A Random Dictator Is All You Need

Itai Arieli, Yakov Babichenko, Inbal Talgam-Cohen and Konstantin Zabarnyi

Technion – Israel Institute of Technology

Abstract

We study an information aggregation setting in which a decision maker makes an informed binary decision by merging together information from several symmetric agents. Each agent provides the decision maker with a recommendation, which depends on her information about the hidden state of nature. While the decision maker has a prior distribution over the hidden state and knows the marginal distribution of each agent's recommendation, the correlation between the recommendations is chosen adversarially. The decision maker's goal is to choose an information aggregation rule that is robustly optimal.

We prove that for a sufficiently large number of agents, for the three standard robustness paradigms – minimax, regret and approximation ratio – the robustly-optimal aggregation rule is identical. Specifically, the optimal aggregation rule is the *random dictator* rule, which chooses an agent uniformly at random and adopts her recommendation. For a small number of agents, this result no longer holds – the random dictator rule can be suboptimal for minimizing the regret even for two agents. We further characterize the minimal regret for any number of agents through the notion of concavification, and demonstrate how to utilize this characterization in the case of two agents.

1 Introduction

Every day, we receive conflicting forecasts regarding such issues as investment, health, or even the weather. Based on these forecasts, we must make decisions like whether to buy or sell, to operate or not to operate, to go out or stay home. In most cases, when basing our decisions on the predictions we have, we are not aware of how the information obtained from different sources is correlated. A fundamental challenge presented recently by [17] is – how to aggregate the forecasts and reach decisions despite this knowledge gap.

The practice of *information aggregation* is a common technique used to improve the accuracy of prediction models. Combining forecasts obtained from multiple sources can help account for the inherent uncertainty and variability present in any forecasting task. There is a variety of methods for combining forecasts, including simple averaging [15], median averaging [25], or weighting each forecast based on its past performance [19]. Forecast aggregation has been extensively studied in many fields, such as finance [9], marketing [4], and meteorology [34].

Given the multitude of approaches, to produce reliable and meaningful results one must carefully consider the underlying assumptions and limitations of each aggregation method. A common way to evaluate a solution to a decision problem under uncertainty in general, and to the problem of information aggregation under uncertainty in particular, is via the methodology of *robustness*. That is, the performance of a decision rule is evaluated according to the worst-case (a.k.a. adversarial) scenario.

Within the robustness methodology, there are two leading paradigms for quantifying the quality of a decision under uncertainty: the *minimax* approach and the *regret* approach. The minimax approach [40, 23] involves choosing the option that minimizes the maximum possible loss. This approach is based on the idea that the decision maker wants to optimize the worst-case outcome, and thus applying it does not require a benchmark. Alternatively, the regret approach [36] sets an ambitious benchmark – the loss from a decision made without the burden of uncertainty. It then aims to minimize the difference (known as the regret) between the performance of the optimal decisions with and without uncertainty – that is, to mimic as closely as possible the optimal informed decision.

In a typical scenario, the minimax course of action differs from the regret-minimizing course of action. For example, the celebrated result of Carroll [13], establishing the minimax-optimality of selling items separately in a combinatorial auction setting, does not hold for regret minimization (see Appendix B). Such conflicting recommendations weaken the predictive, as well as prescriptive, power of the robustness paradigm.

The methodology of robustness, and specifically – the minimax approach, is applied by De Oliveira et al. [17] to an information aggregation setting. In this paper, we follow their lead by applying the complementary approach of regret to a basic information aggregation setting with symmetric agents. Surprisingly, we show that the simple aggregation rule of randomly choosing an agent to make the decision, known as the *random dictator* rule, is optimal for both robustness approaches – minimax and regret – for a wide range of instances. This result can be viewed as strengthening the message of [17]. Moreover, the same rule turns out to be asymptotically (in the number of agents) optimal for the *approximation ratio* robustness approach, which is common in computer science [39] and some game theory applications [35]. Our main takeaway is that in our binary decision setting, the random dictator rule is universally optimal with respect to all standard robustness concepts.

1.1 Our Information Aggregation Setting

A leading example. We now demonstrate our basic information aggregation setting; many of the simplifying assumptions can be relaxed (see Appendix A). Consider a *decision maker (DM)* facing the decision of whether or not to implement a project. An unknown binary *state of nature* (or *state* for short) indicates whether the project is beneficial or detrimental in terms of utility. As a concrete example, consider the decision whether to implement containment measures during a pandemic in a certain region (the project). This project's utility depends on whether illness is on the rise in that region or not (the state of nature). We consider an identical-interest scenario in which the DM and all forecasting agents agree on the utility of the project in both the beneficial and detrimental states.

Each agent is privately exposed to some information about the state. For instance, this can be the number of illness cases in an agent's social network.[1] Based on this information, the agents update the probability with which they believe the beneficial state to hold. The agents are symmetric in the sense that all agents who witness a rise (respectively, decline) in illness around them hold the same belief regarding this probability. Formally, agents have identical marginal distributions of posterior beliefs. Each agent anonymously fills in a survey reporting her true recommendation as to whether the project should be implemented.[2] The DM receives the statistics of the survey – i.e., the number of agents recommending implementation. A (probabilistic) *aggregation rule* maps the survey's outcome to a probability with which the project will be implemented. Importantly, the aggregation rule can rely on knowledge of the marginal distributions of posteriors, but not on the correlation among the agents.

Our goal is to optimize the aggregation rule under the robust (adversarial) approach. On the one hand, the *minimax* approach aims to find an aggregation rule that performs well for *all possible correlations between the private information of different agents*. The minimax-optimal aggregation rule has been studied by De Oliveira et al. [17], who show that the optimal aggregation rule is to follow the best expert. On the other hand, the *regret minimization* approach considers as a benchmark a hypothetical *Bayesian DM*, who knows the true correlations between the agents' private information; given the statistics of the survey, the Bayesian DM can calculate the posterior distribution of the state, and take an optimal action given the survey outcome. The goal is to minimize the difference between the utility of the Bayesian DM and that of our ignorant DM who does not know the correlations.

The assumption of symmetric agents is common in many contexts in the literature [see, e.g., 12]; in our leading example, this assumption is reasonable since the agents are ordinary citizens who obtain exogenous information with an unknown reliability level. Another possible interpretation of our model involves a researcher (DM) repeatedly experimenting – making the symmetry assumption intrinsic. The knowledge of the marginals corresponds to the knowledge of the false-positive and the false-negative rates. Robustness to the correlation may be viewed as robustness to the sampling procedure according to which the samples for the experiments are collected.

1.2 Outline of the Main Results

We begin by revisiting the work of De Oliveira et al. [17] on minimax-optimal aggregation, and developing an analogue of their main result for symmetric agents. According to their result, the aggregator should follow the best expert. In our symmetric analogue, the DM only needs to observe the aggregate statistics of the agents' recommendations. Even though the aggregator cannot follow any single agent (their recommendations are treated anonymously), the aggregator can follow a random agent by adopting a

[1] During the height of the COVID-19 pandemic, such information was found to be extremely useful in forecasting the spread of illness in the short term – see [38].

[2] We assume that agents' recommendations are truthful – an agent recommends implementation of the project if and only if according to her private information, the expected utility of the project is non-negative. This assumption is natural for many surveys, e.g., if the agents are unaware of how the DM will aggregate their recommendations. See Subsection A.5 for further discussion of this issue.

recommendation chosen uniformly at random (the random dictator rule). Since, by symmetry, all agents are qualified to be best experts, such an aggregation rule is minimax-optimal.

Turning from minimax to regret, the analysis is more involved. One challenge for the aggregation task is that the number of possible posterior belief realizations grows exponentially in the number of agents. This is where treating the agents anonymously is helpful. We utilize a recent characterization by [2] of feasible information structures to reduce the problem to a single-dimensional one. In contrast to some cases with a small number of agents (as we demonstrate for two agents in Example 5.2), we prove that for a large class of instances – in particular, for any sufficiently large number of agents – the unique regret-minimizing aggregation function is the random dictator rule (Theorem 4.3).

To deal with the remaining instances, we apply a zero-sum game formulation to deduce a clean formula for the minimal regret the DM can guarantee (Theorem 5.1). The optimal regret turns out to be the maximum distance between a high-dimensional function P^* representing the expected utility of a Bayesian DM and its concavification. This result implies a connection between minimizing regret and the concavification technique that is usually applied in Bayesian persuasion [26]. Interestingly, even though our DM is ignorant to the information structure, her regret is tightly related to the probability with which a Bayesian DM correctly guesses the state of nature. The concavification formula allows us to resolve the regret problem in instances in which the previous approach fails, i.e., cases with a small number of agents.

While the anonymity assumption (by which the DM observes only the aggregate statistics of the agents' recommendations) is natural in the context of information aggregation, we also show that considering this anonymous version of the problem is without loss of generality (Appendix A.3).

1.3 Related Work

The most closely related paper to ours is [17]. The settings considered in the two works are similar. The major difference is that our main focus is on the regret approach, which differs (conceptually and technically) from their minimax approach. Another difference is that we assume symmetric agents. Unlike [17], we also study an anonymous variant of the problem in which the aggregator observes the statistics of recommendations rather than the entire recommendation vector.

A work that adopts the regret approach is Arieli et al. [3]. They study Blackwell-ordered [10, 11] and conditionally i.i.d. information structures. Additional works that consider regret as a measure of robustness include: Babichenko and Garber [7], which explores the class of partial-evidence information structures in a repeated game setting; [14], which considers information structures with informational substitutes; and Neyman and Roughgarden [33], which shows that under the projective substitutes assumption, the principal can improve upon the random dictator mechanism considered by us.

Information aggregation with uncertainty about the information structure has been studied in a voting framework by Levy and Razin [28]. They consider a voter's uncertainty about whether the signals from different sources are fully correlated or conditionally independent. They show that such uncertainty may encourage the voter to rely more on the information she gets and less on her political preference.

Robustness to correlations given fixed marginal distributions over the agents' types (private information) has been previously studied in the framework of mechanism design. Some examples include [13, 24], studying the optimal robust mechanism for a principal, assuming that she knows the marginal distributions of the agents' types, but not the joint distribution.

Forecast aggregation is a well-studied topic within statistics. Some works (see, e.g., [18, 37, 42]) examine how well simple aggregation rules perform. An alternative approach uses a training set to find an optimal aggregation scheme within a parametric family of such schemes (e.g., [32]). In social choice, *judgment aggregation* studies the aggregation of beliefs about whether logical statements are true or false [21]. In economics, Levy and Razin [29] consider forecast aggregation with multiple experts. Their aggregation scheme follows the assumption that the information structure of the experts is the one maximizing the likelihood of observing their realized forecasts. Further recent works on forecast aggregation include Wang et al. [41], who study aggregation without being exposed to the past performance data using peer assessment scores. In the correlation-robust framework, we should mention [30], who study aggregation of information with uncertainty both about the joint information distribution and the set of all possible signals.

The fundamental question of whether social choice rules aggregate well the information of the crowd is an old topic in economics that goes back to Condorcet [16], who proves that the simple majority rule is an asymptotically optimal aggregation rule when agents vote sincerely according to their private information. However, Austen-Smith and Banks [6] demonstrate that Condorcet theorem strongly relies on the property that agents' information is symmetric. In asymmetric information structures, even when the signals are conditionally independent, the simple majority aggregation rule is far from being optimal. These observations have initiated an extensive line of research that aims to understand which aggregation rules aggregate well the information of the crowd (e.g., [31, 22, 27, 1], just to mention a few). Our paper aims to address the same question, but in a setting with the correlation of signals being unknown. This is in contrast to most of the existing research, which commonly assumes that the signals of the agents are independent conditional on the state. Our result indicates the robust optimality of the random dictator rule, and we show in Section 4 that in our symmetric setting, simple majority aggregation is far from being optimal.

Paper organization. In Section 2, we introduce the formal mathematical model and basic definitions. The model includes some simplifying assumptions; generalizations are discussed in Appendix A. Section 3 revisits the minimax robustness paradigm in our model, showing that the random dictator rule is always optimal. In Section 4, we present an approach for analyzing regret minimization that is particularly useful for a large number of agents. This approach allows to upper bound the regret of any specific aggregation rule; we extend the main result of this section to the approximation ratio robustness paradigm in Appendix A.4. Section 5 connects the regret-minimizing aggregation problem to the concavification of a function; and Subsection 5.1 shows an application for two agents. Section 6 concludes and suggests some questions for future work.

2 Preliminaries

Notation. For any set D, let $\Delta(D)$ be the set of all probability distributions over D.

2.1 Setting

We consider a binary space of *states of nature* $\Omega = \{0, 1\}$, equipped with a publicly-known *prior* $\mu \in [0, 1]$ – the probability that the (unknown) true state $\omega \in \Omega$ is equal to 1. We extend our main results to an arbitrary finite state space in Appendix A.2.

The model contains a *decision maker* (or *DM*, also known as an *aggregator*) and n informed *agents*. The agents are numbered $1, \dots, n$, with agent i observing a binary *signal* $s_i \in S := \{L, H\}$ that represents information about the state of nature. The agents truthfully report to the DM either H or L according to their observed signal. The assumption that signals are binary is for simplicity of presentation – our results can be extended to arbitrary sets of signals, including continuum-many signals (see Subsection A.1).

The joint distribution according to which the state and the n signals are generated is denoted by $\pi \in \Delta(\Omega \times S^n)$, and is called the *information structure*. We use $\pi(\cdot)$ to denote the probability of a given event according to the information structure. The information structure must be compatible with the prior distribution – i.e., $\pi(\omega = 1) = \mu$.

The posteriors. Upon observing $s_i = \sigma$, agent i computes her *posterior distribution* over the states as follows: the posterior probability of $\omega = 1$ is $\frac{\pi(\omega=1, s_i=\sigma)}{\pi(s_i=\sigma)}$. The agents are *symmetric* in the following sense – the marginal distributions over posterior beliefs of all the agents are identical. The posterior of agent i after observing the signal $s_i = L$ is assumed to be fixed across all agents. Denote this posterior by p_1 and call it the *low posterior*. Similarly, the *high posterior* of an agent is p_2. Note that for every $1 \le i \le n$ we have:

$$p_1 = \frac{\pi(\omega = 1, s_i = L)}{\pi(s_i = L)}; \; p_2 = \frac{\pi(\omega = 1, s_i = H)}{\pi(s_i = H)}.$$

Throughout the paper we assume that $p_1 < \frac{1}{2} < p_2$ – otherwise, the agents' reports do not affect the DM's guess regarding the state of nature, and the setting is degenerate.

As in [17], we assume that p_1, p_2 are known to the DM; this enables us to focus on the effect of not knowing the correlations, which are captured by the information structure π that is not observed by the DM.

Throughout the paper we use the following parameters a, b: Parameter a (respectively, b) represents the probability of a single fixed agent to have the high posterior conditional on state $\omega = 0$ (respectively,

$\omega = 1$). The parameters are deduced by straightforward calculations to be:[3]

$$a = \frac{(1-p_2)(\mu - p_1)}{(1-\mu)(p_2 - p_1)};\ b = \frac{p_2(\mu - p_1)}{\mu(p_2 - p_1)}. \tag{1}$$

2.2 Information Aggregation

We consider an identical-interest scenario. Therefore, we assume that the agents truthfully report their private signals to the DM.

However, we assume that the DM only observes the fraction of agents who report H, which is denoted by $\nu \in \{0, \frac{1}{n}, \ldots, 1\}$. Following [2], we assume *anonymous* agents – the DM does not observe the identity of the agents who report H. Note that *this assumption is without loss of generality* (see Subsection A.3). It is introduced both for simplicity of presentation and for emphasizing the unimportance of the agents' identities for understanding the optimal DM' strategy.

The DM's goal is to aggregate the information into a guess of the correct state, using her knowledge of the fraction ν, as well as the prior μ and posteriors p_1, p_2. Namely, the DM's utility is 1 if she guesses the state correctly, and is 0 otherwise.[4] The DM can use a mixed guessing strategy, captured by an *aggregation rule* $f : \{0, \frac{1}{n}, \ldots, 1\} \to [0,1]$, where $f(\nu)$ denotes the probability that the DM guesses $\omega = 1$ after observing a fraction of ν agents who report H.

The information structure π and state of nature ω induce a distribution over ν – that is, a distribution over the possible fractions $\{0, \frac{1}{n}, \ldots, 1\}$ of agents with the high posterior. Let $\hat{\pi} \in \Delta\big(\Omega \times \{0, \frac{1}{n}, \ldots, 1\}\big)$ denote the distribution over ω and ν induced by π. The corresponding distributions over ν conditional on the state $\omega = 0$ and $\omega = 1$ are denoted by $\hat{\pi}_0$ and $\hat{\pi}_1$, respectively. For an information structure π and an aggregation rule f, the probability of the DM to guess correctly is denoted by $P(f, \pi)$, and equals:

$$P(f, \pi) := \mu \mathbb{E}_{\nu \sim \hat{\pi}_1}[f(\nu)] + (1-\mu)\mathbb{E}_{\nu \sim \hat{\pi}_0}[1 - f(\nu)].$$

Note that $\mathbb{E}_{\nu \sim \hat{\pi}_1}[f(\nu)]$ is the probability that the DM guesses $\omega = 1$, on average over the observed fraction ν given that the true state is 1. Similarly, $\mathbb{E}_{\nu \sim \hat{\pi}_0}[1 - f(\nu)]$ is the probability that the DM guesses $\omega = 0$, on average over the observed fraction ν given that the true state is 0.

2.3 Regret and Minimax

Consider a *Bayesian DM*, i.e., a DM who knows the information structure π. Such a DM always guesses the more likely state. Hence, her probability of guessing correctly equals:

$$P^*(\pi) := \mathbb{E}_{\nu \sim \hat{\pi}}[\max\{\hat{\pi}(\omega = 0|\nu), \hat{\pi}(\omega = 1|\nu)\}].$$

The *regret of an aggregation rule f for an information structure* π is defined by:

$$\mathrm{Reg}(f, \pi) := P^*(\pi) - P(f, \pi).$$

Since the DM does not know π, she aims to design an aggregation rule with low regret *for all* information structures. We define $\mathrm{Reg}(f) := \max_\pi \mathrm{Reg}(f, \pi)$.[5] That is, the *regret of an aggregation rule* f is the regret – i.e., expected additive utility loss – in the worst-case scenario. We refer to the worst-case information structure π as chosen by an *adversary* given f. Denote by $\mathrm{Reg} := \min_f \mathrm{Reg}(f)$ the regret of an *optimal* aggregation rule. Through most of the paper, we focus on computing Reg, and finding an aggregation rule f (which may depend on p_1, p_2, and μ, but obviously not on π) that minimizes $\mathrm{Reg}(f)$.

We further define

$$\mathrm{Minmax} := \max_f \min_\pi P(f, \pi)$$

to be the maximum probability of a correct guess by the DM under the worst possible information structure chosen by an adversary. We say that an aggregation rule f is *minimax-optimal* if $\min_\pi P(f, \pi) = \mathrm{Minmax}$.

[3]The probability of an agent to have each posterior is uniquely determined by the requirement that the expectation of the posterior must be the prior.

[4]For a binary state space, it is easy to generalize all our results to arbitrary utilities; we further generalize our main results to any finite state space and arbitrary utilities in Subsection A.2.

[5]The maximum exists as $\mathrm{Reg}(f, \cdot)$ is a continuous functional defined on a compact space.

2.4 Concavification and Convexification

We recall the standard definitions of the concavification and convexification of a function [11, 5]. Let D be some compact convex set and let $h : D \to \mathbb{R}$ be a function. The *concavification* $\text{cav}[h] : D \to \mathbb{R}$ is the pointwise minimum of all the concave functions whose graph is weakly above h. Formally, $\text{cav}[h]$ is concave and for every concave function $\tilde{h} : D \to \mathbb{R}$ with $\tilde{h}(x) \geq h(x)$ for every $x \in D$, it holds that $\tilde{h}(x) \geq \text{cav}[h](x) \geq h(x)$ for every $x \in D$.[6] Similarly, the *convexification* $\text{vex}[h] : D \to \mathbb{R}$ is the pointwise maximum of all the convex functions whose graph is weakly below h.

3 Minimax-Optimal Aggregation Revisited

In this section, we revisit the work of De Oliveira et al. [17], who study minimax-optimal information aggregation rules. Unlike us, they do not assume that the agents' distributions over posteriors are identical. The main result of [17] states that the minimax-optimal aggregation rule is to follow the recommendation of the best agent (and ignore the remaining agents); "best agent" here means the agent with the most informative private signal. We now give an analogue of this result in our symmetric setting, and provide a simple standalone proof for completeness.

The simple key observation in our symmetric setting is as follows. Since all agents have identical distributions of posteriors, they all qualify as "best agents". By choosing an agent uniformly at random and following her recommendation (an aggregation rule for which anonymous statistics suffice), we essentially follow the recommendation of a best agent. [7] Therefore, the minimax-optimal aggregation rule is the *random dictator* rule – choosing an agent uniformly at random and relying only on her recommendation. Formally:

Definition 3.1. The *random dictator* aggregation rule is the function $f : \left\{0, \frac{1}{n}, \ldots, 1\right\} \to [0, 1]$ satisfying $f\left(\frac{k}{n}\right) = \frac{k}{n}$ for $0 \leq k \leq n$.

For example, if the fraction of agents with signal H is $\nu = 1/3$, the probability of guessing $\omega = 1$ is $f(1/3) = 1/3$, which is equivalent to choosing one of the n agents uniformly at random and following her recommendation.

In the symmetric setting we study, one can analyze optimal aggregation for the minimax paradigm straightforwardly, without relying on the results of [17]. It is informative to consider in this case the best information structure from the perspective of an adversary who tries to fail the DM. Such an information structure is *fully-correlated*, that is, it sends all the agents the same signal, thus revealing the minimal possible amount of information to the DM. The DM then has nothing better to do than following the unanimous recommendation of all agents.

Proposition 3.2 (A special case of the main result of [17])**.** *For every information aggregation setting with prior μ, the random dictator aggregation rule is minimax-optimal. Moreover,* $\text{Minmax} = (1-\mu) \cdot (1-a) + \mu \cdot b$ *for a, b as defined in Eq.* (1)*.*

Proof. We first prove that the random dictator rule guesses ω correctly with expected probability of exactly $(1-\mu) \cdot (1-a) + \mu \cdot b$. Indeed, when $\omega = 0$, the expected fraction of agents with the low posterior is $1 - a$. Moreover, when $\omega = 1$, the expected fraction of agents with the high posterior is b. Since the DM follows the recommendation of a uniformly chosen agent, the law of total expectation yields that the expected probability of a correct guess is $(1-\mu) \cdot (1-a) + \mu \cdot b$, as desired.

It remains to show that for every aggregation rule f, the adversary can ensure that the probability of a correct guess is at most $(1-\mu) \cdot (1-a) + \mu \cdot b$. Indeed, let the adversary pick the fully-correlated information structure – that is, $s_1 \equiv s_2 \equiv \ldots \equiv s_n$. In this case, the random dictator rule is trivially optimal, as all the agents get the same signal. Since, as shown, this rule guesses the correct state with expected probability of exactly $(1-\mu) \cdot (1-a) + \mu \cdot b$, the claim follows. □

This straightforward analysis no longer holds for the regret minimization paradigm. Indeed, for the fully-correlated information structure, our ignorant DM can perform as well as the hypothetical Bayesian DM, and the regret is 0. The next two sections provide two complementary approaches to tackle regret minimization.

[6] See [20] for algorithmic aspects of computing the concavification of a set function.

[7] We assume the DM only observes anonymous data. As we show in Subsection A.3 for regret minimization – this assumption is not essential. Analogous considerations apply to minimax-optimal aggregation rules.

4 Main Result

In this section, we present a general approach for finding the minimum regret Reg, which – quite surprisingly – turns the optimization into a *single-dimensional* problem. We prove that for many settings, the uniquely optimal aggregation rule f is the random dictator rule (Theorem 4.3). In particular, it is always true for a sufficiently large number of agents n (Corollary 4.4).

We start with a preliminary bound on the regret of a given aggregation rule f; see Figure 1.

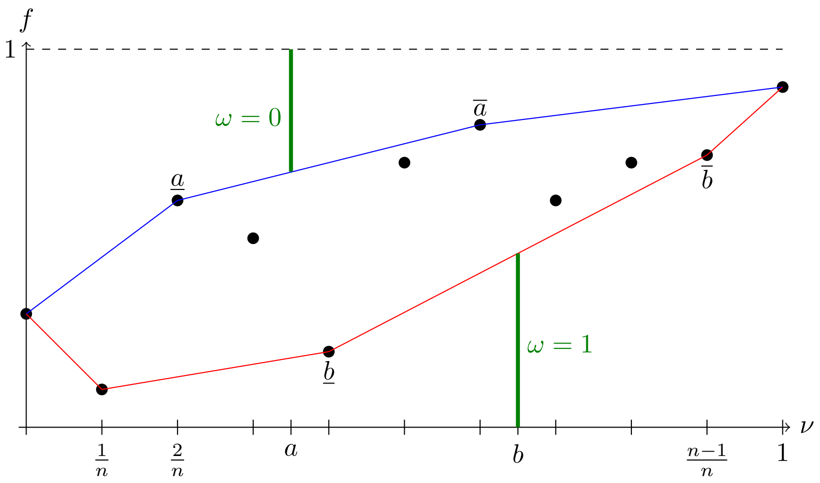

Figure 1: An aggregation rule f. The blue function is cav$[f]$. The red function is vex$[f]$. The length of the left (resp., right) green line captures the minimal probability – across all information structures – of the aggregation function f to guess correctly the state $\omega = 0$ (resp., $\omega = 1$).

Proposition 4.1. *For every a, b, n, μ, and f:*

$$\text{Reg}(f) \leq 1 - (1-\mu) \cdot (1 - \text{cav}[f](a)) - \mu \cdot \text{vex}[f](b).$$

To prove this result, we first note that the set of feasible distributions over frequencies of posteriors has a neat characterization according to a recent result by Arieli and Babichenko [2]. More formally, we say that a pair of distributions $(\hat{\rho}_0, \hat{\rho}_1) \in \left(\Delta\left(\left\{0, \frac{1}{n}, \ldots, 1\right\}\right)\right)^2$ is *feasible* if there exists an information structure π such that $\hat{\rho}_\omega = \hat{\pi}_\omega$ for $\omega = 0, 1$. The parameters a and b defined in Section 2 play a central role in the characterization of feasible distributions.

Theorem 4.2 (Arieli and Babichenko [2])**.** *A pair of conditional distributions $(\hat{\rho}_0, \hat{\rho}_1) \in \left(\Delta\left(\left\{0, \frac{1}{n}, \ldots, 1\right\}\right)\right)^2$ is feasible if and only if $\mathbb{E}[\hat{\rho}_0] = a$ and $\mathbb{E}[\hat{\rho}_1] = b$.*

As the adversary can spread b to the grid of empirical posteriors in any way she prefers provided that the expectation of the spread is b, and the probability of a correct guess by the DM of the state $\omega = 1$ is the expectation of f over the spread – to minimize the correct guess probability, the adversary would prefer to spread b to points on the convexification of f. Similarly, to minimize the correct guess probability for $\omega = 0$, the adversary should spread a to points on the concavification of f.

Proof of Proposition 4.1. Let π be an information structure that induces the distributions of frequencies $(\hat{\pi}_0, \hat{\pi}_1)$. By Theorem 4.2, we know that $\mathbb{E}[\hat{\pi}_0] = a$ and $\mathbb{E}[\hat{\pi}_1] = b$.

Since $\mathbb{E}[\hat{\pi}_0] = a$, the function f guesses the state with probability of at least $1 - \text{cav}[f](a)$. Indeed, cav$[f]$ captures the maximial probability of a mistake (across all possible $\hat{\pi}_0$). Similarly, as $\mathbb{E}[\hat{\pi}_1] = b$, the function f guesses the state with probability of at least $-\text{cav}[-f](b) = \text{vex}[f](b)$: vex$[f]$ captures the minimal probability of a correct guess.

Therefore, the probability of a correct guess of ω under the information structure π is at least:

$$1 - (1-\mu) \cdot (1 - \text{cav}[f](a)) - \mu \cdot \text{vex}[f](b).$$

With the trivial bound of 1 on the correct guess probability of the state, we get for every π:

$$\text{Reg}(f, \pi) \leq 1 - (1-\mu) \cdot (1 - \text{cav}[f](a)) - \mu \cdot \text{vex}[f](b).$$

Since it is true for every π, the proposition follows. $\square$

Tightness of Proposition 4.1. The bound of Proposition 4.1 is tight in many cases. For instance, consider the function f that is depicted in Figure 1. By Theorem 4.2, the adversary can choose the information structure π such that $\hat{\pi}_0$ will be supported on the two points $\underline{a}, \overline{a}$, and $\hat{\pi}_1$ will be supported on the two points $\underline{b}, \overline{b}$. For such an information structure, the Bayesian DM knows to guess the state with probability 1 because $\{\underline{a}, \overline{a}\} \cap \{\underline{b}, \overline{b}\} = \emptyset$. Moreover the aggregation function f guesses the state $\omega = 0$ (respectively, $\omega = 1$) with probability $1 - \mathrm{cav}[f](a)$ ($\mathrm{vex}[f](b)$) exactly. Hence, the bound of Proposition 4.1 is tight. In fact, such an argument will prove tightness of the bound in Proposition 4.1 whenever the supports of the concavification of f at a and of the convexification of f at b are *disjoint*. This is a key observation to deduce the regret-minimizing f for a large class of instances in the case of many agents (i.e., when n is large).

4.1 Regret Minimization of the Random Dictator Rule

In this section, we prove that if $\frac{1}{n} \leq a < b \leq \frac{n-1}{n}$ – then the uniquely optimal aggregation rule is random dictator. As we have already noted in Section 3, this aggregation rule is the natural analogue in our symmetric setting of "following the best agent", which is shown to be minimax-optimal by [17]. Moreover, it follows that random dictator is the *unique approximation ratio-optimizing aggregation rule* when $\frac{1}{n} \leq a < b \leq \frac{n-1}{n}$. Thus, for fixed prior and marginal posteriors, the random dictator rule is asymptotically uniquely optimal *regardless of the studied robustness paradigm.*

Theorem 4.3. *Suppose* $\frac{1}{n} \leq a < b \leq \frac{n-1}{n}$. *Then random dictator is the unique regret-minimizing aggregation rule. Moreover,* $\mathrm{Reg} = 1 - (1-\mu)\cdot(1-a) - \mu \cdot b$.

In particular, this theorem implies an asymptotic result when the marginals remain fixed, but the population of agents grows.

Corollary 4.4. *For every* μ, p_1 *and* p_2*, there exists* N *s.t. for every* $n \geq N$*, the random dictator rule is the unique regret-minimizing aggregation rule. Specifically,* $N = \max\left\{\frac{(1-\mu)(p_2-p_1)}{(1-p_2)(\mu-p_1)}, \frac{\mu(p_2-p_1)}{p_1(p_2-\mu)}\right\}$.

In particular, N is large when one of the marginal posteriors is close to an extreme point of the interval to which it might belong – the prior, 0 or 1. Intuitively, when one of the possible signals either reveals almost no information or reveals almost full information – some intricate aggregation rules might be required, giving, respectively, less or more weight to the recommendations of the agents with this extreme signal.

Intuitively, the proof of the theorem relies on the tightness of Proposition 4.1 discussed above. We show that whenever the constants a and b lie in the segment $\left[\frac{1}{n}, \frac{n-1}{n}\right]$ – the adversary has sufficient flexibility to choose an information structure π *with disjoint supports* for $\hat{\pi}_0$ and $\hat{\pi}_1$; that will drop the DM's probability of guessing (weakly) below the one that is achieved by the random dictator rule. Because of the disjointness property, such an information structure has an ideal performance for the Bayesian DM (that is – always guesses the state). Hence, the DM has no better aggregation rule than the random dictator. The full proof of Theorem 4.3 appears in Appendix C.

Supermajority aggregation. A common class of aggregation functions that is considered both in practice and in the theoretical literature is the class of *supermajority rules*. Namely, $f(\nu) = \mathbf{1}_{\nu \geq \tau}$ for some threshold τ. We note that supermajority rules might perform very bad in terms of regret. For example, fix $a = \frac{1}{2} - \frac{1}{n}, b = \frac{1}{2} + \frac{1}{n}$ and $\mu = \tau = \frac{1}{2}$. The regret of the majority rule is as high as $1 - O\left(\frac{1}{n}\right)$; namely, there exists an information structure for which the Bayesian aggregator guesses the state with probability 1, while the majority rule guesses the state with probability $O\left(\frac{1}{n}\right)$. To show this, one may consider an information structure for the adversary with $\hat{\pi}_0$ supported on 0 and on a grid point $\nu_0 \in \left(\frac{1}{2}, \frac{1}{2} + O\left(\frac{1}{n}\right)\right)$, and with $\hat{\pi}_1$ supported on 1 and on a grid point $\nu_1 \in \left(\frac{1}{2} - O\left(\frac{1}{n}\right), \frac{1}{2}\right)$. Note that conditional on state ω, with probability $1 - O\left(\frac{1}{n}\right)$ the realized fraction of high reports will be ν_ω, which will cause the majority rule to fail. In contrast, the random dictator rule will guess the state correctly with probability of $\frac{1}{2} + \frac{1}{n}$.

Alternative robustness approaches. Our proof of Proposition 4.1 immediately implies a stronger result than Proposition 3.2 – under the minimax paradigm, random dictator is always the *unique* optimal aggregation rule. Moreover, note that it follows from Theorem 4.3 proof that whenever $\frac{1}{n} \leq a < b \leq \frac{n-1}{n}$, the adversary has an optimal strategy that allows the Bayesian DM to guess the state correctly with probability 1. Therefore, the regret-minimizing aggregation rule is also approximation ratio-optimizing

– see Appendix A.4. Hence, given any fixed prior and marginal posteriors, for a large enough number of agents n, *the random dictator rule is uniquely optimal for all three robustness paradigms: minimax, regret and approximation ratio.* It highlights the universal nature of robust optimality of random dictator.

5 General Regret Analysis

In the previous section, we showed how to compute the minimal regret for a large number of agents. In this section, we characterize the minimal regret as the maximal difference between the concavification of a function P^* – representing the correct guess probability of the Bayesian DM – and P^* itself (Theorem 5.1). The formula uses the concavification of the function P^* defined on a 2^{n+1}-dimensional set; it allows to explicitly compute Reg for a small number of agents.

Define, as is standard, $[n] := \{1, \ldots, n\}$. Note that the set of information structures with the given prior μ and given posteriors p_1, p_2 for both agents is a polygon $C \subset \mathbb{R}^{2^{n+1}}$ that is defined by the following equations:

$$C := \{x = (x_D^\omega)_{D \subseteq [n],\, \omega=0,1} : x_D^\omega \geq 0, \sum_{D \subseteq [n]} x_D^0 = 1 - \mu, \sum_{D \subseteq [n]} x_D^1 = \mu, \\ \forall i \in [n] : \sum_{D:\, i \in D} x_D^0 = (1-\mu)a, \sum_{D:\, i \in D} x_D^1 = \mu b\},$$

where the terms $(1-\mu)a$ and μb represent the *unconditional* probability weight that is assigned to the high signal at states 0 and 1, correspondingly. We now present a closed formula for the optimal regret.

We slightly abuse notation and consider $P^*(\pi)$, the probability of a correct guess by a Bayesian DM, to be a function $P^* : C \to [0,1]$. Note that we have for every $x = (x_D^\omega)_{D \subseteq [n],\, \omega=0,1} \in C$:

$$P^*(x) = \sum_{i=0}^{n} \max \left\{ \sum_{D:\, |D|=i} x_D^0, \sum_{D:\, |D|=i} x_D^1 \right\}.$$

Note that P^* is a convex function, as a sum of $n+1$ functions each of which is a maximum of linear functions. The following theorem characterizes the optimal regret.

Theorem 5.1. *For every number of agents $n \geq 2$ it holds that* $\mathrm{Reg} = \max_{x \in C}[\mathrm{cav}[P^*](x) - P^*(x)]$.[8]

The theorem is somewhat surprising because it connects the minimal regret of a DM who is *ignorant* of the information structure to the function P^* that reflects the probability of a correct guess of a Bayesian DM who is *aware* of the information structure.

Proof of Theorem 5.1. The regret minimization problem can be viewed as a zero-sum game between the DM who chooses an aggregation rule f and an adversary who chooses the information structure π (or, equivalently, $x \in C$). Concretely, deterministic aggregation rules $f : \{0, \frac{1}{n}, \ldots, 1\} \to \{0,1\}$ are the pure strategies of the DM, and information structures $x \in C$ are the pure strategies of the adversary. The payoff as a function of the DM's mixed strategy f and the adversary's mixed strategy ϕ is:

$$\mathrm{Reg}(f, \phi) := \mathbb{E}_{x \sim \phi}\left[\sum_{i=0}^{n} \left(\max \left\{ \sum_{D \subseteq \Omega:\, |D|=i} x_D^0, \sum_{D \subseteq \Omega:\, |D|=i} x_D^1 \right\} \right.\right. \\ \left.\left. - \left((1-f_i) \cdot \sum_{D \subseteq \Omega:\, |D|=i} X_D^0 + f_i \cdot \sum_{D \subseteq \Omega:\, |D|=i} x_D^1 \right) \right) \right].$$

The value of this zero-sum game is Reg.

By the minimax theorem, Reg equals to the maximum over all mixed strategies of the adversary followed with a best-response of the DM. Note that the DM's best response to a mixed strategy over

[8] An analogue of this theorem remains valid in a more general setting in which we drop the symmetry assumption among the agents; i.e., the setting that has been considered by De Oliveira et al. [17]: Every agent has a different information structure and reports her recommendation. The probability of a correct guess by a Bayesian DM P^* is well-defined, and the theorem states that $\mathrm{Reg} = \max_{x \in C}[\mathrm{cav}[P^*](x) - P^*(x)]$. The validity of this generalization follows Theorem 5.1 proof relying on minimax arguments only, which can be applied in the non-symmetric case without any modifications.

information structures is her Bayesian guessing of a state given the *known* mixed strategy of the adversary. Formally, let $\phi \in \Delta(C)$ be a mixed strategy of the adversary. Then the DM's utility under best response is $P^*(\mathbb{E}[\phi])$. The hypothetical Bayesian DM will apply a Bayesian guessing in each realization of $x \sim \phi$. Hence, her expected probability of guessing the state is $\mathbb{E}_{x\sim\phi}[P^*(x)]$. Overall, we deduce that:

$$\mathrm{Reg} = \max_{\phi\in\Delta(C)} [\mathbb{E}_{x\sim\phi}[P^*(x)] - P^*(\mathbb{E}[\phi])] = \max_{y\in C}[\mathrm{cav}[P^*](y) - P^*(y)],$$

where the second equality follows from taking $y = \mathbb{E}[\phi]$ at the maximizing mixed strategy. □

5.1 Example: The Two-Agent Case

In this subsection, we demonstrate the use of the duality method on which Theorem 5.1 is based for computing the best aggregation function for $n = 2$ agents and a uniform prior. The result easily generalizes to an arbitrary prior, and a similar methodology can be used to find a closed expression for Reg for small values of n.[9]

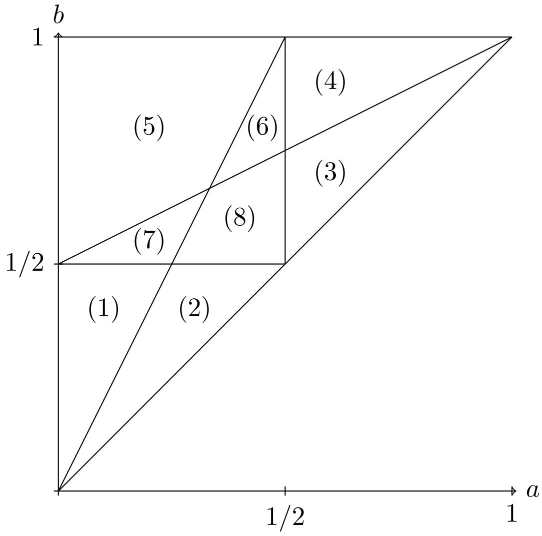

Figure 2: The partition of the values of (a, b) into regions according to the cases in Example 5.2.

Example 5.2. For $n = 2$ agents and a uniform prior $\mu = 1/2$, the optimal aggregation function f satisfies $f(0) = 0$ and $f(1) = 1$. The value of $f\left(\frac{1}{2}\right)$ and the minimal regret Reg are given by:[10]

1. $f\left(\frac{1}{2}\right) = \frac{a+2b}{2(a+b)}$, $\mathrm{Reg} = \frac{a(a+2b)}{2(a+b)}$ if $a < b \le \frac{1}{2}$ and $2a \le b$.

2. $f\left(\frac{1}{2}\right) = \frac{3b-a}{2(a+b)}$, $\mathrm{Reg} = \frac{a^2+4ab-b^2}{2(a+b)}$ if $a < b \le \frac{1}{2}$ and $2a \ge b$.

3. $f\left(\frac{1}{2}\right) = \frac{2-3a+b}{2(2-a-b)}$, $\mathrm{Reg} = \frac{-a^2+4ab+b^2-2a-6b+4}{2(2-a-b)}$ if $\frac{1}{2} \le a < b$ and $1 + a \ge 2b$.

4. $f\left(\frac{1}{2}\right) = \frac{3-2a-b}{2(2-a-b)}$, $\mathrm{Reg} = \frac{(1-b)(3-2a-b)}{2(2-a-b)}$ if $\frac{1}{2} \le a < b$ and $1 + a \le 2b$.

5. $f\left(\frac{1}{2}\right) = \frac{1-b}{1+a-b}$, $\mathrm{Reg} = \frac{a(1-b)}{1+a-b}$ if $a \le \frac{1}{2} \le b$ and $b \ge \max\left\{2a, \frac{a+1}{2}\right\}$.

6. $f\left(\frac{1}{2}\right) = \frac{2-2a-b}{2(1+a-b)}$, $\mathrm{Reg} = \frac{(1-b)(4a-b)}{2(1+a-b)}$ if $a \le \frac{1}{2} \le b$ and $2a \ge b \ge \frac{a+1}{2}$.

7. $f\left(\frac{1}{2}\right) = \frac{-1+a+2b}{2(1+a-b)}$, $\mathrm{Reg} = \frac{a(3+a-4b)}{2(1+a-b)}$ if $a \le \frac{1}{2} \le b$ and $\frac{a+1}{2} \ge b \ge 2a$.

8. $f\left(\frac{1}{2}\right) = \frac{3-a-3b}{2(1+a-b)}$, $\mathrm{Reg} = \frac{a^2-6ab+b^2+5a-b}{2(1+a-b)}$ if $a \le \frac{1}{2} \le b$ and $b \le \min\left\{2a, \frac{a+1}{2}\right\}$.

[9] Note that in the special case of a uniform prior, the assumption $p_1 < \frac{1}{2} < p_2$ trivially holds unless $p_1 = p_2 = a = b = \frac{1}{2}$.

[10] When multiple cases apply – all the corresponding possibilities for f are optimal.

Example 5.2 demonstrates that for a small number of agents ($n = 2$), the regret-minimizing aggregation rule is quite a complex object (see Figure 2 for illustration). This is in a sharp contrast to considering the minimax-optimal aggregation rule, which is always random dictator (Proposition 3.2). Note that in particular, there might be more than one regret-minimizing f. As follows from the previous section, this is a side effect of the number of agents being small.

The proof relies on Theorem 5.1, which allows us to calculate Reg by maximizing the difference $\text{cav}[P^*] - P^*$. The function P^* turns out to be two-dimensional; hence, the calculations are relatively easy. Note that Theorem 5.1 does not characterize the regret-minimizing aggregation rule f. To deduce f, we use the fact that f must induce indifference between the information structures that the adversary chooses with positive probability. This pins down f when it is combined with the educated guess that $f(0) = 0$ and $f(1) = 1$. All these calculations are relegated to Appendix D.

The analysis of Example 5.2 in Appendix D suggests the following interpretation of the optimal f for $n = 2$ and $\mu = \frac{1}{2}$ – the adversary should randomize between two information structures; in each of the two information structures, in one state the adversary should fully correlate the signals, while in the other state she should anticorrelate the signals as much as possible.

6 Conclusions and Future Work

We initiate the study of regret minimization for aggregating information. While the regret minimization approach turns out to be fundamentally different from the minimax approach for a small number of symmetric agents – with the former being much more intricate – they are equivalent for a sufficiently large number of such agents. Moreover, the approximation ratio robustness approach is also equivalent to these approaches when the number of agents is large.

Our work suggests many natural open questions. The tractability of our analysis crucially relies on the fact that the decision of the DM is binary. It remains an open problem how to analyze settings with more than two actions for the DM. Weakening the assumption that the agents are symmetric is another interesting direction for the regret approach. Furthermore, an online variant of our setting – with the DM trying to learn the optimal aggregation rule – remains unexplored. Another interesting challenge is to come up with mild limitations on the power of the adversary under which natural aggregation rules – such as supermajority – perform better than the random dictator and potentially are robustly optimal. Finally, the existence of a *universally* robust strategy (aggregation rule in our case) that is optimal under several robustness approaches is a seemingly rare phenomenon, which we have found to hold in the information aggregation problem. We do not know of any other setting in which such a universally robust strategy exists. It will be interesting to find other settings with this phenomenon and explore underlying reasons.

Acknowledgements. Itai Arieli is supported by the Ministry of Science and Technology grant no. 2028255. Yakov Babichenko is supported by the Binational Science Foundation BSF grant no. 2018397 and by the German-Israeli Foundation for Scientific Research and Development GIF grant no. I-2526-407.6/2019. Inbal Talgam-Cohen is supported by a Google Research Scholar award. This research is supported by the Israel Science Foundation grant no. 336/18. Yakov and Inbal are supported by the Binational Science Foundation BSF grant no. 2021680. Konstantin Zabarnyi is supported by a PBC scholarship for Ph.D. students in data science. The authors are grateful to Kfir Eliaz, Sergiu Hart and the anonymous reviewers for their helpful remarks.

References

[1] David S Ahn and Santiago Oliveros. Approval voting and scoring rules with common values. *Journal of Economic Theory*, 166:304–310, 2016.

[2] Itai Arieli and Yakov Babichenko. A population's feasible posterior beliefs. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, pages 326–327, 2022.

[3] Itai Arieli, Yakov Babichenko, and Rann Smorodinsky. Robust forecast aggregation. *Proceedings of the National Academy of Sciences*, 115(52):E12135–E12143, 2018.

[4] Jon Scott Armstrong. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, volume 30. Springer, 2001.

[5] Robert J Aumann, Michael Maschler, and Richard E Stearns. *Repeated games with incomplete information.* MIT press, 1995.

[6] David Austen-Smith and Jeffrey S Banks. Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, 90(1):34–45, 1996.

[7] Yakov Babichenko and Dan Garber. Learning optimal forecast aggregation in partial evidence environments. *Mathematics of Operations Research*, 46(2):628–641, 2021.

[8] Yakov Babichenko, Inbal Talgam-Cohen, Haifeng Xu, and Konstantin Zabarnyi. Regret-minimizing Bayesian persuasion. *Games and Economic Behavior*, 136:226–248, 2022.

[9] Marta Bańbura, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin. Now-casting and the real-time data flow. In *Handbook of Economic Forecasting*, volume 2, pages 195–237. Elsevier, 2013.

[10] David Blackwell. Comparison of experiments. Technical report, Howard University, Washington, United States, 1950.

[11] David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

[12] Jeremy Bulow and Paul Klemperer. Auctions versus negotiations. *American Economic Review*, 86 (1):180–194, 1996.

[13] Gabriel Carroll. Robustness and separation in multidimensional screening. *Econometrica*, 85(2): 453–488, 2017.

[14] Yiling Chen and Bo Waggoner. Informational substitutes. In *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 239–247, 2016.

[15] Robert T Clemen and Robert L Winkler. Combining economic forecasts. *Journal of Business and Economic Statistics*, 4(1):39–46, 1986.

[16] Marquis de Condorcet. Essai sur l'application de l'analyse a la probability des decisions rendues a la plurality des voix. Paris: De l'imprimerie royale. *Condorcet: Selected Writings, ed. by KM Baker. Indianapolis, IN, Bobbs Merrill*, 1785. Translated in 1976 to "Essay on the Application of Mathematics to the Theory of Decision-Making".

[17] Henrique De Oliveira, Yuhta Ishii, and Xiao Lin. Robust merging of information. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC)*, pages 341–342, 2021.

[18] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345): 118–121, 1974.

[19] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 20(1):134–144, 2002.

[20] Shaddin Dughmi. Submodular functions: Extensions, distributions, and algorithms. A survey. Available at `https://arxiv.org/abs/0912.0322`, 2009.

[21] Ulle Endriss. Judgment aggregation. In *Handbook of Computational Social Choice*, chapter 17, pages 399–426. Cambridge University Press, 2016.

[22] Timothy Feddersen and Wolfgang Pesendorfer. Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 92(1):23–35, 1998.

[23] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. In *Uncertainty in Economic Theory*, pages 141–151. Routledge, 2004.

[24] Wei He and Jiangtao Li. Correlation-robust auction design. *Journal of Economic Theory*, 200: 105403, 2022.

[25] Victor Richmond R Jose and Robert L Winkler. Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24(1):163–169, 2008.

[26] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6): 2590–2615, 2011.

[27] Jean-François Laslier and Jörgen W Weibull. An incentive-compatible Condorcet jury theorem. *The Scandinavian Journal of Economics*, 115(1):84–108, 2013.

[28] Gilat Levy and Ronny Razin. Correlation neglect, voting behavior, and information aggregation. *American Economic Review*, 105(4):1634–45, 2015.

[29] Gilat Levy and Ronny Razin. A maximum likelihood approach to combining forecasts. *Theoretical Economics*, 16(1):49–71, 2021.

[30] Gilat Levy and Ronny Razin. Combining forecasts in the presence of ambiguity over correlation structures. *Journal of Economic Theory*, 199:105075, 2022.

[31] Andrew McLennan. Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents. *American Political Science Review*, 92(2):413–418, 1998.

[32] Paul Newbold and Clive WJ Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, 137(2): 131–146, 1974.

[33] Eric Neyman and Tim Roughgarden. Are you smarter than a random expert? The robust aggregation of substitutable signals. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, pages 990–1012, 2022.

[34] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5):1155–1174, 2005.

[35] Tim Roughgarden and Inbal Talgam-Cohen. Approximately optimal mechanism design. *Annual Review of Economics*, 11:355–381, 2019.

[36] Leonard Jimmie Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46(253):55–67, 1951.

[37] James H Stock and Mark W Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430, 2004.

[38] Hal Varian. Nowcasting COVID-19, 2020. `https://simons.berkeley.edu/talks/nowcasting-covid-19`.

[39] Vijay V Vazirani. *Approximation Algorithms*, volume 1. Springer, 2001.

[40] Abraham Wald. *Statistical Decision Functions*. Wiley, 1950.

[41] Juntao Wang, Yang Liu, and Yiling Chen. Forecast aggregation via peer prediction. In *Proceedings of the 9th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 131–142, 2021.

[42] Victor Zarnowitz. The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting*, 3(1):11–26, 1984.

A Extensions

In this appendix, we discuss the generality of our results from the paper body. We start by showing that all our results hold even without the assumption of binary marginal posteriors (Subsection A.1). We proceed by showing that our main results on the optimality of random dictator (Theorem 4.3 and Corollary 4.4) hold for an arbitrary state space (Subsection A.2). Then we establish that throughout the paper, one can drop the assumption that agents' reports to the DM are anonymous (Subsection A.3). Finally, we show how the proof of Theorem 4.3 implies that for a sufficiently large number of agents, the random dictator aggregation rule also optimizes the approximation ratio (Subsection A.4). We conclude that for a large enough number of agents, the random dictator rule is the unique optimal aggregation rule regardless of the robustness paradigm – minimax, regret or approximation ratio. Finally, we discuss the truthfulness assumption on the agents in Subsection A.5.

A.1 Non-Binary Signals

As we mentioned in Section 2, we assume that $|S| = 2$ purely for simplicity. Here, we show that all our results continue to hold in a much more general setting, provided that the agents still submit binary recommendations.

We consider the model described in Section 2 with the following differences. We take an arbitrary set of signals $S \subseteq [0, 1]$, with each signal labeled by the marginal posterior it induces on an agent observing it. We consider identical distributions over marginal posteriors $\pi_1, \ldots, \pi_n$ of agents no. $1, \ldots, n$ (respectively), with $\pi_i \in \Delta(\Omega \times S)$ being the projection of $\pi \in \Delta(\Omega \times S^n)$ to the first and the $i + 1$-th coordinates. We assume that $\min\{\Pr_{q\sim\pi_1}[q < 1/2], \Pr_{q\sim\pi_1}[q > 1/2]\} > 0$, where q denotes the marginal posterior probability for $\omega = 1$.[11]

Each agent makes a reports in $\{H, L\}$. Specifically, if her marginal posterior $q \in S$ satisfies $q \geq 1/2$ – she chooses the *high report (H)*, which is interpreted as $\omega = 1$ being the likely possibility; otherwise, she chooses the *low report (L)*. Similarly to the basic setting, our ignorant DM only observes the fraction of agents with the H report, but not the report of each agent separately. Moreover, the DM knows the distribution over marginal posteriors, but not the full information structure π. In contrast, the *Bayesian* DM observes the report of each agent separately, and she also knows the full information structure π.[12]

Denote $p_1 := \mathbb{E}_{q\sim\pi_1}[q|q < 1/2]$ and $p_2 := \mathbb{E}_{q\sim\pi_1}[q|q \geq 1/2]$ (both are well-defined by our assumption on the marginal distributions). Then $p_1 < 1/2 < p_2$. We claim that all the proofs we presented in the paper work in this more general setting.

First, note that given an information structure π in the general setting and the induced information structure π' with the only possible marginal posteriors being p_1, p_2 – each aggregation function f satisfies:

$$\mathbb{E}_{\nu\sim\hat{\pi}_1}[f(\nu)] = \mathbb{E}_{\nu\sim\hat{\pi'}_1}[f(\nu)],$$

and:

$$\mathbb{E}_{\nu\sim\hat{\pi}_0}[1 - f(\nu)] = \mathbb{E}_{\nu\sim\hat{\pi'}_0}[1 - f(\nu)].$$

Indeed, the ignorant DM can emulate any aggregation function from the basic setting also in the new setting without making any changes, and vice versa.

Denote by $s^B(\cdot)$ the state that the Bayesian DM guesses as a function of ν. To finish modifying our proofs to the general setting – it is enough to prove an analogous condition to the above for the Bayesian DM, which is provided by the following lemma.[13]

Lemma A.1. *Fix $\omega' \in \Omega$. Then for any information structure π in the general setting and the corresponding induced information structure π' in the binary-posterior setting from Section 2, it holds that:*

$$\mathbb{E}_{\nu'\sim\hat{\pi'}_{\omega'}}\left[\Pr\left[s^B(\nu') \neq \omega'\right]\right] = \mathbb{E}_{\nu\sim\hat{\pi}_{\omega'}}\left[\Pr\left[s^B(\nu) \neq \omega'\right]\right]$$

Proof. The key idea is that the Bayesian DM cannot decrease the mistake probability knowing π compared to π' as she does not observe the exact posterior of each agent, but only her recommendation. We shall only prove the lemma for $\omega' = 1$ (the case $\omega' = 0$ is analogous).

Indeed, we have:

$$\mathbb{E}_{\nu'\sim\hat{\pi'}_1}\left[\Pr\left[s^B(\nu') \neq 1\right]\right] = \Pr_{q\sim\pi'(\omega=1)}[q < 1/2] = \Pr_{q\sim\pi'(\omega=1)}[q = p_1] =_{(*)} \Pr_{q\sim\pi(\omega=1)}[q < 1/2] = \mathbb{E}_{\nu\sim\hat{\pi}_1}\left[\Pr\left[s^B(\nu) \neq 1\right]\right],$$

where $(*)$ holds by definition of π'. □

A.2 Arbitrary State Space and Utilities

In this section, we weaken the assumption that the state space is binary and the DM just tries to guess the correct state. Rather, we just assume that Ω is finite, and the state space is equipped with some prior distribution $\mu = (\mu_\omega)_{\omega\in\Omega}$. The DM has to choose between a *default action* yielding her a constant

[11] If, e.g., all the elements of S induce posteriors not below $1/2$ – the DM clearly should just always guess $\omega = 1$.

[12] We stress that neither the ignorant nor the Bayesian DM observe the exact marginal posteriors of the agents, but only their recommendations.

[13] Note that the expected utility of the ignorant DM in both settings is the convex combination of the two quantities above, with respective weights μ and $1 - \mu$.

utility of 0 regardless of the state, and an *optional action* yielding her a state-dependent utility according to some function $u : \Omega \to \mathbb{R}$.[14]

Each agent recommends taking one of the actions. An agent with a posterior belief $q \in \Delta(\Omega)$ recommends taking the optional action if in expectation over q, the DM is weakly better off taking it. Formally: $\mathbb{E}_{\omega\sim q}[u(\omega)] \geq 0$. Otherwise, she recommends taking the default action. The (marginal) information structure π_i for a single agent is identical for all agents and is given as follows. Agent i ($i \in [n]$) gets a binary private signal $s_i \in \{L, H\} = S$, with the corresponding posteriors being $p^L = \left(p^L_\omega\right)_{\omega\in\Omega}, p^H = \left(p^H_\omega\right)_{\omega\in\Omega}$. We shall assume that $\mathbb{E}_{\omega\sim p^H}[u(\omega)] > 0$ and $\mathbb{E}_{\omega\sim p^L}[u(\omega)] < 0$ – that is, upon observing H the agent would recommend taking the optional action, while upon observing L she would recommend the default action.[15] Denote by $\Omega^H \subseteq \Omega$ the set of states for which the DM is weakly better off taking the optional action, and define $\Omega^L := \Omega \setminus \Omega^H$. That is, $\Omega^H = \{\omega' \in \Omega : u(\omega') \geq 0\}$ and $\Omega^L = \{\omega' \in \Omega : u(\omega') < 0\}$.

The ignorant DM and the Bayesian DM both observe the fraction of agents recommending H. The ignorant DM knows just the identical marginal information structure $\pi_i \in \Delta(\Omega \times S)$, while the Bayesian DM is aware of the joint information structure $\pi \in \Delta(\Omega \times S^n)$. The aggregation rule f gets as input the fraction of agents recommending H and outputs the probability of taking the optional action.

For any state $\omega' \in \Omega$, the probability of a specific agent to have the high posterior p^H conditional on the state ω' is:

$$a_{\omega'} = \frac{p^H_{\omega'}\left(\mu_{\omega'} - p^L_{\omega'}\right)}{\mu_{\omega'}\left(p^H_{\omega'} - p^L_{\omega'}\right)}.$$

The regret of an aggregation rule f, $\mathrm{Reg}(f)$, is the difference between the expected utilities of a Bayesian DM knowing π and the ignorant DM not knowing π. The following generalization of Theorem 4.3 holds.

Theorem A.2. *Suppose $\frac{1}{n} \leq a_{\omega'} \leq \frac{n-1}{n}$ for every $\omega' \in \Omega$. Then the random dictator aggregation rule is uniquely optimal. Moreover:*

$$\mathrm{Reg} = \sum_{\omega'\in\Omega^H} \mu_{\omega'} u(\omega') - \sum_{\omega'\in\Omega} \mu_{\omega'} a_{\omega'} u(\omega').$$

Let us describe the intuition behind the proof of Theorem A.3, which follows similar ideas to the proof of Theorem 4.2 (the binary case). We first note that the set of feasible distributions over frequencies of posteriors has an analogous neat characterization for an arbitrary state space (rather than binary).

Theorem A.3 (Arieli and Babichenko [2] - A generalization of Theorem 4.2)**.** *Let $\left(\hat{\rho}_{\omega'}\right)_{\omega'\in\Omega} \in \left(\Delta\left(\left\{0, \frac{1}{n}, \ldots, 1\right\}\right)\right)^{|\Omega|}$ be a tuple of conditional distributions over frequencies of the H-signal. There exists an information structure π with $\hat{\pi}_{\omega'} = \hat{\rho}_{\omega'}$ for every $\omega' \in \Omega$ if and only if $\mathbb{E}[\hat{\rho}_{\omega'}] = a_{\omega'}$ for every $\omega' \in \Omega$.*

Following the same outline as of Proposition 4.1 proof, with the uses of Theorem 4.2 replaced by Theorem A.3, we deduce the following result (the proof is omitted due to the similarity to Proposition 4.1).

Proposition A.4. *For every $\{a_{\omega'}\}_{\omega'\in\Omega}, n, \mu,$ and f:*

$$\mathrm{Reg}(f) \leq \sum_{\omega'\in\Omega^H} \mu_{\omega'} u(\omega') - \sum_{\omega'\in\Omega^L} \mu_{\omega'} \mathrm{cav}[f](a_{\omega'}) u(\omega') - \sum_{\omega'\in\Omega^H} \mu_{\omega'} \mathrm{vex}[f](a_{\omega'}) u(\omega').$$

Similarly to the binary-state case, whenever there exists an optimal information structure π for which the supports of $\hat{\pi_{\omega'}}$ with $\omega' \in \Omega^H$ are disjoint from the supports of $\hat{\pi_{\omega'}}$ with $\omega' \in \Omega^L$ – the inequality in Proposition A.4 is tight. Turning this idea into Theorem 4.2 proof is similar to Theorem 4.3 proof. We start with a lemma lower-bounding $\mathrm{Reg}(f)$.

Lemma A.5. *Fix f and let l be the line passing through $(0, f(0))$ and $(1, f(1))$. Suppose there exist distinct $x_1, x_2, x_3, x_4 \in \left\{0, \frac{1}{n}, \ldots, 1\right\}$ s.t. $x_1 \leq a_{\omega'} \leq x_2$ for every $\omega' \in \Omega^L$ and $x_3 \leq b \leq x_4$ for every $\omega' \in \Omega^H$. Suppose further that $(x_1, f(x_1)), (x_2, f(x_2))$ is not below l and $(x_3, f(x_3)), (x_4, f(x_4))$ not above l. Then:*

$$\mathrm{Reg}(f) \geq \mathrm{Reg}(l) = \sum_{\omega'\in\Omega^H} \mu_{\omega'} u(\omega') - \sum_{\omega'\in\Omega} \mu_{\omega'} (a_{\omega'}(f(1) - f(0)) + f(0)) u(\omega').$$

[14]Note that the normalization of the DM's utility in one of the actions to be constantly 0 is w.l.o.g.

[15]If an agent would (weakly) recommend the same action regardless of the signal, then the DM can just ignore the recommendations.

Proof. Assume that there exist x_1, x_2, x_3, x_4 as described. As in Lemma C.1 proof, we get that the Bayesian DM's expected utility under l is independent of the information structure π; moreover, the Bayesian DM's expected utility equals its maximum possible value $\sum_{\omega' \in \Omega^H} \mu_{\omega'} u(\omega')$ for π inducing the distributions $(\hat{\pi}_{\omega'})_{\omega' \in \Omega}$ over values of ν with respective expectations equal to $a_{\omega'}$ such that $\mathrm{supp}(\hat{\pi}_{\omega'}) = \{x_1, x_2\}$ for $\omega' \in \Omega^L$ and $\mathrm{supp}(\hat{\pi}_{\omega'}) = \{x_3, x_4\}$ for $\omega' \in \Omega^H$ (note that the distributions $\hat{\pi}_{\omega'}$ are uniquely determined). Thus, it is enough to prove that for the above π, the expected ignorant DM's utility under f is at most as under l.

Indeed, by the assumption on x_1, x_2, x_3, x_4, the ignorant DM's utility loss in some $\omega' \in \Omega^L$ for taking the optional action compared to taking the default action is at least $a_{\omega'}(f(1) - f(0)) - f(0)|u(\omega')|$, while her utility gain in some $\omega' \in \Omega^H$ for taking the optional action compared to the default action is at most $(a_{\omega'}(f(1) - f(0)) + f(0))u(\omega')$. Therefore, the expected utility of the ignorant DM under f is at most $\sum_{\omega' \in \Omega} \mu_{\omega'}(a_{\omega'}(f(1) - f(0)) + f(0))u(\omega')$, which is her expected utility under l. □

Proof of Theorem A.2. Let l be the line passing through $(0, f(0))$ and $(1, f(1))$, and let $f(x) \equiv mx + k$ (for constants m, k) be its equation. The proof that for any choice of f, the inequality ensured by Lemma A.5 holds is done exactly as in Theorem 4.3 proof, with a and b replaced by the sets of $a_{\omega'}$ for which $\omega' \in \Omega^L$ and $\omega' \in \Omega^H$, respectively. In particular, there exists an optimal f which is linear.

Note that $k = f(0)$ and $m = f(1) - f(0)$. From Lemma A.5, we have for any f:

$$\mathrm{Reg}(f) \geq \sum_{\omega' \in \Omega^H} \mu_{\omega'} u(\omega') - m \sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega') - k \sum_{\omega' \in \Omega} \mu_{\omega'} u(\omega').$$

Moreover, for a linear f we have equality.

The only constraints on m, k are $k = f(0) \in [0, 1]$ and $m + k = f(1) \in [0, 1]$. To minimize the bound, we must either have $k = 1$ and $m \in \{-1, 0\}$, or $k = 0$ and $m \in \{0, 1\}$. Therefore, for any f:

$$\mathrm{Reg}(f) \geq \sum_{\omega' \in \Omega^H} \mu_{\omega'} u(\omega') - \max\left\{ \sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega'), \sum_{\omega' \in \Omega} \mu_{\omega'} (1 - a_{\omega'}) u(\omega'), \sum_{\omega' \in \Omega} \mu_{\omega'} u(\omega'), 0 \right\}.$$

Note that $\sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega')$ can be viewed as the expected utility of a DM in a setting with just a single agent. As for any posterior, an agent recommends the weakly better action, this expected utility is greater or equal to the utility of any fixed action. That is:

$$\sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega') \geq \max\left\{ \sum_{\omega' \in \Omega} \mu_{\omega'} u(\omega'), 0 \right\}.$$

Moreover, we have:

$$\sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega') + \sum_{\omega' \in \Omega} \mu_{\omega'} (1 - a_{\omega'}) u(\omega') = \sum_{\omega' \in \Omega} \mu_{\omega'} u(\omega') < \sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega'),$$

which implies $\sum_{\omega' \in \Omega} \mu_{\omega'} (1 - a_{\omega'}) u(\omega') < 0$. Thus, for any f:

$$\mathrm{Reg}(f) \geq \sum_{\omega' \in \Omega^H} \mu_{\omega'} u(\omega') - \sum_{\omega' \in \Omega} \mu_{\omega'} a_{\omega'} u(\omega').$$

Furthermore, for $f(x) \equiv x$ equality holds. Therefore, the random dictator rule is optimal. The uniqueness follows from the same arguments as in Theorem 4.3 proof, with Lemma C.1 replaced by Lemma A.5, which can be done since $\Omega^L = \{\omega' \in \Omega : u(\omega') < 0\}$, $\Omega^H = \{\omega' \in \Omega : u(\omega') \geq 0\}$ and there exists $\omega' \in \Omega^H$ for which a strict inequality holds. □

As a corollary of Theorem A.2, given any fixed prior and marginal posteriors – for a large enough number of agents n, the random dictator rule is uniquely optimal. One can further drop the assumption that each agent gets a binary signal (as was done in Subsection A.1).

A.3 Non-Anonymity

We now prove that the assumption that the DM does not observe the identities of the agents is w.l.o.g. Namely, let us consider a setting in which the DM's aggregation rule $f : 2^{[n]} \to [0, 1]$ depends also on the identity of the agents, where the domain of f is the set of all agents getting the high signal.[16] Denote by $\widehat{P^*}(\pi), \widehat{P}(f,\pi), \widehat{\text{Reg}}(f,\pi), \widehat{\text{Reg}}(f)$ and $\widehat{\text{Reg}}$ the following quantities in the non-anonymous setting, respectively: the probability of a correct guess by the Bayesian DM under the information structure π, the probability of a correct guess by the ignorant DM under the information structure π and the aggregation rule f, the regret under f and π, the maximum regret of f and the minimal regret in the setting. We claim that $\text{Reg} = \widehat{\text{Reg}}$.

For a permutation τ on $[n]$, an aggregation rule f and a distribution over information structures ϕ, let f_τ and ϕ_τ be, respectively, the aggregation rule and the information structure obtained from f and ϕ upon applying τ on the agents' identities. We need the following lemma.

Lemma A.6. *Consider the two-player zero-sum game between the DM and the adversary in which the DM chooses a non-anonymous aggregation rule f, the adversary chooses a distribution over information structures ϕ, and the payoff is $\mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f,\pi)\right]$. Then there exist optimal DM's and adversary's strategies f^* and ϕ^*, respectively, such that $f^* = f^*_\tau$ and $\phi^* = \phi^*_\tau$ for any permutation τ on $[n]$.*

Note that Lemma A.6 immediately implies $\text{Reg} = \widehat{\text{Reg}}$. Indeed, let f^* and ϕ^* be non-anonymous optimal aggregation rule and distribution over information structures (respectively) invariant under agents' permutations. If f^* is suboptimal in the anonymous setting, then there exists ϕ with $\mathbb{E}_{\pi\sim\phi}[\text{Reg}(f^*,\pi)] > \mathbb{E}_{\pi\sim\phi^*}[\text{Reg}(f^*,\pi)]$. However, we have $\mathbb{E}_{\pi\sim\phi}[\text{Reg}(f^*,\pi)] = \mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f^*,\pi)\right]$ and $\mathbb{E}_{\pi\sim\phi^*}[\text{Reg}(f^*,\pi)] = \mathbb{E}_{\pi\sim\phi^*}\left[\widehat{\text{Reg}}(f^*,\pi)\right]$, implying $\mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f^*,\pi)\right] > \mathbb{E}_{\pi\sim\phi^*}\left[\widehat{\text{Reg}}(f^*,\pi)\right]$, a contradiction. Similarly, we get that ϕ^* is optimal in the anonymous setting.

Proof of Lemma A.6. Fix an optimal aggregation rule f. Define $f^* : 2^{[n]} \to [0,1]$ by $f^*(S) := \sum_\tau \frac{f_\tau(S)}{n!}$. Then trivially $f^*(S) = \frac{\sum_{S'\subseteq[n]:|S'|=|S|} f(S)}{\binom{n}{|S|}}$ depends only on $|S|$ – i.e., it is invariant under permutations on the agents' identities. Let ϕ be an adversary's optimal strategy.

By the agents' symmetry, the aggregation rule f_τ is optimal for any permutation τ on $[n]$. Thus, $\widehat{\text{Reg}} = \mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f_\tau,\pi)\right]$ for any τ, implying $\widehat{\text{Reg}} = \mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f^*,\pi)\right]$. Therefore, f^* is an optimal aggregation rule invariant under agents' permutations.

Define ϕ^* to be a distribution over information structures choosing uniformly at random a permutation τ on $[n]$ and then choosing an information structure according to ϕ_τ. Since f^* is invariant under any permutation τ on the agents, and the agents are symmetric – we get $\mathbb{E}_{\pi\sim\phi_\tau}\left[\widehat{\text{Reg}}(f^*,\pi)\right] = \mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f^*,\pi)\right]$ for any permutation τ, implying $\mathbb{E}_{\pi\sim\phi^*}\left[\widehat{\text{Reg}}(f^*,\pi)\right] = \mathbb{E}_{\pi\sim\phi}\left[\widehat{\text{Reg}}(f^*,\pi)\right] = \widehat{\text{Reg}}$. Thus, ϕ^* is an optimal adversary's strategy invariant under agents' permutations, as needed. $\square$

A.4 Approximation Ratio

In this section, we consider an alternative robustness paradigm – *approximation ratio.* This approach differs from regret analysis by considering the *ratio* of the probabilities of the correct guesses of the ignorant and the Bayesian DM, rather than their difference. Formally, the *approximation ratio of the aggregation rule f in an information structure* π is $\text{Appr}(f,\pi) := \frac{P(f,\pi)}{P^*(\pi)}$. The *approximation ratio of an aggregation rule* f is $\text{Appr}(f,\pi) := \min_\pi \text{Appr}(f,\pi)$.[17]

Proposition A.7. *Suppose $\frac{1}{n} \le a < b \le \frac{n-1}{n}$.*[18] *Then random dictator is the unique approximation ratio-maximizing aggregation rule. Moreover,* $\text{Appr} = (1-\mu)\cdot(1-a) + \mu\cdot b$.

Proof. Let f be an aggregation rule. By the proof of Theorem 4.3, for $\frac{1}{n} \le a < b \le \frac{n-1}{n}$ – when the adversary targets maximizing the regret, there always exists an optimal adversary's information structure π for which the supports of $\hat{\pi}_0$ and $\hat{\pi}_0$ are disjoint. Such a strategy reveals the value of ω to a Bayesian DM with the maximum possible probability of 1. Therefore, $\text{Appr}(f)$ is at most the probability of a correct guess of ω under f, which is shown in Proposition 3.2 to be at most $(1-\mu)\cdot(1-a)+\mu\cdot b$, with this bound achieved exclusively by the random dictator rule.

[16]Note that as in the previous two subsections, one can extend the results to arbitrary posteriors, and extend Theorem 4.3 and its corollaries to arbitrary finite state spaces.

[17]Note that this minimum exists since $\text{Appr}(f,\cdot)$ is a continuous functional defined on a compact space.

[18]Note that this condition holds for all $n \ge \max\left\{\frac{(1-\mu)(p_2-p_1)}{(1-p_2)(\mu-p_1)}, \frac{\mu(p_2-p_1)}{p_1(p_2-\mu)}\right\}$.

Conversely, let the DM use the random dictator rule. Then the proof of Proposition 3.2 implies that the DM guesses ω correctly with probability $(1-\mu)\cdot(1-a)+\mu\cdot b$. In particular, the approximation ratio of random dictator is lower-bounded by this value, as needed. $\square$

We should also mention that one can get an analogous result in the setting with arbitrary state space and utilities presented in Subsection A.2 – the random dictator rule remains optimal.

A.5 Strategic Agents

We have assumed that agents simply report the better recommendation conditional on their information. Our robustness analysis advocates the usage of the random dictator rule as an aggregation rule even for strategic agents. Here we highlight another desirable property of the random dictator rule: If agents are strategic and they know that the aggregation rule is the random dictator, then it is a dominant strategy for them to report the better recommendation conditional on their information. Namely, in addition to being robustly optimal the random dictator, it is also immune to scenarios in which some (or all) agents start acting strategically instead of truthfully.

B Minimax versus Regret

In this appendix, we demonstrate that the celebrated result of Carroll [13] on the minimax-optimality of selling goods separately does not extend to regret optimality. We further discuss another setting ([8]) in which different notions of robustness lead to different robustly optimal strategies.

Carroll [13] consider a classical combinatorial auction setting. A single seller suggests n goods to a single buyer. The buyer's valuation for the goods is additive, and it is described by a vector $v=(v_i)_{i\in[n]}$ that specifies her utility for each good. The vector v is drawn according to a distribution π. By the direct revelation principle, a mechanism for selling the goods might be described as a (possibly infinite) menu M from which the buyer picks a single item. Each menu item takes the form $m=(q_1,...,q_n,p)\in M$, where $q_i\in[0,1]$ specifies the probability of the good to be allocated to the buyer, and $p>0$ is the price of this menu item. We denote by $M^*(\pi)$ the optimal mechanism (menu). Let $s(M,\pi)$ denote seller's revenue in case she uses mechanism M.

Carroll [13] study a scenario in which the seller does not know π (and hence cannot pick $M^*(\pi)$), but only knows the marginal distribution of valuations on each good separately. Namely, the seller knows π_i, where π_i is the distribution of v_i. The correlation between the v_is is unknown to the seller. The seller chooses a mechanism M as a function of $((\pi_1,...,\pi_n)$ only. The seller's goal is to choose a mechanism that performs robustly well; namely, for all possible correlations. Therefore, the *minimax* guarantee and the *regret* guarantee of a mechanism M are defined by

$$\mathrm{Minmax}(M)=\inf_{\pi \text{ with marginals } (\pi_1,\dots,\pi_n)} s(M,\pi)$$
$$\mathrm{Reg}(M)=\sup_{\pi \text{ with marginals } (\pi_1,\dots,\pi_n)} [s(M^*(\pi),\pi)-s(M,\pi)]$$

The result of [13] states that selling the goods separately (with an optimal price on each good i that is determined by π_i only), which is denoted by M_{sep}, is minimax-optimal – i.e., it maximizes $\mathrm{Minmax}(M)$ across all mechanisms). Below we demonstrate an example in which M_{sep} is not regret-optimal (i.e., it does not minimize $\mathrm{Reg}(M)$).

Consider the case in which n is even and π_i is the uniform distribution over $\{1,2\}$ for all goods $i\in[n]$. An optimal price for each good separately is 1, and it ensures a deterministic revenue of 1. Hence, the total revenue is n. We argue that $\mathrm{Reg}(M_{sep})=\frac{n}{2}$. Indeed, let π be the uniform distribution over the two vectors $\{v,v'\}$, where $v=(1,2,1,2,\dots,1,2)$ and $v'=(2,1,2,1,\dots,2,1)$. Selling all goods as a single bundle at a price of $\frac{3n}{2}$, which is denoted by M_{bun}, ensures selling with probability 1, yielding a revenue of $\frac{3n}{2}$ for the seller. Moreover, the seller cannot extract more than $\frac{3n}{2}$ revenue, because the revenue is bounded by buyer's expected value for the entire bundle.

Assume by way of contradiction that M_{sep} minimizes regret; i.e., the minimal regret is $\frac{n}{2}$. We view the regret-minimization problem as a zero-sum game between the seller (who chooses M) and the adversary (who chooses π). A (mixed) strategy of the adversary in this zero-sum game is a distribution ϕ over πs. By the minimax theorem, there exists a mixed strategy ϕ for the adversary such that for every M, we have: $\mathbb{E}_{\pi\sim\phi}[s(M^*(\pi),\pi)-s(M,\pi)]\geq\frac{n}{2}$. In particular, this should hold for $M=M_{sep}$. Note, however, that $s(M^*(\pi),\pi)\leq\frac{3n}{2}$ (because it is buyer's expected value for the entire bundle) and that $s(M_{sep},\pi)=n$.

Therefore, to achieve the $\frac{n}{2}$ difference ϕ must be supported on πs for which $s(M^*(\pi), \pi) = \frac{3}{2}$. This can happen only if all goods are sold with probability 1 (otherwise the full revenue cannot be extracted) and we must have $v_1 + ... + v_n = \frac{3n}{2}$ with probability 1 (again because otherwise the full revenue cannot be extracted). Therefore, ϕ is supported on πs for which buyer's value for the entire bundle is $\frac{3n}{2}$ with probability 1. Therefore, against the mixed strategy ϕ the seller can use M_{bun} and have a regret of 0, which is a contradiction.

Further discussion. Another example that demonstrates the differences of the robustness approaches appears in Babichenko et al. [8]. Babichenko et al. [8] study the Bayesian persuasion problem ([26]) with uncertainty of the sender about the receiver's utility. In this setting the minimax approach is very pessimistic: It is hopeless to have a non-trivial upper bound when the number of states is large. The regret approach succeeds in providing non-trivial asymptotic (in the number of states) bounds on the regret. Interestingly, in the same setting providing non-trivial asymptotic bounds on the approximation ratio turns to be impossible, whereas for a constant number of states bounds are provided.

C Proof of Theorem 4.3

We first lower-bound Reg under the assumption that one can separate the supports of $\hat{\pi}_0$ and $\hat{\pi}_1$. We shall show that the DM betters off by replacing f with the segment connecting the points $(0, f(0))$ and $(1, f(1))$.

Lemma C.1. *Fix f and let l be the line passing through $(0, f(0))$ and $(1, f(1))$. Suppose there exist distinct $x_1, x_2, x_3, x_4 \in \left\{0, \frac{1}{n}, \ldots, 1\right\}$ s.t. $x_1 \le a \le x_2$, $x_3 \le b \le x_4$ with $(x_1, f(x_1)), (x_2, f(x_2))$ not below l and $(x_3, f(x_3)), (x_4, f(x_4))$ not above l.*[19] *Then:*

$$\mathrm{Reg}(f) \ge \mathrm{Reg}(l) = 1 - (1-\mu) \cdot (1 - a(f(1) - f(0))) - \mu(b(f(1) - f(0))) - (2\mu - 1)f(0).$$

Proof. Assume that there exist x_1, x_2, x_3, x_4 as described. Note first that since l is a linear function, $P(l, \pi)$ is independent of the information structure π. Moreover, $P^*(\pi) = 1$ for π inducing the distributions $\hat{\pi}_0, \hat{\pi}_1$ over values of ν with $\mathrm{supp}(\hat{\pi}_0) = \{x_1, x_2\}$ and $\mathrm{supp}(\hat{\pi}_1) = \{x_3, x_4\}$, and respective expectations a, b (note that the distributions $\hat{\pi}_0$, $\hat{\pi}_1$ are uniquely determined). Indeed, separating these supports allows the Bayesian DM to guess ω correctly with probability 1. Thus, $\mathrm{Reg}(l) = 1 - P(l, \pi)$ for the above π. As $\mathrm{Reg}(f) \ge \mathrm{Reg}(f, \pi)$, it is enough to prove that $P(f, \pi) \le P(l, \pi)$.

Indeed, by the assumption on x_1, x_2, x_3, x_4, we have $\mathbb{E}_{X_0 \sim \hat{\pi}_0}[1 - f(X_0)] \le (1 - a(f(1) - f(0))) - f(0)$. Similarly, we get $\mathbb{E}_{X_1 \sim \hat{\pi}_1}[f(X_1)] \le b(f(1) - f(0)) + f(0)$.

Therefore:

$$P(f, \pi) \le (1-\mu) \cdot (1 - a(f(1) - f(0)) - f(0)) + \mu(b(f(1) - f(0)) + f(0)) = P(l, \pi),$$

and the lemma statement follows. ☐

Now we are ready to prove Theorem 4.3. We first show, using Lemma C.1, that there always exists an optimal f which is a linear function; the required x_1, x_2, x_3, x_4 from Lemma C.1 statement would be chosen from the set $\left\{0, \frac{1}{n}, 1 - \frac{1}{n}, 1\right\}$. Then we prove that the optimal f must be random dictator.

Proof of Theorem 4.3. Let l be the line passing through $(0, f(0))$ and $(1, f(1))$. We shall first prove that for any choice of f, the inequality ensured by Lemma C.1 holds. To find x_1, x_2, x_3, x_4 as needed in Lemma C.1, consider the possible cases regarding the graph of f:

1. $\left(\frac{1}{n}, f\left(\frac{1}{n}\right)\right)$ is not above l and $\left(\frac{n-1}{n}, f\left(\frac{n-1}{n}\right)\right)$ is not below l. Then one may take $x_1 = 0$, $x_2 = \frac{n-1}{n}$, $x_3 = \frac{1}{n}$, $x_4 = 1$ (note that $n > 2$).

2. $\left(\frac{1}{n}, f\left(\frac{1}{n}\right)\right)$ is strictly above l, and $\left(\frac{n-1}{n}, f\left(\frac{n-1}{n}\right)\right)$ is not below l. This time, one can take $x_1 = \frac{1}{n}$, $x_2 = \frac{n-1}{n}$, $x_3 = 0$, $x_4 = 1$.

3. $\left(\frac{n-1}{n}, f\left(\frac{n-1}{n}\right)\right)$ is strictly below l, and $\left(\frac{1}{n}, f\left(\frac{1}{n}\right)\right)$ is not above l. This case is symmetric to the previous one.

4. Otherwise, $\left(\frac{1}{n}, f\left(\frac{1}{n}\right)\right)$ is strictly above l and $\left(\frac{n-1}{n}, f\left(\frac{n-1}{n}\right)\right)$ is strictly below l. One can take $x_1 = \frac{1}{n}$, $x_2 = 1$, $x_3 = 0$, $x_4 = \frac{n-1}{n}$.

[19] Formally, e.g., $f(x_1) \ge (f(1) - f(0))(x - x_1) + f(0)$.

Therefore, there exists an optimal f which is linear – i.e., f coincides with l. Fix an optimal linear f and write $f(x) = mx + k$ for some constants m, k.

As $k = f(0)$ and $m = f(1) - f(0)$, we get:

$$\begin{aligned}&\mathrm{Reg}(f) = 1 - ((1-\mu)\cdot(1 - a(f(1) - f(0))) + \mu(b(f(1) - f(0))) + (2\mu - 1)f(0)) = \\ &\mu + k\cdot(1-2\mu) + m\cdot((1-\mu)a - \mu b).\end{aligned}$$

The only constraints on m, k are $k = f(0) \in [0,1]$ and $m + k = f(1) \in [0,1]$. Note that $(1-\mu)a - \mu b = \frac{(\mu - p_1)(1-2p_2)}{p_2 - p_1} < 1 - 2\mu$ as $p_1 < \frac{1}{2}, \mu < p_2$. Therefore, to minimize $\mathrm{Reg}(f)$, one should take $k = 0$. Moreover, as $p_2 > \frac{1}{2}$, one should take $m = 1$. Thus:

$$\mathrm{Reg}(f) \geq 1 - (1-\mu)(1-a) - \mu b.$$

Since the right-hand side is exactly the regret of the random dictator rule, the optimality of random dictator follows.

To show uniqueness, consider some optimal f. If f is not the random dictator rule, there exists some $i \in [n]$ with $\left(\frac{i}{n}, f\left(\frac{i}{n}\right)\right)$ either strictly above l or strictly below l. Assume w.l.o.g. That the former holds. The inequality in Lemma C.1 must be tight and the lower bound provided by this lemma must equal $1-(1-\mu)(1-a)-\mu b$. Thus, we must have that for any choices of x_1, x_2, x_3, x_4 satisfying the conditions of Lemma C.1: $i \neq x_1, x_2, x_3, x_4$. In particular, $1 < i < n-1$. Note that either $\frac{i}{n} \leq a$ or $\frac{i}{n} \geq a$. Assume w.l.o.g. that the former holds. Then re-selecting x_1 form the value in $\left\{0, \frac{1}{n}\right\}$ in Lemma C.1 proof to $x_1 = \frac{i}{n}$ prevents from the inequality in Lemma C.1 to be tight, as the restriction of f to x_1, x_2, x_3, x_4 is not linear. □

D Analysis of Example 5.2

Proof. Similarly to the proof of Theorem 5.1, we consider a two-player zero-sum game between the DM and the adversary. The DM's pure strategies are choices of an aggregation function $f : \left\{0, \frac{1}{n}, \ldots, 1\right\} \to \{0,1\}$ and the adversary's pure strategies are choices of $x \in C$ (which represent admissible information structures). We denote the mixed strategy of the adversary by ϕ; note that mixtures over admissible information structures are themselves admissible information structures. The payoff function is:

$$\begin{aligned}\mathrm{Reg}(f,\phi) = \mathbb{E}_{x\sim\phi}\Bigg[\sum_{i=0}^{2}\Bigg(&\max\Bigg\{\sum_{D\subseteq\{0,1\}:\,|D|=i} x_D^0, \sum_{D\subseteq\{0,1\}:\,|D|=i} x_D^1\Bigg\} \\ &-\Bigg((1-f_i)\cdot\sum_{D\subseteq\{0,1\}:\,|D|=i} X_D^0 + f_i\cdot\sum_{D\subseteq\{0,1\}:\,|D|=i} x_D^1\Bigg)\Bigg)\Bigg].\end{aligned}$$

Denote the function f described in the proposition statement by f^*. We shall show that there exists ϕ^* s.t. (f^*, ϕ^*) is a mixed Nash equilibrium, which would complete the proof (after further computing the game value). We divide the proof to three cases according to the order between $a, b, 1/2$.

Case 1: $a < b \leq 1/2$**.** First, note that each $x \in \mathrm{supp}(\phi)$ that is a best reply to f satisfying $f(0) = 0$, $f(1) = 1$ must be a solution of the following convex program:[20]

$$\begin{aligned}\max_{p\in[0,a/2],\,q\in[0,b/2]}\Big\{&\max\{1/2 - a + p, 1/2 - b + q\} + 2\max\{a/2 - p, b/2 - q\} + \max\{p, q\} \\ &- (1/2 - a + p) - 2(a/2 - p)(1 - f(1/2)) - 2(b/2 - q)f(1/2) - q\Big\}.\end{aligned}$$

Denote the above target function by $h(p,q)$. As h is a convex function – it obtains its maximum at an extreme point of the domain. Hence, the only candidates for solutions are $(p,q) \in \{(0,0), (a/2, 0), (0, b/2), (a/2, b/2)\}$. Note that:

- $h(0,0) = (b-a)(1 - f(1/2))$.

[20]We denote $p := x_{1,2}^0$, $q := x_{1,2}^1$ and use the definition of C.

- $h(a/2,0) = a/2 + b(1 - f(1/2)) >_{\text{For } a>0} h(0,0)$.
- $h(0,b/2) = \max\{1/2 - a, 1/2 - b/2\} + a - 1/2 + af(1/2)$.
- $h(a/2,b/2) = 0 <_{\text{For } a>0} h(a/2,0)$.

Therefore, the only remaining candidates for maximizers of h are $(a/2,0)$ and $(0,b/2)$, regardless of f. A straightforward check implies that for the f minimizing $\max\{h(a/2,0), h(0,b/2)\}$, it must hold that $h(a/2,0) = h(0,b/2)$. If $2a \le b$, we get $f(1/2) = \frac{a/2+b}{a+b} = \frac{a+2b}{2(a+b)}$ and $\text{Reg} = \frac{a(a+2b)}{2(a+b)}$. Otherwise, we get $f(1/2) = \frac{3b-a}{2(a+b)}$ and $\text{Reg} = \frac{a^2+4ab-b^2}{2(a+b)}$.

Write $x \in C$ as $x = \left(x^0_\emptyset, x^0_1, x^0_2, x^0_{1,2}, x^1_\emptyset, x^1_1, x^1_2, x^1_{1,2}\right)$. Given an adversary's (pure) strategy $x \in C$, let δ_x be the embedding of x into the space of adversary's mixed strategies; that is, δ_x is a distribution over elements of C assigning probability 1 to x. From the arguments above, it follows that any distribution of the form: $\phi^* = \alpha\delta_{x^a} + (1-\alpha)\delta_{x^b}$, with $x^a := (1/2 - a, a/2, a/2, 0, 1/2 - b/2, 0, 0, b/2)$, $x^b := (1/2 - a/2, 0, 0, a/2, 1/2 - b, b/2, b/2, 0)$ and $\alpha \in [0,1]$, is a best reply to f^*. If remains to show that f^* is a best reply to ϕ^* for a suitable α. To this end, note that from linearity of expectation and the payoff function in the values of f, necessarily there exists a best reply f to ϕ^* getting values in $\{0,1\}$. Take $\alpha = \frac{b}{a+b}$.

Consider the following cases:

- If $2a \le b$ – by inspection, one can check that $f(0) = f(1/2) = 0$, $f(1) = 1$ is a best reply to ϕ^*. Moreover, it yields expected payoff of $0 \cdot \frac{b}{a+b} + \frac{2b+a}{2} \cdot \frac{a}{a+b} = \frac{a^2+2ab}{2(a+b)} = \text{Reg}(f^*, \phi^*)$. Hence, f^* is also a best reply to ϕ^*, as desired.
- Otherwise, a straightforward check implies that $f(0) = f(1/2) = 0$, $f(1) = 1$ is a best reply to ϕ^*, yielding expected payoff of $\frac{2a-b}{2} \cdot \frac{b}{a+b} + \frac{2b+a}{2} \cdot \frac{a}{a+b} = \frac{a^2+4ab-b^2}{2(a+b)} = \text{Reg}(f^*, \phi^*)$; thus, f^* is also a best reply to ϕ^*, as needed.

Case 2: $1/2 \le a < b$. This case is symmetric to the previous one, with a replaced by $1-b$, and b replaced by $1-a$.

Case 3: $a \le 1/2 \le b$. We argue similarly to Case 1. The definition of h stays exactly the same, but this time the constraints on p, q are $p \in [0, a/2]$ and $q \in [b - 1/2, b/2]$. By analogous arguments to Case 1, we get that for the optimal f it holds that $h(a/2, b-1/2) = h(0, b/2)$. That is:

$$\begin{aligned}&1/2 - a/2 + 1 - b + \max\{a/2, b - 1/2\} - (1/2 - a/2) - 2(1/2 - b/2)f(1/2) - b + 1/2 = \\ &\max\{1/2 - a, 1/2 - b/2\} + a - 1/2 + af(1/2).\end{aligned}$$

A straightforward computation yields the value of $f(1/2)$ as in the proposition statement, with the adversary's best reply being $\alpha\delta_{x^a} + (1-\alpha)\delta_{x^b}$, with $x^a := (1/2 - a, a/2, a/2, 0, 1/2 - b/2, 0, 0, b/2)$, $x^b := (1/2 - a/2, 0, 0, a/2, 0, 1/2 - b/2, 1/2 - b/2, b - 1/2)$ and $\alpha = \frac{b}{a+b}$. $\square$

Itai Arieli
Technion – Israel Institute of Technology
Haifa, Israel
Email: iarieli@technion.ac.il

Yakov Babichenko
Technion – Israel Institute of Technology
Haifa, Israel
Email: yakovbab@technion.ac.il

Inbal Talgam-Cohen
Technion – Israel Institute of Technology
Haifa, Israel
Email: italgam@cs.technion.ac.il

Konstantin Zabarnyi
Technion – Israel Institute of Technology
Haifa, Israel
Email: konstzab@gmail.com