Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

# A Geometric Method to Construct Minimal Peer Prediction Mechanisms

**Rafael Frongillo**
CU Boulder
raf@colorado.edu

**Jens Witkowski**
ETH Zurich
jensw@inf.ethz.ch

**Abstract**

Minimal peer prediction mechanisms truthfully elicit private information (e.g., opinions or experiences) from rational agents without the requirement that ground truth is eventually revealed. In this paper, we use a geometric perspective to prove that minimal peer prediction mechanisms are equivalent to power diagrams, a type of weighted Voronoi diagram. Using this characterization and results from computational geometry, we show that many of the mechanisms in the literature are unique up to affine transformations, and introduce a general method to construct new truthful mechanisms.

## 1 Introduction

User-generated content is essential to the effective functioning of many social computing and e-commerce platforms. A prominent example is eliciting information through crowdsourcing platforms, such as Amazon Mechanical Turk, where workers are paid small rewards to do so-called *human computation* tasks, which are easy for humans to solve but difficult for computers. For example, humans easily recognize celebrities in images, whereas even state-of-the-art computer vision algorithms perform significantly worse.

While statistical techniques can adjust for biases or identify noisy users, they are appropriate only in settings with repeated participation by the same user, and when user inputs are informative in the first place. But what if providing accurate information is costly for users, or if users have incentives to lie? Consider an image annotation task (e.g. for search engine indexing), where workers may wish to save effort by annotating with random words, or words that are too generic (e.g. "animal"). Or consider a public health program that requires participants to report whether they have ever used illegal drugs, and where participants may lie about their drug use due to shame or eligibility concerns.

Peer prediction mechanisms address these incentive problems. They are designed to elicit truthful private information from self-interested participants, such as answers to the question "Have you ever used illegal drugs?" Crucially, peer prediction mechanisms cannot use ground truth. In the public health example this means the program cannot verify whether a participant has or has not ever used illegal drugs; it can only use the participants' voluntary reports.



The classical peer prediction method (Miller, Resnick, and Zeckhauser 2005) addresses this challenge by comparing the reported information of a participant with that of another participant, and computing a payment rule which ensures that truth revelation is a strategic equilibrium. The major shortcoming of the classical peer prediction method with regard to practical applications is that it requires too much common knowledge. Bayesian Truth Serum mechanisms (Prelec 2004; Witkowski and Parkes 2012a; Radanovic and Faltings 2013) relax these common knowledge assumptions but require participants to report a probability distribution in addition to the actual information that is to be elicited. That is, they are not *minimal*. The 1/p mechanism (Jurca and Faltings 2008; 2011) and the Shadowing Method (Witkowski and Parkes 2012a; Witkowski 2014) relax the common knowledge assumptions of classical peer prediction to some degree while still being minimal.

**Our Results.** In this paper, we provide a complete characterization of the mechanism design space of minimal peer prediction, which includes the classical peer prediction method, output agreement, the 1/p mechanism, and the Shadowing Method as special cases. While it was known that every minimal mechanism requires some constraint on the agents' belief models (Jurca and Faltings 2011), it was unknown which constraints allow for truthful mechanism and how the constraints of different truthful mechanisms relate to one another. We answer these questions in Section 3 by adapting techniques from property elicitation, allowing us to prove the equivalence of minimal peer prediction mechanisms and power diagrams, a type of weighted Voronoi diagram. In Section 4, we then use this and results from computational geometry to show that all aforementioned mechanisms are unique with respect to their belief model constraints up to positive-affine transformation. One important corollary of this is that maximizing effort incentives for any of these mechanisms reduces to computing the effort-optimal positive-affine transformation. In Section 5, we exemplify how to construct new truthful mechanisms for new conditions. In Section 6, we revisit the classical peer prediction method and show how to compute a mechanism that is maximally-robust with respect to deviations between the mechanism's and the agents' belief models. We conclude with useful directions for future work in Section 7.

## 2 Preliminaries

In this section, we introduce the model, and review concepts in peer prediction and computational geometry.

### 2.1 Model

There is a group of $n \geq 2$ rational, risk-neutral and self-interested agents. When interacting with the environment, each agent $i$ observes a signal $S_i$,[1] which is a random variable with values $[m] := \{1, \ldots, m\}$ and $m \geq 2$. The signal represents an agent's experience or opinion. The objective in peer prediction is to elicit an agent's signal in an incentive compatible way, i.e. to compute payments such that agents maximize their expected payment by reporting their signal to the mechanism (center) truthfully.

To achieve this, all peer prediction mechanisms require that agent $i$'s signal observation tells her something about the signal observed by another *peer* agent $j \neq i$. For example, this could be agent $j = i + 1$ (modulo $n$), so that the agents form a "ring," where every agent is scored using the "following" agent. (Our results hold for any choice of peer agent.) Let then

$$p_i(s_j|s_i) = \Pr_i(S_j = s_j \mid S_i = s_i) \tag{1}$$

denote agent $i$'s *signal posterior* belief that agent $j$ receives signal $s_j$ given agent $i$'s signal $s_i$. We refer to $p_i(\cdot|\cdot)$ as agent $i$'s *belief model*. A crucial assumption for the existence of strictly incentive compatible peer prediction mechanisms is that every agent's belief model satisfies *stochastic relevance* (Johnson, Pratt, and Zeckhauser 1990).

**Definition 1.** *Random variable $S_i$ is* stochastically relevant *for random variable $S_j$ if and only if the distribution of $S_j$ conditional on $S_i$ is different for all possible values of $S_i$.*

That is, stochastic relevance holds if and only if $p_i(\cdot|s_i) \neq p_i(\cdot|s_i')$ for all $i \in [n] := \{1, \ldots, n\}$ and all $s_i' \neq s_i$. Intuitively, one can think of stochastic relevance as correlation between different agents' signal observations.

Similar to the signal posteriors, we denote agent $i$'s *signal prior* belief about signal $s_j$ by $p_i(s_j) = \Pr_i(S_j = s_j)$. Note that for the prior and the posteriors it holds that $p_i(s_j) = \sum_{k=1}^{m} p_i(s_j|k) \cdot p_i(k)$. Moreover, going forward, it is assumed that $p_i(s) > 0$ for all $s \in [m]$ and $i \in [n]$.

### 2.2 Peer Prediction Mechanisms

We are now ready to define peer prediction mechanisms.

**Definition 2.** *A (minimal)* peer prediction mechanism *is a function $M : [m] \times [m] \to \mathbb{R}$, where $M(x_i, x_j)$ specifies the payment to agent $i$ when she reports signal $x_i$ and her peer agent $j$ reports signal $x_j$.*

We use ex post subjective equilibrium (Witkowski and Parkes 2012b), which is the most general solution concept for which truthful peer prediction mechanisms are known.

**Definition 3.** *Mechanism $M$ is* truthful *if we have*

$$s_i = \operatorname*{argmax}_{x_i} \mathop{\mathbf{E}}_{S_j}\Big[M\big(x_i, S_j\big) \mid S_i = s_i\Big],$$

*for all $i \in [n]$ and all $s_i \in [m]$.*

[1] We will drop the subscript to denote a generic signal.

The equilibrium is *subjective* because it allows for each agent to have a distinct belief model, and *ex post* because it allows for (but doesn't require) knowledge of other agents' belief models. Ex post subjective equilibrium is strictly more general than Bayes-Nash equilibrium (BNE) as it coincides with BNE when all agents share the same belief model, i.e. if $p_i(\cdot|\cdot) = p_j(\cdot|\cdot)$ for all $i, j \in [n]$.

**Definition 4.** *A mechanism $M'$ is a* positive-affine transformation *of mechanism $M$ if there exists $f : [m] \to \mathbb{R}$ and $\alpha > 0$ such that for all $x_i, x_j \in [m]$, $M'(x_i, x_j) = \alpha M(x_i, x_j) + f(x_j)$.*

The importance of Definition 4 lies in the fact that if $M$ is truthful, then $M'$ is truthful as well. As we will see, in certain cases these are the *only* possible truthful mechanisms.[2]

**Lemma 1.** *Let $M'$ be a positive-affine transformation of $M$. Then $M'$ is truthful if and only if $M$ is truthful.*

### 2.3 Effort Incentives

Peer prediction mechanisms are especially useful for incentivizing effort, i.e. the costly acquisition of signals. Our modeling of effort follows Witkowski (2014).

**Definition 5.** *Given that agent $j$ invests effort and reports truthfully, the* effort incentive *$e_i(M)$ that is implemented for agent $i$ by peer prediction mechanism $M$ is the difference in expected utility of investing effort followed by truthful reporting and not investing effort, i.e.*

$$e_i(M) = \mathop{\mathbf{E}}_{S_i,S_j}\Big[M(S_i, S_j)\Big] - \max_{x_i \in [m]} \mathop{\mathbf{E}}_{S_j}\Big[M(x_i, S_j)\Big],$$

*where $x_i$ is agent $i$'s signal report that maximizes her expected utility according to the signal prior, and where the expectation is using agent $i$'s subjective belief model $p_i(\cdot|\cdot)$.*

An important observation is that positive-affine transformations of a mechanism simply scale its effort incentives; we will use this fact in Section 5 to optimize effort.

**Lemma 2.** *For any mechanism $M$, and any positive-affine transformation $M' = \alpha M + f$, we have $e_i(M') = \alpha e_i(M)$.*

### 2.4 The Probability Simplex

The intuition for our main results can be provided for $m = 3$ signals already, and so we give such examples throughout the paper. For probability distributions over only 3 signals, there is a convenient graphical representation of the *probability simplex* $\Delta_m$ as an equilateral triangle, where the three corners represent the signals (see Figure 1L). The closer a point is to a corner (the distance from the corner's opposing side), the more probability mass of that corner's signal is on that point.[3] The triangular shape ensures that for any point on the triangle the values of the three dimensions sum up to 1. For example, the point $y(a, b, c) = (1/2, 1/3, 1/6)$ in Figure 1L is at height $1/6$ (since the top corner represents

[2] See the appendix of the full version for all omitted proofs: http://www.cs.colorado.edu/~raf/media/papers/geometry.pdf

[3] This representation is equivalent to the natural embedding into $\mathbb{R}^3$ and viewing in the direction $(-1, -1, -1)$.

signal $c$), and one half away from the right side of the triangle (because the left corner represents signal $a$). Observe that with three signals, there are only two degrees of freedom since the entries have to sum up to 1. Therefore, by fixing the point's position with respect to $a$ and $c$, the value for signal $b$ is fixed as well. (Confirm that $\boldsymbol{y}$ is one third away from the left side.)

### 2.5 Power Diagrams

Our results rely on a concept from computational geometry known as a *power diagram*, which is a type of weighted Voronoi diagram (Aurenhammer 1987a).

**Definition 6.** *A* power diagram *is a partitioning of $\Delta_m$ into sets called* cells*, defined by a collection of points $\{\boldsymbol{v}^s \in \mathbb{R}^m : s \in [m]\}$ called* sites *with associated weights $w(s) \in \mathbb{R}$, given by*

$$\text{cell}(\boldsymbol{v}^s) = \left\{ \boldsymbol{u} \in \mathbb{R}^m : s = \underset{x \in [m]}{\text{argmin}} \{ \|\boldsymbol{u} - \boldsymbol{v}^x\|^2 - w(x) \} \right\}.$$

*We call $\|\boldsymbol{u} - \boldsymbol{v}^x\|^2 - w(x)$ the* power distance *from $\boldsymbol{u}$ to site $\boldsymbol{v}^x$; thus, for every point $\boldsymbol{u}$ in $\text{cell}(\boldsymbol{v}^s)$, it holds that $\boldsymbol{v}^s$ is closer to $\boldsymbol{u}$ in power distance than any other site $\boldsymbol{v}^x$.*

We have defined power diagrams for the special case of the probability simplex, which is the case we need in this paper. The more general definition allows for a different number of sites than dimensions.[4] The usual definition of a Voronoi diagram follows by setting all weights $w(s)$ to 0.

## 3 Mechanisms and Power Diagrams

As with previous work, we would like to make statements of the form, "As long as the belief models satisfy certain constraints, the mechanism is truthful." For example, the Shadowing Method (Witkowski and Parkes 2012a; Witkowski 2014) is truthful if and only if $p_i(s|s) - y(s) > p_i(s'|s) - y(s')$ for all $s, s' \in [m] : s' \neq s$, and some distribution $y(\cdot)$, which is a parameter of the mechanism (also see Figure 1L.) When used directly (and not as a building block for more complex mechanisms), it is often assumed that there is a known, common signal prior $p(\cdot) = p_i(\cdot)$ for all $i \in [n]$, which is then used as $y(\cdot) = p(\cdot)$. As we will see, both the Shadowing Method and the 1/p mechanism (Jurca and Faltings 2008; 2011) are actually robust in that they are truthful even if there is no common signal prior. All that is required is that the agents' possible posteriors fall into the correct regions. While it has been known that the constraints required by the Shadowing Method and the 1/p mechanism are incomparable, i.e. there exist belief models for which the Shadowing Method is truthful but the 1/p mechanism is not, and vice versa (Witkowski 2014), it was not known for which constraints there exist truthful mechanisms. In this section, we answer this question, and characterize all belief model constraints for which truthful mechanisms exist.

As a first step towards this goal, we formally define these constraints on belief models that limit which posteriors are possible following which signal.

[4] Also, note that we exclude cell boundaries; see Theorem 6.

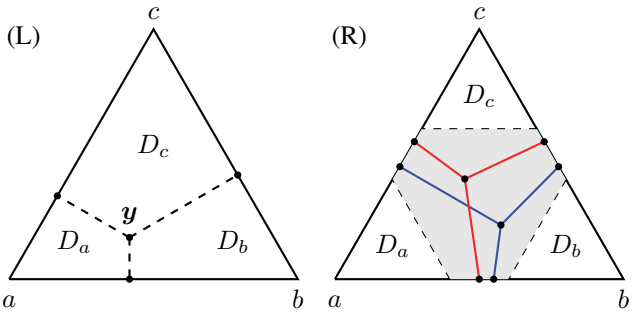


Figure 1: **(L)** Example of the probability simplex with three signals ($a$, $b$, and $c$) and intersection point $\boldsymbol{y} = (1/2, 1/3, 1/6)$. The regions $D_s$ are those for which the Shadowing Method incentivizes the agents to report the respective signal $s$. For example, for any posterior belief falling into $D_c$, agent $i$ should report $x_i = c$. To see that the depicted partitioning is indeed coming from this constraint, first observe that there is indifference at the intersection point, i.e. when $p_i(\cdot|s) = y(\cdot)$, and that the constraints are linear, so that the indifference borders are lines. The only remaining piece is then to determine the points for each pair of signals, where the third (left out) signal's weight is 0, and draw a line between that point and $\boldsymbol{y}$. For example, the indifference point between signals $b$ and $c$, where $a$ has no weight is $(0, 7/12, 5/12) \Leftrightarrow p_i(c|c) - 1/6 = p_i(b|c) - 1/3; p_i(a|c) = 0$. **(R)** A non-maximal constraint, with two consistent power diagrams (red and blue).

**Definition 7.** *A* belief model constraint *is a collection $\mathcal{D} = \{D_s \subseteq \Delta_m : s \in [m]\}$ of disjoint sets $D_s$ of distributions. If additionally we have $\text{cl}(\cup_s D_s) = \Delta_m$, i.e. if $\mathcal{D}$ partitions the simplex, we say $\mathcal{D}$ is* maximal*.*

A belief model constraint $\mathcal{D} = \{D_1, \ldots, D_m\}$ ensures that for each agent $i$, following signal observation $S_i = s_i$, her belief about her peer agent's signal $s_j$ is restricted to be in $D_{s_i}$. It is easy to come up with non-maximal belief model constraints, such as "$\forall s\ p(s|s) > 0.6$" (Figure 1R). Note that under such a constraint, some distributions are not valid posteriors for any signal. In contrast, a maximal constraint covers the simplex, partitioning it into $m$ bordering but non-overlapping regions (Figure 1L).

We can now talk about mechanisms being truthful with respect to belief model constraints.

**Definition 8.** *A mechanism $M(\cdot, \cdot)$ is truthful with respect to belief model constraint $\mathcal{D}$ if $M$ is truthful whenever $p_i(\cdot|s) \in D_s$ for all agents $i \in [n]$ and all signals $s \in [m]$.*

It directly follows from this perspective that all minimal peer prediction mechanisms require a belief model constraint (Jurca and Faltings 2011). Consider, for example, a posterior belief $p_i(\cdot|s) = (3/5, 3/20, 1/4)$. Without any constraint on the belief model, it is not clear if this is the posterior following signal 1, 2, or 3. This choice needs to be made since a given posterior belief can only belong to one signal (stochastic relevance, Definition 1), and so every truthful minimal peer prediction mechanism requires a belief model constraint.

One very natural constraint to consider is to take an arbitrary mechanism $M$ and restrict to only those belief models under which $M$ is truthful. It turns out that this set can be succinctly described by a belief model constraint, which we call the constraint *induced* by $M$. Moreover, the regions of this induced constraint must take a particular shape, that of a power diagram, and conversely, *every* power diagram is an induced constraint of some mechanism.

### 3.1 Induced Constraints

We will now observe that for any mechanism $M$, there is a belief model constraint $\mathcal{D}^M$, which exactly captures the set of belief models for which $M$ is truthful. In other words, not only is $M$ truthful with respect to $\mathcal{D}^M$, but under any belief model that does not satisfy $\mathcal{D}^M$, $M$ will not be truthful. The construction of $\mathcal{D}^M$ is easy: for each signal $s$, $D_s^M$ is the set of distributions $p(\cdot|s)$ under which $x_i = s$ is the unique optimal report for $M$. Note that if $D_s^M$ is empty for any $s$, then $M$ is not truthful for any belief model.

**Lemma 3.** *Let $M : [m] \times [m] \to \mathbb{R}$ be an arbitrary mechanism, and let $\mathcal{D}^M$ be the belief model constraint given by*

$$D_s^M = \left\{ p_i(\cdot|s) : s = \operatorname*{argmax}_{x_i} \mathop{\mathbf{E}}_{S_j \sim p_i(\cdot|s)} M(x_i, S_j) \right\} .$$

*Then $M$ is truthful with respect to $\mathcal{D}^M$, but not truthful for belief models not satisfying $\mathcal{D}^M$. Moreover, if the rows of $M$ are all distinct, $\mathcal{D}^M$ is maximal.*

*Proof.* Suppose $p_i(\cdot|s) \in D_s^M$ for all $i \in [n], s \in [m]$. Then by construction of $D_s^M$, an agent $i$ receiving signal $s$ maximizes expected payoff by reporting $s$, and hence $M$ is truthful. By definition then, $M$ is truthful with respect to $\mathcal{D}^M$. Now suppose $p_i(\cdot|s) \notin D_s^M$ for some $i \in [n], s \in [m]$. Then $s \neq \operatorname{argmax}_{x_i} \left\{ \mathbf{E}_{S_j \sim p_i(\cdot|s)} M(x_i, S_j) \right\}$, and thus $M$ cannot be truthful. Finally, consider the convex function $G(p) = \max_x \mathbf{E}_{S_j \sim p} M(x, S_j)$. By standard results in convex analysis (c.f. (Frongillo and Kash 2014, Theorem 3)) $G$ has subgradient $M(x, \cdot)$ whenever $x$ is in the $\operatorname{argmax}$. As the rows of $M$ are all distinct, multiple elements in the $\operatorname{argmax}$ corresponds to multiple subgradients of $G$, and thus $G$ is nondifferentiable[5] at the set of indifference points $\{p : |\operatorname{argmax}_x \mathbf{E}_{S_j \sim p} M(x, S_j)| > 1\}$. As $G$ must be differentiable almost everywhere (Aliprantis and Border 2007, Thm 7.26), these indifference points must have measure 0 in the probability simplex, and thus $\mathcal{D}^M$ is maximal. □

### 3.2 Equivalence to Power Diagrams

We have seen that every mechanism $M$ induces some belief model constraint $\mathcal{D}^M$, and that $M$ is truthful with respect to $\mathcal{D}^M$. We now show further that $\mathcal{D}^M$ is a power diagram, and conversely, that every power diagram has a mechanism such that $D_s^M = \operatorname{cell}(\boldsymbol{v}^s)$ for all $s$.

[5]Technically speaking, we should restrict to the first $m-1$ coordinates of the distribution, so that $G$ is defined on a full-dimensional subset of $\mathbb{R}^{m-1}$; this can easily be done by a linear transformation without altering the argument.

The concrete mapping is as follows. Given mechanism $M : [m] \times [m] \to \mathbb{R}$, we construct sites and weights by:

$$\boldsymbol{v}^s = M(s, \cdot), \quad w(s) = \|\boldsymbol{v}^s\|^2 = \|M(s, \cdot)\|^2 . \tag{2}$$

Conversely, given a power diagram with sites $\boldsymbol{v}^s$ and weights $w(s)$, we construct the mechanism $M$ as follows:

$$M(x_i, x_j) = \boldsymbol{v}^{x_i}(x_j) - \tfrac{1}{2}\|\boldsymbol{v}^{x_i}\|^2 + \tfrac{1}{2} w(x_i) , \tag{3}$$

where $\boldsymbol{v}^{x_i}(x_j)$ is the $x_j$th entry of $\boldsymbol{v}^{x_i}$. We note that these formulas are more explicit versions of those appearing in *property elicitation*, a domain requiring ground truth; our results are in cases direct translations from that literature (Lambert and Shoham 2009; Frongillo and Kash 2014).

With these conversions in hand, we can now show that they indeed establish a correspondence between minimal peer prediction mechanisms and power diagrams.

**Theorem 4.** *Given any mechanism $M : [m] \times [m] \to \mathbb{R}$, the induced belief model constraint $\mathcal{D}^M$ is a power diagram. Conversely, for every power diagram given by sites $\boldsymbol{v}^s$ and weights $w(s)$, there is a mechanism $M$ whose induced belief model constraint $\mathcal{D}^M$ satisfies $D_s^M = \operatorname{cell}(\boldsymbol{v}^s)$ for all $s$.*

*Proof.* Observe that if either relation (2) or (3) holds, we have the following for all $x, \boldsymbol{p}$:

$$-2\boldsymbol{p} \cdot \boldsymbol{v}^x + \|\boldsymbol{v}^x\|^2 - w(x) = -2 \mathop{\mathbf{E}}_{S_j \sim \boldsymbol{p}} [M(x, S_j)] . \tag{4}$$

To see this, note that $\boldsymbol{p} \cdot M(x, \cdot) = \mathbf{E}_{S_j \sim \boldsymbol{p}}[M(x, S_j)]$. Adding $\|\boldsymbol{p}\|^2$ to both sides of Eq. 4 gives

$$\|\boldsymbol{p} - \boldsymbol{v}^x\|^2 - w(x) = \|\boldsymbol{p}\|^2 - 2 \mathop{\mathbf{E}}_{S_j \sim \boldsymbol{p}} [M(x, S_j)] . \tag{5}$$

Now applying Eq. 5 to the definitions of a power diagram and of $\mathcal{D}^M$, we have

$$\begin{aligned}
\boldsymbol{p} \in \operatorname{cell}(\boldsymbol{v}^s) &\iff s = \operatorname*{argmin}_x \{\|\boldsymbol{p} - \boldsymbol{v}^x\|^2 - w(x)\} \\
&\iff s = \operatorname*{argmin}_x \Big\{ \|\boldsymbol{p}\|^2 - 2 \mathop{\mathbf{E}}_{S_j \sim \boldsymbol{p}} [M(x, S_j)] \Big\} \\
&\iff s = \operatorname*{argmax}_x \mathop{\mathbf{E}}_{S_j \sim p(\cdot|s)} M(x, S_j) \\
&\iff \boldsymbol{p} \in D_s^M .
\end{aligned}$$

Finally, as Eq. 2 defines a power diagram for any mechanism $M$, and Eq. 3 defines a mechanism for any power diagram, we have established our equivalence. □

**Corollary 5.** *Let $\mathcal{D}$ be a maximal belief model constraint. Then there exists a mechanism which is truthful with respect to $\mathcal{D}$ if and only if $\mathcal{D}$ is a power diagram.*

Note that the conversion from mechanisms to power diagrams (Eq. 2) and back (Eq. 3) are inverse operations. In particular, this shows that mechanisms are in one-to-one correspondence with power diagrams on $\Delta_m$. In the following section, we will leverage this tight connection, and use results from computational geometry to show that several well-known mechanisms are unique in the sense that they are the only mechanisms, up to positive-affine transformations, that are truthful for their respective belief model constraints.

## 4 Uniqueness

Consider the standard output agreement mechanism $M(x_i, x_j) = 1$ if $x_j = x_i$ and 0 otherwise. It is easy to see that the mechanism is truthful as long as each agent assigns the highest posterior probability $p_i(\cdot|s_i)$ to their own signal $s_i$, yielding the constraint "$p(s|s) > p(s'|s)\ \forall s' \neq s$." The type of question we address in this section is: are there any other mechanisms than $M$ that are guaranteed to be truthful as long as posteriors satisfy this condition? We will show that, up to positive-affine transformations, the answer is no: output agreement is unique. Moreover, the Shadowing Method and the 1/p mechanism are also unique for their respective conditions on posteriors.

To get some intuition for this result, let us see why output agreement is unique for $m = 3$ signals $a, b, c$. From Theorem 4, we know that $\mathcal{D} = \mathcal{D}^M$, the induced belief model constraint, is a power diagram (which is depicted in Figure 2L). In general, there may be many sites and weights that lead to the same power diagram, and these may yield different mechanisms via Eq. 3. In fact, it is a general result that any positive scaling of the sites followed by a translation (i.e. some $\alpha > 0$ and $\boldsymbol{u} \in \mathbb{R}^m$ so that $\hat{\boldsymbol{v}}^s = \alpha \boldsymbol{v}^s + \boldsymbol{u}$ for all $s$) will result in the same power diagram for an appropriate choice of weights (Aurenhammer 1987a). As it turns out, such scalings and translations exactly correspond to positive-affine transformations when passing to a mechanism through Eq. 3. Thus, we only need to show that the sites for the output agreement power diagram are unique up to scaling and translation. Here, another useful property of sites comes into play: the line between sites of two adjacent cells must be perpendicular to the boundary between the cells. Examining Figure 2L, one sees that after fixing $\boldsymbol{v}^a$, the choice of $\boldsymbol{v}^b$ is constrained to be along the blue dotted line, and once $\boldsymbol{v}^a$ and $\boldsymbol{v}^b$ are chosen, $\boldsymbol{v}^c$ is fixed as the intersection of the red dotted lines. Thus, we can specify the sites by choosing $\boldsymbol{v}^a$ (a translation), and how far away from $\boldsymbol{v}^a$ to place $\boldsymbol{v}^b$ (a positive scaling). We can then conclude that output agreement for three signals is unique up to positive-affine transformations. We now give the general result.

**Theorem 6.** *If there exists a mechanism $M$ that is truthful for some maximal belief model constraint $\mathcal{D}$, and there is some $\boldsymbol{y} \in \Delta_m$ with $y(s) > 0\ \forall s$ such that $\cap_s \mathrm{cl}(D_s) = \{\boldsymbol{y}\}$, then $M$ is the unique truthful mechanism for $\mathcal{D}$ up to positive-affine transformations.*

*Proof.* As $M$ is truthful with respect to $\mathcal{D}$, we have $\mathcal{D} = \mathcal{D}^M$ and thus $\mathcal{D}$ is a power diagram $P$. By the assumption of the theorem, we observe that the only vertex (0-dimensional face) of $P$ must be $\boldsymbol{y}$, the intersection of the $\mathrm{cl}(D_s)$, as there are $m$ cells but $\Delta_m$ has dimension $m - 1$.[6] Thus, by definition of a *simple cell complex*,[7] as the vertex $\boldsymbol{y}$ is in the relative interior of $\Delta_m$, we see that the extension $\hat{P}$ of $P$ onto the affine hull of $\Delta_m$ must also have $\boldsymbol{y}$ as the only vertex. Now following the proof of (Frongillo and Kash 2014, Theorem 4), we note that (Aurenhammer 1987a, Lemma 1) and (Aurenhammer 1987c, Lemma 4) together imply the following: if $\hat{P}$ is represented by sites $\{\boldsymbol{v}^s\}_{s\in[m]}$ and weights $w(\cdot)$, then any other representation of $\hat{P}$ with sites $\{\hat{\boldsymbol{v}}^s\}_{s\in[m]}$ satisfies $\exists \alpha > 0, \boldsymbol{u} \in \mathbb{R}^m$ s.t. $\hat{\boldsymbol{v}}^s = \alpha \boldsymbol{v}^s + \boldsymbol{u}$ for all $s \in [m]$. In other words, all sites must be a translation and scaling of $\{\boldsymbol{v}^s\}$. To complete the proof, we observe that different choices of $\boldsymbol{u}$ and $\alpha$ (with suitable weights) merely yield an affine transformation of $M$ when passed through Eq. 3, and as any positive-affine transformation preserves truthfulness, the result follows. □

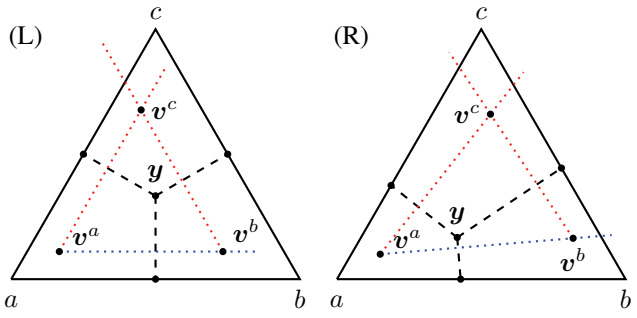


Figure 2: **(L)** Constructing a mechanism that is truthful under the same conditions as output agreement. **(R)** The truthful mechanism with respect to the "complement 1/p" condition in power diagram form with sites $\boldsymbol{v}^a, \boldsymbol{v}^b$, and $\boldsymbol{v}^c$. Notice that while the intersection point $\boldsymbol{y} = (1/2, 1/3, 1/6)$ is the same as in Figure 1L, the belief model constraint as depicted by the dashed partitioning is now given by the new "complement 1/p" condition.

**Corollary 7.** *The following mechanisms are unique, up to positive-affine transformations, with respect to the corresponding constraints (each with "$\forall s' \neq s$" implied):*

1. *Output Agreement,* $p(s|s) > p(s'|s)$
2. *Shadowing Method,* $p(s|s) - y(s) > p(s'|s) - y(s')$
3. *1/p Mechanism,* $p(s|s)/y(s) > p(s'|s)/y(s')$ .

*Proofs.* In all three cases, the given mechanism is known to be truthful for its respective belief model constraint. Moreover, for all three constraints $\mathcal{D}$, one can check that $\cup_s \mathrm{cl}(D_s) = \Delta_m$ and $\cap_s \mathrm{cl}(D_s) = \{\boldsymbol{y}\}$ meaning that $\boldsymbol{y}$ is the unique distribution bordering every set $D_s$ (for Output Agreement, $\boldsymbol{y}$ is the uniform distribution). Hence, the mechanisms are unique up to positive-affine transformations by Theorem 6. □

To conclude this section, we note that the term "maximal" is necessary in Theorem 6. If $\mathcal{D}$ is not maximal, there may be many more mechanisms that are truthful with respect to $\mathcal{D}$. For example, Figure 1R depicts two distinct power diagrams yielding mechanisms that are truthful with respect to the non-maximal constraint "$p(s|s) > 0.6$", and thus they are not merely positive-affine transformations of each other.

[6] Throughout the proof we implicitly work in the affine hull $\mathcal{A}$ of $\Delta_m$. We may assume without loss of generality that all sites lie in $\mathcal{A}$, as we may translate any site to $\mathcal{A}$ while adjusting its weight to preserve the diagram and resulting mechanism.

[7] $P$ is simple if each of $P$'s vertices is a vertex of exactly $m$ cells of $P$, the minimum possible (Aurenhammer 1987c, p.50).

## 5 Optimal Mechanisms for New Conditions

In this section, we exemplify the design of a new mechanism that is truthful with respect to a new condition. Moreover, we find the positive-affine transformation that maximizes effort incentives subject to a budget.[8] It then follows directly from Theorem 6 that the final mechanism's effort incentives are globally optimal given this condition. That is, there is no peer prediction mechanism that is truthful with respect to the new condition providing better effort incentives.

As intuition for the new condition, imagine the mechanism has an estimate of the agents' signal priors $p(a, b, c) = (0.01, 0.04, 0.95)$, which it designates as the intersection point $y(\cdot) = p(\cdot)$ of the belief model constraint. Consider now posterior $p_i(a, b, c|s) = (0.02, 0.01, 0.97)$, where the 1/p mechanism would pick signal $a$ since its relative increase from prior (as estimated by the mechanism) to posterior is highest (it doubles). However, one could also consider the relative decrease in "error": in a world without noise, the posterior would have $p_i(s|s) = 1$ for every $s$, and so signal $a$'s relative decrease from $0.99 = 1-0.01$ to $0.98 = 1-0.02$ is not as "impressive" as signal $c$'s decrease in error from $0.05 = 1 - 0.95$ to $0.03 = 1 - 0.97$ (a reduction of almost one half). Formalizing this intuition yields the "complement 1/p" condition, $\frac{1-y(s)}{1-p_i(s|s)} > \frac{1-y(s')}{1-p_i(s'|s)} \ \forall\, s' \neq s$.

Theorem 6 implies that there is a unique mechanism that is truthful for this new condition, up to positive-affine transformations. We now sketch the construction for the new "complement 1/p" condition, returning to our running example with $m = 3$ signals and intersection point $\boldsymbol{y} = (1/2, 1/3, 1/6)$ as depicted in Figure 2R. For the general procedure, see Appendix B in the full version of this paper.

1. Pick any point for $\boldsymbol{v}^a$, say $\boldsymbol{v}^a = (4/5, 1/10, 1/10)$.[9]
2. Pick any $\boldsymbol{v}^b$ on the blue dotted line, ensuring that the line between $\boldsymbol{v}^a$ and $\boldsymbol{v}^b$ is perpendicular to the $a, b$ cell boundary. Here we choose $\boldsymbol{v}^b = (1/10, 81/110, 9/55)$.
3. For all other signals $s$, $\boldsymbol{v}^s$ is now uniquely determined by $\boldsymbol{v}^a$ and $\boldsymbol{v}^b$ as the lines between any two sites must be perpendicular to their cell boundary. Here we only have one other signal, $c$, so we take $\boldsymbol{v}^c$ to be the unique point at the intersection of the red dotted lines, which is $\boldsymbol{v}^c = (38/275, 111/550, 33/50)$.
4. Calculate the weights by observing that $\boldsymbol{y}$ must be equidistant (in the power distance) to all sites simultaneously: $w(a) = 1, w(b) = 123/100, w(c) = 2423/1875$.
5. We obtain the resulting mechanism by applying Eq. 3:

$$M(\cdot,\cdot) = \frac{1}{1100}\begin{bmatrix} 1067 & 297 & 297 \\ 437 & 1137 & 507 \\ 563 & 633 & 1137 \end{bmatrix}.$$

6. The mechanism $M^*$ which optimizes effort incentives given a budget of 1 among all positive-affine transformations of $M$ is:

$$M^*(\cdot,\cdot) = \frac{1}{20}\begin{bmatrix} 15 & 0 & 0 \\ 0 & 20 & 5 \\ 3 & 8 & 20 \end{bmatrix}.$$

This step can be computed as follows: subtract the largest amount from each column that still preserves nonnegative payments, and then scale so the largest entry is 1.

We conclude by noting that this example condition admits a closed form solution: $M(s, s') = 0$ if $s = s'$, and $\frac{-1}{1-y(s)}$ otherwise. One can check that adding constants to each column to give non-negative payments recovers $M^*$. Of course, our construction applies even when no convenient closed-form solution exists, such as in Figure 3R.

## 6 Maximally-Robust Mechanisms

In the classical peer prediction method (Miller, Resnick, and Zeckhauser 2005), the mechanism is assumed to have full knowledge of the agents' belief models. Recent work relaxes the method's knowledge requirements, e.g. using additional reports (Prelec 2004; Witkowski and Parkes 2012b; 2012a) or using reports on several items (Dasgupta and Ghosh 2013; Witkowski and Parkes 2013). An approach closer to the classical method has been suggested by Jurca and Faltings (2007), who compute a minimal mechanism as the solution of a conic optimization problem that ensures truthfulness as long as the agents' belief models are close to the mechanism's, with respect to Euclidean distance. This restriction, a form of *robustness*, is defined as follows.

**Definition 9.** *(Jurca and Faltings 2007) A mechanism $M$ is $\epsilon$-robust with respect to belief model $p(\cdot|\cdot)$ if $M$ is truthful for $p^*(\cdot|\cdot)$ whenever the following holds for all $s_i \in [m]$,*

$$\sum_{s_j \in [m]} \left(p(s_j|s_i) - p^*(s_j|s_i)\right)^2 \leq \epsilon^2 . \tag{6}$$

While Jurca and Faltings fix the robustness $\epsilon$ as a hard constraint, one may also seek the mechanism that maximizes this robustness. The achievable robustness is of course limited by the mechanism's belief model $p(\cdot|\cdot)$; in particular, the "robustness areas" around the mechanism's posteriors cannot overlap; see Figure 3L. Viewing robustness in geometric terms, we obtain a closed-form solution.

**Theorem 8.** *Let $p(\cdot|\cdot)$ be the mechanism's belief model in classical peer prediction. Then the following mechanism is maximally robust:*

$$M(x_i, x_j) = p(x_j|x_i) - \frac{1}{2}\sum_{s=1}^{m} p(s|x_i)^2. \tag{7}$$

*Proof.* In light of Theorem 4, we may focus instead on power diagrams. From Eq. 6, for all $s$ we must have $B_\epsilon(p(\cdot|s)) \subseteq \text{cell}(\boldsymbol{v}^s)$, where $B_\epsilon(\boldsymbol{u})$ is the Euclidean ball of radius $\epsilon$ about $\boldsymbol{u}$ (restricted to the probability simplex). Letting $d = \min_{s,s' \in [m]} \|p(\cdot|s) - p(\cdot|s')\|$ be the minimum Euclidean distance between any two posteriors, it becomes clear that robustness of $d/2$ or greater cannot be achieved, as

[8] See Appendix C of the full version for the details of this optimization problem.

[9] For intuition and clarity, we choose $\boldsymbol{v}^a \in D_a$, but it could be in any other cell, e.g. $D_c$, or even outside of the simplex.

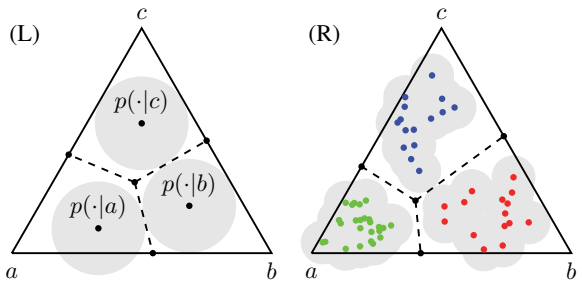


Figure 3: **(L)** Maximally-robust mechanism with respect to deviations from the mechanism's belief model. The sites of the power diagram are $\boldsymbol{v}^s = p(\cdot|s)$, and all weights are 0. **(R)** A mechanism with maximal robustness with respect to labeled posterior data.

$\frac{1}{2}p(\cdot|s)+\frac{1}{2}p(\cdot|s') \in B_{d/2}(p(\cdot|s)) \cap B_{d/2}(p(\cdot|s'))$. (See Figure 3L.) Robustness of any $\epsilon < d/2$ can be achieved, however, by taking a Voronoi diagram with sites $\boldsymbol{v}^s = p(\cdot|s)$; the definition of $d$ ensures that $B_\epsilon(p(\cdot|s)) = B_\epsilon(\boldsymbol{v}^s) \subseteq \text{cell}(\boldsymbol{v}^s)$ for all $s$. One then recovers Eq. 7 via Eq. 3 with $\boldsymbol{v}^s = p(\cdot|s)$ and $w(s) = 0$ for all $s \in [m]$. □

**Corollary 9.** *The classical peer prediction method with the quadratic scoring rule is maximally robust.*

One can easily adapt the above to design maximally robust mechanisms with respect to non-Euclidean distances as well, so long as that distance can be expressed as a Bregman divergence. Each such divergence has a corresponding scoring rule, and one simply replaces the quadratic score with this score (Frongillo and Kash 2014, Appendix F).

## 7 Conclusion

We introduced a new geometric perspective on minimal peer prediction mechanisms, and proved that it is without loss of generality to think of a minimal peer prediction mechanism as a power diagram. This perspective then allowed us to prove uniqueness of several well-known mechanisms up to positive-affine transformations, to construct novel peer prediction mechanisms for new conditions, and to compute the maximally-robust mechanism with respect to agents' subjective belief models deviating from the mechanism's.

We believe the most exciting direction for future work is to construct mechanisms from real-world data. Suppose the mechanism designer wants to collect typical data about the belief models of agents rather than making an educated guess at likely posteriors. A natural way to do this would be to use gold standard labels and elicit (signal, posterior) pairs from agents. Given these data, the designer can then train a classifier within the class of power diagrams, which intuitively predicts the signal associated with a new posterior. Finally, this power diagram can then be converted to a mechanism using Eq. 3. If a max-margin criterion is imposed when training, as depicted in Figure 3R, the resulting mechanism will be maximally robust with respect to the training set. When the data are not linearly separable, a soft-margin solution, such as the one by Borgwardt (2015), may be appropriate. We explore this approach in ongoing work.

## References


Aliprantis, C. D., and Border, K. C. 2007. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

Aurenhammer, F. 1987a. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing* 16(1):78–96.

Aurenhammer, F. 1987b. Recognising polytopical cell complexes and constructing projection polyhedra. *Journal of Symbolic Computation* 3(3):249–255.

Aurenhammer, F. 1987c. A criterion for the affine equivalence of cell complexes in R d and convex polyhedra in R d+1. *Discrete & Computational Geometry* 2(1):49–64.

Borgwardt, S. 2015. On Soft Power Diagrams. *Journal of Mathematical Modelling and Algorithms in Operations Research* 14(2):173–196.

Dasgupta, A., and Ghosh, A. 2013. Crowdsourced Judgement Elicitation with Endogenous Proficiency. In *Proceedings of the 22nd ACM International World Wide Web Conference (WWW'13)*, 319–330.

Frongillo, R., and Kash, I. 2014. General truthfulness characterizations via convex analysis. In *Web and Internet Economics*, 354–370. Springer.

Johnson, S.; Pratt, J. W.; and Zeckhauser, R. J. 1990. Efficiency Despite Mutually Payoff-Relevant Private Information: The Finite Case. *Econometrica* 58(4):873–900.

Jurca, R., and Faltings, B. 2007. Robust Incentive-Compatible Feedback Payments. In *Trust, Reputation and Security: Theories and Practice*, volume 4452 of *LNAI*. Springer-Verlag. 204–218.

Jurca, R., and Faltings, B. 2008. Incentives for Expressing Opinions in Online Polls. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC'08)*, 119–128.

Jurca, R., and Faltings, B. 2011. Incentives for Answering Hypothetical Questions. In *Proceedings of the 1st Workshop on Social Computing and User Generated Content (SC'11)*.

Lambert, N., and Shoham, Y. 2009. Eliciting truthful answers to multiple-choice questions. In *Proceedings of the 10th ACM conference on Electronic commerce*, 109–118.

Miller, N.; Resnick, P.; and Zeckhauser, R. 2005. Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science* 51(9):1359–1373.

Prelec, D. 2004. A Bayesian Truth Serum for Subjective Data. *Science* 306(5695):462–466.

Radanovic, G., and Faltings, B. 2013. A Robust Bayesian Truth Serum for Non-binary Signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI'13)*, 833–839.

Rybnikov, K. 1999. Stresses and Liftings of Cell-Complexes. *Discrete & Computational Geometry* 21(4):481–517.

Witkowski, J., and Parkes, D. C. 2012a. A Robust Bayesian Truth Serum for Small Populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI'12)*, 1492–1498.

Witkowski, J., and Parkes, D. C. 2012b. Peer Prediction Without a Common Prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12)*, 964–981.

Witkowski, J., and Parkes, D. C. 2013. Learning the Prior in Minimal Peer Prediction. In *Proceedings of the 3rd Workshop on Social Computing and User Generated Content (SC'13)*.

Witkowski, J. 2014. *Robust Peer Prediction Mechanisms*. Ph.D. Dissertation, Department of Computer Science, Albert-Ludwigs-Universität Freiburg.