The probability of satisfying axioms: a non-binary perspective on economic design

Pierre Bardier[*]

Paris School of Economics and École Normale Supérieure de Paris,
48 Boulevard Jourdan 75014 Paris, France.

Abstract

We provide a formal framework accounting for a widespread idea in the theory of economic design: analytically established incompatibilities between given axioms should be qualified by the likelihood of their violation. We define the degree to which rules satisfy an axiom, as well as several axioms, on the basis of a probability measure over the inputs of the rules.

Armed with this notion of degree, we propose and characterize:

- *a criterion to evaluate and compare rules given a set of axioms, allowing the importance of each combination of axioms to differ, and*
- *a criterion to measure the compatibility between given axioms, building on a analogy with cooperative game theory.*

JEL classification: D47, D70, D71, D60
Keywords: Degree of satisfaction; probability of satisfaction; ranking of rules; performance of rules; desirability of axioms; compatibility of axioms; non-binary theory of market design, voting and social choice.

1 Introduction

In the theory of economic design, incompatibilities between axioms have given rise to a myriad of notions of the *degree to which a given axiom is satisfied*. However, these are, generally, model-and-axiom-specific. In contrast, this paper explores the potential of defining such a notion as the *probability with which an axiom, as well as a set of axioms, is satisfied*, without restricting the analysis to particular types of properties or problems.

[*] pierre.bardier@ens.fr

In the face of incompatibilities, notions of degree allow to compare in a nuanced way rules that are not comparable when sticking to the binary constraint according to which either a rule satisfies the axioms under consideration, if it meets the requirements for all the elements in its domain of definition, or does not, "at all", satisfy them. Take the example, without getting into details here, of "non-dictatorial" and "non-trivial" voting rules, defined on the universal domain of preferences associated with a finite set of alternatives and a finite set of voters. A consequence of the Gibbard-Satterthwaite theorem is that the relative merit of any of these rules cannot be assessed on the basis of the full-fledged axiom of "strategy-proofness" (Gibbard (1973), Satterthwaite (1975)). However, it is still possible, in principle, and useful, to compare the sensitivity to manipulation[1] of two "non-dictatorial" and "non-trivial" rules. One can, for instance, build an index representing the potential gains faced by voters misrepresenting their preferences in the two rules. Alternatively, one can compare the sets of preference profiles for which these rules are manipulable, using the partial order of set-inclusion. As this example suggests, two prominent interpretations support the use of notions of degree: one in which the parameters selected to measure the departure from a desirable property represent the *intensity* of the violation, and one in which they represent its *plausibility*. Our approach bears on the latter as we propose to compare rules according to the probability that they satisfy an axiom, or a set of axioms.[2]

Studies discussing the likelihood that a rule satisfy a certain axiom, be that through empirical or theoretical analysis, all require that a specific way of *counting the instances* for which the rule meets the considered requirements be chosen. These instances can be composed of theoretical preference profiles, or stochastically generated ones, sets of alternatives, parameters of actual elections, *etc.* In that respect, simulation models, focusing initially mostly on the occurrence of the "Condorcet paradox" in voting (see Gehrlein (1983) and Gehrlein et al. (2017) for reviews of this literature), are now commonly used in diverse settings (*e.g.*, in addition to voting, market design, fair division), under decreasingly restrictive statistical assumptions (Wilson (2019), Diss and Kamwa (2020), Szufa et al. (2020), Boehmer et al. (2021), Boehmer et al. (2023), Böhm et al. (2024)).

In line with these models, while also accounting for other approaches (see Section 2), we consider an *abstract set of instances* —the inputs of a rule— endowed with a probability structure reflecting their relative frequency. The typical example of such a set in our view is the set of preference profiles associated with either a fixed or a varying group of agents.

[1] Or the lack of "strategy-proofness".

[2] Other mathematical objects than (probability) measures can capture the plausibility of the violation. As an illustration, topological notions may be involved, for example, informally, in statements concluding that the set of preference profiles for which a rule does not meet the requirements of an axiom is "Baire-negligible".

We can then measure the mass of instances for which not only "punctual" axioms, but also "relational" axioms (Thomson (2023)) are verified.[3] This general framework applies in any field in which an axiomatic approach is relevant, and in particular, covers a wide spectrum of market design, voting and social choice problems, be they Arrovian aggregation problems, voting problems, fair division selection —or ranking— problems, with divisible or indivisible resources. Importantly, it does so while providing the degree of satisfaction of either single axioms or sets of axioms.

Defining the degree of satisfaction as a probability guarantees, in contrast to defining it on the basis of a notion of intensity, its commensurability across (sets of) axioms. Concretely, the possibility to compare the extent to which a given rule satisfies two different combinations of axioms proves fundamental to **(1)** evaluate and compare rules, and **(2)** measure the compatibility of these axioms.

Let us first illustrate the simple objects around which this work is structured. The questions raised in **(1)** and **(2)** will be addressed on the basis of collections of probabilities presented in arrays of the following form:

Example 1. Consider three axioms, a_1, a_2 and a_3:

$$\begin{array}{ccccccc} a_1 & a_2 & a_3 & a_1a_2 & a_1a_3 & a_2a_3 & a_1a_2a_3 \\ 1 & 0.8 & 0.4 & 0.8 & 0.4 & 0.35 & 0.35 \end{array},$$

where the 6th column, say, reads as "the considered rule satisfies axiom a_2 and axiom a_3 *simultaneously* with probability 0.35" —precise definitions are given in Section 3.

Given a set of instances and a set of axioms, there is an intuitive criterion, to which we alluded in the discussion of the consequences of the Gibbard-Satterthwaite theorem, and with which the one we propose is consistent. According to it, a rule performs better than another one if, for each combination of axioms, the subset of instances for which it violates the requirements is included in the set of instances for which the other rule violates the requirements. Certainly, interesting comparisons of rules can be derived for some types of problems using this criterion (Pathak and Sönmez (2013), Arribillaga and Massó (2016), Abdulkadiroğlu et al. (2020), Abdulkadiroğlu and Grigoryan (2021)).[4] In general, nevertheless, it induces a very partial ranking of rules: by working with a notion

[3] We propose a formal definition of these two types of axioms in Section 3.1. Informally, some axioms are requirements made on outcomes obtained for each instance separately, while others formulate restrictions on outcomes obtained from different instances related in a specific way.

[4] A related approach consists in looking for a rule such that the set of instances for which the rule satisfies the axioms is maximal for inclusion (Dasgupta and Maskin (2008), Barberà and Gerber (2017)).

of degree based on probabilities of satisfaction, one obtains an *extension* of such an order, accounting for the fact that some violations are more likely than others.[5]

(1) *Evaluating and comparing rules.* We actually consider a *completion*[6] of the partial order we just described. Indeed, we introduce and characterise a criterion to quantify the performance of a rule with regard to two key components. The first component is, as expected, probabilities of satisfaction. The second component focuses on the specific normative content of axioms. More precisely, the normative desirability of axioms, and, crucially, that of their combinations, are defined through the use of a *capacity.*[7] Not only can an axiom be more valuable to the eye of a researcher or a policy maker than another one, but *synergies* are likely to emerge in the combination of axioms: conditionally on the satisfaction of a given axiom, the satisfaction of another one may be more or less valued, so that these axioms may be "complementary" or "substitutable". Capacities enable to capture this type of dependence.

Importantly, this formulation has an operational interpretation. A decision maker, for example a policy maker in charge of selecting a mechanism to match students with schools, must distinguish between two rules on the basis of several principles that are logically incompatible. She then asks a team of researchers to estimate, for each rule, how probable the satisfaction of each combination of principles is. After determining it, the research team inquires about the relative importance of each combination of principles for the policy maker. The task is then to integrate these two pieces of information in order to decide on the rule to adopt.

Loosely speaking, we identify the only measure of performance that *i*) consistently extends the natural measure for the case of a single axiom, while *ii*) disentangling the probability with which a rule satisfies a set of axioms and the probability with which it satisfies a superset of it —see Theorem 1. The necessity to do so comes from the monotonicity of capacities: the valuation of the satisfaction of a given combination of axioms is incorporated in that of a superset of it. We show that a measure failing point *ii)* displays some redundancy and may thus wrongly lead to the conclusion that some rule performs better than another one.

(2) *Measuring the compatibility of axioms.* Finally, a collection of probabilities can be analysed in order to determine the degree of compatibility, or, equivalently, the degree of incompatibility, of axioms, given a rule, or given a family of rules. When such a collection

[5] In the school choice context, for example, working with such a notion enables to take into account the correlation between students' preferences.

[6] That is, an extension to a complete order.

[7] A capacity is a real-valued function defined on the power set associated with the axioms, which gives value 0 to the empty set, and is monotonic with respect to inclusion.

is associated with one specific rule, computing how likely this rule is to satisfy a certain axiom, *given that it satisfies some others*, enables to better understand its behaviour. One can also analyse collections of probabilities to identify how (in)compatible axioms are, given a domain of admissible rules.

We introduce and characterise a criterion fulfilling this purpose based on an analogy with cooperative game theory. Admissible collections of probabilities are naturally associated with a unique cooperative game and, *on the obtained restricted set of games*, we identify the Shapley value as the most adequate measure —see Theorem 2.

In Section 2, we situate our approach in relation to the literature. In Section 3, we provide general definitions of rules, of "punctual" and "relational" axioms, and of the degree to which a rule satisfies a given set of axioms. We address question **(1)** in Section 4, after characterising the set of admissible collections of probabilities. We address question **(2)** in Section 5. Finally, in Section 6, we discuss methods to proceed to a robust analysis with respect to the probabilities of satisfaction.

2 Related Literature

Thomson (2001), in a paper in which he seeks to characterise the essential features of the *axiomatic program*, conceives this research as the attempt to draw as precise a frontier as possible between axioms that are compatible and axioms that are not. It is then possible to distinguish, for a given set of axioms, families of problems for which they can all be satisfied, and families for which they cannot. This view has motivated the most standard way of dealing with impossibilities in the theory of economic design: when some axioms are shown to be incompatible on a given domain of parameters, it seems natural to look for restricted domains in which these axioms can actually be combined. Accordingly, the plausibility of the compatibility of these axioms becomes the plausibility of the restricted domains, and it is left to the consumer of the theory to assess how suitable the domain restrictions are in the context at hand. Recently, this type of approach has saliently been described in Moulin (2019), reviewing new developments in the theory of fair allocation, centered around very structured problems such as ones with "one-dimensional single-peaked preferences", "dichotomous preferences", or "preferences with perfect substitutability". Restricted preference domains such as those described above have also received special interest in algorithmic social choice theory, in particular because their simpler structure is likely to decrease the complexity of algorithms (Brandt et al. (2016a)).

Yet, this approach maintains the binary constraint according to which a given condition is satisfied on a whole domain of parameters or is not, "at all", satisfied, whereas constructing a *less partial order* between rules would require to know, when one fails to yield the desired outcomes, by how much it fails. For that matter, the use of parametrically weakened versions is quite classical: one or several parameters indicate the intensity of departure from the original studied property, see Moulin and Thomson (1988), Schummer (2004), Brandt et al. (2011), Chevaleyre et al. (2017) and Skowron (2021) for instance.[8][9] However, parametrizations are model-and-axiom-specific, which makes them, most often, incomparable to each other. In other words, most often, the definition of parametrized versions of two axioms gives no clue on how to define the degree to which they are simultaneously satisfied. In contrast, in this paper, we exploit the commensurability that a probability notion offers when measuring the performance of rules, as well as when studying the compatibility of axioms.

Another theoretical approach, closer to the way we proceed, was adopted in the context of Arrovian social choice theory in Campbell and Kelly (1994, 2015). The method is to "count", using a (probability) measure, the pairs, or triples, of alternatives for which studied axioms are satisfied in order to identify *trade-offs* between them. One can describe this method in the general terms of our paper: the set of instances endowed with a measure structure in these papers is the set of pairs, or triples, of alternatives —and not, for example, the set of preference profiles. For us, the point of considering an abstract set of instances is precisely to be able to account for various approaches *i)* involving different sets on which a measure is defined, and *ii)* deriving degrees of satisfaction through different methods, *e.g.* mathematical analysis, as in Campbell and Kelly (1994, 2015), simulations or econometric estimations.

We already discussed in the introduction how this work relates to simulation models. Their general principle is to derive, from (statistical) assumptions on the behaviour and the preferences of agents involved in a given aggregation problem, the probability of occurrence of certain types of outcomes under different rules. A recent review of this vast literature can be found in Diss and Kamwa (2020). Let us mention a few examples of studies in computational social choice and market design primarily consisting in measuring the empirical frequency of the violation of a given property, different from "Condorcet consistency".[10] In the former literature, Brandt et al. (2014) study the number of solutions

[8] Actually Moulin and Thomson (1988) show how parametric relaxations can be used to demonstrate the salience of the incompatibility of given principles.

[9] The number of papers introducing parametric relaxations is extremely large and this list is by no means exhaustive. All the cited papers belong to a different branch of the economic design literature.

[10] Once again, see Gehrlein (1983) and Gehrlein et al. (2017) for a review of the literature specifically dedicated to the "Condorcet paradox". See also Lepelley et al. (2000) and Laslier (2010).

selected by standard tournament solution concepts, using both real world preference data and simulations, thus testing for their (lack of) decisiveness. Focusing on the occurrence of the "agenda contraction paradox", Brandt et al. (2016b) conclude, based on simulations used to extend theoretical results obtained for problems involving four alternatives, that sensitivity to such contraction is of higher practical relevance than the "Condorcet loser paradox". Aleskerov et al. (2012) study the level of manipulability of multi-valued rules, using computer experiments on problems with four and five alternatives, after extending theoretical indices used for single-valued social choice procedures. In market design, Roth and Peranson (1999) conduct simulations on data from the "National Resident Matching Program" to account for the manipulability of the matching mechanism, and observed that even if it is in principle manipulable, the number of agents who would have an interest in returning a false report vanishes as the size of the market grows. Ghasvareh et al. (2020) (Chapter 4) compare the frequency of "priority violations" in three well-known many-to-one matching mechanisms that satisfy "strategy-proofness" and "efficiency", for different statistical distributions on preference profiles.

Taking stock, we believe that the present work can help analyse and compare rules in a subtle way by providing measures of performance that incorporate the probability to simultaneously satisfy several axioms, as well as their normative desirability and that of their combination. In particular, it provides a way to enrich the use of models based on notions of degree interpreted in terms of frequency of satisfaction.[11]

3 A commensurable notion of degree of satisfaction

The core of our analysis is conducted on the basis of collections of probabilities such as the one in Example 1. The set of all collections is characterised in Section 4.1 and in the Appendix (Section 8.1). In this section, we propose a definition of axioms, and of the degree of satisfaction, which cover the vast majority of axioms studied in the theory of economic design.

3.1 Rules and axioms

Any notion of the frequency with which a given rule satisfies axioms requires considering *instances* over which measuring its behaviour. As suggested above, letting preference profiles vary and analysing the *outcomes* prescribed by a rule, which is often done in practice,

[11] Wilson (2019) highlighted the importance of this issue for computer experiments.

is an obvious way to study instances —and distributions over these instances. That is why, for concreteness, we use the case of varying preference profiles in all the illustrations of subsequent definitions.

However, in order to be as general as possible, we introduce an abstract notion of instance, from which *classes of problems*, *rules* and *axioms* are defined. Informally, the frequency of satisfaction of an axiom will be defined as the measure of the set of instances for which the considered rule meets the stated requirements.

The starting point for evaluating the performance of rules is to specify the relevant domain: we define a **class of problems** as a pair of sets (I, O), and refer to elements of I as instances, and to elements of O as outcomes.[12] These objects respectively represent the arguments and the images of a rule: a **rule** f is a mapping:

$$f : I \to 2^O.$$

As an illustration, in the classical microeconomic division problem, an instance is a profile of continuous, monotonic and convex individual preferences over $\mathbb{R}^l_+$ associated to a group N of $n \in \mathbb{N}$ agents among which a social endowment $\Omega \in \mathbb{R}^l_+$ of $l \in \mathbb{N}$ divisible resources must be allocated.[13] A rule is then a *correspondence*, mapping each such instance i to a set o of vectors in $\mathbb{R}^{nl}_+$.

A slight modification of this definition is needed in order to cover ranking problems.[14]

Before we define axioms, an important distinction should be made between, in the words of Thomson (2023), *punctual* and *relational* axioms. The former are requirements imposed on outcomes obtained for each instance separately, while the latter formulate restrictions on outcomes obtained from different instances related in a specific way. In the microeconomic division framework just mentioned, “efficiency” is a punctual axiom, and so is “no-envy”, while “population monotonicity” is a relational one.[15]

A **punctual axiom** a is a mapping:

$$\begin{aligned} a : I &\to 2^O \\ i &\mapsto O^a_i. \end{aligned}$$

[12] All the objects we consider depend on a specific pair (I, O), but this dependence is most often omitted in the following.

[13] Given a set B and a natural number K, B^K denotes the $K-$fold Cartesian product of B. In addition, $\mathbb{R}^K_+$ $(\mathbb{R}^K_{++})$ denote the set of vectors in $\mathbb{R}^K$ with only non-negative (positive) components.

[14] A rule is then a mapping $f : I \to R(O)$, where $R(O)$ is the set of binary relations on O. All the definitions below are easily adapted by “replacing” 2^O by $R(O)$, and “$\subseteq$” by “$\in$”.

[15] For definitions of these conditions, see, *e.g.*, Moulin (2019) (Section 3.3).

In words, a specifies for each instance a set of admissible outcomes, and the image of an instance under rule f satisfies the requirements of a if and only if it is included in this set of admissible outcomes. Then, the typical exercise in economic design consists in finding a rule f —and, ideally, all rules f— such that:

$$\text{for all } i \in I,\ f(i) \subseteq O_i^a.$$

It is not surprising that a punctual axiom be defined in the same way as a rule, that is, as a mapping from I to 2^O: it is standard to associate a punctual axiom with a correspondence; for example, in the microeconomic division framework, the "efficiency" axiom naturally induces the "Pareto correspondence", which selects, for each admissible preference profile, all the "Pareto efficient" allocations.

A **relational axiom** a is associated with a parameter $K^a \in \mathbb{N}$ and is a mapping:

$$\begin{aligned} a : I^{K^a} &\to 2^{O^{K^a}} \\ (i_1, ..., i_{K^a}) &\mapsto O^a_{i_1,...,i_{K^a}}. \end{aligned}$$

Similarly, one typically looks for rules f such that:

$$\text{for all } (i_1, ..., i_{K^a}) \in I^{K^a},\ (f(i_1), ..., f(i_{K^a})) \subseteq O^a_{i_1,...,i_{K^a}}.$$

This is a general definition, but most relational axioms considered in different domains involve the comparison of outcomes obtained from only two different instances. Returning to the above example, "population monotonicity" requires to consider, for a fixed social endowment, the outcomes of a rule when computed for a profile of preferences of a group of agents N and a profile of a group $N \cup \{k\}$, $k \notin N$, which coincides with the preceding profile for agents in N.

In words, a specifies for each tuple of instances a set of admissible tuples of outcomes, and the image of a tuple of instances under rule f satisfies the requirements of a if and only if it is included this set of admissible tuples of outcomes. According to this definition, there typically are tuples of instances $(i_1, ..., i_{K^a})$ for which $O^a_{i_1,...,i_{K^a}} = 2^{O^{K^a}}$, that is, for which a imposes no restriction whatsoever.[16] In our example, "population monotonicity" is silent about pairs of preference profiles such that none is an extension of the other to a superset of agents.

[16] That is, a relational axiom associates with any tuple of instances *related in a specific way* a *specific* set of admissible tuples of outcomes, and does not associate with any other tuple of instances a restricted set of admissible tuples of outcomes.

The reader can see that taking $K^a = 1$ yields the definition of a punctual axiom; we however maintain this conceptually meaningful distinction for presentation purposes.[17]

3.2 The probability of satisfying axioms

The vast majority of studies using stochastic preference models to generate instances, as well as the vast majority of papers involving a mathematical analysis of a set of instances endowed with a probability structure, have focused on *i)* a single axiom at a time, and *ii)* a punctual one. It is, however, possible to define the mass of instances for which a rule satisfies simultaneously several punctual or relational axioms.

Let A be a finite set of $J \in \mathbb{N}$ axioms and $f : I \to 2^O$ a rule. Collections of probabilities $(p^f_S)_{\emptyset \neq S \subseteq A} \in [0,1]^{2^{J-1}}$ such as the one given in Example 1 are obtained in the following way.

Let us first illustrate what the definition of the degrees of satisfaction would be if A were only made of punctual axioms. Let ξ be a $\sigma-$algebra defined on I and μ a probability measure defined on (I, ξ). Let a be a punctual axiom, and assume $D^f(a) = \{i \in I, f(i) \subseteq O^a_i\}$, the set of instances whose image under f meets the requirements imposed in a, is measurable. Then the degree to which f satisfies a according to μ is simply $\mu(D^f(a))$. Similarly, the degree to which all the axioms in $S \subseteq A$ are simultaneously satisfied is simply $\mu(\bigcap_{a \in S} D^f(a))$.

The general definition covering the case of sets of relational axioms requires additional notation but its principle is the same. In order to cover both types of axioms, set $K^a = 1$ when a is punctual, even if this parameter is not needed for the definition of a punctual axiom. Similarly to what precedes, define $D^f(a) = \{(i_1, ..., i_{K^a}) \in I^{K^a}, (f(i_1), ..., f(i_{K^a})) \subseteq O^a_{i_1,...,i_{K^a}}\}$ for $a \in A$, the set of tuples of instances whose images under f meet the requirements imposed in a.

Let $K^A = \max_{a \in A} K^a$ and consider I^{K^A}, endowed with a $\sigma-$algebra ξ^{K^A}. For a probability measure defined on (I^{K^A}, ξ^{K^A}), denoted by μ, the **degree** to which f satisfies S, a

[17] These definitions do not cover *all* conceivable axioms. Some "existential axioms" (Fishburn (2015)), such as conditions on the range of a rule cannot be formulated in this way. They still can be analysed in the way we propose in Sections 4 and 5, by considering that they are satisfied by a rule either with degree 0 or with degree 1.

In addition, as noted in Thomson (2023), this distinction is *soft* in the sense that some properties can be formulated both as punctual and as relational axioms. In such a case, the choice between these formulations is at the discretion of the researcher. Schmidtlein and Endriss (2023) propose an interesting discussion on the different ways of defining an axiom.

non-empty subset of A, is given by:

$$p_S^f = \mu\Big(\Big\{(i_1, ..., i_{K^A}) \text{ such that } (i_1, ..., i_{K^a}) \in D^f(a) \text{ for all } a \in S\Big\}\Big).^{[18]}$$

To summarise, p_S^f is the proportion —computed from the probability measure μ— of tuples of K^A instances such that, for any axiom $a \in S$, the image under f of their restriction to the K^a relevant instances satisfies the requirements of a.

Remark: For finite classes of problems, *i.e.* for pairs (I, O) such that I and O are finite, the measurability assumptions we introduced are innocuous. This is the case for voting problems with finitely many potential voters and candidates, and for allocation problems with finitely many potential agents and indivisible items. These two examples are of primary importance in our framework as they are studied in two fields of research where simulations are extensively used.

4 How to measure the performance of rules ?

How can one assess the performance of a rule f and, importantly, compare it with that of other rules, based on $p^f = (p_S^f)_{\emptyset \neq S \subseteq A}$, the probabilities of satisfying axioms in $A = \{a_1, ..., a_J\}$, defined in Section 3.2 ?

4.1 Admissible collections of probabilities

We address this issue by constructing a performance criterion defined for any p in the set of possible collections of probabilities, a subset P of $[0,1]^{2^{J-1}}$ characterised by consistency conditions relating, for all $\emptyset \neq S \subseteq A$, the probability to satisfy all subsets of S.

For instance, the probability of satisfying all the axioms in S cannot be larger than any of the probabilities of satisfying all of them but one, that is, to any probability in $(p_{S \setminus a})_{a \in S}$.[19] The probability of satisfying all axioms in S is also constrained below. Select $a \in S$; taking $p_{S\setminus a}$ and p_a as given, what is the worst case in terms of probability of satisfying $S \setminus a$ *and* a? It corresponds to the situation in which the intersection of the sets of tuples of instances for which they are respectively satisfied has minimal measure, and the associated probability is $1 - (1 - p_{S\setminus a}) - (1 - p_a) = p_{S\setminus a} - (1 - p_a)$ if it is positive, 0

[18] We slighly abuse notation here by omitting the permutation used to restrict to the relevant K^a instances for each $a \in S$. In addition, similarly to the case of a punctual axiom, $\Big\{(i_1, ..., i_{K^A})$ such that $(i_1, ..., i_{K^a}) \in D^f(a)$ for $a \in S\Big\}$ is assumed measurable in (I^{K^A}, ξ^{K^A}).

[19] As is standard, we abuse notation by writing $S \setminus a$ rather than $S \setminus \{a\}$.

otherwise. In Example 1, given that the sets of tuples of instances for which a_2 and a_3 are satisfied have measure 0.8 and 0.4 respectively, the set for which they are simultaneously satisfied has at least measure $1 - 0.2 - 0.6 = 0.2$.

We gave some necessary conditions for $p \in [0,1]^{2^J-1}$ to be an admissible collection of probabilities —referred to as *Fréchet inequalities*. They are not, nevertheless, sufficient, as the following example reveals.

Example 2. Let $A = \{a_1, a_2, a_3\}$ and consider:

$$\begin{array}{cccccccc} & a_1 & a_2 & a_3 & a_1a_2 & a_1a_3 & a_2a_3 & A \\ \mathbf{p} & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.4 \end{array}.$$

The collection p meets the two conditions above. However, letting f be a rule associated with p, the set of tuples of instances for which f satisfies a_1 and the set of tuples of instances for which f satisfies a_2 are equal up to a set with measure 0 —if it were not the case, f would not simultaneously satisfy the two axioms with the same probability as it satisfies each of them. Similarly, the set of tuples of instances for which f satisfies a_2 and the set of tuples of instances for which f satisfies a_3 are equal up to a set with measure 0. In addition, all these sets have measure 0.7. It is then inconsistent that f simultaneously satisfy a_1, a_2 and a_3 with probability 0.4 only: it must satisfy them with probability 0.7.

The framework of *probabilistic Boolean satisfyability* (Nilsson (1986), Georgakopoulos et al. (1988)) provides a natural formulation to identify necessary and sufficient conditions for a collection in $[0,1]^{2^J-1}$ to be a collection of probabilities. For the sake of brevity, we develop this idea in the appendix and simply state in this section the characterisation of P, the **set of possible collections of probabilities**.

Before the statement, we need to introduce a specific family of collections. Let $\emptyset \neq S \subseteq A$. We define $p^{1,S}$ by:

$$\begin{aligned} p_T^{1,S} &= 1 \text{ if } \emptyset \neq T \subseteq S, \\ p_T^{1,S} &= 0 \text{ otherwise.} \end{aligned}$$

A rule with degrees of satisfaction given by $p^{1,S}$ satisfies any combination of axioms which is not included in S with probability 0. In addition, it satisfies all the elements of S with probability 1, and *thus*, all subsets of S with probability 1 (S included). We comment more substantially on these collections later, as they correspond to special cases of the *single-axiom-reducible problems* we introduce in Section 4.3.

Let $\mathbf{0}$ denote the null vector in $\mathbb{R}^{2^J-1}$.

Lemma 1. *P is the closed convex hull of $\mathbf{0}$ and $(p^{1,S})_{\emptyset \neq S \subseteq A}$.*

The proof of Lemma 1, as well as all subsequent proofs, is in the appendix.

Importantly, when defining a measure of performance *for all* $p \in P \subseteq [0,1]^{2^J-1}$ (see Section 4.3), we look for a measure that is applicable to *any* set of J axioms. For example, for two specific axioms a_1 and a_2 such that a_1 logically implies a_2 on the considered class of problems, it is impossible to find a probability measure on the set of instances such that a_1 is satisfied with a larger probability than a_2. Thus, a measure of performance that would be specifically tailored to $A = \{a_1, a_2\}$ would not need be defined for all consistent collections in $[0,1]^3$.

4.2 Normative desirability of axioms and sets of axioms

As axioms most often reflect normative principles that matter to different extents, a key additional element for this evaluation needs to be introduced. More precisely, not only can an axiom be more valuable in the eyes of a researcher or a designer than another one, but *synergies* are likely to emerge in the combination of axioms. For instance, in the microeconomic allocation framework, satisfying "efficiency" may be more or less valued than satisfying "no-envy", and, furthermore, the value of satisfying another fairness criterion such as "egalitarian equivalence", given the satisfaction of "no-envy", may be reduced, so that, equivalently, it may become more desirable to satisfy the efficiency condition. In this perspective, the example of a non-manipulability axiom such as "strategy-proofness" is also highly instructive. Indeed, the likelihood of truthful revelation is all the more important as other axioms involving requirements on preferences are satisfied[20], and, conversely, the satisfaction of these other axioms is all the more valuable that it is likely that they are applied to the actual —truthfully revealed— preferences.

This observation leads us to allow the *intrinsic valuation* of a non-empty combination $S \subseteq A$ of the axioms to differ from the sum of the intrinsic valuations of axioms in S. As a consequence, we define the **set of possible intrinsic valuations** as the **set of capacities** on A:

$$U = \left\{ (u_S)_{\emptyset \neq S \subseteq A} \in \mathbb{R}_+^{2^J-1}, u_T \leq u_S \text{ if } T \subset S, \text{ for all } S \subseteq A \right\}.$$[21]

[20] The satisfaction of "strategy-proofness" by itself may, of course, still be appreciated as, for example, it can be interpreted as preventing agents with lower ability to compute optimal actions from being disadvantaged (this interpretation has played an important role in the school choice literature (*e.g.* Artemov et al. (2017)), but the interest of this axiom mainly lies in its interaction with other axioms.

[21] *We abuse language here as a capacity on set A, standardly, is also defined for the empty set, for which it returns value* 0. The set U would be appropriately referred to as a projection on $\mathbb{R}_+^{2^J-1}$ of the set of capacities. We however omit this qualification: an element of U is called a capacity.

By U_{st} we denote the set of strictly monotonic capacities.[22]

The weak monotonicity assumption, with respect to inclusion, embedded in the use of capacities, can be interpreted as meaning that all axioms under consideration are normatively desirable. **Super-additive capacities** on A are of special interest for our analysis. Let

$$U_{s.a} = \Big\{u \in U, u_S \geq u_T + u_{T'} \text{ if } T \cup T' = S \text{ and } T \cap T' = \emptyset \text{ , for all } S \subseteq A\Big\}.$$

A super-additive intrinsic valuation is interpreted as the result of *complementarities* between all the considered axioms and is, for example, well suited to account for the interaction between “strategy-proofness”, “efficiency” and “‘no-envy” as suggested above. From a general point of view, we see the use of super-additive valuations as the one most adequate to the typical problems studied in normative economics where the considered axioms are particular formulations of general and independent principles.[23] We say that axioms in set A are **complementary** when we consider a super-additive capacity on A.[24]

Remark: Clearly, we have a cardinal interpretation of intrinsic valuations: capacities express the intensity of preferences between combinations of axioms. This is key, for example, to capture complementarity through super-additivity. Note however that the rich information offered by the use of capacities in this framework might be partly, or completely, disregarded. One can restrict attention to capacities that only depend on the number of axioms that are satisfied: capacities of the form $u : S \in 2^A \setminus \{\emptyset\} \mapsto g(|S|) \in \mathbb{R}_+$, with $g : \mathbb{R}_+ \to \mathbb{R}_+$ non-decreasing.[25] Such capacities may reflect the view that *all* the axioms in A are complementary, g being convex, or that they all are substitutes, g being concave, but they imply neutrality across combinations of axioms of the same size. One could refer to these cases as *uniform complementarity* and *uniform substitutability*. *Agnosticism* with respect to axioms can be captured through the use of the capacity $u : S \in 2^A \setminus \{\emptyset\} \mapsto |S| \in \mathbb{R}_+$.[26]

[22] That is, $U_{st} = \Big\{(u_S)_{\emptyset \neq S \subseteq A} \in \mathbb{R}_+^{2^J - 1}, u_T < u_S \text{ if } T \subset S, \text{ for all } S \subseteq A\Big\}$.

[23] In the example we gave involving “no-envy” and “egalitarian equivalence”, though, a super-additive valuation would not capture the effect we described.

[24] Axioms in set A are substitutes if the studied capacity is sub-additive. More generally, a capacity reflects complementarities between the axioms of set $T \subseteq A$ if for all disjoint $S, S' \subset T$, $u_T \geq u_S + u_{S'}$, and substitutability if the reverse inequality holds.

[25] For any set B, $|B|$ denotes the cardinality of B.

[26] The analysis would remain the same if we were to restrict attention to the set of normalised capacities, *i.e.* the set of capacities such that A is given value 1.

4.3 Characterisation of the measure

We have introduced the key ingredients for the construction of a performance measure, namely, the degrees of satisfaction, defined through probabilities, and the intrinsic valuations, defined through a capacity.

A **(performance) measure** is a mapping:

$$m : U \times P \to \mathbb{R}_+,$$

satisfying the following normalisation condition: for all $p \in P$,

$$m(\mathbf{0}, p) = 0.$$

This natural property posits that the degrees to which a given rule satisfies the axioms in A do not matter if these axioms are irrelevant to the decision maker —note that a measure only takes non-negative values.

Given a pair of an intrinsic valuation and a collection of probabilities, the most intuitive measure for the performance of a rule associated with the collection arguably consists in taking the standard weighted sum:

$$\begin{aligned} \tilde{m} : U \times P &\to \mathbb{R} \\ (u, p) &\mapsto \sum_{\emptyset \neq S \subseteq A} u_S p_S. \end{aligned}$$

However, considering a capacity that positively depends on the cardinality of combinations, one can see that such a measure *double counts* the satisfaction of some sets of axioms.

Example 3. Let $A = \{a_1, a_2, a_3\}$ and consider:

	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
$\mathbf{u}$	1	1	1	3	3	3	6 .
$\mathbf{p}$	0.7	0.8	0.5	0.7	0.25	0.3	0.25

As $p_{a_1} = p_{a_1a_2}$, the rule associated with p satisfies a_1 with exactly the same probability as it satisfies a combination of a_1 and another axiom while the intrinsic valuation of this combination incorporates the intrinsic valuation of a_1. Then, it is questionable that a_1 should have an impact on the measure under p, as is the case with $\tilde{m}$. *As a consequence, in our axiomatic approach, we will look for measures taking into account the difference* *between the probability with which a rule satisfies a given combination of axioms and the probability with which it satisfies any superset of it.*

We first introduce a measure that would not double count the satisfaction of a_1 in Example 3. For all non-empty $S \subset A$, let $\hat{p}_S = \max_{T:S\subset T} \{p_T\}$ and let $\hat{p}_A = 0$. The value $\hat{p}_S$ gives the maximal probability associated with a superset of S. Consider the following performance measure:

$$\begin{aligned} \hat{m} : U \times P &\to \mathbb{R}_+ \\ (u, p) &\mapsto \sum_{\emptyset \neq S \subseteq A} u_S(p_S - \hat{p}_S). \end{aligned}$$

We will call $\hat{m}$ the “weighted minimal difference measure”.

The redundancy we identified in the way the standard weighted sum $\tilde{m}$ is computed is not merely a cardinal anomaly in the sense that it impacts the ranking of rules: in Example 4 below, a rule associated with p' performs better than a rule associated with p according to the measure $\hat{m}$, which does not double-count in this example either[27], while the standard weighted sum $\tilde{m}$ yields the opposite conclusion. Note that u only depends on the cardinality of combinations.

Example 4. Let $A = \{a_1, a_2, a_3\}$ and consider:

	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
$\mathbf{u}$	1	1	1	5	5	5	15
$\mathbf{p}$	0.7	0.7	0.7	0.6	0.6	0.6	0.6
$\mathbf{p}'$	1	1	0.45	1	0.45	0.45	0.45

One has $\tilde{m}(u, p) = 20.1$, $\tilde{m}(u, p') = 18.7$, while $\hat{m}(u, p) = 9.3$ and $\hat{m}(u, p') = 9.5$.

We now identify another desirable property of a performance measure —which is not satisfied by the standard weighted sum $\tilde{m}$.

Let us place ourselves in the case where A is a singleton. In this *single-axiom case* where the considered rule satisfies axiom a with probability p_a, while a is given valuation u_a, it is fair to say that, given the cardinal interpretation of intrinsic valuations, the natural way to measure the performance of this rule is to take the “expected valuation” $p_a u_a$. However,

[27] The value $\hat{m}(u, p)$ is computed as a weighted sum, so that a given set of axioms impacts this value if and only if it is given a non-null weight. Yet, in this example, all the sets which are satisfied with the same probability as one of their supersets are given weight 0.

$\tilde{m}$ does not constitute an appropriate generalisation to multiple axioms of this natural single-axiom measure.

Let $\lambda \in [0,1]$, $\emptyset \neq S \subseteq A$, and $p^{\lambda,S} \in P$ defined by:

$$\begin{aligned} p_T^{\lambda,S} &= \lambda \text{ if } \emptyset \neq T \subseteq S, \\ p_T^{\lambda,S} &= 0 \text{ otherwise.} \end{aligned}$$

The collection $p^{\lambda,S}$ lies on the edge of P between the extreme points $\mathbf{0}$ and $p^{1,S}$. Here is an illustration of $p = p^{0.6,a_1a_3}$, when $A = \{a_1, a_2, a_3\}$:

$$\begin{array}{cccccccc} & a_1 & a_2 & a_3 & a_1a_2 & a_1a_3 & a_2a_3 & A \\ \mathbf{p} & 0.6 & 0 & 0.6 & 0 & 0.6 & 0 & 0 \end{array}.$$

A rule with degrees of satisfaction given by $p^{\lambda,S}$ satisfies any combination of axioms which is not included in S with probability 0. In addition, it satisfies S with exactly the same probability, λ, as it satisfies all the elements of S. As a consequence, we claim that the problem of measuring the performance of this rule has the same structure as a single-axiom problem in which the considered axiom is satisfied with probability λ and is given valuation u_S. *Hence, a measure that would provide a consistent generalisation to multiple axioms of the natural "expected valuation" defined for the single-axiom case, would, in contrast to $\tilde{m}$, return, for any $(u, p^{\lambda,S}) \in U \times P$, the image $p_S u_S = \lambda u_S$.*[28]

While the observation that a measure should *count once and only once* the satisfaction of a given subset of axioms requires some additional work in order to translate into a mathematical principle, this second observation immediately yields the following requirement:

Expected valuation for single-axiom-reducible problems

Let $\lambda \in [0,1]$, $\emptyset \neq S \subseteq A$.

Let $u \in U$. Then,

$$m(u, p^{\lambda,S}) = \lambda u_S.$$

For example, the weighted minimal difference measure $\hat{m}$ satisfies this property. So does $(u,p) \mapsto \max_{\emptyset \neq S \subseteq A} u_S p_S$.

[28] Of course, this already prevents some form of double-counting: when the sets of axioms that the rule satisfies with *the same positive probability* form a chain, then, only the satisfaction of the maximal set of this chain should be taken into account in the measure of the performance.

In the discussion of Example 3 and the standard weighted sum $\tilde{m}$, we identified an anomaly of double-counting: because there is a superset of a_1 that the rule satisfies with the same probability as the one with which it satisfies a_1, there is no increment, in terms of probability, induced by focusing on the satisfaction of axiom a_1 only, and thus, a_1 should have no impact on the value $m(u,p)$. The fact that P is $(2^J-1)-$dimensional convex polytope (Lemma 1) is important in providing a definition of the increment in probability associated with a non-empty set S under collection p, covering the case in which $p_S > p_T$ for any superset T of S.

For all $p \in P \setminus \{\mathbf{0}\}$, there exists a unique *pair $(\mathcal{I}^p, (\alpha_T^p)_{\emptyset\neq T\subseteq A})$, where $\mathcal{I}^p$ is a non-empty family of non-empty subsets of A, and $(\alpha_T^p)_{\emptyset\neq T\subseteq A}$ a family of real numbers, such that:*

- $0 < \alpha_T^p \leq 1$ *for all $T \in \mathcal{I}^p$ and* $\sum_{T\in\mathcal{I}^p} \alpha_T^p \leq 1$,
- $\alpha_T^p = 0$ *for all $\emptyset \neq T \in 2^A \setminus \mathcal{I}^p$, and*

$$p = \sum_{T\in\mathcal{I}^p} \alpha_T^p p^{1,T} \quad \Big(= \sum_{\emptyset\neq T\subseteq A} \alpha_T^p p^{1,T} + (1 - \sum_{T\in\mathcal{I}^p} \alpha_T^p)\mathbf{0}\Big).$$

We write the trivial equality in parenthesis above in order to stress that, in general, the sum $\sum_{T\in\mathcal{I}^p} \alpha_T^p$ is not one.

For the collection $\mathbf{0}$, we allow for the associated family of subsets to be empty and write $\mathcal{I}^{\mathbf{0}} = \emptyset$.

Importantly, there is a generic procedure enabling one to determine, for any $p \in P$, $(\mathcal{I}^p, (\alpha_T^p)_{\emptyset\neq T\subseteq A})$. *The computation of $(\alpha_T^p)_{\emptyset\neq T\subseteq A}$ involves a recursive equation that makes clear that α_T^p gives the increment we described above.* Let $p \in P$; we proceed inductively:

Step 1. If $p_A > 0$, set $\mathcal{I}_1^p = \{A\}$ and $\alpha_A^p = p_A$, otherwise, set $\mathcal{I}_1^p = \emptyset$ and $\alpha_A^p = 0$;

Step k (for $\mathbf{2 \leq k \leq J-1}$). Set $\mathcal{I}_k^p = \mathcal{I}_{k-1}^p \cup \{T \subseteq A$ with $|T| = J-k$ and $p_T - \sum_{S:T\subset S} \alpha_S^p > 0\}$, and, for all $T \subseteq A$ with $|T| = J-k$, $\alpha_T^p = p_T - \sum_{S:T\subset S} \alpha_S^p$.

Define $\mathcal{I}^{\mathbf{p}} = \mathcal{I}_{\mathbf{J-1}}^{\mathbf{p}}$.

Example 5. Let us illustrate the computation of $\alpha^p = (\alpha_T^p)_{\emptyset\neq T\subseteq A}$, with $A = \{a_1, a_2, a_3\}$:

	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
$\mathbf{p}$	0.7	0.8	0.5	0.7	0.25	0.3	0.25.
$\boldsymbol{\alpha^p}$	0	0.05	0.2	0.45	0	0.05	0.25

Since $\alpha^p_A = p_A$ for all $p \in P$, one sees from the recursive equation $\alpha^p_T = p_T - \sum_{S:T\subset S} \alpha^p_S$ that α^p_T gives the increment in probability, induced by focusing on the satisfaction of axioms in T, rather than focusing on axioms in T together with additional axioms. In other words, α^p_T gives the probability that the rule associated with p satisfies all the axioms in T and no other axiom. That is why we refer to α^p_T as the **contribution** of set T under (the collection of probabilities) p.

Finally, this recursive definition implies that α^p is the solution of a *Möbius inversion problem* (see Theorem 1 just below).

Motivated by the discussion of Example 3, we require that for all $p \in P$, and for all $u \in U$, the impact of $\emptyset \neq T \subseteq A$ on a measure m, *given valuation* u, depend on the contribution of T under p:

Same contribution—same impact

Let $p, p' \in P$ and $\emptyset \neq S \subseteq A$ be such that $\alpha^p_S = \alpha^{p'}_S$.
Let $u \in U$ and $u^S \in U$ be such that $u^S_T = u_T$ for all $T \neq S$.
Then,

$$m(u^S, p) - m(u, p) = m(u^S, p') - m(u, p').$$

Valuations u and u^S above only differ, potentially, in their component associated with the combination S. Then, the differences above measure the impact of S on measure m under p and the impact of S under p', respectively, *given valuation* u*, when one changes* u_S *to* u^S_S. This principle does not imply that the impact of an axiom under p be the same for any $u \in U$.

We are now able to present the characterisation of a unique performance measure, under the additional requirement that the projection of the measure on U be a continuous mapping, where U is endowed with the usual induced topology from $\mathbb{R}_+^{2^J-1}$.[29] We simply write that the measure is continuous on U.

Theorem 1. *A performance measure* $m : U \times P \to \mathbb{R}_+$*, continuous on* U*, satisfies*

- *Same contribution—same impact, and*
- *Expected valuation for single-axiom-reducible problems*

[29] For all $p \in P$, $m^p : u \in U \mapsto m(u, p) \in \mathbb{R}_+$ is a continuous mapping.

if and only if it is the ***weighted Möbius performance measure****,* $\ddot{m}$*:*

$$\begin{aligned} \ddot{m} : U \times P &\to \mathbb{R}_+ \\ (u,p) &\mapsto \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big). \end{aligned}$$

The name of the characterised measure comes from the fact that the function associating $(p_S)_{\emptyset \neq S \subseteq A}$ with $\Big(\sum_{T:S\subseteq T} (-1)^{|T\setminus S|} p_T \Big)_{\emptyset \neq S \subseteq A}$ is the *Möbius transform of the set function* $p : 2^A \setminus \emptyset \to [0,1]$ for the partial order $\geq$ on $2^A \setminus \emptyset$ such that:

$$\text{for all } S, T, \;\; S \geq T \iff S \subseteq T.$$

The reader is referred to the appendix and to Grabisch (2016) (Chapter 2) for more details on the definition of the Möbius transform of a set function, given an arbitrary partial order defined on a (finite) set.[30]

Remark: Let us insist on the precise role of the continuity assumption. If a performance measure $m : U \times P \to \mathbb{R}_+$ satisfies *same contribution—same impact* and *expected valuation for single-axiom-reducible problems*, then, for all $p \in P$, and all $u \in U_{st}$,

$$m(u,p) = \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big) + b^p, \text{ for some } b^p \in \mathbb{R},$$

where $b^{p^{\lambda,S}} = 0$ for all $\lambda \in [0,1]$, and all non-empty $S \subseteq A$.

We conclude this section by illustrating with a comparison between the weighted Möbius performance measure $\ddot{m}$ and the measure $\hat{m} : (u,p) \mapsto \sum_{\emptyset \neq S \subseteq A} u_S(p_S - \hat{p}_S)$, illustrating in particular why $\ddot{m}$ properly accounts for the intrinsic valuation associated with any non-empty subset of A, while $\hat{m}$ does not.

Example 6. Let $A = \{a_1, a_2, a_3\}$; we let $w^{\hat{m}}$ and $w^{\ddot{m}}$ denote the collections of weights

[30] For all $(u,p) \in U \times P$, $m(u,p) = \sum_{\emptyset \neq S \subseteq A} p_S \Big(\sum_{T \subseteq S} (-1)^{|S \setminus T|} u_T \Big)$. The term in parenthesis corresponds to the Möbius transform, for the usual partial order defined by set inclusion, of a capacity $(u_S)_{S \subseteq A} \in \mathbb{R}^{2^J}$, returning 0 for the empty set.

according to $\hat{m}$ and to $\ddot{m}$, respectively:

$$\begin{array}{cccccccc}
 & a_1 & a_2 & a_3 & a_1a_2 & a_1a_3 & a_2a_3 & A \\
\mathbf{p} & 0.55 & 0.6 & 0.2 & 0.35 & 0.05 & 0.15 & 0 \\
\mathbf{w}^{\hat{\mathbf{m}}} & 0.2 & 0.25 & 0.05 & 0.35 & 0.05 & 0.15 & 0 \\
\mathbf{w}^{\ddot{\mathbf{m}}} & 0.15 & 0.1 & 0 & 0.35 & 0.05 & 0.15 & 0
\end{array}.$$

Let u be a capacity defined on A. The two measures return the same weight for any subset of cardinality at least 2. Let us focus on axiom a_3. In weighting u_{a_3} by $0.2 - 0.05 - 0.15 = 0$, $\ddot{m}$ takes into account the fact that, by the monotonicity of u, in giving the weight 0.05 to $u_{a_1a_3}$, and 0.15 to $u_{a_2a_3}$, a portion $0.15+0.05$ of u_{a_3} has already been incorporated in the measure of the performance of the rule. By focusing on the probability of satisfaction of a single superset of a_3 ($\hat{m}$ violates *same contribution—same impact*), namely a_2a_3, $\hat{m}$ *overweights* the valuation of a_3.[31]

The way to measure the performance of a rule was one of the two natural questions raised when considering the collection of the degrees to which it satisfies all combinations of axioms. The second one pertains to measuring the (in)compatibility of axioms, given such a collection.

5 Where does the incompatibility come from?

A collection of probabilities associated with a specific rule may be used in order to analyse how compatible the considered principles are under this rule. Indeed, such a collection indicates how likely the rule is to satisfy a certain axiom, given that it satisfies any subset of other axioms, which enables to better understand its behaviour.

Under this interpretation, as in the previous section, a collection $p \in P$ is associated with a single rule. We see however two additional ways to interpret such a collection. First, p can be obtained as a "summary collection" of multiple collections associated with different rules. As a simple illustration, p may indicate the probability that *all the rules* belonging to a given family satisfy the combinations of axioms in A.

Furthermore, in the case of several punctual axioms, one could inquire about the *existence* of outcomes having all the required properties, rather than about the *selection*,

[31] The reason why $\hat{m}$ did not overweight the valuation of a_1 in Example 3, and did not overweight the valuation of any subset for p' in Example 4, is that for all these subsets, there existed a superset with the same probability of satisfaction: in such cases, the weighting formula of $\hat{m}$ equals that of the Möbius performance measure.

by a specific rule, of such an outcome. Using the notation, and following the reasoning, of Section 3, this amounts to considering, for any non-empty $S \subseteq A$, $D(S) = \{i \in I \text{ such that } \bigcap_{a\in S} O_i^a \neq \emptyset\}$. Then, considering a $\sigma-$algebra ξ on I, such that, for all $\emptyset \neq S \subseteq A$, $D(S)$ is measurable, and μ a probability measure on (I, ξ), we define $p_S = \mu(D(S))$. The set of consistent collections of probabilities can then be described as in Section 4.1.[32]

We address the question of how axioms interact, given a collection, without favouring any of these interpretations.

A first step towards the answer consists in building a measure of how compatible an axiom $a \in A$ is with the other axioms under p, and this can be done by determining how a contributes to the **overall degree of incompatibility** $1 - p_A$, compared to the other axioms. This question is akin to the general purpose of cooperative game theory where one tries to determine ways to allocate the benefits or costs of cooperation/interaction among a given set of agents.[33]

In order to formalise this connection, define

$$P^* = \Big\{(1, (p_S)_{\emptyset\neq S\subseteq A}), (p_S)_{\emptyset\neq S\subseteq A} \in P\Big\}.$$

Each element of P is associated with one and only one element of P^*; hence a generic element of P^* will also be denoted by p to alleviate notation, and we will sometimes write p as $(p_S)_{S\subseteq A}$.[34] The notation $\mathbf{0}$ will now represent the null vector in $\mathbb{R}^{2^J}$.

Define $V = 1 - P^* = \{1 - p,\ p \in P^*\}$, and let $v \in V$. Note that $v_\emptyset = 1 - p_\emptyset = 0$. Thus, *$v$ is a cooperative game associated with the set of axioms A*. The number $v_A = 1 - p_A$ represents the overall degree of violation of axioms in A by a rule associated with p, and this magnitude must be distributed among them.

All definitions below could be equivalently formulated as a requirement on V or on P^*. In order to be consistent with the previous sections, we choose to write definitions on P^*.

An **incompatibility measure** is a mapping:

$$\psi :\ p \in P^* \mapsto (\psi_a(p))_{a\in A} \in \mathbb{R}^J.$$

[32] Under the previous definition, in the appendix (Section 8.1), any $\emptyset \neq S \subseteq A$ is associated with a logical sentence whose truth value stands for the satisfaction, by a given rule, of all the axioms in S. Now, for each $\emptyset \neq S \subseteq A$, take the logical sentence $\tilde{S}$ whose truth value stands for the non-emptyness of $\bigcap_{a\in S} O_i^a$; the analysis of Section 8.1 goes through.

[33] Related to our approach is the literature focusing on the use of tools of cooperative game theory in feature attribution problems (Lundberg and Lee (2017)).

[34] In words, each element of P^* is obtained by choosing a unique consistent collection defined for all non-empty combination of axioms, and by adjoining to it the value 1 for the empty-set, which should be though of as being always satisfied.

For all $a \in A$, ψ_a gives a measure of the incompatibility of a with the axioms in $A \setminus a$.

The connection with cooperative game theory draws attention to specific incompatibility measures. For example, the **Banzhaf incompatibility measure** is defined by:

$$\phi_a(p) = \sum_{S \subseteq A \setminus a} \frac{1}{2^{J-1}} (p_S - p_{S \cup a}), \text{ for all } p \in P^*, \text{ and all } a \in A.$$

The **Shapley incompatibility measure** is defined by:

$$\varphi_a(p) = \sum_{S \subseteq A \setminus a} \frac{|S|!\,(J - |S| - 1)!}{J!} (p_S - p_{S \cup a}), \text{ for all } p \in P^*, \text{ and all } a \in A.$$

The Banzhaf and the Shapley measure differ in the weight associated with $p_S - p_{S \cup a}$. Let us briefly recall the standard interpretation of this weight for both measures. The Banzhaf measure for axiom a weights the value $p_S - p_{S \cup a}$ by the probability that combination S form in a scenario where all combinations are equally likely to form. The Shapley measure of axiom a is constructed by weighting the value $p_S - p_{S \cup a}$ by the probability that the axioms in S, and only the axioms in S, arrive before a in a scenario where axioms arrive one by one, according to a uniformly random permutation of A.

For $a \in A$, and $S \subseteq A \setminus a$, the difference $p_S - p_{S \cup a}$ interprets as the cost in probability of satisfaction that a exerts on S. As a consequence, it is natural to require that the incompatibility measure associated with a be a function of the cost exerted by a on all $S \subseteq A \setminus a$:

Same cost—same incompatibility

Let $p, p' \in P^*$ and $a \in A$ be such that, for all $S \subseteq A \setminus a$, $p_S - p_{S \cup a} = p'_S - p'_{S \cup a}$. Then,

$$\psi_a(p) = \psi_a(p').$$

This principle corresponds to the invariance property implied by Young's *strong monotonicity* axiom in his classical characterisation of the Shapley value on the subspace of games associated with a fixed group of players (Young (1985)). The Banzhaf and the Shapley incompatibility measures satisfy this principle. According to both measures, the greater the value, the greater the incompatibility.

Two classical properties are necessary for ψ to be a relevant measure in the problem we consider. The first one makes it possible to interpret ψ as allocating the incompatibility

$1 - p_A$ among axioms in A.

Allocation of incompatibility
Let $p \in P^*$;

$$\sum_{a \in A} \psi_a(p) = 1 - p_A.$$

This condition corresponds to the standard efficiency principle in cooperative game theory.

The last requirement states that the evaluation should not be biased towards any axiom:

Symmetry
Let $p \in P$ and $\pi : A \to A$, a permutation.
Let p^π the collection defined by $p^\pi_S = p_{(\pi(a))_{a \in S}}$. Then,

$$\psi_a(p) = \psi_{\pi(a)}(p^\pi).$$

These three principles single out the Shapley incompatibility measure:

Theorem 2. *An incompatibility measure $\psi : P^* \to \mathbb{R}^J$ satisfies*

- *Same cost—same incompatibility,*
- *Allocation of incompatibility, and*
- *Symmetry*

if and only if it coincides with the Shapley incompatibility measure: for all $a \in A$,

$$\begin{aligned} \psi_a : P^* &\to \mathbb{R}^J \\ p &\mapsto \sum_{S \subseteq A \setminus a} \frac{|S|!\,(J - |S| - 1)!}{J!} (p_S - p_{S \cup a}). \end{aligned}$$

Remark: The reader will not be surprised that the Shapley incompatibility measure satisfies these three properties, given Young's axiomatisation of the Shapley value (Young

(1985)) on the entire set of games $G = \{u = (u_S)_{S \subseteq A} \in \mathbb{R}^{2^J}, u_\emptyset = 0\}$; however the fact that it is actually characterised by them is not immediate. Indeed, P^* is defined by specific consistency conditions, and it turns out that the family of "unanimity games", used in his proof, does not belong to $V = 1 - P^*$.

Example 7. Let $A = \{a_1, a_2, a_3\}$ and consider the "unanimity game" $u^{a_1a_3}$ defined by $u_T^{a_1a_3} = 1$ if $a_1a_3 \subseteq T$, $u_T^{a_1a_3} = 0$ otherwise. Then $p = 1 - u^{a_1a_3}$ is such that $p_{a_1} = p_{a_3} = 1$ but $p_{a_1a_3} = 0$, that is $p \notin P^*$, and thus $u^{a_1a_3} \notin V$.

However, we build on the fact that each $v \in V$ can be written as a convex combination of points in V —the extreme points of V are identified thanks to the proof of Lemma 1— and build on an induction argument that is analogous to the one that Young proposes.

The Shapley measure is also characterised by *allocation of incompatibility*, *symmetry*, and direct adaptations of the classical "null-player" and "additivity and positive homogeneity" axioms —the reader may find their explicit definitions in the appendix (see Theorem 3). However, *same cost—same incompatibility*, *allocation of incompatibility* and *symmetry* are in our view the very axioms that support the interpretation of ψ as an adequate measure.

The following proposition draws a connection between the Shapley incompatibility measure and the Möbius transform that we used to build a measure of the performance of a rule.

Proposition 1. *Let $a \in A$ and $p \in P^*$. Then,*

$$\varphi_a(p) = \sum_{S \subseteq A : a \notin S} \frac{\sum_{T:S \subseteq T} (-1)^{|T \setminus S|} p_T}{|A \setminus S|}.$$

The numerator of the summand is given by the Möbius transform of the set function $p : 2^A \to [0, 1]$ for the partial order $\geq$ on 2^A such that:

$$\text{for all } S, T, \;\; S \geq T \iff S \subseteq T.$$

The only difference with the previous section, due to the fact that we consider p in P^* and not in P, is that this transform must be defined for the empty set. In line with the previous section, let

$$\begin{aligned}\alpha^* : P^* &\to \mathbb{R}^{2^J}\\ (p_S)_{S\subseteq A} &\mapsto \Big(\sum_{T:S\subseteq T}(-1)^{|T\setminus S|}p_T\Big)_{S\subseteq A},\end{aligned}$$

and α^{*p} denote $\alpha^*(p)$. That is, for all $\emptyset \neq S \subseteq A$, $\alpha^{*p}_S = \alpha^p_S$, and $\alpha^{*p}_\emptyset = 1 - \sum_{\emptyset\neq S\subseteq A}\alpha^p_S$. For $S \subseteq A$ and $p \in P^*$, the value α^{*p}_S gives the probability of satisfying all the axioms in S and only the axioms in S. Thus, from the point of view of an incompatibility measure, $\alpha^{p^*}_S$ represents what one loses by considering the other axioms. Then, the Shapley measure allocates this loss equally between these other axioms: each $a \in A \setminus S$ is allocated an incompatibility value of $\frac{\alpha^{*p}}{|A\setminus S|}$. A similar formula is known for the Shapley value on games.[35]

6 Discussion: robustness with respect to the collection of probabilities

One of the motivations for the use of a notion of degree defined as a probability of satisfaction was to extend a natural partial order between rules. According to it, a rule $f : I \to 2^O$ performs better than a rule $f' : I \to 2^O$, when the set of axioms under consideration is A, if and only if, for all non-empty $S \subseteq A$,

$$\begin{aligned}&\Big\{(i_1, ..., i_{K^A}) \text{ such that } (i_1, ..., i_{K^a}) \in D^{f'}(a) \text{ for all } a \in S\Big\}\\ &\subseteq \\ &\Big\{(i_1, ..., i_{K^A}) \text{ such that } (i_1, ..., i_{K^a}) \in D^{f}(a) \text{ for all } a \in S\Big\},\end{aligned} \tag{1}$$

that is, for all possible combinations S, the set of (tuples of) instances for which f' satisfies all the axioms in S is included in the set of (tuples of) instances for which f does. Allowing combinations of axioms to matter to a different extent in the evaluation of rules through the use of set functions that are monotonic with respect to inclusion, we have characterised a performance measure $\ddot{m} : U \times P \to \mathbb{R}_+$ inducing a *completion* of this partial order.

Fix $u \in U$ throughout this section, if condition (1) holds for f and f', then, *for any probability measure on* (I^{K^A}, ξ^{K^A}) *on the basis of which* p^f *and* $p^{f'}$ *are computed*, $\ddot{m}(u, p^f) \geq \ddot{m}(u, p^{f'})$. However, when condition (1) does not hold, there may of course be probability

[35] For all $u \in G$, all $a \in A$, the Shapley value associated with a for game u is given by $\sum_{S\subseteq A: a\in S} \frac{\sum_{T\subseteq S}(-1)^{|S\setminus T|}u_T}{|S|}$ (see Grabisch (2016)).

measures on (I^{KA}, ξ^{KA}) for which the induced p^f and $p^{f'}$ yield $\dddot{m}(u,p^f) \geq \dddot{m}(u,p^{f'})$ and others for which the induced p^f and $p^{f'}$ yield $\dddot{m}(u,p^f) \leq \dddot{m}(u,p^{f'})$. This is a challenge when several such probability measures must be integrated into the analysis —in particular when there is ambiguity on the measure that should be used to compute collections of probabilities.

Two ways of dealing with such a variability stand out. For exposition purposes, we will describe them under the assumption that only finite sets of probability measures on (I^{KA}, ξ^{KA}) are considered. A finite set of such probability measures is generically denoted by $\Delta = \{\mu_1, ..., \mu_K\}$. Then, given such a set, a rule f is associated with a finite set of collections of probabilities $\Pi^f_\Delta = \{p^f_1, ..., p^f_K\}$.[36] The set of finite sets of collections is denoted by $\mathcal{P}$, and a generic element of it is denoted by $\Pi = \{p_1, ..., p_H\}$.

The first way consists in *selecting* a collection $p \in P$ supposed to represent or summarise the collections under consideration. Given a mapping $\kappa : \mathcal{P} \to P$, and given Δ, one concludes that the rule f performs better than the rule f' if and only if $\dddot{m}\left(u, \kappa(\Pi^f_\Delta)\right) \geq \dddot{m}\left(u, \kappa(\Pi^{f'}_\Delta)\right)$. This order is always complete and κ should be such that it extends the partial order defined by condition (1). It is for instance the case with the following mapping:

$$\kappa : \Pi = \{p_1, ..., p_H\} \mapsto \sum_{h=1,...,H} \beta^\Pi_h p_h,$$

with $\beta^\Pi_h \geq 0$ and $\sum_{h=1,...,H} \beta^\Pi_h = 1$.

The other method consists in comparing two rules f and f', given Δ, on the basis of the sets of values taken by the performance measure as p^f and $p^{f'}$ vary in Π^f_Δ and $\Pi^{f'}_\Delta$, respectively. In that perspective, let us give several (not necessarily complete) standard criteria (see Fishburn (1985), Bewley (2002), Echenique et al. (2022), Bardier et al. (2024)).

***α*-maxmin criterion:** Let $\alpha \in [0,1]$. The rule f performs better than the rule f' if and only if:

$$\begin{aligned}&\alpha\Big(\max_{k=1,...,K} \dddot{m}(u,p^f_k)\Big) + (1-\alpha)\Big(\min_{k=1,...,K} \dddot{m}(u,p^f_k)\Big)\\ &\geq\\ &\alpha\Big(\max_{k=1,...,K} \dddot{m}(u,p^{f'}_k)\Big) + (1-\alpha)\Big(\min_{k=1,...,K} \dddot{m}(u,p^{f'}_k)\Big).\end{aligned}$$

This criterion is complete and consistent with the partial order defined by condition (1) for

[36] One has $p^f_k = \left(\mu_k\left(\left\{(i_1, ..., i_{K^A}) \text{ such that } (i_1, ..., i_{K^a}) \in D^f(a) \text{ for all } a \in S\right\}\right)\right)_{\emptyset \neq S \subseteq A}$, for all $k = 1, ..., K$.

any value of $\alpha \in [0,1]$. When $\alpha = 0$, the criterion focuses on a worst-case analysis, and, when $\alpha = 1$, on a best-case analysis.

All the criteria described so far induce a complete order between rules, but it is arguable that for comparisons to be robust, when facing different possible probability measures, one must account for the possibility that there be no "sufficient evidence" for two rules to be compared. The following criteria capture this idea.

Max-and-min criterion: The rule f performs better than the rule f' if and only if:

$$\begin{cases} \max_{k=1,\dots,K} \ddot{m}(u, p_k^f) \geq \max_{k=1,\dots,K} \ddot{m}(u, p_k^{f'}) \\ \min_{k=1,\dots,K} \ddot{m}(u, p_k^f) \geq \min_{k=1,\dots,K} \ddot{m}(u, p_k^{f'}) \end{cases}.$$

Point-wise criterion: The rule f performs better than the rule f' if and only if:

$$\text{for all } k = 1, \dots, K,\ \ddot{m}(u, p_k^f) \geq m(u, p_k^{f'}).$$

Min-vs-max criterion: The rule f performs better than the rule f' if and only if:

$$\min_{k=1,\dots,K} \ddot{m}(u, p_k^f) \geq \max_{k=1,\dots,K} \ddot{m}(u, p_k^{f'}).$$

The *max-and-min criterion* and the *point-wise criterion* are consistent with the partial order defined by condition (1). While extending this partial order was one of the main motivations of our approach, the *min-vs-max criterion* is not consistent with it. More generally, there is a classical trade-off for these criteria between their degree of incompleteness and the conviction one can have in the comparisons they express —given Δ, the *max-and-min criterion* extends the *point-wise criterion*, which extends the *min-vs-max criterion.*

The primary insight from this section is that once we have identified, through our axiomatization, the measure to use for a single collection of probabilities, numerous standard methods, of which we simply provided a few examples, can naturally be combined with the weighted Möbius performance measure in order to carry out a robust analysis.

7 Conclusion

We defined a general notion for the degree to which a rule satisfies a set of axioms. Armed with it, we proposed and characterised **(1)** a unique criterion to evaluate the performance of rules, taking into account the normative desirability of axioms and, crucially, that of their combinations, and **(2)** a unique criterion to determine, for a given collection of probabilities of satisfaction, the role of each axiom in the overall degree of violation.

Let us further elaborate on the operational illustration we pointed out at the beginning of this work. A policy maker in charge of selecting a mechanism to match students with schools, must distinguish between (variants of) the "Deferred Acceptance" (DA), the "Top Trading Cycle" (TTC) and the "Immediate Acceptance" (IA) rules, on the basis of "strategy-proofness" (for students), "efficiency" and "stability", which are incompatible, even in the standard case of strict preferences and priorities. She then asks a team of researchers to estimate, for each rule, how probable the satisfaction of each combination of principles is. The research team inquires about the relative importance of each combination of principles for the policy maker. Then, *the weighted Möbius performance measure* adequately integrate these two piece of information in order to measure and compare the performance of DA, TTC and IA. Each of these rules satisfies certain combinations of these properties with degree 1. In addition, some constrained-optimality results have been obtained —we refer the reader to Abdulkadiroğlu and Andersson (2023) for a presentation of such results. However, the families of rules over which these results hold do not include all (variants of) the three rules, and thus do not provide a way to compare them.[37] For a fixed school choice problem, given an estimation of student's preferences, and a valuation reflecting the importance of each combination of properties for the involved policy maker, one can compare these rules using our criterion.

As we mentioned, certain notions of degree are defined to express the *intensity* of the violation, rather than its *plausibility*. A problem of comparability across axioms obviously emerges with such notions, and developing an analytical framework able to account for a certain *partial commensurability*, inspired by the one we proposed here based on the *complete commensurability* guaranteed by the use of probabilities, constitutes an important complementary research.

8 Appendix

8.1 The set of collections of probabilities

With each non-empty set of axioms $S \subseteq A$, we associate a unique **logical sentence**, that we denote by $\tilde{S}$, whose truth value stands for the satisfaction of all the axioms in S. There are thus $2^J - 1$ such logical sentences. The set of sentences is denoted by $\mathcal{S}$. It induces a

[37] For example, Abdulkadiroğlu and Grigoryan (2021) show that "when each school has a single seat, the top trading cycles algorithm has less priority violations than any Pareto efficient and strategy-proof mechanisms." Neither DA or IA are among these rules.

set of **possible worlds**, denoted by $\mathcal{W}$, defined as the set of *logically consistent collections of truth values* of all sentences. That is, $\mathcal{W}$ is made of the 2^J collections of truth values $W = (w_{\tilde{S}})_{\tilde{S} \in \mathcal{S}}$, where $w_{\tilde{S}} \in \{T, F\}$ —T standing for *True*, F for *False*— such that, for all $\emptyset \neq S \subseteq A$,

$$w_{\tilde{S}} = T \text{ if and only if, for all } \emptyset \neq S' \subset S, w_{\tilde{S}'} = T.$$

One can illustrate this construction with a table where rows are sentences and columns are possible worlds.

Example 8. Let $A = \{a_1, a_2, a_3\}$, the set possible worlds $\mathcal{W}$ is represented by:

	W_0	W_1	W_2	W_3	W_4	W_5	W_6	W_8
$\tilde{a_1}$	F	T	T	T	T	F	F	F
$\tilde{a_2}$	F	F	T	F	T	T	T	F
$\tilde{a_3}$	F	F	F	T	T	F	T	T
$\widetilde{a_1a_2}$	F	F	T	F	T	F	F	F
$\widetilde{a_1a_3}$	F	F	F	T	T	F	F	F
$\widetilde{a_2a_3}$	F	F	F	F	T	F	T	F
$\tilde{A}$	F	F	F	F	T	F	F	F

.

Consider a collection $p = (p_S)_{\emptyset \neq S \subseteq A}$, with which we associate the collection $\tilde{p} : \tilde{S} \in \mathcal{S} \mapsto p_S \in [0, 1]$, where S is the non-empty subset of A to which the logical sentence $\tilde{S}$ corresponds in the construction above. In words, p is consistent if there exists a probability measure π on the set of possible worlds such that the value, according to $\tilde{p}$, associated with any sentence, is equal to the sum of the values, according to π, associated with the possible worlds where the sentence has truth value T.

Mathematically, take an arbitrary permutation of possible worlds and an arbitrary permutation of sentences, and define the incidence matrix H, of size $(2^J - 1) \times 2^J$, by $H_{ij} = 1$ if sentence $\tilde{S}_i$ has truth value T in world W_j, 0 otherwise. Let $\Delta^{\mathcal{W}} = \{\pi = (\pi_j)_{j=1,\ldots,2^J}, 0 \leq \pi_j \leq 1, \sum_j \pi_j = 1\}$.

We can now define the set of possible collections of probabilities, P. The collection $p = (p_S)_{\emptyset \neq S \subseteq A}$ belongs to P if and only if the associated $\tilde{p}$ is such that there exists $\pi \in \Delta^{\mathcal{W}}$ such that:

$$\tilde{p} = H \cdot \pi,$$

where $(\,\cdot\,)$ denotes the usual scalar product.

The set P is thus defined as the set of collections for which a specific linear system admits a solution. It remains to characterise the family of extreme points of P (Lemma 1):

$$P \text{ is the closed convex hull of } \mathbf{0} \text{ and } (p^{1,S})_{\emptyset \neq S \subseteq A};$$

we prove it now.

Proof of Lemma 1

The set of vectors $\tilde{p} \in [0,1]^{2^J - 1}$ for which there exists a solution in $\Delta^{\mathcal{W}}$ to the *linear* system:

$$\tilde{p} = H \cdot \pi$$

is convex and closed.

Hence, P is a convex compact subset of $[0,1]^{2^J-1}$, and, as such, it is the closed convex hull of its extreme points. It thus remains to prove that the extreme points of P are $\mathbf{0}$ and $(p^{1,S})_{\emptyset \neq S \subseteq A}$.

It is clear that these points are in P, and that they are extreme. These collections are the only $\{0,1\}-$valued collections in P. Indeed, let p be a $\{0,1\}-$valued collection that does not coincide with the null vector, or any $p^{1,S}$, $\emptyset \neq S \subseteq A$. Then, there exist $T, T' \subseteq A$ such that $\emptyset \neq T \subset T'$ and $p_{T'} = 1$ and $p_T = 0$; that is, $p \notin P$.

Let $p \in P$ such that there is $\emptyset \neq \tilde{T} \subseteq A$ such that $0 < p_{\tilde{T}} < 1$. Then there is $\epsilon > 0$ such that the collections $\underline{p}, \overline{p}$, defined by

$$\begin{aligned} \underline{p}_T &= p_T - \epsilon \text{ if } 0 < p_T < 1 \\ \underline{p}_T &= 1 \text{ if } p_T = 1 \\ \underline{p}_T &= 0 \text{ if } p_T = 0 \end{aligned}$$

and

$$\begin{aligned} \overline{p}_T &= p_T + \epsilon \text{ if } 0 < p_T < 1 \\ \overline{p}_T &= 1 \text{ if } p_T = 1 \\ \overline{p}_T &= 0 \text{ if } p_T = 0 \end{aligned}$$

are in P. In addition, $p = \frac{1}{2}\underline{p} + \frac{1}{2}\overline{p}$, that is, p is not extreme. We have thus characterised the set of extreme points of P.

By the linear independance of the family $(p^{1,S})_{\emptyset \neq S \subseteq A}$, we have proved:

For all $p \in P \setminus \{\mathbf{0}\}$, there exists a unique *pair $(\mathcal{I}^p, (\alpha^p_T)_{\emptyset \neq T \subseteq A})$, where $\mathcal{I}^p$ is a non-empty family of non-empty subsets of A, and $(\alpha^p_T)_{\emptyset \neq T \subseteq A}$ a family of real numbers, such that:*

- $0 < \alpha_T^p \leq 1$ *for all* $T \in \mathcal{I}^p$ *and* $\sum_{T \in \mathcal{I}^p} \alpha_T^p \leq 1$,
- $\alpha_T^p = 0$ *for all* $\emptyset \neq T \in 2^A \setminus \mathcal{I}^p$, *and*

$$p = \sum_{T \in \mathcal{I}^p} \alpha_T^p p^{1,T} \quad \Big(= \sum_{\emptyset \neq T \subseteq A} \alpha_T^p p^{1,T} + (1 - \sum_{T \in \mathcal{I}^p} \alpha_T^p) \mathbf{0} \Big).$$

We write the trivial equality in parenthesis above in order to stress that, in general, the sum $\sum_{T \in \mathcal{I}^p} \alpha_T^p$ is not one.

For the collection $\mathbf{0}$, we allow for the associated family of subsets to be empty and write $\mathcal{I}^{\mathbf{0}} = \emptyset$.

8.2 Theorem 1

A performance measure $m : U \times P \to \mathbb{R}_+$, *continuous on* U, satisfies

- Same contribution—same impact, and
- Expected valuation for single-axiom-reducible problems

if and only if it is the weighted Möbius performance measure, $\ddot{m}$:

$$\begin{aligned} \ddot{m} : U \times P &\to \mathbb{R}_+ \\ (u, p) &\mapsto \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big). \end{aligned}$$

Proof of Theorem 1

Only-if part. Let $m : U \times P \to \mathbb{R}_+$ be a *continuous* measure on U satisfying *same contribution—same impact* and *expected valuation for single-axiom-reducible problems*. For all $p \in P$, let $m^p : u \in U \mapsto m(u, p) \in \mathbb{R}_+$.

Let $u \in U_{st}$ and $\emptyset \neq S \subseteq A$. Consider $u^{S,\epsilon} \in U_{st}$ defined by $u_T^{S,\epsilon} = u_T$ if $T \neq S$ and $u_S^{S,\epsilon} = u_S + \epsilon$, for some $\epsilon > 0$. There exists such a $u^{S,\epsilon}$ in U_{st} because u lies in U_{st}.

Fix $p \in P$. For all $p' \in P$ such that $\alpha_S^p = \alpha_S^{p'}$, by *same contribution—same impact*,

$$\begin{aligned} & m^p\Big((u_{a_1}, \ldots, u_S + \epsilon, \ldots, u_A)\Big) - m^p\Big((u_{a_1}, \ldots, u_S, \ldots, u_A)\Big) \\ = {} & m^{p'}\Big((u_{a_1}, \ldots, u_S + \epsilon, \ldots, u_A)\Big) - m^{p'}\Big((u_{a_1}, \ldots, u_S, \ldots, u_A)\Big). \end{aligned}$$

Yet, by definition of a single-axiom-reducible problem, for $\lambda = \alpha^p_S \in [0,1]$, $\alpha^{p^{\lambda,S}}_S = \lambda = \alpha^p_S$. Hence,

$$\begin{aligned}
&m^p\Big((u_{a_1}, \dots, u_S + \epsilon, \dots, u_A)\Big) - m^p\Big((u_{a_1}, \dots, u_S, \dots, u_A)\Big)\\
&= m^{p^{\lambda,S}}\Big((u_{a_1}, \dots, u_S + \epsilon, \dots, u_A)\Big) - m^{p^{\lambda,S}}\Big((u_{a_1}, \dots, u_S, \dots, u_A)\Big)\\
&= \lambda(u_S + \epsilon) - \lambda u_S,
\end{aligned}$$

where the last equality follows from *expected valuation for single-axiom-reducible problems*.

As a consequence, for all $u \in U_{st}$, all $p \in P$, and all non-empty $S \subseteq A$,

$$\frac{m^p\Big((u_{a_1}, \dots, u_S + \epsilon, \dots, u_A)\Big) - m^p\Big((u_{a_1}, \dots, u_S, \dots, u_A)\Big)}{\epsilon} = \alpha^p_S.$$

By the convexity of U_{st}, the fact that $u^{S,\epsilon}$ lies in U_{st} implies that for all $0 < \gamma < \epsilon$, $u^{S,\gamma} \in U_{st}$.

One can thus let ϵ tend to 0 in the $2^J - 1$ equalities above, and this yields, as U_{st} is open, that for all $p \in P$, m^p is (continuously) differentiable on U_{st} and its gradient vector is *constant* and equal to $(\alpha^p_S)_{\emptyset \neq S \subseteq A}$. As U_{st} is connected, we have proved that for all $p \in P$, there exists $b^p \in \mathbb{R}$ such that, for all $u \in U_{st}$:

$$m^p(u) = \sum_{\emptyset \neq S \subseteq A} u_S \alpha^p_S + b^p.$$

As m^p is continuous, and as U is the closure of U_{st}, $m^p(u) = \sum_{\emptyset \neq S \subseteq A} u_S \alpha^p_S + b^p$ for all $u \in U$. We can now use $m(\mathbf{0}, p) = 0$ to conclude that for all $p \in P$, and $u \in U$

$$m(u,p) = \sum_{\emptyset \neq S \subseteq A} u_S \alpha^p_S.$$

It remains to give an explicit formula for α^p_S.

Recall that $\alpha^p_A = p_A$ for all $p \in P$.

For all $\emptyset \neq S \subset A$, and $p \in P$, the term α^p_S is defined recursively by

$$\begin{aligned}
&\alpha^p_S = p_S - \sum_{T : S \subset T} \alpha^p_T\\
\iff &\sum_{T : S \subseteq T} \alpha^p_T = p_S. \qquad (*)
\end{aligned}$$

Equation $(*)$ corresponds to the formula defining the Möbius transform of p associated with the partial order $\geq$ defined on $2^A \setminus \emptyset$, such that, $S \geq T \iff S \subseteq T$.[38] The remaining of the proof is similar to the construction of the Möbius transform associated with a finite set partially ordered by inclusion in the standard way.

By the classical result of Rota (Rota (1964)), $(\alpha_S^p)_{\emptyset \neq S \subseteq A}$ satisfies Equation $(*)$ for all non-empty $S \subseteq A$ if, and only if, there is a unique mapping $\nu : (2^A \setminus \emptyset) \times (2^A \setminus \emptyset) \to \mathbb{R}$ such that,

- $\nu(T, S) = 1$ if $T = S$; $\nu(T, S) = -\sum_{S' : S \subset S' \subseteq T} \nu(T, S')$ if $S \subset T$, and $\nu(T, S) = 0$ otherwise; and,
- for all non-empty $S \subseteq A$, $\alpha_S^p = \sum_{T : S \subseteq T} \nu(T, S) p_T$.

We prove by induction on the value of $|T \setminus S|$ that $\nu(T, S) = (-1)^{|T \setminus S|}$, for all $\emptyset \neq S, T$ with $S \subset T$.

If $|T \setminus S|= 1$, then $\nu(T, S) = -\nu(T, T) = -1$. Assume the result holds for all S', T' with $|T' \setminus S'|= k \in \mathbb{N}$ and let S, T such that $|T \setminus S|= k + 1$. Then,

$$\begin{aligned}
\nu(T, S) &= -\sum_{S' : S \subset S' \subseteq T} \nu(T, S') \\
&= -\sum_{S' : S \subset S' \subseteq T} (-1)^{|T \setminus S'|} \\
&= -\sum_{S' : S \subseteq S' \subseteq T} (-1)^{|T \setminus S'|} + (-1)^{|T \setminus S|} \\
&= (-1)^{|T \setminus S|}.
\end{aligned}$$

The second equality follows from the induction hypothesis. The last one comes from the basic result in combinatorics, according to which $\sum_{S' : S \subseteq S' \subseteq T} (-1)^{|T \setminus S'|}$ is equal to 1 if $S = T$ and to 0 otherwise (see Lemma 1.1 in Grabisch (2016)).

We have proved that m coincides with the weighted Möbius performance measure $\ddot{m}$.

If part. It is obvious that $\ddot{m}$ is continuous on U and satisfies *same contribution—same impact.*

[38] Again, we refer the reader to Grabisch (2016) (Chapter 2) for more details on the definition of the Möbius transform of a set function, given an arbitrary partial order defined on a (finite) set.

Let $\lambda \in [0,1]$, $\emptyset \neq S \subseteq A$, $p^{\lambda,S} \in P$, and $u \in U$. Then,

$$\ddot{m}(u, p^{\lambda,S}) = \sum_{\emptyset \neq B \subseteq A} u_B \lambda \Big(\sum_{T : B \subseteq T \subseteq S} (-1)^{|T \setminus B|} \Big).$$

As, for all $B \subseteq A$, all $S \subseteq A$,

$$\sum_{T : B \subseteq T \subseteq S} (-1)^{|T \setminus B|} = \begin{cases} 1, & \text{if } B = S \\ 0, & \text{otherwise} \end{cases},$$

$m(u, p^{\lambda,S}) = \lambda u_S$. We have proved that $\ddot{m}$ satisfies *expected valuation for single-axiom-reducible problems.*

Independence

- Consider the mapping:

$$\begin{aligned} m : U \times P &\to \mathbb{R} \\ (u,p) &\mapsto \max_{\emptyset \neq S \subseteq A} u_S p_S. \end{aligned}$$

Such a performance measure satisfies *expected valuation for single-axiom-reducible problems* but not *same contribution—same impact.* Consider the following example, with $A = \{a_1, a_2, a_3\}$:

	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
$\mathbf{u}$	1	1	1	3	3	2	6
$\mathbf{u^{a_1,a_2}}$	1	1	1	3	3	5	6 .
$\mathbf{p}$	0.85	0.9	0.9	0.65	0.7	0.8	0.6
$\mathbf{p'}$	0.7	0.55	0.5	0.35	0.3	0.4	0.2

The reader can check that when considering u and $u^{a_2a_3}$, the impact of a_2a_3 on m, as defined in the *same contribution—same impact* principle is $4 - 3.6 = 0.4$ under p while it is $2 - 0.2 = 0.8$ under p' (and $\alpha^{p}_{a_2a_3} = \alpha^{p'}_{a_2a_3} = 0.2$).

The measure $\hat{m}$ defined in Section 4.3 also satisfies *expected valuation for single-axiom-reducible problems* but not *same contribution—same impact.*

- Consider the mapping:

$$\begin{aligned} m : U \times P &\to \mathbb{R} \\ (u,p) &\mapsto \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big)^2. \end{aligned}$$

  Such a performance measure satisfies *same contribution—same impact* but not *expected valuation for single-axiom-reducible problems*: for all $\lambda \in [0,1]$, all $u \in U$, $m(u,p) = \lambda^2$.

Let us now display a performance measure which satisfies the two axioms, but is not continuous on U; $m : U \times P \to \mathbb{R}_+$:

$$m(u,p) = \begin{cases} \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big) + b^p, \text{ for some } b^p \in \mathbb{R}, & \text{if } u \in U_{st} \\ \sum_{\emptyset \neq S \subseteq A} u_S \Big(\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T \Big) + 4b^p & \text{otherwise} \end{cases},$$

where,

$$b^p \begin{cases} > 0 & \text{if } p \neq p^{\lambda, S} \text{ for any } \emptyset \neq S \subseteq A, \text{ and any } \lambda \in [0,1], \\ = 0 & \text{otherwise} \end{cases}.$$

8.3 Theorem 2

An incompatibility measure $\psi : P^* \to \mathbb{R}^J$ satisfies

- Same cost—same incompatibility,
- Allocation of incompatibility, and
- Symmetry

if and only if it coincides with the Shapley incompatibility measure: for all $a \in A$,

$$\begin{aligned} \psi_a : P^* &\to \mathbb{R}^J \\ p &\mapsto \sum_{S \subseteq A \setminus a} \frac{|S|!\,(J - |S| - 1)!}{J!} (p_S - p_{S \cup a}). \end{aligned}$$

Proof of Theorem 2

The *if part* is readily checked.

Suppose $\psi : P^* \to \mathbb{R}^J$ satisfies the three properties. Let

$$\begin{aligned}\tilde{\psi} : V &\to \mathbb{R}^J \\ v &\mapsto \psi(1-v).\end{aligned}$$[39]

Consider φ the Shapley incompatibility measure, and let $\tilde{\varphi}$ denote the restriction of the classical Shapley value to the set of games V: for all $a \in A$,

$$\begin{aligned}\tilde{\varphi}_a : V &\to \mathbb{R} \\ v &\mapsto \varphi(1-v) = \sum_{S \subseteq A \setminus a} \frac{|S|!\,(J-|S|-1)!}{J!}(v_{S\cup a} - v_S).\end{aligned}$$

Consider the family of games $(\hat{v}^S)_{S\subseteq A}$ in V, where each game $\hat{v}^S$ is defined by

$$\hat{v}^S_T = \begin{cases} 1 & \text{if } T \not\subseteq S \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\hat{v}^A = \mathbf{0}$, $\hat{v}^S = 1 - \hat{p}^{1,S}$ for all $\emptyset \neq S \subset A$, and $\hat{v}^\emptyset = \hat{p}^{1,A}$, where $\hat{p}^{1,S}$ denotes a collection of $\mathbb{R}^{2^J}$, such that $\hat{p}^{1,S}_\emptyset = 1$, $\hat{p}^{1,S}_T = 1$ if $\emptyset \neq T \subseteq S$, 0 otherwise.

The following table illustrates this definition, with $A = \{a_1, a_2, a_3\}$, taking $v = \hat{v}^{a_1a_2}$ and the corresponding p:

	$\emptyset$	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
v	0	0	0	1	0	1	1	1.
p	1	1	1	0	1	0	0	0

We see from the proof of Lemma 1 that the set of extreme points of V is the family $(\hat{v}^S)_{S\subseteq A}$ —the set of extreme points of P^* is the family $(\hat{p}^{1,S})_{S\subseteq A}$. Let $v \in V$. There exist a unique family of subsets of A, denoted by $\mathcal{I}^v$, and a unique family of positive real numbers

[39] Briefly, *same cost—same incompatibility* implies that for all $v, v' \in V$ and all $a \in A$ such that $v_{S\cup a} - v_S = v'_{S\cup a} - v'_S$ for all $S \subseteq A \setminus a$, $\tilde{\psi}_a(v) = \tilde{\psi}_a(v')$. Also, *allocation of incompatibility* implies that for all $v \in V$, $\sum_{a\in A} \tilde{\psi}_a(v) = v_A$. Finally, defining, for a permutation π, the game v^π by $v^\pi_S = v_{(\pi(a))_{a\in S}}$, *symmetry* implies that $\tilde{\psi}_a(v) = \tilde{\psi}_{\pi(a)}(v^\pi)$.

$(\alpha^v_T)_{T\in\mathcal{I}^v}$, such that $\sum_{T\in I^v}\alpha^v_T = 1$ and

$$v = \sum_{T\in\mathcal{I}^v}\alpha^v_T\hat{v}^T.$$

In particular, on $V\setminus\hat{v}^A$, for such families, the Shapley value associated with $a\in A$ is given by

$$\tilde{\varphi}_a(v) = \sum_{T\in\mathcal{I}^v}\alpha^v_T\tilde{\varphi}_a(\hat{v}^T) = \sum_{T\in\mathcal{I}^v:a\notin T}\alpha^v_T\frac{1}{|A\setminus T|},$$

and $\tilde{\varphi}_a(\hat{v}^A) = 0$. The reader may refer to Lemma 2 below, where these equalities are proved. Moreover, it is not needed for this proof to explicitly give $(\mathcal{I}^v, (\alpha^v_T)_{T\in\mathcal{I}^v})$; this is simple though and we do so in the proof of Proposition 1 below.

We are now able to prove that the functions $\tilde{\psi}$ and $\tilde{\varphi}$ coincide on V.

In the following, for a game $v\in V$, we say that two axioms $a, a'\in A$ are **symmetric** in v if their transposition defines a symmetry of v. Formally, consider the permutation $\pi: A\to A$ defined by $\pi(\tilde{a}) = \tilde{a}$ for all $\tilde{a}\in A\setminus\{a, a'\}$, $\pi(a) = a'$ and $\pi(a') = a$. Axioms a and a' are symmetric in v if, for all $S\subseteq A$, $v_{(\pi(a)_{a\in S})} = v_S$.

For $v\in V$, let K^v denote the number of non-zero terms in a the convex combination of extreme points of V to which v is equal, described above.

If $K^v = 0$, then $v = \mathbf{0}$ *symmetry* and *allocation of incompatibility* imply that, for all $a\in A$, $\tilde{\psi}_a(v) = 0 = \tilde{\varphi}_a(v)$.

If $K^v = 1$, then there is $T\subseteq A$ such that $v = \hat{v}^T$. If $T = A$, then $v = \mathbf{0}$ and one concludes as in the previous case. If $T = \emptyset$, by *symmetry* and *allocation of incompatibility*, for all $a\in A$, $\tilde{\psi}_a = \frac{1}{J} = \tilde{\varphi}_a$. Assume now $\emptyset\neq T\subset A$. For all $a\in T$, $v_{S\cup a} - v_S = 0$ for all $S\subseteq A\setminus a$. Indeed, either $S\nsubseteq T$ and $v_{S\cup a} = v_S = 1$, or $S\subset T$ and $v_{S\cup a} = v_S = 0$. By *same cost—same incompatibility*, this yields $\psi_a(v) = \psi_a(\mathbf{0}) = 0$. All $a, a'\notin T$ are symmetric in v and we conclude, by *symmetry* and *allocation of incompatibility*, that $\tilde{\psi}_a(v) = \tilde{\psi}_{a'}(v) = \frac{1}{|A\setminus T|} = \tilde{\varphi}_a(v)$.

We now proceed by induction on the value of K^v.

Assume that for all $v\in V\setminus\mathbf{0}$ such that $K^v\leq k\in\mathbb{N}$, for all $a\in A$,

$$\tilde{\psi}_a(v) = \sum_{T\in\mathcal{I}^v:a\notin T}\alpha^v_T\frac{1}{|A\setminus T|}.$$

Let $v = \sum_{T\in\mathcal{I}^v}\alpha^v_T\hat{v}^T\in V\setminus\mathbf{0}$ with $K^v = k+1$. Consider $\mathcal{T}^v = \bigcup_{T\in\mathcal{I}^v}T$ and $a\in\mathcal{T}^v$.

Define the game

$$\nu = \sum_{T \in \mathcal{I}^v : a \notin T} \alpha_T^v \hat{v}^T + (1 - \sum_{T \in \mathcal{I}^v : a \notin T} \alpha_T^v) \mathbf{0}.$$

Clearly, $\nu \in V$ and $K^\nu \leq k$. In addition, $\nu_{S \cup a} - \nu_S = v_{S \cup a} - v_S$ for any $S \subseteq A \setminus a$. Indeed, for all $S \subseteq A \setminus a$, for all $T \in \mathcal{I}^v$,

$$\hat{v}_{S \cup a}^T - \hat{v}_S^T = \begin{cases} 0 & \text{if } S \nsubseteq T \\ 0 & \text{if } S \subset T \text{ and } a \in T \\ 1 & \text{if } S \subseteq T \text{ and } a \notin T, \end{cases}$$

which implies:

$$v_{S \cup a} - v_S = \sum_{T \in \mathcal{I}^v : S \subseteq T, a \notin T} \alpha_T^v = \nu_{S \cup a} - \nu_S.$$

Therefore, if $\nu \neq \mathbf{0}$, *i.e.* if $a \notin \bigcap_{T \in \mathcal{I}^v} T$,

$$\tilde{\psi}_a(v) = \tilde{\psi}_a(\nu) = \sum_{T \in \mathcal{I}^v : a \notin T} \alpha_T^v \frac{1}{|A \setminus T|} = \tilde{\varphi}_a(v),$$

where the first equality follows from *same cost—same incompatibility*, and the second follows from the induction hypothesis. And if $\nu = \mathbf{0}$, *i.e.* if $a \in \bigcap_{T \in \mathcal{I}^v} T$, $\tilde{\psi}_a(v) = 0 = \tilde{\varphi}_a(v)$.

Moreover, all axioms in $A \setminus \mathcal{T}^v$ are symmetric in v.[40] As $\tilde{\psi}_a$ coincides with $\tilde{\varphi}_a$ for $a \in \mathcal{T}^v$, *symmetry* and *allocation of incompatibility* imply that for $a \notin \mathcal{T}^v$, $\tilde{\psi}_a(v) = \tilde{\varphi}_a(v)$.

We have proved that $\tilde{\psi}$ coincides with $\tilde{\varphi}$, which implies that the incompatibility measure ψ coincides with the Shapley incompatibility measure.

Independence

That the three principles, expressed for $\tilde{\psi} : V \to \mathbb{R}$, that is, expressed for the restriction to V of a *solution*, are independent is shown in exactly the same way as when the set of admissible games is $G = \{u = (u_S)_{S \subseteq A} \in \mathbb{R}^{2^J}, u_\emptyset = 0\}$.

8.4 Alternative characterisation of the Shapley incompatibility measure

Consider the following principles.

[40] For all $a \in A \setminus \mathcal{T}^v$, for all $S \subseteq A \setminus a$, $v_{S \cup a} = 1$, thus, any transposition of two elements of $A \setminus \mathcal{T}^v$ is a symmetry of v.

No cost—no incompatibility

Let $p \in P^*$, and $a \in A$ such that $p_{S \cup a} = p_a$ for all $S \subseteq A \setminus a$.
Then,

$$\psi_a(p) = 0.$$

Such an axiom simply exerts no cost in terms of probability of satisfaction and should thus be considered as maximally compatible with the others.

Convex linearity

Let $p, p' \in P^*$, $\lambda \in [0, 1]$.
Then,

$$\psi_a(\lambda p + (1 - \lambda)p') = \lambda \psi_a(p) + (1 - \lambda)\psi_a(p').$$

This is a weakening of the classical *additivity and positive homogeneity* principle required on the whole subspace $\left\{p \in \mathbb{R}^{2^J}, p_\emptyset = 1\right\}$.[41] It is suited for P^*, which is a compact and convex subset of this subspace. It is best interpreted as a simplicity requirement.

Theorem 3. *An incompatibility measure $\psi : P^* \to \mathbb{R}^J$ satisfies*

- *Convex linearity,*
- *Allocation of incompatibility,*
- *Symmetry, and*
- *No cost—no incompatibility*

if and only if it coincides with the Shapley incompatibility measure.

Proof. The *if part* is readily checked.

Suppose $\psi : P^* \to \mathbb{R}^J$ satisfies these four properties.

Lemma 2. *Let $S \subseteq A$. Then, for all $a \in A$, $\tilde{\psi}_a(\hat{v}^S) = \tilde{\varphi}_a(\hat{v}^S)$.*

[41] For completeness, let us state it: let $p, p' \in \left\{p \in \mathbb{R}^{2^J}, p_\emptyset = 1\right\}, \lambda \geq 0$. Then,

$$\psi_a\left(p + \lambda p'\right) = \psi_a(p) + \lambda \psi_a\left(p'\right).$$

Proof. If $S = A$, then $\hat{v}^S = \mathbf{0}$ and *symmetry* and *allocation of incompatibility* imply $\tilde{\psi}_a(\hat{v}^S) = 0 = \tilde{\varphi}_a(\hat{v}^S)$, for all $a \in A$. Let $S \subset A$ and consider $\hat{v}^S$. Let $a \in S$, then, by *no cost—no incompatibility*, $\tilde{\psi}_a(\hat{v}^S) = 0 = \tilde{\varphi}_a(\hat{v}^S)$. In addition, all axioms in $A \setminus S$ are symmetric so that, by *symmetry* and *allocation of incompatibility*, for all $a \in A \setminus S$, $\tilde{\psi}_a(\hat{v}^S) = \tilde{\varphi}_a(\hat{v}^S) = \frac{1}{|A \setminus S|}$. □

By *convex linearity*, for all $v \in V$, there exist a unique family of subsets of A, denoted by $\mathcal{I}^v$, and a unique family of positive real numbers $(\alpha_T^v)_{T \in \mathcal{I}^v}$, such that $\sum_{T \in \mathcal{I}^v} \alpha_T^v = 1$ and

$$\tilde{\psi}_a(v) = \sum_{T \in \mathcal{I}^v} \alpha_T^v \tilde{\psi}_a(\hat{v}^T) \text{ for all } a \in A.$$

Then, by Lemma 2,

$$\tilde{\psi}_a(v) = \sum_{T \in \mathcal{I}^v} \alpha_T^v \tilde{\varphi}_a(\hat{v}^T) = \tilde{\varphi}_a(v) \text{ for all } a \in A.$$

□

That the four principles, expressed for $\tilde{\psi} : V \to \mathbb{R}$, that is, expressed for the restriction to V of a *solution*, are independent is shown in exactly the same way as when the set of admissible games is $G = \{u = (u_S)_{S \subseteq A} \in \mathbb{R}^{2^J}, u_\emptyset = 0\}$.

8.5 Proposition 1

Let $a \in A$ and $p \in P^*$. Then,

$$\varphi_a(p) = \sum_{S \subseteq A : a \notin S} \frac{\sum_{T : S \subseteq T} (-1)^{|T \setminus S|} p_T}{|A \setminus S|}.$$

Proof of Proposition 1

As we noted in the proof of Theorem 2, the set of extreme points of the convex polytope P^* is the family $(\hat{p}^{1,S})_{S \subseteq A}$ defined by $p_T^{1,S} = 1$ if $T \subseteq S$ and 0 otherwise, for all $T \subseteq S$. For all $p \in P^*$, there exists a unique pair $(\mathcal{I}^{*p}, (\alpha_T^{*p})_{T \subseteq A})$, where $\mathcal{I}^{*p}$ is a non-empty family of subsets of A, and $(\alpha_T^{*p})_{T \subseteq A}$ is a family of real numbers such that

- $0 < \alpha_T^{*p} \leq 1$ for all $T \in \mathcal{I}^{*p}$ and $\sum_{T \in \mathcal{I}^{*p}} \alpha_T^{*p} = 1$,

- $\alpha_T^{*p} = 0$ for all $T \in 2^A \setminus \mathcal{I}^{*p}$, and

$$p = \sum_{T \in \mathcal{I}^{*p}} \alpha_T^{*p} \hat{p}^{1,T}.$$

Let $p \in P^*$. Similarly to the procedure described in Section 4, $(\mathcal{I}^{*p}, (\alpha_T^{*p})_{T \subseteq A})$ obtains as follows:

Step 1. If $p_A > 0$, set $\mathcal{I}_1^{*p} = \{A\}$ and $\alpha_A^{*p} = p_A$, otherwise, set $\mathcal{I}_1^{*p} = \emptyset$ and $\alpha_A^{*p} = 0$;

Step k (for 2 ≤ k ≤ J). Set $\mathcal{I}_k^{*p} = \mathcal{I}_{k-1}^{*p} \cup \{T \subseteq A$ with $|T| = J - k$ and $p_T - \sum_{S:T \subset S} \alpha_S^{*p} > 0\}$, and, for all $T \subseteq A$ with $|T| = J - k$, $\alpha_T^{*p} = p_T - \sum_{S:T \subset S} \alpha_S^{*p}$.

Define $\mathcal{I}^{*\mathbf{p}} = \mathcal{I}_{\mathbf{J}}^{*\mathbf{p}}$.

Then, $(\alpha_T^{*p})_{T \subseteq A}$ is the Möbius transform of p for the partial order $\geq$ defined on 2^A such that:

$$\text{for all } S, T,\ S \geq T \iff S \subseteq T.$$

Hence, for all $p \in P^*$, for all $S \subseteq A$, $\alpha_S^{*p} = \sum_{T:S \subseteq T} (-1)^{|T \setminus S|} p_T$.

The set of extreme points of V is the family $(\hat{v}^S)_{S \subseteq A} = (1 - \hat{p}^{1,S})_{S \subseteq A}$. Let $v = 1 - p$, $p \in P^*$, it is easy to see that $v = \sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \hat{v}^S$. Indeed, for all $T \subseteq A$:

$$\begin{aligned} p_T &= \sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \hat{p}_T^{1,S} \\ \iff v_T &= 1 - \sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \hat{p}_T^{1,S} \\ \iff v_T &= 1 - \sum_{S \in \mathcal{I}^{*p}: T \subseteq S} \alpha_S^{*p} \\ \iff v_T &= \sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \hat{v}_T^S. \end{aligned}$$

The last equivalence comes from the fact that $\sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \hat{v}_T^S = \sum_{S \in \mathcal{I}^{*p}: T \nsubseteq S} \alpha_S^{*p}$, and that $\sum_{S \in \mathcal{I}^{*p}: T \nsubseteq S} \alpha_S^{*p} + \sum_{S \in \mathcal{I}^{*p}: T \subseteq S} \alpha_S^{*p} = 1$.

Then, for all $p \in P^*$, all $a \in A$,

$$\varphi_a(p) = \tilde{\varphi}_a(1 - p) = \sum_{S \in \mathcal{I}^{*p}} \alpha_S^{*p} \tilde{\varphi}_a(\hat{v}^S) = \sum_{S \subseteq A: a \notin S} \alpha_S^{*p} \frac{1}{|A \setminus S|}.$$

Acknowledgements: I am deeply grateful to Marc Fleurbaey, Antonin Macé and William

Thomson for their support, as well as for their detailed comments on this work. I thank the participants of the Workshop on Collective Decisions in Economic Analysis of the University of Alicante, the 9th International Workshop on Computational Social Choice in the University of Beersheba, the Conference on Economic Design in the University of Girona, the 16th meeting of the Society for Social Choice and Welfare in the Autonomous Technological Institute of Mexico, and the participants of the Online Social Choice and Welfare seminar, and of the economics seminar of the University of Caen.

References

Abdulkadiroğlu, A., Andersson, T., 2023. School choice, in: Handbook of the Economics of Education. Elsevier. volume 6, pp. 135–185.

Abdulkadiroğlu, A., Grigoryan, A., 2021. Priority-based assignment with reserves and quotas. Technical Report. National Bureau of Economic Research.

Abdulkadiroğlu, A., Che, Y.K., Pathak, P.A., Roth, A.E., Tercieux, O., 2020. Efficiency, justified envy, and incentives in priority-based matching. American Economic Review: Insights 2, 425–442.

Aleskerov, F., Karabekyan, D., Sanver, M.R., Yakuba, V., 2012. On the manipulability of voting rules: The case of 4 and 5 alternatives. Mathematical Social Sciences 64, 67–73.

Arribillaga, R.P., Massó, J., 2016. Comparing generalized median voter schemes according to their manipulability. Theoretical Economics 11, 547–586.

Artemov, G., Che, Y.K., He, Y., 2017. Strategic 'mistakes': Implications for market design research. Working paper .

Barberà, S., Gerber, A., 2017. Sequential voting and agenda manipulation. Theoretical Economics 12, 211–247.

Bardier, P., Dong-Xuan, B., Nguyen, V.Q., 2024. Unanimity of two selves in decision making. arXiv preprint arXiv:2406.11166 .

Bewley, T.F., 2002. Knightian decision theory. part i. Decisions in economics and finance 25, 79–110.

Boehmer, N., Bredereck, R., Faliszewski, P., Niedermeier, R., Szufa, S., 2021. Putting a compass on the map of elections. arXiv preprint arXiv:2105.07815 .

Boehmer, N., Faliszewski, P., Kraiczy, S., 2023. Properties of the mallows model depending on the number of alternatives: A warning for an experimentalist. Working paper .

Böhm, P., Bredereck, R., Gölz, P., Kaczmarczyk, A., Szufa, S., 2024. Putting fair division on the map. Working paper .

Brandt, F., Brill, M., Seedig, H.G., 2011. On the fixed-parameter tractability of composition-consistent tournament solutions, in: Twenty-Second International Joint Conference on Artificial Intelligence.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D. (Eds.), 2016a. Handbook of Computational Social Choice. Cambridge University Press.

Brandt, F., Geist, C., Seedig, H.G., 2014. Identifying k-majority digraphs via sat solving, in: Proceedings of the 1st AAMAS workshop on exploring beyond the worst case in computational social choice (EXPLORE).

Brandt, F., Geist, C., Strobel, M., 2016b. Analyzing the practical relevance of voting paradoxes via ehrhart theory, computer simulations, and empirical data, in: Proceedings of the 2016 international conference on autonomous agents & multiagent systems, pp. 385–393.

Campbell, D.E., Kelly, J.S., 1994. Trade-off theory. The American Economic Review 84, 422–426.

Campbell, D.E., Kelly, J.S., 2015. Social choice trade-off results for conditions on triples of alternatives. Mathematical Social Sciences 77, 42–45.

Chevaleyre, Y., Endriss, U., Maudet, N., 2017. Distributed fair allocation of indivisible goods. Artificial Intelligence 242, 1–22.

Dasgupta, P., Maskin, E., 2008. On the robustness of majority rule. Journal of the European Economic Association 6, 949–973.

Diss, M., Kamwa, E., 2020. Simulations in models of preference aggregation. Œconomia. History, Methodology, Philosophy , 279–308.

Echenique, F., Miyashita, M., Nakamura, Y., Pomatto, L., Vinson, J., 2022. Twofold multiprior preferences and failures of contingent reasoning. Journal of Economic Theory 202, 105448.

Fishburn, P.C., 1985. Interval graphs and interval orders. Discrete mathematics 55, 135–149.

Fishburn, P.C., 2015. The theory of social choice. Princeton University Press.

Gehrlein, W.V., 1983. Condorcet's paradox. Theory and Decision 15, 161–197.

Gehrlein, W.V., Lepelley, D., et al., 2017. Elections, voting rules and paradoxical outcomes. Springer.

Georgakopoulos, G., Kavvadias, D., Papadimitriou, C.H., 1988. Probabilistic satisfiability. Journal of complexity 4, 1–11.

Ghasvareh, P., Pápai, S., Concordia University (Montréal, Q.D.o.E., 2020. Fairness Comparisons of Matching Rules. Concordia University. URL: https://books.google.com/books?id=-bxRzwEACAAJ.

Gibbard, A., 1973. Manipulation of voting schemes: a general result. Econometrica: journal of the Econometric Society , 587–601.

Grabisch, M., 2016. Set functions, games and capacities in decision making. volume 46. Springer.

Laslier, J.F., 2010. In silico voting experiments, in: Handbook on approval voting. Springer, pp. 311–335.

Lepelley, D., Louichi, A., Valognes, F., 2000. Computer simulations of voting systems. Advances in Complex Systems 3, 181–194.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems 30.

Moulin, H., 2019. Fair division in the internet age. Annual Review of Economics 11, 407–441.

Moulin, H., Thomson, W., 1988. Can everyone benefit from growth?: Two difficulties. Journal of Mathematical Economics 17, 339–345.

Nilsson, N.J., 1986. Probabilistic logic. Artificial intelligence 28, 71–87.

Pathak, P.A., Sönmez, T., 2013. School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. American Economic Review 103, 80–106.

Rota, G.C., 1964. On the foundations of combinatorial theory: I. theory of möbius functions, in: Classic Papers in Combinatorics. Springer, pp. 332–360.

Roth, A.E., Peranson, E., 1999. The redesign of the matching market for american physicians: Some engineering aspects of economic design. American economic review 89, 748–780.

Satterthwaite, M.A., 1975. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of economic theory 10, 187–217.

Schmidtlein, M.C., Endriss, U., 2023. Voting by axioms, in: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, pp. 2067–2075.

Schummer, J., 2004. Almost-dominant strategy implementation: exchange economies. Games and Economic Behavior 48, 154–170.

Skowron, P., 2021. Proportionality degree of multiwinner rules, in: Proceedings of the 22nd ACM Conference on Economics and Computation, pp. 820–840.

Szufa, S., Faliszewski, P., Skowron, P., Slinko, A., Talmon, N., 2020. Drawing a map of elections in the space of statistical cultures, in: Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, pp. 1341–1349.

Thomson, W., 2001. On the axiomatic method and its recent applications to game theory and resource allocation. Social Choice and Welfare 18, 327–386.

Thomson, W., 2023. The Axiomatics of Economic Design, Vol. 1: An Introduction to Theory and Methods. Springer Nature.

Wilson, M.C., 2019. Inputs, algorithms, quality measures: More realistic simulation in social choice. The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers , 165–171.

Young, H.P., 1985. Monotonic solutions of cooperative games. International Journal of Game Theory 14, 65–72.