General Truthfulness Characterizations Via Convex Analysis

Rafael M. Frongillo

University of Colorado Boulder

Ian A. Kash

University of Illinois at Chicago

Abstract

We present a model of truthful elicitation which generalizes and extends mechanisms, scoring rules, and a number of related settings that do not qualify as one or the other. Our main result is a characterization theorem, yielding characterizations for all of these settings. This includes a new characterization of scoring rules for non-convex sets of distributions. We combine the characterization theorem with duality to give a simple construction to convert between scoring rules and randomized mechanisms. We also show how a generalization of this characterization gives a new proof of a mechanism design result due to Saks and Yu.

1. Introduction

Information elicitation, the gathering of information from an agent by a principal, is a key problem in economics, statistics, machine learning, and finance. In these settings, one is interested in obtaining the preferences of an agent, a probability distribution from an expert, the desired prediction from an algorithm, and the risk of a portfolio, respectively. The key challenge in all of these settings is to reward the agent in such a way that the agent will truthfully reveal her knowledge rather than being encouraged to reveal some incorrect version. A central question in all these literatures has been to characterize all the ways this can be achieved; that is, to characterize all the truthful mechanisms, proper scoring rules, proper loss functions, or elicitable properties, respectively. Many variants on such characterization theorems have been proved. (See Section 1.2 for a partial list.) Moreover, the proofs of these results all use various tools from convex analysis; in particular, the characterizations tend to be in terms of convex functions and their subgradients.

Despite this commonality of question and technique, the literature on mechanism design has proceeded essentially independent of these other literatures and vice versa. Further, many characterizations are presented for a particular setting, and it is not immediately clear how they would apply to others. As a result, there are many theorems in the literature whose proofs are slight variations on existing results to adapt them to a new setting. While ex post the variations are slight, ex ante the needed changes were frequently not obvious and required significant effort to pin down. An excellent example of this is the scoring rules characterization by Gneiting and Raftery (2007), which has had at least three variants used in various settings (Boutilier, 2012; Chen and Kash, 2011; Cid-Sueiro, 2012).

In this paper we address these two problems by formulating a model of information elicitation which is general enough to encompass all these variants, and providing a characterization for it. This characterization obviates the need for additional new theorems in any future application that fits into our framework. Further, it provides a clearer understanding of the connections between mechanism design and scoring rules which allows us to translate results from one domain to the other.

Our model consists of a single agent endowed with some type t known only to the agent, who is asked to reveal her type to the principal. After doing so, the principal gives the agent a score $\mathsf{A}(t', t)$ that depends on both the agent's reported type t' and her true type t. We allow A to be quite general, with the main requirement being that $\mathsf{A}(t', \cdot)$ is an affine function (linear transformation plus a constant) of the true type t, and seek to understand when it is optimal for the agent to truthfully report her type. Given this truthfulness condition, it is immediately clear why convexity plays a central role—when an agent's type is t, we desire the score for telling the truth to satisfy $\mathsf{A}(t, t) = \sup_{t'} \mathsf{A}(t', t)$, which means this "consumer surplus" function $G(t) := \mathsf{A}(t, t)$ must be convex as the pointwise supremum of affine functions.

One special case of our model is mechanism design with a single agent, where the designer wishes to select an outcome based on the agent's type. In this setting, $\mathsf{A}(t', \cdot)$ can be thought of as the allocation and payment given a report of t', which combine to determine the utility of the agent as a function of her type. In this context, $\mathsf{A}(t, t)$ is the consumer surplus function (or indirect utility function), and Myerson's well-known characterization (1981) states that, in single-parameter settings, a mechanism is truthful if and only if the consumer surplus function is convex and its derivative (or subgradient at points of non-differentiability) is the allocation rule. More generally, this relationship remains true in higher dimensions (see Rochet (1985)). Here the restriction that $\mathsf{A}(t', \cdot)$ be affine is without loss of generality, because we view types as functions and function application is a linear operation. (See Section 3.2 for more details.)

Another special case is a proper scoring rule, also called a *proper loss* in the machine learning literature, where an agent is asked to predict the distribution of a random variable and given a score based on the observed realization of that variable. In this setting, types are distributions over outcomes, and $\mathsf{A}(t', t)$ is the agent's subjective expected score for a report that the distribution is t' when she believes the distribution is t. As an expectation, this score is linear in the agent's type. Gneiting and Raftery (2007) unified and generalized existing results in the scoring rules literature by characterizing proper scoring rules in terms of convex functions and their subgradients.

Further, the generality of our model allows it to include settings that do not quite fit into the standard formulations of mechanisms or scoring rules. These include counterfactual scoring rules for decision-making (Othman and Sandholm, 2010; Chen and Kash, 2011; Chen et al., 2011), proper losses for machine learning with partial labels (Cid-Sueiro, 2012), mechanism design with partial allocations (Cai et al., 2013), and responsive lotteries (Feige and Tennenholtz, 2010).

1.1. Our Contribution

Our main theorem (Theorem 1) is a general characterization theorem that generalizes and extends known characterization theorems for proper scoring rules (substantially) and truthful mechanisms (slightly). We also survey applications to related settings and show our theorem can be used to provide characterizations for them as well, including new results about mechanism design with partial allocation and responsive lotteries. Thus, our theorem eliminates the need to independently derive characterizations for such settings. Our unification also yields a new construction to convert between scoring rules and randomized mechanisms (see Construction 1). Finally, we conclude by examining cases where, rather than reporting their full type, agents select from a finite set of possible reports; Lambert and Shoham (2009) showed that, when the type is a probability distribution, the possible mappings from true type to optimal report(s) correspond to a generalization of Voronoi diagrams (see Section 5). We extend their result to settings where the private information need not be a probability distribution, and give a tight characterization for a particular restricted "simple" case.

1.1.1. Scoring Rules

Our contribution to the scoring rules literature is a characterization of proper scoring rules for non-convex sets of distributions, the first of its kind. As motivation, note that it is very natural to ask for scoring rules which are proper with respect to a particular family of distributions, e.g. if the principal was convinced that the agent's belief came from such a family. Yet many common families are non-convex, meaning they are not closed under mixtures, including most exponential families (Gaussian, Poisson, exponential, Laplace, Pareto, etc.) (Nielsen and Garcia, 2009). Moreover, general non-convex sets of distributions have proven useful as a way of separating informed and uninformed experts (Babaioff et al., 2011; Fang et al., 2010). Despite how natural and useful it is to restrict to the non-convex case, no characterization was known.

We give such a characterization, which surprisingly shows that the only scoring rules which are proper for a non-convex $\mathcal{P}$ are those which can be extended to a proper scoring rule on $\mathsf{Conv}(\mathcal{P})$; in other words, one does not gain flexibility by ruling out distributions in the "interior" of $\mathcal{P}$. Interestingly, the main technical tool we need to extend the proof to this case comes from the mechanism design literature, where characterizations for non-convex type spaces have been previously established. Additionally, we show that properness of a scoring rule is locally checkable, in the sense that it suffices to verify it in a neighborhood around each distribution. See Section Appendix B.2 and Corollary 3.

1.1.2. Mechanism Design

For mechanism design, our contributions are more modest. Our characterization is a minor extension of Archer and Kleinberg's characterization, removing a technical assumption (Archer and Kleinberg, 2008, Theorem 6.1). We show how many previous results about implementability and revenue equivalence can be translated into our framework, but do not introduce significant new results. Instead, the main interest of our approach is that by translating economic questions into convex analysis questions, we can simplify several proofs and expose the underlying intuition. Additionally, we show how known results about scoring rules yield a new proof of an implementability theorem due to Saks and Yu (2005).

1.1.3. Novel Elicitation Settings

Perhaps the most useful direct application of our main theorem is to elicitation settings that do not quite match the standard frameworks of scoring rules or mechanism design, as it immediately provides a characterization for such settings. In Section 4 we demonstrate the versatility of our characterization by surveying four recent such examples (on decision rules, proper losses for partial labels, mechanism design with partial allocations, and responsive lotteries), and showing how our results could have been applied.

1.1.4. Summary of Novel Results

Since our results cover a range of applications and include many reframings or small extensions of existing results, we summarize the main novel results here.

1. A general characterization theorem for many non-standard applications;
2. A characterization of scoring rules for non-convex sets of distributions;
3. Scoring rules are proper iff they are locally proper;
4. A new geometric proof of the Saks–Yu (2005) result on implementable mechanisms;
5. A new construction to convert between scoring rules and randomized mechanisms.

1.2. Related Work

The similarities between mechanisms and scoring rules were noted by (among others) Fiat et al. (2013), who gave a construction to convert mechanisms into scoring rules and vice versa, and Feige and Tennenholtz (2010), who gave techniques to convert both to "responsive lotteries." Further, techniques from convex analysis have a long history in the analysis of both models (see Gneiting and Raftery (2007); Vohra (2011)). However, we believe that our results use the "right" representation and techniques, which leads to more elegant characterizations and arguments. For example, the construction used by Fiat et al. has the somewhat awkward feature that the scoring rule corresponding to a mechanism has one more outcome than the mechanism did, a complication absent from our results. Similarly, the constructions used by Feige and Tennenholtz only handle special cases and they claim "there is no immediate equivalence between lottery rules and scoring rules," while we can give such an equivalence. So while prior work has understood that there is a connection, the nature of that connection has been far from clear.

A large literature in mechanism design has explored characterizations of when allocation rules can be truthfully implemented; see e.g. McAfee and McMillan (1988); Jehiel et al. (1996, 1999); Jehiel and Moldovanu (2001); Saks and Yu (2005); Bikhchandani et al. (2006); Müller et al. (2007); Archer and Kleinberg (2008); Ashlagi et al. (2010); Carroll (2012). Similarly, work on revenue equivalence can be cast in our framework as well (Myerson, 1981; Krishna and Maenner, 2001; Heydenreich et al., 2009; Carbajal and Ely, 2013). For scoring rules, our work connects to a literature that has used non-convex sets of probability distributions to separate (usefully) informed exports from uninformed experts (Fang et al., 2010; Babaioff et al., 2011).

A more general version of our setting looks at eliciting *properties* of the private information rather than the full information. The study of property elicitation in scoring rules can be traced to Leonard Savage in 1971, who considered the problem of eliciting expected values of random variables (Savage, 1971). The general case of arbitrary properties mapping types to reports has received considerable attention (Osband, 1985; Lambert et al., 2008; Gneiting, 2011), and we refer the interested reader to (Frongillo and Kash, 2019, Sections 3 and 4) for an extension of affine scores to this general setting. We instead focus on finite properties, where the set of possible reports is a finite set. Lambert and Shoham (2009) characterized elicitable finite properties, showing a connection to power diagrams from computational geometry. In mechanism design, finite properties correspond to settings where the mechanism has a finite set of allocations. Our results about finite properties provide a new proof of a theorem due to Saks and Yu (2005) that characterizes when allocation rules that select from a finite set of allocations have payments that make them truthful. Such mechanisms can be viewed as eliciting a ranking over outcomes rather than a utility for each outcome (common in, e.g., matching contexts), and our results are related to characterizations due to Carroll (2012). Subsequent to our work, our finite properties characterizations have been applied to characterize minimal peer prediction mechanisms (Frongillo and Witkowski, 2017).

1.3. Notation

We define $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ to be the extended real numbers. Given a set of measures M on a space X with σ-algebra $\mathcal{B}$, a function $f : X \to \overline{\mathbb{R}}$ is M-quasi-integrable if $\int_X f(x) d\mu(x) \in \overline{\mathbb{R}}$ for all $\mu \in M$. Let $\Delta(X)$ be the set of all probability measures on X. We denote by $\mathsf{Aff}(X \to Y)$ and $\mathsf{Lin}(X \to Y)$ the set of functions from $X \subseteq \mathcal{V}_1$ to $Y \subseteq \mathcal{V}_2$ which are restrictions of affine and linear functions (respectively) from vector space $\mathcal{V}_1$ to vector space $\mathcal{V}_2$. We write $\mathsf{Conv}(X)$ to denote the convex hull of a set of vectors X, the set of all (finite) convex combinations of elements of X. Some useful facts from convex analysis are collected in Appendix Appendix A.

2. Model and Main Result

We consider a general model with an agent who has a given type $t \in \mathcal{T}$ and reports some possibly distinct type $t' \in \mathcal{T}$, at which point the agent is rewarded according to some score $\mathsf{A}(t', t)$ which is affine in the true type t. This reward we call an affine score. We wish to characterize all *truthful* affine scores, which incentivize the agent to report her true type t.

Definition 1. *Let $\mathcal{T} \subseteq \mathcal{V}$ for some vector space $\mathcal{V}$ over $\mathbb{R}$, and let $\mathcal{A} \subseteq \mathsf{Aff}(\mathcal{T} \to \overline{\mathbb{R}})$.*[1] *A function $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ is an* affine score with score set $\mathcal{A}$ *if $\mathsf{A}(t, \cdot) \in \mathcal{A}$ for all $t \in \mathcal{T}$. When $\mathcal{A} = \mathsf{Aff}(\mathcal{T} \to \overline{\mathbb{R}})$, we simply call A an* affine score. *We say A is* truthful *if for all $t, t' \in \mathcal{T}$,*

$$\mathsf{A}(t', t) \leq \mathsf{A}(t, t). \tag{1}$$

If this inequality is strict for all $t \neq t'$, then A is strictly truthful.

Our characterization uses convex analysis, a central concept of which is the subgradient of a function, which is a generalization of the gradient yielding a linear approximation that is always below the function.

Definition 2. *Given some function $G : \mathcal{T} \to \mathbb{R}$, a function $d \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})$ is a* subgradient *to G at t if for all $t' \in \mathcal{T}$,*

$$G(t') \geq G(t) + d(t' - t). \tag{2}$$

We denote by $\partial G : \mathcal{T} \rightrightarrows \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})$ the multivalued map such that ∂G_t is the set of subgradients to G at t.[2] *We will occasionally overload the ∂G notation to mean $\partial G = \cup_{t \in \mathcal{T}} \partial G_t$. We say a parameterized family of linear functions $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})\}_{t \in \mathcal{T}'}$ for $\mathcal{T}' \subseteq \mathcal{T}$ is a* selection of subgradients *if $d_t \in \partial G_t$ for all $t \in \mathcal{T}'$; we denote this succinctly by $\{d_t\}_{t \in \mathcal{T}'} \in \partial G$.*

For mechanism design, it is typical to assume that utilities are always real-valued. However, the log scoring rule (one of the most popular scoring rules)

[1] While affine functions to the reals may be defined as linear functions plus a constant, one must be careful to define affine functions to the extended reals. In line with Waggoner (2021), we say $\ell(\cdot) \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})$ if (i) for all $\alpha \in \mathbb{R}$ and $t \in \mathcal{T}$ we have $\ell(\alpha t) = \alpha \ell(t)$, and (ii) for all $t, t' \in \mathcal{T}$ we have $\ell(t + t') = \ell(t) + \ell(t')$ whenever $\{\ell(t), \ell(t')\} \neq \{\infty, -\infty\}$. For (i) we adopt the convention $0 \cdot \infty = 0 \cdot (-\infty) = 0$. We define an affine function $a(\cdot) \in \mathsf{Aff}(\mathcal{V} \to \overline{\mathbb{R}})$ to be one of the form $a(t) = \ell(t_0 - t) + c$ for some $\ell \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}}), t_0 \in \mathcal{V}, c \in \mathbb{R}$. For example the function $1 + \ell_1 + \infty \cdot \ell_2$ for $\ell_1, \ell_2 \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})$ is in $\mathsf{Aff}(\mathcal{V} \to \overline{\mathbb{R}})$. See Waggoner (2021) for a full treatment.

[2] The literature often refers to subgradients taking the value $\pm\infty$ as *subtangents* (cf. (Gneiting and Raftery, 2007)). Waggoner (2021, Definition 3.1) uses the term *extended subgradient* to refer to both finite- and infinite-valued cases; for simplicity we use *subgradient* for both.

has the feature that if an agent reports that an event has probability 0, and then that event does occur, the agent receives a score of $-\infty$. Essentially solely to accommodate this, we allow affine scores and subgradients to take on values from the extended reals. In the next paragraph we provide the relevant definitions, but for most purposes it suffices to ignore these and simply assume that all affine scores are real-valued.

It is standard (cf. Gneiting and Raftery (2007)) to restrict consideration to the "regular" case, where intuitively only things like the log score are permitted to be infinite. In particular, an affine score A is *regular* if $\mathsf{A}(t,t) \in \mathbb{R}$ for all $t \in \mathcal{T}$, and $\mathsf{A}(t',t) \in \mathbb{R} \cup \{-\infty\}$ for $t' \neq t$. A regular affine score may therefore be written $\mathsf{A}(t',t) = \mathsf{A}_\ell(t', t-t') + c_{t'}$ for some $\mathsf{A}_\ell(t',\cdot) \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})$ and $c_{t'} = \mathsf{A}(t',t') \in \mathbb{R}$. Similarly, a parameterized family of linear functions (e.g. a family of subgradients) $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})\}_{t\in\mathcal{T}}$ is $\mathcal{T}$*-regular* if $d_t(t'-t) \in \mathbb{R} \cup \{-\infty\}$ for all $t,t' \in \mathcal{T}$. Recall that $d_t(t-t) = d_t(0) = 0$ by definition of Lin. Likewise, $\mathcal{T}$-regular affine functions have $\mathcal{T}$-regular linear parts with finite constants (i.e. we exclude the constant functions $\pm\infty$). For the remainder of the paper we assume all affine scores and parameterized families of linear or affine functions are $\mathcal{T}$-regular, where $\mathcal{T}$ will be clear from context.

2.1. Main Result

We now state, and prove, our characterization theorem. The proof takes Gneiting and Raftery's (2007) proof for the case of scoring rules on convex domains and extends it to the non-convex case using a variant of a technique introduced by Archer and Kleinberg (2008) for mechanisms with non-convex type spaces. This technique is essentially that used in prior work on extensions of convex functions (Peters and Wakker, 1987; Yan, 2014). The theorem applies for arbitrary score sets $\mathcal{A}$, though as A is given, this fact is not needed in the proof.

Theorem 1. *Let an affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ with score set $\mathcal{A}$ be given. A is truthful if and only if there exists some convex $G : \mathsf{Conv}(\mathcal{T}) \to \overline{\mathbb{R}}$ with $G(\mathcal{T}) \subseteq \mathbb{R}$, and some selection of subgradients $\{d_t\}_{t\in\mathcal{T}} \in \partial G$, such that*

$$\mathsf{A}(t',t) = G(t') + d_{t'}(t-t'). \tag{3}$$

Proof. It is immediate from the subgradient inequality (2) that the proposed form is in fact truthful, as

$$\mathsf{A}(t',t) = G(t') + d_{t'}(t-t') \leq G(t) = G(t) + d_t(t-t) = \mathsf{A}(t,t).$$

For the converse, we are given some truthful $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$.

By definition, each $\mathsf{A}(t,\cdot)$ is the restriction of an affine function defined on all of $\mathcal{V}$. While in general there may be many different affine functions on $\mathcal{V}$ whose restriction to $\mathcal{T}$ is $\mathsf{A}(t,\cdot)$, we now argue that $\mathsf{A}(t,\cdot)$ is well-defined on $\mathsf{Conv}(\mathcal{T})$. Let $\mathsf{A}_\ell(t,\cdot)$ denote the linear part of $\mathsf{A}(t,\cdot)$. Given any $\hat{t} \in \mathsf{Conv}(\mathcal{T})$ we may represent $\hat{t}$ as a finite convex combination $\hat{t} = \sum_{i=1}^m \alpha_i t_i$ where $t_i \in \mathcal{T}$. We have $\mathsf{A}(t,\hat{t}) = \mathsf{A}_\ell(t, \sum_{i=1}^m \alpha_i t_i) + \mathsf{A}(t,t) = \sum_{i=1}^m \alpha_i \mathsf{A}_\ell(t,t_i) + \mathsf{A}(t,t) = \sum_{i=1}^m \alpha_i \mathsf{A}(t,t_i)$, showing that $\mathsf{A}(t,\hat{t})$ does not depend on the choices of t_i and α_i.[3] Thus, $\mathsf{A}(t,\cdot)$ is both determined and well-defined on $\mathsf{Conv}(\mathcal{T})$.

Now we let $G(\hat{t}) \doteq \sup_{t\in\mathcal{T}} \mathsf{A}(t,\hat{t})$, which is convex as the pointwise supremum of convex (in our case affine) functions. Since A is truthful, we in particular have $G(t) = \mathsf{A}(t,t) \in \mathbb{R}$ for all $t \in \mathcal{T}$ by our regularity assumption. Then, also by truthfulness, we have for all $t' \in \mathcal{T}$ and $\hat{t} \in \mathsf{Conv}(\mathcal{T})$,

$$G(\hat{t}) \doteq \sup_{t\in\mathcal{T}} \mathsf{A}(t,\hat{t}) \geq \mathsf{A}(t',\hat{t}) = \mathsf{A}(t',t') + \mathsf{A}_\ell(t',\hat{t}-t') = G(t') + \mathsf{A}_\ell(t',\hat{t}-t'). \tag{4}$$

Hence, $\mathsf{A}_\ell(t',\cdot)$ satisfies (2) for G at t', so A is of the form (3). □

3. Mechanism Design and Scoring Rules as Affine Scores

In this section, we show how scoring rules and mechanisms fit comfortably within our framework.

3.1. Scoring Rules for Non-Convex $\mathcal{P}$

In this section, we show that the Gneiting and Raftery characterization is a simple special case of Theorem 1, and moreover that we generalize their result to the case where the set of distributions $\mathcal{P}$ may be non-convex. We also give a result about local properness derived using tools from mechanism design in Appendix Appendix B.2. To begin, we formally introduce scoring rules and show that they fit into our framework. The goal of a scoring rule is to incentivize an expert who knows a probability distribution to reveal it to a principal who can only observe a single sample from that distribution.

Definition 3. *Given outcome space $\mathcal{O}$ and set of probability measures $\mathcal{P} \subseteq \Delta(\mathcal{O})$, a* scoring rule *is a function $\mathsf{S} : \mathcal{P}\times\mathcal{O} \to \overline{\mathbb{R}}$ such that $\mathsf{S}(p,\cdot)$ is $\mathcal{P}$-quasi-integrable for all $p \in \mathcal{P}$ (see below). We say S is* proper *if for all $p, q \in \mathcal{P}$,*

$$\mathbb{E}_{o\sim p}[\mathsf{S}(q,o)] \leq \mathbb{E}_{o\sim p}[\mathsf{S}(p,o)]. \tag{5}$$

If the inequality in (5) is strict for all $q \neq p$, then S is strictly proper.

To incorporate this into our framework, take the type space $\mathcal{T} = \mathcal{P}$. Thus, we need only construct the correct score set of affine functions available to the scoring rule as payoff functions. Intuitively, these are the functions that describe what payment the expert receives given each outcome, but we have a technical requirement that the expert's expected utility be well defined. Thus, following Gneiting and Raftery, we take $\mathcal{F}$ to be the set of $\mathcal{P}$-quasi-integrable

[3]When working with the extended reals, this last expression follows as $\alpha_i \geq 0$, so by regularity of A, none of the summands in expression $\sum_{i=1}^m \alpha_i \mathsf{A}_\ell(t,t_i)$ can have value ∞. We omit this reasoning for the remainder of the paper, which shows up tacitly when the setting involves the extended reals.

functions $f : \mathcal{O} \to \overline{\mathbb{R}}$, meaning $\int_{\mathcal{O}} f(o)dp(o) \in \overline{\mathbb{R}}$ for all $p \in \mathcal{P}$, and the score set $\mathcal{A} = \{p \mapsto \int_{\mathcal{O}} f(o)\, dp(o) \mid f \in \mathcal{F}\}$. Note that in this case $\mathcal{A}$ actually contains *linear* functions of p, which are trivially affine.

We now apply Theorem 1, which yields the following generalization of Gneiting and Raftery (2007).

Corollary 1. *For an arbitrary set $\mathcal{P} \subseteq \Delta(\mathcal{O})$ of probability measures, a regular*[4] *scoring rule $\mathsf{S} : \mathcal{P} \times \mathcal{O} \to \overline{\mathbb{R}}$ is proper if and only if there exists a convex function $G : \mathsf{Conv}(\mathcal{P}) \to \mathbb{R}$ with functions $G_p \in \mathcal{F}$ such that*

$$\mathsf{S}(p, o) = G(p) + G_p(o) - \int_{\mathcal{O}} G_p(o')\, dp(o'), \tag{6}$$

where $G_p : q \mapsto \int_{\mathcal{O}} G_p(o')\, dq(o')$ is a subgradient of G for all $p \in \mathcal{P}$.

Proof. Truthfulness of the given form follows immediately from Theorem 1 and our definition of $\mathcal{A}$. For the converse, let $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ be a given truthful affine score. From the theorem, $\mathsf{A}(p, \cdot) = G(p) + d_p(\cdot - p) \in \mathcal{A}$, so we must have $d_p \in \mathcal{F}$; the subgradients are then of the form $G_p : q \mapsto \int_{\mathcal{O}} d_p(o)\, dq(o)$ as desired. $\square$

Importantly, Corollary 1 immediately generalizes the characterization of Gneiting and Raftery (2007) to the case where $\mathcal{P}$ is not convex, which is new to the scoring rules literature. One direction of this extension is obvious (if S is truthful on the convex hull of a set then it is truthful on that set), but the other is not, and is an important negative result in that it rules out the possibility of new scoring rules arising by restricting the set of distributions (as long as the restriction does not change the convex hull of the set).

In the absence of a characterization, several authors have worked in the non-convex $\mathcal{P}$ case. For example, Babaioff et al. (2011) examine when proper scoring rules can have the additional feature that uninformed experts do not wish to make a report (have a negative expected utility), while informed experts do wish to make one. They show that this is possible in some settings where the space of reports is not convex. Our characterization shows that, despite not needing to ensure properness on reports outside $\mathcal{P}$, essentially the only possible scoring rules are still those that are proper on all of $\Delta(\mathcal{O})$. We state the simplest version of such a characterization, for perfectly informed experts, here.

Corollary 2. *Let a non-convex set $\mathcal{P} \subseteq \Delta(\mathcal{O})$ and $\bar{p} \in \Delta(\mathcal{O}) - \mathcal{P}$ be given. A scoring rule S is proper and guarantees that experts with a belief in $\mathcal{P}$ receive a score of at least δ_A while experts with a belief of $\bar{p}$ receive a score of at most δ_R if and only if S is of the form (6) with $G(p) \geq \delta_A \forall p \in \mathcal{P}$ and $G(\bar{p}) \leq \delta_R$.*

With a similar goal to Babaioff et al., Fang et al. (2010) find conditions on $\mathcal{P}$ for which every continuous "value function" $G : p \mapsto \mathsf{S}(p, p)$ on $\mathcal{P}$ can be attained

[4] Just as for affine scores, regular scoring rules cannot be ∞ and only incorrect reports can yield $-\infty$.

by some S with the motivation of eliciting the expert's information when it is known to come from some family of distributions (which in general will not be a convex set). As such, they provide sufficient conditions on particular non-convex sets, as opposed to our result which provides necessary and sufficient conditions for all non-convex sets. Beyond these specific applications, our characterization is useful for answering practical questions about scoring rules. For example, suppose we assume that people have beliefs about probabilities in increments of 0.01. Does that change the set of possible scoring rules? No. What happens if they have finer-grained beliefs but we restrict them to such reports? They will end up picking a "nearby" report.[5]

In Appendix Appendix B.2, we show how local truthfulness conditions, where one verifies that an affine score is truthful by checking that it is truthful in a small neighborhood around every point, from mechanism design generalize to our framework. In particular Corollary 10 shows that local properness (i.e. properness for distributions in a neighborhood) is equivalent to global properness for scoring rules on convex $\mathcal{P}$, an observation that is also new to the scoring rules literature. See Appendix Appendix B.2 for the precise meaning of (weak) local properness (i.e. truthfulness).

Corollary 3. *For a convex set $\mathcal{P} \subseteq \Delta(\mathcal{O})$ of probability measures, a scoring rule $\mathsf{S} : \mathcal{P} \times \mathcal{O} \to \overline{\mathbb{R}}$ is proper if and only if it is (weakly) locally proper.*

3.2. Mechanism Design

We now show how to view a mechanism as an affine score. First, we formally introduce mechanisms in the single agent case. (See below for remarks about multiple agents.) Then we show how known characterizations of truthful mechanisms follow easily from our main theorem. This allows us to relax a minor technical assumption from the most general such theorem.

Definition 4. *Given outcome space $\mathcal{O}$ and a type space $\mathcal{T} \subseteq \mathbb{R}^{\mathcal{O}}$, consisting of functions mapping outcomes to reals, a (direct)* mechanism *is a pair (f, p) where $f : \mathcal{T} \to \mathcal{O}$ is an* allocation rule *and $p : \mathcal{T} \to \mathbb{R}$ is a* payment*. The utility of the agent with type t and report t' to the mechanism is $U(t', t) = t(f(t')) - p(t')$; we say the mechanism (f, p) is* truthful *if $U(t', t) \leq U(t, t)$ for all $t, t' \in \mathcal{T}$.*

Here we suppose that the mechanism can choose an allocation from some set $\mathcal{O}$ of outcomes, and there is a single agent whose type $t \in \mathcal{T}$ is itself the valuation function. That is, the agent's net utility upon allocation o and payment p is $t(o) - p$. Thus, following Archer and Kleinberg (2008), we view the type space $\mathcal{T}$ as lying in the vector space $\mathcal{V} = \mathbb{R}^{\mathcal{O}}$. The advantage of this representation is that while agent valuations in mechanism design can generally be complicated functions, viewed this way they are all linear: for any $v_1, v_2 \in \mathcal{V}$, we have

[5] In some cases, there will be no optimal report, such as when $\mathcal{P}$ is the set of Bernoulli(p) distributions with irrational p. Here for rational p, such as $p = 1/2$, there is merely a sequence of reports which approach the true distribution p.

$(v_1 + \alpha v_2)(o) = v_1(o) + \alpha v_2(o)$. Thus, we have an affine score $\mathsf{A}(t', t) \doteq U(t', t)$, with score set $\mathcal{A} = \{t \mapsto t(o) + c \mid o \in \mathcal{O}, c \in \mathbb{R}\}$, so that every combination of outcome and payment a mechanism could choose is an element of $\mathcal{A}$.

As an illustration of our theorem, consider the following characterization, due to Myerson (1981), for a single-parameter setting (i.e. when the agent's type can be described by a single real number). The result states that an allocation rule is implementable, meaning there is some payment rule making it truthful, if and only if it is *monotone* in the agent's type.

Corollary 4 (Myerson (1981))**.** *Let $\mathcal{T} = \mathbb{R}_+$, $\mathcal{O} \subseteq \mathbb{R}$, so that the agent's valuation is $t \cdot o$. Then a mechanism (f, p) is truthful if and only if*

1. *f is monotone non-decreasing in t,*
2. *$p(t) = tf(t) - \int_0^t f(t')dt' + p_0$.*

Proof. By elementary results in convex analysis f is a subgradient of a convex function on $\mathbb{R}$ if and only if it is monotone non-decreasing. By Theorem 1, the mechanism is truthful if and only if f is a subgradient of the particular function $G(t) = U(t, t) = t(f(t)) - p(t)$, which is equivalent to (i) and the condition $G(t) = \int_0^t f(t')dt' + C$. □

More generally, applying our theorem gives the following characterization. It is essentially equivalent to that of Archer and Kleinberg (2008) (their Theorem 6.1), although our approach allows the relaxation of a technical assumption they term "outcome compactness" which they require when the set of types is non-convex.

Corollary 5. *A mechanism (f, p) is truthful if and only if there exists a convex function $G : \mathsf{Conv}(\mathcal{T}) \to \mathbb{R}$ and some selection of subgradients $\{dG_t\}_{t \in \mathcal{T}}$, such that for all $t \in \mathcal{T}$, $f(t) = dG_t$ and $G(t) = t(f(t)) - p(t)$.*

We have written $f(t) = dG_t$, yet the allocation is an outcome $f(t) \in \mathcal{O}$ while the subgradient is a linear function $dG_t \in \mathsf{Lin}(\mathbb{R}^{\mathcal{O}} \to \mathbb{R})$. Recall the natural isomorphism between outcomes o and evaluation functions $E_o : t \mapsto t(o)$, which means that we can always represent an outcome as a function sending types to their value for that outcome. In that sense, we mean that dG_t is the function $dG_t : t' \mapsto t'(o)$ where $o = f(t)$. The intuition that the subgradient is the allocation function still holds up to this technicality.

While we have thus far dealt with single-agent deterministic mechanisms implemented in dominant strategies, this characterization actually applies significantly more broadly. In a sense, extending our characterizations to multiple agents is trivial: a mechanism is truthful if and only if it is truthful for agent i when fixing the reports of the other agents. Hence, we merely apply our characterization to each single-agent mechanism induced by reports of the other agents. This is sufficient for our present study, but there are certainly reasons to take a more nuanced approach to the multi-agent setting; see Section 6 for further discussion. To extend to randomized mechanisms, one can take $f : \mathcal{T} \to \Delta(\mathcal{O})$ and define $U(t', t) = \mathbb{E}_{o \sim f(t')}[t(o)] - p(t')$, which is still affine in t. We can even extend to non-risk-neutral agents by taking the outcome space to be $\mathcal{O}' \doteq \Delta(\mathcal{O})$. Finally, we can extend to Bayesian agents; in the above discussion of the multi-agent setting, take expectations instead of fixing specific types for the other agents.

Of course, mechanism design asks many questions beyond whether a particular mechanism is truthful, and some of these can be reframed as questions in convex analysis. Implementability focuses on the question of when there exist payments that make a given allocation rule truthful. Figure 1(a) illustrates known characterizations and how they were proved. As it shows, several of them rely on showing equivalence to a condition from convex analysis known as *cyclic monotonicity*. Instead, in Appendix Appendix B, we reprove them in our more general framework by showing equivalence to the condition of being subgradients of a convex function (see Figure 1(b)). This reframing has three main benefits. First, by exposing the essential convex analysis question, we are able to greatly simplify the proofs of some of these results. For example, the original proof of Theorem 11 relies on representing the allocation rule using a graph and arguing about the limit behavior of a process of creating paths in that graph. In contrast, our proof simply requires defining a function and showing it is convex with the correct subgradients by elementary arguments. Second, this reframing reveals that these results actually yield, we believe, new results in convex analysis (in particular, Theorems 10, 11, and 12 and Corollary 11). Third, this approach shows us how to translate known results from mechanism design into new results about scoring rules, as we saw in Section 3.1. While elements of a subgradient-based approach can be found in a variety of work on characterizing implementability (see, e.g.,McAfee and McMillan (1988); Jehiel et al. (1996, 1999); Jehiel and Moldovanu (2001); Krishna and Maenner (2001); Milgrom and Segal (2002); Bikhchandani et al. (2006)), this work has tended to use individual facts applied to particular settings, in contrast to our approach of translating mechanism design questions into convex analysis questions. Nevertheless, as these are essentially reframings of known results that do not directly provide new insights for mechanism design, we defer this material to Appendix Appendix B.

Revenue equivalence is the question of when all mechanisms with a given allocation rule charge the same prices (up to a constant). Translating this question into convex analysis terms, given a selection of a subgradient, when is the associated convex function unique up to a constant? We ask the more general question: what are all the convex functions consistent with a given selection of a subgradient? The result is a theorem, extending a result due to Kos and Messner (2013), that characterizes the possible payments of every truthful mechanism, even those that do not satisfy revenue equivalence. As their analysis essentially applies the natural convex analysis technique, we again defer this material to Appendix Appendix C.

3.3. Duality Between Scoring Rules and Mechanisms

Using our framework, we can also see a deeper connection between mechanisms and scoring rules than has been observed in the literature. In essence,

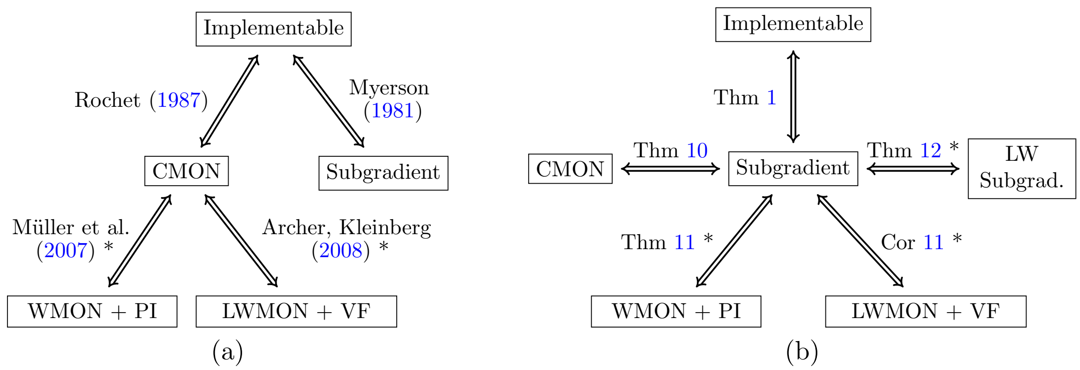

Figure 1: Proof structure of existing mechanism design literature (a), and the new proof structure presented in this paper (b). Asterisks (*) denote the requirement that $\mathcal{T}$ be convex. We write CMON for cyclic monotonicity, WMON for weak monotonicity, and PI for path independence. For the other abbreviations, L is local, W is weak, and VF is vortex freeness, a condition weaker than path independence introduced in Archer and Kleinberg (2008). Definitions of these conditions can be found in Appendix Appendix B.

scoring rules are dual mechanisms. In the scoring rule setting, an agent has a private distribution (their belief) and the principal gives the agent a payoff vector (the score), which assigns the agent a real-valued payment for each possible outcome. Dually, in a randomized mechanism, the agent possesses a private payoff vector (their type) encoding their value for each possible outcome, and the principal assigns a distribution (the allocation) over these outcomes. This observation allows us to give a very simple and natural construction to convert between scoring rules and mechanisms. Unlike previous constructions (e.g., Fiat et al. (2013)) we can establish a direct bijection between the two objects, and do not require any normalization.

Fix a finite outcome space $\mathcal{O}$.[6] Let $\mathcal{S} = \{\mathsf{S} : \mathcal{P}\times\mathcal{O} \to \overline{\mathbb{R}} \mid \mathcal{P} \subseteq \Delta(\mathcal{O}) \text{ is convex}\}$ denote the set of all scoring rules defined on convex sets. Similarly, let $\mathcal{M} = \{(f : \mathcal{T} \to \Delta(\mathcal{O}), \pi : \mathcal{T} \to \mathbb{R}) \mid \mathcal{T} \subseteq \mathbb{R}^{\mathcal{O}} \text{ is convex}\}$ denote the set of all randomized mechanisms defined on convex sets. We use π for prices to avoid confusion with probabilities. In what follows, we will write expectations as dot products, such as $t \cdot f(t') = \mathbb{E}_{o\sim f(t')}[t(o)]$. We also write $\mathsf{S}(p', p) \doteq \mathbb{E}_{o\sim p}\mathsf{S}(p', o)$. Recall that for a mechanism (f, π) we define $U(t', t) \doteq t \cdot f(t') - \pi(t')$.

Construction 1. *Given a scoring rule* $\mathsf{S} : \mathcal{P} \times \mathcal{O} \to \overline{\mathbb{R}}$*, let* $G_{\mathsf{S}} : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$ *be defined by* $G_{\mathsf{S}}(p) = \sup_{p'\in\mathcal{P}} \mathsf{S}(p', p)$ *for* $p \in \mathcal{P}$ *and* $G_{\mathsf{S}}(p) = \infty$ *otherwise, and set* $\mathcal{T}_{\mathsf{S}} \doteq \partial G_{\mathsf{S}}(\mathcal{P}) \cap \mathbb{R}^{\mathcal{O}}$*. Similarly, given a randomized mechanism* $\mathsf{M} = (f, \pi)$

[6] We believe this construction can be extended to infinite outcomes. For a suitable dual pair of vector spaces, and topology for part (2) (Aliprantis and Border, 2007), Theorem 2 extends immediately. The discussion following the construction relies on convex functions of Legendre type, i.e., essentially smooth and strictly convex. The results alluded to there hold for a suitable definition of Legendre type for infinite-dimensional settings, such as for Banach spaces (Bauschke et al., 2001, Section 5).

on type space $\mathcal{T}$, let $G_{\mathsf{M}} : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$ be given by $G_{\mathsf{M}}(t) = \sup_{t' \in \mathcal{T}} U(t', t)$ for $t \in \mathcal{T}$ and $G_{\mathsf{M}}(t) = \infty$ otherwise, and set $\mathcal{P}_{\mathsf{M}} \doteq \partial G_{\mathsf{M}}(\mathcal{T}) \cap \Delta(\mathcal{O})$. Define the multivalued maps $\Phi : \mathcal{S} \rightrightarrows \mathcal{M}$ and $\Psi : \mathcal{M} \rightrightarrows \mathcal{S}$ as follows,

$$\begin{aligned} \Phi(\mathsf{S}) &\doteq \{\mathsf{M} = (f, \pi) : \mathcal{T}_{\mathsf{S}} \to \Delta(\mathcal{O}) \times \mathbb{R} \mid \forall t \in \mathcal{T}_{\mathsf{S}}\ f(t) \in \partial G_{\mathsf{S}}^*(t),\ \pi(t) = t \cdot f(t) - G_{\mathsf{S}}^*(t)\} \\ \Psi(\mathsf{M}) &\doteq \{\mathsf{S} : \mathcal{P}_{\mathsf{M}} \times \mathcal{O} \to \overline{\mathbb{R}} \mid \forall p \in \mathcal{P}_{\mathsf{M}}, o \in \mathcal{O}\ \mathsf{S}(p, o) = G_{\mathsf{M}}^*(p) + d_p(o) - d_p \cdot p,\ d_p \in \partial G_{\mathsf{M}}^*(p)\}\ , \end{aligned} \tag{7}$$

where for any function $F : \mathbb{R}^k \to \overline{\mathbb{R}}$ we write $F^* : \mathbb{R}^k \to \overline{\mathbb{R}}$ for the convex conjugate, $F^*(x) = \sup_{d \in \mathbb{R}^k} d \cdot x - F(d)$.

Our construction takes a scoring rule and produces a set of mechanisms, and vice versa. We now show that these are inverse operations, under mild regularity assumptions. The set-valued nature of our construction arises from the choices in selecting subgradients in Theorem 1; as we discuss below, when there is only one choice for the subgradients, the maps Φ, Ψ become single-valued. Note that the duality in Construction 1 is different from prediction market duality; there the menu of possible reports changes, akin to posted-price mechanisms, whereas here the private type itself changes, from a belief to a valuation function.[7] See Frongillo and Kash (2019) for a treatment of both notions of duality, which includes prediction markets and posted-price mechanisms.

Theorem 2. *Let $\mathcal{S}_{\text{prop}} \subseteq \mathcal{S}$ be the set of proper scoring rules in $\mathcal{S}$, and $\mathcal{M}_{\text{truth}} \subseteq \mathcal{M}$ be the set of truthful mechanisms in $\mathcal{M}$. We have the following:*

1. *For any regular proper scoring rule S, any $\mathsf{M} \in \Phi(\mathsf{S})$ is a truthful mechanism; conversely, for any truthful mechanism M, any $\mathsf{S} \in \Psi(\mathsf{M})$ is a regular proper scoring rule.*
2. *Let $\mathcal{C}$ be the set of closed convex functions.[8] Restricting to the sets $\{\mathsf{S} \in \mathcal{S}_{\text{prop}} : G_{\mathsf{S}} \in \mathcal{C}\}$ and $\{\mathsf{M} \in \mathcal{M}_{\text{truth}} : G_{\mathsf{M}} \in \mathcal{C}\}$, the maps Φ and Ψ are inverses as multivalued maps.*

Proof. For part (1), note that by Corollaries 1 and 5, the forms are proper/truthful if we show that the convex functions are finite on the correct domain, and subdifferentiable in the case of G_{S}^* (recall that we allow vertical subgradients for scoring rules). First, let S be given. The set $\mathcal{T}_{\mathsf{S}}$ of finite subgradients of G_{S} is a convex set. As S is regular and proper, we have for all $p \in \mathcal{P}$ that $G_{\mathsf{S}}(p) = \mathsf{S}(p, p) \in \mathbb{R}$. In particular, G_{S} is a proper convex function. From (Rockafellar, 1997, Theorem 7.4), $\operatorname{cl} G_{\mathsf{S}}$ is therefore closed and proper, where cl denotes the closure of a function. From (Rockafellar, 1997, Corollary 13.3.1) we now have that $G_{\mathsf{S}}^* = (\operatorname{cl} G_{\mathsf{S}})^*$ is finite on all of $\mathbb{R}^{\mathcal{O}}$ as the domain of $\operatorname{cl} G_{\mathsf{S}}$ is bounded. (In fact, propriety of S is only used to ensure that $\operatorname{cl} G_{\mathsf{S}}$ is a proper convex function; any

[7] See Abernethy, Chen, and Vaughan (2013) for the model and results to which we refer, though the key ideas behind this duality appeared earlier (Hanson, 2003; Chen and Vaughan, 2010).

[8] A convex function is closed if its epigraph is a closed set, or equivalently, if it is lower semi-continuous.

non-proper scoring rule satisfying this very weak assumption would also suffice.) Subdifferentiability of G^*_{S} is then implied by (Rockafellar, 1997, Theorem 23.4).

Conversely, given some mechanism M, a similar argument shows that G^*_{M} is finite on $\mathcal{P}_{\mathsf{M}}$. Applying (Rockafellar, 1997, Theorem 23.5), we have $p \in \partial G_{\mathsf{M}}(t) \iff G_{\mathsf{M}}(t) + G^*_{\mathsf{M}}(p) = p \cdot t$, and as $p \cdot t$ and $G_{\mathsf{M}}(t)$ are both finite, we conclude that $G^*_{\mathsf{M}}(p)$ is finite. The result then follows from the Corollaries.

For (2), for any convex $G : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$, let $\mathcal{S}_G$ and $\mathcal{M}_G$ be the set of proper scoring rules and truthful mechanisms, respectively, with consumer surplus $G : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$. Observe that Φ and Ψ depend on their argument (scoring rule or mechanism) only through its consumer surplus function. In particular, $\Phi(\mathsf{S}) = \Phi(\mathsf{S}')$ if $\mathsf{S}, \mathsf{S}' \in \mathcal{S}_G$ for some $G : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$, and moreover $\Phi(\mathsf{S}) = \mathcal{M}_{G^*}$ by the construction (7). Similarly, $\Psi(\mathsf{M}) = \mathcal{S}_{G^*}$ for any consumer surplus function $G : \mathbb{R}^{\mathcal{O}} \to \overline{\mathbb{R}}$ and any $\mathsf{M} \in \mathcal{M}_G$. From (Rockafellar, 1997, Theorem 12.2), we have $(G^*)^* = G$ for any $G \in \mathcal{C}$, which gives the result. □

To illustrate Theorem 2, consider the log scoring rule $\mathsf{S}(p, o) = \log p(o)$, which is strictly proper. The expected score function is negative Shannon entropy, $G_{\mathsf{S}}(p) = \sum_{o \in \mathcal{O}} p(o) \log p(o)$, which is a closed convex function. In this case, $\Phi(\mathsf{S}) = \{\mathsf{M}\}$ is a unique mechanism $\mathsf{M} = (f, \pi)$ which chooses the allocation probabilities according to the familiar multiplicative weights formula:

$$f(t) = \left(\frac{e^{t(o)}}{\sum_{o' \in \mathcal{O}} e^{t(o')}} \right)_{o \in \mathcal{O}} \in \Delta(\mathcal{O}), \quad \pi(t) = \log \sum_{o \in \mathcal{O}} e^{t(o)} - \frac{\sum_{o \in \mathcal{O}} t(o) e^{t(o)}}{\sum_{o \in \mathcal{O}} e^{t(o)}} \, . \quad (8)$$

The prices π bear resemblance to the Log Market Scoring Rule (LMSR) (Hanson, 2003), but as remarked above, the private information here is a fixed valuation function t, not a belief in the form of a probability distribution. Moreover, we have $\Psi(\mathsf{M}) = \{\mathsf{S}\}$, meaning we recover the log score. Observe that, while $\partial G_{\mathsf{M}}(p)$ is not a singleton since we may shift any subgradient by the all-ones vector, this shift does not change the resulting score, so the choice is effectively unique. (See also Footnote 10.)

The log score example raises a natural question: under what conditions are Φ and Ψ single-valued, and thus bijections in the usual sense? This question essentially boils down to conditions under which the choice of subgradient in eq. (7) is unique, or in other words, when the functions G_{S} and G_{M} are guaranteed to both be differentiable. One suitable set of mechanisms and scoring rules are those whose consumer surplus functions are of *Legendre type*, meaning they are strictly convex and differentiable on the (relative) interior of their effective domain, and the norm of the gradient diverges to ∞ as one approaches the (relative) boundary (Rockafellar, 1997, Theorem 26.3 & Corollary 26.3.1).[9] Moreover, strict convexity translates to the scores and mechanisms being strictly

[9] If a convex function f has an effective domain D with empty interior, f cannot be differentiable. Nonetheless, one can define differentiability of f with respect to the affine span of D: one could say a convex function f is differentiable at a point $x \in \mathrm{relint}(D)$ if the directional derivative of f at x is linear on the subspace $\mathrm{span}(D - \{x\})$. As above, if f

proper and strictly truthful. This Legendre condition is somewhat restrictive, however, as the scoring rules and mechanisms satisfying it must look something like the log score example above. In particular, the mechanism must be defined on an unrestricted type space, as otherwise the conditions would force an infinite allocation or price on the boundary of the type space.

Intuitively, if we remove the gradient divergence condition from Legendre type, and work with convex functions which are strictly convex and differentiable, the only problem that could arise is a non-unique choice of subgradients at the (relative) boundary of the effective domain. Yet, for convex functions which are differentiable on their (relative) interior, they are continuously differentiable, and we may canonically fill in the choices on the (relative) boundary by taking the limit of the gradient. To illustrate this approach, let $\mathcal{O} = \{0, 1\}$ represent being given an item or not, and consider the "linear" mechanism $\mathsf{M} = (f, \pi)$ on restricted type space $\mathcal{T} = [0, 1]$ given by $f(t) = t$, $\pi(t) = t^2/2$. Here f assigns the probability of being allocated the item. Then M is strictly truthful, with $G_{\mathsf{M}}(t) = t^2/2$ for $t \in \mathcal{T}$ and $G_{\mathsf{M}}(t) = \infty$ otherwise. Applying our original construction gives $\Psi(\mathsf{M}) = \{\mathsf{S}_{a,b} \mid a \le 0, b \ge 1\}$ where $\mathsf{S}_{a,b}(p, o) = po - p^2/2$ for $p \in (0, 1)$, and $\mathsf{S}_{a,b}(0, o) = ao$ and $\mathsf{S}_{a,b}(1, o) = bo - 1$. For all of these choices of a and b, we have $G_{\mathsf{S}}(p) = G^*_{\mathsf{M}}(p) = p^2/2$ for $p \in [0, 1]$ and $G_{\mathsf{S}}(p) = \infty$ otherwise. While Ψ does give a range of possible scoring rules, all strictly proper, we may canonically select the one respecting the limit of the gradients of G_{S} on $(0, 1)$, that is, the scaled Brier score $\mathsf{S}_{0,1}$. Returning to mechanisms, our construction would give $\mathcal{T}_{\mathsf{S}} = \mathbb{R}$, since $\partial G_{\mathsf{S}}(0) = (-\infty, 0]$, $\partial G_{\mathsf{S}}(p) = \{p\}$ for $p \in (0, 1)$ and $\partial G_{\mathsf{S}}(1) = [1, \infty)$. In other words, we could arrive at an unrestricted mechanism. Applying the same logic, however, and setting $\mathcal{T}_{\mathsf{S}} = \operatorname{cl} \partial G_{\mathsf{S}}(\,(0, 1)\,) = [0, 1]$, we would arrive at $\Phi(\mathsf{S}_{0,1}) = \{\mathsf{M}\}$. Establishing this kind of bijection for restricted type spaces more broadly is an interesting direction for future work.

The mappings in Theorem 2 are not arbitrary; a mechanism and scoring rule which are related by Φ or Ψ satisfy certain identities. For example, for S and $\mathsf{M} = (f, \pi)$ satisfying part (2) of the theorem, and $\mathsf{M} \in \Phi(\mathsf{S})$, a standard result of convex analysis states that $G_{\mathsf{S}}(p) + G_{\mathsf{M}}(t) = p \cdot t$ whenever either $t = d$ (from eq. (7)) or $p = f(t)$ hold. To provide further intuition for this relationship with a somewhat whimsical example, suppose a gambler in a casino examines the rules of a dice-based game of chance and forms belief p about the probabilities of possible outcomes, assuming the dice are fair. The gambler is then offered a proper scoring rule S to predict the outcome of the game, and reports truthfully. Before the game is played, however, the casino informs the gambler that the dice used need not be fair, and offers the gambler the opportunity to select from among different choices of dice using a truthful mechanism M where the gambler's private information is the valuation function t given by $t(o) = \mathsf{S}(p, o)$;[10]

is differentiable on $\operatorname{relint}(D)$ in this sense, then although $\partial f(x)$ will not be a singleton for $x \in \operatorname{relint}(D)$, the score itself will be unique.

[10]Strictly speaking, to directly apply the convex analysis result referenced above, we would need to take $t(o)$ to be the linear part of $\mathsf{S}(p, o)$; in our setting, however, the allocation and prices of the mechanism are invariant to adding a constant to the valuation of every outcome.

note that $t(o)$ truly is the gambler's valuation of outcome o. If $\mathsf{M} \in \Phi(\mathsf{S})$, and the gambler again reports truthfully, then the dice chosen by the mechanism will be fair. And what will be the gambler's profit in expectation from both the scoring rule and mechanism? Zero. This follows from the above identity, and the observation that $t \cdot f(t) - \pi(t) = \mathsf{S}(p,p) - (p \cdot t - G_{\mathsf{M}}(t)) = G_{\mathsf{S}}(p) + G_{\mathsf{M}}(t) - p \cdot t$. The power of our construction is that these relationships hold regardless of the initial scoring rule S.

4. Affine Scores for Other Elicitation Settings

A number of other application domains do not quite fit into mechanism design or scoring rules, forcing researchers to adapt results to their particular setting. For example, one may wish to elicit several distributions at once, or partially allocate items or rewards. Fortunately, many such settings can be easily expressed in terms of the more general framework of affine scores. We now briefly survey four such domains, and in each show how our main theorem could have directly provided the characterization ultimately used, rather than requiring effort to conceptualize and prove it.

4.1. Decision Rules

A line of work has considered a setting where a decision maker needs to select from a finite set of decisions or actions and so desires to elicit the distribution over outcomes conditional on selecting each alternative (Othman and Sandholm, 2010; Chen and Kash, 2011; Chen et al., 2011). Since only one decision will be made and so only one conditional distribution can be sampled, simply applying a standard proper scoring rule does not result in truthful behavior. Applying Theorem 1 to this setting characterizes what expected scores must be, from which many of the results in these papers follow.

For example, consider the model of proper scoring rules for decision rules (Chen and Kash, 2011), which we now describe:

- The decision maker's goal is to select an action from $\mathcal{A} = \{1, \ldots, n\}$;
- After $i \in \mathcal{A}$ is chosen, an outcome from $\mathcal{O} = \{o_1, \ldots o_m\}$ will be realized, where intuitively the probability of each outcome depends on the action chosen;
- To help decide, the decision maker asks an expert for the probabilities $P_{i,o} = \Pr[o \text{ realized} | i \text{ chosen}]$; we will denote by $\mathcal{P}$ the set of all allowable such probability matrices;

To see this, note that the subgradients of $G_{\mathsf{M}} = G_{\mathsf{S}}^*$ are probability distributions, so the directional derivative is 1 in the direction of $\mathbb{1}$, the all-ones vector; integrating gives the result. See also Abernethy et al. (2013).

- The action is then chosen according to a fixed decision rule $D : \mathcal{P} \to \Delta(\mathcal{A})$ selected in advance by the decision maker, where $D_i(P)$ is the probability of choosing action i given the expert reported the matrix P;
- The decision maker scores the expert's report based on the chosen action i and realized outcome o, according to the function $\mathsf{S}_{i,o} : \mathcal{P} \to \mathbb{R} \cup \{-\infty\}$.

Given a belief P and report Q, we can therefore write the expert's expected score as $V(Q, P) \doteq \sum_{i,o} D_i(Q) P_{i,o} \mathsf{S}_{i,o}(Q)$. The definition of (strict) properness for a particular decision rule then follows naturally: a regular scoring rule S is *proper* for a decision rule D if $V(P, P) \geq V(Q, P)$ for all P and all $Q \neq P$. It is *strictly proper* for the decision rule if the inequality is strict.

A key novelty of this setting is the *type* of the agent: a matrix of conditional probabilities. This is a different object from a distribution over $\mathcal{O}$ or even $\mathcal{O} \times \mathcal{A}$, and thus one cannot use standard scoring rule characterizations. Similarly, the type does not describe the utility of the agent, and hence mechanism design characterizations cannot be applied. For this reason, Chen and Kash (2011) derive a characterization from scratch. Fortunately, the utility of the expert in this setting is an affine score, as the expected score $V(Q, P)$ can be written as a linear combination of the entries of P, and therefore is affine in P. We can therefore immediately apply our theorem to derive Chen and Kash's (2011) characterization, and even extend it to the case where the set $\mathcal{P}$ of probability matrices is not convex. In the theorem statement we make use of the Frobenius inner product $P : Q \doteq \sum_{i,o} P_{i,o} Q_{i,o}$.

Corollary 6. *Given a set of probability matrices $\mathcal{P} \subseteq \Delta(\mathcal{O})^n$ A regular scoring rule is (strictly) proper for a decision rule D if and only if*

$$\mathsf{S}_{i,o}(Q) = \begin{cases} G(Q) - G_Q : Q + \frac{G_{Q,i,o}}{D_i(Q)} & D_i(Q) > 0 \\ \Pi_{i,o}(Q) & D_i(Q) = 0 \end{cases}$$

where $G : \mathsf{Conv}(\mathcal{P}) \to \mathbb{R} \cup \{-\infty\}$ is a (strictly) convex function, G_Q is a subgradient of G at the point Q with $G_{Q,i,o} = 0$ when $D_i(Q) = 0$, and $\Pi_{i,o} : \mathcal{P} \to \mathbb{R} \cup \{-\infty\}$ is an arbitrary function that can take a value of $-\infty$ only when $Q_{i,o} = 0$.

Proof. By Theorem 1, S is (strictly) proper for D if and only if there exists a (strictly) convex G such that $V(Q, P) = G(Q) + dG_Q(P - Q)$. That is,

$$\sum_{i,o} D_i(Q) P_{i,o} \mathsf{S}_{i,o}(Q) = G(Q) - G_Q : Q + \sum_{i,o} G_{Q,i,o} P_{i,o} \ ,$$

or for all i such that $D_i(Q) \neq 0$,

$$\mathsf{S}_{i,o}(Q) = G(Q) - G_Q : Q + \frac{G_{Q,i,o}}{D_i(Q)} \ .$$

When $D_i(Q) = 0$, S is unconstrained (other than the minimal requirements regarding $-\infty$ for regularity). However, note that our affine score is restricted in that, because $D_i(Q)$ is fixed, some choices in $\mathcal{A}$ are not possible to select as subgradients. In particular, it must be that $G_{Q,i,o} = 0$ when $D_i(Q) = 0$. $\square$

4.2. Proper Losses for Partial Labels

Several variants of proper losses have appeared in the machine learning literature, one of which is the problem of estimating the probability distribution of labels for a new data point when the training data may contain several *noisy* labels, possibly not even including the correct label. (This is frequently the case, for example, when using crowdsourced labels for training data.) More formally, one wishes to estimate $p \in \Delta_n$ where the true label $y \in \{1, \dots, n\}$ is drawn from p. However, instead of observing a sample $y \sim p$ and designing a proper loss $\ell(\hat{p}, y)$, one instead only observes some noisy set of labels $S \subseteq \{1, \dots, n\}$. Hence, the task is to design a loss $\ell(\hat{p}, S)$ which when minimized over one's data yields accurate estimates of the true p.

Recently this problem was studied by Cid-Sueiro (2012) under the assumption that the labels in S are drawn i.i.d. from distribution $q = Mp$ for some known $M \in \mathbb{R}^{2^n \times n}$, where p is the true label distribution. That is, if the actual label is drawn from p, the noisy set of labels is drawn from Mp (using some indexing of the sets, say lexographical). Formally, a loss in this setting takes the form $\ell : \Delta_n \times 2^{\{1,\dots,n\}} \to \mathbb{R}$, and is proper if the expected score $\mathbb{E}_{S \sim Mp}[\ell(\hat{p}, S)] = \ell(\hat{p}, \cdot)^\top Mp$ is minimized at $\hat{p} = p$. Cid-Sueiro provides a characterization (his Theorem 4.3) of all proper losses for an even more general version of this setting where M is not known exactly but assumed to be a member of some known class; the loss should be proper for any M in this class. As the (negative) expected loss is linear in the underlying distribution p, our Theorem 1 applies and allows us to recover his characterization result. This construction holds more generally for latent outcome settings; any observable (here a set of labels) whose distribution has an affine relationship with the latent outcome would suffice to apply our theorem.

Rather than introducing the full, general model used by Cid-Sueiro, we show how our theorem applies to yield a characterization for a single, fixed M. This is a special case of his Theorem 4.3, generalized to allow restricted sets of probability distributions $\mathcal{P}$.

Corollary 7. *Let number of labels n, matrix $M \in \mathbb{R}^{2^n \times n}$, and $\mathcal{P} \subseteq \Delta_n$ be given such that $M\mathcal{P}$ has full dimension in the column space of M. Let outcome set $\mathcal{O} = 2^{\{1,\dots,n\}}$ be the power set of $\{1, \dots, n\}$. A regular score $\mathsf{S} : \mathcal{P} \times \mathcal{O} \to \overline{\mathbb{R}}$ is proper if and only if there exists a convex function $G : \mathsf{Conv}(\mathcal{P}) \to \mathbb{R}$ such that*

$$\ell(\hat{p}, S) = -G(\hat{p}) - G_{\hat{p}}^\top (M^+ e_S) + G_{\hat{p}}^\top \hat{p}, \tag{9}$$

where e_S is a one-hot encoding (indicator vector) of S, M^+ is a left inverse of M, and G_p is a subgradient of G for all $p \in \mathcal{P}$.

Proof. For the form given, the expected loss is

$$\begin{aligned}\mathbb{E}_{S \sim Mp}[\ell(\hat{p}, S)] &= \mathbb{E}_{S \sim MP}[-G(\hat{p}) - G_{\hat{p}}^\top (M^+ e_S) + G_{\hat{p}}^\top \hat{p}] \\ &= -G(\hat{p}) - G_{\hat{p}}^\top M^+ Mp + G_{\hat{p}}^\top \hat{p} \\ &= -G(\hat{p}) - G_{\hat{p}} \cdot (p - \hat{p}) \ ,\end{aligned}$$

which is linear in p. Propriety of the given form then follows immediately from Theorem 1. For the converse, let $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ be a given truthful affine score for $\mathcal{T} = \mathcal{P}$ and score set $\mathcal{A} = \{p \mapsto \mathbb{E}_{S \sim Mp} f(S) \mid f : 2^{\{1,\ldots,m\}} \to \mathbb{R}\}$. From the theorem, $\mathsf{A}(p, \cdot) = G(p) + d_p(\,\cdot\, - p) \in \mathcal{A}$. Thus we can write $\ell(\hat{p}, \cdot)^\top Mp = \mathbb{E}_{S \sim Mp}[\ell(\hat{p}, S)] = -G(\hat{p}) - G_{\hat{p}} \cdot (p - \hat{p})$. Taking $\hat{\ell}(\hat{p}, \cdot) = \ell(\hat{p}, \cdot) + (G(\hat{p}) - G_{\hat{p}} \cdot \hat{p})\mathbb{1}$, this means $\hat{\ell}(\hat{p}, \cdot)^\top Mp = -G_{\hat{p}} \cdot p$ for all $p \in \mathcal{P}$. By our assumption that $M\mathcal{P}$ has full dimension in the column space of M, $\hat{\ell}(\hat{p}, \cdot)^\top = -M^+ G_{\hat{p}}$ for some left inverse M^+, showing ℓ is of the desired form. (If $M\mathcal{P}$ does not have full dimension there may be additional choices of ℓ with the correct expected value on $\mathcal{P}$, but the proof otherwise applies.) $\square$

4.3. Mechanism Design with Partial Allocation

Several mechanism design settings considered in the literature have some form of *exogenous* randomization, in that "nature" chooses some outcome ω according to some (often unknown) distribution, and this distribution may depend on the allocation chosen by the mechanism. Examples include sponsored search auctions (Feldman and Muthukrishnan, 2008), multi-armed bandit mechanisms (Babaioff et al., 2009), and recent work on daily deals (Cai et al., 2013). The work of Cai, Mahdian, Mehta, and Waggoner (2013) introduces a very general model for such settings, which will be our focus in this subsection.

In the setting of Cai, et al., the mechanism designer wants to elicit two pieces of information: the agent's (expected) value for an item in an auction and the probability distribution of a random variable conditional on that agent winning. Their goal is to understand how the organizer of a daily deal site can take into account the value that will be created for users, as opposed to just the advertiser (the agent), when a particular deal is chosen to be advertised. For example, the site operator may prefer deals that sell to many users over equally profitable deals that sell only to a few because this keeps users interested for future days. The authors characterize the possible implementable ways of quantifying user welfare as a function of the agent's probabilistic belief as to the outcome (of nature) resulting from each allocation. We will show how to recover this characterization as a special case of a more general setting in which a mechanism designer wishes to elicit two pieces of information, but the second need not be restricted to probability distributions.

We begin with a description of the setting of Cai et al. (2013). Let $\mathcal{O}$ be a set of outcomes, and for each outcome o and each agent i, let $\Omega^{i,o}$ be some set of events. For example, o could determine which agent wins an auction for the opportunity to advertise a special offer from its business and $\Omega^{i,o}$ could represent the set of numbers of customers that may purchase the deal. Agents each have a valuation function $v^i : \mathcal{O} \to \mathbb{R}$ and a set of beliefs $p^{i,o} \in \Delta(\Omega^{i,o})$ for each allocation $o \in \mathcal{O}$, e.g., an expected value for getting to advertise and a probability distribution over the number of customers who accept the deal. The mechanism aggregates all of this information into a single outcome o, and additionally choses some payoff function $s^i : \Omega^{i,o} \to \mathbb{R}$, so that the final utility of agent i is $v^i(o) + \mathbb{E}_{p^{i,o}}[s^i]$; that is, the winning agent both gets to advertise and accepts a scoring rule contract regarding its prediction of the number of customers. A mechanism is truthful if for all values of v and p for the other agents, agent i maximizes her total utility by reporting v^i and $p^i \doteq (p^{i,o})_{o\in\mathcal{O}}$ truthfully. For additional examples, the standard sponsored search setting has $\Omega^{i,o} = \{\text{click}, \text{no click}\}$ for o such that i is allocated a slot, and the probabilities $p^{i,o}$ are assumed to be public knowledge. Moreover, the decision rules framework discussed above is a single-agent special case with $v \equiv 0$ and $\Omega^o = \Omega^{o'} = \Omega$ for all $o \in \mathcal{O}$; of course, unlike the interpretation above, in this setting $o \in \mathcal{O}$ is the allocation/decision while Ω is the set of outcomes.

Motivated by incorporating the utilities of the end consumers in a daily deal setting, Cai et al. (2013) ask when one can implement an allocation rule of the form $f(v,p) = \operatorname{argmax}_{o\in\mathcal{O}} v(o) + g^o(p^o)$, which they interpret as maximizing welfare of the winner plus a term that captures something about the welfare of consumers. In other words, when does there exist some choice of payment making f truthful. The authors conclude that this can be done if and only if g^o is convex for each $o \in \mathcal{O}$. In what follows, we will recover this result using our affine score framework.

We first observe that this model can easily be cast as an affine score, as follows. For simplicity, we fix some agent i and focus on the single-agent case; as discussed several times above, this is essentially without loss of generality. The type is simply the combined private information of the agent, with type space

$$\mathcal{T} \subseteq \left\{ (v,p) : v \in \mathbb{R}^{\mathcal{O}},\ p \in \prod_{o\in\mathcal{O}} \Delta(\Omega^{i,o}) \right\}. \tag{10}$$

By assumption, the utility of the agent upon allocation o and payoff s is simply $v(o) + \mathbb{E}_{p^o}[s]$, which is linear in the type $t = (v,p)$ and therefore affine. Thus the net payoff $\mathsf{A}(t',t) = v(o(t')) + \mathbb{E}_{p^{o(t')}}[s(t')]$ is an affine score, with score set $\mathcal{A} = \{(v,p) \mapsto v(o) + \mathbb{E}_{p^o}[s] : o \in \mathcal{O}, s \in \mathbb{R}^{\Omega^{i,o}}\}$, where again we have fixed i.

To answer the implementability question of Cai et al. (2013), we consider a general type space of the form $\mathcal{T} \subseteq \mathcal{V} = \mathcal{V}^X \times \mathcal{V}^Y$ for linear subspaces $\mathcal{V}^X$ and $\mathcal{V}^Y$. In the daily deal setting, we would have $\mathcal{V}^X = \mathbb{R}^{\mathcal{O}}$ and $\mathcal{V}^Y = \prod_{o\in\mathcal{O}} \mathbb{R}^{\Omega^{i,o}}$, and $\mathcal{T} = \mathcal{V}^X \times \mathcal{T}^Y$ where $\mathcal{T}^Y = \prod_{o\in\mathcal{O}} \Delta(\Omega^{i,o}) \subseteq \mathcal{V}^Y$.

We wish to know when a function $f^X : \mathcal{T} \to \mathsf{Lin}(\mathcal{V}^X \to \mathbb{R})$ is *extendable* with respect to $\mathcal{A} \subseteq \mathsf{Aff}(\mathcal{V} \to \mathbb{R})$, in the sense that there exists some truthful affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ with score set $\mathcal{A}$, and some $f^Y : \mathcal{T} \to \mathsf{Aff}(\mathcal{V}^Y \to \mathbb{R})$ such that $\mathsf{A}(t',t) = f^X(t')(t^X) + f^Y(t')(t^Y)$, where $t = (t^X, t^Y)$. In the daily deals context, we have $t^X = v$ and $t^Y = p$. Here f^X selects the outcome o, formally represented as the linear map $E_o : v \mapsto v(o)$, and f^Y is an expected score of the form $\mathbb{E}_{\omega\sim p^o}[s]$, which is affine (here linear) in p. Thus, asking whether f^X is extendable is equivalent to asking which rules for selecting the outcome o are implementable.

To answer this question, let us introduce some notation. For each $a \in \mathcal{A}$ we write $X(a) \in \mathsf{Lin}(\mathcal{V}^X \to \mathbb{R})$ to be the linear part of a on $\mathcal{V}^X$, and $Y(a)$ to be the *affine* part of a on $\mathcal{V}^Y$. For any $\mathcal{A}' \subseteq \mathcal{A}$ we will write $X(\mathcal{A}') \doteq \{X(a) : a \in \mathcal{A}'\}$.

In this very general framework, we can show the following.

Theorem 3 (Informal)**.** *Partial allocation rule f^X is extendable with respect to $\mathcal{A}$ if and only if there exists $\mathcal{A}' \subseteq \mathcal{A}$ such that, for all $t \in \mathcal{T}$,*

$$f^X(t) \in \operatorname*{argsup}_{x \in X(\mathcal{A}')} \left\{ x(t^X) + \sup_{\substack{a \in \mathcal{A}' \\ X(a)=x}} \left\{ Y(a)(t^Y) \right\} \right\} . \tag{11}$$

One direction follows from the fact that an affine score is truthful if and only if $\mathsf{A}(t) \in \operatorname{argsup}\{a(t) : a \in \mathcal{A}'\}$ where $\mathcal{A}' = \{A(t,\cdot) \mid t \in \mathcal{T}\}$; we simply take the supremum first over $X(\mathcal{A})$ and then over the rest. For the other direction, for any $\mathcal{A}' \subseteq \mathcal{A}$, taking $\mathsf{A}(t', t) = f(t')(t^X) + y(t')(t^Y)$ where y is in the argsup of the supremum of eq. (11), provided one exists, gives a truthful affine score.

Returning to the special case of daily deals, let us denote by $a_{o,s} \in \mathcal{A}$ the function $(v, p) \mapsto v(o) + \mathbb{E}_{p^o}[s]$. We now see that f is implementable if and only if it satisfies

$$f(v,p) \in \operatorname*{argsup}_{o \in \mathcal{O}} \left\{ v(o) + \sup_{s \,:\, a_{o,s} \in \mathcal{A}'} \{\mathbb{E}_{p^o}[s]\} \right\}, \tag{12}$$

for some $\mathcal{A}' \subseteq \mathcal{A}$. Thus, letting $g^o(p^o) = \sup\{\mathbb{E}_{p^o}[s] : a_{o,s} \in \mathcal{A}'\}$, we see that g^o is convex as the supremum of affine functions. Moreover, given any collection of convex functions $\{g^o\}_{o \in \mathcal{O}}$, where $g^o : \Delta(\Omega^{i,o}) \to \mathbb{R}$, we can define $S^o \doteq \{\omega \mapsto g^o(p) + dg^o(\mathbb{1}_\omega - p) : p \in \operatorname{dom}(g^o)\}$ and $\mathcal{A}' \doteq \{a_{o,s} : o \in \mathcal{O},\, s \in S^o\} \subseteq \mathcal{A}$, thus recovering each g^o in the above expression. It then only remains to show that no other nonconvex function can serve in the argsup; for this one may appeal to the argument of Cai et al. (2013) which observes that the indifference points between different allocations are fixed, thus determining the function in the argsup up to a constant.

A special case of the above, but closer to classical mechanism design, is captured in the following scenario. The mechanism designer has two distinct sets of goods to allocate and wants to design a truthful mechanism that is consistent with a partial allocation rule that determines how the primary goods should be allocated given the agent's preferences over both types of goods. We can capture this setting with the score set $\mathcal{A} = \{a_{o_1,o_2,p} : t \mapsto t_1(o_1) + t_2(o_2) + p \mid o_1 \in \mathcal{O}_1, o_2 \in \mathcal{O}_2, p \in \mathbb{R}\}$. Such mechanisms are characterized by the following informal theorem.

Theorem 4 (Informal)**.** *Consider an agent with type $t = (t_1, t_2)$. A truthful affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ with score set $\mathcal{A}$ (above) is consistent with a partial allocation rule $f : \mathcal{T} \to \mathcal{O}_1$, if and only if there exists $\mathcal{A}' \subseteq \mathcal{A}$ such that*

$$f(t) \in \operatorname*{argsup}_{o_1' \in \mathcal{O}_1} \left\{ t_1(o_1') + \sup_{\substack{a_{o_1,o_2,p} \in \mathcal{A}' \\ o_1' = o_1}} \{t_2(o_2) + p\} \right\} . \tag{13}$$

In particular, analogous to the daily deals setting, the mechanism designer is restricted to mechanisms that make decisions based on a convex function of t_2 (the inner supremum is a pointwise supremum over affine functions and thus convex).

We conclude by noting similarity to work of Chambers and Lambert (2014), where a center wishes to elicit an agent's belief about some future event, and to do so multiple times as the event gets closer. The solution proposed is essentially to have the agent choose a menu of scoring functions at time 0 from a menu of menus, and then at time 1 choose a score from the menu chosen at time 0. Both the present setting and theirs share this sense of "menu of menus", however one can check that the two are incomparable. In particular, the relationship between the final score and the beliefs of the agent can be nonlinear in the Chambers–Lambert model.

4.4. Responsive Lotteries

Feige and Tennenholtz (2010) study the problem of how an agent can be incentivized to indirectly reveal their utility function over outcomes by being given a choice of lotteries over those outcomes, an approach with applications to experimental psychology, market research, and multiagent mechanism design. Formally, a lottery rule is a function $f : \mathbb{R}^n \to \Delta_n$, i.e., from utility vectors over n outcomes to probability distributions (lotteries) over those outcomes. The authors give several examples of effective lottery rules under the assumption of risk neutrality. In contrast, our approach allows us to give a complete characterization, which highlights the relationship between natural desiderata and underlying geometric features of the set of possible lotteries, as we now describe.

The authors ask when a lottery rule is *truthful dominant*, which is defined as having the following three features: *incentive compatibility*, meaning the true utility vector is among the optimal reports, *rational uniqueness*, meaning the optimal lottery for a given utility is unique, and *rational invertibility*, meaning every report can be optimal for at most one utility. We would like to relate these notions to simple geometric features, which we introduce now informally, and formalize later in Definition 6. A convex set K is *strictly convex* if no point on its boundary can be expressed as a convex combination of other points in K, and K is *smooth* if each point on its boundary has a unique unit normal vector. Using the results to follow, we can show that strict truthfulness and continuity of the lottery rule jointly correspond to strict convexity of the lottery set, and uniqueness of the utility given the optimal lottery corresponds to smoothness of the boundary.

Corollary 8. *A lottery rule f satisfies incentive compatibility and rational uniqueness if and only if $\{f(x)\} = \operatorname{argmax}_{p \in K} \langle x, p \rangle$ for $K \subset \Delta_n$ compact and strictly convex relative to Δ_n. Moreover, f additionally satisfies rational invertibility (and thus is truthful dominant) if and only if K is additionally smooth.*

Proof. In this setting, we can see that utility vectors are equivalent up to positive-affine transformations. (Since there are no payments, multiplying the utility of each outcome by a positive constant or adding a constant to the utility for each outcome has no effect on the optimal lottery for an agent.) Thus, we may linearly project the utilities and probability simplex onto the set $V = \{x \in \mathbb{R}^n : \sum_i x_i = 0\}$, which only changes the expected utilities by a constant. We then write these vectors in a basis for $V \cong \mathbb{R}^{n-1}$, normalize the utilities (only scaling them) to the unit sphere in V, and apply Theorem 5. $\square$

We now turn to the geometric statements needed to establish Corollary 8, beginning with some formal definitions. We denote by ∂K the boundary of the set $K \subseteq \mathbb{R}^n$.

Definition 5. *Given a compact convex set $K \subset \mathbb{R}^n$, we define the* exposed face *$F_K(t)$ in direction $t \neq 0$ and the* normal cone *$N_K(k)$ at point $k \in \partial K$ by*

$$F_K(t) = \operatorname*{argmax}_{k \in K} \langle t, k \rangle \,, \qquad N_K(k) = \{t \in \mathbb{R}^n : k \in F_K(t)\}. \tag{14}$$

An exposed face $F_K(t)$ is simply the set of points as far in direction t as possible; for example, on a triangle ABC the exposed faces are vertices A,B,C and the edges $\overline{AB}$, $\overline{BC}$, $\overline{CA}$. The normal cone $N_K(k)$ is simply the set of all valid normal vectors to K at a point k on the boundary of K; for the triangle, the normal cone at a point on an edge is simply a ray perpendicular to it, but on a vertex there is a closed cone spanning the exterior angle of the triangle.

Definition 6. *We say K is* strictly convex *if $F_K(t)$ is a singleton for all $t \neq 0$. Dually, we say K is* smooth *if $N_K(k)$ is a ray (i.e. $\{\alpha t : \alpha \geq 0\}$ for some $t \neq 0$) for all $k \in \partial K$.*

To illustrate: a disc is smooth and strictly convex, a triangle is neither, a rounded rectangle is smooth but not strictly convex, and the intersection of two discs is strictly convex but not smooth. We now establish the tight connection between these geometric concepts and the truthfulness conditions in this setting, which will imply Corollary 8. Throughout, we identify linear functions $h \in \mathsf{Lin}(\mathbb{R}^n \to \mathbb{R})$ with vectors $v \in \mathbb{R}^n$ such that $h(\cdot) = \langle v, \cdot \rangle$.

Theorem 5. *Let $\mathcal{T} = \{t \in \mathbb{R}^n : \|t\|_2 = 1\}$ be the unit sphere in $\mathbb{R}^n$, and let a truthful affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ be given, with score set $\mathcal{A} = \mathsf{Lin}(\mathbb{R}^n \to \mathbb{R}) \cong \mathbb{R}^n$. Let $S : \mathcal{T} \to \mathbb{R}^n, t \mapsto \mathsf{A}(t, \cdot)$. Then S is continuous and A is strictly truthful, if and only if $S(\mathcal{T})$ is the boundary of a compact and strictly convex set $K \subset \mathbb{R}^n$. S is additionally injective if and only if K is additionally smooth.*

Proof. Define $\mathcal{A}' \doteq S(\mathcal{T})$. We begin with the first part of the theorem. Let K be compact and strictly convex, and $\mathcal{A}' = \partial K$. Then as A is truthful, we must have $S(t) \in \operatorname{argsup}_{a \in \mathcal{A}'} \langle t, a \rangle$. As $\mathcal{A}' = \partial K$, we may also write $S(t) \in \operatorname{argmax}_{k \in K} \langle t, k \rangle$. Now by strict convexity of K, we have for every $a \in \mathcal{A}' = \partial K$, there exists some $t \in \mathcal{T}$ such that $\{a\} = \operatorname{argmax}_{k \in K} \langle t, k \rangle$. We conclude $S(t) = a$, giving strict truthfulness. Continuity follows immediately from Berge's Maximum Theorem (Ok, 2007).

For the converse, let A be strictly truthful and S continuous. By standard arguments, since $\mathcal{T}$ is a compact subset of $\mathbb{R}^n$, we have $\mathcal{A}' = S(\mathcal{T})$ is compact as a continuous image of a compact set. Thus, $K \doteq \mathsf{Conv}(\mathcal{A}')$ is a compact convex set. Letting $F_K(t) \doteq \mathrm{argmax}_{k\in K} \langle t, k\rangle$ be the exposed face of K in direction t, we will now show $F_K(t) = \{S(t)\}$. First, observe that the extreme points of K, $\mathrm{ext}(K)$, are a subset of $\mathcal{A}'$; otherwise we have $k \in \mathrm{ext}(K) \setminus \mathcal{A}'$, so $K \setminus \{k\}$ is a convex set containing $\mathcal{A}'$, contradicting the definition of $K = \mathsf{Conv}(\mathcal{A}')$. Now we may apply (Urruty and Lemaréchal, 2001, Proposition A.2.4.6) to express the argmax in terms of the extreme points of K, giving us

$$F_K(t) \doteq \operatorname*{argmax}_{k\in K} \langle t, k\rangle = \mathsf{Conv}\left(\operatorname*{argmax}_{k\in \mathrm{ext}(K)} \langle t, k\rangle\right) \subseteq \mathsf{Conv}\left(\operatorname*{argmax}_{a\in \mathcal{A}'} \langle t, a\rangle\right) = \{S(t)\},$$

where the last equality uses strict truthfulness. As K is compact, $F_K(t)$ is nonempty. Thus $F_K(t) = \{S(t)\}$, and we conclude $S(t) \in \mathrm{ext}(K)$. Hence $\mathcal{A}' = S(\mathcal{T}) \subseteq \mathrm{ext}(K)$. Together with the reverse inclusion above, we have $\mathcal{A}' = \mathrm{ext}(K)$. We now apply (Urruty and Lemaréchal, 2001, Proposition C.3.1.5) to obtain $\partial K = \bigcup_{t\in\mathcal{T}} F_K(t) = \bigcup_{t\in\mathcal{T}}\{S(t)\} = \mathcal{A}'$. Finally, as $\mathrm{ext}(K) = \mathcal{A}' = \partial K$, we have strict convexity of K.

For the final statement of the theorem, we note that by (Urruty and Lemaréchal, 2001, Proposition C.3.1.4), we have $k \in F_K(t) \iff t \in N_K(k)$. By the above, we already have $F_K(t) = \{S(t)\}$ for all $t \in \mathcal{T}$, which implies $N_K(k) \cap \mathcal{T} = \{t : S(t) = k\}$. Hence, $N_K(a)$ is a ray for all $a \in \mathcal{A}'$ if and only if S is injective. $\square$

5. Extension to Finite-Valued Properties

We wish to generalize the notion of truthful elicitation from eliciting private information from some set $\mathcal{T}$ to accept reports from a space $\mathcal{R}$ which may be different from $\mathcal{T}$. To even discuss truthfulness in this setting, we need a notion of a truthful report r for a given type t. Drawing on the scoring rules literature, we encapsulate this notion by a general multivalued map which specifies all (and only) the correct values for t.

Definition 7. *Let $\mathcal{T}$ be a given type space, where $\mathcal{T} \subseteq \mathcal{V}$ for some vector space $\mathcal{V}$ over $\mathbb{R}$, and $\mathcal{R}$ be some given report space. A* property *is a multivalued map $\Gamma : \mathcal{T} \rightrightarrows \mathcal{R}$ which associates a nonempty set of correct report values to each type. We let $\Gamma_r \doteq \{t \in \mathcal{T} \,|\, r \in \Gamma(t)\}$ denote the set of types t corresponding to report value r.*

One can think of Γ_r as the "level set" of Γ corresponding to value r, a concept which is especially useful for finite $\mathcal{R}$, the case we focus on in this section.

We extend the notion of an affine score to properties, where the report space is $\mathcal{R}$ instead of $\mathcal{T}$ itself, and we require $\mathsf{A}(r, \cdot) \in \mathsf{Aff}(\mathcal{V} \to \overline{\mathbb{R}})$ for all $r \in \mathcal{R}$.

Definition 8. *An affine score $\mathsf{A} : \mathcal{R} \times \mathcal{T} \to \overline{\mathbb{R}}$* elicits *a property $\Gamma : \mathcal{T} \rightrightarrows \mathcal{R}$ if for all t,*

$$\Gamma(t) = \operatorname*{argsup}_{r\in\mathcal{R}} \mathsf{A}(r, t). \tag{15}$$

If we merely have $\Gamma(t) \subseteq \mathrm{argsup}_{r\in\mathcal{R}} \mathsf{A}(r,t)$*, we say* A weakly elicits Γ*. A property* $\Gamma : \mathcal{T} \rightrightarrows \mathcal{R}$ *is* elicitable *if there exists some affine score* $\mathsf{A} : \mathcal{R} \times \mathcal{T} \to \overline{\mathbb{R}}$ *eliciting* Γ*.*

Note that it is certainly possible to write down A such the argsup in (15) is not well defined. This corresponds to some types not having an optimal report, which we view as violating a minimal requirement for a sensible affine score. Thus, in order for A to be an affine score, we require (15) to be well defined for all $t \in \mathcal{T}$.

In the remainder of this section, we examine the special case of *finite properties*, where $\mathcal{R}$ is a finite set of reports. We treat the general case in Frongillo and Kash (2019), and here leverage additional structure of finite report sets to provide stronger characterizations. In the scoring rules literature, Lambert and Shoham (2009) view finite properties as eliciting answers to multiple-choice questions. There are also applications to mechanism design, discussed in Section 5.1. Assume throughout that $\mathcal{R}$ is finite and that $\mathcal{T}$ is a convex subset of $\mathcal{V} = \mathbb{R}^d$. In this setting, we will use the concept of a power diagram from computational geometry.

Definition 9. *Given a set of points* $P = \{p_i\}_{i=1}^m \subset \mathcal{V}$*, called* sites*, and weights* $w \in \mathbb{R}^m$*, a* power diagram $D(P,w)$ *is a collection of cells* $\mathrm{cell}(p_i) \subseteq \mathcal{T}$ *defined by*

$$\mathrm{cell}_{P,w}(p_i) = \left\{ t \in \mathcal{T} \;\middle|\; i \in \operatorname*{argmin}_j \left\{ \|p_j - t\|^2 - w_j \right\} \right\}. \tag{16}$$

The following result is a straightforward generalization of Theorem 4.1 of Lambert and Shoham (2009), and is essentially a restatement of results due to Aurenhammer (1987a; 1987b).

Theorem 6. *A property* $\Gamma : \mathcal{T} \rightrightarrows \mathcal{R}$ *for finite* $\mathcal{R}$ *is elicitable if and only if the level sets* $\{\Gamma_r\}_{r\in\mathcal{R}}$ *form a power diagram* $D(P,w)$*.*

Proof. Let us examine the condition that t is an element of $\mathrm{cell}_{P,w}(p_i)$ for some power diagram $D(P,w)$:

$$\begin{aligned} t \in \mathrm{cell}_{P,w}(p_i) &\iff i \in \operatorname*{argmin}_j \left\{ \|p_j - t\|^2 - w_j \right\} \\ &\iff i \in \operatorname*{argmin}_j \left\{ \|p_j\|^2 - 2\langle p_j, t\rangle - w_j \right\}. \end{aligned} \tag{17}$$

Note that eq. (17) is affine in t. Now given some $D = D(P,w)$ with index set $\mathcal{R}$, we simply let $\mathsf{A}(r,t) = 2\langle p_r, t\rangle + w_r - \|p_r\|^2$. By (17) we immediately have $r \in \mathrm{argsup}_{r'} \mathsf{A}(r',t) \iff t \in \mathrm{cell}_{P,w}(p_r)$, as desired.

Conversely, let an affine score A eliciting Γ be given. As we are in finite dimensions, we may write $\mathsf{A}(r,t) = \langle x_r, t\rangle + c_r$ for $x_r \in \mathcal{V}$ and $c_r \in \mathbb{R}$. Letting $p_r = x_r/2$ and $w_r = \|p_r\|^2 + c_r$, we see by (17) again that $\Gamma_r = \mathrm{cell}(p_r)$ of the diagram $D(\{p_r\}, w)$. Hence, Γ is a power diagram. $\square$

We have now seen what kinds of finite-valued properties are elicitable, but how can we elicit them? More precisely, as the proof above gives sufficient conditions, what are all ways of eliciting a given power-diagram? In general, it is difficult to provide a "closed form" answer to this question, so we restrict to the *simple* case, where essentially the cells of a power diagram are as constrained as possible.

Definition 10 (Aurenhammer (1987c))**.** *A* j-polyhedron *is the intersection of dimension* j *of a finite number of closed halfspaces of* $\mathcal{V} = \mathbb{R}^d$*, where* $0 \le j \le d$*. A* tiling C *in* $\mathcal{V}$ *is a covering of* $\mathcal{V}$ *by finitely many* j*-polyhedra, called* j-faces *of* C*, whose (relative) interiors are disjoint. If furthermore their non-empty intersections are faces of* C *then* C *is a* cell complex*. A cell complex* C *is called* simple *if each of its* j*-faces is in the closure of exactly* $(d-j+1)$ d*-faces (cells).*

With this definition in hand, we can now characterize the ways to elicit *simple* properties, those whose level sets form a simple cell complex.

Theorem 7. *Let* $\mathcal{V} = \mathbb{R}^d$ *and let finite-valued, elicitable, simple property* $\Gamma : \mathcal{T} \rightrightarrows \mathcal{R}$ *be given. Then there exist points* $\{p_r\}_{\mathcal{R}} \subseteq \mathcal{V}$ *such that the following holds: for any affine score* $\mathsf{A} : \mathcal{R} \times \mathcal{T} \to \overline{\mathbb{R}}$ *eliciting* Γ*, there exist* $\alpha > 0$*,* $p_0 \in \mathcal{V}$*, and* $w \in \mathbb{R}^{\mathcal{R}}$ *such that*

$$\mathsf{A}(r,t) = 2\,\langle \alpha p_r + p_0, t\rangle - \|\alpha p_r + p_0\|^2 + w_r \ , \tag{18}$$

and conversely, for all such α *and* p_0 *there exists* $w \in \mathbb{R}^{\mathcal{R}}$ *making* A *in eq.* (18) *elicit* Γ*.*

Proof. A result of Aurenhammer for simple cell complexes, given in Lemma 1 of Aurenhammer (1987b) and the proof of Lemma 4 of Aurenhammer (1987a), states the following: given sites P and P' and weights w, there exist weights w' such that $D(P', w') = D(P, w)$ if and only if P' is a homothet (translated and positively scaled copy) of P. We simply apply this fact to the proof of Theorem 6. □

Detecting elicitable finite properties. As a practical matter, it is natural to ask if we can efficiently determine whether a given finite-valued property Γ is elicitable. By Theorem 6, we need only test whether the cells $C = \{\Gamma_r\}_{r\in\mathcal{R}}$ form a power diagram. For the *simple* case, Aurenhammer gives the "Orthogonal Dual" algorithm for this task; see § 2.2 of Aurenhammer (1987c) and comments thereafter. The orthogonal dual algorithm assumes that the cells are stored in an *incidence lattice*, with nodes for each face of C, and edges when faces are incident (a j-dimensional face which contains a $(j-1)$-dimensional face). The runtime of the algorithm is $O(m)$, where m is the number of facets (faces of dimension $d-1$). More generally, Borgwardt and Frongillo (2019) present a weakly polynomial-time algorithm to detect power diagrams in the general case, via a simple linear program.

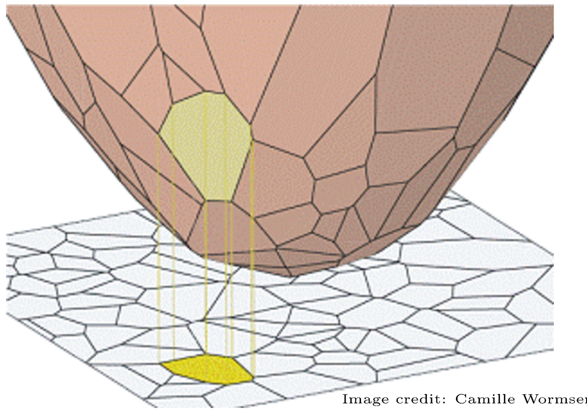

Figure 2: A consumer surplus function G and its corresponding partition of the type space, Γ. The proof of Theorem 6 leverages the fundamental relationship between projections of convex functions and power diagrams.

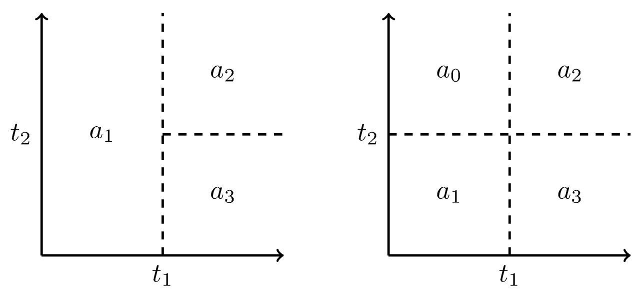

Figure 3: An allocation rule which cannot be implemented, for any distinct choices of a_i (left), and a rule which could be implemented for appropriate choices of a_i (right).

5.1. Finite Properties in Mechanism Design

Mechanisms with a finite set of allocations are common. Carroll (2012) examines them and observes they give rise to polyhedral typespaces. Theorem 6 strengthens this characterization to power diagrams, which rules out polyhedral examples such as the one shown in Figure 3. In particular, the example of the left of the figure fails to be a power diagram because all power diagrams in $\mathbb{R}^d$ are cell complexes (Aurenhammer, 1987a), whereas the intersection of the a_1 and a_2 cells is not a face of the a_1 cell, making it merely a tiling.

Suppose we are in a such a mechanism design setting with a finite set of allocations $\mathcal{X}$ and we have picked an allocation rule a. Under what circumstances is a implementable, i.e., when is there a payment rule that makes the resulting mechanism truthful? For convex type sets, Saks and Yu (2005) showed that the following condition is necessary and sufficient.

Definition 11. *Allocation rule a satisfies* weak monotonicity *(WMON) if $a(t) \cdot (t' - t) \leq a(t') \cdot (t' - t)$ for all $t, t' \in \mathcal{T}$.*

From Theorem 1, we know that a being implementable means that there exists a convex function G such that a is a selection of its subgradients. From the characterization of finite properties, however, we also know that a must encode the sites of a power diagram. This gives us a new proof of Saks–Yu by showing that WMON characterizes power diagrams. In particular, we can leverage the following characterization of power diagrams due to Aurenhammer (1987a). This result assumes that a is defined on all of $\mathcal{V} = \mathbb{R}^d$, as until recently power diagrams have not been studied on restricted domains. However, recent results for restricted domains imply that Aurenhammer's argument generalizes in a straightforward way (Borgwardt and Frongillo, 2019). For completeness, we provide a sketch of the proof for the restricted version. Another standard assumption in the literature on power diagrams is that all cells are full dimensional (as was done in Definition 10), although the machinery developed for them typically does not rely on this fact. The full-dimensional case is typically the interesting one in mechanism design as well because otherwise the set of types with such an allocation is of measure zero. Therefore we treat only this case for the remainder of this section.

Definition 12. *A* tiling C *of a convex set* $\mathcal{T} \subseteq \mathcal{V}$ *is a covering of* $\mathcal{T}$ *by finitely many polyhedra* Z_i *(where the* i *index the polyhedra) whose interiors when restricted to* $\mathcal{T}$ *(and relative to* $\mathcal{T}$ *if* $\mathcal{T}$ *is not full dimensional in* $\mathcal{V}$*) are non-empty and disjoint.*

Theorem 8 (Aurenhammer (1987a))**.** *Let* C *be a tiling of a convex set* $\mathcal{T} \subseteq \mathcal{V} = \mathbb{R}^d$ *by* n *polyhedra and* $\{p_1, \dots, p_n\}$ *be a point set. Then* C *is the restriction to* $\mathcal{T}$ *of a power diagram with sites* $\{p_1, \dots, p_n\}$ *defined on all of* $\mathcal{V}$*, if and only if*

1. *Orthogonality: For* $Z_i \neq Z_j$*, the line* L *that contains* p_i *and* p_j *(and is directed from* p_i *to* p_j*) is orthogonal to each face common to* Z_i *and* Z_j*.*
2. *Orientation: Any directed line that can be obtained by translating* L *and that intersects* Z_i *and* Z_j *first meets* Z_i*.*

Proof. One direction is trivial: if C is a restriction of a power diagram to $\mathcal{T}$ then by Aurenhammer's original theorem all choices of point set consistent with an unrestricted power diagram satisfy orthogonality and orientation on all of $\mathcal{V}$ and thus also on $\mathcal{T}$. For the other direction, Borgwardt and Frongillo (2019) show that a tiling of $\mathcal{T}$ is a restriction of a power diagram if and only if the following LP is feasible:

$$\begin{array}{rcll} \lambda_{ij} \cdot a_{ij} & = & p_i - p_j & \forall i \leq n, \forall j \in J_i \\ \lambda_{ij} \cdot \gamma_{ij} & = & \gamma_j - \gamma_i & \forall i \leq n, \forall j \in J_i \\ \lambda_{ij} & > & 0 & \forall i \leq n, \forall j \in J_i \end{array}$$

Here $i, j \in \{1, \dots, n\}$ index into the cells, J_i is the indices of cells adjacent to cell i, and the cells are given by constants $a_{ij} \in \mathbb{R}^d$, $\gamma_{ij} \in \mathbb{R}$, so that the ith cell is defined as $\{t \in \mathcal{T} \mid a_{ij} \cdot t \leq \gamma_{ij} \ \forall j \in J_i\}$. Thus, the variables are the sites

$\{p_1, \ldots, p_n\} \subseteq \mathbb{R}^d$ and pseudo-weights $\{\gamma_1, \ldots, \gamma_n\} \subseteq \mathbb{R}$ which are in bijection with the true weights.

In our setting, by contrast, we are given the p_i along with the a_{ij} and γ_{ij}, and need only find real numbers γ_i and λ_{ij} for which the program is feasible. By orthogonality, there is a unique choice of λ_{ij} satisfying the first constraint. Furthermore, by orientation it is strictly positive, satisfying the third constraint. For the second constraint, for any pair (i, j) with $j \in J_i$ and choice of γ_i there is a unique γ_j satisfying the constraint. Establishing the existence of a globally consistent set of choices is the heart of Aurenhammer's argument. In particular, he shows that if i, j, k share a vertex of C then for arbitrary γ_i the unique choices of γ_j and γ_k which satisfy the second constraint for (i, j) and (i, k) also satisfy it for (j, k). Global consistency then follows by a simple inductive construction. Start by choosing a cell i and arbitrary γ_i. At each step we assign some γ_j. If there is a j for which γ_j is unassigned and j has a vertex which is shared with two assigned cells then, per Aurenhammer's argument, we can assign γ_j consistent with all cells assigned so far. Otherwise every unassigned cell has a most a single face in common with a single assigned cell and therefore we can choose one which has such a face and it can trivially be assigned consistently. □

Theorem 9. *A tiling C is a restriction of a power diagram with sites $\{p_1, \ldots, p_n\}$ if and only if for all $t \in Z_i$ and $t' \in Z_j$ we have $p_i \cdot (t' - t) \leq p_j \cdot (t' - t)$ (i.e. C satisfies WMON).*

Proof. If C is a restriction of a power diagram to $\mathcal{T}$, then by definition

$$2p_i \cdot t - w_i \geq 2p_j \cdot t - w_j$$
$$2p_j \cdot t' - w_j \geq 2p_i \cdot t' - w_i.$$

Adding these shows C satisfies WMON.

Now suppose C satisfies WMON. We show orthogonality and orientation. For orthogonality, let $t, t' \in Z_i \cap Z_j$. Then $p_i \cdot (t' - t) = p_j \cdot (t' - t)$, or $(p_i - p_j) \cdot (t' - t) = 0$. Thus, the face is orthogonal to L. For orientation, let $t \in Z_i$ and $t' \in Z_j$ be on such a translated L. That is, we can write $t' = t + c(p_j - p_i)$ for some $c \in \mathbb{R}$. By WMON, $(p_j - p_i) \cdot (t' - t) \geq 0$, or $c(p_j - p_i) \cdot (p_j - p_i) \geq 0$. Thus $c \geq 0$. Therefore such a translated L first meets Z_i. □

Corollary 9 (Saks and Yu (2005))**.** *If $\mathcal{X}$ is finite, $\mathcal{T}$ is convex, and a satisfies WMON, then a is implementable.*

Proof. We apply Theorem 9 to conclude that a defines a power diagram with sites $\mathcal{X}$. Theorem 6 then gives implementability because the construction used in the proof uses the sites as the allocations (when rescaled to remove the 2s). In order to apply Theorem 9, it remains only to show that an allocation rule a satisfying WMON further implies that it defines a tiling. This follows by a straightforward geometric argument that has been used in a number of previous proofs (see, e.g., Lemma 4.2 of Archer and Kleinberg (2008)). For completeness, provide it here.

Let $x \in \mathcal{X}$ be given. We can define a polyhedron P_x associated with x by the intersection of the constraints $t \cdot (x-y) \geq \inf_{t' s.t. a(t')=x} t' \cdot (x-y)$ for all $y \in X$. By WMON, $\inf_{t' s.t. a(t')=x} t' \cdot (x-y) \geq \sup_{t s.t. a(t)=y} t \cdot (x-y)$, so for any distinct $x, y \in \mathcal{X}$, P_x and P_y are separated by the hyperplane $t \cdot (x-y) = \inf_{t' s.t. a(t')=x} t' \cdot (x-y)$, so their interiors with respect to $\mathcal{T}$ are disjoint. By construction $t \in P_{a(t)}$ for all t, which implies that these polyhedra cover $\mathcal{T}$. $\square$

6. Discussion

We have presented a model of truthful elicitation which generalizes and extends both mechanisms and scoring rules. On the mechanism design side, we have seen how our framework provides simpler, more general, or more constructive proofs of a number of known results about implementability and revenue equivalence, some of which lead to new results about scoring rules. On the scoring rules side, we have provided the first characterization for scoring rules for non-convex sets of probability distributions. We also show how results about power diagrams in the scoring rules literature lead to a new proof of the Saks–Yu result in mechanism design.

Our analysis makes use of the fact that $\mathsf{A}(t', t)$ is affine in t to ensure that $G(t) = \sup_{t'} \mathsf{A}(t', t)$ is a convex function. However, this convexity continues to hold if $\mathsf{A}(t', t)$ is instead a convex function of t. Thus, a natural direction for future work is to investigate characterizations of convex scores. While mechanisms can always be represented as affine functions by taking the types to be functions from allocations to $\mathbb{R}$, it may be more natural to treat the type as a parameter of a (convex) utility function. While many such utility functions are affine (e.g. dot-product valuations), others such as Cobb-Douglas functions are not. Berger, Müller, and Naeemi (2009; 2010) have investigated such functions and given characterizations that suggest a more general result is possible. Another potential application is scoring rules for alternate representations of uncertainty, several of which result in a decision maker optimizing a convex function (Halpern, 2003).

In one sense getting such a characterization is straightforward. In the affine case we want $\mathsf{A}(t', t)$ to be an affine function such that $\mathsf{A}(t', t) \leq G(t)$ and $\mathsf{A}(t', t') = G(t')$. Since we have fixed its value at a point, the only freedom we have is in the linear part of the function, and being such a linear function is exactly the definition of a subgradient. So while our characterization of affine scores is in some sense vacuous, it is also powerful in that it allows us to make use of the tools of convex analysis. A similarly vacuous characterization is possible for the convex case: $\mathsf{A}(t', t)$ is a convex function such that $\mathsf{A}(t', t) \leq G(t)$ and $\mathsf{A}(t', t') = G(t')$. The challenge is to find a restatement that is useful and naturally handles constraints such as those imposed by the form of a utility function.

Theorem 6 shows that scoring rules for finite properties are essentially equivalent to the weights and points that induce a power diagram. This characterization has already been applied to understand minimal peer prediction mechanisms (Frongillo and Witkowski, 2017). As power diagrams are known to be connected to the spines of amoebas in algebraic geometry, aspects of toric geometry used by string theorists, and tropical hypersurfaces in tropical geometry (Van Manen and Siersma, 2005), there may be useful characterization results in those fields as well. The last is particularly suggestive given the recent use of tropical geometry techniques in mechanism design (Baldwin and Klemperer, 2012).

Appendix A. Convex Analysis Primer

In this appendix, we review some facts from convex analysis that are used in the paper.

Fact 1. *Let* $\{f_t \in \mathsf{Aff}(\mathcal{V} \to \mathbb{R})\}_{t \in \mathcal{T}}$ *be a parameterized family of affine functions. Then* $G(t) = \sup_{t' \in T} f_{t'}(t)$ *is convex as the pointwise supremum of convex functions.*

This follows because convex functions are those with convex epigraphs. The epigraph of this supremum is the intersection of the epigraphs of the individual functions, which is a convex set as the intersection of convex sets.

Fact 2. $d : \mathbb{R} \to \mathbb{R}$ *is a selection of subgradients of a convex function on* $\mathbb{R}$ *if and only if it is monotone non-decreasing.*

See (Rockafellar, 1997, Theorem 24.3) for a proof of a slightly more general statement.

Fact 3. *For convex* G *on convex* $\mathcal{T}$, $\{dG_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t \in \mathcal{T}} \in \partial G$ *satisfies path independence.*

Informally, path independence means that integrals of dG_t do not depend on the path through $\mathcal{T}$ chosen (see Equation (B.3) and following for a formal definition). Since the restriction of G to a line is a one-dimensional convex function, $G(y) - G(x) = \int_{L_{xy}} dG_t(y - x)dt$ (Rockafellar, 1997, Corollary 24.2.1). Summing along the individual lines in a path from x to y gives that the value of the path integral is $G(y) - G(x)$ regardless of the path chosen.

The following is a classic result in convex analysis (cf. (Urruty and Lemaréchal, 2001, Thm E.1.4.1)) which we prove for completeness. In its statement, given a convex function $G : \mathcal{V} \to \overline{\mathbb{R}}$ we make use of its *convex conjugate* $G^* : \mathcal{V}^* \to \overline{\mathbb{R}}$, where $\mathcal{V}^*$ is the dual vector space to $\mathcal{V}$,[11] which is defined as $G^*(d) = \sup_{v \in \mathcal{V}} d(v) - G(v)$.

Fact 4. *Let* $G : \mathcal{V} \to \overline{\mathbb{R}}$ *be convex. Then for all* $v \in \mathcal{V}, d \in \mathcal{V}^*$,

$$G^*(d) = d(v) - G(v) \iff d \in \partial G_v.$$

[11] More specifically, we require $(\mathcal{V}, \mathcal{V}^*)$ to be a dual pair of vector spaces; cf. (Aliprantis and Border, 2007, Definition 5.90).

Proof. We can simply break down the conditions step by step:

$$\begin{aligned} G^*(d) = d(v) - G(v) &\iff v \in \operatorname{argsup}_{v' \in \mathcal{V}}\ d(v') - G(v') \\ &\iff \forall v' \in \mathcal{V},\ d(v) - G(v) \geq d(v') - G(v') \\ &\iff \forall v' \in \mathcal{V},\ G(v') \geq G(v) + d(v' - v), \end{aligned}$$

where in the last step we merely negated and added $d(v') \in \mathbb{R}$ to both sides. □

Fact 5. *For any convex function G, the set $\partial G^{-1}(d) \doteq \{x \in \mathrm{dom}(G) : d \in \partial G_x\}$ is convex.*

Proof. Let $x, x' \in \partial G^{-1}(d)$; then one easily shows (cf. Fact 4) that $G(x) - G(x') = d(x - x')$. Now let $\hat{x} = \alpha x + (1-\alpha)x'$; we have,

$$\begin{aligned} G(\hat{x}) &\leq \alpha G(x) + (1-\alpha)G(x') && \text{(A.1)} \\ &= \alpha(G(x) - G(x')) + G(x') \\ &= \alpha d(x - x') + G(x') \\ &= d(\hat{x} - x') + G(x') && \text{(A.2)} \\ &\leq G(\hat{x}), && \text{(A.3)} \end{aligned}$$

where we applied convexity of G in (A.1) and the subgradient inequality for d at x' in (A.3). Hence, by eq. (A.2) we have shown $G(\hat{x}) - G(x') = d(\hat{x} - x')$, so by Fact 4, $d \in \partial G_{\hat{x}}$. □

Appendix B. Characterizing Truthful Mechanisms

While our theorem provides a characterization of truthful mechanisms in terms of convex consumer surplus functions, this is not always the most natural representation for a mechanism. In this section, we examine two other approaches to characterizing truthful mechanisms that have been explored in the literature and show that they have insightful interpretations in convex analysis, which allows us to greatly simplify their proofs. Furthermore, our phrasing of these results is as conditions for a parameterized family of linear functions to be a selection of subgradients of a convex function. We believe this phrasing converts known results in mechanism design into new results in convex analysis. It also shows how any such result in convex analysis would give a characterization of implementable mechanisms. Note that certain results in this section require an assumption that the relevant parameterized families are in fact real-valued, which is natural given our focus on mechanism design.

Appendix B.1. Subgradient characterizations

From an algorithmic perspective, it may be more natural to focus on the design of the allocation rule f. There is a large literature that focuses on when there exists a choice of payments p to make f into a truthful mechanism (e.g. Saks and Yu (2005); Ashlagi et al. (2010)). Viewed through our theorem, this becomes a very natural convex analysis question: when is a function f a subgradient of a convex function?[12] Unsurprisingly, the central result in the literature is closely connected to convex analysis.

Definition 13. *A family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t\in\mathcal{T}}$ satisfies* cyclic monotonicity (CMON) *if for all finite sets $\{t_0, \ldots, t_k\} \subseteq \mathcal{T}$,*

$$\sum_{i=0}^{k} d_{t_i}(t_{i+1} - t_i) \leq 0, \tag{B.1}$$

where indices are taken modulo $k+1$. The weaker condition that (B.1) hold for all pairs $\{t_0, t_1\}$ is known as weak monotonicity (WMON).

A well known characterization from convex analysis is that a function f defined on a convex set is a subgradient of a convex function on that set iff it satisfies CMON Rockafellar (1997). Rochet's (1987) proof that payments exist to implement f on a possibly non-convex $\mathcal{T}$ iff f satisfies CMON is effectively a proof of a generalization of this theorem. Rochet notes that his proof is adapted from the one given in Rockafellar's text (1997) of the weaker theorem where $\mathcal{T}$ is restricted to be convex. We adapt Rochet's proof to highlight how its core is a construction of G. As we use this basic construction several times, we first analyze it independently.

Given any family $\{d_t\}_{t\in\mathcal{T}}$ of linear functions in $\mathsf{Lin}(\mathcal{V} \to \mathbb{R})$, define $P_d : \mathcal{T} \times \mathcal{V} \to \overline{\mathbb{R}}$ as follows:[13]

$$P_d(t, t') \doteq \sup_{\substack{k\in\mathbb{N},\ \{t_1,\ldots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_{k+1}=t'}} \sum_{i=0}^{k} d_{t_i}(t_{i+1} - t_i). \tag{B.2}$$

One way to interpret $P_d(t, t')$ is as the length of the shortest path from t to t' in a graph with edge weights determined by $-d$, and in that form has seen extensive use in mechanism design Vohra (2011). We interpret it somewhat differently, as the best lower bound on $G(t') - G(t)$ for an arbitrary convex function G with subgradients d (and infinity if there is no such convex function). In particular, computing the best lower bound at every point yields a convex function.

Lemma 1. *Let $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t\in\mathcal{T}}$ be given. If d satisfies CMON, then for all $t, t' \in \mathcal{T}$ and all $t'' \in \mathcal{V}$, the following hold:*

1. $P_d(t, t') + P_d(t', t'') \leq P_d(t, t'')$
2. $d_t(t'' - t) \leq P_d(t, t'')$
3. $P_d(t, t) = 0$

[12] More precisely, we want for all t the allocation $f(t)$ to be a subgradient at t. Equivalently, we can view f as a parameterized family of functions, which is how we state our results.

[13] Note that the second argument of P_d is from $\mathcal{V}$ rather than $\mathcal{T} \subset \mathcal{V}$ because we wish to apply this when, e.g., $t' \in \mathsf{Conv}(\mathcal{T})$.

4. $P_d(t,t') + P_d(t',t) \le 0$
5. $P_d(t,\cdot)$ *is convex and real-valued on* $\mathsf{Conv}(\mathcal{T})$*, with* $d \in \partial P_d(t,\cdot)$ *on* $\mathcal{T}$

Otherwise, $P_d \equiv \infty$ *on all inputs.*

Proof. If CMON is not satisfied, then there is a cycle $C = t_0, \ldots, t_k, t_0$ with positive sum. Then for any t and t' the path tC^jt' that consists of starting at t, going to t_0, going around the cycle j times, then going to t' has a sum that goes to infinity as j goes to infinity. For the remainder, assume that CMON is satisfied.

1. $P_d(t,t') + P_d(t',t'') \le P_d(t,t'')$

$$\begin{aligned}
&P_d(t,t') + P_d(t',t'') \\
&= \sup_{\substack{k\in\mathbb{N},\ \{t_1,\ldots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_{k+1}=t'}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) + \sup_{\substack{k\in\mathbb{N},\ \{t_1,\ldots,t_k\}\subseteq\mathcal{T} \\ t_0=t',\ t_{k+1}=t''}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&= \sup_{\substack{j,k\in\mathbb{N},\ \{t_1,\ldots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_j=t',\ t_{k+1}=t''}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&\le \sup_{\substack{k\in\mathbb{N},\ \{t_1,\ldots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_{k+1}=t''}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&= P_d(t,t'')
\end{aligned}$$

2. $d_t(t''-t) \le P_d(t,t'')$
   Taking $k=0$ shows that $d_t(t''-t)$ is an element of set over which the supremum is taken.
3. $P_d(t,t) = 0$
   By CMON, $P_d(t,t) \le 0$. By claim (2), $d_t(t-t) = 0 \le P_d(t,t)$.
4. $P_d(t,t') + P_d(t',t) \le 0$
   By claims (1) and (3), $P_d(t,t') + P_d(t',t) \le P(t,t) = 0$.
5. $P_d(t,\cdot)$ is convex and real-valued on $\mathsf{Conv}(\mathcal{T})$, with $d \in \partial P_d(t,\cdot)$ on $\mathcal{T}$
   By CMON, for $t' \in \mathcal{T}$ $P_d(t,t') \le -d_t(t-t')$. Thus, $P_d(t,t')$ is finite on $\mathcal{T}$. $P_d(t,\cdot)$ is a pointwise supremum of convex functions, so is convex. By

convexity, it is also finite on $\mathsf{Conv}(\mathcal{T})$. For any $t' \in \mathcal{T}$ and $t'' \in \mathsf{Conv}(\mathcal{T})$,

$$
\begin{aligned}
P_d(t,t') + d_{t'}(t''-t') &= d_{t'}(t''-t') + \sup_{\substack{k\in\mathbb{N},\ \{t_1,\dots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_{k+1}=t'}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&= \sup_{\substack{k\in\mathbb{N},\ \{t_1,\dots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_k=t'\ t_{k+1}=t''}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&\le \sup_{\substack{k\in\mathbb{N},\ \{t_1,\dots,t_k\}\subseteq\mathcal{T} \\ t_0=t,\ t_{k+1}=t''}} \sum_{i=0}^{k} d_{t_i}(t_{i+1}-t_i) \\
&= P_d(t,t''),
\end{aligned}
$$

so d_t satisfies (2).

□

Having extracted the construction at the core of Rochet's proof, the rephrasing of his result as a statement about convex functions now follows easily.

Theorem 10 (Adapted from Rochet (1987))**.** *A family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t\in\mathcal{T}}$ satisfies CMON if and only if there exists a convex $G : \mathsf{Conv}(\mathcal{T}) \to \mathbb{R}$ such that $\{d_t\}_{t\in\mathcal{T}} \in \partial G$.*

Proof. Given such a G, by (2) we have $d_{t_i}(t_{i+1}-t_i) \le G(t_{i+1}) - G(t_i)$. Summing gives (B.1). Given such a family $\{d_t\}_{t\in\mathcal{T}}$, fix some $t_0 \in \mathcal{T}$ and set $G : t \mapsto P_d(t_0, t)$. The result follows from Lemma 1(5). □

A number of papers have sought simpler and more natural conditions than CMON that are necessary and sufficient in special cases, e.g. Saks and Yu (2005); Archer and Kleinberg (2008); Ashlagi et al. (2010). These results are typically proven by showing they are equivalent to CMON. However, it is much more natural to directly construct the relevant G. As an example, we show one such result has a simple proof using our framework. This particular proof also has the advantage of providing a characterization of the payments that is more intuitive than the supremum in Rochet's construction.

As in Myerson's (1981) construction for the single-parameter case, we construct a G by integrating over d_t. In particular, for any two types x and y our construction makes use of the line integral

$$\int_{L_{xy}} d_t(y-x)dt = \int_0^1 d_{(1-t)x+ty}(y-x)dt. \tag{B.3}$$

As Berger et al. (2009) and Ashlagi et al. (2010) observed, if $\{d_t\}_{t\in\mathcal{T}}$ satisfies WMON and $\mathcal{T}$ is convex, this (Riemann) integral is well defined because it is the integral of a monotone function. If these line integrals vanish around all triangles (equivalently $\int_{L_{xy}} d_t(y-x)dt + \int_{L_{yz}} d_t(z-y)dt = \int_{L_{xz}} d_t(z-x)dt$) we say $\{d_t\}$ satisfies *path independence*.

Theorem 11 (adapted from Müller et al. (2007)). *For convex $\mathcal{T}$, a family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t\in\mathcal{T}}$ is a selection of subgradients of a convex function if and only if $\{d_t\}_{t\in\mathcal{T}}$ satisfies WMON and path independence.*

Proof. Given a convex function G and selection of subgradients $\{d_t\}$, $\{d_t\}$ satisfies CMON and thus WMON. Path independence also follows from convexity (Rockafellar (1997) p. 232). Now given a $\{d_t\}$ that satisfies WMON and path independence, fix a type $t_0 \in \mathcal{T}$ and define $G(t') = \int_{L_{t_0 t'}} d_t(t'-t_0)dt$ (well defined by WMON as the integral of a monotone function). Given $x, y, z \in \mathcal{T}$ such that $z = \lambda x + (1-\lambda)y$, by path independence and the linearity of d_z we have

$$\begin{aligned}
&\lambda G(x) + (1-\lambda)G(y) \\
&= G(z) + \lambda \int_{L_{zx}} d_t(x-z)dt + (1-\lambda)\int_{L_{zy}} d_t(y-z)dt \\
&\geq G(z) + \lambda d_z(x-z) + (1-\lambda)d_z(y-z) = G(z),
\end{aligned}$$

so G is convex. Similarly, for $x, y \in \mathcal{T}$, d_t satisfies (2) because

$$d_x(y-x) \leq \int_{L_{xy}} d_t(y-x)dt = G(y) - G(x). \quad \square$$

Appendix B.2. Local Characterizations

In many settings, it is easier to reason about the behavior of a mechanism given small changes to its input rather than arbitrary changes, so several authors have sought to characterize truthful mechanisms using local conditions Archer and Kleinberg (2008); Berger et al. (2009); Carroll (2012). We show in this section how many of these results are in essence a consequence of a more fundamental statement, that convexity is an inherently local feature. For example, in the twice differentiable case it can be verified by determining whether the Hessian is positive semidefinite at each point. We start with a local convexity result, and use it to show that an affine score is truthful if and only if it satisfies a very weak local truthfulness condition introduced by Carroll (2012). Afterwards we turn to a characterization by Archer and Kleinberg (2008) that proved a similar theorem for a different notion of local truthfulness. Our results (specifically Theorem 12) show that these two notions of local truthfulness are equivalent because Archer and Kleinberg's definition corresponds to the condition of being a local subgradient, while Carroll's corresponds to the condition of being a weak, local subgradient, which we now define.

Definition 14. *Let $\mathcal{T}$ be convex. A family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})\}_{t\in\mathcal{T}}$ is a* weak local subgradient (WLSG) *of a convex function $G : \mathcal{T} \to \overline{\mathbb{R}}$ if for all $t \in \mathcal{T}$ there exists an open neighborhood U_t of t such that for all $t' \in U_t$,*

$$G(t) \geq G(t') + d_{t'}(t-t') \quad \text{and} \quad G(t') \geq G(t) + d_t(t'-t). \tag{B.4}$$

Furthermore, if for every $s \in \mathcal{T}$, eq. (B.4) holds for all $t, t' \in U_s$, we say $\{d_t\}_{t\in\mathcal{T}}$ is a local subgradient (LSG) *of G.*

We now show that being a WLSG is a necessary and sufficient condition for a family of functions to be a selection of subgradients. The proof is heavily inspired by Carroll (2012).

Theorem 12. *Let $\mathcal{T}$ be convex. A family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \overline{\mathbb{R}})\}_{t\in\mathcal{T}}$ is a selection of subgradients of a convex function $G : \mathcal{T} \to \overline{\mathbb{R}}$ if and only if it is a WLSG of G.*

(Adapted from Carrol (2012)). As usual, the forward direction is trivial. For the other, let $t, t' \in \mathcal{T}$ be given; we show that the subgradient inequality for d_t holds at t'. By compactness of $\mathsf{Conv}(\{t, t'\})$, we have a finite set $t_i = \alpha_i t' + (1-\alpha_i)t$, where $0 = \alpha_0 \leq \cdots \leq \alpha_{k+1} = 1$, such that WLSG holds between each t_i and t_{i+1}. (The cover $\{U_s \mid s \in \mathsf{Conv}(\{t, t'\})\}$ has a finite subcover; take t_{2j} from the subcover and $t_{2j+1} \in U_{t_{2j}} \cap U_{t_{2j+2}}$.) By the WLSG condition (B.4), we have for each i,

$$0 \geq G(t_{i+1}) - G(t_i) + d_{t_{i+1}}(t_i - t_{i+1}) \tag{B.5}$$
$$0 \geq G(t_i) - G(t_{i+1}) + d_{t_i}(t_{i+1} - t_i). \tag{B.6}$$

Now using the identity $t_{i+1} - t_i = (\alpha_{i+1} - \alpha_i)(t' - t)$ and adding (B.5) to (B.6) yields

$$d_{t_{i+1}}(t' - t) \geq d_{t_i}(t' - t)$$

Chaining these inequalities yields that for all i,

$$d_{t_i}(t' - t) \geq d_{t_0}(t' - t) = d_t(t' - t). \tag{B.7}$$

Again using the identity $t_{i+1} - t_i = (\alpha_{i+1} - \alpha_i)(t' - t)$, we can apply (B.7) to (B.6) yielding

$$0 \geq \ G(t_i) - G(t_{i+1}) + d_t(t_{i+1} - t_i). \tag{B.8}$$

Summing (B.8) over $0 \leq i \leq k$ gives

$$0 \geq \ G(t_0) - G(t_{k+1}) + d_t(t_{k+1} - t_0).$$

Recalling our definitions for t_i yields the result. □

The WLSG condition translates to an analogous notion in terms of truthfulness, *weak local truthfulness*.

Definition 15. *An affine score is* weakly locally truthful *if for all $t \in \mathcal{T}$ there exists some open neighborhood U_t of t, such that truthfulness holds between t and every $t' \in U_t$, and vice versa. That is,*

$$\forall t \in \mathcal{T},\ \forall t' \in U_t,\ \ \mathsf{A}(t', t) \leq \mathsf{A}(t, t) \ \ \textit{and} \ \ \mathsf{A}(t, t') \leq \mathsf{A}(t', t'). \tag{B.9}$$

Corollary 10 (Generalization of Carroll (2012))**.** *An affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ for convex $\mathcal{T}$ is truthful if and only if it is weakly locally truthful.*

Proof. Defining $G(t) := \mathsf{A}(t,t)$, by weak local truthfulness we may write

$$G(t) = \mathsf{A}(t,t) \geq \mathsf{A}(t',t) = G(t') + \mathsf{A}_\ell(t', t-t')$$
$$G(t') = \mathsf{A}(t',t') \geq \mathsf{A}(t,t') = G(t) + \mathsf{A}_\ell(t, t'-t),$$

where t' is local to t and $\mathsf{A}_\ell(t,\cdot)$ is the linear part of $\mathsf{A}(t,\cdot)$. This says that $d_t = \mathsf{A}_\ell(t,\cdot)$ satisfies WLSG for convex function G; the rest follows from Theorem 12 and Theorem 1. $\square$

Finally, in the spirit of Section Appendix B.1, Archer and Kleinberg (2008) characterized local conditions under which an allocation rule can be made truthful. A key condition from their paper is *vortex-freeness*, which is a condition they show to be equivalent to local path independence (analogous to our terminology of weak local subgradients it can be thought of as weak local path independence). The other condition, local WMON, means that WMON holds in some neighborhood around each type. Their result then follows from the observation that local WMON and local path independence imply local subgradient. While this particular proof is not significantly simpler than the original, we believe it is somewhat more natural and clarifies the connection between the underlying reasons a notion of local truthfulness suffices both here and in Carroll's setting.

Corollary 11. *Let $\mathcal{T}$ be convex. A family $\{d_t \in \mathsf{Lin}(\mathcal{V} \to \mathbb{R})\}_{t\in\mathcal{T}}$ is a selection of subgradients of a convex function if and only if it satisfies local WMON and is vortex-free.*

Proof. We prove the reverse direction; suppose $\{d_t\}_{t\in\mathcal{T}}$ satisfies local WMON and is vortex-free. From Lemma 3.5 of Archer and Kleinberg (2008) we have that vortex-freeness is equivalent to path independence, so by Theorem 11 for all t there exists some open U_t such that $\{d_{t'}\}_{t'\in U_t}$ is the subgradient of some convex function $G^{(t)} : U_t \to \mathbb{R}$. We need only show the existence of some G such that $\{d_t\}_{t\in\mathcal{T}}$ is the subgradient of G on each U_t; the rest follows from Theorem 12.

Fix some $t_0 \in \mathcal{T}$ and define $G(t) = \int_{L_{t_0\,t}} d_{t'}dt'$, which is well defined by compactness of $\mathsf{Conv}(\{t_0,t\})$ and the fact that a locally increasing real-valued function is increasing. But for each t' and $t \in U_{t'}$ we can also write $G^{(t')}(t) = \int_{L_{t'\,t}} d_{t''}dt''$ by (Rockafellar, 1997, p. 232), and now by path independence we see that G and $G^{(t')}$ differ by a constant. Hence $\{d_t\}_{t\in\mathcal{T}}$ must be a subgradient of G on $U_{t'}$ as well, for all $t' \in \mathcal{T}$. $\square$

Appendix C. Revenue Equivalence

Perhaps the most celebrated result in auction theory is the revenue equivalence theorem, which states that, in a single item auction, the revenue from an agent (equivalently that agent's consumer surplus) is determined up to a constant by the equilibrium probability that each possible type of that agent will receive the item Myerson (1981). A large body of work has looked for more general conditions under which this holds (see, e.g., Krishna and Maenner (2001)) or what can be said when it does not Carbajal and Ely (2013). One general approach is due to Heydenreich et al. (2009), who use a graphical representation related to CMON. Given our main theorem, this is unsurprising. In convex analysis terms, asking whether an implementable allocation rule satisfies revenue equivalence is asking whether all convex functions that have a selection of their subgradients that corresponds to that allocation rule are the same up to a constant. As we saw in the proof of Lemma 1, CMON permits the natural construction of a convex function from its subgradient via (B.1). Intuitively, if we know the payments we want for some subset of types, we can check if those are consistent with a desired payment for some other type by checking whether this construction still works, both in terms of the constraints of the existing types on the new one and the new one on the existing ones. The following theorem applies this insight to get a result that is stronger than revenue equivalence as iteratively applying it characterizes the possible payments for *every* mechanism.

Theorem 13. *Let G be a convex function on $\mathsf{Conv}(\mathcal{T})$, $d = \{d_t\}_{t\in\mathcal{T}}$ a selection of its subgradients on $\mathcal{T}$, $S \subseteq \mathcal{T}$ non-empty, $t^* \in \mathcal{T} \setminus S$, and c be given. Then there exists a convex G' on $\mathsf{Conv}(\mathcal{T})$ agreeing with G on S, with $\{d_t\}_{t\in\mathcal{T}} \in \partial G'$ and $G'(t^*) = c$, if and only if*

$$\sup_{t_0\in S} G(t_0) + P_d(t_0, t^*) \leq c \leq \inf_{t_0\in S} G(t_0) - P_d(t^*, t_0) \tag{C.1}$$

Proof. Given such a G', the LHS of (C.1) becomes $\sup_{t_0\in S} G'(t_0) + P_d(t_0, t^*) \leq G'(t^*)$. Applying the definition of P_d (B.2) and then repeatedly applying the subgradient inequality (2) yields the desired inequality. Similarly, the RHS of (C.1) can be rewritten as $G'(t^*) + P_d(t^*, t_0) \leq G'(t_0)$ for all $t_0 \in S$, and the definition and subgradient inequality applied.

Now suppose (C.1) holds. Let $G'(t) \doteq \max\left\{c + P_d(t^*, t),\ \sup_{t_0\in S} G(t_0) + P_d(t_0, t)\right\}$. By Theorem 10, d satisfies CMON, so by Lemma 1 G' is convex, finite-valued on $\mathsf{Conv}(\mathcal{T})$, and has $\{d_t\} \in \partial G'$. Hence, we need only show that G' agrees with G on S and has $G'(t^*) = c$.

First, fixing any $t \in S$, we will establish the following:

$$G(t) = \sup_{t_0\in S} G(t_0) + P_d(t_0, t). \tag{C.2}$$

As $P_d(t,t) = 0$ from Lemma 1(3), we have $G(t) = G(t) + P_d(t,t) \leq \sup_{t_0\in S} G(t_0) + P_d(t_0, t)$. Furthermore, $G(t_0) + P_d(t_0, t) \leq G(t)$ for all $t_0 \in \mathcal{T}$ by repeated application of the subgradient inequality (2). Hence, we have $\sup_{t_0\in S} G(t_0) + P_d(t_0, t) \leq G(t)$ as well.

By eq. (C.2), we can write $G'(t) = \max\{c + P_d(t^*, t), G(t)\}$ when $t \in S$. But by the RHS of eq. (C.1), we see $c + P_d(t^*, t) \leq G(t)$, so $G'(t) = G(t)$. Similarly, applying the LHS of eq. (C.1) and $P_d(t^*, t^*) = 0$ to the definition of $G'(t^*)$, we have $G'(t^*) = c$. $\square$

Viewed through Theorem 13, revenue equivalence holds when the upper and lower bounds from (C.1) match after the value of G is fixed at a single point. This allows us to derive a necessary and sufficient condition for revenue equivalence that is equivalent to that given by Heydenreich et al. (2009) and actually applies to all affine scores. For example, this gives a revenue equivalence theorem for mechanisms with partial allocation.

Corollary 12 (Revenue Equivalence)**.** *Let a truthful affine score* $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ *be given, and* $d = \{d_t\}_{t \in \mathcal{T}}$ *be the corresponding selection of subgradients from* (3)*. Then every truthful affine score* $\mathsf{A}' : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ *with the same corresponding selection of subgradients differs from* A *by a constant (i.e.* $\mathsf{A}(t', t) = \mathsf{A}'(t', t) + c$*) if and only if* $P_d(t', t) + P_d(t, t') = 0$ *for all* $t, t' \in \mathcal{T}$*.*

Proof. We will show that the convex function G from eq. (3) is unique up to a constant if and only if $P_d(t', t) + P_d(t, t') = 0$ for all $t, t' \in \mathcal{T}$.

For the forward direction, let $t_0 \in \mathcal{T}$ be arbitrary. Then for all $t \in \mathcal{T}$, taking (C.1) with $S = \{t_0\}$ and $G(t) \doteq c + P_d(t_0, t)$ gives the condition $G(t_0) + P_d(t_0, t) \leq G'(t) \leq G(t_0) - P_d(t, t_0)$ for the value of $G'(t)$. But as $P_d(t, t_0) = -P_d(t_0, t)$ we have $G'(t) = P_d(t_0, t) + G(t_0) = G(t)$ for all t.

For the reverse direction, assume $P_d(t^1, t^2) \neq -P_d(t^2, t^1)$ for some $t^1, t^2 \in \mathcal{T}$, and let $G^1(t) \doteq P_d(t^1, t)$ and $G^2(t) \doteq P_d(t^1, t^2) + P_d(t^2, t)$. We easily check from Lemma 1(3) that $G^1(t^2) = G^2(t^2) = P_d(t^1, t^2)$, but we have $G^1(t^1) = 0$ while $G^2(t^1) = P_d(t^1, t^2) + P_d(t^2, t^1) \neq 0$. $\square$

We note that these two results are similar to results of Kos and Messner (2013). The main novelties in our version are showing that every value in the interval yields a convex function (as opposed to merely the extremal ones), the ability to characterize possible values after the values at multiple points are fixed (as opposed to a single point), and the framing in terms of convex analysis.

The conditions given by Theorem 13 and Corollary 12, while general, are not particularly intuitive. However, there are a number of special cases where they do have natural interpretations for mechanism design. The first is when the set of types is finite. In this setting (explored in an auction theory context in, e.g., Diakonikolas et al. (2012)) it is well known that revenue equivalence does not hold. The finite set of constraints (C.1) can be used in general as a linear program to, e.g., maximize revenue (see Section 6.5.2 of Vohra (2011) for an example). In particular cases, they may become simple enough to have a nice characterization. For example, in the single-parameter setting only a linear number of paths need be considered. This setting is illustrated in Figure C.4.

More broadly, as we saw in the proof of Theorem 11, the (supremum over the) sum can often be interpreted as an integral. In particular, the fact that G is convex guarantees that (under mild conditions) integrals of a selection of its subgradient are path independent and the integral from t to t' gives $G(t') - G(t)$. If $\mathcal{T}$ is connected by smooth paths (e.g. if it is convex), this means that $\mathcal{T}$ satisfies revenue equivalence for all implementable mechanisms (previously shown under a somewhat different notion of the set of types Heydenreich et al. (2009)). As it is particularly simple to prove, we state the version for convex $\mathcal{T}$.

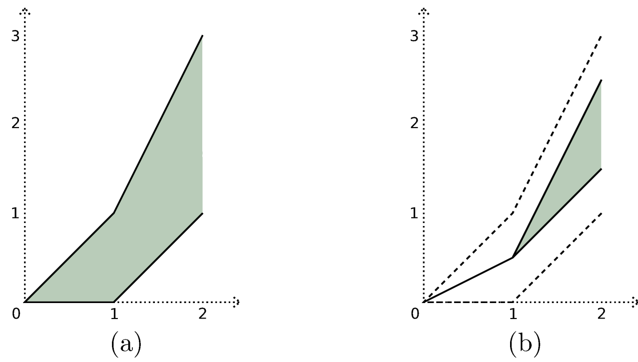

Figure C.4: Consider a one-dimensional setting with type space $\mathcal{T} = \{0, 1, 2\}$ and $d_0 = 0, d_1 = 1, d_2 = 2$. In (a), we fix $G(0) = 0$, yielding a range of possible values dictated by the subgradients: $0 \leq G(1) \leq 1$ and $1 \leq G(2) \leq 3$. We can pick any point in the resulting set and fix G there. However, we cannot pick any increasing function: in (b), we fix $G(1) = 0.5$, restricting $G(2)$ to the interval $[1.5, 2.5]$.

Corollary 13. *Let $\mathcal{T}$ be convex, a truthful affine score $\mathsf{A} : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ be given, and $\{dG_t\}_{t \in \mathcal{T}}$ be the corresponding selection of subgradients from (3). Then any truthful affine score $\mathsf{A}' : \mathcal{T} \times \mathcal{T} \to \overline{\mathbb{R}}$ with the same corresponding selection of subgradients differs from A by a constant (i.e. $\mathsf{A}(t', t) = \mathsf{A}'(t', t) + c$).*

Proof. By Theorem 1, we know that A and A' only differ only in their choice of convex function G. However, each choice has the same selection of subgradients, and two convex functions with the same selection of subgradients differ by a constant Rockafellar (1997). For intuition, see the construction of G by integrating its subgradients in the proof of Theorem 11. □

Acknowledgments

We would like to thank Aaron Archer, Steffen Borgwardt, Gabriel Carroll, Yiling Chen, Philip Dawid, Felix Fischer, Hu Fu, Philippe Jehiel, Nicolas Lambert, Peter Key, David Parkes, Colin Rowat, Greg Stoddard, Matus Telgarsky, Rakesh Vohra, and Bo Waggoner for helpful discussions, and anonymous reviewers for helpful feedback. This work was supported in part by National Science Foundation Grants CCF-1657598, CCF-1934915, and IIS-2045347.

References

Abernethy, J., Chen, Y., Vaughan, J.W., 2013. Efficient market making via convex optimization, and a connection to online learning. ACM Transactions on Economics and Computation 1, 12. URL: `http://dl.acm.org/citation.cfm?id=2465777`.

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

Archer, A., Kleinberg, R., 2008. Truthful germs are contagious: a local to global characterization of truthfulness, in: Proceedings of the 9th ACM Conference on Electronic Commerce, pp. 21–30. URL: http://dl.acm.org/citation.cfm?id=1386796.

Ashlagi, I., Braverman, M., Hassidim, A., Monderer, D., 2010. Monotonicity and implementability. Econometrica 78, 1749–1772. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA8882/abstract.

Aurenhammer, F., 1987a. A criterion for the affine equivalence of cell complexes in R d and convex polyhedra in R d+1. Discrete & Computational Geometry 2, 49–64. URL: http://link.springer.com/article/10.1007/BF02187870, doi:10.1007/BF02187870.

Aurenhammer, F., 1987b. Power diagrams: properties, algorithms and applications. SIAM Journal on Computing 16, 78–96. URL: http://epubs.siam.org/doi/pdf/10.1137/0216006.

Aurenhammer, F., 1987c. Recognising polytopical cell complexes and constructing projection polyhedra. Journal of Symbolic Computation 3, 249–255. URL: http://www.sciencedirect.com/science/article/pii/S0747717187800032, doi:10.1016/S0747-7171(87)80003-2.

Babaioff, M., Blumrosen, L., Lambert, N.S., Reingold, O., 2011. Only valuable experts can be valued, in: Proceedings of the 12th ACM conference on Electronic commerce, pp. 221–222. URL: http://www.stanford.edu/~nlambert/papers/screening-experts-full.pdf.

Babaioff, M., Sharma, Y., Slivkins, A., 2009. Characterizing truthful multi-armed bandit mechanisms, in: Proceedings of the 10th ACM conference on Electronic commerce, pp. 79–88. URL: http://dl.acm.org/citation.cfm?id=1566386.

Baldwin, E., Klemperer, P., 2012. Tropical geometry to analyse demand. Technical Report. Working paper, Oxford University. URL: http://www.nuff.ox.ac.uk/users/klemperer/Tropical.pdf.

Bauschke, H.H., Borwein, J.M., Combettes, P.L., 2001. Essential smoothness, essential strict convexity, and legendre functions in banach spaces. Communications in Contemporary Mathematics 3, 615–647.

Berger, A., Müller, R., Naeemi, S.H., 2009. Characterizing incentive compatibility for convex valuations, in: International Symposium on Algorithmic Game Theory, Springer. pp. 24–35.

Berger, A., Müller, R., Naeemi, S.H., 2010. Path-Monotonicity and Incentive Compatibility. Technical Report. Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization. URL: http://econpapers.repec.org/RePEc:dgr:umamet:2010035.

Bikhchandani, S., Chatterji, S., Lavi, R., Mu'alem, A., Nisan, N., Sen, A., 2006. Weak Monotonicity Characterizes Deterministic Dominant-Strategy Implementation. Econometrica 74, 1109–1132. URL: http://dx.doi.org/10.1111/j.1468-0262.2006.00695.x, doi:10.1111/j.1468-0262.2006.00695.x.

Borgwardt, S., Frongillo, R.M., 2019. Power diagram detection with applications to information elicitation. Journal of Optimization Theory and Applications 181, 184–196. doi:10.1007/s10957-018-1442-y.

Boutilier, C., 2012. Eliciting forecasts from self-interested experts: scoring rules for decision makers, in: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2, pp. 737–744.

Cai, Y., Mahdian, M., Mehta, A., Waggoner, B., 2013. Designing markets for daily deals, in: International Conference on Web and Internet Economics, Springer. pp. 82–95.

Carbajal, J.C., Ely, J.C., 2013. Mechanism design without revenue equivalence. Journal of Economic Theory 148, 104–133.

Carroll, G., 2012. When Are Local Incentive Constraints Sufficient? Econometrica 80, 661–686. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA9454/abstract.

Chambers, C.P., Lambert, N.S., 2014. Dynamically eliciting unobservable information, in: Proceedings of the fifteenth ACM conference on Economics and computation, ACM. pp. 987–988. URL: http://dl.acm.org/citation.cfm?id=2602859.

Chen, Y., Kash, I., Ruberry, M., Shnayder, V., 2011. Decision Markets With Good Incentives. Internet and Network Economics , 72–83.

Chen, Y., Kash, I.A., 2011. Information elicitation for decision making, in: The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pp. 175–182.

Chen, Y., Vaughan, J., 2010. A new understanding of prediction markets via no-regret learning, in: Proceedings of the 11th ACM conference on Electronic commerce, pp. 189–198.

Cid-Sueiro, J., 2012. Proper losses for learning from partial labels, in: Advances in Neural Information Processing Systems 25, pp. 1574–1582. URL: http://books.nips.cc/papers/files/nips25/NIPS2012_0738.pdf.

Diakonikolas, I., Papadimitriou, C., Pierrakos, G., Singer, Y., 2012. Efficiency-revenue trade-offs in auctions, in: Automata, Languages, and Programming. Springer, pp. 488–499. URL: http://link.springer.com/chapter/10.1007/978-3-642-31585-5_44.

Fang, F., Stinchcombe, M.B., Whinston, A.B., 2010. Proper scoring rules with arbitrary value functions. Journal of Mathematical Economics 46, 1200–1210. URL: http://www.sciencedirect.com/science/article/pii/S0304406810000984.

Feige, U., Tennenholtz, M., 2010. Responsive lotteries, in: International Symposium on Algorithmic Game Theory, Springer. pp. 150–161.

Feldman, J., Muthukrishnan, S., 2008. Algorithmic Methods for Sponsored Search Advertising, in: Liu, Z., Xia, C.H. (Eds.), Performance Modeling and Engineering. Springer US, pp. 91–122. URL: http://link.springer.com/chapter/10.1007/978-0-387-79361-0_4.

Fiat, A., Karlin, A., Koutsoupias, E., Vidali, A., 2013. Approaching Utopia: strong truthfulness and externality-resistant mechanisms, in: Proceedings of the 4th conference on Innovations in Theoretical Computer Science, pp. 221–230.

Frongillo, R., Witkowski, J., 2017. A Geometric Perspective on Minimal Peer Prediction. ACM Transactions on Econonimcs and Computation 5, 17:1–17:27. URL: http://doi.acm.org/10.1145/3070903, doi:10.1145/3070903.

Frongillo, R.M., Kash, I.A., 2019. General truthfulness characterizations via convex analysis. arXiv preprint arXiv:1211.3043 .

Gneiting, T., 2011. Making and Evaluating Point Forecasts. Journal of the American Statistical Association 106, 746–762.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378.

Halpern, J., 2003. Reasoning about uncertainty. MIT Press.

Hanson, R., 2003. Combinatorial Information Market Design. Information Systems Frontiers 5, 107–119.

Heydenreich, B., Müller, R., Uetz, M., Vohra, R.V., 2009. Characterization of revenue equivalence. Econometrica 77, 307–316. URL: http://onlinelibrary.wiley.com/doi/10.3982/ECTA7168/abstract.

Jehiel, P., Moldovanu, B., 2001. Efficient Design with Interdependent Valuations. Econometrica 69, 1237–1259. URL: http://dx.doi.org/10.1111/1468-0262.00240, doi:10.1111/1468-0262.00240.

Jehiel, P., Moldovanu, B., Stacchetti, E., 1996. How (Not) to Sell Nuclear Weapons. The American Economic Review 86, pp. 814–829. URL: http://www.jstor.org/stable/2118306.

Jehiel, P., Moldovanu, B., Stacchetti, E., 1999. Multidimensional Mechanism Design for Auctions with Externalities. Journal of Economic Theory 85, 258 – 293. URL: http://www.sciencedirect.com/science/article/pii/S0022053198925017, doi:http://dx.doi.org/10.1006/jeth.1998.2501.

Kos, N., Messner, M., 2013. Extremal incentive compatible transfers. Journal of Economic Theory 148, 134–164.

Krishna, V., Maenner, E., 2001. Convex Potentials with an Application to Mechanism Design. Econometrica 69, 1113–1119. URL: http://www.jstor.org/stable/2692257, doi:10.2307/2692257.

Lambert, N.S., Pennock, D.M., Shoham, Y., 2008. Eliciting properties of probability distributions, in: Proceedings of the 9th ACM Conference on Electronic Commerce, pp. 129–138.

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

McAfee, R., McMillan, J., 1988. Multidimensional incentive compatibility and mechanism design. Journal of Economic Theory 46, 335–354. URL: http://www.sciencedirect.com/science/article/pii/0022053188901354, doi:10.1016/0022-0531(88)90135-4.

Milgrom, P., Segal, I., 2002. Envelope Theorems for Arbitrary Choice Sets. Econometrica 70, 583–601. URL: http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00296/abstract, doi:10.1111/1468-0262.00296.

Myerson, R.B., 1981. Optimal auction design. Mathematics of operations research 6, 58–73.

Müller, R., Perea, A., Wolf, S., 2007. Weak monotonicity and Bayes-Nash incentive compatibility. Games and Economic Behavior 61, 344–358. URL: http://www.sciencedirect.com/science/article/pii/S0899825607000152.

Nielsen, F., Garcia, V., 2009. Statistical exponential families: A digest with flash cards. arXiv:0911.4863 [cs] URL: http://arxiv.org/abs/0911.4863. arXiv: 0911.4863.

Ok, E.A., 2007. Real analysis with economic applications. volume 10. Princeton University Press. URL: http://books.google.com/books?hl=en&lr=&id=f2V7ppFuE7oC&oi=fnd&pg=PP2&dq=related:NIiwYtHBU1IHeM:scholar.google.com/&ots=Vg8RTYE-9-&sig=R5sj24AHZEUOQd3SHrlMhP4jViY.

Osband, K.H., 1985. Providing Incentives for Better Cost Forecasting. University of California, Berkeley.

Othman, A., Sandholm, T., 2010. Decision rules and decision markets, in: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, pp. 625–632.

Peters, H.J.M., Wakker, P.P., 1987. Convex Functions on Non-Convex Domains. Economics Letters 22, 251–255. URL: http://repub.eur.nl/pub/23240/.

Rochet, J.C., 1985. The taxation principle and multi-time hamilton-jacobi equations. Journal of Mathematical Economics 14, 113–128.

Rochet, J.C., 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. Journal of Mathematical Economics 16, 191–200. URL: http://www.sciencedirect.com/science/article/pii/0304406887900073.

Rockafellar, R., 1997. Convex analysis. volume 28 of *Princeton Mathematics Series*. Princeton University Press.

Saks, M., Yu, L., 2005. Weak monotonicity suffices for truthfulness on convex domains, in: Proceedings of the 6th ACM conference on Electronic commerce, pp. 286–293. URL: http://dl.acm.org/citation.cfm?id=1064040.

Savage, L.J., 1971. Elicitation of personal probabilities and expectations. Journal of the American Statistical Association 66, 783–801.

Urruty, J.B.H., Lemaréchal, C., 2001. Fundamentals of Convex Analysis. Springer.

Van Manen, M., Siersma, D., 2005. Power diagrams and their applications. arXiv preprint math/0508037 URL: http://arxiv.org/abs/math/0508037.

Vohra, R.V., 2011. Mechanism design: a linear programming approach. Cambridge University Press, Cambridge; New York.

Waggoner, B., 2021. Linear functions to the extended reals. arXiv preprint arXiv:2102.09552 .

Yan, M., 2014. Extension of convex function. Journal of Convex Analysis 21, 965–987.