Robust forecast aggregation

Itai Arieli[a,1], Yakov Babichenko[a,1], and Rann Smorodinsky[a,1]

[a]Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa 3200003, Israel

Edited by Paul R. Milgrom, Stanford University, Stanford, CA, and approved November 9, 2018 (received for review August 13, 2018)

Bayesian experts who are exposed to different evidence often make contradictory probabilistic forecasts. An aggregator, ignorant of the underlying model, uses this to calculate his or her own forecast. We use the notions of scoring rules and regret to propose a natural way to evaluate an aggregation scheme. We focus on a binary state space and construct low regret aggregation schemes whenever there are only two experts that either are Blackwell-ordered or receive conditionally independent and identically distributed (i.i.d.) signals. In contrast, if there are many experts with conditionally i.i.d. signals, then no scheme performs (asymptotically) better than a (0.5, 0.5) forecast.

information aggregation | one-shot regret minimization | conditionally independent information structure | Blackwell-ordered information structure

Just the other day, we were planning our weekend activities and looked at the weather forecast for New York on Sunday, August 12. In particular, what interested us was the probability of rain. Yahoo!'s precipitation forecast was 35%, the Weather Channel's was 50%, and TimeAndDate weather was only 6% (all three screenshots are provided in *SI Appendix*, Section 5). It was unclear to us how to aggregate these conflicting forecasts, although we knew that all three were from reputable sources that were using sound weather models and reliable data.

Our dilemma is not unique. In fact, many of us face conflicting advice from experts on a daily basis: forecasts from reliable pollsters on the outcome of presidential elections, medical prognoses from trusted physicians, investment advice from experienced financial pundits, and more.

This challenge is, in fact, inherent in many governing bodies. In the political arena, we often see ministers and legislators who, as elected officials, must decide on critical issues and policies while lacking subject-matter expertise. These publicly elected officials dictate health care policies and decide on military development and deployment, financial regulation, and so on, without any medical/military/financial background. As a result, they reach out to experts for advice, such as ad hoc committees, civil servants with years of experience, lobbyists, and more. Similar to elected officials, board members of commercial companies are often seasoned business people with managerial experience who often lack industry-specific knowledge. These board members essentially need to aggregate input from various experts to make a decision.

We consider a model with two types of agents. A set of nonstrategic experts share a common prior over the state space. Each expert receives a private signal that induces a posterior distribution (the expert's forecast). In contrast, an ignorant aggregator is not familiar with the common prior and the signal structure (we refer to this pair as the information structure). The aggregator observes the experts' forecasts and must aggregate them into a single forecast. How should we evaluate the aggregator, and what is his or her best course of action? These are the questions in which we are interested.

To study this, we first elucidate four aspects of the model:

- How to measure the accuracy of a forecast: A natural and prevalent family of measures of forecast accuracy, and the one we adopt here, is that of proper scoring rules and, in particular, the square loss function (see ref. 1). The appealing property of proper scoring rules is that they induce a Bayesian expert to be truthful about his or her forecast.
- How to model an expert: An expert has some prior distribution over the state space and receives a private signal that he or she then uses to compute a posterior distribution using Bayes rule. All experts share a common prior, but signals are private.
- How to model an ignorant aggregator: An aggregator is ignorant if his or her forecast is a function of a vector of experts' forecasts only. In particular, an ignorant aggregator's forecast is not a function of the underlying information structure. Obviously, his or her forecast is inferior compared with some hypothetical omniscient expert. This omniscient expert knows the vector of forecasts and in addition knows the information structure and the vector of private signals observed by the experts. The score attained by the hypothetical omniscient expert's forecast serves as a benchmark for our ignorant aggregator.
- How to evaluate an ignorant aggregator's performance: In an ideal world, the ignorant aggregator would have access to all relevant information, and so he or she would produce an optimal forecast using Bayes rule. This hypothetical forecast is no other than the one produced by our omniscient expert. Thus, the cost of being ignorant is the expected marginal performance associated with the difference in information between the ignorant aggregator and the omniscient expert, where the expectation is taken with respect to the information available, ex ante, to the omniscient expert.

Finally, note that the proposed evaluation measure depends on the specific information structure, which is unknown at the outset. This begs the question of which information structure should be used. Here we adopt the robust (or, equivalently, the

Significance

The problem of conflicting advice from multiple experts is common to policy makers, corporate governors, patients facing medical prognosis, individuals requiring financial advice, and more. We ask, How should an ignorant advisee (one that observes experts' forecasts only) aggregate information when this is provided in probabilistic terms (such as forecasts over events)? We propose a robustness criterion based on the classical notions of scoring rule and regret. Under reasonable assumptions on the underlying information structure of the experts, we provide formulas that allow an ignorant aggregator to perform almost as well as an omniscient expert (one that aggregates perfectly all of the information) whenever there are two experts. We also show that this is hopeless when facing many experts.

Author contributions: I.A., Y.B., and R.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: yakovbab@technion.ac.il, iarieli@technion.ac.il, or rann@technion.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1813934115/-/DCSupplemental.

Published online December 11, 2018.

ECONOMIC SCIENCES

adversarial) approach. We say that the ignorant aggregator can guarantee a regret of α in a given class of information structures if his or her relative loss is at most α, for all information structures in the given class.

In addition to its inherent merit, the reference to the omniscient expert as a benchmark provides two other methodological advantages. First, as this benchmark is conceivably the strictest, any upper bound (e.g., theorems 1 and 2) obtained on the regret with respect to the omniscient expert implies the same bound for all other competitive benchmarks. Second, it turns out that the lion's share of the analysis carried out for the omniscient-expert benchmark carries over to other reasonable benchmarks, as for instance the best expert benchmark (see *SI Appendix*, Section A).

One important aspect of our model is that it pertains to a single interaction. In particular, the aggregator has no prior experience with the experts, and he or she cannot observe past realizations. We argue that this single interaction condition is realistic in some settings. Aggregating prognoses from physicians is typically a one-off challenge, for example. Similarly, aggregating economic forecasts is important when deciding on a mortgage, and for many of us, this is the only time it is called for. However, even if aggregators repeatedly interact with experts and have an opportunity to learn, there is always the challenge of the first interaction, and often, typically with publicly elected officials, the first impression matters.

Our Contribution. We first demonstrate that without any restriction on the plausible information structure, there is no hope for an aggregator to perform well (see proposition 1). We provide an example where the ignorant aggregator cannot guarantee any regret below $\frac{1}{4}$, which, in turn, is trivially achievable by constantly announcing $\frac{1}{2}$ irrespective of experts' forecasts (recall that we measure accuracy by the square loss function). Thus, to allow any hope for positive results, we impose additional restriction on the plausible information structure.

The first family of information structures we consider is when there are only two experts, of which one is better than the other, yet the ignorant aggregator does not know which of the two it is. Following Blackwell (2), we say that two experts are Blackwell-ordered whenever one is strictly better informed than the other. To perform as well as the omniscient expert, one only needs to know which of the two experts is better. It turns out that even without this information, much can be done. Theorem 1 provides an optimal aggregation scheme for this setting (Eq. **2**) with a regret of $\frac{1}{8}(5\sqrt{5}-11)\approx 0.0225$.

We then study the case where the experts' signals are distributed independently conditional on the realized state. It is well known that in such an environment having only knowledge of the common prior allows for Bayesian aggregation of the experts' forecasts. The ignorant aggregator, not knowing even the prior, can nevertheless perform quite well. To do so, he or she uses the following natural approach. He or she "guesses" a prior and performs the aggregation as if the guess equaled the actual prior. One particular guess is to use an average of the two posteriors, a scheme we refer to as the average prior aggregation scheme (see Eq. **3**). This, apparently, results in a regret of 0.0260 (see theorem 2). Although one can actually do better (see proposition 2), the same theorem shows it is almost optimal as the regret is bounded below by $\frac{1}{8}(5\sqrt{5}-11)\approx 0.0225$, the exact same regret as that of the Blackwell-order setting. We discuss the gap of $0.0260-0.0225=0.0035$ and some related conjectures in *SI Appendix*, Section 2.

When the number of experts, n, who receive conditionally independent and identically distributed (i.i.d.) signals grows, the performance of the ignorant aggregator deteriorates. Theorem 4 shows that the regret bound approaches $\frac{1}{4}$ as $n\rightarrow\infty$. This suggests that one cannot aggregate the information into some intelligent forecast other than $\frac{1}{2}$. In other words, for the worst-case information structure, the aggregator's (approximately) best course of action is to ignore the forecasts and predict $\frac{1}{2}$, which guarantees him a loss of $\frac{1}{4}$. There is no procedure that significantly improves upon this one. Obviously, the ignorant aggregator decreases his or her loss with more available forecasts. However, the omniscient expert is far better equipped to take advantage of these additional forecasts and, as the proof demonstrates, improves at a much faster rate. This negative result has a conceptual implication as it highlights the significance of the common prior assumption in models of information aggregation in large markets. Even if all experts share a common prior and that fact itself is known to the aggregator, he or she may not perform better than a simple guess that ignores the data. Unlike standard settings in statistics and machine learning where additional sample data allow for improved performance, in our adversarial setting, structural data become more acute.

Our negative results indicate that an aggregator should invest in learning about the underlying information structure to produce reasonable forecasts. For example, Prelec et al. (3) propose that the aggregator implicitly learns about the information structure by eliciting additional information from the experts. In particular, experts are asked to provide their guess of the average answer of the other experts. They go on and experimentally demonstrate the benefit of their proposal.

Related Literature. Forecast aggregation is a well-studied topic within the statistics community. One approach examines how well simple (possibly natural) aggregation schemes perform. For example, refs. 4–6 study the performance of the averaging scheme, whereas others focus on what was inspired by the Bayesian paradigm. Similar to our approach, the performance of the various schemes is often evaluated in the context of some particularly defined (although empirically important) family of information structures, such as independent log-normal posteriors (e.g., refs. 7 and 8). Another approach is data-driven and uses a training set to find an optimal aggregation scheme within a given parametric family of such schemes (e.g., ref. 9). We refer the interested reader to the comprehensive reviews by refs. 10 and 11.

Our approach significantly differs from that extensive literature in a few aspects. We begin with a family of information structures and pursue optimal aggregation schemes and performance bounds without imposing any constraints on the scheme. The classes of information structures we study, Blackwell-ordered and conditionally independent forecasters, are broadly speaking wider than what is typical in the aforementioned literature. Finally, we adopt an adversarial or robust approach to evaluate the performance of an aggregation scheme. Although the adversarial approach has been used by statisticians for estimation of parameters (see, e.g., ref. 12), it has not been applied, to the best of our knowledge, to the forecast aggregation problem.

A more recent strand of literature that studies how to integrate the advice of multiple experts is offered by the machine-learning community. Such expert advice may take the form of forecasts, as we do here, but can also take the form of proposed action (such as portfolio selection in a financial market setting). The goal of these techniques is regret minimization. In that model, an ignorant aggregator (the "machine" in their jargon) repeatedly receives input from multiple experts and takes an action based on a vector of the experts' advice. At each stage, the aggregator is paid according to some function that depends on his or her action and the temporal state of nature. The "regret" measures how much worse, in hindsight, the aggregator performs compared with the best expert. The literature provides a variety of settings and schemes for choosing actions such that the average per-stage regret goes to zero. The reader is referred to ref. 13 for a review on this topic. The major distinction with our work is that it considers a repeated setting and emphasizes the learning aspect, whereas we study a one-shot model, about which the machine learning literature is mute.

Another related research topic is that of expert testing and, in particular, multiple expert testing (14–17). Typically, a naive agent faces a single or multiple set of self-proclaimed experts, who often provide conflicting advice. Rather than inquiring how the agent can aggregate the information, this strand of literature studies whether the agent can pick out the true experts from the charlatans. One natural test for ranking experts, in the context of prediction in financial markets, is that of portfolio returns. Sandroni (16) shows that indeed the better informed expert outperforms the less informed one in the long run.

In a companion paper (18), we use a similar model to study the conditions under which an aggregator can perfectly learn the state of the world (and in particular obtain a regret of zero) whenever there are many i.i.d. experts. The fact that there are circumstances where an ignorant aggregator, observing many i.i.d., forecasters, cannot identify the realized state has been pointed out in ref. 3. Forecast aggregation in a repeated setting is studied in ref. 19. They focus on the class of partial evidence information structures where each expert is exposed to a different subset of conditionally independent signals. In contrast to our negative result with many experts in the one-shot case, it is shown that an aggregator can learn to aggregate forecasts optimally under general conditions.

Model

Let $\Omega=\{0,1\}$ denote the binary state of nature. An information structure for n experts, denoted by $(S,\mathbf{P})$, consists of some n-dimensional signal space, $S=S_1\times\cdots\times S_n$, and a distribution $\mathbf{P}\in\Delta(\Omega\times S)$. Let $\mu=\mathbf{P}(\omega=1)$ denote the prior probability of the state $\omega=1$. Expert i receives a signal $s_i\in S_i$, drawn according to $\mathbf{P}$, and announces his forecast, $x_i(s_i)=\mathbf{P}(\omega=1|s_i)$ (his conditional probability for the state $\omega=1$).

We consider an ignorant aggregator who is ignorant with respect to the information structure and observes only the vector of experts' forecasts $x(s)=(x_1(s_1),\ldots,x_n(s_n))$. An aggregation scheme of n forecasts is a function $f:[0,1]^n\to[0,1]$. We study settings where the ignorant aggregator may have partial knowledge about the information structure. This partial knowledge takes the form of a subset of such structures (a class of information structures). We compare the performance of the ignorant aggregator with that of the omniscient expert—that is, an expert who knows $\mathbf{P}$ and observes all of the signals of all of the agents. Note that this is the most competitive benchmark we can set to evaluate an aggregator's performance. For a discussion of less competitive benchmarks, see *SI Appendix*, Section A.

The basic building block for evaluating the performance of the aggregators is a scoring rule, which is a function $l:[0,1]\times\Omega\to$. It assigns a loss to any pair of realization and forecast (probability of the state $\omega=1$). A proper scoring rule is a scoring rule for which the minimal expected loss is obtained when the forecast is equal to the actual distribution. Hence, a proper scoring rule incentivizes the omniscient expert to report the posterior probability. One prominent example of a proper scoring rule, which is central to our analysis, is the square loss function (1): $l(x,\omega)=(x-\omega)^2$.

Conditional on the information structure $(S,\mathbf{P})$, the ignorant aggregator can only hope to do as well as the omniscient expert. The omniscient expert's best prediction is obtained using Bayes rule and is equal to $\hat{x}(s)=\mathbf{P}(\omega=1|s)$, where $s=(s_1,\ldots,s_n)$. Hence, the expected relative loss of the aggregation scheme f is

$$L(f,\mathbf{P})=E_{\mathbf{P}}[l(f(x(s)),\omega)-l(\hat{x}(s),\omega)].$$

Given a class of information structures, $\mathcal{C}$, the regret of the aggregation scheme f over $\mathcal{C}$ is the expected relative loss in the worst case scenario* :

$$R_{\mathcal{C}}(f)=\sup_{\mathbf{P}\in\mathcal{C}} L(f,\mathbf{P}). \qquad \textbf{[1]}$$

We start with a preliminary observation that provides an alternative formula for the relative loss of an aggregation scheme, $f:[0,1]^n\to\mathbb{R}$, given an arbitrary information structure, $\mathbf{P}$:

Lemma 1:

$$L(f,\mathbf{P})=\mathbb{E}_{(\omega,s)\sim\mathbf{P}}[(f(x_1(s_1),\ldots,x(s_n))-\hat{x}(s_1,\ldots,s_n))^2].$$

Proof: For every realized vector of signals $s=(s_1,\ldots,s_n)$,

$$\begin{aligned}
&\mathbb{E}_\omega[(f(x(s))-\omega)^2-(\hat{x}(s))-\omega)^2|s]\\
&=\mathbf{P}(\omega=1|s)[(f(x(s))-1)^2-(\hat{x}(s))-1)^2]\\
&\quad+\mathbf{P}(\omega=0|s)[(f(x_1(s)))^2-(\hat{x}(s))^2]\\
&=\hat{x}(s)[(f(x(s))-1)^2-(\hat{x}(s)-1)^2]\\
&\quad+(1-\hat{x}(s))[(f(x(s)))^2-(\hat{x}(s))^2]\\
&=[(f(x(s))^2-2\hat{x}(s)f(x(s))+(\hat{x}(s))^2]\\
&=(f(x(s))-\hat{x}(s))^2.
\end{aligned}$$

Since the equation holds for every $s=(s_1,\ldots,s_n)$, it holds also in expectation over $s=(s_1,\ldots,s_n)$.

General Information Structures

The trivial aggregation scheme, $f(x_1,x_2)=\frac{1}{2}$, ignores the forecasts made by the two experts yet guarantees a regret of $\frac{1}{4}$. Our first observation is that no other aggregation scheme can outperform this. In fact, the following is a slightly stronger result:

Proposition 1: There exists an information structure $\mathbf{P}$, such that for every aggregation scheme f, $L(f,\mathbf{P})\geq\frac{1}{4}$.

Proof: Let $S_i=\{s_i,s_i'\}$ for $i=1,2$ and let $\mathbf{P}$ be the following distribution:

$\omega=0$

	s_2	s_2'
s_1	1/4	0
s_1'	0	1/4

$\omega=1$

	s_2	s_2'
s_1	0	1/4
s_1'	1/4	0

It is easy to check that $x_i(s_i)=x_i(s_i')=\frac{1}{2}$ and $\hat{x}(s_1,s_2)=\hat{x}(s_1',s_2')=0$ and $\hat{x}(s_1,s_2')=\hat{x}(s_1',s_2)=1$. Namely, each one of the signals separately is uninformative, but together they reveal

*The term *regret* is inspired by terminology introduced by ref. 20 in the context of measuring success under a worst case scenario. In a way, this is also reminiscent of ref. 21's notion of MinMax expectation.

the state of nature. The ignorant aggregator always observes two forecasts of $\frac{1}{2}$ and has no better action than forecasting $\frac{1}{2}$. On the other hand, the omniscient expert always knows the state (with probability 1). Therefore, for every aggregation scheme f, the relative loss is at least $\frac{1}{4} - 0$.

Thus, ignorantly aggregating forecasts without any restriction on the family of information structures is impossible. What can be done when we consider special classes of information structures? Apparently, for some natural classes of information structures, there are aggregation schemes that guarantee a surprisingly low regret.

Blackwell-Ordered Experts

An interesting case to analyze in our settings is the scenario where one expert is more informed than the other.

Definition 1: Expert 1 is more informed in an information structure $(S_1, S_2, \mathbf{P})$ if $S_1 = S_1' \times S_2'$, $S_2' = S_2$, and $\mathbf{P}(s_2' = s_2) = 1$, where $s_1 = (s_1', s_2')$ is the signal of expert 1. Namely, they both observe the same signal s_2, but expert 1 observes in addition the signal s_1'.

Expert 2 is more informed in an information structure $(S_1, S_2, \mathbf{P})$ if $S_2 = S_1' \times S_2'$, $S_1' = S_1$, and $\mathbf{P}(s_1' = s_1) = 1$, where $s_1 = (s_1', s_2')$ is the signal of expert 2.

An information structure $(S_1, S_2, \mathbf{P})$ is Blackwell-ordered if one of the experts is more informed. Let $\mathcal{BO}$ denote the set of all Blackwell-ordered information structures.

Specifically, the better informed expert has access to the signal available to the less informed expert, and he or she receives an additional private signal. This notion is equivalent to the notion of Blackwell domination and Blackwell ordering (see ref. 2).

Similarly to Eq. **1**, we define the regret of an aggregation scheme f in the Blackwell environment to be

$$R_{\mathcal{BO}}(f) = \sup_{\mathbf{P} \in \mathcal{BO}} L(f, \mathbf{P}).$$

To gain some intuition about the problem, we study the regret of two simple and naive aggregation schemes.

Naive Aggregation Schemes.

The DeGroot scheme. Consider the naive aggregation scheme $f(x_1, x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2$, which coincides with the celebrated DeGroot opinion formation function (see ref. 4). Recall that our criterion of success computes the regret under an adversarial information structure. Consider the following information structure: Assume that the prior is $\mu = \frac{1}{2}$ and that expert 1 receives no additional information while expert 2 learns the realized state ω—namely, expert 1 is perfectly uninformed, while expert 2 is perfectly informed. The resulting pair of forecasts will be $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, 1)$, each with probability $\frac{1}{2}$. The aggregator's forecast, under the naive DeGroot scheme, will be either $\frac{1}{4}$ or $\frac{3}{4}$, each with probability $\frac{1}{2}$. In both cases, the forecast will differ by $\frac{1}{4}$ from that of the better expert, and hence, the regret in the square loss utilities is at least $\frac{1}{16} = 0.0625$.[†]

The minimal entropy scheme. In a Bayesian framework, whenever one of the experts' forecasts is extreme [$x_i \in \{0, 1\}$], he is correct with probability 1 and the aggregator should adopt his forecast independently of the information structure. A naive generalization of this is to follow the expert whose forecast is more informative, in terms of entropy. This implies adopting the more extreme forecast. Formally:

$$f(x_1, x_2) = \begin{cases} x_1 \text{ if } \left|x_1 - \frac{1}{2}\right| > \left|x_2 - \frac{1}{2}\right| \\ x_2 \text{ otherwise.} \end{cases}$$

As it turns out, this aggregation scheme does not always perform well. To see this, we first note that by the splitting lemma of Aumann and Maschler (22), there is an identification between Blackwell-ordered information structures and martingales (X_0, X_1, X_2) of posteriors. $X_0 = \mu$ is the prior, X_1 is the posterior of the less informed expert, and X_2 is the posterior of the more informed expert. Consider the posterior belief martingale (X_0, X_1, X_2) with expectation $X_0 = \frac{1}{2}$ and where $X_1 = 0.2, 0.8$ with equal probabilities. The conditional probabilities for X_2 are $P(X_2 = 0 | X_1 = 0.2) = \frac{5}{7}$, $P(X_2 = 0.7 | X_1 = 0.2) = \frac{2}{7}$, and symmetrically, $P(X_2 = 1 | X_1 = 0.8) = \frac{5}{7}$, $P(X_2 = 0.3 | X_1 = 0.8) = \frac{2}{7}$. Fig. 1 visualizes this martingale.

In this information structure with probability $\frac{1}{7}$, the ignorant aggregator observes the pair of forecasts $(0.2, 0.7)$. Based on f, the ignorant aggregator predicts the more extreme forecast 0.2, whereas the omniscient expert forecast is 0.7. Symmetrically, with probability $\frac{1}{7}$, the ignorant aggregator observes the pair $(0.8, 0.3)$ and predicts 0.8, which, once again, is 0.5 away from the forecast of the better informed expert. Thus, the induced regret is at least $\frac{2}{7} \cdot \frac{1}{4} \approx 0.0714$, which is even worse than that of the simple average aggregation scheme.

Optimal Aggregation. The analysis of the two naive forecast aggregation schemes and the corresponding information structures suggests that a regret-minimizing aggregation scheme should assign weights to the forecasts that do depend on their distance from $\frac{1}{2}$ (greater distance translates to more weight) but not too radically. The formula of the precision scheme, which we now discuss, follows this intuition. We denote by $\phi(x) = \frac{1}{x(1-x)}$

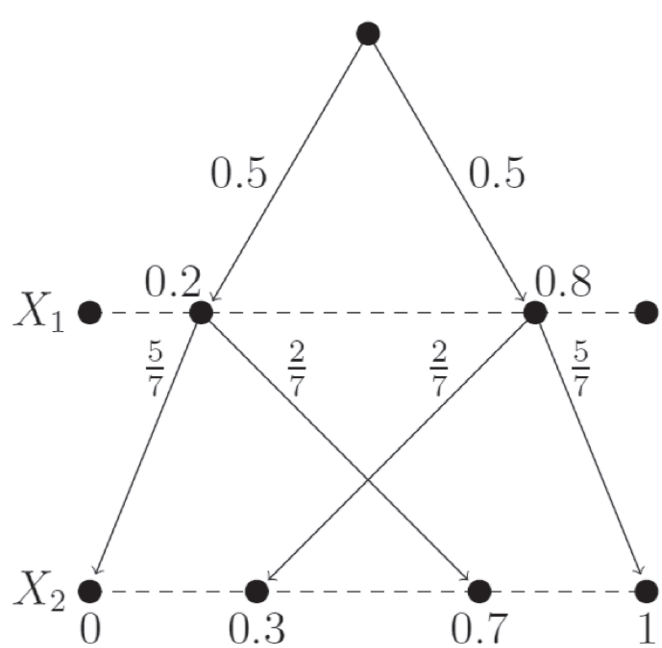

Fig. 1. The martingale X_1, X_2.

[†]In fact, by solving the appropriate optimization function, it follows that this information structure leads to the worst case relative loss, and so the regret of the DeGroot scheme is exactly $\frac{1}{16} = 0.0625$. The underlying intuition for this is that, on the one hand, the omniscient expert incurs no loss due to the existence of a perfectly informed expert, while on the other hand, we maximize the loss of the DeGroot scheme by maximizing the gap between the two experts. This is done by introducing a completely uninformed expert.

the precision of a forecast x.[‡] The idea is to assign weights to the two forecasts proportional to their precision. More concretely, the definition of the precision scheme is given in Eq. **2**.

$$f_{pre}(x_1, x_2) = \begin{cases} \frac{\phi(x_1)}{\phi(x_1)+\phi(x_2)} x_1 + \frac{\phi(x_2)}{\phi(x_1)+\phi(x_2)} x_2 & \text{if } |x_1 - x_2| \le 0.4 \\ \frac{\sqrt{\phi(x_1)}}{\sqrt{\phi(x_1)}+\sqrt{\phi(x_2)}} x_1 + \frac{\sqrt{\phi(x_2)}}{\sqrt{\phi(x_1)}+\sqrt{\phi(x_2)}} x_2 & \text{if } |x_1 - x_2| > 0.4. \end{cases} \quad [2]$$

To complete the definition whenever the precision is not a well-defined set, $f_{pre}(0, x_2) = f_{pre}(x_1, 0) = 0$ whenever $x_1, x_2 < 1$ and $f_{pre}(1, x_2) = f_{pre}(x_1, 1) = 1$ whenever $x_1, x_2 > 0$. Finally, set $f_{pre}(0, 1) = f_{pre}(1, 0) = \frac{1}{2}$.[§]

A clear peculiarity of the precision scheme is the role of separating the case where $|x - y| < 0.4$. We discuss and explain this in *SI Appendix*, Section 1.

Our main result for this section is the following:

Theorem 1: For Blackwell-ordered information structures, the precision scheme guarantees a regret of $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225425$. Moreover, no aggregation scheme guarantees a lower regret. In other words, $R_{\mathcal{BO}}(f_{pre}) = \frac{1}{8}(5\sqrt{5} - 11) \le R_{\mathcal{BO}}(f)$ for any aggregation scheme f.

Note that the interaction can be modeled as a zero-sum game between an ignorant aggregator (who chooses f) and an adversary (who chooses $\mathbf{P}$). The proof relies on an explicit formulation of the maxmin strategies of both the adversary and the aggregator in this zero-sum game. Once formulated, the proof is relatively easy, as it is then sufficient to show that the presented strategies guarantee the value $\frac{1}{8}(5\sqrt{5} - 11)$ for both sides. In the *SI Appendix*, section 1, we provide some intuition as to how we derived these maxmin strategies.

Proof: We start by presenting an optimal strategy for the adversary. Namely, we present a distribution over two Blackwell-ordered information structures, such that an aggregator who knows the mixed strategy of the adversary cannot achieve a regret below $\frac{1}{8}(5\sqrt{5} - 11)$. This obviously implies that our ignorant aggregator cannot achieve a better regret either.

We set the before $\mu = \frac{1}{2}$. The less informed expert receives one of two signals that yield posteriors of $x \in (0, \frac{1}{2})$ and $1 - x$ with equal probability $\frac{1}{2}$; that is, the less informed agent observes a noisy binary signal that is compatible with the correct state with probability $1 - x$. Conditional on the posterior x, the more informed expert observes an additional signal that yields posteriors of 0 and $1 - x$ with probabilities $\frac{1-2x}{1-x}$ and $\frac{x}{1-x}$, respectively. Such an information structure exists by the Aumann–Maschler splitting lemma (22). Symmetrically, conditional on the posterior $1 - x$, the more informed expert observes an additional signal that yields posteriors of x and 1 with probabilities $\frac{x}{1-x}$ and $\frac{1-2x}{1-x}$, respectively. Fig. 2 demonstrates the martingale of posteriors for the less and more informed experts.

Now consider the mixed strategy where the more informed expert is chosen to be expert 1 or expert 2 with equal probability $\frac{1}{2}$. In the case where the ignorant aggregator observes the pair of forecasts $\{x, 1 - x\}$, which occurs with probability $\frac{x}{1-x}$, he or she does not know who the better informed expert is. In fact, he or she assigns equal probability $\frac{1}{2}$ to the event that expert $i = 1, 2$ is the more informed expert. Therefore, his or her optimal prediction in such a case is $\frac{1}{2}$, which is $(\frac{1}{2} - x)$-far from the omniscient expert forecast. Thus, the relative loss of any aggregation scheme against this mixed strategy is at least $\frac{x}{1-x}(\frac{1}{2} - x)^2$. Maximizing over $x \in (0, \frac{1}{2})$ yields a regret of $\frac{1}{8}(5\sqrt{5} - 11)$, which is obtained for $x = \frac{1}{4}(3 - \sqrt{5})$.

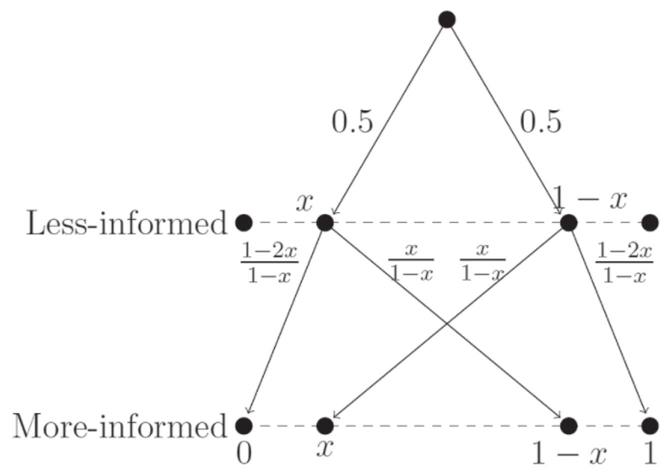

Fig. 2. The martingale of posteriors.

We now prove that the average prior scheme guarantees a regret of at most $\frac{1}{8}(5\sqrt{5} - 11)$. Note that the precision scheme is anonymous—namely, $f_{pre}(x_1, x_2) = f_{pre}(x_2, x_1)$. Therefore, the adversary's best reply against the precision scheme contains an information structure, where expert 2 is the more informed expert. Henceforth, we restrict attention to such information structures. Now the adversary's strategy can be viewed as a martingale X_0, X_1, X_2 of length 2, where $X_0 = \mu$ is the prior and X_i is the posterior of expert i. By Lemma 1, the relative loss is given by $L(f_{pre}, (X_0, X_1, X_2)) = \mathbb{E}_{x_i \sim X_i}[(f_{pre}(x_1, x_2) - x_2)^2]$. The payoff L is a convex combination of payoffs with a fixed realization x_1 when the weights of the payoffs are according to the distribution X_1. Therefore, we can assume without loss of generality that the adversary's best reply is such that X_1 is a Dirac measure or equivalently $X_0 = X_1 = \mu$ with probability 1.

For $x < y < z$, we denote by $M_{x,y,z}$ the martingale where $X_0 = X_1 = y$ with probability 1, $X_2 = x$ with probability $\frac{z-y}{z-x}$, and $X_2 = z$ with probability $\frac{y-x}{z-x}$ (see Fig. 3). Note that the set of martingales of length 2 is a convex set whose extreme points are exactly $\{M_{x,y,z}\}_{0 \le x \le y \le z \le 1}$. Moreover, note that the adversary's utility is linear in the representation of the martingale. Namely, for a martingale that is given by the convex combination $M = \sum_{(x,y,z) \in W} \alpha_{x,y,z} M_{x,y,z}$, we have

$$L(M, f) = \sum_{(x,y,z) \in W} \alpha_{x,y,z} L(M_{x,y,z}, f).$$

Therefore, given the aggregation scheme f_{pre}, the adversary has a best reply to f_{pre} of the form $M_{x,y,z}$. From this, we deduce that

$$R_{\mathcal{BO}}(f_{pre}) = \sup_{(x,y,z) \in [0,1]^3 : x \le y \le z} L(M_{x,y,z}, f_{pre}).$$

Consider the following four compact ranges in $[0, 1]^3$:

$$K_1 = \{(x, y, z) | x \le y \le z, \ |x - y| \ge 0.4, \ |z - y| \ge 0.4\}$$
$$K_2 = \{(x, y, z) | x \le y \le z, \ |x - y| \le 0.4, \ |z - y| \ge 0.4\}$$
$$K_3 = \{(x, y, z) | x \le y \le z, \ |x - y| \ge 0.4, \ |z - y| \le 0.4\}$$
$$K_4 = \{(x, y, z) | x \le y \le z, \ |x - y| \le 0.4, \ |z - y| \le 0.4\}.$$

[‡]In statistics, the precision of a random variable is the reciprocal of the variance. The forecast x means that the experts believe that the state is a Bernoulli random variable with a success probability x. Thus, $\phi(x) = \frac{1}{x(1-x)}$ is the precision of the forecast.

[§]Note that the probability that the experts' forecasts are either (1, 0) or (0, 1) is zero in any information structure.

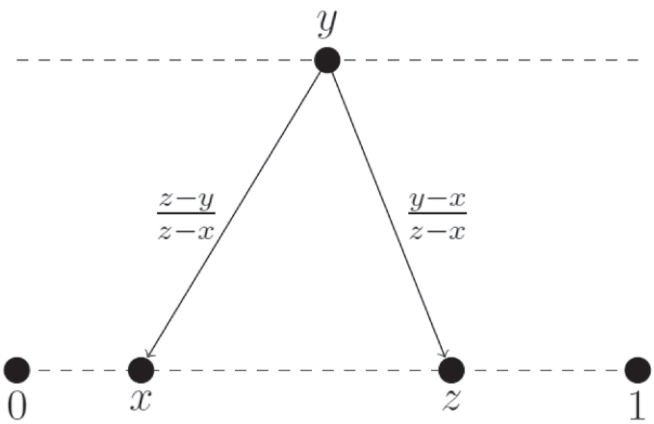

Fig. 3. $M_{x,y,z}$, the extreme points of the class of martingales.

We note that for every triplet (x, y, z) where $x \le y \le z$, there exists $1 \le m \le 4$ such that $(x, y, z) \in K_m$ and the relative loss of f on each of the K_m is determined by a fixed loss function. For example, for (x, y, z) in K_1, we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{P}_{x,y,z}}[l(x_2, \omega) - l(f_{pre}(x_1, x_2), \omega)]\\&=\frac{z-y}{z-x}\left(\frac{\sqrt{x(1-x)}y + \sqrt{y(1-y)}x}{\sqrt{x(1-x)} + \sqrt{y(1-y)}} - x\right)^2\\&\quad+\frac{y-x}{z-x}\left(\frac{\sqrt{y(1-y)}z + \sqrt{z(1-z)}y}{\sqrt{y(1-y)} + \sqrt{z(1-z)}} - z\right)^2.\end{aligned}$$

Similarly, in range K_2, we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{P}_{x,y,z}}[l(x_2, \omega) - l(f_{pre}(x_1, x_2), \omega)]\\&=\frac{z-y}{z-x}\left(\frac{x(1-x)y + y(1-y)x}{x(1-x) + y(1-y)} - x\right)^2\\&\quad+\frac{y-x}{z-x}\left(\frac{\sqrt{y(1-y)}z + \sqrt{z(1-z)}y}{\sqrt{y(1-y)} + \sqrt{z(1-z)}} - z\right)^2.\end{aligned}$$

Similar expressions can be obtained for ranges K_4 and K_3. Thus, to show that f_{pre} guarantees a regret of at most $\frac{1}{8}(5\sqrt{5} - 11)$ to the ignorant aggregator, one need only solve four 3D optimization problems in the four compact domains $\{K_i\}_{i=1,2,3,4}$. These optimization problems have been solved numerically by Matlab, which shows that the global maximum of $R(x, y, z)$ is obtained at two points $(0, \frac{1}{4}(3 - \sqrt{5}), 1 - \frac{1}{4}(3 - \sqrt{5}))$ and $(\frac{1}{4}(3 - \sqrt{5}), 1 - \frac{1}{4}(3 - \sqrt{5}), 1)$ and is equal to $\frac{1}{8}(5\sqrt{5} - 11)$.

Two Conditionally Independent Experts

Another family of information structures that is prevalent in the economics literature is that of experts who receive independent signals, conditional on the realized state. We refer to this as a conditionally independent information structure. In such settings, knowing the prior $\mu = \mathbf{P}(\omega = 1)$ together with the posteriors $x_1, x_2, \ldots, x_n$ is sufficient for applying the optimal Bayesian aggregation (that of the omniscient expert).

It is straightforward to verify the following (see, e.g., ref. 23): If $\mathbf{P}(\omega = 1|s_i) = x_i$ for $1 \le i \le n$, then

$$\begin{aligned}\mathbf{P}(\omega = 1|s_1, \ldots, s_n) &= g(\mu, x_1, \ldots, x_n)\\&= \frac{(1-\mu)^{n-1}\prod_{i=1}^n x_i}{(1-\mu)^{n-1}\prod_{i=1}^n x_i + \mu^{n-1}\prod_{i=1}^n (1-x_i)}.\end{aligned}$$

The ignorant aggregator, however, does not know the prior. Nevertheless, this observation induces a natural family of aggregation schemes, where Bayes rule is applied to a "dummy" prior (or a guess). The guess of the prior can be based on the experts' forecasts. It turns out that the resulting regret is surprisingly low when the number of experts is 2 and the aggregator uses the standard average of the two forecasts as his or her guess for the prior. This guess entails the following aggregation scheme, which we refer to as the average-prior scheme:

$$\begin{aligned}f_{avg}(x_1, x_2) &= g\left(\frac{x_1 + x_2}{2}, x_1, x_2\right)\\&= \frac{x_1 x_2 (1 - \frac{x_1+x_2}{2})}{x_1 x_2 (1 - \frac{x_1+x_2}{2}) + (1-x_1)(1-x_2)\frac{x_1+x_2}{2}}.\end{aligned} \quad [3]$$

Let $\mathcal{CI}$ be the class of all independent information structures for two experts.

Theorem 2: $R_{\mathcal{CI}}(f_{avg}) = 0.0260$—that is, the average prior scheme guarantees a regret of 0.0260. Moreover, for every aggregation scheme f, $R_{\mathcal{CI}}(f) \ge \frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225$. That is, no aggregation scheme guarantees a regret lower than $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225$.

In fact, our lower bound on the regret is stronger. Even for the more restricted class of conditionally i.i.d. information structures (not only conditionally independent), there is no aggregation scheme that guarantees a regret below $\frac{1}{8}(5\sqrt{5} - 11)$.

Proof: We begin by introducing a strategy for the adversary that guarantees him a regret of at least $\frac{1}{8}(5\sqrt{5} - 11)$. We present a distribution over two conditionally i.i.d. information structures, such that an aggregator who knows the mixed strategy of the adversary cannot achieve a regret below $\frac{1}{8}(5\sqrt{5} - 11)$.

This obviously implies that our ignorant aggregator cannot achieve a better regret either.

Consider the mixed strategy that randomizes uniformly over the following two conditionally independent structures. In the first information structure, the prior is $x \in (0, \frac{1}{2})$. The signals are conditionally identically distributed signals that induce their posterior belief 0 and $\frac{1}{2}$ with probabilities $1 - 2x$ and $2x$, respectively [by Aumann and Maschler's splitting lemma (22), such a signal structure exists]. In the second information structure, the prior is $1 - x \in (\frac{1}{2}, 1)$. The posterior beliefs are 1 and $\frac{1}{2}$ with probabilities $1 - 2x$ and $2x$, respectively. A visualization of the adversary's mixed strategy appears in Fig. 4.

When the realizations of the two forecasts turn out to be $x_1 = x_2 = \frac{1}{2}$, an ignorant aggregator does not know whether the prior was x or $1 - x$, while an omniscient expert does. Simple symmetry arguments show that the best forecast for an ignorant aggregator in such an event is $\frac{1}{2}$, whereas an omniscient expert forecasts

$$\frac{\frac{1}{2}\frac{1}{2}(1-x)}{\frac{1}{2}\frac{1}{2}(1-x) + \frac{1}{2}\frac{1}{2}x} = 1 - x \text{ and } \frac{\frac{1}{2}\frac{1}{2}x}{\frac{1}{2}\frac{1}{2}x + \frac{1}{2}\frac{1}{2}(1-x)} = x,$$

depending on whether the prior was x or $1 - x$, respectively. By Lemma 1, the relative loss in the event of $x_1 = x_2 = \frac{1}{2}$ is $(\frac{1}{2} - x)^2$. A simple calculation shows that the probability of the event $x_1 = x_2 = \frac{1}{2}$ is $\frac{x}{1-x}$. Therefore, the relative loss of an aggregator who knows the adversary's strategy is

$$\frac{x}{1-x}\left(\frac{1}{2} - x\right)^2.$$

Maximizing the relative loss over $x \in (0, \frac{1}{2})$ yields a regret of $\frac{1}{8}(5\sqrt{5} - 11)$, which is obtained at $x = \frac{1}{4}(3 - \sqrt{5})$. This proves that no aggregation scheme can guarantee a regret below $\frac{1}{8}(5\sqrt{5} - 11)$.

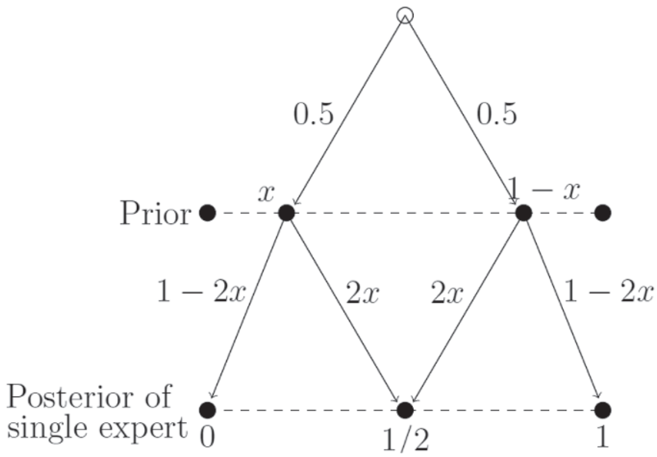

Fig. 4. Distribution over priors and posteriors of a single agent.

Next we shall prove that any strategy of the adversary against the average prior scheme yields a relative loss of at most 0.0260 to the aggregator. An adversary can be viewed as a triple $(\mu, \mathbf{P}^1, \mathbf{P}^2)$ where μ is the prior and $\mathbf{P}^i \in \Delta([0,1])$ is a distribution of expert i's posterior beliefs with expectation $\mathbb{E}(\mathbf{P}_i) = \mu$. Assume that $\mathbf{P}_i$ assigns a prior probability of p_i to the posterior x_i. Simple calculations show that the probability of the pair of posteriors being (x_1, x_2) is given by the following expression that is multilinear in p_1 and p_2:

$$h(p_1, p_2, \mu, x_1, x_2) = p_1 p_2 \left(\frac{(1-x_1)(1-x_2)}{1-\mu} + \frac{x_1 x_2}{\mu} \right).$$

By Lemma 1, the relative loss of the aggregator in the case where (x_1, x_2) is the realized posterior probability is

$$r(\mu, x_1, x_2) = \left(\frac{x_1 x_2 (1-\mu)}{x_1 x_2 (1-\mu) + (1-x_1)(1-x_2)\mu} - \frac{x_1 x_2 (1 - \frac{x_1+x_2}{2})}{x_1 x_2 (1 - \frac{x_1+x_2}{2}) + (1-x_1)(1-x_2)\frac{x_1+x_2}{2}} \right)^2. \quad [4]$$

For every $y \le \mu \le z$, let $\mathbf{P}^i_{\mu,y,z}$ be a posterior distribution with support $\{y, z\}$ such that x is realized with probability $\frac{z-\mu}{z-y}$ and y is realized with probability $\frac{\mu-y}{z-y}$. These are the extreme points of the convex set of all posterior distributions. Namely, any distribution $\mathbf{P}^i$ can be written as $\mathbf{P}^i = \sum_j \alpha^i_j \mathbf{P}^i_{\mu, y^i_j, z^i_j}$, where $\sum_j \alpha^1_j = \sum_j \alpha^2_j = 1$. The multilinearity of the relative loss implies the following formula on the expected relative loss:

$$L(f_{avg}, (\mu, \mathbf{P}^1, \mathbf{P}^2)) = \sum_{j,k} \alpha^1_j \alpha^2_k L((\mu, \mathbf{P}^1_{\mu, y^1_j, z^1_j}, \mathbf{P}^2_{\mu, y^2_k, z^2_k}), f_{avg}), \quad [5]$$

which is a convex combination of relative losses of the form $L(f_{avg}, (\mu, \mathbf{P}^1_{\mu,y^1,z^1}, \mathbf{P}^2_{\mu,y^2,z^2}))$. Therefore, the regret that an adversary can achieve against the average aggregation scheme is given by

$$\max_{\mu, y_1 \le \mu \le z_1, y_2 \le \mu \le z_2} L((\mu, \mathbf{P}^1_{\mu,y^1,z^1}, \mathbf{P}^2_{\mu,y^2,z^2}), f_{avg}).$$

Note also that the objective function has a closed formula:

$$\begin{aligned} &L((\mu, \mathbf{P}^1_{\mu,y^1,z^1}, \mathbf{P}^2_{\mu,y^2,z^2}), f_{avg}) \\ &= h\left(\frac{z_1-\mu}{z_1-y_1}, \frac{z_2-\mu}{z_2-y_2}, \mu, y_1, y_2 \right) \cdot r(\mu, y_1, y_2) \\ &+ h\left(\frac{z_1-\mu}{z_1-y_1}, \frac{\mu-y_2}{z_2-y_2}, \mu, y_1, z_2 \right) \cdot r(\mu, y_1, z_2) \\ &+ h\left(\frac{\mu-y_1}{z_1-y_1}, \frac{z_2-\mu}{z_2-y_2}, \mu, z_1, y_2 \right) \cdot r(\mu, z_1, y_2) \\ &+ h\left(\frac{\mu-y_1}{z_1-y_1}, \frac{\mu-y_2}{z_2-y_2}, \mu, z_1, z_2 \right) \cdot r(\mu, z_1, z_2). \end{aligned} \quad [6]$$

In summary, we have shown that, given the aggregation scheme f, the adversary's best reply problem can be reduced to a concrete maximization problem over five parameters, μ, y_1, y_2, z_1, z_2. Numerical Matlab calculations show that the global maximum of this five-variable fraction is obtained at the point $\mu = 0.120$, $y_1 = 0.120$, $z_1 = 0.120$, $y_2 = 0$, and $z_2 = 0.746$ and is equal to 0.0260.[¶]

Mind the Gap. Theorem 2 leaves a gap between the upper and lower regret bounds whenever the information structure is conditionally independent. We are not able to close this gap; however, we can slightly improve the upper bound by using a nonintuitive variant of the average prior scheme. That is, instead of updating the two posteriors with respect to their average, we determine the dummy prior as follows:

$$ep(x_1, x_2) = \begin{cases} 0.49x_1 + 0.49x_2 & \text{if } x_1 + x_2 \le 1 \\ 0.49x_1 + 0.49x_2 + 0.02 & \text{otherwise.} \end{cases} \quad [7]$$

Proposition 2: For conditionally independent information structures, the aggregation scheme, $f(x_1, x_2) = g(ep(x_1, x_2), x_1, x_2)$, guarantees a regret of 0.0250.

The proof of proposition 2, which bears a similarity to the proof of the first part of theorem 2, is relegated to *SI Appendix*, Section 2.

In many economic models, signals, in addition to being conditionally independent, are also identical. For this case, we conjecture that the regret bound of $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225$ is tight and that the average prior scheme is indeed optimal.

Conjecture 3: For conditionally i.i.d. information structures, the minimal regret that can be guaranteed is equal to $\frac{1}{8}(5\sqrt{5} - 11) \approx 0.0225$, and the average prior scheme guarantees this regret.

We discuss this conjecture further in *SI Appendix*, Section 2.

Many Conditionally Independent Experts

We turn to study how the regret of the ignorant aggregator is affected as the number of independent experts grows. Obviously, when the ignorant aggregator observes more forecasts of experts with conditionally i.i.d. signals, his or her expected loss decreases (the aggregator can choose to ignore the forecasts of additional experts). However, this does not imply that the guaranteed regret decreases. We recall that the ignorant aggregator is benchmarked against the omniscient expert whose loss decreases as well. Theorem 4 shows that the omniscient aggregator's loss shrinks much faster in the worst case information structure and, in fact, the regret bound approaches $\frac{1}{4}$ as $n \to \infty$. In particular, this implies that the loss of the omniscient expert goes to 0 (roughly speaking, he or she identifies the state), while the ignorant aggregator remains completely uncertain about the state

¶Note that the adversarial best reply corresponds to an information structure where the first expert receives no information.

(roughly speaking, he or she has no intelligent forecast other than $\frac{1}{2}$). This result is in contrast to the positive result for the two i.i.d. experts (theorem 2). In case of two experts, whenever the ignorant aggregator is clueless, so is (almost) the omniscient expert.

Let $\mathcal{D}_n$ be the class of all i.i.d. information structures with n experts. The interplay between the improvement of the ignorant aggregator and that of the omniscient expert is given in the following theorem.

Theorem 4: For any number of conditionally independent and identical experts, n, and any aggregation scheme, $f:[0,1]^n \to \mathbb{R}$, the following bound on regret holds: $R_{\mathcal{D}_n}(f) \geq \frac{1}{4} - 3\sqrt{\frac{\log n}{n}}$.

Thus, as the number of agents grows, no aggregation scheme can guarantee a performance that does better than the fixed scheme that always forecasts 0.5. This stands in sharp contrast to the $n=2$ case. The proof of theorem 4 is relegated to *SI Appendix*, Section 3. Below we discuss the key ideas of the proof.

Idea of the Proof of Theorem 4. We start the discussion with the following example (which appears also in refs. 3 and 18). Consider the following two information structures for a single expert: The priors in these two information structures are $\mu_1 = \frac{1}{2}$ and $\mu_2 = \frac{7}{10}$, respectively.

Table 1. The information structures *I*(1), *I*(2)

$I(1)$:		$\omega=0$	$\omega=1$
	s_0	3/8	1/8
	s_1	1/8	3/8

$I(2)$:		$\omega=0$	$\omega=1$
	s_0	3/40	1/40
	s_1	9/40	27/40

Let us denote by $(e.m.\omega)$, where $m=1,2$ and $\omega=0,1$, the event that the information structure is $I(m)$ and the state is ω. A straightforward computation shows that the distribution over the experts' posteriors (forecasts) is identical for the two events $(e.1.1)$ and $(e.2.0)$. In both cases, every expert will either observe s_0 with probability $\frac{1}{4}$ and forecast $\frac{1}{4}$ or will observe s_1 with probability $\frac{3}{4}$ and will forecast $\frac{3}{4}$. Let us denote this posterior distribution over forecasts by ψ.

Consider an adversary's mixed strategy over these two information structures with weights $\frac{1-\mu_2}{1-\mu_2+\mu_1} = \frac{3}{8}$ assigned to information structure $I(1)$ and $\frac{\mu_1}{1-\mu_2+\mu_1} = \frac{5}{8}$ to information structure $I(2)$. Our ignorant aggregator cannot perform better than a hypothetical aggregator, who knows the two information structures, adversary's mixed strategy, and the forecasts (note that, unlike the omniscient expert, this hypothetical aggregator cannot observe the actual signals and does not know the realization of the adversary's mixed strategy). We call such an aggregator "Bayesian."

Assume that n is large enough so that the empirical distribution of the sample of n forecasts is "essentially" equal to the precise posterior distribution. Thus, in events $(e.1.1)$ and $(e.2.0)$, the hypothetical aggregator will observe ψ and his or she Bayesian belief about the event $\omega=1$ will be $\frac{3/8\cdot\mu_1}{3/8\cdot\mu_1+5/8\cdot(1-\mu_2)} = \frac{1}{2}$. In other words, the adversary can set the probabilities over $I(1)$ and $I(2)$ such that even the hypothetical aggregator who observes the distribution over posteriors precisely will have complete uncertainty about the state in cases $(e.1.1)$ and $(e.2.0)$. This is especially true for the ignorant aggregator.

Unfortunately for the adversary, there are two additional cases, $(e.1.0)$ and $(e.2.1)$, where the hypothetical aggregator succeeds in determining the state. However, if we now introduce a third information structure, $I(3)$, such that the posteriors in the events $(e.2.1)$ and $(e.3.0)$ coincide, then a mixed strategy over the three information structures can be constructed such that the same uncertainty will prevail when this posterior is observed.

For concreteness, the new information structure, $I(3)$, is given in Table 2, and the probabilities over $I(1), I(2), I(3)$ are $\frac{9}{65}, \frac{15}{65}, \frac{41}{65}$.

Table 2. The information structure *I*(3)

$I(3)$:		$\omega=0$	$\omega=1$
	s_0	3/328	1/328
	s_1	81/328	243/328

We proceed to define the information structure $I(m)$ and the corresponding mixed strategy iteratively such that the hypothetical aggregator faces complete uncertainty unless the events $(e.1.0)$ and $(e.m.1)$ are realized. In this construction, the posterior of all of the experts in all of the information structures is either $\frac{1}{4}$ or $\frac{3}{4}$.

It turns out that in order for the probability of the events $(e.1.0)\cup(e.m.1)$ to vanish, we should repeat this construction with the pair of forecasts $\frac{1}{2}\pm\epsilon$ (for sufficiently small $\epsilon>0$) instead of $\{\frac{1}{4},\frac{3}{4}\}$.

On the other hand, in order for the omniscient expert to perform well by observing a sample from the distribution, the value of ϵ cannot be too small (note that if we set $\epsilon=0$, all of the experts' forecasts will be equal to $\frac{1}{2}$, in which case the omniscient expert is also left clueless). Some tedious (yet standard) calculations show that if we set $\epsilon=\Theta(\sqrt{\frac{\log n}{n}})$ and let the mixed strategy for the adversary be with support of size $k=\Theta(\sqrt{\frac{n}{\log(n)}})$, then we have the following two phenomena:

- The events $(e.1.0)\cup(e.k.1)$ occur with small probability (and thus the hypothetical aggregator cannot perform well).
- The omniscient expert can determine the state with high probability, because the conditional distributions over posteriors $I_0(m)$ [i.e., $I(m)$ conditional on $\omega=0$] and $I_1(m)$ are sufficiently "far" from one another, for all $1\leq m\leq k$.

Discussion

We study non-Bayesian forecast aggregation and introduce an evaluation criterion for such aggregation schemes. This criterion is based on the notion of regret with respect to a square loss utility function. There are various degrees of freedom in the model as well as in choosing an evaluation criterion that provide room for further research.

Choice of Benchmark. In this paper, we study the regret of an ignorant aggregators while using an omniscient expert as our benchmark. Recall that the omniscient expert knows the information structure and also observes the actual signals experts received. This benchmark, one may argue, is too challenging. A less challenging benchmark should improve the aggregator's performance. We discuss two such alternatives.

The Bayesian aggregator. Recall the notion of a Bayesian aggregator introduced in the *Model* section. This Bayesian aggregator knows the information structure and the forecasts of the experts but does not observe the experts' private signals. Seemingly the regret associated with a Bayesian aggregator is smaller than that associated with the omniscient expert. For the families of information structures studied here (theorems 1 and 2), the regret is equal in both cases—with respect to the omniscient expert and with respect to the Bayesian aggregator. This follows from the fact that for such information structures both benchmarks entail an equal loss when the adversary uses his Maxmin mixed strategy.

The disappointing regret obtained in proposition 1, when the adversary has no restriction on the choice of a strategy and can use correlated signals, cannot be improved when using the Bayesian aggregator as a benchmark. This, however, is not a straightforward observation. In fact, for information structure designed in the proof of the proposition, the omniscient expert does remarkably well while the Bayesian aggregator fails miserably. Nevertheless, a mild modification of the information structure in proposition 1 provides the desired outcome, where the ignorant aggregator suffers a square loss of $\frac{1}{4} - \epsilon$ (where $\epsilon > 0$ is arbitrarily small), while the Bayesian aggregator suffers a square loss of zero.

Below we formalize the notion of a Bayesian aggregator and state a proposition analogous to proposition 1.

A Bayesian aggregator's posterior is

$$\tilde{x}(x_1, x_2) = \mathbf{P}(\omega = 1 | x_1(s_1) = x_1, x_2(s_2) = x_2).$$

The corresponding regret is given by

$$R^{AO}(f, \mathbf{P}) = \mathbb{E}_{(\omega, s_1, s_2) \sim \mathbf{P}}[(f(x_1(s_1), x_2(s_2)) - \omega)^2] - \mathbb{E}_{(\omega, s_1, s_2) \sim \mathbf{P}}[(\tilde{x}(x_1(s_1), x_2(s_2))) - \omega)^2].$$

Proposition 3: For general information structures, there is no aggregation scheme that guarantees a regret below $\frac{1}{4}$ with respect to the Bayesian aggregator benchmark.

The proof is relegated to *SI Appendix*, Section 4.

Best expert benchmark. Inspired by the regret-minimization literature in repeated expert advice settings (13), one may compare the square loss of the ignorant aggregator with the square loss of the better expert—namely:

$$R^{B}(f, \mathbf{P}) = \mathbb{E}_{(\omega, s_1, s_2) \sim \mathbf{P}}[(f(x_1(s_1), x_2(s_2)) - \omega)^2] - \min_{i=1,2} \mathbb{E}_{(\omega, s_i) \sim \mathbf{P}}[(x_i(s_i) - \omega)^2].$$

The result on Blackwell-ordered information structures remains the same, because the best expert coincides with the omniscient expert in this environment. Interestingly, this also holds for conditionally independent information structures with two experts, where the minimal regret is between 0.0225 and 0.0260.[#]

For conditionally i.i.d. information structures, on the other hand, the ignorant aggregator obviously can perform as well as the best expert (in expectation) simply by mimicking expert 1. By the symmetry of the problem, all experts suffer the same expected loss. This simple observation holds for any number of agents.

The minimal regret with respect to the best expert benchmark in general information structures, allowing for correlation, remains an interesting open problem (even for the case of two experts).

Extensions. In addition to various conjectures mentioned in the paper, we note that our model is restricted to a simple setting where there are only two states of nature; the scoring rule is set to be the square loss function, and we mostly focus on the case of two experts. For conditional i.i.d. signals, we provided a lower bound on the regret that the aggregator can guarantee as a function of the number of experts n. The simplicity of the problem was crucial for our ability to crack it. Extending our results to a larger state space, other scoring rules, and any number of experts is by no means straightforward. One particular question of interest is whether the increasing regret observed in the i.i.d. information structure as the number of experts grows holds for other information structures, such as the Blackwell one.

Two additional interesting research directions are to study the price of ignorance for strategic experts, in contrast with our truth-telling experts, and also to study a dual problem in which the challenge is to characterize the set of information structures that performs well for a given natural aggregation scheme (e.g., averaging or averaging the log-likelihoods). We hope to further understand these issues in future work.

ACKNOWLEDGMENTS. This research is supported by Israel Science Foundation Grants 2021296 and 2018889, United States–Israel Binational Science Foundation and National Science Foundation Grant 2016734, German-Israel Foundation Grant I-1419-118.4/2017, Ministry of Science and Technology Grant 19400214, Technion Vice President for Research grants, and the Bernard M. Gordon Center for Systems Engineering at the Technion.

[#]We omit the underlying reasoning that leads to this result.

ECONOMIC SCIENCES

1. Brier GW (1950) Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 78:1–3.
2. Blackwell D (1953) Equivalent comparisons of experiments. *Ann Math Stat* 24:265–272.
3. Prelec D, Seung HS, McCoy J (2017) A solution to the single-question crowd wisdom problem. *Nature* 541:532–535.
4. DeGroot MH (1974) Reaching a consensus. *J Am Stat Assoc* 69:118–121.
5. James SH, Watson MW (1984) Combination forecasts of output growth in a seven-country data set. *J Forecast* 23:405–430.
6. Zarnowitz V (1984) The accuracy of individual and group forecasts from business outlook surveys. *J Forecast* 3:11–26.
7. Min Ck, Zellner A (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *J Econom* 56:89–118.
8. Winkler RL (1981) Combining probability distributions from dependent information sources. *Manag Sci* 27:479–488.
9. Newbold P, Granger CW (1974) Experience with forecasting univariate time series and the combination of forecasts. *J R Stat Soc Ser A* 137:131–164.
10. Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Int J Forecast* 5:559–583.
11. Elliott G, Timmermann A (2006) Forecast combinations. *Handbook of Economic Forecasting* (North Holland, Amsterdam), Vol 1, pp 135–196.
12. Bickel PJ (1984) Parametric robustness: Small biases can be worthwhile. *Ann Stat* 12:864–879.
13. Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge Univ Press, New York).
14. Al-Najjar NI, Weinstein J (2008) Comparative testing of experts. *Econometrica* 76:541–559.
15. Feinberg Y, Stewart C (2008) Testing multiple forecasters. *Econometrica* 76:561–582.
16. Sandroni A (2000) Do markets favor agents able to make accurate predictions? *Econometrica* 68:1303–1341.
17. Olszewski W, Sandroni A (2009) Manipulability of comparative tests. *Proc Natl Acad Sci USA* 106:5029–5034.
18. Arieli I, Babichenko Y, Smorodinsky R (2017) When is the crowd wise? SSRN:3083608. Preprint, posted December 11, 2017.
19. Babichenko Y, Garber D (2018) Learning of optimal forecast aggregation in partial evidence environments. arXiv:1802.07107. Preprint, posted February 20, 2018.
20. Hannan J (1957) Approximation to Bayes risk in repeated play. *Contrib Theor Game* 3:97–139.
21. Gilboa I, Schmeidler D (1989) Maxmin expected utility with a non-unique prior. *J Math Econ* 18:141–153.
22. Aumann RJ, Maschler M, Stearns RE (1995) *Repeated Games with Incomplete Information* (MIT Press, Cambridge, MA).
23. Bordley RF (1982) A multiplicative formula for aggregating probability assessments. *Manag Sci* 28:1137–1148.