Non Binary Social Choice

Pierre BARDIER

Abstract

We provide a formal framework accounting for a very common intuition in the (computational) social choice literature: analytically established incompatibilities should be qualified by the plausibility of axioms' violations. In consistency with the increasing use of simulation models aiming at assessing the empirical performance of rules, we define the degree to which a given rule satisfies an axiom based on a probability distribution defined over the inputs that this rule takes. We then propose and characterise a criterion to evaluate and compare rules given a set of axioms, and a criterion to measure axioms' compatibility between each other for a given rule, building on an analogy with cooperative game theory.

1 Introduction

Incompatibilities between axioms have given rise to numerous attempts to provide specific notions of the *degree to which an axiom is satisfied*, as alternatives to the typical domain restriction method. Such notions enable in turn to compare in a subtle way rules that where not comparable when sticking to the binary constraint according to which a rule satisfies an axiom if it satisfies its requirements on a whole parameter domain, or simply does not satisfy the axiom. Two predominant interpretations support their use: one in which the parameters selected to measure the departure from a desirable property represent the *intensity* of the violation of axioms, and one in which they represent its *plausibility*. We adopt the latter approach in this paper as we propose a criterion to compare rules based on the probability that they satisfy axioms and sets of axioms.

Studies discussing the empirical frequency, or the theoretical likelihood, of a given rule satisfying a certain property consist in *counting the instances* for which the rule meets the considered requirements —these instances can be composed of theoretical or stochastically generated preference profiles, real elections various parameters, sets of alternatives *etc.* In that respect, simulation models, focusing initially mostly on the occurrence of the Condorcet paradox in voting situations (see [11] and [12] for a review of this literature), are now commonly used, in increasingly diverse settings (*e.g* in voting, fair division, market design problems), and under less restrictive statistical assumptions ([9], [22]).

In consistency with these models, we consider an abstract set of instances, endowed with a probability structure used to account for their relative frequency. The typical example of such a set in our view is the set of preference profiles associated either to a fixed or to a varying group of agents. We can then measure the mass of instances for which *punctual and relational axioms*, in the words of William Thomson ([21]), are verified. This general framework may be applied in any field in which an axiomatic approach is relevant and, in particular, covers a large spectrum of social choice problems, be they Arrovian aggregation, voting problems, fair division selection - or ranking - models, with divisible or indivisible goods, while, importantly, providing the degree of satisfaction of both single axioms and of combinations of axioms.

Defining the degree of satisfaction as a probability guarantees, in contrast to definitions based on intensity notions, its commensurability across (sets of) axioms. Concretely, the possibility to compare by how much a given rule satisfies two different sets of axioms proves fundamental to *i)* evaluate and compare rules and *ii)* decompose the logical performance of a given rule.

As we already mentioned, for a given set of axioms and a given rule, we consider the probability to satisfy all possible combinations of axioms. We then propose and characterise a criterion to assess the performance of a rule on the basis of this collection of probabilities, taking into account the way in which the specific normative content of axioms should influence this evaluation. More precisely, the normative desirability of axioms and their interaction are introduced through the use of a *capacity*, and the way the probability to satisfy a given set of axioms influences the measure of the performance depends on the specific structure of this set function.

Finally, these probability vectors can be analysed in order to determine the level of compatibility of axioms given a specific rule. We propose and characterise a criterion fulfilling this purpose based on an analogy with cooperative game theory.

2 Related Literature

William Thomson, in a paper in which he seeks to characterise the essential features of the *axiomatic program* [21] conceives this research as the attempt to draw as precise a frontier as possible between families of problems for which given properties are compatible and families for which they are not. This view has motivated the most standard way of dealing with impossibilities in social choice theory: when some axioms are shown to be incompatible on a given parameter domain, it seems natural to look for restricted domains in which these axioms actually can be combined. Accordingly, the plausibility of these axioms' compatibility becomes the plausibility of the restricted domains, and it is left to the *theory consumer* ([21]) to assess how suitable the domain restrictions are in the context at hand. Recently, this type of approach has saliently been described in Hervé Moulin's review of new developments of fair allocation theory, centered around very structured problems such as ones with one-dimensional single-peaked preferences, dichotomous preferences, or preferences with perfect substituability [17]. Restricted preference domains have also proved fruitful in algorithmic social choice theory, for example as tools to circumvent the NP-hardness of winner determination in famous voting rules ([10]).

Yet, this approach maintains the binary constraint according to which a given condition is satisfied on a whole parameter domain or simply is not satisfied, whereas, in order to construct a *less partial* order between rules, when a procedure fails to yield good outcomes, one would like to know 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 [16] and [3] for instance.[1] However, a great number of principles guiding the formulation of axioms are not satisfactorily parametrizable based on this interpretation in terms of intensity.[2]

Another type of theoretical approach, closer to the way we proceed in this paper, was adopted in the context of Arrovian social choice theory in [7], [6] and [8]. The point is to *count the number* of pair, or triples, of alternatives for which studied axioms are satisfied in order to identify *trade-offs* between them.

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

[1] In this classical paper, the authors of [16] propose such a parametric relaxation of a proportionality axiom —where, informally, the parameter defines a portion of the endowment divided by the number of agents, and it is required that for any preference profile, the considered rule selects an allocation for which each agent prefers his or her bundle to this portion— and then actually show that the impossibility they established for the non parametric version still holds for any value adopted in the parametric relaxation.

[2] If it is probably always possible to find parametric relaxations of that type, it is more complex to find ones that cater to a non-equivocal, intuitive and convincing —with respect to the principle that motivates them— sense in which the weaker conditions depart from the stronger one.

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 [9]. Let us mention a few example of studies in computational social choice and market design consisting in measuring the empirical frequency of the violation of a given property, different from Condorcet consistency.[3] In the former literature, [4] studied the number of solutions 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, [5] concluded, based on simulations used to extend theoretical results obtained for problems involving four alternatives, that sensibility to such contraction is of higher practical relevance than the Condorcet loser paradox. After extending theoretical indices used for single-valued social choice procedures, [1] study the level of manipulability of multi-valued rules, using computer experiments on problems with four and five alternatives. In market design, [18] conducted 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 *de jure* manipulable, the number of agents who would have an interest in returning a false report, when every other agent is truthful, vanishes as the size of the market grows.

We believe that this work can help analyse and compare rules in a subtle way by providing measures of performance that incorporate both the probability to satisfy a set of axioms, their normative desirability and the way they interact. In particular, it provides a way to enrich the use of models based on notions of degree interpreted in terms of frequency of violation, especially of simulation models.[4]

3 A commensurable notion of degree of satisfaction

Let us give an illustration of the simple objects we briefly mentioned in the introduction around which this work is structured. We aim first at selecting rules on the basis of arrays of the following form:

a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
1	0.8	0.4	0.8	0.4	0.35	0.35

which represents *how often* a (choice) rule satisfies combinations of axioms in $A = \{a_1, a_2, a_3\}$ —precise definitions will be given subsequently.

3.1 Rules and axioms

Any notion of the frequency at which a given rule satisfies axioms requires considering *instances* over which measuring its behaviour. As suggested above, making preference profiles vary and studying the *outcomes* prescribed by a rule, which is often done in practice, is an evident way to generate instances —and distributions over these instances. However, in order to be both as general and as concise as possible, we directly 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 required properties.

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 i of I as instances,

[3]Once again, see [11] and [12] for a review of the literature specifically dedicated to the Condorcet paradox. See also [13] and [14].

[4][22] highlighted the importance of this issue for computer experiments.

and to elements o of O as outcomes.[5] These objects respectively represent the arguments and the images of a rule: a **rule** f is a mapping

$$f : I \to O.$$

As an illustration, in the classical microeconomic division problem, an instance is made of a group of $n \in \mathbb{N}$ agents, a social endowment of $l \in \mathbb{N}$ infinitely divisible resources $\Omega \in \mathbb{R}^l_+$, and a profile of continuous, monotonic and convex individual preferences over $\mathbb{R}^l_+$.[6] A rule is then a correspondence, mapping each such instance i to a set o of vectors in $\mathbb{R}^{nl}_+$.[7]

An axiom can then be described as defining a subset of admissible rules. Before proceeding to definitions, an important distinction should be made between, in the words of William Thomson [21], *punctual* and *relational* axioms. The former are requirements made 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 we just mentioned as an illustration, 'efficiency' is a punctual axiom, and so is 'envy-freeness', while 'population monotonicity' and 'ressource monotonicity' are relational ones.[8]

A **punctual axiom** a is defined as a set of rules such that for a given family $(O^a_i)_{i \in I}$, $O^a_i \subseteq O$,

$$f \in a \text{ if and only if } \big[f(i) \in O^a_i \text{ for all } i \in I \big].$$

In words, a specifies for each instance a set of admissible outcomes, and the image of an instance by rule f satisfies the requirement of a if and only if it belongs to an element of this family.

A **relational axiom** a is defined as a set of rules such that for a given $K \in \mathbb{N}$ and a family $(O^a_{i_1,...,i_K})_{(i_1,...,i_K) \in I^K}$, $O^a_{i_1,...,i_K} \subseteq O^K$,

$$f \in a \text{ if and only if } \big[\big(f(i_1), ..., f(i_K)\big) \in O^a_{i_1,...,i_K} \text{ for all } (i_1, ..., i_K) \in I^K \big].$$

This is a general definition, but most of relational axioms considered in different domains of social choice theory 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[9] $N \cup \{k\}$ which coincides with the preceding profile for agents in N. According to this definition, there typically are tuples of instances $(i_1, ..., i_K)$ for which $O^a_{i_1,...,i_K} = O$, that is, for which a imposes no restriction whatsoever. 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.

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

3.2 The probability to satisfy axioms

The vast majority of studies using stochastic preference models to generate instances have focused on *i)* punctual axioms and *ii)* a single axiom at a time. It is, however, possible to define the *mass* of instances for which a rule satisfies simultaneously several relational or punctual axioms.

[5] (I, O) is fixed throughout this paper. All the objects we consider depend on it but this dependence is omitted.

[6] Throughout this paper, 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.

[7] We stress that, as in this example, typically, singletons are only special cases of elements of O.

[8] Our purpose is not here to give an explicit definition of these conditions. See [17] (section 3.3) or [20] (sections 7.1 and 7.4).

[9] $k \notin N$.

Let A be a finite set of axioms and $f : I \to O$ a rule. Vectors $(p^f_S)_{S \subseteq A} \in [0,1]^{2^J}$ such as the one opening this section[10] are obtained in the following manner.

Assume $a \in A$ is punctual. Let $D^f(a) = \{i \in I, f(i) \in O^a_i\}$ denote the set of instances whose image by f meets the requirements imposed in a. If there is a $\sigma-$algebra ξ^I defined on I such that $D^f(a)$ is measurable, then the degree to which f satisfies a according to the probability measure μ defined on (I, ξ^I) is simply $\mu\big(D^f(a)\big)$. More generally, for any $a \in A$, let $K^a \in \mathbb{N}$ denote the number of instances associated to a in the definition of a relational axiom above —once again, $K^a = 1$ when a is punctual. For all $\emptyset \neq S \subseteq A$, we let $K^S = max_{a \in S} K^a$. Under the appropriate **measurability assumption**[11], letting μ^{K^S} denote the product measure on I^{K^S} obtained from μ, and $D^f(a) = \big\{(i_1, ..., i_{K^a}) \in I^{K^a}, \big(f(i_1), ..., f(i_{K^a})\big) \in O^a_{i_1,...,i_{K^a}}\big\}$ for $a \in A$, the **degree** to which f satisfies S is given by:

$$p^f_S = \mu^{K^S}\left(\left\{\big(i_1, ..., i_{K^S}\big) \text{ such that } \big(i_1, ..., i_{K^a}\big) \in D^f(a) \text{ for } a \in S\right\}\right)$$

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

4 How to measure the performance of rules ?

How can one assess the performance of a rule f and, importantly, compare it with other rules, based on the probabilities to satisfy axioms in A, $(p^f_S)_{S \subseteq A}$?

We address this issue by constructing a performance criterion defined for any p in the **set of possible probability vectors**, a subset of $[0,1]^{2^J}$ characterised by consistency conditions relating the probability to simultaneously satisfy all axioms in S and the probability to simultaneously satisfy all axioms but one in S:

$$P = \left\{(p_S)_{S \subseteq A}, p_\emptyset = 1, max\Big\{0; max_{a \in S}\big\{p_{S \setminus a} - (1 - p_a)\big\}\Big\} \leq p_S \leq min_{a \in S}\big\{p_{S \setminus a}\big\} \text{ for all } S \neq \emptyset\right\}.$$

The right-hand side inequality states that for $S \subseteq A$, the probability to satisfy all the axioms in S cannot be superior to any of the probabilities to satisfy all of them but one, that is to any probability in $(p_{S \setminus a})_{a \in S}$.

The probability to satisfy all axioms in S is also constrained below by $(p_{S \setminus a})_{a \in S}$. Select $a \in S$; taking $p_{S \setminus a}$ and p_a as given, the worst case in terms of probability to satisfy $S \setminus a$ and a happens when the intersection of the sets of profile 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 otherwise.[12] Hence the left-hand side inequality.

[10] For any rule, the value associated to the empty set is 1, that is why we omitted it in the opening example.

[11] For any $a \in A$, I^{K^a} is endowed with the product $\sigma-$algebra inherited from ξ^I, denoted $\xi^{I^{K^a}}$, and $D^f(a)$, defined in the main text, is measurable in (I^{K^a}, ξ^{K^a}).

[12] In the example we presented at the beginning, given that the sets of profiles 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 0.2.

4.1 Normative desirability

As axioms most often reflect normative principles of (collective) decision 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 to the eye of a researcher or a designer than another one, but *synergies* are likely to emerge in the combination of axioms, if only through the fact that axioms are particular formulations of general principles that may be deemed normatively independent from one another. For instance, in the microeconomic allocation framework, the satisfaction of 'efficiency' may be more or less valued than the one of 'envy-freeness', and, furthermore, the value of the satisfaction of another fairness criterion such as 'egalitarian equivalence', given the satisfaction of 'envy-freeness', 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 plausibility of truthful revelation implied in this axiom is all the more important as other axioms involving requirements on preferences are satisfied[13], and, conversely, the satisfaction of these other axioms seems all the more valuable that it is likely that they are applied to *real* preferences.

This observation leads us to allow the *intrinsic valuation* of a combination $S \subseteq A$ of 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** u on A:

$$U = \Big\{(u_S)_{S \subseteq A}, u_\emptyset = 0, u_T \leq u_S \text{ if } T \subseteq S, \text{ for all } S\Big\} \subset \mathbb{R}^{2^J}.$$

The weak monotonicity assumption, as regard to inclusion, embedded in the use of capacities, can be interpreted as relying on the assumption that all axioms under consideration are normatively desirable. The **set of super-additive capacities** will play an important role in the subsequent developments; it is given by

$$\hat{U} = \Big\{u \in U, u_S \geq u_T + u_{T'} \text{ if } T, T' \subseteq S \text{ and } T \cap T' = \emptyset\Big\}.$$

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 'envy-freeness' as we suggested above. From a general point of view, we see the use of super-additive valuations as the one most adapted to the typical theoretical problems in normative economics where each considered axiom is a particular formulation of general and independent ethical principles, in other words, where no two axioms are *subtitutes.*[14] The set of strictly super-additive capacities is denoted $\hat{U}_{st}$.

4.2 Characterisation of the measure

4.2.1 A measure for the reference case of complementary axioms

Considering $(u,p) \in U \times P$, as in example (1) below, the most intuitive **measure** $m : U \times P \to \mathbb{R}$, for the overall performance of a rule to which p is associated would consist in taking the standard arithmetic mean $\bar{m}(u,p) = \sum_{S \subseteq A} u_S.p_S$. However, the reader will see in this example that such

[13]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*[2])), but, clearly, the value of this axiom mainly lies in its interaction with other properties.

[14]In the example we gave involving 'envy-freeness' and 'egalitarian equivalence', a super-additive valuation would not capture the effect we described.

a measure *double counts* the satisfaction of a_1. Indeed, *i)* $p_{a_1} = p_{a_1 a_2}$ while *ii)* $u(a_1a_2) \geq u(a_1) + u(a_2)$ —actually, $u \in \hat{U}$:[15]

	$\emptyset$	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
u	0	1	1	1	4	3	3	6
p	1	0.7	0.9	0.5	0.7	0.3	0.4	0.25

(1)

In words, *i)* the rule satisfies a_1 with exactly the same probability as it satisfies a combination of a_1 and an other axiom while *ii)* the intrinsic valuation of the combination incorporates the separate intrinsic valuations of a_1 and of the other axiom. As a consequence, it is questionable that a_1 should *influence* the measure.

Based on this observation, restricting attention first to valuations in $\hat{U}$, for which the analogue of *ii)* holds for any combination, we propose a measure that keeps this simple additive form while 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 a superset of it. For any $S \subset A$, let $\hat{p}_S = max_{T:S\subset T}\, \{p_T\}$ and let $\hat{p}_A = 0$ —$\hat{p}_S$ thus gives the maximal probability associated to a superset of S. We define the following measure:

$$\hat{m} : \hat{U} \times P \to \mathbb{R}$$
$$(u,p) \mapsto \sum_{S \subseteq A} u_S.(p_S - \hat{p}_S).$$

The weight associated to a_1 in example (1) according to $\hat{m}$ is 0. This measure weights the valuation of each combination by the difference between its probability of satisfaction and the maximal probability of satisfaction among its supersets —informally, when $u \in \hat{U}$, the intrinsic valuation of S is incorporated in the valuation of any superset T of S, so that the real impact of S on the measure of performance of this rule should be determined by the differential in probability of satisfaction between S and T ($p_S \geq p_T$). In the following example, a rule associated to p' performs better than a rule associated to p according to $\hat{m}$, while the initial weighted average $\bar{m}$ yields the opposite conclusion.

	$\emptyset$	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
u	0	1	1	1	5	5	5	15
p	1	0.7	0.7	0.7	0.6	0.6	0.6	0.6
p′	1	0.7	0.7	0.7	0.7	0.7	0.7	0.4

(2)

The measure $\bar{m}$ weights each u_S, of $S \subset A$ with $|S| = 2$, by the associated probability p_S, while, according to p, the satisfaction of A happens with exactly the same probability as the one of S, and while, according to u, the intrinsic valuation of A incorporates that of S and of $a \notin S$. On the contrary, this fact is taken into account in measure $\hat{m}$ according to which each such u_S is given weight 0, whereas is is given weight 0.3 when p' is considered.

The measure $\hat{m}$ can be simply characterised by the three following properties.

Additivity and Positive Homogeneity (APH): For all $u, u' \in \hat{U}$, $\lambda > 0$, $p \in P$, we have $m(\lambda u + u', p) = \lambda m(u, p) + m(u', p)$.

As we mentioned earlier, we look for a measure which keeps the simple form of the arithmetic mean above; (APH) guarantees such a feature as it stipulates that for a fixed probability vector, the performance measure associated to an affine combination of two intrinsic valuations is simply the affine combination of the performance measures obtained for each of these valuations.

[15] Any set of axioms $S = \{a_1, ..., a_K\} \subseteq A$ is denoted $a_1...a_K$ in all the tables we use.

Marginal Probability (MP): Let $u \in \hat{U}_{st}$, $p, p' \in P$ and $\emptyset \neq S \subseteq A$ such that $p_S - \hat{p}_S = p'_S - \hat{p}'_S$. Let $u^S \in \hat{U}$ such that $u^S_S \neq u_S$, $u^S_T = u_T$ for all $T \neq S$. Then,

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

The introduction of this property roots in the anomaly of double counting we identified. Valuations u and u^S only differ in their component associated to S. This principle states that the impact of S on measure m under p and the impact of S under p', measured by the difference $m(u^S, p) - m(u, p)$ and $m(u^S, p') - m(u, p')$ respectively, are equal whenever the considered rule presents the same *marginal probability for S* under p and under p', *i.e*, whenever $p_S - \hat{p}_S = p'_S - \hat{p}'_S$. It implies in particular that whenever a rule presents a 0 marginal probability for a certain combination of axioms S under p, such as a_1 in example (1), then, for u and u^S, S has the same impact on m under p as under the vector $(1, \mathbf{0}_{2^J-1})$.[16] The assumption that u belongs to $\hat{U}_{st}$ ensures that we can always define such a valuation u^S in $\hat{U}$.

Remark: We consider that, given this interpretation of (MP), it makes more sense to require that the stated property hold *for any* $u \in \hat{U}_{st}$ and any associated u^S; it would, however, be sufficient to require that there exists such $u \in \hat{U}_{st}$ and a corresponding $u^S \in \hat{U}$, as inspection of the proof shows.

Expectation for Perfect Superposition (EPS): For any $\lambda \in [0, 1]$, $\emptyset \neq S \subseteq A$, let $p^{\lambda,S} \in P$ defined by $p^{\lambda,S}_{S'} = \lambda$ if $S' \subseteq S$ and $p^{\lambda,S}_{S'} = 0$ otherwise, $S' \neq \emptyset$, and $p^{\lambda,S}_{\emptyset} = 1$. For all $u \in \hat{U}$, we have $m(u, p^{\lambda,S}) = \lambda u_S$.

Here is a representation of $p = p^{0.6,a_1a_3}$:

	$\emptyset$	a_1	a_2	a_3	a_1a_2	a_1a_3	a_2a_3	A
p	1	0.6	0	0.6	0	0.6	0	0

It is not difficult to see that such $p^{\lambda,S}$ indeed belongs to P. We use the expression *perfect superposition* because $p^{\lambda,S}$ obtains when the sets of profiles of instances for which each axiom in S is satisfied are identical —up to a 0-probability-measure-set— and the set of profiles of instances for which each axiom outside S is satisfied has a 0 probability measure. As all the subsets T of S are satisfied with the same probability, and as no other combination of axioms present a strictly positive probability, the case we made above for the prevention of redundancy implies that only S should impact measure $m(\cdot, p^{\lambda,S})$ on $\hat{U}$. In other words, only u_S should be taken into account under such probability vector, and, thus, the 'expected valuation' λu_S is the natural candidate for the measure's value.

We obtain the following characterisation result, the proof of which, as that of subsequent results, can be found in the appendix:

Theorem 1. *A measure* $m : \hat{U} \times P \to \mathbb{R}$ *satisfies additivity and positive homogeneity, marginal probability and expectation for perfect superposition if and only if it coincides with* $\hat{m}$.

The interpretation of how weights associated to $p \in P$ are computed according to $\hat{m}$ is akin to that of a Choquet integral, except that *inclusion* is taken into account in $\hat{m}$: u_A is weighted by p_A, then we turn to any S with $|S| = J - 1$ and u_S is weighted by $p_S - p_A$, then to any T with $|T| = J - 2$ where u_T is weighted by $p_T - max_{S \supset T}\{p_S\}$, *etc.*

When commenting on example (1), the identification of a redundancy in the measure was not based on the fact that u was supper-additive, but simply on the fact that there existed a combination of axioms a_1a_2 such that $u_{a_1a_2} \geq u_{a_1} + u_{a_2}$ and $p_{a_1} = p_{a_1a_2}$. How should the measure behave when such inequality is not verified, that is when a substitution effect between axioms arises, such as the one we described between 'envy-freeness' and 'egalitarian equivalence' ?

[16] For all $b \in \mathbb{R}$, $K \in \mathbb{N}$, $\mathbf{b}_K$ denotes the element of $\mathbb{R}^K$ whose components are all equal to b.

4.2.2 A measure for the general case

Consider the computation of $(u,p) \in U \times P \mapsto \sum_{S \subseteq A} u_S.p_S \in \mathbb{R}$ in example (3) below.

$$\begin{array}{ccccccccc} & \emptyset & a_1 & a_2 & a_3 & a_1a_2 & a_1a_3 & a_2a_3 & A \\ \mathbf{u} & 0 & 3 & 2 & 3 & 6 & 6 & 4 & 9 \\ \mathbf{p} & 1 & 0.7 & 0.9 & 0.5 & 0.7 & 0.3 & 0.4 & 0.25 \end{array} \quad (3)$$

The measure double counts the satisfaction of a_2a_3. Indeed, *i)* $p_{a_2a_3} = 1-(1-p_{a_2})-(1-p_{a_3}) > 0$, while *ii)* $u(a_2a_3) < u(a_2) + u(a_3)$. In words, *i)* the rule satisfies a_2a_3 with the minimum *possible* probability considering the *actual* probabilities with which it satisfies respectively a_2 and a_3 while *ii)* this combination's intrinsic valuation is incorporated in the sum of the separate intrinsic valuations of a_2 and of a_3. Therefore, it is questionable that a_2a_3 should influence the measure of the performance of this rule.

Let us use the shortcut 'u_S is super-additive' to refer to the fact that $u_S \geq \sum_{k \in K} u_{T_k}$ where $(T_k)_{k \in K}$ is any partition of S.[17] We now propose a measure $m^* : U \times P \to \mathbb{R}$ that, once again, keeps the same simple form of the arithmetic mean while taking into account *i)* when u_S is supper-additive, the difference between the probability with which a rule satisfies a given combination of axioms S and the probability with which it satisfies supersets of it, and *ii)* when it is not, the difference between the probability with which a rule satisfies S and the probability with which it satisfies supersets of it, and subsets of it. More precisely, m^* is defined based on the requirement that the impact of a combination S on the performance measure of a given rule associated to $p \in P$ should only depend on whether u_S is super-additive or not, so that in particular the measure coincides with $\hat{m}$ on $\hat{U} \times P$.

For $u \in U$, $S \subseteq A$, let $Supper_{S,u} = \{T = S \cup S', u_T \geq u_S + u_{S'}, S' \neq \emptyset, S \cap S' = \emptyset\}$, and $Sub_{S,u} = \{(T_k)_{k \in K}, \sum_{k \in K} u_{T_k} > u_S, (T_k)_{k \in K} \text{ is a partition of } S\}$. For any $p \in P$, we define

$$\overline{p}^u_S = \underset{T \in Supper_{S,u}}{max} \{p_T\} \text{ if } Supper_{S,u} \neq \emptyset$$
$$0 \text{ otherwise.}$$
$$\underline{p}^u_S = max\left\{0; \underset{(T_k)_{k \in K} \in Sub_{S,u}}{max} \{1 - \sum_{k \in K} (1 - p_{T_k})\}\right\} \text{ if } Sub_{S,u} \neq \emptyset$$
$$0 \text{ otherwise.}$$

That is, $\overline{p}^u_S$ denotes the maximal probability among the sets that are composed of S and of axioms that are complementary to those in S, and $\underline{p}^u_S$ denotes the minimum *possible* probability to satisfy S given the *actual* probabilities to satisfy disjoint sets of axioms that are substituable and whose reunion is S.

The measure we define weights the intrinsic valuation of a combination S, u_S, by the difference between p_S and $max\{\overline{p}^u_S; \underline{p}^u_S\}$, the value of the latter depending crucially on whether u_S is supper-additive:

$$m^* : U \times P \to \mathbb{R}$$
$$(u,p) \mapsto \sum_{S \subseteq A} u_S.\left(p_S - max\{\overline{p}^u_S; \underline{p}^u_S\}\right).$$

Note that whenever u_S is super-additive, the associated weight $p_S - max\{\overline{p}^u_S; \underline{p}^u_S\}$ is equal to $p_S - \overline{p}^u_S$, which, in particular, equals $p_S - \hat{p}_S$ if $u \in \hat{U}$. Note also that $u_a = u_\emptyset + u_a$, so that the

[17] 'u_S is strictly super-additive' refers to the case where this inequality is strict.

weigh associated to any axiom a is $p_a - \overline{p}^u_a$. Finally, the weight associated to a_2a_3 in example (3) according to m^* is 0, as $p_{a_2a_3} = max\{\overline{p}^u_{a_2a_3}; \underline{p}^u_{a_2a_3}\} = \underline{p}^u_{a_2a_3} = 0.4$.

The interpretation of the weights when u_S is super-additive is, of course, the same as for $\hat{m}$ on $\hat{U}$. Simply note that to define m^*, we have to account for cases in which u_S is super-additive and $Supper_{S,u}$ is empty, *i.e*, in which there is no set of axioms that are complementary to those in S. In that case, the weight associated to S is p_S.

Consider now S such that $u_S < \sum_{k\in K} u_{T_k}$ for a partition $(T_k)_{k\in K}$ of S. For the same reason as above, it is still necessary to take into account the probability to satisfy supersets of S, that is, to compare p_S and $\overline{p}^u_S$. Furthermore, the intrinsic valuation of S is incorporated in the sum of the separate valuations of each T_k, while, by definition of a capacity, $u_S \geq u_{T_k}$, for all $k \in K$. Therefore, the real impact of S on the measure of the performance of the rule should also depend on the differential between the actual probability p_S and the minimal possible probability for S to be satisfied given the actual probabilities $(p_{T_k})_{k\in K}$.

To summarise, $p_S - max\{\overline{p}^u_S; \underline{p}^u_S\}$ appears as the *relevant marginal probability* to consider since $max\{\overline{p}^u_S; \underline{p}^u_S\}$ corresponds to the maximal probability associated to an object whose intrinsic valuation incorporates that of S.

The measure m^* is characterised by four intuitive properties; three of them corresponding to generalisations of the ones we introduced in the previous section.

We say that $u, u' \in U$ belong to the same **interaction class** if for all partition $(T_k)_{k\in K}$ such that $u_S \geq \sum_{k\in K} u_{T_k}$, we have $u'_S \geq \sum_{k\in K} u'_{T_k}$. That is, two intrinsic valuations belong to the same interaction class if any disjoint sets of axioms who are (weakly)[18] complementary under one valuation are complementary under the other. The same correspondance holds for substitutes.

Additivity and Positive Homogeneity on Interaction classes (APHI): For all $u, u' \in U$ belonging to the same interaction class, $\lambda > 0$, $p \in P$, we have $m(\lambda u + u', p) = \lambda m(u, p) + m(u', p)$.

This simplicity requirement states that if u and u' represent the same interaction between axioms, in the sense given in the definition of an interaction class, then, for a fixed probability vector, the measure value associated to an affine combination of u and u' is simply the affine combination of the measure values taken on u and u'.

Marginal Probability Generalised (MPG): Let $u \in U$ such that, for all $\emptyset \neq S \subseteq A$, u_S is either strictly super-additive, or not super-additive. Let $\emptyset \neq S \subseteq A$ and $p, p' \in P$ such that $p_S - max\{\overline{p}^u_S; \underline{p}^u_S\} = p'_S - max\{\overline{p}'^u_S; \underline{p}'^u_S\}$. Let $u^S \in U$ with $u^S_S \neq u_S$ and $u^S_{S'} = u_{S'}$ for $S' \neq S$, such that u^S belongs to the same interaction class as u. Then,

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

The assumption on u simply guarantees that we can always find such $u^S \in U$. This principle states, as its counterpart in the previous subsection, that the impact that a set of axioms S should have on the measure of the performance of a rule associated to p should only depend on the relevant marginal probability, the value of the latter depending on whether u_S is super-additive.

Expectation for Perfect Superposition Generalised (EPSG): For all $S \subseteq A$, $u \in U$ such that u_S is supper-additive, for $\lambda \in [0, 1]$, we have $m(u, p^{\lambda,S}) = \lambda u_S$.

Comparability (C): Let $u, u' \in U$, for all $\emptyset \neq S \subseteq A$ such that u_S is strictly supper-additive and u'_S is not supper-additive. Let $p \in P$ such that $p_S - max\{\overline{p}^u_S; \underline{p}^u_S\} = p_S - max\{\overline{p}^{u'}_S; \underline{p}^{u'}_S\}$. Then,

$$m(u^{\epsilon,S}, p) - m(u, p) = m(u'^{\epsilon,S}, p) - m(u', p),$$

[18] We use this term to refer to the case where the preceding inequality is an equality.

where capacities $u^{\epsilon,S}$ and $u'^{\epsilon,S}$ are defined by $u^{\epsilon,S}_S = u_S - \epsilon$ and $u^{\epsilon,S}_{S'} = u_{S'}$ for $S' \neq S$, and $u'^{\epsilon,S}_S = u'_S - \epsilon$ and $u'^{\epsilon,S}_{S'} = u'_{S'}$ for $S' \neq S$, $\epsilon > 0$ being such that $u^{\epsilon,S}_S$ is super-additive and $u'^{\epsilon,S}_S$ is not.

This principle guarantees that relevant marginal probabilities are treated symmetrically: p is such that the relevant marginal probability under u is equal to the relevant marginal probability under u' and it is required that when we affect u_S and u'_S by the same $\epsilon-$reduction, the measure should be impacted to the same extent.

Remark: Given this interpretation, requiring this property to hold for all $p \in P$ such that the relevant marginal probabilities are identical seems legitimate; we could however restrict it to hold only for $p^{\lambda,S}$, $\lambda \in [0,1]$, $S \subseteq A$, without changing the characterisation result below (we refer to this restricted version as *comparability for perfect superposition* and define it explicitly in the appendix).

Theorem 2. *A measure $m : U \times P \to \mathbb{R}$ satisfies additivity and positive homogeneity on interaction classes, marginal probability generalised , expectation for perfect superposition generalised and comparability (or comparability for perfect superposition) if and only if it coincides with m^*.*

5 How to explain the logical performance of rules ?

The *logical* performance of any rule f with respect to the set of axioms A is defined as the probability for f to simultaneously satisfy all the axioms in A, p^f_A —we focus only on the probabilities of satisfaction in this section. How to account for p^f_A using $(p^f_S)_{S\subseteq A}$? In other words, given the overall degree of compatibility p^f_A, how can we analyse how axioms interact with each other under f ? A first step towards the answer consists in building a measure of how compatible an axiom $a \in A$ is with the other axioms in $A \setminus a$ under f, and this can be done by determining how a contributes to p^f_A compared to axioms in $A \setminus a$. 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.[19]

Define $V = 1 - P = \{1 - p,\ p \in P\}$, of generic element v. Note that $v_\emptyset = 1 - p_\emptyset = 0$: v is a *cooperative game associated to the set of axioms A*. $v_A = 1 - p_A$ represents the incompatibility between axioms in A, and this magnitude must be distributed among them. An **incompatibility measure** with respect to the set A of axioms[20] is defined as a **solution** on (A, V), that is a mapping

$$\begin{aligned} \psi : V &\to \mathbb{R}^J \\ v &\mapsto \big(\psi_a(v)\big)_{a\in A}. \end{aligned}$$

For $a \in A, S \subseteq A \setminus a$, given the definition of V, $v_{S\cup a} - v_S = p_S - p_{S\cup a}$ 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 to an axiom a be a function of the cost exerted by a on all $S \subseteq A \setminus a$:

Marginality (M): For any $a \in A$, $v, v' \in V$, if $v_{S\cup a} - v_S = v'_{S\cup a} - v'_S$ for all $S \subseteq A \setminus a$, $\psi_a(v) = \psi_a(v')$.

This principle corresponds to the essential property expressed in Young's *strong monotonicity* axiom in his classical characterisation of the Shapley value on the subspace of games associated to a fixed group of players ([23]).

[19]Closely related to our approach is the literature focusing on the use of (modified versions of) the Shapley value in feature attribution problems ([15]).

[20]As a reminder, we fixed $A = \{a_1, ..., a_J\}$.

Two classical axioms introduced in cooperative game theory are necessary to interpret ψ as a relevant measure in the problem we consider.

Incompatibility Allocation (IA): For any $v \in V$, $\sum_{a \in A} \psi_a(v) = v(A)$.

Of course, this principle corresponds to the standard efficiency principle. In our framework, it is this requirement that makes it possible to interpret ψ as allocating the incompatibility $1 - p_A$ among axioms in A.

Symmetry (S): For any $v \in V$ any permutation $\pi : A \to A$, let v^π the game defined by $v^\pi_S = v_{(\pi(a))_{a \in S}}$. Then, $\psi_a(v) = \psi_{\pi(a)}(v^\pi)$.

This principle states that the evaluation should not be biased towards some axiom.

We define the **Shapley incompatibility measure** as the restriction to (A, V) of the usual Shapley value:

$$\varphi_a(v) = \sum_{S \subseteq A \setminus a} \frac{|S|!(J - |S| - 1)!}{J!}(v_{S \cup a} - v_S) \text{ for all } v \in V, a \in A.$$

The reader will not be surprised that the Shapley incompatibility measure satisfies these three properties; the fact that it is actually characterised by them is not direct however. Indeed, P is defined by a very specific consistency condition, and it turns out that no usual basis of the set of cooperative games $G = \{u \in \mathbb{R}^{2^J}, u_\emptyset = 0\}$ —such as the family of *unanimity games*, of *conjugate unanimity games*, of *Yokote games*, or of *Walsh games*— belongs to $V = 1 - P$. Consider for example $A = \{a_1, a_2, a_3\}$ and the unanimity game $u^{a_1a_3}$ defined by $u^{a_1a_3}_T = 1$ if $a_1a_3 \subseteq T$, $u^{a_1a_3}_T = 0$ otherwise, then $p = 1 - v$ is such that $p_{a_1} = p_{a_3} = 1$ but $p_{a_1a_3} = 0$, that is $p \notin P$, and thus $u^{a_1a_3} \notin V$. Therefore, common techniques using standard bases can not be directly applied here.

We use the extreme points of V to proceed to the following characterisation (see appendix):

Theorem 3. *An incompatibility measure $\psi : V \to \mathbb{R}^J$ satisfies marginality, incompatibility allocation and symmetry if and only if it coincides with the Shapley incompatibility measure $\varphi : V \to \mathbb{R}^J$.*

Remark: The Shapley measure is also characterised by *incompatibility allocation*, *symmetry*, and direct adaptations for a solution restricted to V of the *null-player* and *additivity and positive homogeneity* axioms considered by Shapley in his original work ([19]) —the reader may find their explicit definitions in the appendix. However, *marginality*, *incompatibility allocation* and *symmetry* are in our view the very axioms that support the interpretation of ψ as an adapted measure.

6 Conclusion

In consistency with increasingly used simulation methods, we defined a general notion for the degree to which a rule satisfies a set of axioms. Armed with it, we proposed *i)* a criterion to evaluate rules' performance taking into account the normative desirability of axioms and of synergies effects between them, and *ii)* a criterion to determine, for a given set of axioms, the role of each axiom in the degree to which a certain rule simultaneously satisfies all of them.

In further stages of this work, priority will be given to applications of these two criteria in problems characterised by analytically established incompatibilities between axioms.

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.

References

[1] Fuad Aleskerov, Daniel Karabekyan, M Remzi Sanver, and Vyacheslav Yakuba. On the manipulability of voting rules: The case of 4 and 5 alternatives. Mathematical Social Sciences, 64(1):67–73, 2012.

[2] Georgy Artemov, Yeon-Koo Che, and Yinghua He. Strategic 'mistakes': Implications for market design research. Working paper.

[3] Felix Brandt, Markus Brill, and Hans Georg Seedig. On the fixed-parameter tractability of composition-consistent tournament solutions. In Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[4] Felix Brandt, Christian Geist, and Hans Georg Seedig. 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), 2014.

[5] Felix Brandt, Christian Geist, and Martin Strobel. 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, pages 385–393, 2016.

[6] Donald Campbell and Jerry Kelly. The possibility-impossibility boundary in social choice. In Social Choice Re-examined: Volume 1 Proceedings of the IEA Conference held at Schloss Hernstein, Berndorf, near Vienna, Austria, pages 179–204. Springer, 1997.

[7] Donald E Campbell and Jerry S Kelly. Trade-off theory. The American Economic Review, 84(2):422–426, 1994.

[8] Donald E Campbell and Jerry S Kelly. Social choice trade-off results for conditions on triples of alternatives. Mathematical Social Sciences, 77:42–45, 2015.

[9] Mostapha Diss and Eric Kamwa. Simulations in models of preference aggregation. Œconomia. History, Methodology, Philosophy, (10-2):279–308, 2020.

[10] Edith Elkind. Restricted preference domains in social choice: Two perspectives. In Algorithmic Game Theory: 11th International Symposium, SAGT 2018, Beijing, China, September 11-14, 2018, Proceedings 11, pages 12–18. Springer, 2018.

[11] William V Gehrlein. Condorcet's paradox. Theory and Decision, 15(2):161–197, 1983.

[12] William V Gehrlein, Dominique Lepelley, et al. Elections, voting rules and paradoxical outcomes. Springer, 2017.

[13] Jean-François Laslier. In silico voting experiments. In Handbook on approval voting, pages 311–335. Springer, 2010.

[14] Dominique Lepelley, Ahmed Louichi, and Fabrice Valognes. Computer simulations of voting systems. Advances in Complex Systems, 3(01n04):181–194, 2000.

[15] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017.

[16] Hervé Moulin and William Thomson. Can everyone benefit from growth?: Two difficulties. Journal of Mathematical Economics, 17(4):339–345, 1988.

[17] Hervé Moulin. Fair division in the internet age. Annual Review of Economics, 11:407–441, 2019.

[18] Alvin E Roth and Elliott Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. American economic review, 89 (4):748–780, 1999.

[19] Lloyd S Shapley. A value for n-person games. 1953.

[20] William Thomson. Fair allocation rules. Working paper.

[21] William Thomson. On the axiomatic method and its recent applications to game theory and resource allocation. Social Choice and Welfare, 18(2):327–386, 2001.

[22] Mark C Wilson. 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, pages 165–171, 2019.

[23] H Peyton Young. Monotonic solutions of cooperative games. International Journal of Game Theory, 14(2):65–72, 1985.

Pierre BARDIER
Paris School of Economics (PSE) and Ecole Normale Supérieure de Paris (ENS Ulm)
Paris, France
Email: pierre.bardier@ens.fr

Appendix

Theorem 1

Proof of Theorem 1

Theorem 1. *A measure $m : \hat{U} \times P \to \mathbb{R}$ satisfies additivity and positive homogeneity, marginal probability and expectation for perfect superposition if and only if it coincides with $\hat{m}$.*

Proof. The *if* part is readily checked.

For any $p \in P$, we let $m^p : \hat{U} \to \mathbb{R}$ denote $u \mapsto m(u,p)$.

Let $p \in P$. $\hat{U}$ is a pointed convex cone of $\mathbb{R}^{2^J}$. Then, m^p, which, by (AC), is additive and positively homogeneous on this convex cone, can be uniquely extended to a linear form $\tilde{m}^p : Vect(\hat{U}) \to \mathbb{R}$, where $Vect(\hat{U})$ denotes the vectorial subspace of $\mathbb{R}^{2^J}$ spanned by $\hat{U}$. Therefore[21], there are unique real numbers $\alpha_S^p, S \subseteq A$, such that for all $p \in P$, $u \in \hat{U}$,

$$m(u,p) = \sum_{S \subseteq A} u_S.\alpha_S^p.$$

Now fix[22] $\emptyset \neq S \subseteq A$ and let $\alpha_S : p \in P \mapsto \alpha_S^p \in \mathbb{R}$. Let $p, p' \in P$ such that $p_S - \hat{p}_S = p'_S - \hat{p}'_S$. Consider $u \in \hat{U}_{st}$ and $u^S \in \hat{U}$ as introduced in (MP). Then (MP) implies that $\sum_{T \subseteq A} u_T^S.\alpha_T^p - \sum_{T \subseteq A} u_T.\alpha_T^p = \sum_{T \subseteq A} u_T^S.\alpha_T^{p'} - \sum_{T \subseteq A} u_T.\alpha_T^{p'} \Leftrightarrow \big(u_S^S - u_S\big)\alpha_S(p) = \big(u_S^S - u_S\big)\alpha_S(p') \Leftrightarrow \alpha_S(p) = \alpha_S(p')$.

In other words, the function α_S can be written as a function of the difference[23] $p_S - \hat{p}_S$: more precisely,

$$\alpha_S(p) = \hat{\alpha}_S(p_S - \hat{p}_S) \text{ for all } p \in P$$

with $\hat{\alpha}_S : [0,1] \to \mathbb{R}$. Moreover, $\hat{\alpha}_S(0) = 0$. Indeed, for these same u and u^S, considering $p^{0,S}$ as defined in (EPS) and applying this axiom yields $\sum_{S \subseteq A} u_T \hat{\alpha}_T(0) = 0 = \sum_{S \subseteq A} u_T^S \hat{\alpha}_T(0)$; which implies $\hat{\alpha}_S(0) = 0$.

Consider now any $p^{\lambda,S} \in P$, $\lambda \in [0,1]$, as introduced in (EPS), and $u \in \hat{U}$ such that $u_S \neq 0$. The equality $m(u, p^{\lambda,S}) = \lambda u_S$ then writes

$$\sum_{T \subseteq A} u_T \hat{\alpha}_S\big(p_T^{\lambda,S} - \hat{p}_T^{\lambda,S}\big) = \lambda u_S$$
$$\Leftrightarrow u_S \hat{\alpha}_S(\lambda) = \lambda u_S.$$

As $p^{\lambda,S} \in P$ can be defined for any $\lambda \in [0,1]$, the preceding equality implies that $\hat{\alpha}_S(\lambda) = \lambda$ for all $\lambda \in [0,1]$; which concludes. ☐

Independence of the axioms

Adopting the following modification of the marginal probability principle, and requiring that a null probability vector should give a null measure, readily yields a general class of measures, those for which there is $\alpha : [0,1] \to \mathbb{R}$, $\alpha(0) = 0$, such that $m(u,p) = \sum_{S \subseteq A} u_S.\alpha(p_S - \hat{p}_S)$.

[21] Because $\tilde{m}^p$ is a linear form on $Vect(\hat{U})$, there is a unique $\alpha^p \in \mathbb{R}^{2^J}$ such that $\tilde{m}^p(a) = a \cdot \alpha^p$ for all $a \in Vect(\hat{U})$—$(\,\cdot\,)$ is used to denote the usual scalar product on $\mathbb{R}^{2^J}$.

[22] The linear expression of m and the fact that $u_\emptyset = 0$ allow to restrict to sets different from $\emptyset$.

[23] With the previously mentioned convention $\hat{p}_A = 0$.

Symmetric Marginal probability (SMP): Let $u \in \hat{U}_{st}$, $p, p' \in P$ and $\emptyset \neq S, S' \subseteq A$ such that $p_S - \hat{p}_S = p'_{S'} - \hat{p}'_{S'}$. Let $\epsilon > 0$ such that for $u^S \in \hat{U}$ defined by $u^S_S = u_S + \epsilon$ and $u^S_T = u_T$ for all $T \neq S$ and $u^{S'} \in \hat{U}$ such that $u^{S'}_S = u_{S'} + \epsilon$ and $u^{S'}_T = u_T$ for all $T \neq S$. Then,

$$m(u^S, p) - m(\bar{u}, p) = m(u^{S'}, p') - m(\bar{u}, p').$$

Normalisation (N): For all $u \in \hat{U}$, $m\big(u, (1, \mathbf{0}_{2^J-1})\big) = 0$.

The proof is a very simple adaptation of the preceding one and is omitted.

Any such measure which does not coincide with $\hat{m}$ satisfies (APH) and (MP) but not (EPS). The following measure satisfies (APH) and (EPS) but not (MP): $(u, p) \mapsto \sum_{S \subseteq A} u_S(p_S - \bar{p}_S)$, where $\bar{p}_S = \frac{\sum_{T=S\cup a, a\notin S} p_T}{\sum_{T=S\cup a, a\notin S} 1}$ for all $S \subset A$, and $\bar{p}_A = 0$, that is $\bar{p}_S$ is the average probability of satisfaction among supersets of S of cardinal $|S|+1$. The measure $(u, p) \mapsto max_{S\subseteq A} u_S(p_S - \tilde{p}_S)$, satisfies (MP) and (EPS) but not (APH).

Theorem 2

Proof of Theorem 2

Comparability for Perfect Superposition (CPS): Let $u, u' \in U$, $\emptyset \neq S \subseteq A$ such that u_S is strictly supper-additive and u'_S is not supper-additive. For all $\lambda \in [0, 1]$ such that such that $p^{\lambda,S}_S - max\{\overline{p}^{u\lambda,S}_S, \underline{p}^{u\lambda,S}_S\} = p^{\lambda,S}_S - max\{\overline{p}^{u'\lambda,S}_S, \underline{p}^{u'\lambda,S}_S\}$,

$$m(u^{\epsilon,S}, p^{\lambda,S}) - m(u, p^{\lambda,S}) = m(u'^{\epsilon,S}, p^{\lambda,S}) - m(u', p^{\lambda,S}),$$

where capacities $u^{\epsilon,S}$ and $u'^{\epsilon,S}$ are defined by $u^{\epsilon,S}_S = u_S - \epsilon$ and $u^{\epsilon,S}_{S'} = u_{S'}$ for $S' \neq S$, and $u'^{\epsilon,S}_S = u'_S - \epsilon$ and $u'^{\epsilon,S}_{S'} = u'_{S'}$ for $S' \neq S$, $\epsilon > 0$ being such that $u^{\epsilon,S}_S$ is super-additive and $u'^{\epsilon,S}_S$ is not.

Theorem 2. *A measure $m : U \times P \to \mathbb{R}$ satisfies additivity and positive homogeneity on interaction classes, marginal probability generalised, expectation for perfect superposition generalised and comparability (or comparability for perfect superposition) if and only if it coincides with m^*.*

Proof. The *if* part is readily checked.

For any $p \in P$, we let $m^p : U \to \mathbb{R}$ denote $u \mapsto m(u, p)$. For any $u \in U$, we let $C^u \subset U$ denote the interaction class to which u belongs.

Let $p \in P$. C^u is a convex cone of $\mathbb{R}^{2^J}$. Then, m^p, which, by (APHI), is additive and positively homogeneous on this convex cone, can be uniquely extended to a linear form $\tilde{m}^p : Vect(C^u) \to \mathbb{R}$. Therefore, for all $(u', p) \in C^u \times P$, there are unique real numbers α^{p,C^u}_S, $S \subseteq A$ such that

$$m(u', p) = \sum_{S\subseteq A} u'_S.\alpha^{p,C^u}_S.$$

Now fix[24] $\emptyset \neq S \subseteq A$ and let, for all $u \in U$, $\alpha^{C^u}_S : p \in P \mapsto \alpha^{p,C^u}_S \in \mathbb{R}$.

Consider $p, p' \in P, u, u^S \in U$ as introduced in (MPG), with u_S strictly super-additive, guaranteeing that we can find such u^S belonging to C^u. Then, by the same simple computation as the one used for (MP) in the preceding proof, and as $\overline{p}^u_S = max\{\overline{p}^u_S; \underline{p}^u_S\}$,

$$\alpha^{C^u}_S(p) = \hat{\alpha}^{C^u}_S(p_S - \overline{p}^u_S) \text{ for all } p \in P$$

[24]The linear expression of m and the fact that $u_\emptyset = 0$ allow to restrict to sets different from $\emptyset$.

with $\hat{\alpha}_S^{C^u}:[0,1]\to\mathbb{R}$. Moreover, $\hat{\alpha}_S^{C^u}(0)=0$, by the same argument as in the preceding proof, using (EPSG).

Consider now any $p^{\lambda,S}\in P$, $\lambda\in[0,1]$, as introduced in (EPSG), and $u\in U$ such that $u_S\neq 0$ is strictly supper-additive. The equality $m(u,p^{\lambda,S})=\lambda u_S$ then writes

$$\sum_{T\subseteq A} u_T\hat{\alpha}_S^{C^u}\big(p_T^{\lambda,S}-\overline{p}_T^{u\lambda,S}\big)=\lambda u_S$$
$$\Leftrightarrow u_S\hat{\alpha}_S^{C^u}(\lambda)=\lambda u_S.$$

As $p^{\lambda,S}\in P$ can be defined for any $\lambda\in[0,1]$, the preceding equality implies that $\hat{\alpha}_S^{C^u}(p_S-\overline{p}_S^u)=p_S-\overline{p}_S^u$. Finally, note that if $u_S=\sum_{k\in K}u_{T_k}$ for some partition $(T_k)_{k\in K}$ of S (and $u_S\geq\sum_{k'\in K'}u_{T_{k'}}$ for all partition $(T_{k'})_{k'\in K'}$ of S) there is $u'\in U$ belonging to C^u such that $u'_S\neq 0$ is strictly supper-additive.

A symmetric argument to the one of the second to last paragraph concludes to the fact that

$$\alpha_S^u(p)=\tilde{\alpha}_S^{C^u}\big(p_S-max\{\overline{p}_S^u;\underline{p}_S^u\}\big)\text{ for all }p\in P$$

with $\tilde{\alpha}_S^{C^u}:[0,1]\to\mathbb{R}$, for all $u\in U$ such that u_S is not supper-additive.

Let $u\in U$ such that u_S is not supper-additive and let $\bar{u}\in U$ such that $\bar{u}_S$ is strictly supper-additive. Let $0\leq\lambda\leq 0.5$ and consider $p^{\lambda,S}$; we have $p_S-max\{\overline{p}_S^{u\lambda,S},\underline{p}_S^{u\lambda,S}\}=\lambda=p_S^{\lambda,S}-max\{\overline{p}_S^{\bar{u}\lambda,S},\underline{p}_S^{\bar{u}\lambda,S}\}$. By (C), or (CPS), $\tilde{\alpha}_S^{C^u}\big(p_S-max\{\overline{p}_S^{u\lambda,S},\underline{p}_S^{u\lambda,S}\}\big)=\hat{\alpha}_S^{C^{\bar{u}}}\big(p_T^{\lambda,S}-max\{\overline{p}_S^{\bar{u}\lambda,S},\underline{p}_S^{\bar{u}\lambda,S}\}\big)=p_S^{\lambda,S}-\overline{p}_S^{\bar{u}\lambda,S}=p_S^{\lambda,S}-max\{\overline{p}_S^{\bar{u}\lambda,S},\underline{p}_S^{\bar{u}\lambda,S}\}$; which implies $\tilde{\alpha}_S^{C^u}(\lambda)=\lambda$. Let $0.5<\lambda\leq 1$; applying this same argument using $p^{1-\lambda,S}$ yields the desired conclusion. □

Independence of the axioms

Direct adaptations of the examples we gave to show the independence of our axioms in the previous section give, respectively, examples of axioms satisfying (APHI), (MPG), (C) but not (EPSG); (APHI), (C), (EPSG) but not (MPG); (MPG), (C), (EPSG) but not (APHI). The measure $(u,p)\mapsto\sum_{S\subseteq A}u_S.\beta_S^{u,p}$ with $\beta_S^{u,p}=p_S-\overline{p}_S^u$ if u_S is supper-additive and $\beta_S^{u,p}=3\big(p_S-max\{\overline{p}_S^u;\underline{p}_S^u\}\big)$ otherwise satisfies (APHI), (MPG), (EPSG) but not (C).

Theorem 3

Proof of Theorem 3

Theorem 3. *An incompatibility measure $\psi:V\to\mathbb{R}^J$ satisfies marginality, incompatibility allocation and symmetry if and only if it coincides with the Shapley incompatibility measure $\varphi:V\to\mathbb{R}^J$.*

Proof. The *if* part is readily checked. Consider the family of games $(\hat{v}^S)_{S\subseteq A}$ where each game v^S is defined by $\hat{v}_T^S=1$ if $T\not\subseteq S$ and $\hat{v}_T^S=0$ otherwise. Note that $\hat{v}^S$ belongs to V for all $S\subseteq A$ as $\hat{v}^S=1-p^{1,S}$ where $p^{1,S}$, defined in the previous section, belongs to P. We provide the table summarising $v=v^{a_1a_2}$ and $p=p^{1,a_1a_2}$ where $A=\{a_1,a_2,a_3\}$:

	$\emptyset$	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

Lemma 1. *P is a non-empty compact convex subset of $[0,1]^{2^J}$, and thus so is V.*

Proof. Convexity: Let $p^1, p^2 \in P$ and consider $0 < \alpha < 1$ and $\tilde{p} = \alpha p^1 + (1-\alpha)p^2$. Obviously, $\tilde{p}_\emptyset = 1$ and $\tilde{p}_S \geq 0$ for all $S \subseteq A$. Moreover,

$$\begin{aligned}\tilde{p}_S &\leq \alpha min_{a\in S}\{p^1_{S\backslash a}\} + (1-\alpha)min_{a\in S}\{p^2_{S\backslash a}\}\\ &\leq min_{a\in S}\{\alpha p^1_{S\backslash a} + (1-\alpha)p^2_{S\backslash a}\}.\end{aligned}$$

and

$$\begin{aligned}\tilde{p}_S &\geq \alpha max_{a\in S}\{p^1_{S\backslash a} - (1-p_a)\} + (1-\alpha)max_{a\in S}\{p^2_{S\backslash a} - (1-p_a)\}\\ &\geq max_{a\in S}\{\alpha(p^1_{S\backslash a} - (1-p_a)) + (1-\alpha)(p^2_{S\backslash a} - (1-p_a)\}.\end{aligned}$$

As a consequence, $\tilde{p} \in P$.

Compactness: P is bounded and closed in $[0,1]^{2^J}$. □

Lemma 2. *The set of extreme points of V is the family $(\hat{v}^S)_{S\subset A}$ (note the strict inclusion here).*

Proof. We already observed that this family lies in V, and it is obvious that each such $\hat{v}^S$ is extreme. These games —in addition to $\hat{v}^A$ which is not extreme obviously, as it is the null vector— are the only $\{0,1\}-$valued games in V. Indeed, let v a $\{0,1\}-$valued game which does not coincide with any $\hat{v}^S$, $S \subseteq A$. Then, there must exist $T, T' \subseteq A$ such that $T \subset T'$ and $v_{T'} = 0$ and $v_T = 1$, but then $1 - v_{T'} = 1$ and $1 - v_T = 0$, *i.e*, $1 - v \notin P$.

Let $v \in V$ such that there is $\tilde{T} \subseteq A$ such that $0 < v_{\tilde{T}} < 1$. Define the games $\tilde{v}, \bar{v}$ by

$$\begin{aligned}&\tilde{v}_T = v_T - \epsilon \text{ if } 0 < v_T < 1\\ &\tilde{v}_T = 1 \text{ if } v_T = 1\\ &\tilde{v}_T = 0 \text{ if } v_T = 0\end{aligned}$$

and

$$\begin{aligned}&\bar{v}_T = v_T + \epsilon \text{ if } 0 < v_T < 1\\ &\bar{v}_T = 1 \text{ if } v_T = 1\\ &\bar{v}_T = 0 \text{ if } v_T = 0\end{aligned}$$

where $\epsilon > 0$ such that $min_{T:v_T>0}\{v_T - \epsilon\} > 0$ and $max_{T:v_T<1}\{v_T + \epsilon\} < 1$. Both $\tilde{v}, \bar{v}$ are in V and $v = \frac{1}{2}\tilde{v} + \frac{1}{2}\bar{v}$, that is, v is not extreme. □

These two lemma imply the following intermediate important conclusion[25]:

Lemma 3. *Let $v \in V$, there exists $I^v \subseteq 2^A$ and positive real numbers $(\alpha^v_T)_{T\in I^v}$ such that $\sum_{T\in I^v} \alpha^v_T = 1$ and*

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

In particular, the Shapley measure associated to $a \in A$ is given by

$$\varphi_a(v) = \sum_{T\in I^v} \alpha^v_T \varphi_a(\hat{v}^T) = \sum_{A\neq T\in I^v : a\notin T} \alpha^v_T \frac{1}{|A\setminus T|}$$

on $V \setminus \hat{v}^A$ and $\varphi_a(\hat{v}^A) = 1$.[26]

[25] In $\mathbb{R}^{2^J}$ any non-empty convex compact subset is the convex hull of its extreme points.

[26] If these equalities are not clear to the reader, he or she may refer to Lemma 4 below, where they are proved.

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

If $K^v = 0$, then $v = \mathbf{0}_{\mathbf{2^J}}$ and (S) and (IA) imply that $\psi_a(v) = 0 = \varphi_a(v)$ for all $a \in A$. If $K^v = 1$, then there is $T \subseteq A$ such that $v = \hat{v}^T$. Suppose $T = A$, then, by (S) and (IA), $\psi_a(v) = 1 = \varphi_a(v)$. Suppose now $\emptyset \neq T \subset A$. For all $a \in T$, $v_{S\cup a} - v_S = 0$ for all $S \subseteq A \setminus a$, which yields $\psi_a(v) = \psi_a(\mathbf{0}_{\mathbf{2^J}}) = 0$ by (M). All $a, a' \notin T$ are symmetric for v and we conclude, by (S) and (IA), that $\psi_a(v) = \psi_{a'}(v) = \frac{1}{|A\setminus T|} = \varphi_a(v)$. We now proceed by induction on the value of K^v.

Suppose $\psi_a(v) = \sum_{A\neq T\in I^v : a\notin T} \alpha_T^v \frac{1}{|A\setminus T|}$ for all $v \in V \setminus \hat{v}^A$ such that $K^v \leq k \in \mathbb{N}$. Let $v = \sum_{T\in I^v} \alpha_T^v \hat{v}^T \in V \setminus \hat{v}^A$ with $K^v = k+1$. Consider $\mathcal{T}^v = \bigcup_{T\in I^v} T$ and $a \in \mathcal{T}^v$. Define the game

$$\nu = \sum_{T\in I^v : a\notin T} \alpha_T^v \hat{v}^T.$$

Because $V \setminus \hat{v}^A$ is convex —$\hat{v}^A$ is extreme in V— $\nu \in V \setminus \hat{v}^A$. In addition, $K^\nu \leq k$. Finally, observe that $\nu_{S\cup a} - \nu_S = v_{S\cup a} - v_S$ for any $S \subseteq A \setminus a$. Therefore,

$$\psi_a(v) = \psi_a(\nu) = \sum_{A\neq T\in I^v : a\notin T} \alpha_T^v \frac{1}{|A\setminus T|} = \varphi_a(v),$$

where the first equality follows from (M) and the second follows from the induction hypothesis.

Moreover, all axioms in $A \setminus \mathcal{T}^v$ are symmetric for v. As ψ_a coincides with φ_a for $a \in \mathcal{T}^v$, (S) and (IA) imply $\psi_a(v) = \varphi_a(v)$ for $a \notin \mathcal{T}^v$.

□

Independence of the axioms

These three axioms being independent when the set of admissible games is restricted to V is shown in exactly the same way as when the set of admissible games is G.

Alternative characterisation of the Shapley incompatibility measure

While our main characterisation was based on Young's axioms, we provide now a characterisation based on Shapley's axioms, building on the first steps of the preceding proof.

Null-Axiom (NA): For any $v \in V$, all $a \in A$, if $v_{S\cup a} = v_a$ for any $S \subseteq A\setminus a$, then $\psi_a(v) = 0$.

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

Conditional Additivity and Positive Homogeneity (CAPH): For any $v, v' \in V$, $\lambda \geq 0$ such that $v + \lambda v' \in V$, $\psi_a(v + \lambda v') = \psi_a(v) + \lambda\psi_a(v')$.

As in the standard cooperative game theory framework, and as the axiom (APH) we considered in the previous section, this principle is best interpreted as a simplicity requirement. We use the term *conditional* because, obviously, the affine combination of elements of V need not be in V.

Theorem 4. *The incompatibility measure $\psi : V \to \mathbb{R}^J$ satisfies null-axiom, incompatibility allocation, symmetry and conditional additivity and positive homogeneity if and only if it coincides with the Shapley incompatibility measure $\varphi : V \to \mathbb{R}^J$.*

Proof.

Lemma 4. *For any $S \subseteq A$, $\psi_a(\hat{v}^S) = \varphi_a(\hat{v}^S)$ for all $a \in A$.*

Proof. The case where $S = A$ follows from Lemma 3 and (CAPH). Let $S \subset A$ and consider $\hat{v}^S$. If $a \in S$, then a is a null-axiom in $\hat{v}^S$ so that, by (NA), $\psi_a(\hat{v}^S) = 0 = \varphi_a(\hat{v}^S)$. In addition, all $a \notin S$ are symmetric so that, by (S) and (IA),$\psi_a(\hat{v}^S) = \varphi_a(\hat{v}^S) = \frac{1}{|A \setminus S|}$. $\square$

Conclusion. In virtue of Lemma 3 and (CAPH), for all $v \in V$, there exists $I^v \in 2^A$ and positive real numbers $(\alpha_T^v)_{T \in I^v}$ such that $\sum_{T \in I^v} \alpha_T^v = 1$ and

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

Then, by Lemma 4,

$$\psi_a(v) = \sum_{T \in I^v} \alpha_T^v \varphi_a(\hat{v}^T) = \varphi_a(v) \text{ for all } a \in A.$$

$\square$

These four axioms being independent when the set of admissible games is restricted to V is shown in exactly the same way as when the set of admissible games is G.