American Economic Review: Papers & Proceedings 2014, 104(5): 463–468
http://dx.doi.org/10.1257/aer.104.5.463

How Sharing Information Can Garble Experts' Advice†

By Matthew Elliott, Benjamin Golub, and Andrei Kirilenko*

Executives often employ experts who draw on rich information to recommend an appropriate action. For instance, a commander-in-chief is advised by several intelligence experts about how to respond to potential threats.

When such interactions are discussed in policy circles, a common worry is that there is not enough sharing of information among experts, resulting in bad advice and inefficient outcomes. On the one hand, it is clear that combining information can result in better decisions. On the other hand, integrating experts' capacities could potentially interact with their incentives in ways that harm a principal. In this article, we focus on the latter force, and describe a class of examples in which a principal is worse off when experts share information.

Consider a situation where there is an unknown binary state of interest (e.g., whether a terrorist attack will be attempted). There is a correct decision for each state (e.g., whether to take costly preemptive measures). Each of several experts (e.g., the directors of the CIA and the FBI) chooses from a set of procedures, which generate information about the state. They then use that information to recommend an action to the executive, who decides which action to take. Each expert finds it costly to recommend the ex post incorrect action. The key payoff variables are the costs of two kinds of error: predicting the unusual event when it is not going to happen (a false positive, or type I error) and failing to predict it when it is about to happen (a false negative, or type II error).

Experts' incentives, which are exogenous in our model, are allowed not to be aligned with each other or with the principal's interests. In particular, while everyone dislikes making errors, the way they trade off false positives and false negatives may differ. In making his choice, the principal aggregates both experts' recommendations optimally given his own payoffs.

One may conjecture that giving experts access to more information improves decisions. For example, if the principal and both experts trade off type I and type II errors in the same way, then the principal can delegate the decision to the experts, and—because the problem is essentially a single-agent one—sharing information never makes the outcome worse.

We study whether it holds more generally that information sharing is beneficial for the principal. Sharing information is modeled as giving each expert access to some of the procedures possessed by the other. We show that in a natural class of examples (where information structures are particularly simple), sharing brings about worse outcomes for the principal whenever it is desired by the experts.

In taking experts' payoffs as exogenous, we abstract from some interesting contracting issues. However, given that experts often have large and privately known payoffs associated with their reputations, this is probably a closer fit to reality than a model in which (for example) a government offers the CIA director state-contingent payments to align his incentives.

Our work relates to a literature on committees of experts giving advice in the presence of conflicts of interest—e.g., Li, Rosen, and Suen (2001); Li and Suen (2004); Suen (2004); and Ottaviani and Sørensen (2006). As in Kartik (2009), language has an intrinsic meaning in our model, and there is an exogenous cost to making statements that turn out to be incorrect.

*Elliott: Division of Humanities and Social Sciences 228-77, Caltech, Pasadena, CA 91125 (e-mail: melliott@hass.caltech.edu); Golub: Society of Fellows and Department of Economics, Littauer Center, Harvard, Cambridge, MA 02138 (e-mail: ben.golub@gmail.com); Kirilenko: MIT Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: ak67@mit.edu). We thank Jeremy Bulow, Matthew O. Jackson, Paul Milgrom, and Michael Ostrovsky for helpful discussions. We are also grateful to Heski Bar-Isaac, Alex Frankel, and Ben Olken for their suggestions.

† Go to http://dx.doi.org/10.1257/aer.104.5.463 to visit the article page for additional materials and author disclosure statement(s).

For another recent study of strategic issues in the provision of advice, see Kamenica and Gentzkow (2011).

The structure of this note is as follows. In Section I, we set up the primitives and formally define sharing information. In Section II, we present an example showing how permitting sharing can hurt the principal. Section III discusses the result and offers some caveats.

I. Model

A. *The Environment*

Agents.—The *agents,* also called experts, are indexed by $i \in I$.

States.—The state of interest is $s \in \mathcal{S} = \{0, 1\}$, corresponding to whether an event will occur. The random variable corresponding to the state is denoted by S. The "null hypothesis"—for the purpose of defining type I and type II errors—is state 0.

Technologies.—Each expert can choose one of many *procedures* for generating a recommendation $x_i \in \mathcal{S}$. Formally, a procedure is summarized by a function $\pi : \mathcal{S} \to [0, 1]$, where $\pi(s)$ is the probability of making recommendations conditional on state s occurring—i.e., of being correct in state s. Thus, coordinatewise larger procedures correspond to better information about the state. The set of procedures available to expert i is called his *technology* and is denoted by Π_i. We will sometimes write procedures as ordered pairs: $(\pi(0), \pi(1))$.

Timing and Payoffs

(i) Each expert $i \in I$ chooses a procedure $\pi_i \in \Pi_i$.

(ii) The state is realized, with the following distribution:

$$S = \begin{cases} 1 & \text{with probability } \mu_1 \\ 0 & \text{with probability } \mu_0 = 1 - \mu_1. \end{cases}$$

(iii) Each expert simultaneously generates a random recommendation X_i, taking values in $\mathcal{S}$ (the state space). The recommendations are conditionally independent given the state S, and satisfy $\mathbf{P}(X_i = s \,|\, S = s) = \pi_i(s)$.

(iv) Expert i's (random) payoff is $\theta^i_S \mathbf{1}\{X_i = S\}$. The amount $\theta^i_S > 0$ of the reward depends both on the expert's identity and on the (random) state. The expert receives the reward only when his recommendation matches the state. We denote by $U_i(\pi_i)$ expert i's ex ante expected payoff from selecting procedure π_i.

Principal.—The principal sees the vector $(X_i)_{i\in I}$ of recommendations and chooses an action $a \in \mathcal{S}$. His payoff is $\theta^p_S \mathbf{1}\{S = a\}$, where $\theta^p_s > 0$ for every $s \in \mathcal{S}$.

The agents and principal are all risk-neutral.

An *environment* is defined as a tuple $e = (I, \boldsymbol{\mu}, (\Pi_i)_{i\in I}, (\boldsymbol{\theta}^i)_{i\in I}, \boldsymbol{\theta}^p)$.

B. *Technologies Arising from Signals*

There are several ways to generate procedures; here we present a simple way for a procedure to come about as a use of a noisy signal. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space on which the random variable S (the state) is defined. Suppose an expert i observes a random variable (a signal) taking values in a space $\mathcal{Z}$, i.e., $Z : \Omega \to \mathcal{Z}$. Any mixed strategy $\sigma : \mathcal{Z} \to \Delta(\mathcal{S})$, which generates (random) recommendations based on this signal, induces a procedure $\pi^{Z,\sigma}$ via the definition $\pi^{Z,\sigma}(s) = \mathbf{E}[\sigma_i(Z)(s) \,|\, S = s]$.

The right-hand side is a conditional expectation given that state s obtains. The quantity whose expectation is being taken is the probability put on s by the mixed strategy mapping signals Z to predictions. In other words: expert i's strategy describes, for every signal he might see, a probability with which to make the recommendation $s = 1$. This induces a probability that his recommendation is correct conditional on any state, and is summarized by the procedure $\pi^{Z,\sigma} : \mathcal{S} \to [0, 1]$.

For a fixed random variable Z, let the technology $\Pi(Z)$ be the set of procedures $\pi^{Z,\sigma}$ as σ ranges over all mixed strategies.

C. *Technologies: Assumptions and Definitions*

This section focuses on assumptions about technologies. We want to permit experts to randomize among procedures; to make recommendations that don't rely on information about the state; and to implement the "opposite" of a given procedure. The following assumptions on technologies formalize these conditions.

ASSUMPTION 1: *For each i*, Π_i *is convex.*[1]

This follows from allowing experts to implement lotteries over the procedures available to them. Next, we assume each expert has access to trivial procedures, which always issue the same recommendation (and therefore do a perfect job of avoiding one kind of error).

ASSUMPTION 2: *For each i, the technology* Π_i *contains the procedures* $(0, 1)$ *and* $(1, 0)$.

Finally, if a procedure can be implemented, then so can the procedure obtained by taking the output of that procedure and issuing the opposite recommendation.

ASSUMPTION 3: *For each i, if* $\pi \in \Pi_i$, *then*[2] $\mathbf{1} - \pi \in \Pi_i$.

It will be helpful to define the set of procedures that must (under the foregoing three assumptions) be available to an expert, whenever his technology contains an arbitrary set of procedures. To this end, let $\mathcal{P}$ be the family of all technologies Π satisfying the above assumptions. As $\mathcal{P}$ is closed under intersections[3] we can define the closure of Π, written $\mathrm{cl}(\Pi)$, as the smallest set containing Π and satisfying all the assumptions:

$$\mathrm{cl}(\Pi) = \bigcap_{\substack{\Pi' \in \mathcal{P} \\ \Pi \subseteq \Pi'}} \Pi'.$$

Finally, we define a simple technology as one that is generated by (i.e., is the closure of) a single procedure.

DEFINITION 1: *A technology* Π *is* simple *if there is some procedure* π *so that* $\pi = \mathrm{cl}(\pi)$.

It can readily be checked that a technology Π is simple if and only if $\Pi = \Pi(Z)$ for some random variable Z taking values in $\mathcal{Z} = \{0, 1\}$. Thus, simple technologies correspond to binary experiments.

D. *Sharing Information*

We will study the effects of experts sharing their technologies with each other. Technology sharing is assumed to be imperfect and so we first define a class of procedures that yield strictly less information than a given one.

DEFINITION 2: *For a procedure* π, *we define the set of garblings of* π *to be* $L(\pi) = \{\alpha\pi + (1-\alpha)\pi' : \pi' \in C, \alpha \in [0, 1)\}$, *where* $C = \mathrm{cl}(\emptyset)$ *is the set of uninformative procedures.*[4]

The set $L(\pi)$ is simply the set of procedures obtained by taking π and mixing it with a positive amount of some uninformative procedure—i.e., effecting a nontrivial Blackwell garbling. We can now define what it means for experts to share their technologies.

DEFINITION 3: *Given a tuple of technologies*, $\mathbf{\Pi} = (\Pi_i)_{i \in I}$, *we say that another tuple* $\mathbf{\Pi}' = (\Pi_i)_{i \in I}$, *is a sharing of* $\mathbf{\Pi}$ *if, for every* $i \in I$, *we have*

$$\Pi_i' = \mathrm{cl}\left[\Pi_i \cup \left(\bigcup_{j \in I \setminus \{i\}} \tilde{\Pi}_{ij}\right)\right],$$

where $\tilde{\Pi}_{ij} \subseteq L(\Pi_j)$.

The definition stipulates that, after a sharing of information, each expert maintains access to his technology, and also has access to some technologies that are less informative (possibly only slightly) than those that were originally available to the other experts. Because of the definition of the operator L, sharing permits any expert to come arbitrarily close to replicating the procedures used by the others.

[1] A convex combination of two procedures π and π' is any procedure of the form $\alpha\pi + (1-\alpha)\pi'$, with $\alpha \in [0, 1]$. Addition and scaling of procedures is defined pointwise.

[2] The subtraction in the ensuing expression is defined pointwise.

[3] This is true of the set of technologies satisfying any single assumption.

[4] This is the set of all convex combinations of the procedures $(0, 1)$ and $(1, 0)$.

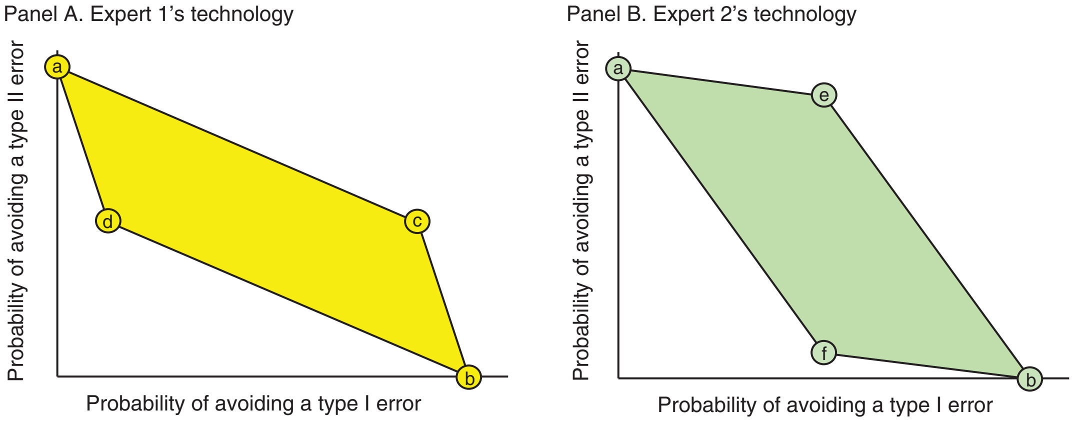

FIGURE 1. EXPERTS' TECHNOLOGIES CORRESPOND TO THEIR ABILITIES TO AVOID TYPE I AND TYPE II ERRORS

Note: Panels A and B show the type I and type II error rates available to the two experts through their technologies.

E. *Preferences over Environments*

Given an environment, an agent i's utility from using a technology $\pi \in \Pi_i$ is

$$U_i(\pi) = \sum_{s \in \mathcal{S}} \mu_s \, \pi(s) \theta_s^i.$$

When a profile $\boldsymbol{\pi}$ of procedures is played, the principal's utility is

$$U_p(\boldsymbol{\pi}) = \sup_{A \in \mathcal{A}} \mathbf{E}[\theta_S^p \mathbf{1}\{S = A\}],$$

where $\mathcal{A}$ is the family of all $\{0, 1\}$-valued random variables (action choices) measurable with respect to the vector of recommendations $(X_i)_{i \in I}$.

A given environment e induces a (nonempty) set of optimal outcomes $\mathbf{P}^*(e) = \Pi_{i \in I} P_i^*$, where P_i^* consists of all those $\pi \in \Pi_i$ that are optimal for agent i. We induce agents' and the principal's preferences over sets of outcomes from their basic preferences via the strong set order.

II. Example

In this section we consider a simple example demonstrating that a principal may not want to allow experts to share information. In particular, we construct two environments, distinguished only by the fact that the technologies in the second are a sharing of those present in the first. Both experts strictly prefer the post-sharing environment, but the principal strictly prefers the pre-sharing environment.

Suppose there are two experts $i \in \{1, 2\}$. Fixing the common prior $\boldsymbol{\mu}$, agents' technologies Π_i allow them to avoid making type I (false positive) and type II (false negative) errors. Their technologies correspond to production possibility sets in error-avoidance space. For the case in which both experts' technologies are simple, this is shown in Figure 1, panels A and B. Expert 1 has access to a single procedure π_1 that results in the error rates given by point c. However, expert 1 could "flip" this procedure (making the opposite recommendation), resulting in the error rates given by point d. Expert 1 can also make constant recommendations. This allows him to avoid making one type of error or the other for sure, generating the type I and type II error rates given by points a and b. Mixing among these four points, expert 1 can achieve anything in the shaded convex hull of Figure 1, panel A.[5] This shaded region corresponds to the error avoidance rates possible using technology $\Pi_1 = \mathrm{cl}(\pi_1)$. Expert 2's initial technology is constructed similarly in panel B.

[5] Of course, as an expert dislikes making both type I and type II errors, he will only ever choose error rates on the northeast boundary of this technology.

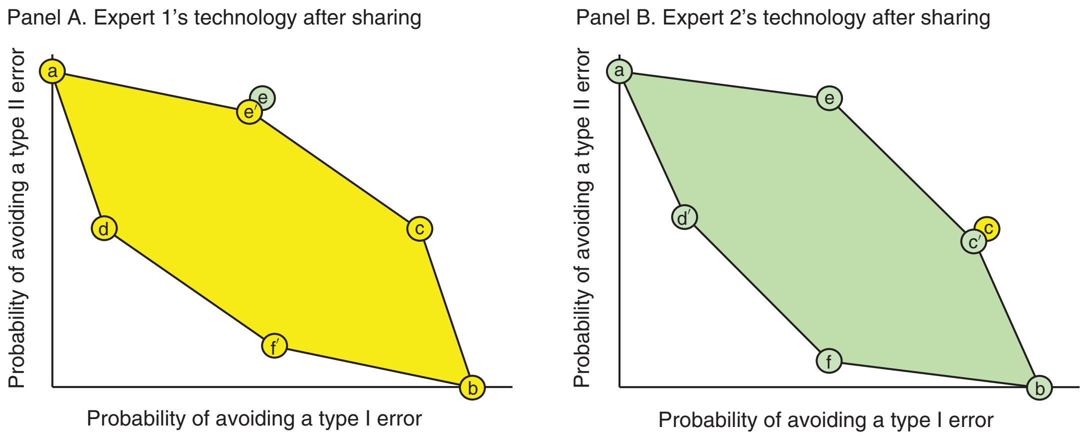

Figure 2. Experts' Technologies After Sharing

Note: Sharing technologies permits both experts to make, potentially, fewer type I and fewer type II errors. Nevertheless, this can result in recommendations that are strictly less informative for the principal.

We will suppose that in the initial environment, both experts strictly prefer making informative recommendations: that their optimal choices are c and e.

Suppose now that the experts share their information. As sharing is imperfect, suppose that expert 1 gains access to error avoidance rates e' and that expert 2 gains access to error avoidance rates c'. The new error avoidance production possibilities for the experts are shown in Figure 2, panels A and B. We assume the experts will share their information in this way only when they both want to. Information sharing will then occur if and only if: (i) expert 1 prefers the error avoidance rates e' to c; and (ii) expert 2 prefers the error avoidance rates c' to e. Suppose then that we are in a situation where the experts agree to exchange information. Is such an exchange beneficial for the principal?

After the exchange of information, expert 1 will choose to avoid making errors at rates e' instead of c while expert 2 will choose to avoid making errors at rates c' instead of e. Despite the improvement in the technologies of each expert, collectively the predictions the experts make after sharing information are unambiguously worse for *any* principal preferences. This is because at point c' more type I *and* more type II errors are made than at point c; meanwhile, at point e' both more type I and more type II errors are made than at point e. No matter how the principal chooses to use the experts' predictions after they have exchanged information, the principal could have done better with the predictions the experts would have made had they not shared information. Interestingly, it is because both the experts want to share information that information sharing is bad for the principal.

III. Discussion

The main aspects distinguishing our model from many other communication games are: (i) the requirement that experts make coarse predictions in the same space as the state space of interest; and (ii) the assumption that they enjoy exogenously given rewards when their predictions match the state. This perspective makes sense when the language in which experts give advice is not controlled by the principal who uses the information, and when "predicting correctly" has a meaning—and a value—beyond the interaction between the principal and experts.

Our model of experts' technologies requires each of them to use a *single* procedure in making a recommendation (both before and after a sharing). We call this a "no synergies" assumption. To take a concrete example, suppose that expert 1's procedure allows him to track one suspect, and expert 2's procedure allows him to track another suspect. Suppose that after a sharing of information, expert 1 can simultaneously track both suspects. Then if, for example, the two suspects meet the same (unknown) third party, such a meeting may provide a strong indication that an attack is imminent. Expert 1 can use this information to issue a warning. In our model there are no such synergies from information sharing.

The assumption of no synergies fits a case where each procedure[6] is a complete description of an indivisible bureaucratic routine, and there are not enough resources for a given expert to run multiple such routines in parallel. Alternatively, one can think of the model as holding constant each expert's "quantity" of signals, while varying their "quality" via a sharing. In any case, after a sharing, an expert can choose from a larger set of procedures than before, but can still run only one.

If instead each agent could use several of the experiments available to him and exploit synergies, then information sharing would result in different information improvements. In contrast to our result, these information improvements could lead to the experts' collectively making more informative predictions after a sharing. Synergies would be *necessary* in order for a sharing to be beneficial in the example we have outlined. Including them in the model would lead to a substantially richer analysis of when information sharing is desirable. We conjecture that, even in the presence of synergies, experts may choose not to take advantage of them. In that case, the principal could still be worse off, due to phenomena similar to those demonstrated in our example. However, a full exploration of these ideas is left for future work.

REFERENCES

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615.

Kartik, Navin. 2009. "Strategic Communication with Lying Costs." *Review of Economic Studies* 76 (4): 1359–95.

Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and Common Interests in Committees." *American Economic Review* 91 (5): 1478–97.

Li, Hao, and Wing Suen. 2004. "Delegating Decisions to Experts." *Journal of Political Economy* 112 (1): S311–35.

Ottaviani, Marco, and Peter Norman Sørensen. 2006. "Professional Advice." *Journal of Economic Theory* 126 (1): 120–42.

Suen, Wing. 2004. "The Self-Perpetuation of Biased Beliefs." *Economic Journal* 114 (495): 377–96.

[6] Including the process of interpreting its results.