False Consensus, Information Theory, and Prediction Markets*

Yuqing Kong
Peking University

Grant Schoenebeck
University of Michigan

Abstract

Our main result shows that when agents' private information about an event are independent conditioning on the event's outcome, then, after an initial announcement, whenever agents have similar beliefs about the outcome, their information is aggregated. That is, there is no false consensus.

Our main result has a short proof based on a natural information theoretic framework. A key ingredient of the framework is the equivalence between the sign of the "interaction information" and a super/sub-additive property of the value of people's information. This provides an intuitive interpretation and an interesting application of the interaction information, which measures the amount of information shared by three random variables.

We illustrate the power of this information theoretic framework by reproving two additional results within it: 1) that agents quickly agree when while announcing beliefs in round robin fashion [Aaronson 2005]; and 2) results from [Chen et al 2010] on when prediction market agents should release information to maximize their payment. We also interpret the information theoretic framework and the above results in prediction markets by proving that the expected reward of revealing information is the conditional mutual information of the information revealed.

1 Introduction

Initially Alice thinks Democrats will win the next presidential election with probability 90% and Bob thinks Democrats will win with 10%. The players then alternate announcing their beliefs of the probability the Democrats will win the next presidential election. Alice goes first and declares, "90%". Bob, then updates his belief rationally based on some commonly held information, some private information, and what he can infer from Alice's declaration (e.g. to 30%) and announces that, "30%". Alice then updates her belief and announces it, and so forth.

Formally, we have the following definition.

Definition 1.1 (Agreement Protocol [Aaronson, 2005]). *Alice and Bob share a common prior over the three random variables W, X_A, and X_B. W denotes the event to be predicted, and $X_A = x_A$ and $X_B = x_B$ are Alice's and Bob's private signals respectively.*

The players alternate announcing their beliefs of W's realization.

In round 1, Alice declares her rational belief $p_A^1 = \Pr[W|X_A = x_A]$. [1] *Then Bob updates and declares his belief $p_B^1 = \Pr[W|X_B = x_B, p_A^1]$ rationally conditioning on his private information and what he can infer from Alice's declaration.*

Similarly, at round i, Alice announces her updated belief $p_A^i = \Pr[W|X_A = x_A, p_A^1, p_B^1, \ldots, p_A^{i-1}, p_B^{i-1}]$; and subsequently, Bob updates and announces his belief $p_B^i = \Pr[W|X_B = x_B, p_A^1, p_B^1, \ldots, p_A^{i-1}, p_B^{i-1}, p_A^i]$.

This continues indefinitely.

*This is a write up of the theorem and proof which appeared in Kong and Schoenebeck's EC 2017 tutorial `https://drive.google.com/open?id=18QifPXMezN42FgYnsYEgen7tjbxeuoEI`

[1] Here and elsewhere we use the notation $\Pr[W]$ to denote a vector whose $w \in W$th coordinate indicates $\Pr[W = w]$.

Two fundamental questions arise from this scenario:

1. Will Alice and Bob ever agree or at least approximately agree, and if so will they (approximately) agree in a reasonable amount of time?

2. If they (approximately) agree, will their agreement (approximately) aggregate their information? That is, will they (approximately) agree on the posterior belief conditioning on Alice and Bob's private information.

Aumann [1976] famously showed that rational Alice and Bob will have the same posterior belief given that they share the same prior and their posteriors are a common knowledge. In particular, if the agents in the agreement protocol ever stop updating their beliefs, they must agree. While this may seem counter-intuitive, a quick explanation is that it is not rational for Alice and Bob to both persistently believe they know more than the the other person. This result does not fully answer the first question because the common knowledge requires a certain amount of time to be achieved.

Aaronson [2005] answers the first question in the affirmative: rational Alice and Bob will take at most $O(\frac{1}{\delta\epsilon^2})$ rounds to have (ϵ,δ)-close beliefs,[2] regardless of how much they disagree with each other initially.

Alas, it is known the second question cannot always be answered in the affirmative, and thus agreement may not fully aggregate information.

Example 1.2 (False consensus)**.** *Say Alice and Bob each privately and independently flip a fair coin, and the outcome is the XOR of their results. Alice and Bob immediately agree, both initially proclaiming the probability* 0.5*. However, this agreement does not aggregate their information; pooling their information, they could determine the outcome.*

Nonetheless, we answer the second question affirmatively for a large class of structures in a generalized context with more than two agents that we call the consensus protocol. Notice that in the above false consensus example, Alice's and Bob's private information are independent but once we condition on the outcome they are dependent. Chen et al. [2010] call this independent structure "complements" for reasons will become clear. They also propose another independent structure, "substitutes" where both Alice and Bob's private information are independent conditioning on the outcome. Chen and Waggoner [2016] further develop these concepts.

We will show that in the "substitutes" setting, i.e., when Alice and Bob's information are conditionally independent, (approximate) agreement implies (approximate) aggregation. We prove the results in the n agents setting which is a natural extension of the Alice and Bob case. Our proof is direct and short based on the information-theoretic tools.

High Level Proof First, we denote the value of an agent's information as the mutual information between their private information and the outcome conditioning on the public information.

The main lemma shows that the "substitute" structure implies the *sub-additivity* of the values of people's private information.

When people approximately agree with each other conditioning on the history, the remaining marginal value of each individual's private information is small $\leq \epsilon$. Under the "substitutes" structure, the sub-additive property of the main lemma implies the total value of information that has not been aggregated is most $n\epsilon$ where n is the number of agents ($n=2$ in the Alice and Bob case). Therefore, (approximate) agreement implies (approximate) aggregation.

[2] $\Pr[|\text{Alice's expectation} - \text{Bob's expectation}| > \epsilon] < \delta$

To show the main lemma, a key ingredient is the equivalence between the sign of the 'interaction information' and the super/sub-additive property of the value of people's information. Interaction information is a generalized mutual information which measures the amount of information shared by three random variables. Unlike mutual information, interaction information can be positive or negative and thus is difficult to interpret and does not yet have a broad applications. Our framework provides an intuitive interpretation and an interesting application of the interaction information.

We illustrate the power of this information theoretic framework by reproving two additional results within it: 1) that agents quickly agree when while announcing beliefs in round robin fashion [Aaronson, 2005]; and 2) results from Chen et al. [2010] that to maximize their payment in a prediction market, when signals are substitutes, agents should reveal them as soon as possible, and when signals are complements, agents should reveal them as late as possible. We also interpret our information theoretic framework, our main result, and our quick convergence reproof in the context of prediction markets by proving that the expected reward of revealing information is the conditional mutual information of the information revealed.

The reproof that agents quickly agree uses the aggregated information as a potential function and observes that each round in which their is ϵ disagreement, the aggregated information must increase by ϵ. The result of when agents should reveal information in a prediction market follows from the sub/super-additivity of mutual information in each of these cases, which can be established using the sign of the interaction information.

Independent Work Frongillo et al. [2021] independently prove that agreement implies aggregation of agents' information under similar special information structures. However, the analyses are very different. Frongillo et al. [2021] employ a very delicate analysis that allows the results to be extended to general divergence measures. Our information-theoretic framework's analysis provides a direct, short, and intuitive proof.

2 Preliminaries

2.1 Complements and Substitutes

Following Chen et al. [2010] we will be interested in two main types of signals.

Definition 2.1. *(Substitutes and Complements [Chen et al., 2010]) W denotes the event to be predicted.*

Substitutes *Agents' private information $X_1, X_2, \cdots, X_n$ are independent conditioning on W.*

Complements *Agents' private information $X_1, X_2, \cdots, X_n$ are independent.*

2.2 Information Theory Background

This section introduces multiple concepts in information theory that we will use to analyze the consensus protocol and, later, prediction markets.

Definition 2.2 (Entropy Shannon [1948]). *We define the entropy of a random variable X as*

$$H(X) := -\sum_x \Pr[X = x] \log(\Pr[X = x]).$$

Moreover, we define the conditional entropy of X conditioning on an additional random variable $Z = z$ as

$$H(X|Z = z) := -\sum_x \Pr[X = x|Z = z] \log(\Pr[X = x|Z = z])$$

We also define the conditional entropy of X conditioning on Z as

$$H(X|Z) := \mathbb{E}_Z[H(X|Z = z)].$$

The entropy measures the amount of uncertainty in a random variable. A useful fact is that when we condition on an additional random variable, the entropy can only decrease. The follows immediately from the concavity of log.

Definition 2.3 (Mutual information Shannon [1948]). *We define the mutual information between two random variables X and Y as*

$$I(X;Y) := \sum_{X,Y} \Pr[X = x, Y = y] \log \left(\frac{\Pr[X = x, Y = y]}{\Pr[X = x] Pr[Y = y]} \right)$$

Moreover, we define the conditional mutual information between random variables X and Y conditioning on an additional random variable $Z = z$ as

$$I(X;Y|Z = z) := \sum_{x,y} \Pr[X = x, Y = y|Z = z] \log \left(\frac{\Pr[X = x, Y = y|Z = z]}{\Pr[X = x|Z = z] \Pr[Y = y|Z = z]} \right)$$

We also define the conditional mutual information between X and Y conditioning on Z as

$$I(X;Y|Z) := \mathbb{E}_Z I(X;Y|Z = z).$$

Fact 2.4 (Facts about mutual information). *Cover and Thomas [2006]*

Symmetry: $I(X;Y) = I(Y;X)$

Relation to entropy: $H(X) - H(X|Y) = I(X;Y)$, $H(X|Z) - H(X|Y,Z) = I(X;Y|Z)$

Non-negativity: $I(X;Y) \geq 0$

Chain rule: $I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$

Monotonicity: *when X and Z are independent conditioning on Y, $I(X,Z) \leq I(X;Y)$.*

The first two facts follows immediately from the formula. The third follows from the second, and the fact that conditioning can only decreases entropy. The first three allow one to understand mutual information as the amount of uncertainly of one random variable that is eliminated once knowing the other (or vice versa). The chain rule follows from the second because $I(X,Y;Z) = H(Z) - H(Z|X,Y) = H(Z) - H(Z|Y) + H(Z|Y) - H(Z|X,Y) = I(X;Z) + I(Y;Z|X)$. The last property follows from the fact that $I(X;Y,Z) = I(X;Y)$ which can be proved by algebraic calculus and the chain rule which says $I(X;Y,Z) = I(X;Z) + I(X;Y|Z)$.

The interaction information, sometimes called co-information, is a generalization of the mutual information for three random variables. Previous work Bell [2003] connects this concept to a memory related concept in brain theory to illustrate the degrees of "hanging togetherness".

Definition 2.5 (Interaction information [Ting, 1962]). *For three random variables X, Y and Z, we define the interaction information among them as*

$$I(X;Y;Z) := I(X;Y) - I(X;Y|Z)$$

Fact 2.6 (Symmetry [Ting, 1962]). $I(X,Y,Z) = I(Y,X,Z) = I(X,Z,Y)$

This can be verified through algebraic manipulations.

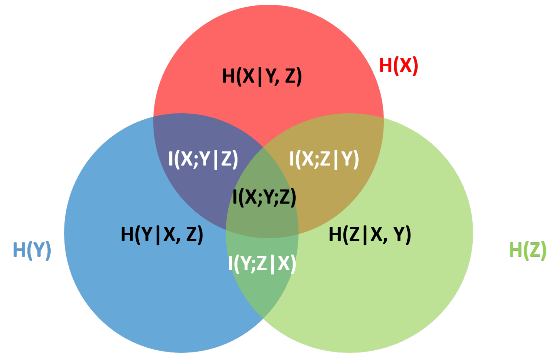

Figure 1: Venn diagram

Venn diagram As shown in figure 1, random variables X, Y, Z can be visualized as sets $H(X), H(Y), H(Z)$ where the set's area represents the uncertainly of the random variable, and:

Mutual Information: operation ";" corresponds to intersection " $\cap$" and is symmetric;

Joint Distribution: operation "," corresponds to union " $\cup$" and is symmetric;

Conditioning: operation " |" corresponds to difference " |";

Disjoint Union: operation " +" corresponds to the disjoint untion " $\sqcup$";

For example, $H(X, Y) = H(X) + H(Y|X)$ because the LHS is $H(X) \cup H(Y)$. Note that the interaction information corresponds to the center of the Venn diagram in figure 1. However, despite the intuition that area is positive, the interaction information is not always positive.

3 Information-theoretic Consensus

We will analyze the consensus protocol, a multi-agent generalization of the agreement protocol, and use the information-theoretic framework to show that 1) consensus is quickly achieved; 2) the sub-additivity of "substitutes" guarantees that approximate consensus implies approximate aggregation.

Definition 3.1 (Consensus Protocol)**.** *Agents $A = \{1 \ldots, n\}$ share a common prior over the $n+1$ random variables W, X_1, X_2, and X_n. W denotes the event to be predicted, and for $i \in A$, $X_i = x_i$ denotes agent i's private signal and its realization.*

The players take turns announcing their beliefs of W's outcome.

In round 1, agent 1 declares her rational belief $p_1^1 = \Pr[W|X_1 = x_1]$. [3] *Then, in sequence, for $i \in \{2, \ldots, n\}$, agent i updates and declares her belief $p_i^1 = \Pr[W|X_i = x_i, p_1^1, \ldots, p_{i-1}^1]$ rationally conditioning on her private information and what she can infer from the previous agents' declarations.*

Similarly, at round t, the agents cycle through the rotation. Agent i announces $p_i^t = \Pr[W|X_i = x_i, H_i^t]$ where H_i^t denotes the historical declarations. Thus $H_1^t = p_1^1, p_2^1, \ldots, p_n^{t-1}$ and for $i = 2$ to n $H_i^t = p_1^1, p_2^1, \ldots, p_{i-1}^t$.

This continues indefinitely. More generally, let $H^t = p_1^1, p_2^1, \ldots, p_n^1, \ldots, p_1^t, p_2^t, \ldots, p_n^t$ denote the history of the first t rounds. Also, it will be convenient to interpret H_{n+1}^t as H_1^{t+1}.

[3] Recall we use the notation $\Pr[W]$ to denote a vector whose $w \in W$th coordinate indicates $\Pr[W = w]$.

Note that $H^t = H_1^{t+1} = H_{n+1}^t$.

We define the amount of information aggregated at any time as the mutual information between W and the historical declarations. The following lemma shows three intuitive properties: 1) the amount of information aggregated is non-decreasing and 2) the growth rate depends on the marginal value of the agent's private information; 3) agreement implies no marginal value.

Lemma 3.2 (Information-Theoretic Properties of Consensus Protocol). *In the consensus protocol,*

Non-decreasing Historical Information *any agent's declaration does not decrease the amount of information so*

$$I(H_{i+1}^t; W) \geq I(H_i^t; W), \text{ for all } i \in A, t \in \mathbb{N}.$$

Therefore, the information increases after each round:

$$I(H^{t+1}; W) \geq I(H^t; W), \text{ for all } t \in \mathbb{N}.$$

Growth Rate = Marginal Value *the change in the historical information is the conditional mutual information between the acting agent's private information and the predicted event conditioning on the history, i.e.,*

$$I(H_{i+1}^t; W) - I(H_i^t; W) = I(X_i; W|H_i^t), \text{ for all } i \in A, t \in \mathbb{N}.$$

Moreover, one round's declarations will increase the amount of historical information at least any single agent's mutual information with the predicted event conditioning on the history of the previous rounds,

$$I(H^{t+1}; W) - I(H^t; W) \geq I(X_i; W|H^t), \text{ for all } i \in A.$$

Agreement ⇒ No Marginal Value *If for some realizable history h_t after t rounds, an agent i perfectly agrees with the current declaration after t rounds, that is,*

$$\Pr[W|X_i = x_i, H^t] = \Pr[W|H^t], \text{ for all possible } x_i,$$

then her immediate marginal value, $I(X_i; W|H^t)$, is zero.

We defer the proof to Section A. The last property inspires our definition of approximate consensus.

Definition 3.3 (ϵ-MI consensus). *For every t, round t achieves ϵ-MI consensus if for all i,*

$$I(X_i; W|H^t) \leq \epsilon.$$

With the above definition, the analysis of consensus time and false consensus becomes intuitive.

3.1 Quick Consensus

Lemma 3.2 shows that the amount of aggregated information increases and the growth rate is the marginal value of the agent's private information. Moreover, disagreement implies $\geq \epsilon$ marginal value which implies $\geq \epsilon$ growth rate. This almost immediately leads to the quick consensus result.

Theorem 3.4 (Convergence rate). *The consensus protocol achieves ϵ-MI-agreement in at most $\frac{I(X_1,X_2,\ldots;W)}{\epsilon} \leq \frac{H(W)}{\epsilon}$ rounds.*

Proof of Theorem 3.4. Because mutual information is monotone, the amount of information aggregated $I(H^t; W)$ is an increasing function with respect to t. Moreover, we will show that when a round does not achieve ϵ-MI consensus, the expected amount of aggregated information increases by at least ϵ.

Formally, when a round t does not achieve ϵ-MI consensus, there exists an agent i such that $I(X_i; W|H^t) > \epsilon$. Thus, the expected amount of aggregated information increases by at least ϵ in round t because Lemma 3.2 shows that $I(H^{t+1}; W) - I(H^t; W) \geq I(X_i; W|H^t) > \epsilon$.

Finally, because for all t, $I(H^t; W) \leq I(X_1, X_2, ...; W) \leq H(W)$, with at most $\frac{I(X_1,X_2,...;W)}{\epsilon} \leq \frac{H(W)}{\epsilon}$ rounds, the consensus protocol achieves ϵ-MI consensus. □

Though we use a different definition of approximate agreement from Aaronson [2005], our quick consensus results match those of Aaronson [2005] in cases where both are well defined. This can be proved by analyzing the relationship between two commonly used divergence measures, the total variation distance D_{TV} and KL divergence D_{KL}. We defer a detailed comparison to Section **??**.

3.2 No False Consensus with Substitutes

We present our main result in this section: when agent's information are substitutes, there is no false consensus. A key ingredient is a subadditivity property for substitutes.

Theorem 3.5 (Convergence ⇒ Aggregation)**.** *For all priors where agents' private information are substitutes, if agents follow the consensus protocol and achieve ϵ-MI consensus after t rounds, then the amount of information that has not been aggregated is bounded by $n\epsilon$.*

We will use the following Lemma in the proof:

Lemma 3.6 (Subadditivity for substitutes)**.** *When X and Y are independent conditioning on Z:*

Nonnegative Interaction Information $I(X; Y; Z) \geq 0$*;*

Conditioning Reduces Mutual Information $I(Y; Z|X) \leq I(Y; Z)$*;*

Subadditivity of Mutual Information $I(X, Y; Z) \leq I(X; Z) + I(Y; Z)$.

Moreover, when $X_1, \ldots, X_n$ are independent conditioning on W,

$$I(X_1, \ldots, X_n; W) \leq \sum_{i=1}^{n} I(X_i; W).$$

Proof of Lemma 3.6. Note that $I(X; Y|Z) = 0$ because X and Y are independent after conditioning on Z.

Nonnegative Interaction Information follows because $I(X; Y; Z) = I(X; Y) - I(X; Y|Z) = I(X; Y) \geq 0$.

Next $I(Y; Z|X) \geq I(Y; Z)$ follows because $0 \leq I(X; Y; Z) = I(Y; Z; X) = I(Y; Z) - I(Y; Z|X)$ where the inequality is by nonnegative interaction information, the first equality is from the symmetry of interaction information, and the second equality is from the definition of interactive information.

Next, subadditivity immediately follows because: $I(X, Y; Z) = I(X; Z) + I(Y; Z|X) \geq I(X; Z) + I(Y; Z)$ where the equality is from the chain rule and the inequality is because conditioning reduces mutual information.

The moreover follows by using induction and subadditivity.

□

We require an additional observation that no history will disrupt the special information structure.

Observation 3.7. *For any fixed history in the consensus protocol, when $X_1, X_2, \cdots, X_n$ are substitutes (complements), they are still substitutes (complements) conditioning on any history.*

We defer the proof to Section A. Aided by the above information-theoretic properties, we are ready to show our direct and short proof.

Proof for Theorem 3.5. If after t rounds the consensus protocol achieves ϵ-MI consensus, then for all i,

$$I(X_i; W|H^t) \leq \epsilon.$$

After t rounds, the amount of information we have aggregated is $I(H^t; W)$. We can aggregate at most $I(X_1, X_2, ..., X_n; W)$ information when all agents' private information are revealed explicitly. Thus, the amount of information that has not been aggregated is

$$\begin{aligned}
&I(X_1, X_2, ..., X_n; W) - I(H^t; W) \\
=&I(X_1, X_2, ..., X_n, H^t; W) - I(H^t; W) && \text{(The history } H^t \text{ is a function of } X_1, \dots, X_n\text{)} \\
=&I(X_1, X_2, ..., X_n; W|H^t) && \text{(Chain rule)} \\
\leq&\sum_i I(X_i; W|H^t) \leq n\epsilon && \text{(Sub-additivity)}
\end{aligned}$$

The third inequality is from subadditivity: note that because $X_1, \dots, X_n$ are independent after conditioning on W that is still true after conditioning on H from Observation 3.7. □

4 Illuminating Prediction Markets with Information Theory

4.1 Prediction Market Overview

A prediction market is a place to trade information. For example, a market maker can open a prediction market for the next presidential election with two kinds of shares, D-shares and R-shares. If Democrats wins, each D-share pays out one dollar but R-shares are worth nothing. If Republicans win, each R-share is worth one dollar and D-shares are worth nothing. People reveal their information by trading those shares. For example, if you think democrats win with probability 0.7, you should buy D-shares as long as its price is strictly lower than \$ 0.7. If people on balance believe that the current price is less than the probability with which the event will occur, the demand for shares (at that price) will outstrip the supply, driving up the price. Similarly if the price is too high they will buy the opposite share as the prices should sum to 1 since exactly one of the two will payout 1.

Hanson [2003, 2012] propose a model of predictions markets that is theoretic equivalent to the above which we describe below. Instead of buying/selling shares to change the price, the agents simply change the market price directly. We define this formally below.

4.2 Preliminaries for Prediction Markets

We introduce prediction markets formally and relate them to the agreement and consensus protocols. We focus on prediction markets which measures the accuracy of a prediction using the logarithmic scoring rule.

Definition 4.1. *(Logarithmic scoring rule) Fix an outcome space Σ for a signal σ. Let $\mathbf{q} \in \Delta_\Sigma$ be a reported distribution.*

The Logarithmic Scoring Rule maps a signal and reported distribution to a payoff as follows:

$$L(\sigma, \mathbf{q}) = \log(\mathbf{q}(\sigma)).$$

Definition 4.2 (Market scoring rule Hanson [2003, 2012], Chen and Pennock [2012])**.** *We build a market for random variable W as follows: the market sets an initial belief p_0 for W. An sequence of agents is fixed. Activity precedes in rounds. In the ith round, the corresponding agent can change the market price from p_i to p_{i+1}, and will be compensated $L(W, p_{i+1}) - L(W, p_i)$ after W is revealed.*

Let the signal σ be drawn from some random process with distribution $\mathbf{p} \in \Delta_\Sigma$.

Then the expected payoff of the Logarithmic Scoring Rule is:

$$\mathbb{E}_{\sigma \leftarrow \mathbf{p}}[L(\sigma, \mathbf{q})] = \sum_\sigma \mathbf{p}(\sigma) \log \mathbf{q}(\sigma) = L(\mathbf{p}, \mathbf{q}) \tag{1}$$

It is well known (and easily verified) that this value will be uniquely maximized if and only if $\mathbf{q} = \mathbf{p}$. Because of this, the logarithmic scoring rule is called a *strictly proper scoring rule.*

Agreement Protocol & Myopic Agents An agent behaves myopically in the prediction market if she always changes the current price to her Bayesian posterior belief conditioning on her private information and what she can infer from the market's history. When Alice and Bob participate the prediction market alternatively and myopically, the list of market price correspond to their declarations in the agreement protocol. This also naturally extends to n agent consensus protocol setting. Of course, in general, Alice and Bob need not behave myopically and in these cases the market price may not correspond exactly to the agreement protocol declarations.

4.3 Expected Payment = Conditional Mutual Information

Amazingly, we show that the expected payment to an agent, is just a function of the conditional mutual information of the random variables that he reveals. This connects the expected payment in prediction markets and amount of information.

Lemma 4.3 (Expected payment = conditional mutual information)**.** *In a prediction market with log scoring rule, the agent who changes the belief from $\Pr[W|H]$ to $\Pr[W|X = x, H]$ for all x will be paid $I(X; W|H)$ in expectation.*

Proof of Lemma 4.3. Fix any history H:

$$\begin{aligned}
&\mathbb{E}_{X,W|H} L(W, \Pr[W|H, X]) - L(W, \Pr[W|H]) \\
&= \sum_{W,X} \Pr[W, X|H] \log(\frac{\Pr[W|H, X]}{\Pr[W|H]}) \\
&= I(X; W|H)
\end{aligned}$$

$\square$

Interpretation of Previous Results As mentioned, when agents participate the market one by one and behave myopically, the list of market prices correspond to their declarations in the agreement/consensus protocol. The connection between the expected payment and conditional mutual information leads to the following interpretations.

ϵ-MI Consensus If round t achieves ϵ-MI consensus, the expected payment of any particular agent obtained by changing the market price to her Bayesian posterior is at most ϵ.

Quick consensus For any round t that does not achieve ϵ-MI consensus, in round $t+1$ at least one agent will obtain at least an ϵ payment in expectation. However, Lemma 4.3 also implies that the total expected payment in the prediction market is bounded by the entropy of W. Thus, at most $\frac{H(W)}{\epsilon}$ rounds are needed for ϵ-MI consensus.

No False Consensus with Substitutes If round t achieves ϵ-MI consensus, no agent has information with value greater than ϵ. Then subadditivity implies that currently, the expected payment of all agents' private information together is bounded by $n\epsilon$, thus is not valuable.

A Relaxation of Myopia Lemma 4.3 holds in a broader setting than just when agents behave myopically. It also holds in the scenario when agents may hide information but do not try to maliciously trick other agents by misreporting their information early on to try to trick agents later. Formally, agent i who has private information X_i can hide information by changing the market price to $\Pr[W=w|S(X_i),H]$ rather than $\Pr[W=w|X_i,H]$ where H is the history and S is a possibly random function. That is, she declares the rational belief conditioning on the realization of $S(X_i)$ and what she can infer from the history. In this case Lemma 4.3 shows her payment will be $I(S(X_i),W|H)$, which we will use in the next section to analyze the strategic behavior of agents.

A natural question is whether we should expect that fully strategic agents to behave maliciously rather than just hiding information. Previous work Anunrojwong et al. [2019] has shown misreporting or tricking other agents is unlikely to happen in equilibrium.

4.4 Strategic Revelation

This section will the above information-theoretic properties to provide an alternative proof for the results proved in Chen et al. [2010].

Definition 4.4 (Alice-Bob-Alice (ABA))**.** *Alice and Bob holds private information X_A, X_B related to event W respectively. There are three stages. Alice can change the market belief at stage 1 and stage 3. Bob can change the market belief at stage 2.*

We assume that the strategic players can hide their information but not behave maliciously. A strategy profile is a profile of the players' strategies.

Proposition 4.5 (ABA & Information structures Chen et al. [2010])**.** *In prediction market based on log scoring rule,*

Substitutes *when Alice and Bob's private information are substitutes, the strategy profile that Alice reveals X_A at stage 1 and Bob reveals X_B at stage 2 is an equilibrium;*

Complements *when Alice and Bob's private information are complements, the strategy profile that Bob reveals X_B at stage 2 and Alice reveals X_A at stage 3 is an equilibrium.*

We will use the following lemma, which is a direct analogue of Lemma 3.6

Lemma 4.6 (Superadditivity for complements)**.** *When X and Y are independent:*

Nonpositive Interaction Information $I(X;Y;Z) \leq 0$*;*

Conditioning Increases Mutual Information $I(Y;Z|X) \geq I(Y;Z)$*;*

Superadditivity of Mutual Information $I(X,Y;Z) \geq I(X;Z) + I(Y;Z)$*.*

Moreover, when $X_1, \ldots, X_n$ are independent conditioning on W,

$$I(X_1, \ldots, X_n; W) \geq \sum_{i=1}^{n} I(X_i; W).$$

The proof is directly analogous to that of Lemma 3.6 and we defer it to Section A.

Proof of Proposition 4.5. First, for Bob, to maximize his expected payment, it's always optimal to reveal X_B in stage 2. It's left to analyze Alice's optimal strategy given that Bob reveals X_B in stage 2.

If Alice reveals all her information either at the first stage or at the third stage, we only need to compare $I(X_A;W)$ and $I(X_A;W|X_B)$. But $I(X_A;W) - I(X_A;W|X_B) = I(X_A;X_B;W)$. Thus, the results immediately follow from the fact that the sign of the interaction information is nonnegative/nonpositive when the information are substitutes/complements.

It's left to consider the general strategy where Alice reveals part of her information at stage 1, say $S_1(X_A)$, and part of her information at stage 3, say $S_3(X_A)$. In this case, she will be paid $I(S_1(X_A);W) + I(S_3(X_A);W|X_B,S_1(X_A))$ according to Lemma 4.3.

First, it is optimal for Alice to reveal all her remaining information at the last stage since $I(S_3(X_A);W|X_B,S_1(X_A)) \leq I(X_A;W|X_B,S_1(X_A))$ due to monotonicity of mutual information. Thus, Alice will reveal $S_3(X_A) = X_A$.

Under the substitutes structure, Alice's expected payment in stage 3 is

$$I(X_A;W|X_B,S_1(X_A)) \leq I(X_A;W|S_1(X_A))$$

because 1) fixing any $S_1(X_A)$, X_B and X_A are still independent conditioning on W thus are substitutes (Observation 3.7) and 2) conditioning decreases mutual information for substitutes. Therefore,

$$\begin{aligned}
&I(S_1(X_A);W) + I(X_A;W|X_B,S_1(X_A)) \\
\leq &I(S_1(X_A);W) + I(X_A;W|S_1(X_A)) && \text{(conditioning decreases mutual information)} \\
= &I(X_A,S_1(X_A);W) = I(X_A;W). && \text{(chain rule)}
\end{aligned}$$

Thus $I(X_A;W)$ upper bounds Alice's total payment, but this is what she receives if she reveals all her information in the first round.

Under complements structure, Alice's expected payment in stage 1 is

$$I(S_1(X_A);W) \leq I(S_1(X_A);W|X_B)$$

because 1) $S_1(X_A)$ and X_B are independent thus are complements and 2) conditioning decreases mutual information for complements. Therefore,

$$
\begin{aligned}
& I(S_1(X_A);W) + I(X_A;W|X_B,S_1(X_A)) \\
\leq & I(S_1(X_A);W|X_B) + I(X_A;W|X_B,S_1(X_A)) && \text{(conditioning increases mutual information)} \\
= & I(X_A,S_1(X_A);W|X_B) = I(X_A;W|X_B) && \text{(chain rule)}
\end{aligned}
$$

Thus $I(X_A;W|X_B)$ upper bounds Alice's total payment, but this is what she receives if she reveals all her information in the second round.

□

5 Conclusion and Discussion

We have developed an information theoretic framework to analyze the aggregation of information both in the agreement/consensus protocol and prediction markets. We then showed that when agents' private information about an event are independent conditioning on an event's outcome, then, when the agents are in near agreement in the consensus protocol, their information is nearly aggregated. We additionally reproved 1) the consensus protocol quickly converges [Aaronson 2005]; and 2) results from Chen et al. [2010] on when prediction market agents should release information to maximize their payment.

Our analysis of the Alice Bob Alice prediction market straightforwardly extends to any sequence of an arbitrary number of agents. By applying the same argument, it is easily shown that: 1) every agent revealing all their information immediately is an equilibrium in the substitutes case; and 2) every agent revealing all their information as late as possible is an equilibrium in the complements case.

One possible extension is to look at generalizations of Shannon mutual information. For example, starting with any strictly proper scoring rule one can develop a Bregman mutual information [Kong and Schoenebeck, 2019] and ask whether our proofs will go through using this new mutual information definition. When using the logarithmic scoring rule, one arrives at the standard Shannon mutual information used in this paper. However, different scoring rules are possible as well. All such mutual information will still obey the chain rule and non-negativity. As such, generalized versions of Lemma 3.2 will hold. However they may not be symmetric (and similarly their interactive information may not be symmetric). Thus, our techniques cannot be straightforwardly adjusted to reprove Theorem 3.5 or Proposition 4.5 in this manner. So while this generalizes our framework, finding a good application remains future work.

We hope that our framework can analyze additional settings of agents aggregating information. One example would be more inclusive classes of agent signals than in the settings of this paper. However, perhaps our framework could also be applied to analyze broader settings such as social learning [Golub and Sadler, 2016] or rewards for improvements in machine learning outcomes [Abernethy and Frongillo, 2011], where either individually developed machine learning predictors are eventually combined in ensembles or the training data is augmented by individually procured training data.

References

Scott Aaronson. The complexity of agreement. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 634–643. ACM, 2005.

Jacob D Abernethy and Rafael Frongillo. A collaborative mechanism for crowdsourcing prediction problems. *Advances in Neural Information Processing Systems*, 24, 2011.

Jerry Anunrojwong, Yiling Chen, Bo Waggoner, and Haifeng Xu. Computing equilibria of prediction markets via persuasion. In *International Conference on Web and Internet Economics*, pages 45–56. Springer, 2019.

Robert J Aumann. Agreeing to disagree. *The annals of statistics*, pages 1236–1239, 1976.

Anthony J Bell. The co-information lattice. In *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA*, volume 2003. Citeseer, 2003.

Yiling Chen and David M Pennock. A utility framework for bounded-loss market makers. *arXiv preprint arXiv:1206.5252*, 2012.

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

Yiling Chen, Stanko Dimitrov, Rahul Sami, Daniel M Reeves, David M Pennock, Robin D Hanson, Lance Fortnow, and Rica Gonen. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 58(4):930–969, 2010.

Thomas M Cover and Joy A Thomas. Elements of information theory 2nd edition. 2006.

Rafael Frongillo, Eric Neyman, and Bo Waggoner. Agreement implies accuracy for substitutable signals. *arXiv preprint arXiv:2111.03278*, 2021.

Benjamin Golub and Evan Sadler. Learning in social networks. *The Oxford Handbook of the Economics of Networks*, 2016.

Robin Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1): 107–119, 2003.

Robin Hanson. Logarithmic markets coring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets*, 1(1):3–15, 2012.

Yuqing Kong and Grant Schoenebeck. An information theoretic framework for designing information elicitation mechanisms that reward truth-telling. *ACM Trans. Econ. Comput.*, 7(1):2:1–2:33, January 2019. ISSN 2167-8375. doi: 10.1145/3296670. URL http://doi.acm.org/10.1145/3296670.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

Hu Kuo Ting. On the amount of information. *Theory of Probability & Its Applications*, 7(4): 439–447, 1962.

A Additional Proofs

Proof of Lemma 3.2. Because (conditional) mutual information is always non-negative, once we have established the second property of growth rate equal marginal value the first property of non-decreasing information follows immediately. Thus, we start by proving the second property, growth rate equal marginal value.

$$I(H_{i+1}^t; W) - I(H_i^t; W) = I(H_{i+1}^t; W|H_i^t) = I(X_i; W|H_i^t). \tag{2}$$

The first equality holds due to chain rule and the fact that H_{i+1}^t contains H_i^t.

Agent i will rationally update her belief to p_i^t where $\forall w, p_i^t(w) = \Pr[W = w|X_i, H_i^t]$ and declares it. This declaration will be added to $H_{i+1}^t = (H_i^t, p_i^t)$. Moreover, in the agreement protocol, given the history, the Bayesian posterior for predicting W is the current declaration, $\forall w, \Pr[W = w|H_{i+1}^t] = p_i^t(w) = \Pr[W = w|X_i, H_i^t]$. Therefore,

$$\begin{aligned}
I(H_{i+1}^t; W|H_i^t) =& H(W|H_i^t) - H(W|H_{i+1}^t, H_i^t) \\
=& H(W|H_i^t) - H(W|X_i, H_i^t) \quad (\forall w, \Pr[W = w|H_{i+1}^t] = p_i^t(w) = \Pr[W = w|X_i, H_i^t].) \\
=& I(X_i; W|H_i^t)
\end{aligned}$$

Thus, the second inequality of (2) holds.

We now move to the "moreover" part of growth rate equal marginal value.

$$\begin{aligned}
I(H^{t+1}; W) - I(H^t; W) &\geq I(H_{i+1}^{t+1}; W) - I(H^t; W) && (H^{t+1} \text{ contains } H_{i+1}^{t+1}.) \\
&= I(X_i, H_{i+1}^{t+1}; W) - I(H^t; W) \\
& && (\forall w, \Pr[W = w|H_{i+1}^{t+1}] = \Pr[W = w|X_i, H_{i+1}^{t+1}].) \\
&= I(X_i, H_{i+1}^{t+1}; W|H^t) && \text{(Chain rule)} \\
&\geq I(X_i; W|H^t) && \text{(Monotonicity)}
\end{aligned}$$

Finally, we show that agreement implies no marginal value. When an agent i perfectly agrees with the current declaration at round t, that is,

$$\forall w, \Pr[W = w|X_i, H^t] = p_{i-1}^t(w) = \Pr[W = w|H^t]$$

Therefore, $H(W|H^t) = H(W|X_i, H^t)$ which implies that $I(X_i; W|H^t) = 0$. □

Proof of Lemma 4.6. The proof is directly analogous to that of Lemma 3.6

First, $I(X;Y) = 0$ because X and Y are independent.

Nonpositive Interaction Information follows because $I(X;Y;Z) = I(X;Y) - I(X;Y|Z) = -I(X;Y|Z) \leq 0$.

Next $I(Y;Z|X) \geq I(Y;Z)$ because $0 \geq I(X;Y;Z) = I(Y;Z;X) = I(Y;Z) - I(Y;Z|X)$ where the inequality is by nonpostive interaction information, the first equality is from the symmetry of interaction information, and the second equality is from the definition of interactive information.

Third, superadditivity immediately follows because: $I(X,Y;Z) = I(X;Z) + I(Y;Z|X) \geq I(X;Z) + I(Y;Z)$ where the equality is from the chain rule and the inequality is because conditioning increases mutual information.

The moreover follows by using induction and superadditivity.

□

Proof of Observation 3.7. Let D be some distribution over Σ^n where $X_1, X_2, \cdots, X_n$ are all independent. We first observe that after any "independent" restrictions on the realizations of $X_1, X_2, \cdots, X_n$, they are still independent. That is, fix $\Sigma_1, \ldots, \Sigma_n$ where $\Sigma_i \subseteq \Sigma$ for all $i \in \{1, \ldots, n\}$ and let $\xi \subset \Sigma^n$ be the event where $x_i \in \Sigma_i$ for all $i \in \{1, \ldots, n\}$. Then conditioning on ξ, $X_1, X_2, \cdots, X_n$ are independent as well.

For all $(x_1, x_2, \cdots, x_n) \in \Sigma_1 \times \Sigma_2 \times \cdots \Sigma_n$,

$$
\begin{aligned}
&\Pr_D[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n | \xi] \\
&= \frac{\Pr_D[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n]}{\Pr_D[\xi]} \\
&= \frac{\Pi_i \Pr_D[X_i = x_i]}{\Pi_i \Pr_D[X_i \in \Sigma_i]} \\
&= \Pi_i \Pr_D[X_i = x_i | X_i \in \Sigma_i]
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\Pr_D[X_i = x_i | \xi] \\
&= \frac{\Pr_D[X_i = x_i, \xi]}{\Pr_D[\xi]} \\
&= \frac{\Pr_D[X_i = x_i, \forall j \neq i, X_j \in \Sigma_i]}{\Pr_D[\forall j, X_j \in \Sigma_j]} \\
&= \frac{\Pr_D[X_i = x_i] \Pi_{j \neq i} \Pr_D[X_j \in \Sigma_j]}{\Pr_D[X_i \in \Sigma_i] \Pi_{j \neq i} \Pr_D[X_j \in \Sigma_j]} \\
&= \frac{\Pr_D[X_i = x_i]}{\Pr_D[X_i \in \Sigma_i]} \Pi_{j \neq i} \frac{\Pr_D[X_j \in \Sigma_j]}{\Pr_D[X_j \in \Sigma_j]} \\
&= \frac{\Pr_D[X_i = x_i]}{\Pr_D[X_i \in \Sigma_i]}
\end{aligned}
$$

Therefore, we have $\Pr[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n | \xi] = \Pi_i \Pr[X_i = x_i | \xi]$. For all $(x_1, x_2, \cdots, x_n) \notin \Sigma_1 \times \Sigma_2 \times \cdots \Sigma_n$, we have $\Pr[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n | \xi] = 0 = \Pi_i \Pr[X_i = x_i | \xi]$. Thus conditioning on ξ, $X_1, X_2, \cdots, X_n$ are independent as well.

In the consensus protocol, after the first agent makes her declaration p_1^1 which is a function of the prior and X_1, we have a restriction for X_1. The second agent's declaration p_2^1 is a function of the prior, p_1^1 and X_2. For fixed p_1^1, we have an independent restrictions for X_2, and so forth. Thus, for any fixed history, we will have independent restrictions for X_1, X_2, ..., and X_n. Therefore, when X_1, X_2, ..., and X_n are independent, they are still independent given any fixed history as well.

The same analysis shows that when X_1, X_2, ..., and X_n are independent conditioning any $W = w$, they are still independent conditioning on both $W = w$ and any fixed history.

$\square$