2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Informational Substitutes

Yiling Chen
Harvard SEAS
Cambridge, MA, USA
Email: yiling@seas.harvard.edu

Bo Waggoner
Harvard SEAS
Cambridge, MA, USA
Email: bwaggoner@fas.harvard.edu

***Abstract*—We propose definitions of substitutes and complements for pieces of information ("signals") in the context of a decision or optimization problem, with game-theoretic and algorithmic applications. In a game-theoretic context, substitutes capture diminishing marginal value of information to a rational decision maker. There, we address the main open problem in a fundamental strategic-information-revelation setting, prediction markets. We show that substitutes characterize "best-possible" equilibria with immediate information aggregation, while complements characterize "worst-possible", delayed aggregation. Game-theoretic applications also include settings such as crowdsourcing contests and question-and-answer forums. In an algorithmic context, where substitutes capture diminishing marginal improvement of information to an optimization problem, substitutes imply efficient approximation algorithms for a very general class of (adaptive) information acquisition problems.**

In tandem with these broad applications, we examine the structure and design of informational substitutes and complements. They have equivalent, intuitive definitions from disparate perspectives: submodularity, geometry, and information theory. We also consider the design of scoring rules or optimization problems so as to encourage substitutability or complementarity, with positive and negative results. Taken as a whole, the results give some evidence that, in parallel with substitutable items, informational substitutes play a natural conceptual and formal role in game theory and algorithms.

***Keywords*-value of information, decision problems, substitutes, complements, prediction markets, information acquisition, submodularity**

I. Motivation

An agent living in an uncertain world wishes to make some decision (whether to bring an umbrella on her commute, how to design her company's website, ...). She can improve her expected utility by obtaining pieces of information about the world prior to acting (a weatherman's forecast or a barometer reading, market research or automated A/B testing, ...). This naturally leads her to assign *value* to different pieces of information and combinations thereof. The value of information arises as the expected improvement it imparts to her optimization problem.

We would like to generally understand, predict, or design algorithms to guide such agents in acquiring and using information. Consider the analogous case where the agent has value for *items* or goods, represented by a valuation function over subsets of items. A set of items are substitutes if, intuitively, each's value decreases given more of the others; they are complements if it increases. Here, we have rich game-theoretic and algorithmic theories leveraging the structure of substitutes and complements (S&C). For instance, in many settings, foundational work shows that substitutability captures positive results for existence of market equilibria, while complements capture negative results [1]–[5]. When substitutes are captured by submodular valuation functions [6], algorithmic results show how to efficiently optimize (or approximately optimize) subject to constraints imposed by the environment (*e.g.* [7]). For example, an agent wishing to select from a set of costly items with a budget constraint has a $(1 - \frac{1}{e})$-approximation algorithm if her valuation function is submodular [8].

Can we obtain similar structural and algorithmic results for information? Here, a piece of information is modeled as a *signal* or random variable that is correlated in some way with the state of the world that the agent cares about (whether it will rain, how profitable are different website designs, ...). Intuitively, one might often expect information to satisfy substitutable or complementary structure. For instance, a barometer reading and an observation of whether the sky is cloudy both yield valuable information about whether it will rain to an umbrella-toting commuter; but these are substitutable observations for our commuter in that each is probably worth less once one has observed the other. On the other hand, the dew point and the temperature tend to be complementary observations for our commuter: Rain may be only somewhat correlated with dew point and only somewhat correlated with temperature, but is highly correlated with cases where temperature and dew point are close (*i.e.* the relative humidity is high).

Despite this appealing intuition, there are significant challenges to overcome in defining informational S&C. Pieces of information, unlike items, may have complex probabilistic structure and relationships. But on the other hand, this structure alone cannot capture the value of that information, which (again unlike items) seemingly must arise from the context in which it is used. Next, even given a measure of value, it is unclear how to formalize an intuition such as "diminishing marginal value". Finally, it remains to demonstrate that the definitions are tractable and have game-theoretic and/or algorithmic applications. These challenges seem to have prevented a successful theory of informational S&C thus far.

A. This paper: summary and contributions

This paper has four components.

1. We propose a definition of informational substitutes and complements (S&C). Beginning from the very general notion of *value of information* in the context of any specific decision or optimization problem, we define S&C in terms of diminishing (increasing) marginal value for that problem. This requires a definition of "marginal unit" of information. We consider a hierarchy of three kinds of marginal information: learning another signal, learning some deterministic function of another signal, and learning some randomized function ("garbling") of another signal. These give rise to *lattice structures* on the space of signals in a given context; we formalize S&C by *submodularity* and *supermoduarity* on these lattices. The three lattices are respectively very coarse, moderately coarse, and fine; they correspond to *weak*, *moderate*, and *strong* versions of the definitions.

2. We give game-theoretic applications of these definitions, focusing primarily on information aggregation in markets. When strategic agents have heterogeneous, valuable information, we would like to understand when and how their information is revealed and aggregated in an equilibrium of strategic play. Prediction markets, which are toy models of financial markets, are possibly the simplest setting capturing the essence of this question. However, although the efficient market hypothesis states that information is quickly aggregated in financial markets [9], despite much research on this question in economics (*e.g.* [10], [11]) and computer science (*e.g.* [12]–[14]), very little was previously known about how quickly information is aggregated in markets except in very special cases.

We address the main open question regarding strategic play in prediction markets: When and how is information aggregated? We show that informational substitutes imply that all equilibria are of the "best possible" form where information is aggregated immediately, while complements imply "worst possible" equilibria where aggregation is delayed as long as possible. Furthermore, the respective converses hold as well; *e.g.*, if an information structure guarantees the "best possible" equilibria, then it must satisfy substitutes. Informational S&C thus seem as fundamental to equilibria of (informational) markets as substitutable items are in markets for goods.

We believe that informational S&C have the potential for broad applicability in other game-theoretic settings involving strategic information revelation, and toward this end, give some additional example applications. We show that S&C characterize analogous "rush/delay" equilibria in some models of machine-learning or crowdsourcing contests [15], [16] and question-and-answer forums [17]. These results resolve open questions raised by previous work.

3. We give algorithmic applications, focusing on the complexity of approximately-optimal information acquisition. Namely, we define a very broad class of problems, termed SignalSelection, in which a decision maker wishes to acquire information prior to making a decision, but has constraints on the acquisition process. For instance, a company wishes to purchase heterogeneous, pricey data sets subject to a budget constraint, or to place up to k sensors in an environment. We show that substitutes imply efficient approximation algorithms in many cases, including a budget constraint; this extends to an adaptive version of the problem as well. We also show that the problem is hard in general and in the complements case, even when signals are independent uniform bits. These results offer a unifying perspective on a variety of similar "submodularity-based" solutions in the literature [18]–[21].

4. We consider the structure and design of informational S&C. The goal is to both identify substitutable structure and *design* for it. We take some initial steps in these directions. For instance, we provide natural geometric and information-theoretic definitions of S&C and show they are equivalent to the submodularity-based definitions.

We address two fundamental questions: Are there (nontrivial) signals that are substitutes for *every* decision problem? Second, given a set of signals, can we always design a decision problem for which they are substitutes? In the game-theoretic settings above, this corresponds to design of mechanisms for immediate aggregation, somewhat of a holy grail for prediction markets. In algorithmic settings, it has relevance for the design of *submodular surrogates* [22]. Unfortunately, we give quite general negative answers to both questions. Surprisingly, more positive results arise for complements. We give the geometric intuition behind these results and point toward heuristics for substitutable design in practice.

In summary, the contributions of this paper are twofold: (a) in the definitions of informational S&C, along with a body of evidence that they are natural, tractable, and useful; and (b) in the applications, in which we resolve a major open problem on strategic information revelation as well as give a unifying and general framework for a broad algorithmic problem. Our results on structure and design of informational S&C points to potential for these very general definitions and results to have concrete applications.

Taken all together, we believe these results give evidence that informational S&C, in analogy with the successful theories of substitutable goods, have a natural and useful role to play in game theory, algorithms, and in connecting the two.

II. Technical summary

Here, we give a more formal summary of the contributions of this paper. For brevity, we will give only the main or representative results, as well as deferring much intuition or discussion to the relevant sections.

1) Setting and definitions of S&C: We begin from a general *decision* or *optimization* problem, given by a utility function $u(d, e)$ when decision d is taken (*e.g.* bring umbrella) and e is the outcome of some random event E (*e.g.* rain). E and $A_1, \ldots, A_n$ are random events that are drawn jointly from a prior distribution P. $A_1, \ldots, A_n$ are referred to as "base signals", and observing them reveals possibly-useful information about E (*e.g.* A_1 is a barometer reading, …). One can also obtain other signals by combining these, *e.g.* in general a signal may correspond to observing multiple base signals or some partial information about a base signal (this will be formalized shortly). For simplicity we assume all A_i and E have a finite number of outcomes. We use p for the prior distribution on E and *e.g.* p_{a_i} for the posterior on E conditioned on $A_i = a_i$.

An agent would like to select the decision maximizing expected utility, over the randomness of E, given the information she has. Define the "value" function $\mathcal{V}^{u,P}$ for this decision problem and prior as follows: for any signal A,

$$\mathcal{V}^{u,P}(A) = \mathop{\mathbb{E}}_{a \sim A}\left[\max_{d} \mathop{\mathbb{E}}_{e \sim E}[u(d, e) \mid A = a]\right].$$

This is the expected utility for first observing A, then selecting the utility-maximizing action conditioned on the observed realization a. When u and P are clear from context, we drop them from the superscript. Note that the signal A in the definition above can be *e.g.* the "null" signal denoted $\perp$, a pair of base signals denoted $A_i \vee A_j$ for reasons that will become clear soon, etc.

Notice that the marginal value of A versus just knowing the prior is $\mathcal{V}^{u,P}(A) - \mathcal{V}^{u,P}(\perp)$, and the marginal value of signal B given that the agent already has access to A is $\mathcal{V}^{u,P}(A \vee B) - \mathcal{V}^{u,P}(A)$.

The marginal unit of information: We would like to define substitutes and complements in terms of the *marginal value* of information. We have defined value above, but this raises the question: What is a "marginal unit" of information? In different contexts, different marginal units are appropriate, so we define three notions ranging from very "large" units, to more fine-grained, to very fine-grained. These will respectively give rise to weak, moderate, and strong versions of substitutability. (Note: all discussion applies equally to complements.) Strong substitutes will imply moderate substitutes, which will imply weak substitutes. This will correspond to the fact that if all of the "fine-grained" units of information display diminishing marginal value, then "larger" units must as well.

1) **Observing an additional signal.** In a game-theoretic setting, an agent may control whether or not a source of information (signal) is revealed. In an algorithmic setting, an algorithm may need to decide either to acquire additional signals or not. For both cases, the natural unit of marginal information is an entire signal (which is quite "large" or "coarse").
2) **Observing a deterministic function of a signal.** Here, notice that the information revealed by a deterministic function is essentially determined by how it *combines* different outcomes of the signal. An injective function would completely reveal the original signal; a function mapping all realizations to the same output reveals nothing. Thus, this case captures an agent possibly "pooling" outcomes of a signal together to reveal only partial information; or similarly, algorithms optimizing over these discrete post-processings of a signal. The unit of information is significantly more fine-grained.
3) **Observing a randomized function ("garbling") of a signal.** Just as the previous case corresponded to deterministic strategies of an agent, this case corresponds to randomized strategies. Algorithmically, it captures optimization over all "signals about signals". Here, the marginal unit of information can be extremely fine; for instance, consider "output the original signal with probability p, else something random" as $p \to 0$ or $p \to 1$.

In the full version of this paper, we show that each of these notions corresponds to a *lattice* of signals generated by $A_1, \ldots, A_n$. Informally, a lattice here is a set of possible signals having a partial ordering $\preceq$ with $A' \preceq A$ meaning that A' conveys less information than A. Lattices have a *join* operation $A \vee B$ meaning "the signal that conveys the outcome of both A and B" and a *meet* operation $A \wedge B$ meaning "the signal that conveys information common to both A and B".

The first case discussed above corresponds to the *"subset" lattice* of subsets of $\{A_1, \ldots, A_n\}$, with join operation as set union and meet as set intersection.

The second case corresponds to the *"discrete" lattice* defined utilizing Aumann's partition model [23], where nature randomly selects a state of the world and each signal is modeled as a partition of these possible states. Here intuitively $A' \preceq A$ if the signal A' is a deterministic function of the signal A; this is formalized via a lattice of partitions.

For the third case, we construct the *"continuous" lattice* where, intuitively, $A' \preceq A$ if A' is a randomized function or "garbling" of A. This is formalized by extending Aumann's model to allow partial information about any signal. (The partial ordering here is very similar to Blackwell's ordering [24].)

The definitions: Now that we have notions for marginal units of information, we will use sub- and super-modularity to capture diminishing and increasing marginal value. While these definitions may be most familiar in terms of set functions $f : 2^{\{1,\dots,n\}} \to \mathbb{R}$, they have precise generalizations to lattice functions as well.

Definition II-.1. A function f from a lattice to the reals is *submodular* if it exhibits diminishing marginal value: For all A', A, B on the lattice with $A' \preceq A$,

$$f(B \vee A') - f(A') \geq f(B \vee A) - f(A).$$

It is *supermodular* if it exhibits increasing marginal value: For all $A' \preceq A$ and all B, the above inequality is reversed. The sub- or super-modularity is *strict* if, whenever A and B are incomparable on the lattice's ordering, the inequality is strict.

Definition II-.2. In the context of a decision problem u and prior P, the signals on a corresponding lattice $\mathcal{L}$ are *substitutes* if $\mathcal{V}^{u,P}$ is submodular on $\mathcal{L}$. The signals are *complements* if $\mathcal{V}^{u,P}$ is supermodular on $\mathcal{L}$. Substitutes or complements are *weak / moderate / strong* if $\mathcal{L}$ is respectively the subset / discrete / continuous lattice. They are *strict* substitutes (complements) if $\mathcal{V}^{u,P}$ is strictly submodular (supermodular) on $\mathcal{L}$.

The most basic example of substitutes are a pair of signals that are perfectly correlated, *i.e.* take on the same outcome with probability one. In this case, for any decision problem, these signals will be substitutes, as neither can possibly give additional information given the other. Hence, we might call them "universal" substitutes, although trivial ones.

The most basic example of complements might be a pair of independent uniform bits, when the event E of interest is the XOR of the bits. In this case, either bit alone gives no information about E, while both together completely determine E.

In many contexts, signals will neither be substitutes nor complements. This matches our experience with valuations for goods: although substitutes are a broad class, many (in fact most) valuation functions are neither submodular nor supermodular. One example is to simply "pair" the two examples above. Let signals A_1 and A_2 each consist of a pair of independent uniform bits; the first bits of each are perfectly correlated and always the same, while the second are completely independent. The event E also consists of a pair of bits, the first equal to the first bits of A_1 and A_2 and the second equal to the XOR of their second bits.

We will later provide some tools for applying these definitions along with geometric intuition and results on classes of S&C.

2) Game-theoretic applications:

Overview of prediction markets: A prediction market is a very basic model of strategic information revelation and aggregation. There are models of prediction markets in which participants buy and sell shares of securities, as in financial markets [25], [26]. These are strategically equivalent to the model we will adopt, based on *proper scoring rules*, and our results automatically apply to these as well.

A *scoring rule* is a utility function $S(q, e)$ where q is a probability distribution over E. It is (strictly) proper if, when $E \sim q$, choosing $\hat{q} = q$ (uniquely) maximizes the expected score $\mathbb{E}_{e \sim q} S(\hat{q}, e)$. One strictly proper example is the log scoring rule $S(q, e) = \log q(e)$ where $q(e)$ is the probability assigned to $E = e$.

In a prediction market, an event E, strictly proper scoring rule S, and initial prediction $p^{(0)}$ are chosen by the market designer. We suppose that E and some base signals $A_1, \dots, A_n$ are drawn jointly by nature from a common prior P. There are m traders, each observing some signal on the continuous lattice, who take turns participating in a fixed order. At each time $t = 1, \dots, T$, participant i_t updates the prediction from $p^{(t-1)}$ to $p^{(t)}$. After T, a realization of an event $E = e$ is observed, and for each update i made at some time t, i is paid $S(p^{(t)}, e) - S(p^{(t-1)}, e)$. Notice that the entire payment of the mechanism telescopes into $S(p^{(T)}, e) - S(p^{(0)}, e)$.

The prediction market is an extensive-form Bayesian game. In a *Bayes-Nash equilibrium* (BNE), each participant's strategy maximizes expected utility given the others' strategies and her own signal realization. A natural refinement is *perfect Bayesian equilibrium* (PBE), loosely a subgame-perfect version of BNE.

If each participant arrives only once, then properness of the scoring rule would imply that at each opportunity, participants are *truthful*: they reveal their posterior beliefs conditioned on the signal and reports in the market thus far. Then, the relatively weak condition of *distinguishability* – roughly, that others' signals can be uniquely inferred from their truthful predictions [14], [27] – implies information aggregation.

The question is what happens when participants have multiple opportunities. This question is extremely relevant to the study of general markets, where this is a simplified model. But it is also relevant for more general strategic contexts as well, as strategic information withholding and revelation is a very general phenomenon. We later give other game-theoretic examples.

The best possible scenario for aggregation is an *all-rush* equilibrium: When participants are numbered sequentially in order of their first opportunity, then each participant i reports truthfully, fully revealing her information, before $i + 1$'s first opportunity. The worst possible scenario is an *all-delay* equilibrium: When participants are numbered sequentially in order of their *final* opportunity, participant i does not reveal any information until after $i - 1$'s final opportunity.

Challenges and prior work: Prior work [12] observes that the key challenge can be boiled down to the "Alice-Bob-Alice" market where Alice participates at times 1 and 3 while Bob participates at $t = 2$. Bob only participates once, hence is truthful by the observation above. Similarly, Alice is truthful at $t = 3$. The only question left is how Alice reports at $t = 1$.

Despite the problem being narrowed to this extent almost 10 years ago [12], it still seems quite difficult to address *a priori*. It would be ideal if Alice predicts truthfully; but she might withhold information, release some garbling of her signal, or even attempt to mislead or "bluff" Bob into making a mistake that she can profitably correct. It is unclear how to characterize her behavior with any generality considering the large array of possible scoring rules and information structures.

So far, only two special cases have been successfully resolved. [12] considered the log scoring rule and *conditionally independent* signals, *i.e.* for all e, $p(a_1, \ldots, a_n \mid e) = p(a_1 \mid e) \cdots p(a_n \mid e)$. The authors showed that, in this case, in the Alice-Bob-Alice game, Alice reports truthfully at $t = 1$; they extended this to show that any perfect Bayesian equilibrium, with any number of participants, is always all-rush with conditionally independent signals. Subsequently, [13] considered the log scoring rule with *(unconditionally) independent* signals, *i.e.* $p(a_1, \ldots, a_n) = p(a_1) \cdots p(a_n)$. The authors showed in this case that Alice does *not* report truthfully in the first round, hence in general markets, equilibria cannot be all-rush.

[14] again considered the log rule and independent signals, showing that Alice does not report *anything* at $t = 1$ and extending this to prove that any PBE is all-delay.

Unfortunately, it was not clear how one would generalize either the intuition or the technical approach of these works, which relied on special properties of the log scoring rule related to Shannon's entropy. For instance, another popular scoring rule is the *quadratic rule*, $S(p, e) = 2p - \|p\|_2^2$. Might Alice report truthfully for this rule when signals are conditionally independent? Unfortunately, the answer is no even for the popular quadratic ("Brier") scoring rule, as we show in the full version. So conditional independence is not the right general condition, and without that, it is very unclear what such a condition could be. Meanwhile, for the quadratic scoring rule, independent signals *do* induce all-delay equilibria (it is a corollary of a more general sufficient condition for complementarity, in the full version). It seems very intimidating to try to infer from examples such as these (even were they known prior to this work) any general formal, mathematical principle.

Our results: We show that informational substitutes and complements characterize, respectively, all-rush and all-delay equilibria in prediction markets.

Theorem II-.1. *With respect to the market's proper scoring rule S:*

1) *If signals are strong, strict substitutes, then for every set of participants and order of participation,* every *Bayes-Nash equilibrium is all-rush.*
2) *If signals are not strong substitutes, then there is a set of participants and order where* no *perfect Bayesian equilibrium is all-rush.*
3) *If signals are strong, strict complements, then for every set of participants and order,* every *perfect Bayesian equilibrium is all-delay.*
4) *If signals are not strong complements, then there is a set of participants and order where* no *perfect Bayesian equilibrium is all-delay.*

Sketch of ideas: A scoring rule S is a kind of decision problem where $S(p, e)$ is the utility for prediction p and outcome e of nature. Given S and an information structure, one has the associated value function $\mathcal{V}$, with $\mathcal{V}(A)$ the expected score for reporting truthfully after observing signal A.

Now consider for intuition the Alice-Bob-Alice market where Alice has signal A and Bob has signal B. For example, if the initial market prediction $p^{(0)} = p$, the prior on E, and Alice reports truthfully, then her expected net reward from this participation is

$$\mathop{\mathbb{E}}_{a,e} S(p_a, e) - \mathop{\mathbb{E}}_{e} S(p, e) = \mathcal{V}(A) - \mathcal{V}(\bot).$$

(Here, $\bot$ refers to a null, uninformative signal.) This is a general phenomenon: In a prediction market, *a report's expected profit is the marginal value of its information*. An important supporting insight is that, in equilibrium, no participant is deceived or "bluffed" by any other: Each knows the distribution of signals and the equilibrium strategies. This implies that Alice cannot really mislead Bob in equilibrium, but only withhold information from him. (Of course, an equilibrium must be *robust* to Alice's deceptions.)

We describe the intuition for perfect Bayesian equilibrium although, in the case of substitutes, we can also obtain the stronger Bayes-Nash result. In PBE, Bob will always report truthfully according to his beliefs at time $t = 2$ even if Alice deviates from equilibrium, and Alice will always report truthfully at time $t = 3$. This implies that the final market prediction is always p_{ab}, the posterior conditioned on $A = a, B = b$ when Alice and Bob receive these signals. Hence one can almost view the market as a constant-expected-sum game, where the total profit up for grabs is $\mathcal{V}(A \vee B) - \mathcal{V}(\bot)$.

Now let Alice report according to any strategy at $t = 1$; this corresponds to revealing some $A' \preceq A$ on the strong signal lattice. (Because Alice can reveal any less-informative signal than A via a randomized strategy.) Here, in equilibrium, Bob's utility will be $\mathcal{V}(B \vee A') - \mathcal{V}(A')$. Due to the constant-sum property described above, Alice wants to be choosing the $A' \preceq A$ that minimizes Bob's utility. Now the substitutes condition is precisely that the minimizing choice is $A' = A$; that is, Alice reports truthfully at $t = 1$. The complements condition is that $A' = \bot$, *i.e.* Alice reveals nothing at $t = 1$.

This line of reasoning doesn't yet prove the forward directions of the theorem (1 and 3), because we want to show that no other type of equilibria exist beyond the rush (delay) ones. However, we can use it to construct deviation strategies against non-rush (non-delay) strategy profiles. The complements case is intuitively more difficult or less robust here: With substitutes, it is easy to deviate to reporting truthfully early, but with complements, deviating to reporting later depends on how others react to the deviation strategy. The reverse directions (2 and 4) are relatively straightforward. ∎

The results of [14], [27] are implied by showing that their signal structures are substitutes/complements. This is done in the full version of the paper.

One note is that the fact that participants can employ randomized strategies corresponds directly to the need for the strong definitions of S&C here. If restricting to deterministic strategies, the proof would go through for moderate S&C; if restricting to the binary decision of reveal or not at each step, it would go through with weak S&C. And the difference has real bite: If we have only *e.g.* weak substitutes, we can easily construct counterexamples to rush equilibria where participants "partially" reveal the substitutable pieces of their information while withholding the complementary pieces.

Other game-theoretic results: In the full version of the paper, we show analogous results for more general or related settings involving strategic revelation and aggregation of information. In particular, we consider the crowdsourcing and machine-learning contest model of [15], [16] and the question-and-answer forums model of [17]. In both cases we can give new results and our other results have implications for identifying and designing good mechanisms in those settings.

3) Algorithmic results: In the SIGNALSELECTION problem, more formally defined in the full version, we are given a utility function u and prior P on an event E and signals $A_1, \dots, A_n$. We are also given a system of constraints such as a cardinality constraint (at most k signals may be selected) or a budget ("knapsack") constraint, where each signal has a price tag. The task is to output some feasible subset of signals to obtain subject to these constraints, so as to maximize the expected utility for making the decision given that information. To focus on the complexity of the information acquisition problem, we assume oracles for optimizing the utility function and interacting with the prior (which may have exponential size). This focus differs from a variety of related results (see *e.g.* [28]) for special cases, that address complexity of both problems at once.

The main positive result is that weak informational substitutes imply efficient $(1 - 1/e - o(1))$ approximations to this problem for cardinality, budget, and *matroid* constraints. The key idea is that the value function $\mathcal{V}$ is always monotone (because more information never hurts utility) and, under weak substitutes, is submodular on subsets of $\{1, \dots, n\}$. This allows us to apply algorithms for monotone submodular maximization [7], [8], [29].

On the other hand, we give a very strong reduction in the opposite direction: The maximization of any monotone $f : 2^{\{1,\dots,n\}} \to \mathbb{R}$ reduces to a special case of SIGNALSELECTION where all signals are independent, uniform bits and the decision problem is trivial to optimize. To build this reduction, we first apply a deep characterization from theory of proper scoring rules [30]–[32], which allows us to transparently construct a decision problem (in this case, a scoring rule) from a convex function $G : [0, 1]^n \to \mathbb{R}$. We then construct a convex G having the properties we need on the hypercube inductively. This result implies hardness, via known [33] or straightforward examples, even in this trivial setting of independent bits and a trivial-to-optimize decision problem: An exponential number of oracle queries are required to improve upon $1 - 1/e$ for substitutes or to get any approximation better than zero in general or even in the case of complements. Please see the full version for formal statements and proofs.

Under a strengthening of the substitutes condition, the positive results also extend (at least for cardinality constraints) to the adaptive setting where the decisionmaker selects a signal, observes its realization, then selects another, and so on. An extremely similar result already appears in [21], [34] under different terminology and slightly different model. The contribution of our results here is not just in formalizing this model or the generic positive and negative results, but also in drawing connections to the general and formal notion of informational substitutes, which (we have shown) also have broad implications in game theory. One upshot for algorithm designers is that our results described next may have consequences for identifying substitutable structure or even designing for it.

4) Structure and design: So far, we have seen that informational S&C are in a sense "unavoidable" both when studying equilibria in strategic information-revelation settings and when studying the complexity of information acquisition. Given this, their usefulness is greatly enhanced by improved understanding of the innate structure, applicability, and tractability of the definitions. Here, we describe our general convexity-based approach to these problems and some of our results, focusing on the following questions:

- What information structures are *nontrivial*, *universal* substitutes: informational substitutes for every decision problem?
- Given an information structure, is it always possible to design a decision problem for which signals are "somewhat strict" substitutes? Here, *somewhat strict* means that the signals are substitutes and sometimes have strictly positive marginal value.

Notice that universal substitutes would be a counterexample for being able to always design for complementarity, and vice versa. Perhaps surprisingly, we get somewhat different results for these two cases. However, we will give an intuitive geometric reason for why S&C fundamentally differ.

Theorem II-.2.

1) *Given any information structure, it is always possible to design a decision problem so that signals are somewhat strict moderate complements. Hence, all universal moderate substitutes are trivial. (Universal weak substitutes exist, but are "somewhat trivial".)*
2) *There exist nontrivial universal strong complements; an example is XORs of noisy independent bits. Hence, there exist information structures for which it is impossible to design a decision problem so that signals are somewhat strict weak substitutes.*

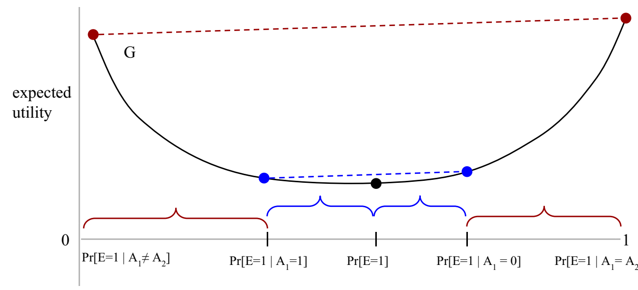
Here, *somewhat trivial* structures have a very special form that we can describe. Roughly, with some probability, we have a fully trivial case where all signals are completely redundant, *i.e.* after observing any one signal, the others provide no information. In other words, the structure must be a mixture distribution between this fully-trivial case and a potentially nontrivial one. (It also has other properties of interest, for instance, the realizations of the fully-trivial case are more "informative" than those of the nontrivial case.)

To show these kinds of results, we turn back to the problems of *prediction* with proper scoring rules (a special case of decision problems). It turns out that some straightforward observations about general decision problems and classic results on proper scoring rules [30]–[32] yield the following useful fact: *For every decision problem u there is a convex function $G : \Delta_E \to \mathbb{R}$, and for every such G there is a u, with $\mathcal{V}(A) = \mathbb{E}_a\, G(p_a)$.* In other words, to study substitutes and complements, it is in a sense enough to study convex functions on probability spaces. A key point is that this fact is constructive in both directions, so for instance, for our negative algorithmic results, we were able to build a convex G with the properties we needed, then use it to construct a decision problem.

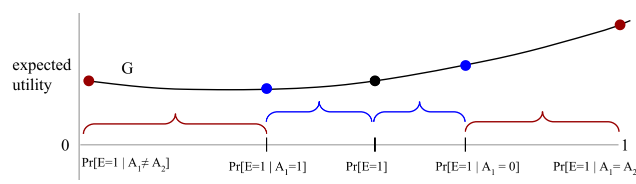
Given this approach, the intuition behind the above theorem can be entirely captured geometrically by properties of convex functions, and this is done in Figure 1.

Alternate definitions: Before closing, we highlight some additional tools for understanding informational S&C that appear in the full version. The first is to consider a *generalized entropy function* h and define substitutes for that entropy measure in terms of *diminishing bits of information* revealed by a signal. We show that each h corresponds to a decision problem u such that substitutes for h are substitutes for u. Shannon entropy corresponds to the problem of making a prediction when scored with the $\log$ scoring rule.

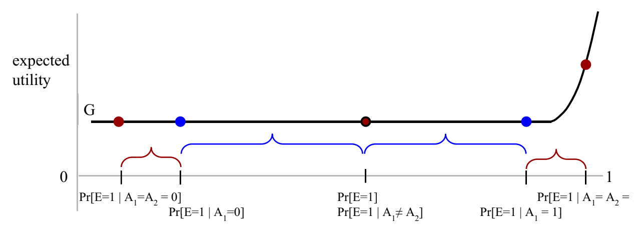
Similarly, the second is to consider a measure of distance between a prior belief and a posterior due to a Bayesian update; for this measure, signals are substitutes if they exhibit *diminishing distance* between prior and posterior beliefs. When the distance measure is a *Bregman divergence*,



(a) **Curvature increases complementarity.** Here A_1, A_2 are i.i.d. noisy bits and $E = A_1 \oplus A_2$, the XOR. Because G is convex, it is more extreme at more extreme beliefs, which tend to correspond to having multiple signals. Here the marginal value of a second signal is much higher than that of a first; complementarity.



(b) **Flattening G cannot remove complementarity.** For these signals (the same as in (a)), the Bayesian update on the second signal is always larger than on the first. Intuitively, even for flat G, this structure will always exhibit complementarity. These signals are universal complements.



(c) **Curvature destroys substitutability.** Here E is a uniform bit conditioned on which each $A_i = E$ independently with large probability. These are often substitutes, but by introducing high curvature at an extreme point, we create a problem where useful decisions can only be made with access to multiple signals. The first signal has no marginal value over the prior, while the second signal has some marginal value, so they are now complements. In general, this argument shows that universal substitutes must involve a "triviality" ruling out this construction.

Figure 1: **Universal substitutes and complements.** In each plot, there are two binary signals A_1, A_2 and a binary event E. The x-axis is the probability of E. G is the expected utility function for some decision problem (different in each plot). The circles correspond to the value of no signals (black), one signal (blue), and two signals (red). The blue braces measure the distance from the prior to posteriors on one signal; red measure additional distance to the posterior on two.

we show that it corresponds to a decision problem u such that the substitutes coincide. In particular, *KL-divergence* (relative entropy) corresponds to predicting against the $\log$ scoring rule.

III. Summary

The main contributions of this paper are to (1) propose a definition of informational substitutes and complements; (2) apply it to resolve an open problem in a basic game-theoretic setting, as well as obtain other game-theoretic and (3) algorithmic results; (4) give tools and results for understanding, identifying, and designing for informational S&C.

Concretely, we resolve several broad open problems in prediction markets: we show when information is immediately aggregated, when it is aggregated as late as possible, and whether it is possible to design markets for these cases (a problem that previously seemed out of reach). To do so, we developed very general definitions of informational S&C and gave tools for making them tractable along with geometric intuition. These definitions were shown to have applications in very general algorithm-design problems as well as other game-theoretic settings, addressing questions raised by prior work. This evidence suggests that informational S&C has a natural and useful role to play in both strategic and algorithmic settings, as well as in connecting the two.

Acknowledgment

Thanks to Thomas Steinke, Mark Bun, and Robert Kleinberg for their suggestions and feedback. We thank the support of the Siebel Foundation; and the National Science Foundation under grant CCF-1301976. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

References

[1] A. S. Kelso Jr and V. P. Crawford, "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, vol. 50, no. 6, pp. 1483–1504, 1982.

[2] A. E. Roth, "Stability and polarization of interests in job matching," *Econometrica: Journal of the Econometric Society*, vol. 52, no. 1, pp. 47–57, 1984.

[3] F. Gul and E. Stacchetti, "Walrasian equilibrium with gross substitutes," *Journal of Economic Theory*, vol. 87, no. 1, pp. 95–124, 1999.

[4] J. W. Hatfield and P. R. Milgrom, "Matching with contracts," *American Economic Review*, vol. 95, no. 4, pp. 913–935, 2005.

[5] M. Ostrovsky, "Stability in supply chain networks," *The American Economic Review*, vol. 98, no. 3, pp. 897–923, 2008.

[6] B. Lehmann, D. Lehmann, and N. Nisan, "Combinatorial auctions with decreasing marginal utilities," in *Proceedings of the 3rd ACM Conference on Electronic Commerce*, ser. EC '01. ACM, 2001, pp. 18–28.

[7] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM Journal on Computing*, vol. 40, no. 6, pp. 1740–1766, 2011.

[8] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operations Research Letters*, vol. 32, no. 1, pp. 41–43, 2004.

[9] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[10] A. S. Kyle, "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, vol. 53, no. 6, pp. 1315–1335, 1985.

[11] M. Ostrovsky, "Information aggregation in dynamic markets with strategic traders," *Econometrica*, vol. 80, no. 6, pp. 2595–2647, 2012.

[12] Y. Chen, D. M. Reeves, D. M. Pennock, R. D. Hanson, L. Fortnow, and R. Gonen, "Bluffing and strategic reticence in prediction markets," in *Proceedings of the 3rd International Conference on Internet and Network Economics*, ser. WINE '07. Springer-Verlag, 2007, pp. 70–81.

[13] S. Dimitrov and R. Sami, "Non-myopic strategies in prediction markets," in *Proceedings of the 9th ACM Conference on Electronic Commerce*, ser. EC '08. ACM, 2008, pp. 200–209.

[14] X. A. Gao, J. Zhang, and Y. Chen, "What you jointly know determines how you act: strategic interactions in prediction markets," in *Proceedings of the 14th ACM Conference on Electronic Commerce*, ser. EC '13. ACM, 2013, pp. 489–506.

[15] J. D. Abernethy and R. M. Frongillo, "A collaborative mechanism for crowdsourcing prediction problems," in *Advances in Neural Information Processing Systems 25*, ser. NIPS '11, 2011, pp. 2600–2608.

[16] B. Waggoner, R. Frongillo, and J. D. Abernethy, "A market framework for eliciting private data," in *Advances in Neural Information Processing Systems 29*, ser. NIPS '15, 2015, pp. 3492–3500.

[17] S. Jain, Y. Chen, and D. C. Parkes, "Designing incentives for online question-and-answer forums," *Games and Economic Behavior*, vol. 86, pp. 458 – 474, 2014.

[18] A. Krause and C. Guestrin, "Optimal nonmyopic value of information in graphical models - efficient algorithms and theoretical limits," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, ser. IJCAI '05, 2005.

[19] ——, "Near-optimal nonmyopic value of information in graphical models," in *21st Conference on Uncertainty in Artificial Intelligence*, ser. UAI '05, 2005.

[20] ——, "Optimal value of information in graphical models," *Journal of Artificial Intelligence Research*, vol. 35, pp. 557–591, 2009.

[21] D. Golovin and A. Krause, "Adaptive submodularity: theory and applications in active learning and stochastic optimization," *Journal of Artificial Intelligence Research*, vol. 42, pp. 427–486, 2011.

[22] Y. Chen, S. Javdani, A. Karbasi, J. A. Bagnell, S. Srinivasa, and A. Krause, "Submodular surrogates for value of information," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, ser. AAAI '15, 2015.

[23] R. J. Aumann, "Agreeing to disagree," *Annals of Statistics*, vol. 4, no. 6, pp. 1236–1239, 1976.

[24] D. Blackwell, "Equivalent comparisons of experiments," *The Annals of Mathematical Statistics*, vol. 24, no. 2, pp. 265–272, 1953.

[25] R. Hanson, "Combinatorial information market design," *Information Systems Frontiers*, vol. 5, no. 1, pp. 107–119, 2003.

[26] J. Abernethy, Y. Chen, and J. W. Vaughan, "Efficient market making via convex optimization, and a connection to online learning," *ACM Transactions on Economics and Computation*, vol. 1, no. 2, p. 12, 2013.

[27] Y. Chen, S. Dimitrov, R. Sami, D. M. Reeves, D. M. Pennock, R. D. Hanson, L. Fortnow, and R. Gonen, "Gaming prediction markets: equilibrium strategies with a market maker," *Algorithmica*, vol. 58, no. 4, pp. 930–969, 2010.

[28] A. Krause and D. Golovin, "Submodular function maximization," *Tractability: Practical Approaches to Hard Problems*, vol. 3, p. 19, 2012.

[29] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[30] J. McCarthy, "Measures of the value of information," *Proceedings of the National Academy of Sciences*, vol. 42, no. 9, pp. 654–655, 1956.

[31] L. J. Savage, "Elicitation of personal probabilities and expectations," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 783–801, 1971.

[32] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[33] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Mathematics of Operations Research*, vol. 3, no. 3, pp. 177–188, 1978.

[34] A. Asadpour, H. Nazerzadeh, and A. Saberi, "Stochastic submodular maximization," in *Proceedings of the 4th International Workshop on Internet and Network Economics*, ser. WINE '08. Springer Berlin Heidelberg, 2008, pp. 477–489.