathbb{E}_q[H\mathbf{1}_{D(t)}]$, so it is absolutely continuous and differentiable almost everywhere. Multiplying by β and differencing over $q = B, A$ gives (i). For (ii), Proposition 15 gives the planner's gain as $\Delta_S(n) + \beta^2 n(\mathbb{E}_B[w] - \mathbb{E}_A[w])$, and $\Delta_S(n) - \Delta_S(\rho^* n)$ is the integral of (i). (iii) follows from (i), since $H\mathbf{1}_{D(t)} \ge 0$. □ The check in the text is computed by `verify_v10.py`.

E.8 Proof of Proposition 12

Because both pilots reveal α and the assignment is independent of α, the recipient's value after a pilot and a report depends on the assignment only through the total precision of what it observes. Under the linked pilot the log-odds are $\Delta D_0/\sigma_0^2$ with $D_0 \sim N(\pm\Delta, 2\sigma_0^2)$, so given the state they

have mean $\pm S_B/2$ and variance S_B, with $S_B = 2\Delta^2/\sigma_0^2$. The report adds log-odds with mean $\pm S_Y/2$ and variance S_Y, $S_Y = \Delta^2/\sigma_1^2$. Independent normal log-odds of this form add, so the total has the same form with precision $S_B + S_Y$. The recipient installs iff $m - \Delta p \geq 0$, so its value is $s(\alpha) + \Delta\,\mathbb{E}\max\{0, p^* - p_S\}$.

(i). Write ℓ for the log-odds and $p = 1/(1+e^{-\ell})$, and let $a_\mp = (\ell^* \mp S/2)/\sqrt{S}$. Since $p = P(\text{state } 1 \mid \ell)$, $\mathbb{E}[p\mathbf{1}\{\ell < \ell^*\}] = \frac{1}{2}\Phi(a_-)$, and $P(\ell < \ell^*) = \frac{1}{2}[\Phi(a_-) + \Phi(a_+)]$. So $G(S) = \frac{1}{2}[p^*\Phi(a_+) - (1-p^*)\Phi(a_-)]$. Differentiate: $a'_- = -\ell^*/(2S^{3/2}) - 1/(4\sqrt{S})$ and $a'_+ = -\ell^*/(2S^{3/2}) + 1/(4\sqrt{S})$. Since $a_+^2 - a_-^2 = 2\ell^*$, $\varphi(a_+) = e^{-\ell^*}\varphi(a_-)$, and $p^* e^{-\ell^*} = 1 - p^*$, so $G'(S) = \frac{1}{2}(1-p^*)\varphi(a_-)(a'_+ - a'_-) = \frac{(1-p^*)\varphi(a_-)}{4\sqrt{S}}$. With $\varphi(a_-) = (2\pi)^{-1/2}\exp(-\ell^{*2}/(2S) + \ell^*/2 - S/8)$ and $(1-p^*)e^{\ell^*/2} = \sqrt{p^*(1-p^*)}$, this is the stated expression.

(ii). $\frac{d}{dS}\log G'(S) = -\frac{1}{2S} + \frac{\ell^{*2}}{2S^2} - \frac{1}{8} = \frac{4\ell^{*2} - 4S - S^2}{8S^2}$, which is positive iff $S^2 + 4S - 4\ell^{*2} < 0$, that is, iff $S < S_c = 2(\sqrt{1+\ell^{*2}} - 1)$. So G' increases on $(0, S_c)$ and decreases after: G is convex and then concave. For $\ell^* = 0$, $G'(S) \propto S^{-1/2}e^{-S/8}$ is strictly decreasing, and $S_c = 0$.

(iii). If $r_1(0,\alpha) \geq 0$, then since $r_1(p,\alpha)$ is increasing in p and $H \geq 0$, the deployment set is the whole interval of beliefs for every $t \geq 0$, after both pilots. In Proposition 11(i) the contribution of α is therefore $\beta^2\lambda\{\mathbb{E}_B[H] - H(1/2,\alpha)\}$. Now $\mathbb{E}_B[H] = \Delta[G(S_B + S_Y) - G(S_B)]$, the value with pilot and report minus the value with the pilot only, and $H(1/2,\alpha) = \Delta[G(S_Y) - G(0)]$.

(iv). $\kappa = \int_{S_B}^{S_B+S_Y} G' - \int_0^{S_Y} G'$. If $\ell^* = 0$, G' is strictly decreasing, so $\kappa < 0$ for $S_B > 0$. If $S_B + S_Y \leq S_c$, G' is strictly increasing on $(0, S_c)$, so $\kappa > 0$. □ ## Proof of Proposition 13

Part (i). Write $Z_t = \theta_t + \varepsilon_t$. For a permutation σ of the types, the vector $(Z_{\sigma(1)}, \ldots, Z_{\sigma(K)})$ is the same function of $((\theta_{\sigma(t)}, \varepsilon_{\sigma(t)}))_t$ that Z is of $((\theta_t, \varepsilon_t))_t$. By Assumption E, the pair $(\theta_\sigma, Z_\sigma)$ has the same law as (θ, Z). The pooled report S is the unordered collection of the entries of Z, so S is unchanged by permuting its entries, and (θ_σ, S) has the same law as (θ, S). Hence $\mathbb{E}[\theta_{\sigma(t)} \mid S] = \mathbb{E}[\theta_t \mid S]$ almost surely for every t and σ. Taking σ to be the transposition of s and t proves (i). Let $\bar{m}(S)$ denote the common value.

Part (ii). Given S, every type has posterior mean $\bar{m}(S)$, so the recipient's pooled value is $\varphi(\bar{m}(S))$. Given Z, the best choice yields $\varphi(\max_t m_t(Z))$ because φ is nondecreasing. By the tower property, $\mathbb{E}[m_t(Z) \mid S] = \mathbb{E}[\theta_t \mid S] = \bar{m}(S)$ for each t. Let $X = \max_t m_t(Z)$. Then $\mathbb{E}[X \mid S] \geq \max_t \mathbb{E}[m_t(Z) \mid S] = \bar{m}(S)$. Since φ is convex and nondecreasing, conditional Jensen's inequality gives

$$\mathbb{E}[\varphi(X) \mid S] \geq \varphi(\mathbb{E}[X \mid S]) \geq \varphi(\bar{m}(S)).$$

Taking expectations proves (ii).

Part (iii). Write $\varphi(x) = a - c + \max\{x^*, x\}$ with $x^* = s - a + c$. For a random variable X with finite mean, Jensen's inequality $\mathbb{E}\max\{x^*, X\} \geq \max\{x^*, \mathbb{E}X\}$ is strict if and only if $P(X < x^*) > 0$

and $P(X > x^*) > 0$. Applying this conditionally on S on a set of positive probability makes the first inequality in the display strict, so the expectation inequality is strict. □

Numerical values in Appendix G.1. With θ_t, ε_t independent standard normal and $K = 3$, $m_t(Z) = Z_t/2$ and $\bar{m}(S) = \bar{Z}/2$, where $\bar{Z}$ is the mean of the Z_t and has variance $2/3$. The linked value is $\int \max\{0, 0.1 + 0.5\sqrt{2}x\}\, 3\phi(x)\Phi(x)^2\, dx = 0.717473$, using the density of the maximum of three standard normals. The pooled value is $\mathbb{E}\max\{0, 0.1 + 0.5\bar{Z}\} = 0.217729$. The script `verify_v07.py` reproduces both numbers, and a Monte Carlo with four million draws gives 0.71731 and 0.21747. The Monte Carlo also confirms that the regression coefficients of θ_1 and θ_2 on the order statistics of Z coincide up to sampling error, as part (i) requires.

E.9 Proof of Proposition 14

Part (i). Each $w_j(\pi) = \sup_x \int u_j(x, \omega)\pi(d\omega)$ is a supremum of affine functions of π, so it is convex. The posterior π_Y is a martingale: $\mathbb{E}[\pi_Y \mid \pi] = \pi$. By Jensen's inequality, $\mathbb{E}[w_j(\pi_Y) \mid \pi] \geq w_j(\mathbb{E}[\pi_Y \mid \pi]) = w_j(\pi)$. With a compact action set and payoffs continuous in the action, a countable dense set of actions suffices, so the supremum is of countably many affine functions. Summing over j gives $H \geq 0$.

Part (ii). The source compares its payoff with and without deployment, which differs by $r(\pi)$. The planner adds the change in the recipients' expected value. With probability λ the report arrives and the recipients' joint value is $\sum_j \mathbb{E}[w_j(\pi_Y)]$. Otherwise it is $\sum_j w_j(\pi)$. The expected change is $\lambda H(\pi)$ and, discounted, $\beta\lambda H(\pi)$. Since $\beta\lambda H \geq 0$ by part (i), $r \geq 0$ implies $r + \beta\lambda H \geq 0$, which gives the containment.

Part (iii). Let $g = r + \beta\lambda H$. At π^*, $g(\pi^*) > 0$ and $r(\pi^*) < 0$. Both functions are continuous, so both strict inequalities hold on an open neighborhood of π^*. On it the planner deploys and the source does not. □

E.10 Proof of Proposition 15

Part (i). For fixed (π), the map $\lambda \mapsto P(r + \beta\lambda H)$ is the maximum of an affine function and zero, with slope $\beta H \geq 0$ by Proposition 14(i). It is therefore nondecreasing and convex. Expectation preserves both properties, and the remaining terms of W_q do not depend on λ. Boundedness of r, H allows differentiation under the expectation away from ties, which gives the derivative. Corollary 3 uses only this derivative and $H = 0$ at degenerate beliefs.

Part (ii). The source's date-1 payoff from deploying exceeds its payoff from not deploying by $r + \rho\beta\lambda H$. The set where this is nonnegative expands in ρ because $H \geq 0$, and at $\rho = 1$ it is the planner's rule from Proposition 14(ii).

Part (iii). Subtract U_q^{ρ} from W_q. The terms $b_q - C$ and $\beta \mathbb{E}_q[s]$ cancel. What remains is the stated expression, and at $\rho = 1$ the first term vanishes.

Part (iv). Let $\Delta_U = (U_B^1 - b_B) - (U_A^1 - b_A)$. The source prefers B if and only if $\Delta_U > d = b_A - b_B$. By part (iii) at $\rho = 1$, $W_q - U_q^1 = \beta^2 \mathbb{E}_q[w]$, so the planner's gain from B is $\Delta_U + \beta^2(\mathbb{E}_B[w] - \mathbb{E}_A[w])$. If B is Blackwell-more informative than A, then the belief distribution under B is a mean-preserving spread of that under A, and w is convex, so $\mathbb{E}_B[w] \geq \mathbb{E}_A[w]$. If w is affine, $\mathbb{E}_B[w] = \mathbb{E}_A[w]$. □

Numerical values. The script `verify_v08.py` computes $\mathbb{E}_A[w] = 0.100000$ and $\mathbb{E}_B[w] = 0.238080$ in the example of Section 7.3. The identity in part (iii) at $\rho = 1$ holds to 10^{-16} for both pilots. The planner's continuation values 0.254591 and 0.447541 equal the source's full-capture values 0.164341 and 0.232674 plus $\beta^2 \mathbb{E}_q[w]$, and $\Delta_P(1) - \Delta_U = 0.124617 = \beta^2(\mathbb{E}_B[w] - \mathbb{E}_A[w])$.

E.11 Proof of Proposition 16

If π is degenerate, the posterior after any report equals the prior, so $\mathbb{E}[w_j(\pi_Y) \mid \pi] = w_j(\pi)$ and $H(\pi) = 0$. The linked and pooled versions then have equal value because both equal w. With $H = 0$ the planner's rule $r + \beta\lambda H \geq 0$ reduces to the source's rule $r \geq 0$. If beliefs about the correspondence are degenerate before both pilots, then π_A and π_B are the same degenerate belief, so $\mathbb{E}_A[w] = \mathbb{E}_B[w]$, and Proposition 15(iv) shows the two rankings coincide. □

E.12 Proof of Proposition 17

(i) For each α, $w(p, \alpha) = \max\{s(\alpha), \alpha + h - \Delta p - c_{11}\}$ is convex in the belief p. The posterior belief is a martingale, and a higher precision $x' > x$ yields a posterior distribution that is a mean-preserving spread of the one at x, because normal location experiments are ordered by precision in the sense of Blackwell. Convexity then gives $\mathbb{E}_{x'}[w] \geq \mathbb{E}_x[w]$. (ii) Write $g(x) = V_S(x; n)$ and $F(x) = \beta^2 n \mathbb{E}_x[w]$, which is nondecreasing by (i). Let $\bar{x}$ be the largest maximizer of g, which exists because X is compact and g is upper semicontinuous. For $x < \bar{x}$, $g(x) \leq g(\bar{x})$ and $F(x) \leq F(\bar{x})$, so $g(x) + F(x) \leq g(\bar{x}) + F(\bar{x})$. The planner's value $g + F$ is therefore at least as large at $\bar{x}$ as at any lower precision, so it has a maximizer weakly above $\bar{x}$. If F is strictly increasing the inequality is strict for $x < \bar{x}$ and no maximizer lies below. (iii) By Proposition 15 the planner's value equals the source's at full capture plus F, so the difference in derivatives is $F'(x)$. □ The numerical values are computed by `verify_v11.py`. # Supporting extensions This appendix collects the results that check how far the discrete model depends on its two-type structure, on the recipient's decision problem, and on the way precision is chosen. They support the main results; they are not needed to follow them. G.1 allows many types and any exchangeable dependence among returns. G.2 allows any recipient decision problem and states the pilot-choice and contracting results in that environment. G.3

treats the case in which the correspondence is known and adds a Gaussian benchmark in which covariance governs report relevance. G.4 lets the source choose the precision of its record on a discrete grid, the counterpart of Section 3. G.5 repeats the key calculations with three types. ## Many types and exchangeable returns

Let there be $K \geq 2$ observable types. The interaction returns $\theta = (\theta_1, \ldots, \theta_K)$ have any joint distribution. A linked report is $Z_t = \theta_t + \varepsilon_t$ for each type t, and the pooled report is the unordered collection $S = \{Z_1, \ldots, Z_K\}$. All statements are conditional on common information, such as α, that does not distinguish types.

Assumption E (exchangeability). The joint law of the pairs $(\theta_t, \varepsilon_t)$, $t = 1, \ldots, K$, is invariant to permutations of the type labels.

Assumption E says that ex ante the investor has no reason to expect one type to earn more than another. The two-type benchmark under the symmetric prior satisfies it. So do independent returns, and returns that share a common factor. Complete negative dependence is not required, and neither is a fixed set of return levels.

Proposition 13 (Type correspondence with many types)**.** *Under Assumption E:*

(i) $\mathbb{E}[\theta_s \mid S] = \mathbb{E}[\theta_t \mid S]$ *almost surely for all types* s, t*. The pooled report cannot rank types.*

(ii) Let a recipient install at most one type, where installing at type t *pays* $a - c + \theta_t$ *and the alternative pays* s*, all three known and common across types. Let* $m_t(Z) = \mathbb{E}[\theta_t \mid Z]$ *and* $\varphi(x) = \max\{s, a - c + x\}$*. The expected optimized value under the linked report,* $\mathbb{E}\,\varphi(\max_t m_t(Z))$*, is at least the value under the pooled report,* $\mathbb{E}\,\varphi(\mathbb{E}[\theta_1 \mid S])$*.*

(iii) The inequality is strict if, with positive probability of S*, the conditional distribution of* $\max_t m_t(Z)$ *given* S *puts positive mass strictly above and strictly below the threshold* $s - a + c$*.*

Appendix F.10 gives the proof. Part (i) is the general form of Proposition 5. The pooled report is a permutation-invariant function of the linked one, so exchangeability leaves nothing in it that favors a type.

Intuition. If nothing in the prior distinguishes the types, and the report forgets which number belongs to which type, then the report cannot distinguish them either. Only the pairing can.

For an illustration, let $K = 3$, with θ_t and ε_t independent standard normal for every type, so that returns are independent across types. Let installing at a type pay $0.1 + \theta_t$, and let the alternative pay zero. With no report the recipient installs at an arbitrary type and earns 0.1. A pooled report raises this to 0.2177, because it is informative about the common level of returns. A linked report raises it to 0.7175, an increase of 0.4997 beyond the pooled value (exact quadrature, matched by a Monte

Carlo with four million draws). With positive dependence, $\theta_t = f + u_t$ with $\mathrm{Var}(f) = \mathrm{Var}(u_t) = 0.5$, the values are about 0.318 (pooled) and 0.554 (linked), by Monte Carlo with an error near 0.001. The matching value of the report is therefore not an artifact of negative dependence. Negative dependence only makes the posterior move in opposite directions across the two types.

E.13 General recipients and the deployment wedge

Let ω be an unknown state with belief π common to the source and the planner. The source chooses whether to deploy. Deployment generates a report Y with a law that depends on ω, and the report reaches each recipient with probability λ. Let $r(\pi)$ be the source's incremental private return from deploying. A recipient $j = 1, \ldots, n$ faces any decision problem with payoff $u_j(x, \omega)$ over a compact action set. Its value at belief π is $w_j(\pi) = \max_x \mathbb{E}_\pi[u_j(x, \omega)]$. Let π_Y be the posterior after the report, and define

$$H(\pi) = \sum_{j=1}^{n} \left(\mathbb{E}[w_j(\pi_Y) \mid \pi] - w_j(\pi) \right).$$

The planner sums the payoffs of the source and the recipients, and discounts the recipients' date-2 payoffs by β.

Proposition 14 (The deployment wedge for general recipients)**.** *(i) $H(\pi) \geq 0$ for every report, every belief, and every collection of recipient decision problems.*

(ii) The source deploys if and only if $r(\pi) \geq 0$. The planner deploys if and only if $r(\pi) + \beta\lambda H(\pi) \geq 0$. The source's deployment set is contained in the planner's.

(iii) If r and H are continuous in π, and there is a belief π^* with $-\beta\lambda H(\pi^*) < r(\pi^*) < 0$, then the planner deploys and the source does not on an open neighborhood of π^*.

Appendix F.11 gives the proof. Part (i) uses only the convexity of each w_j and the fact that posteriors are a martingale, so the sign of the wedge does not depend on the structure of any recipient's problem. Only its size does. Proposition 6 is the two-type special case. Proposition 15 states the dynamic results in the same environment.

The proposition describes a report that imposes no cost on others. If a report also harms competitors or consumers, for example through market power or loss of privacy, the planner's rule contains an additional negative term and private deployment need not be too rare. Part (ii) is therefore a statement about information externalities, not a general welfare result.

Now add a first stage. Each pilot q induces a distribution over public beliefs π_q, and both pilots reveal the same state variables that do not affect the beliefs about the correspondence, such as α. Let $s(\pi)$ be the source's value without deploying, $w(\pi) = \sum_j w_j(\pi)$ the recipients' value without

the report, and assume r, s, w, H are bounded. With fee rate $\rho \in [0,1]$, define the source's and the planner's date-0 values by

$$U_q^\rho(\lambda) = b_q - C + \beta \mathbb{E}_q\big[s + P(r + \rho\beta\lambda H)\big], \qquad W_q(\lambda) = b_q - C + \beta \mathbb{E}_q\big[s + P(r + \beta\lambda H)\big] + \beta^2 \mathbb{E}_q[w],$$

where $P(x) = \max\{x, 0\}$. Setting $\rho = 0$ gives the no-payment benchmark of Section 7.1.

Proposition 15 (Pilot choice and contracting in the general environment)**.** *(i) $W_q(\lambda)$ is nondecreasing and convex in λ, with derivative $\beta^2 \mathbb{E}_q[H\mathbf{1}\{r + \beta\lambda H > 0\}]$ wherever it exists. Corollary 3 holds with K_q defined from these objects.*

(ii) Under fee rate ρ the source deploys if and only if $r + \rho\beta\lambda H \geq 0$. The deployment set expands in ρ, and at $\rho = 1$ it equals the planner's.

(iii) For every ρ,

$$W_q(\lambda) - U_q^\rho(\lambda) = \beta\, \mathbb{E}_q\big[P(r + \beta\lambda H) - P(r + \rho\beta\lambda H)\big] + \beta^2 \mathbb{E}_q[w].$$

At $\rho = 1$ the difference equals $\beta^2 \mathbb{E}_q[w]$.

(iv) At full capture, the source prefers pilot B to pilot A if and only if $\Delta_U > d$, whereas the planner prefers B if and only if $\Delta_U + \beta^2(\mathbb{E}_B[w] - \mathbb{E}_A[w]) > d$. Here Δ_U is the source's continuation gain from B at $\rho = 1$. If B is Blackwell-more informative than A, then $\mathbb{E}_B[w] \geq \mathbb{E}_A[w]$, so pricing the later report alone leaves the planner's preference for B weakly stronger than the source's. The two rankings coincide if w is affine in beliefs.

Appendix F.12 gives the proof. Part (iii) shows exactly what full payment for deployment evidence does not price: the recipient's own value of the pilot's posterior, $\beta^2 \mathbb{E}_q[w]$. In the example of Section 7.3 this is $\mathbb{E}_A[w] = 0.1000$ and $\mathbb{E}_B[w] = 0.2381$, so $\beta^2(\mathbb{E}_B[w] - \mathbb{E}_A[w]) = 0.1246$. That is the gap between the planner's gain 0.1930 and the source's full-capture gain 0.0683, as part (iv) requires.

Appendix E states and proves a Gaussian covariance benchmark, which shows that an activity is informative for other investors when its outcomes covary with theirs, whatever its own return.

E.14 When the correspondence is known

Proposition 16 (The known-correspondence limit)**.** *Suppose the beliefs about the correspondence are degenerate before the pilot, so that the assignment of returns to types is known. Then $H(\pi) = 0$ for every report, the linked and pooled versions of any report have the same value, the source's and*

the planner's deployment rules coincide, and the two pilots induce the same value of w, so the pilot rankings coincide.

Appendix F.13 gives the proof. The proposition is trivial, and it is useful for that reason. It says that in an environment with a known spillover kernel, the information channel studied in this paper is inactive, and that all of the wedges derived here come from unresolved uncertainty about which type benefits. A model of data externalities with known cross-task usefulness, such as the accumulation model discussed in Section 2, therefore contains a different distortion.

E.15 Choosing how precisely to record: a discrete grid

So far the informative pilot is a fixed alternative whose cost advantage d is given. This subsection lets the source choose how much to record and derives the cost difference from that choice. Let the type-linked pilot have precision $x \in X \subset [0,\infty)$, with X compact, so that the contrast errors have variance $1/x$ and $x = 0$ is uninformative. Recording at precision x costs the source $c(x)$, with c nondecreasing, so its immediate proceeds are $b(x) = b_0 - c(x)$. Higher precision is more informative in the sense of Blackwell. Let $V_S(x;t) = b(x) - C + \beta\,\mathbb{E}_x[s + P(r + t\beta\lambda H)]$ be the source's date-0 value at total capture t, where $\mathbb{E}_x$ averages over the posterior induced by precision x. By Proposition 15, the planner's value with n recipients is $V_S(x;n) + \beta^2 n\,\mathbb{E}_x[w]$.

Proposition 17 (Recording precision)**.** *(i) $\mathbb{E}_x[w]$ is nondecreasing in x.*

(ii) Let $\bar{x}$ be the largest maximizer of the source's value at full capture, $V_S(\cdot;n)$. The planner's value has a maximizer weakly above $\bar{x}$, and every maximizer is weakly above $\bar{x}$ if $\mathbb{E}_x[w]$ is strictly increasing. The source records at most as precisely as the planner even when it is paid the recipients' full gain from the later report.

(iii) The gap between the planner's and the source's marginal return to precision, at any x at which both are differentiable, equals $\beta^2 n \frac{d}{dx}\mathbb{E}_x[w]$ when capture is full. It is proportional to n.

Appendix F.14 gives the proof. The comparison in (ii) is only for full capture. At partial capture the difference between the planner's and the source's values contains $V_S(x;n) - V_S(x;t)$, which by Proposition 11 need not be monotone in x, so the argument does not sign it.

Example. Take the parameter set of Section 7.3 with linear cost $c(x) = kx$ and the grid $X = \{0.5, 1, 1.5, 2, 2.78, 4, 6, 8, 12, 16, 24, 32\}$, where $x = 2.78$ is the precision $1/\sigma_0^2$ of Section 7.3. For $k = 0.01$ the source chooses $x = 4$ when it receives no fee, 2.78 at full capture with one recipient, and 2 at full capture with two recipients. The planner chooses $x = 6$ for one recipient and for two. For $k = 0.005$ the source chooses 6, 6, 4, 4 at $t = 0, 0.5, 1, 2$, and the planner chooses 6 and 8 for $n = 1, 2$. For $k = 0.02$ and 0.05 the same ordering holds. In all eight pairs of full-capture

source choice and planner choice that I computed, the planner records more precisely. The expected recipient value $\mathbb{E}_x[w]$ rises from 0.158 at $x = 0.5$ to 0.275 at $x = 32$. The source's precision also falls as the capture rate rises, in line with Proposition 11. The grid is coarse and I report optima on the grid, not exact optima.

Table 4. Chosen recording precision x^* with linear cost $c(x) = kx$ (12-point grid).

k	Source, $t = 0$	Source, $t = 1$	Source, $t = 2$	Planner, $n = 1$	Planner, $n = 2$
0.005	6	4	4	6	8
0.01	4	2.78	2	6	6
0.02	2	2	0.5	4	4
0.05	1	0.5	0.5	1.5	2

Notes. The parameter set is that of Section 7.3. $t = \rho n$ is total capture; the source at $t = n$ is at full capture and is compared with the planner for the same n. The grid is $\{0.5, 1, 1.5, 2, 2.78, 4, 6, 8, 12, 16, 24, 32\}$, so entries at 0.5 may be constrained by the grid. Computed by `verify_v11.py`.

E.16 A three-type dynamic example

The dynamic example of Section 7.3 has two types. This subsection computes the same objects with three observable types, so that the results of Propositions 6, 15, and 11 are checked outside the two-type structure. The interaction levels $(0.8, 0.2, -0.4)$ are known and are assigned to types 1, 2, and 3 by one of the six permutations, each with probability $1/6$. The source is type 1, and the recipients are the separately owned sites of types 2 and 3, so the recipient value w and the information value H are sums over the two. The remaining parameters are those of Section 7.3: $\alpha \in \{-0.3, 0.3\}$ with equal probabilities, $c_{10} = 0.1$, $c_{11} = 0.35$, $\beta = 0.95$, $\sigma_0 = 0.6$, $\sigma_1 = 0.5$, and full access. The linked pilot reveals each type's level with independent normal error; the uninformative pilot leaves the six assignments equally likely. The expectation over the linked pilot is a Monte Carlo average over 60,000 draws with a fixed seed, and the report is integrated numerically.

Four results carry over. First, the deployment rules are nested: after the uninformative pilot, the source never deploys, while the planner deploys in the high-α state, which is half the cases; after the linked pilot, the source deploys with frequency 0.302 and the planner with 0.308. This is Proposition 6. Second, the source's continuation gain from the linked pilot is 0.0675 with no fee and 0.0468 at full capture, and the planner's is 0.1749, so the source and planner disagree about the pilot whenever d lies in $(0.0468, 0.1749)$. The recipients' expected baseline value is 0.200 after the uninformative pilot and 0.342 after the linked pilot, which gives the planner's extra term $\beta^2(\mathbb{E}_B[w] - \mathbb{E}_A[w]) = 0.1281$, as in Proposition 15. The Monte Carlo standard errors of the

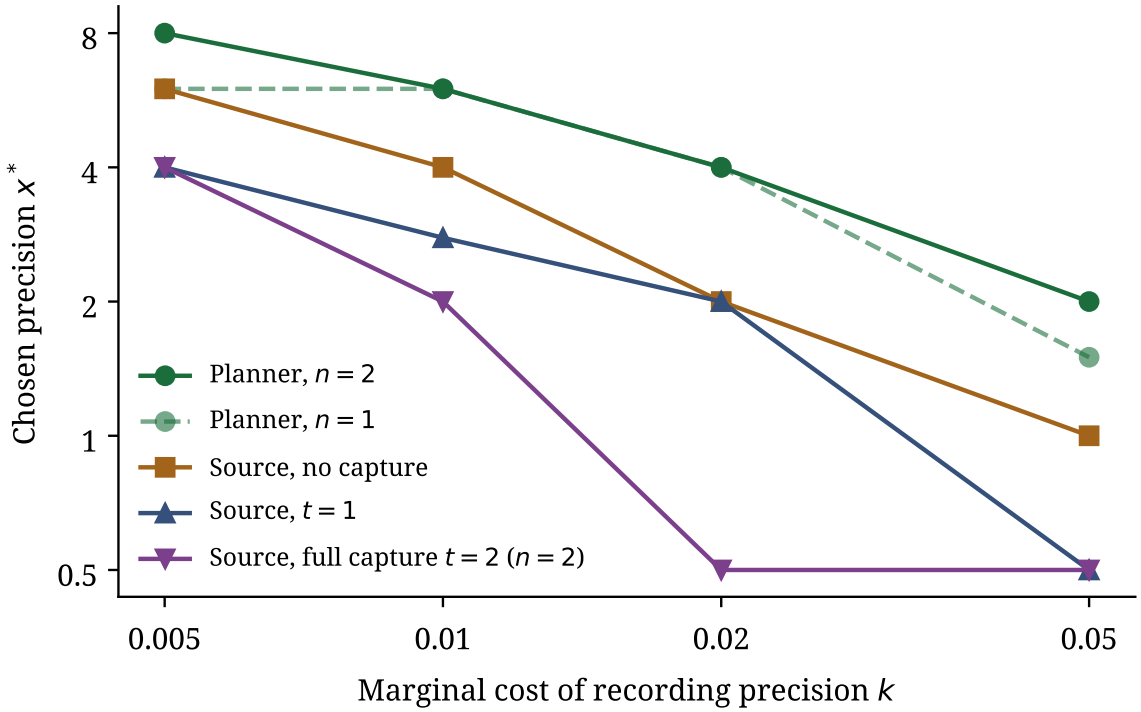

Figure 8: Chosen recording precision as the marginal cost of precision varies.

Notes. The figure plots the entries of Table 4. In every full-capture pair the planner records at least as precisely as the source, and the source's precision falls as capture rises; the figure is a numerical illustration of Proposition 17 on a grid, not a general result for partial capture.

two gains, from ten batches, are 0.0004 and 0.0002. Third, the slope of the source's gain in total capture is $+0.0040$ at $t = 0$ and $+0.0040$ at $t = 0.25$, where the report is not deployed after the uninformative pilot, and -0.0445, -0.0442, and -0.0439 at $t = 0.5, 1, 2$, once it is. This is the sign pattern of Proposition 11, with the activation point again between 0.25 and 0.5. The slopes are computed from the formula of Proposition 11(i) on the simulated draws and carry no separate error estimate. Fourth, the source's gain from the linked pilot falls from 0.0675 at no capture to 0.0027 at $t = 2$.

The three-type example confirms the signs of the two-type calculation. It is a numerical check on one parameter set, not a proof, and the Monte Carlo error is small relative to the gaps reported but is not zero. The code is `verify_v13.py`.

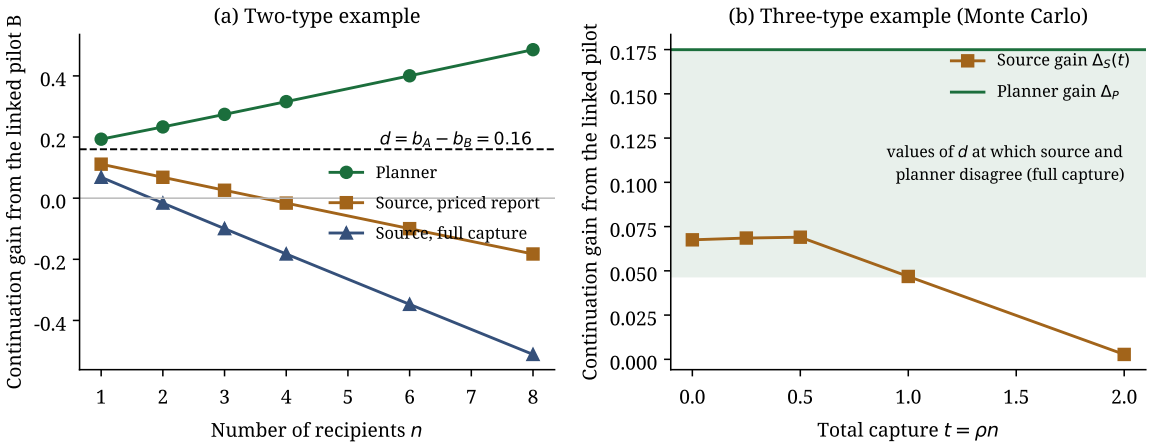

Figure 9: Market size and the pilot-choice wedge: two-type and three-type examples.

Notes. The figure plots continuation gains from the linked pilot. Panel (a) is the two-type example of Table 3. Panel (b) is the three-type example of Appendix G.5 by Monte Carlo (60,000 draws, fixed seed); the shaded band is the set of values of d for which the source and the planner choose differently at full capture. Both panels use one parameter set each.