

The Value of Information in a Congested Fishery

Gabriel Englander

Larry Karp

Leo Simon



WORLD BANK GROUP

Development Economics

Development Research Group

August 2023

Abstract

Congestion can reduce the value of a fishery, resulting in a lower total catch for the same amount of labor, fuel, and equipment expended in fishing activities. Absent the congestion externality, better information about the location and size of fish stocks enables fishers to make more efficient decisions. However, more precise information can cause fishers to converge on the same location or increase fishing at the same time. The cost of the resulting increased congestion can outweigh the direct benefit of better information. This paper identifies the circumstances where an increase in the precision of public and/or private information (about stock size or location) lowers industry profits. Using high-resolution data from Peru's anchoveta fishery, the world's largest by catch volume, the research reveals that despite considerable congestion, more precise private

information would increase expected profits. On the other hand, the profit impact of more precise public information is positive but significantly smaller. This difference reflects the fact that public information increases congestion to a much greater extent, compared to private information. The policy implications are that improving private information about fish stocks—for example through firms investing in forecasting and decision-making technology—could increase industry profits. But anchoveta fishers would not necessarily benefit from more precise public information. As fishery managers control the accessibility and disclosure of information, decisions to make private information public, such as publishing near real-time catch data, could potentially lower fisher profits.

This paper is a product of the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at aenglander@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

The value of information in a congested fishery*

Gabriel Englander[†] Larry Karp[‡] Leo Simon[§]

KEYWORDS: value of information, fishery congestion, Peruvian anchoveta

JEL CLASSIFICATIONS: D83, Q22, Q28, Q56, O13

*We thank Philippe Marcoul, Martin Smith, Kate Pennington, and seminar participants at UC Santa Barbara, UC Berkeley, the London School of Economics and University of Florida for helpful comments. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

[†]Development Research Group, World Bank. 1818 H St NW, Washington, DC 20433. aenglander@worldbank.org

[‡]Department of Agricultural and Resource Economics; University of California, Berkeley. 207 Giannini Hall Berkeley, CA, 94720-3310, USA. karp@berkeley.edu.

[§]Department of Agricultural and Resource Economics; University of California, Berkeley. We deeply regret that Leo died before this paper's completion. Leo, we will miss you.

1 Introduction

Congestion can reduce the value of a fishery, lowering aggregate catch for a given level of fishing “effort” (Brown, 1974). Absent the congestion externality, better information about the location and size of fish stocks enables fishers to make more efficient decisions. However, more precise information can cause fishers to converge on the same location or increase effort at the same time. The cost of the resulting increased congestion can outweigh the direct benefit of better information. Adapting a model by Angeletos and Pavan (2007) (hereafter, AP07), we identify the circumstances where an increase in the precision of public and/or private information (about stock size or location) lowers fishery profits.

We estimate our structural model using high-resolution data from Peru’s anchoveta fishery, the world’s largest fishery, accounting for 8% of global marine fish catch (FAO, 2018). To our knowledge, ours is the first paper that estimates a structural model capable of distinguishing between the value of more precise public and/or private information in a setting where better information might lower equilibrium payoffs. We find that, despite empirical evidence of substantial congestion, more precise private information would increase expected profits. Our point estimate of the welfare effect of more precise public information is also positive, but small in magnitude. Given parameter uncertainty, it is plausible that more precise public information lowers welfare. The point estimate of the elasticity of welfare with respect to private information is over 20 times as large as the estimate of the welfare elasticity with respect to public information. This difference reflects the fact that public information increases congestion to a much greater extent, compared to private information.

It is well-understood that strategic interactions or externalities can cause more precise information to lower profits or welfare. Vives (1984) provides an early analysis of this possibility, which illustrates the theorem of the second best. Marcoul (2020) builds on Vives’ model to describe a fishery in which vessels choose their individual level of effort to maximize their own welfare. Due to congestion, each vessel’s effort creates a negative externality. Vessels can potentially band together in clubs to share information, but not to collectively choose effort. Marcoul identifies the parameter region for which the increased information resulting from this pooling lowers equilibrium profits, eliminating vessels’ incentive to form clubs. Gatewood (1984) provides examples where fishers share information about stock location.

Most empirical studies on the value of information start from the premise that better information improves efficiency. For example, Dufflo et al. (2018) study a regulator with imperfect information about plants’ pollution emissions. They find that perfect information about emissions would lead to a 30% increase in equilibrium abatement, the same magnitude as a 30% increase in the inspection budget. In different settings, firms might respond to the regulator’s better information either by reducing emissions or by increasing defensive

measures such as bribery or litigation. Firms' equilibrium response reinforces the value of information in the first case, and erodes it in the second. Sallee (2014) studies consumers' rational inattention when purchasing energy-intensive durables. A reduction in consumers' cost of learning about the product's energy efficiency might make energy efficiency a more salient characteristic, leading to more efficient choices. Firms might respond to consumers' changed behavior by offering more energy-efficient products, thereby reinforcing the initial beneficial change. These examples illustrate that general equilibrium responses to improved information might be of first order importance.

In settings where these kinds of equilibrium effects are important, the literature on the value of information distinguishes between public and private information.¹ Both are correlated with the true state of nature, e.g. the location or the size of the stock of fish. All agents receive the public information, but agents' private information is uncorrelated, conditional on the state of nature. Greater precision of either type of information enables agents to tailor their actions more precisely to the true state of nature, a change that tends to increase their profits. However, increased precision of either type of information can also increase congestion, so the net effect on profits is ambiguous.

More precise public information causes fishers to make similar decisions, e.g. to move toward the same location or to increase effort at the same time. The resulting increase in the correlation of fishers' actions increases congestion. Private information depends on the true state of nature, so its greater precision can have the same qualitative effect. However, because private information is (conditionally) uncorrelated across agents, this tendency is less pronounced; increased precision of private information is more likely to improve efficiency.

AP07 provide our starting point, but our focus and analytic results differ. AP07 consider a more general model, and they also compare the Nash equilibrium with the outcome under a team. Agents in the Nash equilibrium choose actions to maximize their own welfare. In the team setting, they choose actions to maximize aggregate welfare, but without being able to share information.² We are interested in a special case of their model, relevant to the fishery, and our research question does not involve the team problem. For these reasons, we are able to provide more definitive results and also to extend the domain of parameter space. In the case closest to our model, AP07 obtain either necessary or sufficient conditions (but not both) for more precise public or private information to improve welfare. We obtain a condition that is both necessary and sufficient, thus sharpening the results. Our extension of the parameter set over which the analysis is valid is useful because it includes regions with greater congestion

¹The literature we follow studies games with a unique equilibrium under certainty. A distinct literature studies the role of information in coordination games with multiple equilibria under certainty; imperfect information can render the equilibrium unique (Carlsson & Van Damme, 1993; Morris & Shin, 1998).

²Marcoull (2020) consider the opposite case; there, firms can share information but not coordinate actions.

costs. We also allow the public and private signals to be correlated. Private signals remain uncorrelated with each other, conditional on the state of nature and the public signal.

This analysis provides the foundation for our empirical study. The Peruvian anchoveta fishery is economically and ecologically important, and the setting provides an opportunity to empirically study a class of problems that has received widespread attention from theorists, but not from empiricists. We find substantial evidence of congestion costs and a positive correlation between public and private information. This correlation affects the magnitude, but not the sign of the welfare effects of more precise public or private information.³

Huang and Smith (2014) and Sanz and Diop (2021) provide empirical evidence of the importance of congestion in fisheries. Huang and Smith note that greater congestion can improve efficiency by partially offsetting a stock externality: excessive harvest that reduces the future stock. Unlike Huang and Smith (2014), we estimate only spatial congestion, not a dynamic congestion-related externality.⁴ Regulation in the Peruvian anchoveta fishery limits aggregate and individual vessel catch. To the extent that these limits are set efficiently, they solve the stock externality. Here, because of the absence of an offsetting distortion, congestion always lowers efficiency because it increases the cost of harvesting a given level of catch.

2 Institutional context

The regulator of Peru’s North-Central anchoveta fishery implements a management regime that promotes biologically sustainable and economically efficient fishing.⁵ There are two fishing seasons per year, each of which lasts about three months. The regulator (PRODUCE), in consultation with the marine science agency (IMARPE), sets an industry-wide limit on the total tons that can be landed each season, called the total allowable catch (TAC). Tons “landed” refers to vessels’ transfer of their catch to a processing plant. Since the TAC usually binds, more precise information would not increase aggregate landings in most seasons (Kroetz

³There are interesting parallels between the team problem and the effect of correlation between public and private signals. However, these topics are tangential to our focus, so we raise them only in footnotes and in Appendix B. With high correlation between public and private signals and relatively noisy public signals, the best linear unbiased estimator of the state of nature puts negative weight on the public signal. Even with zero correlation, but sufficiently high congestion costs, agents in the team problem put negative weight on the public signal.

⁴Congestion is a static externality, in that it reduces efficiency conditional on the stock size. An open access fishery tends to reduce rents by lowering the stock. Congestion has the effect of throwing sand in the wheels; it is analogous to an apparently inefficient gear restriction, e.g. requiring vessels to carry a ton of cement. By discouraging fishing, this policy can protect the fish stock, thereby reducing the dynamic externality. The net effect might be to increase efficiency: another example of the theory of the second best.

⁵Peru’s North-Central stock occurs entirely within Peruvian jurisdiction. The stock ranges from Peru’s northern boundary to the 16th parallel south. Peru’s Southern anchoveta stock is shared with Chile and is subject to different regulations. The North-Central stock accounts for 95% of tons caught per year (Englander, 2022). This paper analyzes data from the North-Central fishery only.

et al., 2016). Individual vessel quotas (IVQs) divide the TAC among vessels.⁶ More precise information could help vessels reach their quota at lower cost, for example by reducing search costs. Search comprises 20% of time spent during fishing trips (Joo et al., 2015).

All vessels receive the same price for their catch (Englander, 2022; Hansman et al., 2020).⁷ There are 17 main ports in the North-Central fishery at which vessels can land their catch. Landings data for 2016 enable us to calculate that 47% of vessels’ landings occur at their most common port. The location of catch does not affect the TAC or IVQs.

We view information sharing as occurring within firm, rather than among vessels who land their catch at the same ports. Firms typically consist of several vessels.⁸ Fishers tend to avoid revealing the locations of productive fishing grounds, but fishing firms promote information sharing within firms (Marcoul, 2020; Welch et al., 2022). For example, the two large firms visited by one of the authors in December 2019 had central “command centers” in which staff receive catch data from individual vessels, aggregate the information, and send daily reports with suggested fishing locations to their vessels.

3 The model

Every vessel receives a public and a private signal and then decides where to locate, not (for example) its level of effort. This modeling choice maintains consistency with our data set. We examine the equilibrium welfare effects of greater precision for both types of information. Individual vessels would like to be close to the ideal fishing ground, where the fish are densest. Vessels face either congestion costs or benefits from operating close to each other. If there are congestion costs, each vessel would like to be far from other vessels; if there are positive externalities, each vessel wants to be close to other vessels. Using their public and private information, vessels estimate the ideal location and they form beliefs about other vessels’ location decisions. Each vessel bases its location decision on this estimate and these beliefs.

We use a special case of AP07’s model, but we consider a larger parameter set and allow for correlation between public and private signals. Section 3.2 summarizes the relation between our results and those in AP07 and discusses the team problem; Appendix B contains details.

⁶For example, the IVQ of TASA 71 entitles that vessel to land 0.439043% of the TAC each season. In the second season of 2018, when the TAC was 2.1 million tons, the vessel could land up to 9,220 tons.

⁷The price per ton is a fixed fraction of the price of fishmeal in Hamburg (Fréon et al., 2014). Thus, more precise information will not change prices, as occurred in South Indian fisheries (Jensen, 2007). 97% of tons landed of anchoveta are processed into fishmeal and fish oil (PRODUCE, 2018).

⁸Seven large firms own at least 19 vessels each; 271 “singleton” vessels belong to a firm that owns only one vessel. The remaining vessels belong to firms that own between 2 and 10 vessels.

3.1 The fishing context

Each fishing vessel decides where to locate on the real line. Vessels would like to locate close to the ideal point θ , where the stock of fish is densest. If there are safety benefits or information spillovers, each vessel wants also to locate close to other vessels. If congestion is important – the case we emphasize – there is an incentive to locate far from other vessels. The vessel’s payoff consists of two components, a cost that depends on the distance between that vessel and the ideal location, θ , and a benefit that arises from either being far from (in the case of congestion) or close to (in the case of positive externalities) other vessels.

Before deciding where to locate, all vessels receive the public signal $y = \theta + \varepsilon_y$, with $\varepsilon_y \sim (0, \sigma_y^2)$; each vessel i receives a private signal $x_i = \theta + \varepsilon_{x_i}$, with $\varepsilon_{x_i} \sim (0, \sigma_x^2)$. Vessels’ private signals are uncorrelated, conditional on y . However, the public and private signals may be correlated, with $\mathbf{E}\varepsilon_y\varepsilon_{x_i} = \rho\sigma_y\sigma_x$, $-1 < \rho < 1$. Our analytic results emphasize uncorrelated signals, the case considered by the previous literature. Here we obtain sharp results that can be compared to those in the literature. However, correlation is empirically important in the anchoveta fishery, so our model needs to include this feature.

Denote the ratio of standard deviations as $r \equiv \frac{\sigma_y}{\sigma_x}$; a larger r corresponds to a relatively less precise public signal. After receiving the public and private signals, vessel i ’s Best (minimum variance) Unbiased Linear Estimator (BLUE) for θ is $\delta y + (1 - \delta)x_i$, with⁹

$$\delta \equiv \frac{1 - \rho r}{1 + r^2 - 2\rho r}. \quad (1)$$

The BLUE weight on public information, δ , is positive unless $\rho r > 2$. When the public and private signals are highly positively correlated and in addition the public signal is relatively noisy (so that ρr is large), the public signal contributes a great deal of noise relative to its additional information. In that case, the BLUE gives negative weight to the public signal. The weight on the private signal is positive ($1 - \delta > 0$) unless $\rho > 2r$. When the public and private signals are highly correlated and r is small, the private signal contributes a great deal of noise relative to its additional information, so the BLUE gives that signal negative weight.

When vessel i locates at point k_i its expected payoff is

$$\frac{B}{2}\mathbf{E}_{\{k_j\}} \int (k_i - k_j)^2 dj - \frac{A}{2}\mathbf{E}_\theta (k_i - \theta)^2, \quad (2)$$

⁹We use a quadratic model and restrict attention to linear strategies, so only the first two moments of the distribution matter. We denote δ as the BLU estimator, without reference to Bayesian weights. Therefore, we do not need to specify the distribution associated with the state of nature, θ . In particular, we do not require the commonly invoked assumption of normality. Although the higher moments of the distribution are irrelevant in the linear-quadratic model, the appropriateness of the linear quadratic model does depend on those moments. For example, if the distribution is not single-peaked, both the linear quadratic model and the normality assumption would provide poor approximations.

where $\{k_j\}$ is the profile of other vessels' locations. The parameters A and B are constants; $A > 0$ determines a vessel's cost of being far from the ideal location. If fish are spatially diffuse, a vessel's position relative to the ideal location matters little, and $A \approx 0$. However, if fish are concentrated near the ideal location, the penalty of being distant from it is significant; here, A is large. The magnitude of A also depends on the total biomass relative to fishers' capacity. If the biomass is very large, and a vessel's capacity to catch fish is limited, distance from the ideal location might be unimportant; in this case, also, A is small. The integral in Equation 2 measures the impact on a vessel's payoff of *dispersion*, the distance between an arbitrary vessel and the average location of other vessels. For $B > 0$ there is congestion: vessel i benefits from being far from other vessels, so that its payoff increases with the average distance between it and other vessels. For $B < 0$, proximity to other vessels benefits i .¹⁰ Our payoff function implies that the benefit of dispersion and the benefit of proximity to θ are additive. It is not possible to include an "interaction term" (one involving $(k_i - k_j)(k_i - \theta)$) within a quadratic model, because the expectation of that bilinear term equals 0.

We consider only equilibria that are linear in information, that is, i 's location is $k_i = \gamma y + (1 - \gamma) x_i$. The absence of a constant in this decision rule, and the fact that the weights sum to 1, imply that the decisions are "unbiased", in that a vessel's expected location equals the ideal: $\mathbf{E}_{y, x_i} k_i = \theta$. AP07 show that unbiasedness is a result, not an assumption. Appendix B confirms that if we restrict equilibria to be linear in information, then the formulation in which vessels choose their locations is equivalent to our formulation, in which they choose the parameter of their linear decision rule. This linear-unbiased formulation means that vessel i believes that vessel $j \neq i$ chooses $k_j = \eta y + (1 - \eta) x_j$. In a symmetric equilibrium, $\gamma = \eta$.

We define $\tau \equiv \frac{B}{A}$, a measure of congestion costs relative to the cost of deviating from the ideal location. Appendix A shows that the second order condition to the agent's problem requires $\tau < 1$. Hereafter we assume this inequality holds. Using the definitions of τ and δ (from Equation 1), familiar calculations show that the vessel's weight on the public signal in the unique symmetric linear decision rule is γ^{NE} , with

$$\gamma^{NE} = \frac{1 - \tau - \rho r}{1 - \tau + r^2 - 2\rho r}. \quad (3)$$

(See Appendix A.) When $\tau = 0$, an agent's payoff is unaffected by other agents' locations. In this limiting case, we confirm (using Equations 1 and 3) that $\gamma_{\tau=0}^{NE} = \delta$. In this case, agents use the BLUE weight on the public signal. For another special case, $\rho = 0$, Equation 1 implies that $1 + r^2 = \delta^{-1}$. This result and Equation 3 imply that for $\rho = 0$, the weight on public

¹⁰Although we do not assume a particular distribution for the fish density (conditional on θ), we do assume that density decreases monotonically with the distance from θ , and catch increases monotonically with the density. The quadratic function is an approximation of the combination of these two forces.

information in the equilibrium decision rule simplifies to

$$\gamma_{|\rho=0}^{NE} = \delta \frac{1 - \tau}{1 - \tau\delta} < \delta \text{ iff } \tau > 0. \quad (4)$$

With $\rho = 0$ the Nash equilibrium weight on the public signal is smaller than the BLUE weight, if and only if there is congestion ($\tau > 0$). This comparison need not hold for large ρ . Just as with δ , the Nash equilibrium weight lies in $(0, 1)$ when $\rho = 0$, but it can lie outside that interval when ρ is large; it is non-monotonic and discontinuous in parameters. Appendix C provides comparative static results for γ^{NE} and δ .

With non-zero correlation, we need numerical methods to evaluate the effect of better information on welfare (industry profit). However, for $\rho = 0$, Appendix A establishes:

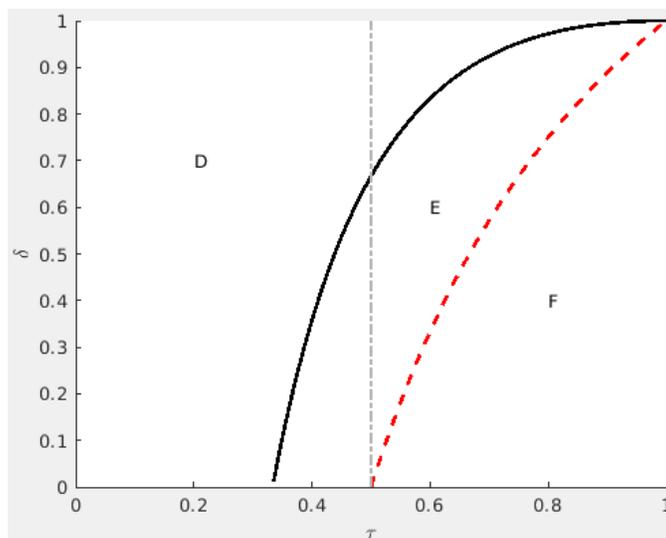
Proposition 1. *Assume $\rho = 0$ and $\tau < 1$. (i) For $\tau < 0$ (where vessels benefit from proximity to other vessels) an increase in either public or private information raises welfare. (ii) For $\tau > 0$ (where there are congestion costs), equilibrium welfare increases with the precision of public information if and only if $\delta > \frac{3\tau-1}{\tau^2+\tau}$. Equilibrium welfare increases with the precision of private information if and only if $\delta > \frac{2\tau-1}{\tau}$. (iii) The higher the relative precision of the public signal (the larger is δ) the larger is the critical level of congestion above which more precise (public or private) information lowers welfare.*

Figure 1 graphs the thresholds identified in Proposition 1 for $0 < \tau < 1$. Greater precision of either public or private information has two effects. It increases accuracy, enabling vessels to move closer to the ideal location, and it reduces dispersion because other vessels also move closer to the ideal location. When a vessel benefits from proximity to other vessels ($B < 0$) both of these changes increase the vessel's payoff. Thus, for $B < 0$ ($\tau < 0$) an increase in the precision of either signal increases welfare.

For $\tau > 0$, where vessels benefit from dispersion (i.e. they are harmed by congestion), the welfare effect of greater precision of either signal is ambiguous. Nevertheless, it is intuitive that greater precision of the public signal, compared to greater precision of the private signal, has a more pronounced effect on decreasing dispersion than on increasing accuracy.¹¹ Therefore, for $\tau > 0$, if greater precision of the public information raises welfare, so does greater precision of the private signal. In region D , where congestion costs are low, greater precision of either signal raises welfare, just as is the case when $\tau < 0$. In region E , with moderate congestion costs, welfare increases with the precision of private information but falls with the precision of public information. The explanation turns on the differing relative effects of the two types

¹¹ AP07 confirm this interpretation, noting that greater precision of the public information, for a given level of precision of the private information, increases the correlation of agents' decisions. Greater precision of the private signal, for a given level of precision of the public signal, lowers this correlation.

Figure 1: Welfare effects of more precise public or private information when $\rho = 0$



Notes: The solid graph is the threshold above which welfare increases with the precision of public information. The dashed graph is the threshold above which welfare increases with the precision of private information. In region D , welfare increases with the precision of both public and private information. In region E , welfare increases with the precision of private information but falls with the precision of public information. In region F , an increase in the precision of either public or private information lowers equilibrium welfare.

of signals on accuracy and dispersion. As noted above, greater precision of either signal increases accuracy and lowers dispersion, but a more precise public signal has a relatively greater effect on the reduction of dispersion. As the public signal becomes more precise, fishers put more weight on it and less on their private signals, thereby reducing information heterogeneity and increasing congestion. Thus, with moderate congestion costs, more precise private information raises welfare and more precise public information lowers welfare.

In region F , where congestion costs are high, the additional cost of greater proximity to other vessels outweighs any benefit of greater proximity to the ideal location. Here, an increase in the precision of both public and private information lowers equilibrium welfare.

Parts (ii) and (iii) of Proposition 1 give the formulae for the graphs of the thresholds shown in Figure 1. For example, for $\delta \approx 0$, where the public information is uninformative, and at the moderate congestion parameter $\tau = 0.34$ (just inside region E), more precise public information contributes very little to getting closer to the ideal location, but, by increasing vessels' dependence on the public signal, it increases congestion by enough to lower welfare. For the same level of τ but a significantly higher δ , the improved accuracy justifies the increase in congestion, so here more precise public information raises welfare.

The proposition implies

Corollary 1. *The upper boundary of region E (the solid curve in Figure 1) identifies the combination of parameters at which welfare is minimized with respect to the precision of public information. The lower boundary of region E identifies the combination of parameters at which welfare is maximized with respect to the precision of private information.*

For example, beginning in region E , holding fixed the precision of private information and increasing the precision of public information, increases δ , resulting in a northward movement in the (τ, δ) plane. This change lowers welfare until we reach the upper boundary of E , beyond which a further increase in the precision of public information raises equilibrium welfare. Thus, welfare is minimized with respect to the precision of public information on the boundary between regions E and D . Similarly, an increase in the precision of private information, holding the precision of public information fixed, lowers δ , resulting in a southward movement in the (τ, δ) plane. For $\tau > 0.5$, this process eventually reaches the lower boundary of E , beyond which further increases in precision of private information lower welfare. Thus, for given τ and σ_y , welfare is maximized on the boundary between regions E and F .

3.2 Relation to AP07 and the team problem

In the team problem, agents do not share information, but each agent chooses its action to maximize expected group welfare rather than individual welfare. AP07 describe the team outcome as efficient. Our model belongs to the class that they describe as being “inefficient only under incomplete information” (Section 5.4 of AP07). An example shows that this efficiency result is conditioned on the assumption that agents use symmetric decision rules: agents with the same information make the same decision. We then summarize the relation between our and AP07’s analytic results for this class of models under the symmetry assumption.

Both the non-cooperative and the team equilibria assume linearity and symmetry, implying that both decisions rules are unbiased: $\mathbf{E}_{y,x_i} k_i = \theta$. This conclusion means that the two decision rules both converge to $k_i = \theta$ under complete information. Thus, the non-cooperative and the team decision rules are equally efficient under common knowledge. However, neither is, in general, first best. When congestion is important, full efficiency may require otherwise similar agents to choose different locations, violating the symmetry assumption.

A two-vessel example under complete information illustrates this point. Suppose that θ is common knowledge; we normalize by setting $\theta = 0$. The aggregate payoff is

$$2\frac{B}{2}(k_1 - k_2)^2 - \frac{A}{2}(k_1^2 + k_2^2). \quad (5)$$

The social optimum must involve $k_1 = -k_2$ because it cannot be efficient to locate both vessels on the same side of $\theta = 0$. Denoting $k_1 = \epsilon = -k_2$, the planner’s maximand simplifies

to $(4B - A)\epsilon^2$. Thus, it is optimal to locate both vessels at $\theta = 0$ if $4B < A$ ($\tau < 0.25$) and to locate them as far apart as is feasible if the inequality is reversed. If we impose symmetry, but allow vessels to randomize their location decisions, then we can represent their decision rule using the distribution $k_i \sim iid(\theta, \sigma^2)$. (Due to the quadratic structure, only the first two moments affect the payoff.) With this symmetric decision rule, the expectation of the payoff in Equation 5 is $(2B - A)\sigma^2$. AP07 impose the restriction $2B < A$ ($\tau < 0.5$) to guarantee concavity of the planner’s problem. In summary, vessels that are required to use symmetric strategies in the team problem should not randomize if and only if $\tau < 0.5$. In the team problem, the restriction to using symmetric strategies bites for $0.25 < \tau$.

Our research question does not involve the team problem. Therefore, we can use the weaker restriction $\tau < 1$, which guarantees concavity of the agent’s problem in the non-cooperative setting, instead of the restriction $\tau < 0.5$ needed for concavity of the planner’s problem. This explains why we can consider higher levels of congestion, those corresponding to $0.5 \leq \tau < 1$. AP07 obtain either necessary or sufficient conditions (but not both) to sign the welfare effect of more precise public or private information for $0 < \tau < 0.5$ (to the left of the vertical line in Figure 1). In contrast, our ranking criteria is both necessary and sufficient and it holds for $\tau < 1$. For $\tau < 0$, where vessels benefit by being close to other vessels, we reproduce AP07’s results. Appendix B provides details.

4 Data

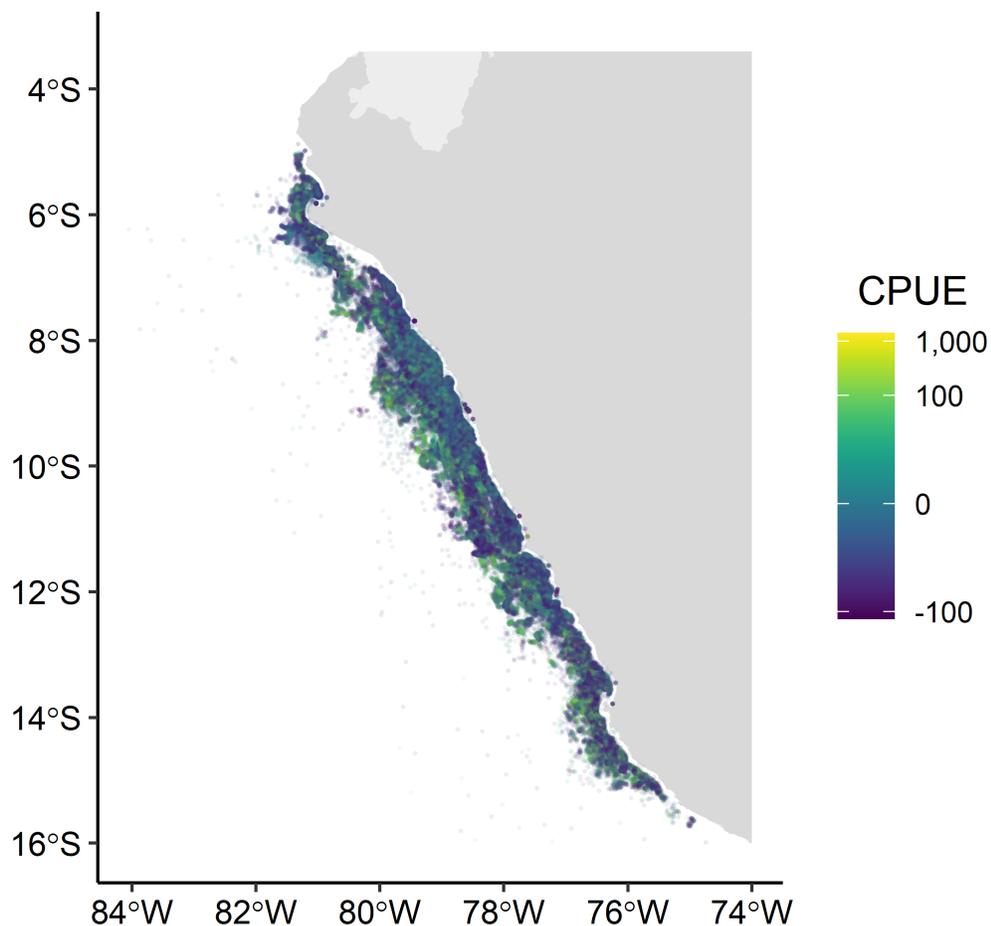
The value of more precise public or private information depends on the relative precision of the two types of information, r , the correlation between the two signals, ρ , and on the severity of congestion relative to the importance of being close to where fish are densest, τ . Here we describe the data used to estimate these three parameters.

We observe the catch of all 806 industrial vessels in Peru’s North-Central anchoveta fishery every time they “set” their net in the water. The unit of observation in this “electronic logbook” data is a set, a vessel-level fishing operation, which occurs at a specific time, longitude, and latitude. Our electronic logbook data spans the six fishing seasons of 2017, 2018, and 2019, containing 247,024 sets. We drop 97 sets that occur on fishing trips that last longer than two weeks (because the median trip lasts 17 hours and the 99th percentile trip lasts 3 days), as well as 7 sets where tons caught is more than 10 times the 99th percentile of tons caught per set by that vessel. Our electronic logbook data therefore contains 246,920 sets.

We construct a measure of Catch Per Unit Effort (CPUE) using vessel characteristics. The simplest measure of CPUE, tons per set, is one measure used by Peru’s marine science agency (IMARPE, 2017). We adjust tons per set by vessel characteristics to account for the

fact that sets by larger and more powerful vessels require more energy than sets by smaller and less powerful vessels. We do so by regressing tons per set on the length (in meters), engine horsepower, and gross tonnage of each vessel.¹² The residuals from this regression are our preferred measure of CPUE because they condition catch on effort. Figure 2 plots CPUE by location. A robustness check that repeats our analysis using a simpler measure of CPUE (tons per set minus vessel-level average tons per set) produces similar parameter estimates.

Figure 2: Peruvian electronic logbook data, 2017 to 2019

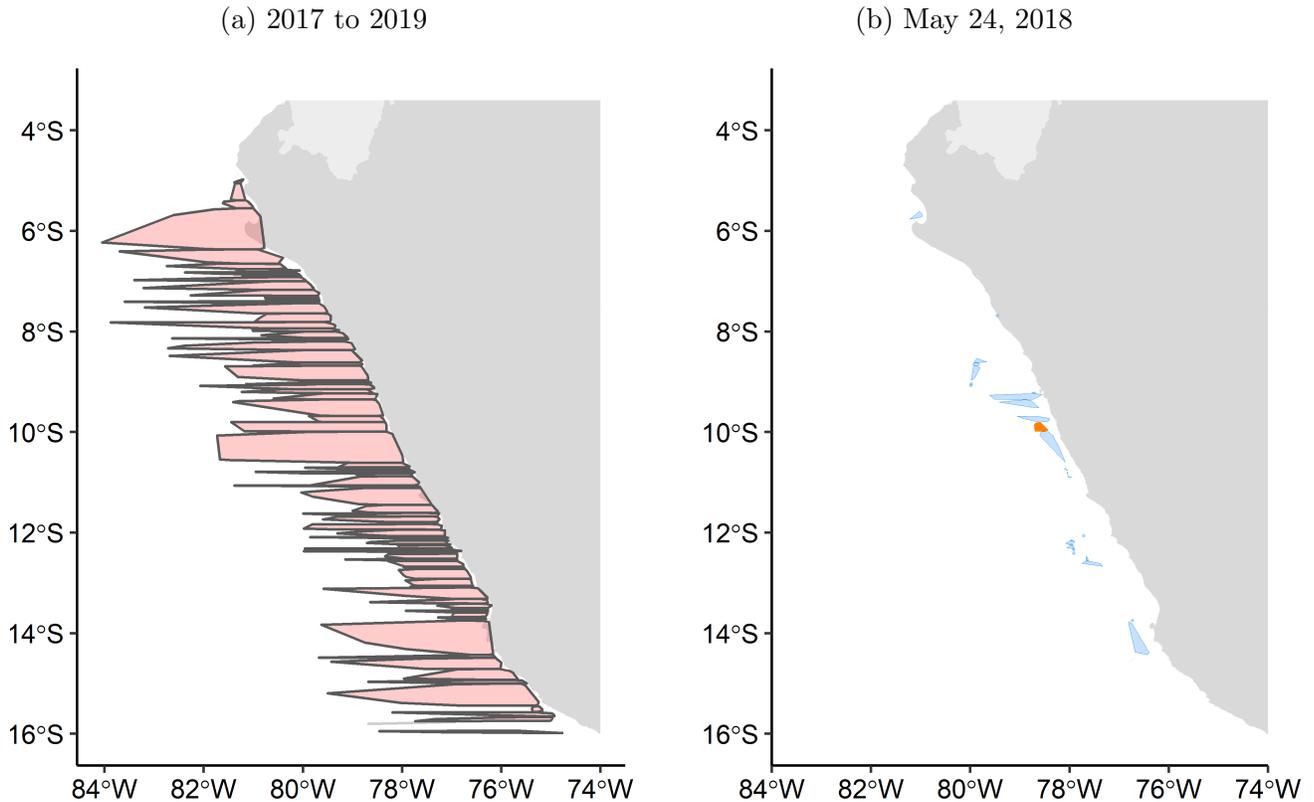


Notes: Each point is a set (vessel-level fishing operation). The color of each point is the catch per unit effort (CPUE) of that set, which we calculate by adjusting tons caught by vessel characteristics. There are 246,920 sets reported by 806 unique vessels in the electronic logbook data. All vessels are prohibited from fishing within 5 nautical miles (9.3 km) of the coast. Peru is dark grey and Ecuador is light grey.

Each set in the electronic logbook data occurs in 1 of 90 fishing zones, which are defined

¹²Gross tonnage is a non-linear measure of a vessel’s internal volume.

Figure 3: Fishing zone polygons



Notes: (a) The regulator defines 90 fishing zones (red). Our electronic logbook data contains the fishing zone each set occurs in. Here we plot the convex hull of each fishing zone as defined by the sets that occur in that fishing zone during our three years of data. (b) In our analysis, we create zone boundaries based on the sets that occur each day. The orange polygon is the convex hull around the sets in the zone with the highest average CPUE on May 24, 2018. The orange polygon is the global best location on this day.

by the regulator and represent distinct, ecologically meaningful fishing grounds. Figure 3(a) shows the convex hull of each fishing zone as defined by the sets that occur in that zone during our three years of data. We identify the ideal location based on each zone's average CPUE each day. We define zones' polygons each day based on the sets that occurred that day; Figure 3(b) displays an example. The polygon boundaries in Figure 3(a) do not affect our estimate of the distance between a vessel and the best location.

The zone-level average CPUE uses multiple sets, providing a less noisy measure of the ideal location, compared to using the single set with the highest CPUE. Since daily data on the actual density of anchoveta do not exist, we use CPUE as its proxy, a standard practice in fishery economics and fishery science (Lynham et al., 2020; Medoff, Lynham, & Raynor,

2022; Nieto et al., 2017; Zhang & Smith, 2011). Using CPUE as a proxy for anchoveta density assumes that catch conditional on effort is higher in places with greater biomass.

Vessels' public information consists of oceanic variables measured daily by NASA and NOAA satellites: chlorophyll, sea surface temperature (SST), SST anomaly, and sea surface salinity. These variables predict anchoveta abundance (Castillo et al., 2019; Silva et al., 2016).¹³ We obtain eight variables in total because NASA and NOAA have multiple satellites. Satellites obtain different measurements at different times because they have different orbital paths. Moreover, some satellites image only part of the earth each day. By using multiple measurements of SST, for example, we obtain a more complete measurement of daily SST. We use chlorophyll, daytime SST, and nighttime SST from VIIRS aboard the SNPP satellite; chlorophyll, daytime SST, and nighttime SST from MODIS aboard the Aqua satellite; sea surface salinity from the SMAP satellite; and SST anomaly from NOAA via Instituto Humboldt, a Peruvian NGO.¹⁴ The resolution of chlorophyll, daytime SST, and nighttime SST data is 4 km (for both VIIRS and MODIS), while the resolution of sea surface salinity and SST anomaly is 25 km. Firms monitor these data and communicate them to their fishers.

Our measure of private information is yesterday's CPUE by all vessels belonging to the same firm.¹⁵ We assume that all vessels within a firm know each other's lagged CPUE, because firms share catch information amongst their vessels (Section 2). This assumption, which affects our estimation of the distributional parameters but not of τ , implies that all vessels within a firm have the same private information. Because agents in our model are vessels, not firms, our definition of private information violates our model assumption that vessels' private signals are uncorrelated. However, each firm comprises a small part of the market, so our model assumption that private signals are uncorrelated with each other, conditional on the state of nature and the public signal, holds approximately.

We could have avoided this disconnect in two ways. First, we could assume that vessels know only their own lagged CPUE, but this would ignore the reality that boats in the same firm share information. Second, we could identify firms rather than vessels as the agents in our model, but this change would make the model less accurate; although firms provide information on suggested fishing locations, our fieldwork indicates that captains of individual vessels ultimately decide where to fish. In addition, our model assumes that agents are homogeneous. This assumption is approximately true when agents are vessels (after controlling

¹³Dissolved oxygen also predicts anchoveta abundance, but we exclude it from our analysis because it is only available from NOAA at the monthly level. Excluding dissolved oxygen should not meaningfully alter our estimates because it is highly correlated with chlorophyll and sea surface temperature (Kim et al., 2020).

¹⁴We use the previous day's value for both nighttime SST variables since a fisher deciding where to fish today only knows last night's SST value.

¹⁵We do not include as private information the locations of other fishing vessels because in our model vessels choose where to fish simultaneously.

for vessel characteristics), but it would be less accurate if agents were firms.

5 Estimation of the distributional parameters, r and ρ

Here we estimate r , the relative precision of public and private signals about the day’s location with the highest stock density, and ρ , the correlation between these signals. The unobserved state of nature in our analytic model is θ , the location with the highest stock density. The empirical application generalizes the analytic specification by assuming that there might be several “local best zones” instead of a single global best location. The generalization takes into account the fishery’s large area and transportation costs, which prevent many boats from traveling to the global best zone in a day. It also recognizes that the stock distribution may be patchy, a feature of many fisheries (Sanchirico & Wilen, 2002). That is, instead of assuming that stock density falls monotonically with the distance to the global best zone, we merely assume that the density falls monotonically with the distance from the local best zone. It is as if the single fishery is broken into several smaller fisheries. The data identifies these smaller fisheries, whose boundaries change daily.

The unit of observation, k_{ifzt} , is a set i , performed by a vessel belonging to firm f , and occurring in zone z on day t . Our measure of the distance between a vessel and its ideal location takes into account that the fishery is too large for a vessel to traverse during a single trip.¹⁶ We suppose that each set could have occurred in a group of feasible zones, which we define as those within 126 km of the zone where the set actually occurred. We chose this radius because it is the median distance among fishing trips calculated in Joo et al. (2015).¹⁷ For a set in a given zone, we define its “local best zone” as the zone with the highest average CPUE within 126 km. We denote this local best zone as $\hat{\theta}_{izt}$. The dependent variable in our regressions is $\left(\|k_{ifzt} - \hat{\theta}_{izt}\|\right)^2$, the squared distance between the zone set i occurs in and the local best zone. Sets in the same zone therefore have the same value of the dependent variable. Of the 33 zones with sets on May 24, 2018 depicted in Figure 3(b), 11 are within 126 km of the orange polygon, the zone with the globally highest average CPUE. The other 22 zones have different locally best zones.

In our empirical model, the variable of interest for a boat is the local best zone. Our regressions measure the public and private signals’ ability to predict the deviation between sets and this local best zone. To the extent that public and private signals predict this

¹⁶The length of the fishery exceeds 1,500 km. At a typical cruising speed of 20 km per hour, it would take more than 3 days to travel along the coast from the fishery’s southern boundary to its northern boundary (Peraltilla & Bertrand, 2014).

¹⁷Joo et al. (2015) use hourly vessel location data to calculate the distance vessels travel between leaving port and landing their catch at processing plants. They also calculate each trip’s maximum distance from the coast. The median maximum distance from the coast is 25 km.

deviation, they can also provide information about the best location. The variance of this prediction is analogous to the variance of the signal; both measure deviations between the predicted best location and the actual best location.

We estimate σ_y^2 with the residual sum of squares (RSS) from the following linear regression:

$$\left(\|k_{ifzt} - \hat{\theta}_{izt}\|\right)^2 = \beta X_{ifzt} + \epsilon_{ifzt} \quad (6)$$

where $i = \text{set}$, $f = \text{firm}$, $z = \text{zone}$, $t = \text{calendar date}$, β are regression coefficients, X_{ifzt} is a matrix of public signals, and ϵ_{ifzt} is the error term.

The matrix X_{ifzt} contains 12,880 public signal predictor variables. For each set and oceanic variable, we record the value of the oceanic variable at the nearest location to set i . (Recall that some satellites only image part of the earth each day.) We also interpolate each oceanic variable over the entire fishery each day using inverse distance weighting. Then we extract the interpolated value of the oceanic variable for each set. We thus create two variables from each oceanic variable – the nearest value and the interpolated value – which gives us 16 oceanic variables (8 times 2). Then we calculate three new variables for each of these 16 original variables. We calculate the average value for set i 's best local zone, the difference between set i 's value and the average value in its best local zone, and the difference between set i 's value and the average value that day (in any zone). After this procedure we have 64 predictor variables (16 plus 3 times 16). Next, we add 16 more predictor variables to reach a total of 80. First, we record the distance between set i and the nearest location at which each of the 8 original, non-interpolated oceanic variables were measured, since closer measurements may be more predictive. Second, we calculate the average of these distances in set i 's best local zone. We calculate the square of each these 80 predictor variables, increasing the number to 160.¹⁸ Finally, we interact all 160 of these variables with each other, obtaining a matrix X_{ifzt} with 12,880 variables (160 choose 2, plus the original 160 variables).¹⁹

To avoid overfitting Equation 6, we randomly divide the data at the day-level into a training set and a test set. We divide the data at the day-level because we want to assess how predictive public information would be of squared distance to the best zone on a given day if we did not observe the actual best zone that day. We randomly assign 75% of days to the training set and the remaining 25% of days to the test set. Of the 427 days in our data, we assign 321 days to the training set, which results in a training set with 183,055 observations.²⁰

We estimate Equation 6 with lasso, a type of penalized linear regression. We use 10-

¹⁸Including additional higher order terms causes our estimation procedure to exceed the RAM of our server.

¹⁹If an oceanic variable has no measurements inside Peru's anchoveta fishery on a given day, sets on this day have missing values for all variables that are a function of that variable.

²⁰This random assignment is irrespective of which fishing season a day occurs in, so the training and test sets are temporally balanced within fishing seasons.

fold cross-validation in choosing the optimal penalty term.²¹ In each iteration of our cross-validation procedure, we impute missing predictor values with that variable’s mean value in the analysis set, remove zero variance predictors (if any), and standardize all predictor variables. We define the optimal penalty term as in Breiman et al. (1984) as the penalty that returns the simplest model (fewest predictor variables) that is within one standard error of the numerically optimal penalty (the smaller penalty that results in a more complex model and the highest in-sample R^2 among all penalties).

We then use our optimal penalty term, 599, to fit Equation 6 on the entire training set. This lasso regression retains 20 predictor variables with non-zero coefficients. We use the regression coefficients to predict squared distance to the best zone in the test set.

The RSS in the test set—the squared sum of the difference between the actual and the predicted squared distance to the best zone—is 8.55654e+11, compared to a total sum of squares for the test set of 1.10697e+12. Our test R^2 is therefore 0.227.²² Fishing opportunities are highly variable within a day and over small spatial scales, and our public signals are coarse and incomplete (e.g., missing a satellite image for a variable in a given location-day). Our public signals are nonetheless somewhat informative. We could obtain a much higher R^2 if we predicted squared distance to the best zone in-sample (without evaluating our predictions in a test set), but doing so would overestimate the predictive power of the public signals.

To estimate σ_x^2 , we replace X_{ifzt} in Equation 6 with a matrix of private signals and then repeat the same procedure we used to estimate σ_y^2 .

We create our matrix of private signals as follows. First, we record the CPUE of the nearest set by vessels in the same firm in the previous day.²³ Second, for each firm-day, we interpolate the previous day CPUE of sets by the firm’s vessels using inverse distance weighting. Then for each set we extract lagged firm-level CPUE at that set’s location. This procedure yields two variables from firms’ lagged CPUE: the nearest value and the interpolated value. Our third and fourth variables are firm-day demeaned nearest and interpolated lag CPUE. Fifth, we record the distance between set i and the nearest set by vessels in the same firm on the previous day. Sixth, we calculate the average interpolated firm-level lag CPUE in set i ’s best local zone (within 126 km of the zone set i occurs in). Our seventh and eighth variables

²¹We train the model using the analysis set, consisting of 90% of days in the training set; we then assess accuracy by predicting squared distance to the best zone for the assessment set, consisting of the remaining 10% of days in the training set. We repeat the procedure 10 times, using different splits of the training data into analysis and assessment sets each time. We save the prediction accuracy—the R^2 in the assessment sets—at different values of the penalty term.

²²We do not include vessel fixed effects in our matrix of public signals or in our matrix of private signals because vessel indicator variables are neither public nor private signals. We do not include vessel fixed effects in our matrix of private signals for the same reason.

²³For each set i by a vessel in firm f on day t , we record the CPUE of the nearest set among the sets that occur on day $t - 1$ by vessels in firm f .

Table 1: Public and private signal parameter estimates

σ_y^2	σ_x^2	r	ρ	δ
8.55654e+11	1.07979e+12	0.890	0.266	0.579

are the differences between this sixth variable and set i 's nearest and interpolated firm-level lag CPUE. Our ninth and tenth variables are the squared distances between the zone set i occurs in and the *predicted* best local zones, which are the zones within 126 km that have the highest average interpolated and nearest firm-level lag CPUE.²⁴ Our eleventh predictor variable is the number of zones that vessels belonging to firm f fished in on day $t - 1$. We add six more predictor variables by calculating third-order polynomials for the three (non-differenced) CPUE variables (the first, second, and sixth variables).²⁵ Finally, we interact all 17 of our private predictor variables to obtain a matrix of private signals with 153 variables (17 choose 2, plus the original 17 variables). If firm f had no sets on day $t - 1$, then we record all predictor variables as missing, except for the number of zones that vessels belonging to firm f fished in on day t .

Estimating Equation 6 with our optimal shrinkage penalty (140) retains 8 predictor variables with non-zero coefficients. Our estimate of σ_x^2 is 1.07979e+12, compared to the same total sum of squares as above (1.10697e+12). The private signals are less informative than the public signals; the private information R^2 in the test set is 0.025.

Our estimate of r is therefore 0.890 ($\sqrt{8.55654e+11}$ divided by $\sqrt{1.07979e+12}$). We estimate ρ , the correlation between the error in the public signals and the error in the private signals, as 0.266. Table 1 displays our parameter estimates.

Using our estimates of r and ρ in the formula for δ (Equation 1), we obtain the estimate $\delta = 0.579$; the BLUE of θ assigns 0.579 weight to the public signal and 0.421 weight to the private signal. This intermediate value of δ means that the welfare effect of increased precision of public or private information depends on τ .

Other specifications and robustness checks for estimating r and ρ

We considered several other dependent variables for our regressions, but we believe that squared distance to the best zone matches our model most closely. For example, we could choose CPUE of set i as our dependent variable, but predicting CPUE (which is not itself a location) is less similar to predicting the best location than our preferred dependent variable.

²⁴By contrast, the best local zone for our sixth predictor variable is the one within 126 km that has the highest actual CPUE among all sets on day t .

²⁵We can calculate third-order instead of second-order polynomials because unlike in our estimation of the precision of the public signal, we are not constrained by our server's RAM.

Alternatively, we could estimate r and ρ with a two-step procedure in which our dependent variable equals 1 if the set occurs in the best local zone and equals 0 otherwise. In the first step, we would regress this variable on public or private signals and then use the resulting predicted values to obtain the predicted best local zone(s) each day. In the second step, we would calculate the precision of the public or private signal as the average squared distance between the predicted best and actual best local zone(s), where the predicted best local zone(s) comes from either the public signal regression or the private signal regression. This procedure is similar to our preferred dependent variable, but we do not implement it because its two steps will degrade the precision of the signals relative to our preferred one-step procedure.

We repeat our estimation procedure with an alternative measure of CPUE: tons per set minus vessel-level average tons per set. We obtain similar results compared to our preferred specification of CPUE. In this robustness check, we estimate $\sigma_y^2 = 8.82187e+11$, $\sigma_x^2 = 1.07862e+12$, $r = 0.904$, $\rho = 0.210$, and $\delta = 0.563$.

As a further robustness check, we repeat our estimation of the distributional parameters with squared distance to the globally best zone as the dependent variable. This specification matches our model but represents reality less well because it ignores travel costs and the large area of the fishery, and the possibility that the stock density is patchy. In the global best zone specification, the dependent variable for all sets on May 24, 2018 is the squared distance between the zone the sets occur in and the orange polygon in Figure 3(b).

Our global best zone robustness check follows the same procedure as the local best zone specification. The only difference is there is now a single best zone each day (the one with the highest average CPUE). Sets on the same day have the same best zone, even if some of the zones those sets occur in are more than 126 km away. In terms of Equation 6, the estimated ideal location is $\hat{\theta}_t$ as opposed to $\hat{\theta}_{izt}$. The median and average distances to the best zone in this specification are 247 km and 309 km. We estimate $\sigma_y^2 = 2.69459e+15$, which corresponds to an R^2 in the test set of 0.271; $\sigma_x^2 = 3.35710e+15$, which corresponds to an R^2 in the test set of 0.092; $r = 0.896$; $\rho = 0.766$; and $\delta = 0.729$. Note that σ_y^2 and σ_x^2 are much larger in the global best zone specification because the dependent variable is larger; the test R^2 are similar to the results of the local best zone specification. Though the relative precision of the two signals (r) is similar in the global best zone specification, there is a much higher degree of correlation between the errors in the signals (ρ). For this reason the BLUE of θ , the parameter δ , places more weight on the public signal in the global best zone specification than in the local best zone specification.

6 Estimation of τ

The parameter τ measures the benefit of dispersion—fewer vessels fishing nearby—relative to the cost of deviating from the ideal location. Estimation of τ relies on the following:

Claim 1. *A vessel’s expected payoff decreases with its distance from the ideal point θ .*

Verification of this claim is almost immediate from the concavity of the vessels’ maximand, guaranteed by the second order condition $B - A < 0$; see Appendix B. As a vessel moves closer to the ideal location, θ , the density of other vessels, and therefore congestion costs, also increase. However, the second order condition guarantees that in expectation a vessel does better by being close to the ideal location. We use the same local best zone specification from Section 5 to estimate the ideal location $\hat{\theta}_{izt}$ for each set.

Treating fishing costs per set as fixed, we use harvest as a measure of a vessel’s payoff. Claim 1 then implies that aggregate harvest per unit of effort at location j , denoted $CPUE_j$, is higher (in expectation) the closer location j is to the ideal location. Given the estimate of the ideal location, $\hat{\theta}_{izt}$, and data on the location of set i , k_{izt} , we have a proxy for the set’s distance from the ideal location, $\left(\|k_{izt} - \hat{\theta}_{izt}\|\right)^2$, where $\|\cdot\|$ is the Euclidean distance. This squared distance to the best location is identical to the dependent variable in our estimation of the distributional parameters (Equation 6), except we omit the firm f subscript because it is not relevant for our estimation of τ .

The first term in Equation 2 equals *expected* dispersion, with actual dispersion equal to $\frac{B}{2} \int (k_{izt} - k_{jzt})^2 dj$. The inverse measure of congestion facing set i is

$$D_{izt} = \frac{1}{N_{it}} \sum_j (k_{izt} - k_{jzt})^2, \quad (7)$$

where N_{it} is the number of sets on day t within 126 km of set i . The median set has 311 other sets within 126 km of it on the day it occurs. 126 km is the same radius we used in Section 5, since sets farther than 126 km might not affect CPUE of set i . We refer to D_{izt} as dispersion because it measures the distance between set i and all other sets on that day within 126 km.²⁶ We divide $\sum_j (k_{izt} - k_{jzt})^2$ by the number of sets on day t within 126 km because our model normalizes the measure of vessels to 1 (Appendix A). We standardize dispersion and squared distance to the best location (subtracting their mean and dividing by their standard deviation) so that their regression coefficients are directly comparable to each other.

We estimate Equation 2 with the following ordinary least squares regression:

²⁶We calculate dispersion for each set; sets in the same zone have (slightly) different dispersion values.

Table 2: Estimation of τ

	CPUE	
	(1)	(2)
Dispersion	1.628 (0.979)	0.157 (0.915)
DistToBest ²	-4.151 (0.789)	-4.139 (0.708)
τ	0.392 (0.236)	0.038 (0.221)
Controls		X
Adj. R ²	0.007	0.07
N	246,920	246,920

Notes: The dependent variable is catch per unit effort (CPUE). Mean CPUE is 0 by construction; the standard deviation is 51.8. We estimate τ as the Dispersion coefficient divided by the negative of the DistToBest² coefficient. We estimate the standard error of τ with the delta method. Our post-double-selection procedure retains 292 variables that predict CPUE or Dispersion (Column 2). Standard errors are two-way clustered at the date and zone level.

$$CPUE_{izt} = \alpha + BD_{izt} + A \left(\left\| k_{izt} - \hat{\theta}_{izt} \right\| \right)^2 + \epsilon_{izt} \quad (8)$$

where α , A , and B are coefficients and ϵ_{izt} is the error term. We two-way cluster standard errors at the level of date and zone. We estimate τ as B divided by $-A$. The negative of A measures the benefit of being closer to the best location.

We expect that the effect of dispersion on CPUE is positive ($B > 0$). All else equal, more dispersion should increase CPUE because there are fewer sets by other vessels near set i (less physical congestion from nearby vessels and less depletion of the local anchoveta population). We expect $A < 0$; as set i occurs farther from the best location, CPUE should decrease.

Column 1 of Table 2 displays our estimates of B , A , and τ corresponding to Equation 8. Our estimates of B and A have the expected signs. A 1 standard deviation increase in dispersion increases CPUE by 1.628 tons, while a 1 standard deviation increase in squared distance to the best location decreases CPUE by 4.151 tons. We estimate $\tau = 0.392$.

While the negative relationship between CPUE and squared distance to the best location is partly mechanical (since the best location is the local zone with the highest average CPUE), our estimate of the relationship between CPUE and dispersion could be confounded by the fact that we do not observe the biomass (abundance) of anchoveta at every location-time. Biomass likely increases CPUE and decreases dispersion: location-times with high biomass

will likely have both higher CPUE and lower dispersion (vessels clustered close to each other). Our estimate of B in Column 1 of Table 2 may therefore be biased downward.

We attempt to alleviate this omitted variable bias by implementing the post-double-selection method of Belloni, Chernozhukov, and Hansen (2014). Since omitted variables bias is caused by variables that affect both the dependent variable (CPUE) and the independent variable of interest (dispersion), post-double-selection identifies the predictors of either variable, and controls for these predictors in a third regression of the dependent variable on the independent variable of interest. First, we run a lasso regression with CPUE as the dependent variable (omitting standardized dispersion as a predictor variable). Second, we run a lasso regression with standardized dispersion as the dependent variable. Finally, we regress (via ordinary least squares) CPUE on standardized dispersion and include as controls all of the variables that lasso retains as predictors of CPUE or standardized dispersion. Standardized squared distance to the best zone is one of the control variables in this final regression because it was retained as a significant predictor in both lasso regressions.

We include as predictors all of the public signal variables (including higher-order terms) that we used to estimate σ_y^2 in Section 5. We also construct the following measures of local stock depletion, which is negatively correlated with current biomass. For a given set i in zone z on day t , we calculate eight lags of tons caught in zone z : tons caught in zone z on day t by other vessels prior to the start of set i ,²⁷ tons caught in zone z on day $t-1$, tons caught in zone z on day $t-2$, and so on until tons caught in zone z on day $t-7$. For each of these 8 variables, we calculate squared terms and indicators for 0 tons caught, yielding a matrix of 24 predictor variables. We add these 24 variables to the 160 public signal variables, and then interact all 184 variables with each other. We include calendar date indicator variables and zone indicator variables as predictors as well, but we do not include them in the interaction due to computational constraints. Zone indicator variables partially capture differences in the cost of fishing across locations. We thus obtain a matrix of 17,247 predictor variables. Of these, the two lasso regressions retain 292 unique predictor variables with non-zero coefficients.

Column 2 of Table 2 displays our estimates when we control for these 292 predictor variables. Instead of becoming larger, which would be consistent with a downward biased coefficient in Column 1, the coefficient on dispersion becomes smaller. This result provides evidence against the concern that omitting biomass from Equation 8 biases the dispersion coefficient downward. Thus, we use the Column 1 estimate of τ in our welfare calculations in Section 7 because that estimate derives from an equation that matches our model more closely (i.e., Equation 2 does not contain control variables).

As a robustness check, we re-estimate Equation 8 with our simpler measure of CPUE,

²⁷A set takes about 90 minutes to complete.

vessel-demeaned tons per set. We obtain similar estimates compared to our preferred specification of CPUE. In this robustness check, we estimate $B = 1.585$, $A = -4.223$, and $\tau = 0.375$.

As an additional robustness check, we re-estimate Equation 8 with a global definition of dispersion and squared distance to the best zone. In this global best zone specification, we measure dispersion of all sets that occur on day t , regardless of their distance to set i . Squared distance to the best zone is the same as the dependent variable of our global best zone robustness check in Section 5. We estimate $B = 2.372$, $A = -4.227$, and $\tau = 0.561$. While this τ is larger than the τ from our preferred local specification, they are not statistically different (t-statistic of difference is 0.54).

Given the estimates from our preferred specification, $\tau = 0.392$ and $\delta = 0.579$, and under the assumption of zero correlation between public and private signals, the Peruvian anchoveta fishery would lie in Region D of Figure 1, where increased precision of both types of information would increase welfare. If we used our Column 2 estimate of $\tau = 0.038$, the fishery would also lie in Region D. However, the non-zero correlation between the public and private signals requires us to use numerical methods to evaluate the welfare effects of increased precision of public and private information.

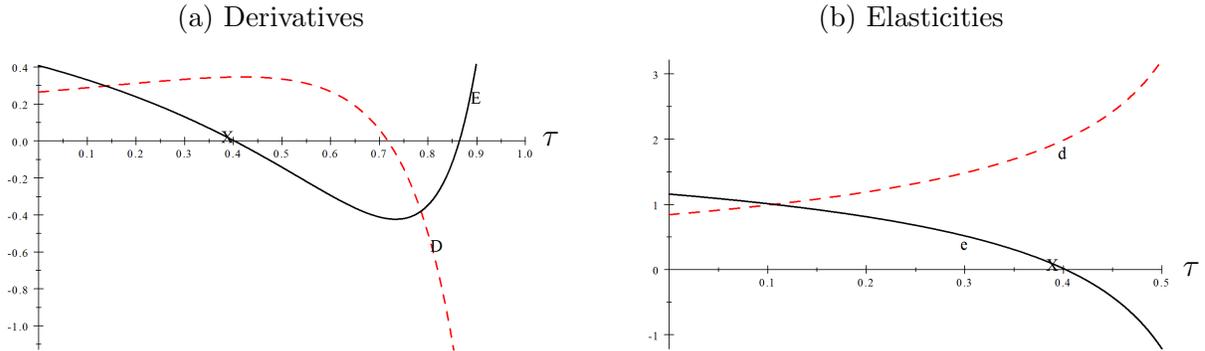
7 Welfare effects of better information

Appendix D provides the derivatives of welfare with respect to the precision of public and private information, as functions of (τ, ρ, r) ; using these formulae, we obtain the elasticities of welfare with respect to the precision of both private and public information. These functions are highly nonlinear, precluding a useful analysis of the entire parameter space, even with numerical methods. We therefore fix (ρ, r) at their point estimates $(0.266, 0.89)$ and show how the welfare effects of more precise information depend on the congestion parameter, τ .

Figure 4(a) shows the derivatives of welfare (scaled by $A\sigma_y$) with respect to the precision of private information (dashed graph) and public information (solid graph). Both derivatives are positive at the point estimate $\tau = 0.392$ (identified by the X in the figure), but the derivative with respect to the precision of public information is close to zero. For $\tau > 0.401$ (a 2% increase over our point estimate, or 4% of the standard error), more precise public information lowers welfare. Our results therefore suggest that there is a medium likelihood that more precise public information lowers welfare.

For our point estimate $(\rho, r) = (0.266, 0.89)$, the derivative of welfare with respect to the precision of public information has two zeros, $\tau = 0.401$ and $\tau = 0.86$. For $\tau > 0.86$ – two standard errors higher than our point estimate of τ – more precise public information raises welfare. Proposition 1 shows that for $\rho = 0$ more precise public information lowers welfare if

Figure 4: Welfare effects of more precise public or private information



Notes: (a) The left panel shows the derivatives of welfare with respect to the precision of private information (dashed graph labeled D) and public information (solid graph labeled E), as functions of τ . Both functions are scaled by $A\sigma_y$. (b) The right panel graphs the elasticities of welfare with respect to the precision of private and public information (dashed graph labeled d and the solid graph labeled e , respectively). In both panels, X identifies our point estimate $\tau = 0.392$.

and only if τ exceeds a critical value; there, the derivative of welfare has a single zero. Figure 4(a) illustrates the complex welfare effects arising from $\rho \neq 0$.

Figure 4(b) graphs the elasticities of welfare with respect to increased precision of private information (dashed curve) and public information (solid curve). Here we consider a range of τ over which welfare is not close to zero; as welfare approaches zero, both elasticities approach infinity, so their graphs become uninformative.²⁸ At the point estimate $\tau = 0.392$, the elasticity for private information is 1.92 and the elasticity for public information is 0.08; the elasticities differ by a factor of 24.

For our point estimates, the BLUE weight on the public signal is $\delta = 0.579$, and the Nash weight is $\gamma^{NE} = 0.40$. Congestion leads to a 30% reduction in the weight on the public signal. Given our point estimates of (r, τ) , our estimate $\rho = 0.266$ reduces the BLUE weight by 4% and it reduces the Nash weight by 6%. For our point estimate of r , we saw that with $\rho = 0.266$, more precise public information increases welfare if and only if $\tau < 0.401$ or $\tau > 0.86$. More precise private information increases welfare if and only if $\tau < 0.72$ (1.4 standard errors above our point estimate). In contrast, for $r = 0.89$ and $\rho = 0$ (i.e., ignoring the correlation between signals), more precise public information increases welfare if and only if $\tau < 0.46$, and more precise private information increases welfare if and only if $\tau < 0.69$.

²⁸The model is designed to examine the effect of congestion on the value of more precise information, not the effect of congestion on the *level* of welfare. Therefore, the graph of welfare with respect to τ conveys little information, except for explaining why we do not graph the elasticities over the entire range of τ .

Thus, failing to properly account for the correlation between signals, increases the (lower) critical bound, below which more precise public information raises welfare by 14%, and it lowers the critical value for private information by 4%.

In summary, our estimates strongly suggest that more precise private information about stock location increases fishers' equilibrium welfare. Our point estimates suggest that more precise public information also increases welfare, but by a negligible amount; given the estimates' uncertainty, it is plausible that more precise public information lowers equilibrium welfare. Had we (incorrectly) ignored the correlation between public and private signals, we would have assessed a substantially higher critical τ , thereby overstating our confidence that more precise public information raises welfare. Ignoring the correlation would not have altered our qualitative conclusions about the value of more precise private information.

8 Conclusion

A rich theoretical literature recognizes the differing equilibrium effects of more precise public and private information, and the possibility that either might lower welfare by changing equilibrium behavior. A rich empirical literature estimates the value of information under the assumption that if information is not useful, agents can ignore it without cost: information always has (weakly) positive value. Thus, the theory on the ambiguous value of information has rarely been implemented empirically; the empirical literature on the value of information has generally ignored the insights from theory.

To obtain clear results, the existing theory assumes that public and private signals are uncorrelated. However, in an empirical setting, these signals are likely to be correlated. To take the model to the data, we extend it to include correlated signals. The increased complexity makes it difficult to obtain general insights, but the comparative statics are easily evaluated given three estimated parameters: the relative precision of private versus public information; the correlation between public and private signals; and the importance of congestion relative to proximity to the ideal location.

We estimate the model using high resolution data from the world's largest fishery by catch volume, the Peruvian anchoveta fishery. We find that congestion is about 40% as important to profits as is fishing close to the ideal location; the correlation between signals is 0.27; and private signals are about 10% less precise than public signals. With these estimates, our model implies that more precise private information is valuable, with an elasticity of welfare equal to 2. Our point estimate for the value of more precise public information is positive but small, with an elasticity of 0.1. Given this small magnitude and the uncertainty in the parameter estimates, it is plausible that more precise public information lowers welfare. Had we ignored

the correlation between signals, we would have substantially over-estimated the probability that more precise public information raises welfare, but would not have qualitatively altered our conclusion that more precise private information is beneficial.

The policy implication is that anchoveta fishers would benefit from more precise private information about the location of fish stocks. More precise public information might increase welfare, but by a small magnitude; the welfare effect could well be negative. It is important to correctly account for the correlation between public and private signals, particularly in assessing the welfare effect of more precise public information. Fishery managers choose the accessibility of public information and choose which private information becomes public. For example, the Peruvian fisheries ministry could improve the precision of public information by publishing near real-time data on catch by all industrial fishing vessels (Englander, 2022). Our results suggest that making this private information public could lower fisher profits.

These conclusions are in line with theory's recognition that, in the presence of congestion (or a similar negative externality) more precise public information is less likely to be beneficial, compared to more precise private information. Public information encourages agents to all act in the same manner, e.g., all move to the same fishing ground, thereby exacerbating the externality. The particularity of our results (to the context of the Peruvian anchoveta fishery) cautions against attempts to generalize them to other fisheries, or other settings where externalities are important. This particularity also shows the value of combining theory with empirical tools. This combination enables us to address a type of policy question not accessible to an atheoretical approach.

References

- Angeletos, G.-M., & Pavan, A. (2007). Efficient use of information and social value of information. *Econometrica*, *75*, 1103–1142.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.
- Brown, G. (1974). An optimal program for managing common property resources with congestion externalities. *Journal of Political Economy*, *82*, 163–173.
- Carlsson, H., & Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, *61*, 989–1018.
- Castillo, R., Rosa, L. D., Diaz, W. G., Madureira, L., Gutierrez, M., Vásquez, L., & Koppelman, R. (2019). Anchovy distribution off Peru in relation to abiotic parameters: A 32-year time series from 1985 to 2017. *Fisheries Oceanography*, *28*(4), 389–401.
- Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in India. *Econometrica*, *86*(6), 2123–2160.
- Englander, G. (2022). Information and Spillovers from Targeting Policy in Peru’s Anchoveta Fishery. *American Economic Journal: Economic Policy*, Forthcoming.
- FAO. (2018). The State of World Fisheries and Aquaculture 2018 - Meeting the Sustainable Development Goals [Food and Agricultural Organization of the United Nations (FAO)].
- Fréon, P., Sueiro, J. C., Iriarte, F., Evar, O. F. M., Landa, Y., Mittaine, J.-F., & Bouchon, M. (2014). Harvesting for food versus feed: A review of Peruvian fisheries in a global context. *Reviews in Fish Biology and Fisheries*, *24*(1), 381–398.
- Gatewood, J. (1984). Cooperation, competition, and synergy: Information-sharing groups among southeast Alaskan salmon seiners. *American Ethnologist*, *11*, 350–370.
- Hansman, C., Hjort, J., León-Ciliotta, G., & Teachout, M. (2020). Vertical Integration, Supplier Behavior, and Quality Upgrading among Exporters. *Journal of Political Economy*, *128*(9), 3570–3625.
- Huang, L., & Smith, M. (2014). The dynamic efficiency costs of common-pool resource exploitation. *American Economic Review*, *104*, 4071–4103.
- IMARPE. (2017). *Situación del Stock Norte-Centro de la anchoveta peruana (Engraulis ringens) al 01 de Noviembre de 2017 y perspectivas de explotación para la segunda temporada de pesca de 2017*. Instituto del Mar del Perú (IMARPE).
- Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The Quarterly Journal of Economics*, *122*(3), 879–924.
- Joo, R., Salcedo, O., Gutierrez, M., Fablet, R., & Bertrand, S. (2015). Defining fishing spatial strategies from VMS data: Insights from the world’s largest monospecific fishery. *Fisheries Research*, *164*, 223–230.
- Kim, Y. H., Son, S., Kim, H.-C., Kim, B., Park, Y.-G., Nam, J., & Ryu, J. (2020). Application of satellite remote sensing in monitoring dissolved oxygen variabilities: A case study for coastal waters in Korea. *Environment International*, *134*, 105301.

- Kroetz, K., Sanchirico, J. N., Contreras, E. G., Novoa, D. C., Collado, N., & Swiedler, E. W. (2016). *Examination of the Peruvian anchovy Individual Vessel Quota (IVQ) System* (tech. rep. Working Paper 749). Inter-American Development Bank.
- Lynham, J., Nikolaev, A., Raynor, J., Vilela, T., & Villaseñor-Derbez, J. C. (2020). Impact of two of the world’s largest protected areas on longline fishery catch rates. *Nature Communications*, *11*(1), 1–9.
- Marcoul, P. (2020). Fish abundance information and fishery’s aggregate behavior. *JAERE*, *7*(5), 933–964.
- Medoff, S., Lynham, J., & Raynor, J. (2022). Spillover benefits from the world’s largest fully protected MPA. *Science*, *378*(6617), 313–316.
- Morris, S., & Shin, H. (1998). Unique equilibrium in a model of self-fulfilling expectation. *American Economic Review*, *88*, 587–597.
- Nieto, K., Xu, Y., Teo, S. L., McClatchie, S., & Holmes, J. (2017). How important are coastal fronts to albacore tuna (*Thunnus alalunga*) habitat in the Northeast Pacific Ocean? *Progress in Oceanography*, *150*, 62–71.
- Peraltilla, S., & Bertrand, S. (2014). In situ measurements of the speed of Peruvian anchovy schools. *Fisheries Research*, *149*, 92–94.
- PRODUCE. (2018). Anuario Estadístico Pesquero y Acuicola 2017 [Ministerio de la Producción (PRODUCE)].
- Sallee, J. M. (2014). Rational inattention and energy efficiency. *The Journal of Law and Economics*, *57*(3), 781–820.
- Sanchirico, J., & Wilen, J. (2002). Bioeconomics of spatial exploitation in a patchy environment. In *Fisheries economics* (pp. 177–198). Routledge.
- Sanz, N., & Diop, S. (2021). Endogenous catch per unit effort and congestion externalities between vessels in a search-matching model: Evidence from the French Guiana shrimp fishery. *Bulletin of Economic Research*, *11*, 1–16.
- Silva, C., Andrade, I., Yáñez, E., Hormazabal, S., Barbieri, M. Á., Aranis, A., & Böhm, G. (2016). Predicting habitat suitability and geographic distribution of anchovy (*Engraulis ringens*) due to climate change in the coastal areas off Chile. *Progress in Oceanography*, *146*, 159–174.
- Vives, X. (1984). Duopoly information equilibrium: Cournot and Bertrand. *Journal of Economic Theory*, *34*, 71–94.
- Welch, H., Clavelle, T., White, T. D., Cimino, M. A., Van Osdel, J., Hochberg, T., Kroodsma, D., & Hazen, E. L. (2022). Hot spots of unseen fishing vessels. *Science Advances*, *8*(44), eabq2109.
- Zhang, J., & Smith, M. (2011). Estimation of a generalized fishery model. *Review of Economics and Statistics*, *93*, 690–699.

Online Appendices

A Derivations and Proofs

The derivation of Equation 3 begins by noting that vessel i 's distance from arbitrary j , with $i \neq j$, is

$$\begin{aligned} k_i - k_j &= \gamma y + (1 - \gamma) x_i - (\eta y + (1 - \eta) x_j) \\ &= \gamma (\theta + \varepsilon_y) + (1 - \gamma) (\theta + \varepsilon_{x_i}) - (\eta (\theta + \varepsilon_y) + (1 - \eta) (\theta + \varepsilon_{x_j})) \\ &= (\eta - 1) \varepsilon_{x_j} + \varepsilon_y (\gamma - \eta) - (\gamma - 1) \varepsilon_{x_i}. \end{aligned}$$

The expectation of the squared distance is

$$\begin{aligned} &\mathbf{E}_{\varepsilon_{x_i}, \varepsilon_{x_j}, \varepsilon_y} \left((\eta - 1) \varepsilon_{x_j} + \varepsilon_y (\gamma - \eta) - (\gamma - 1) \varepsilon_{x_i} \right)^2 \\ &= ((\eta - 1)^2 + (\gamma - 1)^2) \sigma_x^2 + (\gamma - \eta)^2 \sigma_y^2 + 2(\gamma - \eta)(\eta - 1 - (\gamma - 1)) \rho \sigma_x \sigma_y \end{aligned} \quad (9)$$

Using Equation 9 and the assumption of a unit mass of vessels, we can replace the expectation of the integral in Equation 2 with

$$\frac{B}{2} \left(((\eta - 1)^2 + (\gamma - 1)^2) \sigma_x^2 + (\gamma - \eta)^2 \sigma_y^2 - 2(\gamma - \eta)^2 \rho \sigma_x \sigma_y \right).$$

For $B > 0$ vessels face congestion costs, and for $B < 0$ they benefit from the proximity of other vessels.

Vessel i 's deviation from the ideal location is

$$k_i - \theta = \gamma y + (1 - \gamma) x_i - \theta = \gamma \varepsilon_y + (1 - \gamma) \varepsilon_{x_i}$$

The vessel's cost of this deviation is

$$\begin{aligned} \frac{A}{2} \mathbf{E}_{\varepsilon_y, \varepsilon_{x_i}} (k_i - \theta)^2 &= \frac{A}{2} \mathbf{E}_{\varepsilon_y, \varepsilon_{x_i}} (\gamma \varepsilon_y - \gamma \varepsilon_{x_i} + \varepsilon_{x_i})^2 \\ &= \frac{A}{2} (\gamma^2 \sigma_y^2 + (1 - \gamma)^2 \sigma_x^2 + 2(1 - \gamma) \gamma \rho \sigma_x \sigma_y). \end{aligned}$$

Thus, vessel i 's expected payoff when other vessels use the policy η is

$$\begin{aligned} P(\gamma, \eta) &\equiv \frac{B}{2} \left(((\eta - 1)^2 + (\gamma - 1)^2) \sigma_x^2 + (\gamma - \eta)^2 \sigma_y^2 - 2(\gamma - \eta)^2 \rho \sigma_x \sigma_y \right) \\ &\quad - \frac{A}{2} (\gamma^2 \sigma_y^2 + (1 - \gamma)^2 \sigma_x^2 + 2(1 - \gamma) \gamma \rho \sigma_x \sigma_y). \end{aligned} \quad (10)$$

The vessel's first and second order conditions are, respectively

$$\frac{dP}{d\gamma} = -(A - B) (\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y) \gamma + B (2\rho\sigma_x\sigma_y - \sigma_y^2) \eta + (A - B) \sigma_x^2 - A\rho\sigma_x\sigma_y = 0 \quad (11)$$

$$\frac{d^2P}{d\gamma^2} = -(A - B) (\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y) = -A\sigma_x^2 (1 - \tau) (1 + r^2 - 2\rho r) < 0 \quad (12)$$

The second order condition guarantees that the vessel's maximization problem is concave. This condition holds if and only if $\tau < 1$ (given the maintained assumption $A > 0$). This claim is obvious for $\rho \leq 0$, so we need establish it only for $\rho > 0$. Here we have (using $\rho < 1$)

$$1 + r^2 - 2\rho r > 1 + r^2 - 2r = (1 - r)^2 \geq 0. \quad (13)$$

Therefore, the second order condition holds if and only if $1 - \tau > 0$, i.e. where the cost of deviating from the target is more important than the congestion costs.

The first order condition, Equation 11, defines vessel i 's best response function to η . To guarantee a unique (linear) symmetric Nash equilibrium, we need only to exclude the possibility that the slope of the vessel's best response function is 1. Equivalently, we need to exclude the possibility that the coefficients of γ and η in Equation 11 have the same magnitude but different signs.²⁹ That knife-edge case occurs if and only if $1 - \tau = r(2\rho - r)$. This equality cannot hold for $\rho = 0$ (because $\tau < 1$ by the second order condition). However, it might hold for positive ρ . Therefore, for the general case we assume henceforth that $1 - \tau \neq r(2\rho - r)$. Our parameter estimates satisfy this condition.

Evaluating the first order condition at $\gamma = \eta$, we obtain the equation for the symmetric equilibrium

$$((B - A) \sigma_x^2 - A\sigma_y^2 + 2A\rho\sigma_x\sigma_y) \gamma - (B - A) \sigma_x^2 - A\rho\sigma_x\sigma_y = 0. \quad (14)$$

Using the definitions of τ and δ , and solving Equation 14 for γ we obtain the symmetric equilibrium control rule

$$\gamma^{NE} = \frac{(B - A) \sigma_x^2 + A\rho\sigma_x\sigma_y}{((B - A) \sigma_x^2 - A\sigma_y^2 + 2A\rho\sigma_x\sigma_y)} = \frac{1 - \tau - \rho r}{1 - \tau + r^2 - 2\rho r}, \quad (15)$$

shown as Equation 3 in the text.

Before proving the proposition, we note that the equilibrium payoff of the representative

²⁹In that case, there is no symmetric solution to Equation 11, unless in addition $1 - \tau = \rho r$. That additional equality implies that the intercept of the vessel's best response function is 0. That special case, together with a slope of the best response function equal to 1, implies that every $\gamma = \eta$ satisfies the first order condition.

agent (in the unique symmetric linear equilibrium) is

$$Q(\gamma; \sigma_x^2, \sigma_y^2, \rho) \equiv P(\gamma, \gamma; \sigma_x^2, \sigma_y^2; \rho) = \frac{B}{2} \left((\gamma - 1)^2 + (\gamma - 1)^2 \right) \sigma_x^2 - \frac{A}{2} \left(\gamma^2 \sigma_y^2 + (1 - \gamma)^2 \sigma_x^2 + 2(1 - \gamma) \gamma \rho \sigma_x \sigma_y \right). \quad (16)$$

Substituting the NE decision rule, from Equation 15, into $Q(\cdot)$, we obtain welfare as a function of $\tau, \sigma_x^2, \sigma_y^2$ and ρ .

Proof. (Proposition 1) We can express the Nash equilibrium payoff in terms of primitives, using the expression for the planner's payoff in equation (16), evaluated at the NE policy, $\gamma = \sigma_x^2 \frac{A-B}{-B\sigma_x^2 + A\sigma_y^2 + A\sigma_x^2}$, given by Equation 15. This substitution gives the NE payoff

$$Z \equiv -\frac{1}{2} \sigma_y^2 A \sigma_x^2 \frac{-2BA\sigma_y^2 + A^2\sigma_x^2 - 2B\sigma_x^2 A + \sigma_x^2 B^2 + A^2\sigma_y^2}{(-B\sigma_x^2 + A\sigma_y^2 + A\sigma_x^2)^2} = -\frac{1}{2} \sigma_y^2 A \sigma_x^2 \frac{\sigma_x^2 (\tau^2 - 2\tau + 1) + \sigma_y^2 (1 - 2\tau)}{(-\tau\sigma_x^2 + \sigma_y^2 + \sigma_x^2)^2} \quad (17)$$

Part i. The derivative of Z with respect to σ_y^2 is

$$\begin{aligned} \frac{dZ}{d\sigma_y^2} &= \left(\frac{\frac{1}{2} A \sigma_x^2 (1 - \tau)}{((\tau - 1)\sigma_x^2 - \sigma_y^2)^3} \right) \left((1 - 3\tau) \sigma_y^2 + \sigma_x^2 (\tau^2 - 2\tau + 1) \right) \\ &= \frac{1}{2} A \delta^2 \frac{1 - \tau}{(1 - \delta\tau)^3} (3\tau - \tau\delta - \tau^2\delta - 1). \end{aligned} \quad (18)$$

This expression implies that $\frac{dZ}{d\sigma_y^2} > 0 \Leftrightarrow (3\tau - \tau\delta - \tau^2\delta - 1) > 0$, i.e. welfare increases with the precision of public information if and only if

$$3\tau - 1 < \delta\tau(\tau + 1). \quad (19)$$

We now consider three cases. (a) For $\tau > 0$, Inequality 19 holds if and only if $\delta > \frac{3\tau - 1}{\tau^2 + \tau}$. (b) For $0 > \tau > -1$, Inequality 19 holds for $\delta < \frac{3\tau - 1}{\tau^2 + \tau}$. Because $\frac{3\tau - 1}{\tau^2 + \tau} > 1$, greater precision of public information increases welfare for $0 > \tau > -1$. (c) For $\tau < -1$, Inequality 19 holds for $\delta > \frac{3\tau - 1}{\tau^2 + \tau}$. This inequality always holds because for $\tau < -1$, $\frac{3\tau - 1}{\tau^2 + \tau} < 0$. Therefore, for $\tau < 0$, an increase in the precision of public information increases welfare.

Part ii. We also have

$$\begin{aligned} \frac{dZ}{d\sigma_x^2} &= -\frac{1}{2} A \frac{(\sigma_y^2)^2}{(\sigma_x^2 + \sigma_y^2 - \sigma_x^2 \tau)^3} (\sigma_x^2 + \sigma_y^2 - \sigma_x^2 \tau - 2\sigma_y^2 \tau) \\ &= -\frac{1}{2} A \frac{(\delta - 1)^2}{(1 - \tau\delta)^3} (\tau\delta - 2\tau + 1). \end{aligned} \quad (20)$$

Therefore, welfare increases with the precision of private information (i.e. $\frac{dZ}{d\sigma_x^2} < 0$) if and only if $(\tau\delta - 2\tau + 1) > 0$. For $\tau > 0$ this inequality requires $\delta > \frac{2\tau-1}{\tau}$. For $\tau < 0$, the inequality requires $\delta < \frac{2\tau-1}{\tau}$, which always holds.

Part iii. This claim is apparent from Figure 1. □

Proof. (Corollary 1) The corollary is essentially a description of Figure 1. More precise private information corresponds to a decrease in σ_y and an increase in δ . At a point below the upper boundary of region E , more precise public information lowers welfare (so $\frac{dZ}{d\sigma_y^2} > 0$), and above this boundary, more precise public information raises welfare (so $\frac{dZ}{d\sigma_y^2} < 0$). The function Z is continuously differentiable in its arguments. Therefore, for given τ , Z reaches its minimum on the boundary, where $\frac{dZ}{d\sigma_y^2} = 0$. A parallel argument establishes the claim regarding the lower boundary of region E . □

B Relation to AP07

AP07 study the general linear-quadratic-Gaussian problem. As we noted in Section 3, we drop the Gaussian assumption, because we use the minimum variance estimator instead of the Bayesian estimator. However, we retain the linear-quadratic structure. In an important special case of their model, the Nash equilibrium and the planner’s solution are identical under perfect information, but differ under incomplete information (Section 5.4 of their paper). Our model is a special case of that special case. Therefore, it is not surprising that we obtain, for this case, a more complete description of the welfare effects of more precise public and/or private information. It is worth understanding why our reduced generality increases clarity.

We first confirm that our formula for the equilibrium is the same as theirs, and then contrast our comparative statics results (Proposition 1) with those reported in Section 5.4 of their paper. We digress to confirm Claim 1, using notation introduced in this appendix. We close the appendix by explaining why the admissible parameter set in our formulation ($\tau < 1$) is larger than the admissible set in theirs ($\tau < 0.5$).

AP07 introduce a social planner who solves a “team problem”. This planner can instruct vessels how to use information, but cannot share private information across vessels. The planner wants to maximize the expected welfare of the representative vessel, $Q(\gamma; \sigma_x^2, \sigma_y^2)$ in Equation 16. AP07 impose two pairs of restrictions on payoff parameters. The first pair guarantees concavity of the agents’ payoff in its action, k_i , and implies uniqueness of the Nash equilibrium; the second pair is sufficient for concavity of the planner’s payoff. We confirm

that our restrictions $A > 0$ and $\tau < 1$ are equivalent to their first pair of restrictions. We then show that AP07's second pair of restrictions is sufficient, but not necessary for concavity of the planner's maximand, $Q(\gamma; \sigma_x^2, \sigma_y^2)$. Because our analysis does not rely on existence of the planner's solution, we are able to examine the comparative statics of the precision of information over a larger set of parameters.

We begin by showing that our expression for the NE policy weight, γ^{NE} , coincides with the value obtained by specializing their formulae. Define $K \equiv \int_{k_{-i}} k_{-i} p(k_{-i}) dk_{-i}$, vessel i 's expectation of other vessels' average location, conditional on the public information and vessel i 's private information; $p(k_{-i})$ represents vessel i 's subjective probability of any other vessel's location. Vessel i obtains the expected strategic benefit of locating at k_i equal to

$$\begin{aligned} \frac{B}{2} \int_{k_{-i}} (k_i - k_{-i})^2 p(k_{-i}) dk_{-i} &= \frac{B}{2} \int_{k_{-i}} (k_i - K - (k_{-i} - K))^2 p(k_{-i}) dk_{-i} = \\ \frac{B}{2} \left[(k_i - K)^2 + \int_{k_{-i}} (k_{-i} - K)^2 p(k_{-i}) dk_{-i} - 2(k_i - K) \int_{k_{-i}} (k_{-i} - K) p(k_{-i}) dk_{-i} \right] &= \\ \frac{B}{2} [(k_i - K)^2 + \sigma_k^2] & \end{aligned} \tag{21}$$

The last line uses the definition $\sigma_k^2 \equiv \int_{k_{-i}} (k_{-i} - K)^2 p(k_{-i}) dk_{-i}$, a measure of the dispersion of agents' actions, and the fact that $\int_{k_{-i}} (k_{-i} - K) p(k_{-i}) dk_{-i} = 0$ from the definition of K .

The agent's payoff equals the strategic component minus the expected cost of being distant from the ideal location, $\frac{A}{2} (k_i - \theta)^2$. This payoff is

$$u_i = U(k_i, K, \sigma_k^2, \theta) = \frac{B}{2} [(k_i - K)^2 + \sigma_k^2] - \frac{A}{2} (k_i - \theta)^2. \tag{22}$$

To simplify notation, we sometimes replace k_i with k . With this notation, we write U_{kk} as the second derivative of an agent's payoff with respect to their action. AP07 assume that $U_{kk} < 0$, $\frac{-U_{kK}}{U_{kk}} < 1$; the first inequality implies that the vessel's problem is concave in their decision, and the second guarantees uniqueness of the linear equilibrium. For our problem, $U_{kk} = B - A < 0 \Leftrightarrow \tau \equiv \frac{B}{A} < 1$, and $U_{kK} = -B \Rightarrow \frac{-U_{kK}}{U_{kk}} = \frac{B}{B-A} < 1$.

We briefly digress to use the definitions introduced above to establish Claim 1.

Proof. (Claim 1) Using the definition $g_i \equiv k_i - \theta$ and Equation 22, we write vessel i 's payoff as

$$\begin{aligned} u_i = U(k_i, K, \sigma_k^2, \theta) &= \frac{B}{2} [(k_i - K)^2 + \sigma_k^2] - \frac{A}{2} (k_i - \theta)^2 \\ &= \frac{B}{2} [(k_i - \theta - (K - \theta))^2 + \sigma_k^2] - \frac{A}{2} (k_i - \theta)^2 \\ &= \frac{B}{2} [(g_i - (K - \theta))^2 + \sigma_k^2] - \frac{A}{2} (g_i)^2 \Rightarrow . \end{aligned}$$

The average location of other vessels, K , is a random variable with conditional mean $\mathbf{E}(K|\theta) = \theta$ and variance σ_K^2 . Therefore, vessel i 's expected payoff, conditional on θ , is a decreasing concave function of the vessel's distance from the ideal location, $|g_i|$:

$$\mathbf{E}(u_i|\theta) = \frac{B-A}{2}g_i^2 + \frac{B}{2}(\sigma_K^2 + \sigma_k^2).$$

□

Having confirmed Claim 1, we now return to the main topic of this Appendix: the comparison between our paper and AP07. AP07 show that the full information privately optimal decision rule is $k = k_0 + k_1\theta$, and for this problem $k_0 = 0$ and $k_1 = 1$. (AP07 page 1111). The full information (Nash) equilibrium is therefore $k = \theta$.

AP07 use z to represent the Bayesian estimate of θ conditional on the public signal, y , and prior information. We simplify the notation by using y to represent the BLUE estimator of θ conditional on the public signal. This simplification reduces the number of parameters without any loss of generality. We can make this usage consistent with AP07 by assuming that the prior beliefs about θ are diffuse. The decision rule that involves the private signal, x_i and the Bayesian estimate z in their setting, involves x_i and y in our setting.

AP07's formula for the incomplete information equilibrium decision rule is $k(x, y) = (1 - \gamma)x + \gamma y$ (their equation 8 pg 1112, replacing z by y), with

$$\gamma = \delta + \frac{\alpha\delta(1-\delta)}{1-\alpha(1-\delta)} \text{ and } \alpha = \frac{B}{B-A}. \quad (23)$$

A calculation establishes that the value of γ in Equation 23 equals γ^{NE} defined in Equation 15.

AP07 also impose the restrictions $U_{kk} + 2U_{kK} + U_{KK} < 0$, and $U_{kk} + U_{\sigma\sigma} < 0$, noting that these are sufficient for the concavity of the planner's problem. The first of these inequalities holds if and only if $B - A - 2B + B = -A < 0$, which reproduces our earlier assumption. The second holds if and only if $B - A + B < 0 \Rightarrow \tau \equiv \frac{B}{A} < 0.5$. Thus, AP07's full set of parameter restrictions is satisfied for our problem if and only if $A > 0$ and $\tau < 0.5$. Because we do not use the social planner, our comparative statics results hold for $A > 0$ and $\tau < 1$.

Specializing AP07 Proposition 2 to our setting, we have the planner's decision rule

$$k(x, y) = (1 - \gamma^*)x + \gamma^*y$$

with

$$\gamma^* = \delta + \frac{\alpha^*\delta(1-\delta)}{1-\alpha^*(1-\delta)} \text{ and } \alpha^* = \frac{2B}{2B-A}. \quad (24)$$

Although we do not use the planner, merely as a consistency check we confirm that the

decision rule for the planner obtained using our formulation is the same as AP07's. The planner's maximand is $Q(\gamma; \sigma_x^2, \sigma_y^2)$, implying the FOC and decision rule:

$$\begin{aligned} \frac{dQ}{d\gamma} &= (2B\sigma_x^2 - A\sigma_y^2 - A\sigma_x^2)\gamma - 2B\sigma_x^2 + A\sigma_x^2 = 0 \Rightarrow \\ \gamma^{SP} &= \delta \frac{A-2B}{A-2B\delta} = \delta \frac{1-2\tau}{1-2\tau\delta}. \end{aligned} \tag{25}$$

A calculation shows that γ^* , given in Equation 24, and γ^{SP} from Equation 25 are equivalent. The second order condition to the problem of maximizing $Q(\gamma; \sigma_x^2, \sigma_y^2)$ is

$$\begin{aligned} \frac{d^2Q}{d\gamma^2} &= -(\sigma_y^2 + \sigma_x^2)A + 2B\sigma_x^2 = -(\sigma_y^2 + \sigma_x^2)(A - 2B\delta) < 0 \\ &\Leftrightarrow \tau < \frac{1}{2\delta}. \end{aligned} \tag{26}$$

Thus, AP07's sufficient condition for the concavity of the planner's problem in our setting, $\tau < 0.5$, is stronger than our necessary and sufficient condition $\tau < \frac{1}{2\delta}$. However, our condition involves information, via δ . AP07 do not allow the planner to manipulate information. Our condition holds for all information if and only if $\tau < 0.5$, thus returning AP07's condition. (Our analysis does not use the planner's problem, so this difference has no bearing on our comparative statics results.)

As mentioned above, our model is a special case of the setting where equilibrium actions are efficient under full information but not under incomplete information. Their analysis uses a mapping between the information parameter pair (σ_x^2, σ_y^2) and $(\delta, \sigma_x^{-2} + \sigma_y^{-2})$ ("commonality" and "accuracy" in their parlance). Their Proposition 7.i states that welfare increases with precision (holding commonality fixed). Their Proposition 7.ii states that for $\tau < 0$, welfare increases with δ ; and for $0 < \tau < 0.5$, welfare falls with δ (holding accuracy fixed).³⁰ Using their Proposition 7 in our setting, and restricting $\tau < 0.5$, we can restate the results in their Corollary 4, using our notation, as:

- i** $\tau < 0$ is a sufficient condition for welfare to increase with the precision of both private and public information
- ii** $0 < \tau$ is necessary for welfare to fall with the precision of public information and sufficient for welfare to rise with the precision of private information.

AP07 study situations where the planner's decision can be decentralized as a Nash equilibrium by manipulating only the agents' preferences, not by changing their information. For

³⁰They state their results in terms of α and α^* . Using the formulae for these given above, we verify that (over their parameter set $\tau < 0.5$) $\alpha^* > \alpha > 0$ if and only if $\tau < 0$ and $\alpha^* < \alpha < 0$ if and only if $0 < \tau < 0.5$.

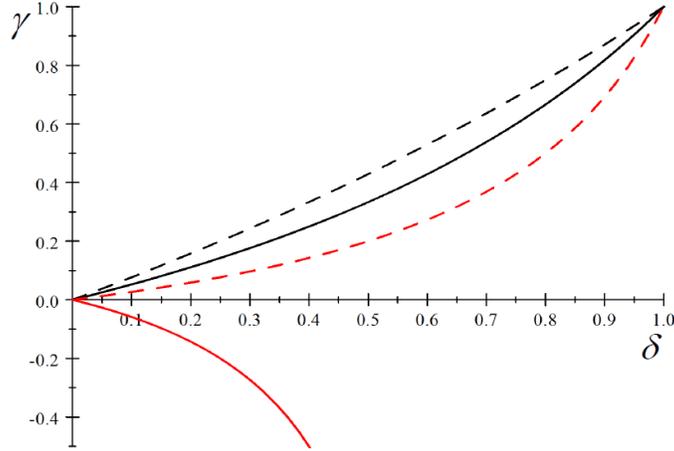


Figure B1: The solid curves show γ^{SP} for $\tau = 0.25$ (black) and $\tau = 0.75$ (red). The dashed curves show γ^{NE} for these two values of τ . In both cases, $\gamma^{SP} < \gamma^{NE}$. However, for $\tau = 0.75$, $\gamma^{SP} < 0$ and decreases with δ . For $\tau = 0.75$, γ^{SP} exists only for $\delta < 0.67$, where the planner's problem is concave.

this reason, AP07 restrict parameters to satisfy $\tau < 0.5$, instead of $\tau < 1$ as in our formulation. A comparison of the expressions for α and α^* shows that the Nash equilibrium supports the socially optimal decision if the agent's parameter B is replaced by $B' = 2B$. However, for $B > \frac{A}{2}$, $B' - A > 0$; here, the agent's problem with the “modified” preferences is convex, so there is no symmetric pure strategy equilibrium. Thus, for $B > \frac{A}{2}$, the planner would have to manipulate both the agent's preferences (τ) and their information (δ) for the Nash equilibrium to support the socially optimal plan.

Even if the planner could manipulate both the agent's preferences and information, that ability would not be useful if $1 > \tau > \frac{1}{2\delta}$. Over this region, the planner's problem is convex, i.e. the maximand is unbounded. For $0 < \tau < 0.5$, both γ^{SP} and γ^{NE} are increasing in δ and $\gamma^{SP} < \gamma^{NE}$. In both equilibria, agents put more weight on the public signal as its precision (relative to that of the private signal) increases. However, the planner instructs vessels to put less weight on the public signal, compared to the Nash Equilibrium, in order to reduce congestion. For $0.5 < \tau < \frac{1}{2\delta}$ (where both γ^{SP} and γ^{NE} exist), it is still the case that $\gamma^{SP} < \gamma^{NE}$; the explanation is unchanged. However, over this region $\gamma^{SP} < 0$ and it decreases in δ . When congestion is severe ($\tau > 0.5$) the planner would like vessels to put negative weight on the public signal in order to reduce congestion. Figure B1 illustrates these comparative statics. The solid curves show γ^{SP} for $\tau = 0.25$ (black) and for $\tau = 0.75$ (red). The dashed curves show the corresponding graphs of γ^{NE} for these two values of τ .

C Comparative statics

The BLUE weight δ is discontinuous in parameters, and changes sign, at $\rho = \frac{1+r^2}{r}$. The function is non-monotonic in parameters, with

$$\frac{d\delta}{d\rho} < 0 \Leftrightarrow r > 1 \text{ and } \frac{d\delta}{dr} > 0 \Leftrightarrow \rho > 4\frac{r}{r^2+1}. \quad (27)$$

A higher correlation between signals increases the BLUE weight on the public signal if and only if the public signal is relatively precise, compared to the private signal ($r < 1$). For ρ close to 0, a relatively more precise public signal (smaller r) increases the BLUE weight on the public signal, as we would expect. However, for $r > 3$ and sufficiently large ρ , a relatively less precise public signal increases the BLUE weight on the public signal.

The function γ^{NE} is non-monotonic in parameters, with

$$\begin{aligned} \frac{d\gamma^{NE}}{d\rho} < 0 &\Leftrightarrow 1 - r^2 - \tau < 0, \quad \frac{d\gamma^{NE}}{d\tau} < 0 \Leftrightarrow \rho < r, \\ \text{and } \frac{d\gamma^{NE}}{dr} < 0 &\Leftrightarrow (r^2 - \tau + 1)\rho < 2r(1 - \tau). \end{aligned} \quad (28)$$

D Figures 4(a) and 4(b)

We provide formulae for the change in welfare arising from more precise private or public information. Using these formulae and the expression for equilibrium welfare, we obtain the elasticities of welfare with respect to the precision of public and private welfare.

Using Equation 16, welfare in the symmetric equilibrium is

$$P = \frac{A}{2} \left(\tau \left(2 \left((\gamma - 1)^2 \right) \sigma_x^2 \right) - \left(\gamma^2 \sigma_y^2 + (1 - \gamma)^2 \sigma_x^2 + 2(1 - \gamma) \gamma \rho \sigma_x \sigma_y \right) \right).$$

We evaluate this function at the Nash equilibrium decision rule, here denoted simply as γ (rather than γ^{NE}).

We first consider the effect of changing the precision of private information, σ_x . We have

$$\begin{aligned} \frac{dP}{d\sigma_x} = & -A\sigma_x\sigma_y^3(\rho\sigma_x - \sigma_y) \times \\ & \frac{2\rho^3\sigma_x\sigma_y - \rho^2\sigma_x^2 - \tau\rho^2\sigma_x^2 - \rho^2\sigma_y^2 + 4\tau\rho\sigma_x\sigma_y - 2\rho\sigma_x\sigma_y + \sigma_x^2 - \tau\sigma_x^2 - 2\tau\sigma_y^2 + \sigma_y^2}{(\tau\sigma_x^2 - \sigma_x^2 + 2\rho\sigma_x\sigma_y - \sigma_y^2)^3}. \end{aligned}$$

We simplify the right side of this equation using the definition $r = \frac{\sigma_y}{\sigma_x}$, to produce $\frac{dP}{d\sigma_x} = -A\sigma_y D$, using the definition

$$D \equiv r^2(\rho - r) \frac{(2\rho^3 r - \rho^2 - \tau\rho^2 - \rho^2 r^2 + 4\tau\rho r - 2\rho r + 1 - \tau - 2\tau r^2 + r^2)}{(\tau - 1 + 2\rho r - r^2)^3}.$$

More precise private information corresponds to a smaller σ_x . Because $-A\sigma_y < 0$, we conclude that more precise private information raises welfare if and only if $D > 0$.

To examine the effect of an increase in the precision of public information, we use $\frac{dP}{d\sigma_y} = -A\sigma_y E$, with $E \equiv \frac{a_0 + a_1 r + a_2 r^2 + a_3 r^3}{(\tau - 1 + 2\rho r - r^2)^3}$, and the definitions

$$\begin{aligned} a_0 &= 3\tau - 3\tau^2 - 1 + \rho^2 - \tau\rho^2 + \tau^3 \\ a_1 &= 3\rho - 3\rho^3 - 9\rho\tau + 6\tau^2\rho + 3\tau\rho^3 \\ a_2 &= -1 + 2\rho^4 + 2\rho^2\tau - 3\tau^2 - \rho^2 + 4\tau \\ a_3 &= -2\rho\tau + \rho - \rho^3. \end{aligned}$$

More precise public information implies a smaller value of σ_y . Because $-A\sigma_y < 0$, we conclude that more precise public information increases equilibrium welfare if and only if $E > 0$.