Criteria to compare mechanisms that partially satisfy a property

Decerf, Benoit; Woitrin, Francois

Published in:
Social Choice and Welfare

DOI:
10.1007/s00355-021-01376-1

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Social Choice and Welfare (2022) 58:835–862
https://doi.org/10.1007/s00355-021-01376-1

ORIGINAL PAPER

Criteria to compare mechanisms that partially satisfy a property: an axiomatic study

Benoit Decerf[1,2] · **Francois Woitrin**[1]

Received: 2 November 2020 / Accepted: 5 November 2021 / Published online: 20 November 2021

Abstract
We study criteria that compare mechanisms according to a property (e.g., Pareto efficiency or stability) in the presence of multiple equilibria. The multiplicity of equilibria complicates such comparisons when some equilibria satisfy the property while others do not. We axiomatically characterize three criteria. The first criterion is intuitive and based on highly compelling axioms, but is also very incomplete and not very workable. The other two criteria extend the comparisons made by the first and are more workable. Our results reveal the additional robustness axiom characterizing each of these two criteria.

1 Introduction

From the assignment of seats at public schools to the allocation of goods against payment in auctions, economics repeatedly faces the problem of choosing among outcomes based on the preferences of a set of agents over these outcomes. To guide such collective choices, outcomes are often sorted according to desirable properties, formalized as *social choice correspondences*. If the agents' preferences are known, the set of outcomes can, for example, be sorted into subsets of Pareto efficient and Pareto inefficient outcomes, or in some applications, into subsets of "fair" and "unfair" outcomes.

Of course, preferences are often private information which makes it impossible for the social planner to directly compute whether outcomes satisfy a desirable property. Instead, the planner must setup a mechanism through which agents interacting strategically determine the selected alternative. Guiding the planner's design

✉ Francois Woitrin
francois.woitrin@unamur.be

Benoit Decerf
benoit.decerf@unamur.be

[1] University of Namur, Namur, Belgium

[2] World Bank, Washington, USA

requires determining which mechanism better provides agents with the incentives to select strategies that, given their preferences, lead to the selection of outcomes satisfying the desirable property.

The comparison of competing mechanisms then follows a three-step procedure. The first step consists in predicting the strategies agents might use in each mechanism as a function of their preferences. Formally, these predictions are captured by *solution concepts* such as undominated strategy, Nash equilibrium, or dominant strategy. Second, one must compute the outcomes selected by the mechanisms when agents play the strategies predicted by the solution concepts selected in the first step. Finally, one must evaluate the resulting outcomes according to the property of interest. Example applications of this three-step procedure can be found in Ergin and Sönmez (2006) and Abdulkadiroğlu et al. (2011).

In practice, given a preference profile, it is common for solutions concepts to make multiple predictions about the strategies agents might use in a mechanism. When this is the case, some of the predicted outcomes might satisfy the property of interest, whereas others might not. When this is the case, it is often unclear how mechanisms should be compared. For example, on a given preference profile, how does a mechanism with one desirable and one undesirable outcomes compare with a mechanism with four desirable and two undesirable outcomes?

In this paper, we propose, characterize and compare three criteria to perform such comparisons. First, the "Proportion" criterion compares, on a profile-by-profile basis, the *fraction* of desirable outcomes reached by each mechanism (the higher the fraction, the better the mechanism performs in terms of the property at stake). We show that this natural criterion is characterized by three compelling axioms. Unfortunately, the Proportion criterion only provides a very partial ranking of mechanisms and often concludes that mechanisms cannot be compared. Moreover, this criterion is not very workable because it requires counting the number of equilibria and identifying the fraction of desirable equilibria. Doing so becomes increasingly difficult as the number of equilibria grows.

Our two other criteria improve on both limitations and therefore constitute our main contribution. Both criteria satisfy the same three axioms as the Proportion criterion, and therefore agree with it on all pairs that the Proportion criterion is able to rank. To provide more complete orderings, each of these two additional criteria also satisfy an additional robustness axiom. Loosely put, these two mirror robustness axioms require that a comparison between two mechanisms would not be altered if both mechanisms had one additional desirable (undesirable) outcome.

Importantly, the two "extended" criteria compare mechanisms by focusing on preference profiles for which outcomes are either all desirable, or all undesirable. By doing so, they yield more affirmative comparisons because they are not necessarily bogged down by the existence of a few preference profiles for which the proportions of desirable outcomes are reversed. Moreover, these criteria do not require counting the number of equilibria nor computing the fraction of desirable outcomes.

Of course, the strength of an affirmative comparison between two mechanisms depends on the criterion used. One can be more confident that a mechanism will perform better than another when they can be ranked by the Proportion criterion than when this can only be done using our two other criteria. Yet, when the Proportion

criterion is silent, comparisons in terms of our dual criteria provide interesting indications about the respective performance that should be expected from two alternative mechanisms. In other words, the improvement on the limitations associated with the Proportion criterion comes at some cost. Our robustness axiom can lead to comparisons between mechanisms that are somewhat more debatable. This axiom can be viewed as capturing the cost of increasing the completeness of the partial order.

We illustrate the different discriminative powers and workabilities of these criteria for the comparison of the stability of two school choice mechanisms on a narrow domain.

The paper is organized as follows. We integrate our work in the literature in Sect. 2. We present the framework in Sect. 3. We derive axiomatically our criteria and discuss their shortcomings in Sect. 4. We then illustrate how those criteria can be use in the school choice problem in Sect. 5 and conclude in Sect. 6.

2 Related literature

Our three criteria use a "profile by profile" approach to compare mechanisms, which is common in the literature on voting procedures (Dasgupta and Maskin 2008; Gerber and Barberà 2016; Arribillaga and Massó 2015).[1] This approach is also common in the matching literature. Whereas our paper focuses on comparing the properties of outcomes, the matching literature has proposed a number of criteria to compare the manipulability of mechanisms. Pathak and Sönmez (2013) for example rank mechanisms by comparing the set of preference profiles for which the mechanisms admit a truthful Nash equilibirum. If a mechanism admits a truthful Nash equilibrium in every profile for which another mechanism also does, then Pathak and Sönmez (2013) conclude that the latter is less manipulable than the former. Similarly, Andersson et al. (2014) study manipulability by comparing the *number* of preference profiles at which each mechanism is manipulable. This type of manipulability comparisons avoids the issue induced by multiple solutions since it relies on binary evaluations: For any given preference profile, either a mechanism is manipulable or it is not.

In contrast, the multiplicity issue is key when evaluating the efficiency or fairness of outcomes. For example, Chen and Kesten (2017) compare school choice mechanisms with respect to the stability of their Nash equilibrium outcomes. The criterion implied by their analysis relies on the comparison of the number of stable equilibria in each type profile. Ergin and Sönmez (2006) show that the multiple equilibria of the Boston mechanism are all Pareto dominated by that of the Deferred Acceptance mechanism.

[1] In Gerber and Barberà (2016), the solution concept is "iterated elimination of weakly dominated strategies" and the correspondence is the possibility of agenda manipulation. In Dasgupta and Maskin (2008), the solution concept is "truthful revelation" and the correspondence is a collection of five voting properties.

When mechanisms do not perfectly satisfy a property of interest, another approach consists in comparing them using a criterion formalizing "by how much" each solution violates the property. In the case of stability, which requires the absence of *blocking pairs*, Combe et al. (2017), Abdulkadiroglu et al. (2019), Dogan and Ehlers (2020b) and Bonkoungou and Nesterov (2020) compare mechanisms by measuring, in different ways, the number of blocking pairs or the number of players participating to a blocking pair in each profile. Dogan and Ehlers (2020a) axiomatically characterize criteria for stability comparisons based on axioms specific to this property. Current research along this approach has so far abstracted from the multiplicity issue that we aim at tackling here.

Going even further away from the binary nature of social choice correspondences, which only acknowledge two desirable or undesirable categories, some authors propose to compare mechanisms using fine-grained normative tool, e.g., a social welfare function. Fleurbaey (2012) axiomatically characterize a criterion that compares how alternative mechanisms perform in the light of a fined-grained ranking of outcomes. Again, the setting considered by that author abstracts from the multiplicity issue.

Finally, our work shares important similarities with the literature on the measurement of predictive success (Selten 1991). We derive criteria that compare mechanisms as a function of their ability to yield outcomes that are selected by a correspondence. That literature derives rules that compare theories as a function of their ability to make predictions that are in line with observations. There are fundamental differences between these two objectives, which imply that our criterion are unrelated to these rules. Indeed these differences in objectives makes the relevant primitives different as well.[2]

3 Framework and notation

This section introduces the terminology and notation for our axiomatic results. We let $N = \{1, \dots, n\}$ denote the set of players, and $o \in O$ denote the set of outcome. Each player $i \in N$ is characterized by a type $y_i \in Y_i$, e.g., the player's preference over the outcomes in O. A type profile is denoted by $y \in Y := \times_{i \in N} Y_i$. Let $X : Y \to 2^O$ be a **social choice correspondence**, sometimes *correspondence*, for short.

A **mechanism** is a game form $M : S \to O$ that associates every strategy profile $s \in S := \times_{i \in N} S_i$ with an outcome in O, where S_i is the *finite* strategy space of $i \in N$. The set of mechanisms is $\mathcal{M}$ ($\mathcal{M}$ includes both direct and indirect mechanisms).

[2] The relevant primitives for our criteria are the numbers of equilibria that yield an outcome that is (resp. not) selected by the correspondence. In contrast, the relevant primitives for these rules include the "hit rate", i.e. the fraction of observations predicted by the theory, and the "area", i.e. the fraction of potential outcomes predicted by the theory. Neither the "hit rate" nor the "area" are relevant primitives for our criteria. We provide here the intuition why the "area" is not a relevant primitive for our criteria. The following two mechanisms have different area but should be considered equivalent by our criteria. The first mechanism has a unique equilibrium that yields an outcome that is selected by the correspondence. The second mechanism has multiple equilibria, all of which yield outcomes that are selected by the correspondence.

Let $C : Y \times \mathcal{M} \to 2^S$ denote a **solution concept**. The set $C(y, M)$ corresponds to the set of strategy profiles that C predicts could be played in mechanism M when the type profile is y. As is common, we henceforth refer to $C(y, M)$ as the set of **equilibria** of M under C when the type profile is y (whether or not C is an "equilibrium" solution concept). Since we assume that strategy spaces are finite, the number of equilibria is always finite. We focus on solution concepts that admit at least one equilibrium for each type profile. The set of such solution concepts is $\mathcal{C}$.

For a given correspondence X, let $\succeq$ be a **partial order** on $\mathcal{M} \times \mathcal{C}$. A partial order is a binary relation that is reflexive, asymmetric, and transitive.[3] The relation $(M, C) \succeq (M', C')$ indicates that mechanism M satisfies the property corresponding to X at least as well as mechanism M' when the former is played according to solution concept C and the latter according to solution concept C'. The symmetric and anti-symmetric relations, i.e., $(M, C) \succ (M', C')$ and $(M, C) \sim (M', C')$, are defined accordingly. Because $\succeq$ is partial, there may exists pairs $[(M, C), (M', C')]$ for which the relation is undefined.

Observe that we require the partial order to compare pairs (M, C), (M', C') that are potentially based on different solution concepts. This recognizes the fact that the behavior and coordination possibilities of players may depend on the mechanism. This is especially true when one of the mechanisms under consideration admits dominant strategies whereas the other does not (in which case, it is reasonable to use dominant strategies as a solution concept for the mechanisms where the latter is non-empty, and use the next best solution concept for the other mechanism, see, e.g., Ergin and Sönmez 2006; Abdulkadiroğlu et al. 2011).

Our objective is to identify partial orders satisfying compelling properties. Throughout, we restrict our attention to partial orders that satisfy an independence property we call *Outcome Neutrality*. This property forces partial orders to compare mechanisms based only on the *number* of equilibria whose outcome are selected (or not) by the social choice correspondence.[4] This captures the idea that the only aspect of equilibrium outcomes that matters to $\succeq$ is whether or not they are selected by the correspondence X. For any set A, we let $\#A$ denote the cardinality of set A.

Axiom 1 (*Outcome Neutrality*) For all $C, C' \in \mathcal{C}$ and all $M, M' \in \mathcal{M}$, if for all $y \in Y$ we have

(i) $\#\{s \in C'(y, M') \mid M'(s) \in X(y)\} = \#\{s \in C(y, M) \mid M(s) \in X(y)\}$, *and*
(ii) $\#\{s \in C'(y, M') \mid M'(s) \notin X(y)\} = \#\{s \in C(y, M) \mid M(s) \notin X(y)\}$,

then $(M, C) \sim (M', C')$.

All partial orders satisfying *Outcome Neutrality* can be reformulated as partial orders over particular "*counting*" functions. Any pair (M, C) defines an associated **counting function** F that associates any y with a function $F(y)$ such that

[3] In particular, the weak relation is transitive.

[4] This property assumes that all equilibria count the same. This is a natural assumption if one believes that all equilibria are equally likely to occur.

$F_0(y) := \#\{s \in C(y, M) \mid M(s) \notin X(y)\}$ and $F_1(y) := \#\{s \in C(y, M) \mid M(s) \in X(y)\}$. When no confusion on the types profile is possible, we simply write the components of the function F_0 and F_1.

Outcome Neutrality implies that any two (M, C) and (M', C') whose associated functions F and F' are the same perform equally well in terms of correspondence X (formally, $(M, C) \sim (M', C')$ whenever $F = F'$). Therefore, the partial order $\succeq$ on domain $\mathcal{M} \times \mathcal{C}$ is equivalent to a partial order on domain $\mathcal{F} = \{F : Y \to Z\}$, where $Z = \{(z_0, z_1) \in \mathbb{N}_0^2 | z_0 + z_1 \geq 1\}$. Observe that set Z is unbounded, a feature that is necessary for some of our results.[5]

Slightly abusing the notation, we also denote the latter partial order by $\succeq$. For the sake of improved readability, all remaining properties on the partial order are expressed on domain $\mathcal{F}$.

4 Criteria

We start by presenting three basic axioms for partial orders. When no confusion is possible, we ignore the role of solution concepts and simply say that we compare two mechanisms. Also, we write that an equilibrium is "in X" ("not in X") if its outcome is selected (not selected) by correspondence X. Finally, we say that two mechanisms M and M' are *equivalent* on a type profile y if $F_0(y) = F'_0(y)$ and $F_1(y) = F'_1(y)$.

Our first axiom, *Domination*, requires that if two mechanisms are equivalent on all but one type profile for which all the equilibria of one mechanism are in X whereas all the equilibria of the other mechanism are not in X, then the former performs better than the latter in terms of X.

Axiom 2 (*Domination*) For all $F, F' \in \mathcal{F}$, if (i) $F_1(y^*) = 0$ and $F'_0(y^*) = 0$ for some y^*, and (ii) $F'(y) = F(y)$ for all $y \neq y^*$, then $F' \succ F$.

Second, *Monotonicity* captures the idea that a larger number of equilibria not in X does not improve performance, while a larger number of equilibria in X does not worsen it. If two mechanisms are equivalent on all but one type profile for which they are not exactly equivalent because one mechanism has either one more equilibrium in X or one less equilibrium not in X than the other mechanism, then the axiom concludes that the former performs weakly better in terms of X.

Axiom 3 (*Monotonicity*) For all $F, F' \in \mathcal{F}$, if (i) for some y^* we have either $F'_0(y^*) = F_0(y^*)$ and $F'_1(y^*) = F_1(y^*) + 1$, or $F_0(y^*) = F'_0(y^*) + 1$ and $F'_1(y^*) = F_1(y^*)$, and (ii) $F'(y) = F(y)$ for all $y \neq y^*$, then $F' \succeq F$.

[5] In particular, Parts 2 of Theorems 1, 2 and 3 require the construction of intermediate mechanisms that may have, for some type profiles, more numerous equilibria than the number of equilibria of the mechanisms being compared. However, Parts 1 of Theorems 1, 2 and 3 do not require Z to be unbounded.

These two axioms are based on demanding preconditions and therefore, on their own, only impose relatively weak restrictions on $\succeq$. Hence, many implausible partial orders are not ruled out by these two alone. To illustrate the need for a third restriction, consider the following example and the following criterion.

Definition 1 (*Absolute Number criterion (AN)*) For any two $F, F' \in \mathcal{F}$, we have $F' \succeq_{AN} F$ whenever

$$F'_1(y) \geq F_1(y) \qquad \text{for all } y \in Y.$$

Moreover, $F' \succ_{AN} F$ if, in addition,

$$F'_1(y^*) > F_1(y^*) \qquad \textit{for some} y^* \in Y.$$

The Absolute Number criterion compares mechanisms based on their respective numbers of equilibria in X. This criterion satisfies our first two basic properties and its logic is implicitly used by Chen and Kesten (2017) (Theorem 2) when comparing the stability of school choice mechanisms.

To see why $\succeq_{AN}$ may be problematic, assume that there is a unique type profile y, which for two mechanisms $\tilde{F}$ and $\tilde{F}'$ is such that $\tilde{F}(y) = (1, 1)$ and $\tilde{F}'(y) = (4, 2)$. Clearly, the AN criterion concludes that $\tilde{F}'$ performs strictly better than $\tilde{F}$ because $\tilde{F}_1(y) = 1 < 2 = \tilde{F}'_1(y)$. This strict comparison is debatable because it ignores the fact that both mechanisms admit equilibria not in X and $\tilde{F}'$ admits more equilibria not in X than $\tilde{F}$ ($\tilde{F}_0(y) = 1 < 4 = \tilde{F}'_0(y)$). Even if $\tilde{F}'$ has twice as many equilibria in X as $\tilde{F}$, it not clear one should conclude that $\tilde{F}'$ performs *strictly* better than $\tilde{F}$ because $\tilde{F}'$ has four times as many equilibria not in X as $\tilde{F}$.

The issue with the AN criterion is that it violates a third basic property. *Replication Invariance* requires that two mechanisms that have the same proportion of their equilibria in X be viewed as performing equally well in terms of X. More precisely, if two mechanisms are equivalent on all but one type profile where one mechanism has k times as many equilibria in X and k times as many equilibria not in X as the other mechanism, then *Replication Invariance* concludes that the two mechanisms perform equally well in terms of X. When this is the case, we say that the former is a *k-replication* of the latter.

Axiom 4 (*Replication Invariance*) For all $F, F' \in \mathcal{F}$ and $k \in \mathbb{N}$, if (i) $F'_0(y^*) = kF_0(y^*)$ and $F'_1(y^*) = kF_1(y^*)$ for some y^*, and (ii) $F'(y) = F(y)$ for all $y \neq y^*$, then $F' \sim F$.

It is easy to see how the axioms introduced thus far reach a different comparison of $\tilde{F} = (1, 1)$ and $\tilde{F}' = (4, 2)$ than $\succeq_{AN}$. Consider a third mechanism $\tilde{F}''$ such that $\tilde{F}''(y) = (2, 1)$. By *Monotonicity*, $\tilde{F}$ performs weakly better than $\tilde{F}''$. By *Replication Invariance*, because $\tilde{F}'$ has twice as many equilibria in X and twice as many equilibria not in X as $\tilde{F}''$, they perform equally well. Together, we must conclude that $\tilde{F}$ performs weakly better than $\tilde{F}'$, in contradiction with the comparison obtained with $\succeq_{AN}$. The debatable comparison obtained with $\succeq_{AN}$ follows from its violation of *Replication Invariance*.

4.1 The Proportion criterion

As we show in Theorem 1, these three axioms jointly characterize the *Proportion* criterion. It compares mechanisms based on the proportion of their equilibria in X.[6] This criterion does not come as a surprise given its reliance on *Replication Invariance*.

Definition 2 [*Proportion criterion* (*PROP*)] For any two $F, F' \in \mathcal{F}$, we have $F' \succeq_{PROP} F$ whenever

$$\frac{F'_1(y)}{F'_0(y) + F'_1(y)} \geq \frac{F_1(y)}{F_0(y) + F_1(y)} \qquad \text{for all } y \in Y.$$

Moreover, $F' \succ_{PROP} F$ if, in addition,

$$\frac{F'_1(y^*)}{F'_0(y^*) + F'_1(y^*)} = 1 \text{ and } \frac{F_1(y^*)}{F_0(y^*) + F_1(y^*)} = 0 \qquad \text{for some } y^* \in Y.$$

Observe that, in line with *Domination*, the Proportion criterion yields strict comparisons only if there is a type profile where one mechanism has all its equilibria in X while all the equilibria of the other mechanism are not in X.

Our first result shows that the Proportion criterion is the *coarsest relation* satisfying our axioms.

Definition 3 (*Coarsest relation*) A partial order $\succeq_{co}$ is the coarsest relation satisfying a set of axioms if

1. $\succeq_{co}$ satisfies the set of axioms.
2. For all $F, F' \in \mathcal{F}$ and all $\succeq$ satisfying the set of axioms,

$$F' \succeq_{co} F \Rightarrow F' \succeq F, \text{ and} \tag{1}$$

$$F' \succ_{co} F \Rightarrow F' \succ F. \tag{2}$$

A partial order that is the coarsest relation satisfying a set of axioms is not necessarily the only partial order that satisfies this set of axioms.

Yet, the coarsest relation is the only partial order that satisfies the set of axioms while remaining silent on all pairs (of functions) that are not ranked by the joint implications of the axioms.

[6] Under our assumptions, the proportion is always well-defined. Indeed, we assume that solution concepts admit at least one equilibrium for each type profile. As a result, the denominateur of the proportion is never zero.

Theorem 1 identifies the close connection between the Proportion criterion and our three basic axioms.[7]

Theorem 1 *The partial order $\succeq_{PROP}$ is the coarsest relation satisfying Domination, Monotonicity and Replication Invariance.*

Proof **Part 1 of Definition** 3: The proof that $\succeq_{PROP}$ satisfies these three axioms is straightforward, and is therefore omitted.

Implication (2) **in part 2 of Definition** 3: $F' \succ_{PROP} F \Rightarrow F' \succ F$

We slightly abuse notation and often write F and F instead of $F(y)$ and $F'(y)$ whenever there is no ambiguity on y. Let $Y^1 = \{y \in Y \mid F'_0 = 0 \text{ and } F_1 = 0\}$ be the set of type profiles for which all equilibria of F' are in X while all the equilibria of F are not in X. Since $F' \succ_{PROP} F$, we have that Y^1 is not empty and also that $\frac{F'_1}{F'_0+F'_1} \geq \frac{F_1}{F_0+F_1}$ for all $y \in Y$.

We show that any partial order $\succeq$ satisfying the list of axioms is such that $F' \succ F$ by constructing two sequences of functions $(L^p)_{p\in\{0,1\}}$ and $(K^p)_{p\in\{0,1\}}$ with $L^p, K^p \in \mathcal{F}$ such that

- $L^0 \succ K^0$,
- $L^1 \succeq L^0$ and $K^0 \succeq K^1$,
- $L^1 = F'$ and $K^1 = F$.

If these two sequences exist, then we have indeed that $F' \succ F$.

We construct each function in the sequence type profile by type profile. First, we construct L^0 and K^0. For all $y \in Y^1$, we take $L^0 = F'$ and $K^0 = F$. For all $y \in Y \backslash Y^1$, we take $L^0_1 = K^0_1 = (F'_0 + F'_1) * F_1$ and $L^0_0 = K^0_0 = (F_0 + F_1) * F'_0$. By successive applications of *Domination* we have $L^0 \succ K^0$. By "successive applications" of *Domination*, we mean that it is straightforward to construct a sequence of functions $(F^p)_{p\in\{0,\ldots,P\}}$ with $F^0 = K^0$, $F^P = L^0$ and such that $F^{p+1} \succ F^p$ by the virtue of *Domination* for all $p \in \{0, \ldots, P-1\}$.

Then, we construct L^1 and K^1 from L^0 and K^0 by changing their images on $Y \backslash Y^1$. For all $y \in Y^1$, we take $L^1 = L^0$ and $K^1 = K^0$. For their construction on $Y \backslash Y^1$, we define two sequences $(\hat{L}^q)_{q\in\{0,1\}}$ and $(\hat{K}^q)_{q\in\{0,1\}}$ with $\hat{L}^q, \hat{K}^q \in \mathcal{F}$ such that

- $K^0 \succeq \hat{K}^0$,
- $\hat{L}^0 \succeq L^0$,
- $\hat{K}^1 \sim \hat{K}^0$,
- $\hat{L}^1 \sim \hat{L}^0$,

[7] These three axioms are independent. Showing independence of *Monotonicity* is the most difficult part. We propose the criterion *I*2, which satisfies all these axioms except *Monotonicity*. Criterion *I*2 is based on the following function $f : [0,1] \to [0,1]$ defined as $f(x) = 1 - x$ for $x \in \{0,1\}$ and $f(x) = x$ for all $x \in (0,1)$. That is, function f is strictly increasing for all $x \in (0,1)$, but returns the smallest value for $x = 1$ and the greatest for $x = 0$. For any two $F, F' \in \mathcal{F}$, we have $F' \succeq_{I2} F$ whenever $f\left(\frac{F'_1(y)}{F'_0(y)+F'_1(y)}\right) \geq f\left(\frac{F_1(y)}{F_0(y)+F_1(y)}\right)$ for all $y \in Y$, and we have $F' \succ_{I2} F$ if in addition the inequality is strict for some $y^* \in Y$.

and we take $L^1 = \hat{L}^1$ and $K^1 = \hat{K}^1$, which implies $L^1 \succ K^1$. For any $y \in Y \backslash Y^1$, we take $\hat{L}^0_0 = L^0_0$ and $\hat{L}^0_1 = L^0_1 + (F'_1 * F_0 - F'_0 * F_1)$, where we have $F'_1 * F_0 - F'_0 * F_1 \geq 0$ because $\frac{F'_1}{F'_1+F'_0} \geq \frac{F_1}{F_1+F_0}$. We have $\hat{L}^0 \succeq L^0$ by (successive applications of) *Monotonicity*. For any $y \in Y \backslash Y^1$, we also take $\hat{K}^0_0 = K^0_0 + (F'_1 * F_0 - F'_0 * F_1)$ and $\hat{K}^0_1 = K^0_1$. We have $K^0 \succeq \hat{K}^0$ by (successive applications of) *Monotonicity*.

Then, we construct $\hat{L}^1$ from $\hat{L}^0$ and $\hat{K}^1$ from $\hat{K}^0$. For any $y \in Y \backslash Y^1$, let $\hat{L}^0$ be a $(F_0 + F_1)$-replication of $\hat{L}^1$ and $\hat{K}^0$ a $(F'_0 + F'_1)$-replication of $\hat{K}^1$ so that we have $\hat{L}^1 \sim \hat{K}^0$ and $\hat{K}^1 \sim \hat{L}^0$ by (successive applications of) *Replication Invariance*.

By construction, we have $L^1 = F'$ and $K^1 = F$ which completes the proof.

Implication (1) in part 2 of Definition 3: $F' \succeq_{PROP} F \Rightarrow F' \succeq F$

The proof can straightforwardly be adapted from the argument provided above, and is therefore omitted. □

Theorem 1 calls for three remarks.

First, observe that one can also find complete orders satisfying this set of axioms.[8]

Second, Theorem 1 would still hold if we restrict ourselves to solution concepts with only one outcome per type profile, such as for instance the "truthfulness" solution concept. For this special case, the criteria must rank functions whose domain of images is $Z' = \{(z_0, z_1) \in \mathbb{N}_0^2 | z_0 + z_1 = 1\}$. All issues associated with having multiple equilibria are ruled out. For this special case, only *Domination* has bite because the remaining three axioms are trivially satisfied. Observe that the Proportion criterion would still yield a partial ranking of mechanisms. This illustrates that the difficulty to characterize a complete order is also present even when the equilibrium is unique.

Third, using the strict versions of axioms *Monotonicity*, i.e., if one mechanism has one more equilibrium in (resp. not in) X than the other mechanism, it performs strictly better (resp. worse) in terms of X, would lead to an impossibility because this stronger axiom is directly incompatible with *Replication Invariance*.

Although the Proportion criterion is very natural, it is affected by two important limitations. Since the Proportion criterion relies on relatively weak axioms, it provides a very partial ranking and is thus often silent. Moreover, the Proportion criterion is not very workable. Indeed, this criterion requires computing the exact number of equilibria in each type profile, which can get quite challenging as even very simplified type profiles can admit multiple strategy profiles.

[8] Consider the following complete order. *For any two* $F, F' \in \mathcal{F}$, *we have* $F' \succeq_{COMP} F$ *whenever*

$$\sum_{y \in Y} \frac{F'_1(y)}{F'_0(y) + F'_1(y)} \geq \sum_{y \in Y} \frac{F_1(y)}{F_0(y) + F_1(y)}$$

Observe that these three axioms do not jointly imply this order. Additional properties would be required, typically imposing some form(s) of anonymity.

4.2 Dual extension of the Proportion criterion

To obtain more complete partial orders, we maintain the axioms imposed thus far while imposing additional restrictions that increase the number of pairs a partial order can compare. In this sense, our new partial orders extend the comparisons from $\succeq_{PROP}$ (they compare in the same way all pairs for which $\succeq_{PROP}$ makes an affirmative comparison, and reach affirmative comparisons for some pairs for which $\succeq_{PROP}$ is silent).

First, we consider an additional robustness axiom that we call *Consistency to Additional* $\in X$. Loosely put, *Consistency to Additional* $\in X$ requires that a comparison would not be altered if both mechanisms had one additional equilibrium in X. More precisely, assume that one mechanism M performs better than another M' (in terms of X). Consider slight variants of these two mechanisms such that, on a single type profile, both variants have one additional equilibrium in X. *Consistency to Additional* $\in X$ requires that the variant of M also performs better than the variant of M'.

Axiom 5 (*Consistency to Additional* $\in X$) For all $F, F', \hat{F}, \hat{F}' \in \mathcal{F}$, if (i) $\hat{F}_0(y^*) = F_0(y^*)$, $\hat{F}'_0(y^*) = F'_0(y^*)$, $\hat{F}_1(y^*) = F_1(y^*) + 1$ and $\hat{F}'_1(y^*) = F'_1(y^*) + 1$ for some y^*, and (ii) $\hat{F}(y) = F(y)$ and $\hat{F}'(y) = F'(y)$ for all $y \neq y^*$, then $F' \succeq F \Rightarrow \hat{F}' \succeq \hat{F}$ and $F' \succ F \Rightarrow \hat{F}' \succ \hat{F}$.

Even if one may consider that *Consistency to Additional* $\in X$ is somewhat less compelling than our three basic axioms, we believe that it constitutes a plausible way of extending their affirmative comparisons. Observe in particular that *Consistency to Additional* $\in X$ does not impose any affirmative comparison on its own. It is only in combination with other axioms that it extends their pre-existing affirmative comparisons to more pairs.

Theorem 2 presented below shows that *Consistency to Additional* $\in X$ is exactly the difference between $\succeq_{PROP}$ and our second criterion. This criterion compares mechanisms by focusing exclusively on those type profiles for which all equilibria are in X or those for which all equilibria are not in X. More precisely, the criterion considers that a mechanism performs at least as well as another if the latter has no equilibria in X whenever the former has no equilibria in X *and* if the former has all its equilibria in X whenever the latter has all its equilibria in X. The comparison becomes strict if for some type profile, the former has all its equilibria in X whereas the latter has not.

Definition 4 [*Profiles with Homogeneous Outcomes criterion* (*PHO*)] For any two $F, F' \in \mathcal{F}$, we have $F' \succeq_{PHO} F$ if for all $y \in Y$

$$F'_1(y) = 0 \Rightarrow F_1(y) = 0, \textit{and}$$
$$F_0(y) = 0 \Rightarrow F'_0(y) = 0.$$

Moreover, we have $F' \succ_{PHO} F$ if in addition

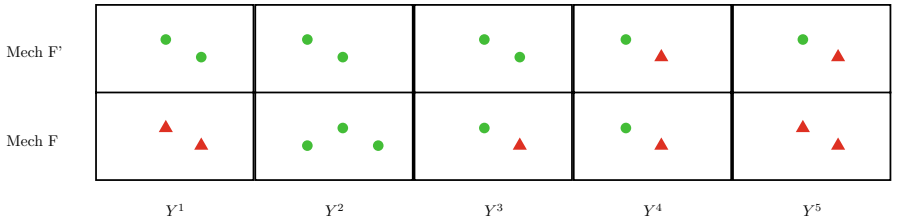

Fig. 1 Illustration of type profiles for each subset of the partition. *Each green dot represents an equilibrium in X and each red triangle represents an equilibrium not in X*

$$F_0'(y^*) = 0 \text{ and } F_0(y^*) > 0 \qquad \text{for some } y^* \in Y.$$

The partial order $\succeq_{PHO}$ is more workable than $\succeq_{PROP}$. Indeed, obtaining affirmative comparisons with $\succeq_{PHO}$ never requires computing the proportion of equilibria in X. Even better, it is not even necessary to compute the number of equilibria associated to each type profile. The reason is that the affirmative comparisons of $\succeq_{PHO}$ are only based on type profiles for which all equilibria are equivalent in terms of X.

Observe that $\succeq_{PHO}$ yields strict comparisons only if there is a type profile where one mechanism has all its equilibria in X while the other's are *not all* in X. In contrast, in the case of $\succeq_{PROP}$, strict comparisons require that there is a type profile where one mechanism has all its equilibria in X while the other's are *all not* in X. Clearly, the weaker condition for strict comparisons under $\succeq_{PHO}$ derives from *Consistency to Additional* $\in X$, which extends the strict comparisons obtained from *Domination*.

Theorem 2 identifies the close connection between $\succeq_{PHO}$ and our four axioms.

Theorem 2 *The partial order* $\succeq_{PHO}$ *is the coarsest relation satisfying Domination, Monotonicity, Replication Invariance and Consistency to Additional* $\in X$.

Proof **Part 1 of Definition** 3:

The PHO criterion clearly satisfies *Domination*, *Monotonicity* and *Replication Invariance*. We only prove that the PHO criterion satisfies *Consistency to Additional* $\in X$. We must show that, when its preconditions are met, we have $F' \succeq_{PHO} F \Rightarrow \hat{F}' \succeq_{PHO} \hat{F}$ and $F' \succ_{PHO} F \Rightarrow \hat{F}' \succ_{PHO} \hat{F}$. As the proof of the two implications are very similar, we only prove the latter.

Again, we slightly abuse notation and write F_0 and F_1 instead of $F_0(y)$ and $F_1(y)$. Given that $F' \succ_{PHO} F$, we can partition $Y = Y^1 \cup Y^2 \cup Y^3 \cup Y^4 \cup Y^5$, where

$$
\begin{aligned}
Y^1 &= \left\{y \in Y \mid F_0' = 0 \text{ and } F_1 = 0\right\},\\
Y^2 &= \left\{y \in Y \mid F_0' = 0 \text{ and } F_0 = 0\right\},\\
Y^3 &= \left\{y \in Y \mid F_0' = 0 \text{ and } F_0 > 0 \text{ and } F_1 > 0\right\},\\
Y^4 &= \left\{y \in Y \mid F_0' > 0 \text{ and } F_1' > 0 \text{ and } F_0 > 0 \text{ and } F_1 > 0\right\},\\
Y^5 &= \left\{y \in Y \mid F_0' > 0 \text{ and } F_1 = 0\right\},
\end{aligned}
$$

and where $Y^1 \cup Y^3$ is not empty. Such partition is illustrated in Fig. 1.

We show that when comparing $\hat{F}$ and $\hat{F}'$, we can also partition $Y = \hat{Y}^1 \cup \hat{Y}^2 \cup \hat{Y}^3 \cup \hat{Y}^4 \cup \hat{Y}^5$ with the same definitions as above, except that these definitions consider functions $\hat{F}$ and $\hat{F}'$ instead of F and F', i.e., $\hat{Y}^1 = \{y \in Y \mid \hat{F}_0' = 0 \text{ and } \hat{F}_1 = 0\}$, $\hat{Y}^2 = \{y \in Y \mid \hat{F}_0' = 0 \text{ and } \hat{F}_0 = 0\}$, and so on. Moreover $\hat{Y}^1 \cup \hat{Y}^3$ is not empty. If we can partition Y in this way, then we have $\hat{F}' \succ_{PHO} \hat{F}$.

There remains to show that the preconditions of *Consistency to Additional* $\in X$, which link F and F' to $\hat{F}$ and $\hat{F}'$, are such that any $y \in Y^1 \cup Y^2 \cup Y^3 \cup Y^4 \cup Y^5$ is such that $y \in \hat{Y}^1 \cup \hat{Y}^2 \cup \hat{Y}^3 \cup \hat{Y}^4 \cup \hat{Y}^5$ and any $y \in Y^1 \cup Y^3$ is such that $y \in \hat{Y}^1 \cup \hat{Y}^3$. For all $y \neq y^*$, we have $\hat{F}(y) = F(y)$ and $\hat{F}'(y) = F'(y)$, which directly implies that for all $p \in \{1, \ldots, 5\}$ we have $y \in \hat{Y}^p$ when $y \in Y^p$. For y^*, we have $\hat{F}_0(y^*) = F_0(y^*)$ and $\hat{F}_0'(y^*) = F_0'(y^*)$, as well as $\hat{F}_1(y^*) = F_1(y^*) + 1$ and $\hat{F}_1'(y^*) = F_1'(y^*) + 1$. These preconditions are such that $y^* \in Y^1 \Rightarrow y^* \in \hat{Y}^3$, $y^* \in Y^2 \Rightarrow y^* \in \hat{Y}^2$, $y^* \in Y^3 \Rightarrow y^* \in \hat{Y}^3$, $y^* \in Y^4 \Rightarrow y^* \in \hat{Y}^4$ and $y^* \in Y^5 \Rightarrow y^* \in \hat{Y}^4$. Finally, as $Y^1 \cup Y^3$ is non-empty, $y^* \in Y^1 \Rightarrow y^* \in \hat{Y}^3$ and $y^* \in Y^3 \Rightarrow y^* \in \hat{Y}^3$, $\hat{Y}^1 \cup \hat{Y}^3$ is not empty, the desired result.

Implication (2) in part 2 of Definition 3: $F' \succ_{PHO} F \Rightarrow F' \succ F$

Since $F' \succ_{PHO} F$, we can partition $Y = Y^1 \cup Y^2 \cup Y^3 \cup Y^4 \cup Y^5$ using the same definitions used for part 1, and moreover $Y^1 \cup Y^3$ is not empty.

We show that any partial order $\succeq$ satisfying the list of axioms is such that $F' \succ F$ by constructing two sequences of functions $(L^p)_{p\in\{0,\ldots,5\}}$ and $(K^p)_{p\in\{0,\ldots,5\}}$ with $L^p, K^p \in \mathcal{F}$ such that

- $L^0 \succ K^0$,
- $L^{p+1} \succeq L^p$ and $K^p \succeq K^{p+1}$ for all $p \in \{0, \ldots, 4\}$,
- $L^5 = F'$ and $K^5 = F$.

If such two sequences exist, then we have indeed that $F' \succ F$.

First, we define L^0 and K^0. We construct these two functions type profile by type profile. For all $y \in Y^1 \cup Y^3$ we take $L_0^0 = K_1^0 = 0$ and $L_1^0 = K_0^0 = 1$. For all $y \in Y^2$ we take $L_0^0 = K_0^0 = 0$ and $L_1^0 = K_1^0 = 1$. For all $y \in Y^4 \cup Y^5$ we take $L_0^0 = K_0^0 = 1$ and $L_1^0 = K_1^0 = 0$. By (successive applications of) *Domination*, we have $L^0 \succ K^0$.

We define the remaining elements of the two sequences in 5 successive steps, one for each subset in the partition of Y. Functions L^p and K^p are constructed from L^{p-1} and K^{p-1} in step p in such a way that for all $a \in \{1, \ldots, p\}$ and all $y \in Y^a$ we have $L^p(y) = F'(y)$ and $K^p(y) = F(y)$. When the construction of a function is left

unspecified on a type profile, it means that this function takes the same image as the function from which it is constructed.

- **Step 1**: Define L^1 and K^1 from L^0 and K^0 by changing their images on Y^1. For any $y \in Y^1$, take $L^1_0 = K^1_1 = 0$ and $L^1_1 = F'_1$ and $K^1_0 = F_0$. That is, $L^1(y)$ is a F'_1-replication of $L^0(y)$ and $K^1(y)$ is a F_0-replication of $K^0(y)$. By (successive applications of) *Replication Invariance*, we have $L^1 \sim L^0$ and $K^1 \sim K^0$.
- **Step 2**: Define L^2 and K^2 from L^1 and K^1 by changing their images on Y^2. For any $y \in Y^2$, take $L^2_0 = K^2_0 = 0$ and $L^2_1 = F'_1$ and $K^2_0 = F_1$. That is, $L^2(y)$ is a F'_1-replication of $L^1(y)$ and $K^2(y)$ is a F_1-replication of $K^1(y)$. By (successive applications of) *Replication Invariance*, we have $L^2 \sim L^1$ and $K^2 \sim K^1$.
- **Step 3**: Define L^3 and K^3 from L^2 and K^2 by changing their images on Y^3. We define two sequences $(\hat{L}^q)_{q\in\{0,1\}}$ and $(\hat{K}^q)_{q\in\{0,1\}}$ with
  - $\hat{K}^0 \sim K^2$,
  - $\hat{L}^0 \succ \hat{K}^1$,
  - $\hat{L}^1 \sim \hat{L}^0$,

  and we take $L^3 = \hat{L}^1$ and $K^3 = \hat{K}^1$, which implies $L^3 \succ K^3$. For any $y \in Y^3$, we take $\hat{K}^0_0 = F_0$ and $\hat{K}^0_1 = 0$. As $\hat{K}^0(y)$ is a F_0-replication of $K^2(y)$, by (successive applications of) *Replication Invariance*, we have $\hat{K}^0 \sim K^2$. We then construct $\hat{L}^0$ from L^2 and $\hat{K}^1$ from $\hat{K}^0$ by addition of the same number of equilibria in $X(y)$. For any $y \in Y^3$, we take $\hat{L}^0_0 = L^2_0$, $\hat{K}^1_0 = \hat{K}^0_0$, $\hat{L}^0_1 = L^2_1 + F_1$ and $\hat{K}^1_1 = \hat{K}^0_1 + F_1$. By transitivity we have from the previous steps that $L^2 \succ \hat{K}^0$. Therefore, we get $\hat{L}^0 \succ \hat{K}^1$ by (successive applications of) *Consistency to Additional* $\in X$. For any $y \in Y^3$, we take $\hat{L}^1_0 = 0$ and $\hat{L}^1_1 = F'_1$. As $\hat{L}^1(y)$ is a $\frac{F'_1}{1+F_1}$-replication of $\hat{L}^0(y)$, by (successive applications of) *Replication Invariance*, we have $\hat{L}^1 \sim \hat{L}^0$. If $\frac{F'_1}{1+F_1}$ is not an integer, then an intermediary function $\hat{L}^*$ must be defined such that $\hat{L}^*(y)$ is a F'_1-replication of $\hat{L}^0(y)$ and, also, such that $\hat{L}^*(y)$ is a $(1 + F_1)$-replication of $\hat{L}^1(y)$.
- **Step 4**: Define L^4 and K^4 from L^3 and K^3 by changing their images on Y^4. We define two sequences $(\hat{L}^q)_{q\in\{0,\ldots,3\}}$ and $(\hat{K}^q)_{q\in\{0,1\}}$ with
  - $\hat{L}^0 \sim L^3$ and $\hat{K}^0 \sim K^3$,
  - $\hat{L}^1 \succ \hat{K}^1$,
  - $\hat{L}^2 \succeq \hat{L}^1$ and $\hat{L}^3 \sim \hat{L}^2$,

  and we take $L^4 = \hat{L}^3$ and $K^4 = \hat{K}^1$, which implies $L^4 \succ K^4$. For any $y \in Y^4$, we take $\hat{L}^0_0 = F'_0F_1$, $\hat{K}^0_0 = F_0$ and $\hat{L}^0_1 = \hat{K}^0_1 = 0$. As $\hat{L}^0(y)$ is a F'_0F_1-replication of $L^3(y)$ and $\hat{K}^0(y)$ is a F_0-replication of $K^3(y)$, by (successive applications of) *Replication Invariance*, we have $\hat{L}^0 \sim L^3$ and $\hat{K}^0 \sim K^3$. We then construct $\hat{L}^1$ from $\hat{L}^0$ and $\hat{K}^1$ from $\hat{K}^0$ by addition of the same number of equilibria in $X(y)$. For any $y \in Y^4$, we take $\hat{L}^1_0 = \hat{L}^0_0$, $\hat{K}^1_0 = \hat{K}^0_0$, $\hat{L}^1_1 = \hat{L}^0_1 + F_1$ and $\hat{K}^1_1 = \hat{K}^0_1 + F_1$. By transitivity we have from the previous steps that $\hat{L}^0 \succ \hat{K}^0$. Therefore, we get $\hat{L}^1 \succ \hat{K}^1$ by (successive applications of) *Consistency to Additional* $\in X$. For any $y \in Y^4$, we take $\hat{L}^2_0 = \hat{L}^1_0$ and $\hat{L}^2_1 = \hat{L}^1_1 + F_1(F'_1 - 1)$. By (successive applications of) *Monotonicity*, we have $\hat{L}^2 \succeq \hat{L}^1$. Finally, for any $y \in Y^4$, we take $\hat{L}^3_0 = F'_0$ and $\hat{L}^3_1 = F'_1$.

As $\hat{L}^2(y)$ is a F_1-replication of $\hat{L}^3(y)$, by (successive applications of) *Replication Invariance*, we have $\hat{L}^3 \sim \hat{L}^2$.

- **Step 5**: Define L^5 and K^5 from L^4 and K^4 by changing their images on Y^5. We define a sequence $(\hat{L}^q)_{q\in\{0,1\}}$ and a function $\hat{K}^0$ with
  - $\hat{L}^0 \sim L^4$ and $\hat{K}^0 \sim K^4$,
  - $\hat{L}^1 \succeq \hat{L}^0$,

  and we take $L^5 = \hat{L}^1$ and $K^5 = \hat{K}^0$, which implies $L^5 \succ K^5$. For any $y \in Y^5$, we take $\hat{L}^0_0 = F'_0$, $\hat{K}^0_0 = F_0$ and $\hat{L}^0_1 = \hat{K}^0_1 = 0$. As $\hat{L}^0(y)$ is a F'_0-replication of $L^4(y)$ and $\hat{K}^0(y)$ is a F_0-replication of $K^4(y)$, by (successive applications of) *Replication Invariance*, we have $\hat{L}^0 \sim L^4$ and $\hat{K}^0 \sim K^4$. For any $y \in Y^5$, we take $\hat{L}^1_0 = \hat{L}^0_0$ and $\hat{L}^1_1 = \hat{L}^0_1 + F'_1$. By (successive applications of) *Monotonicity*, we have $\hat{L}^1 \succeq \hat{L}^0$.

By construction, we have $L^5 = F'$ and $K^5 = F$, which completes the proof.

Implication (1) **in part 2 of Definition** 3:

The proof can straightforwardly be adapted from the argument provided above, and is therefore omitted. □

Interestingly, imposing *Consistency to Additional* $\in X$ allows comparing two mechanisms by only focusing on the subset of type profiles for which all equilibria are equivalent in terms of X. In a sense, type profiles for which both mechanisms yield some outcomes in X and some outcomes not in X are "irrelevant" for the comparison.[9] This greatly increases the number of pairs that can be compared because, unlike $\succeq_{PROP}$, the partial order $\succeq_{PHO}$ is not necessarily bogged down by the existence of a few type profiles for which proportions of equilibria in X are reversed.

This reduction of the domain of "relevant" type profile is rather surprising. We emphasize that *Consistency to Additional* $\in X$ is not sufficient by itself to yield such reduction. In fact, the Absolute Number criterion satisfies *Consistency to Additional* $\in X$ together with *Domination* and *Monotonicity*, but still bases its comparisons on all type profiles in the domain. This reduction is the result of the combination of the list of axioms used.

Finally, we show that an axiom that is dual to *Consistency to Additional* $\in X$ leads to a criterion that is dual to $\succeq_{PHO}$. *Consistency to Additional* $\notin X$ preserves the logic of *Consistency to Additional* $\in X$, but the former focuses on equilibria not in X, whereas the latter focuses on equilibria in X. More precisely, assume that one mechanism M performs better than another M' (in terms of X). Consider slight variants of these two mechanisms such that, on a single type profile, both variants have one additional equilibrium *not* in X. *Consistency to Additional* $\notin X$ requires that the variant of M also performs better than the variant of M'.

Axiom 6 (*Consistency to Additional* $\notin X$) For all $F, F', \hat{F}, \hat{F}' \in \mathcal{F}$, if (i) $\hat{F}_0(y^*) = F_0(y^*) + 1$, $\hat{F}'_0(y^*) = F'_0(y^*) + 1$, $\hat{F}_1(y^*) = F_1(y^*)$ and $\hat{F}'_1(y^*) = F'_1(y^*)$ for

[9] These type profiles are irrelevant in the sense that the exact fraction of outcomes in X of each mechanism does not matter.

some y^*, and (ii) $\hat{F}(y) = F(y)$ and $\hat{F}'(y) = F'(y)$ for all $y \neq y^*$, then $F' \succeq F \Rightarrow \hat{F}' \succeq \hat{F}$ and $F' \succ F \Rightarrow \hat{F}' \succ \hat{F}$.

Unsurprisingly, the partial order $\succeq_{PHO^*}$ associated to *Consistency to Additional* $\notin X$ is very similar to $\succeq_{PHO}$. In fact, the weak comparisons of these two criteria are based on the same conditions. The difference comes from the condition for strict comparisons. The partial order $\succeq_{PHO^*}$ yields strict comparisons only if there is a type profile where one mechanism has *some* of its equilibria in X while the other has *none* of its equilibria in X.

Definition 5 [*Profiles with Homogeneous Outcomes criterion* PHO*] For any two $F, F' \in \mathcal{F}$, we have $F' \succeq_{PHO^*} F$ if for all $y \in Y$

$$F'_1(y) = 0 \Rightarrow F_1(y) = 0, \textit{and}$$
$$F_0(y) = 0 \Rightarrow F'_0(y) = 0.$$

Moreover, we have $F' \succ_{PHO^*} F$ if in addition

$$F'_1(y^*) > 0 \text{ and } F_1(y^*) = 0 \qquad \text{for some } y^* \in Y.$$

Theorem 3 identifies the close connection between $\succeq_{PHO^*}$ and the four axioms.

Theorem 3 *The partial order $\succeq_{PHO^*}$ is the coarsest relation satisfying Domination, Monotonicity, Replication Invariance and Consistency to Additional $\notin X$.*

Proof The proof can straightforwardly be adapted from the proof of Theorem 2, and is therefore omitted. □

5 Illustration with school choice mechanisms

For illustrative purposes, we compare two matching mechanisms for the allocation of school seats. In this context, a matching algorithm determines the allocation of seats based on the preferences reported by the students and the priorities that students receive at the different schools. The players are the students and their strategy set is the set of preferences they can report. A type profile consists in a preference profile together with a priority profile. The complete description of the school choice model considered is given in Appendix 7.1.

We focus on an extremely simplified domain of school choice problems, with only three students and three schools, each endowed with one seat. On this narrow domain, we compare two school choice mechanisms with respect to the "stable" social choice correspondence, which is central in the school choice literature. This fairness property essentially requires that no blocking pair exists in the assignment.[10]

[10] An assignment has a blocking pair if a student is assigned to a school that another student prefers to her assignment and the other student has higher priority at this school than the first student.

Table 1 Type profile y for which all undominated strategy profiles under DA^2 are stable but not all undominated strategy profiles under BOS^2

R_{i_1}	R_{i_2}	R_{i_3}
s_1^*	s_1	s_2
s_2	s_2^*	s_3^*
s_3	s_3	s_1

$\trianglerighteq_{s_1}$	$\trianglerighteq_{s_2}$	$\trianglerighteq_{s_3}$
i_1	i_2	i_3
$\vdots$	$\vdots$	$\vdots$

The two mechanisms we compare are constrained versions of the Deferred Acceptance (DA) and Boston (BOS) mechanisms, for which students are allowed to report preferences on *two* schools only (Haeringer and Klijn 2009). We denote these mechanisms as DA^2 and BOS^2.[11] For both mechanisms, we use *undominated strategy profile* as a solution concept. Although it is widely used, the Nash equilibrium solution concept might not be credible for such mechanisms. In school choice, Nash equilibrium may require a degree of coordination that goes beyond what can reasonably be expected from parents who play the corresponding game, often as a one-shot game. Experimental evidence also suggest that Nash equilibria are rarely reached in these mechanisms (Calsamiglia et al. 2010).

Under both DA^2 and BOS^2, many type profiles admit multiple undominated strategy profiles, some of which lead to stable assignments while others do not. However, there are reasons to believe that DA^2 should be deemed more stable than BOS^2. First, theoretical results have shown that *un*constrained *DA* is stable in dominant strategies, whereas *un*constrained *BOS* is stable only in Nash equilibrium, i.e., when assuming complete coordination among the players. Second, experimental evidence shows that constrained versions of *DA* are more stable than constrained versions of *BOS*. In a constrained environment, i.e., when players can report preferences on a limited number of schools, Calsamiglia et al. (2010) show that, even though stable assignments rarely occur, there are significantly more blocking pairs arising in constrained versions of *BOS* than in constrained versions of *DA*. (Recall that the "stable" correspondence essentially selects assignments that do not contain any blocking pairs.) Also, Klijn et al. (2013) show that, independently of players' risk aversion, *BOS* is less likely to produce stable assignments than *DA*.

Unfortunately, Proposition 1 shows that the Proportion criterion cannot compare the stability of these two mechanisms: There exist type profiles for which the proportion of stable equilibria is greater under DA^2 than under BOS^2, as well as other type profiles for which the converse is true.

Proposition 1 *Let the solution concept C be undominated strategy profiles. Let X denote the stable correspondence. Let* F^{DA^2} *and* F^{BOS^2} *be the functions respectively associated to* DA^2 *and* BOS^2 *by C and X. There exists a type profile* $y \in Y$ *such that*

[11] See Appendix 7.2 for the description of both mechanisms.

Table 2 Type profile y' for which the proportion of stable undominated strategy outcomes is larger under BOS^2 than under DA^2

R_{i_1}	R_{i_2}	R_{i_3}
s_2	s_1	s_1
$\boxed{s_1^*}$	$\boxed{s_2^*}$	s_2
i_1	i_2	$\boxed{s_3}$

$\trianglerighteq_{s_1}$	$\trianglerighteq_{s_2}$	$\trianglerighteq_{s_3}$
i_1	i_2	i_1
i_3	$\vdots$	$\vdots$
i_2		

$$\frac{F_1^{DA^2}(y)}{F_0^{DA^2}(y)+F_1^{DA^2}(y)} = 1 \text{ and } \frac{F_1^{BOS^2}(y)}{F_0^{BOS^2}(y)+F_1^{BOS^2}(y)} < 1,$$

and a type profile $y' \in Y$ *such that*

$$\frac{F_1^{DA^2}(y')}{F_0^{DA^2}(y')+F_1^{DA^2}(y')} < \frac{F_1^{BOS^2}(y')}{F_0^{BOS^2}(y')+F_1^{BOS^2}(y')}.$$

Proof Type profile y is presented in Table 1. For visual convenience, the schools at which a student has top-priority are starred.

First, we show for y that all undominated strategy profiles under DA^2 are stable. As i_1 has a top-priority at her most-preferred school, she has a dominant strategy and is assigned to that school under any undominated strategy profile (Lemma 3). As i_2 and i_3 have a top-priority at their second most-preferred school, they have a dominant strategy (Lemma 4) that ranks their two most-preferred schools according to their true preference (Lemma 5). Therefore, only one assignment can be reached under undominated strategy profiles. This assignment is such that each student is assigned to her most-preferred top-priority school. This assignment is stable.

Second, we show for y that one undominated strategy profiles under BOS^2 is not stable. Consider the reported profile Q shown herebelow.

$$\begin{array}{l} Q_{i_1} : s_1\ s_2 \\ Q_{i_2} : s_1\ s_2 \\ Q_{i_3} : s_2\ s_3 \end{array}$$

Profile Q is an undominated strategy profile under BOS^2 (Lemma 6). The assignment $BOS^2(Q)$ is such that i_1 is assigned to s_1, i_2 is unassigned and i_3 is assigned to s_2. This assignment is unstable as i_2 prefers s_2 over being unassigned and i_2 has a higher priority at s_2 than i_1.

Type profile y' is presented in Table 2. For visual convenience, the schools at which a student has top-priority are starred and the only stable assignment is boxed.

Under DA^2, one-third of undominated strategy outcomes are stable. It is a dominant strategy for both i_1 and i_2 to truthfully report their preference because they each have a top-priority at their second favorite school (Lemma 3). In turn, student i_3 has three undominated strategies (Lemma 5):

$$\begin{array}{l} Q_{i_3} : s_1\ s_2 \\ Q'_{i_3} : s_1\ s_3 \\ Q''_{i_3} : s_2\ s_3 \end{array}$$

If i_3 reports Q_{i_3}, then the assignment is unstable because i_3 is unassigned while the seat at s_3 is vacant. If i_3 reports Q''_{i_3}, then the assignment is again unstable because i_2 is assigned to s_1 even if i_2 has a lower priority at s_1 than i_3. If i_3 reports Q'_{i_3}, then the assignment is stable.

Under BOS^2, more than one-third of undominated strategy outcomes are stable. Students i_1 and i_2 have two undominated strategies whereas i_3 has six undominated strategies (Lemma 6):

$$\begin{array}{ll} Q_{i_1} : s_2\ s_1^* & Q_{i_2} : s_1\ s_2^* \\ Q'_{i_1} : s_1^*\ s_2 & Q'_{i_2} : s_2^*\ s_1 \end{array} \qquad \begin{array}{l} Q_{i_3} : s_1\ s_2 \\ Q'_{i_3} : s_1\ s_3 \\ Q''_{i_3} : s_2\ s_3 \\ Q'''_{i_3} : s_2\ s_1 \\ Q''''_{i_3} : s_3\ s_1 \\ Q'''''_{i_3} : s_3\ s_2 \end{array}$$

We show that a proportion 10/24 of BOS^2 assignments are stable, which is larger than the proportion 1/3 obtained under DA^2.

First, we consider the six undominated strategy profiles for which i_1 and i_2 report Q_{i_1} and Q_{i_2}. None of the six assignments are stable, because for all of them we have either that i_3 is unassigned or i_2 is assigned to s_1.

Second, we consider the six undominated strategy profiles for which i_1 and i_2 report Q'_{i_1} and Q'_{i_2}. Under these profiles, i_1 is assigned to s_1 and i_2 is assigned to s_2. The assignment is stable if i_3 reports s_3, which is the case in all her undominated strategies but Q_{i_3} and Q'''_{i_3}. Hence, four out of these six assignments are stable.

Third, we consider the six undominated strategy profiles for which i_1 and i_2 report Q_{i_1} and Q'_{i_2}. Under these profiles, i_2 is assigned to s_2. The assignment is stable if i_3 reports s_3 and does not report s_1 first. Hence, three out of these six assignments are stable.

Fourth, we consider the six undominated strategy profiles for which i_1 and i_2 report Q'_{i_1} and Q_{i_2}. Under these profiles, i_1 is assigned to s_1. The assignment is stable if i_3 reports s_3 and does not report s_2 first. Hence, three out of these six assignments are stable. □

The second important limitation of $\succeq_{PROP}$ is that this criterion is not very workable. Even in our extremely simplified domain with only three students and three schools, computing the exact number of equilibria in each type profile and identifying the proportion of these equilibria that are in X can be challenging. As we show in the proof of Proposition 1, the relatively simple type profile y' admits 24 different undominated strategy profiles under BOS^2. Investigating the stability of all 24

is quite cumbersome. What is more, the proof only shows that the two mechanisms cannot be compared by $\succeq_{PROP}$, which requires considering only two type profiles.

This example illustrates the need for partial orders that are less partial and more workable than $\succeq_{PROP}$.

The increased discriminatory power of $\succeq_{PHO}$ provides a sense for which we can affirmatively compare DA^2 and BOS^2 in terms of stability. This comparison is in line with our expectations.

Proposition 2 *Let the solution concept C be undominated strategy profiles. Let X denote the stable assignments correspondence. Letting F^{DA^2} and F^{BOS^2} be the functions respectively associated to DA^2 and BOS^2 by C and X, we have $F^{DA^2} \succ_{PHO} F^{BOS^2}$.*[12]

Proof **Part 1.** $F^{DA^2} \succeq_{PHO} F^{BOS^2}$.

First, we show that for all $y \in Y$ for which no undominated strategy profiles under DA^2 leads to a stable assignment, no undominated strategy profiles under BOS^2 leads to a stable assignment. To do so, we show that there exists no $y \in Y$ for which no undominated strategy profiles under DA^2 leads to a stable assignment. Consider the contradiction assumption that, for some type profile $y^* \in Y$, no undominated strategy profile under DA^2 leads to a stable assignment. Let μ^* denote the most-efficient stable assignment for type profile y^*.

Assume first that all students are assigned to a school under μ^*. Consider any undominated strategy profile $Q = (Q_{i_1}, Q_{i_2}, Q_{i_3})$ under DA^2 for which each student i reports $\mu^*(i)$, the school to which she is assigned under μ^*. Any student i has an undominated strategy with this property. Indeed, if $\mu^*(i)$ is not her third favorite acceptable school, then reporting her two favorite acceptable schools in the order of her truthful preference is clearly undominated under DA^2. If $\mu^*(i)$ is her third favorite acceptable school, then i has no dominant strategy under DA^2 and reporting any two acceptable schools in the same order as the order of preference is undominated (Lemma 5).

Since strategies in Q are undominated, they report the schools in the order of the students' truthful preference (Lemma 5). Hence, if $\mu^*(i)$ is not reported first in Q_i, then i prefers the school reported first in Q_i over $\mu^*(i)$. Then, because μ^* is a stable assignment, either the assignment $DA^2(Q)$ is μ^*, which violates the contradiction assumption, or $DA^2(Q)$ is a Pareto improvement over μ^*. In the latter case, $DA^2(Q)$ is unstable because μ^* is the most-efficient stable assignment. As $DA^2(Q)$ is an unstable Pareto improvement over μ^*, we have that one student, say i_3, is assigned under $DA^2(Q)$ to the same school as under μ^*, while i_1 and i_2 have exchanged the schools they are assigned to under μ^*. If all students are assigned to a different school as

[12] Note that with F^{DA} and F^{BOS} the functions respectively associated to unrestricted DA and BOS by C and X, we also have $F^{DA} \succ_{PHO} F^{BOS}$. Indeed, in DA all students have a single dominant strategy which consist in ranking all their acceptable schools without switches. There is therefore only one undominated strategy profile in DA, and this profile is always stable. It is then sufficient to show that some of the many undominated strategy profiles in BOS are not stable.

under μ^*, then μ^* cannot be the most-efficient stable assignment. This implies that $Q_{i_1} : \mu^*(i_2)\ \mu^*(i_1)$ and $Q_{i_2} : \mu^*(i_1)\ \mu^*(i_2)$. Assignment $DA^2(Q)$ is unstable because there is a school $s \in \{\mu^*(i_1), \mu^*(i_2)\}$ that i_3 prefers over $\mu^*(i_3)$ and i_3 has a higher priority at s than the student assigned to s under $DA^2(Q)$. As i_3 prefers s over $\mu^*(i_3)$ we have that $Q'_{i_3} : s\ \mu^*(i_3)$ is undominated under DA^2 (Lemma 5). As i_3 has a higher priority at s than the student assigned to s under $DA^2(Q)$, we must have that $DA^2(Q_{i_1}, Q_{i_2}, Q'_{i_3}) = \mu^*$, which violates the contradiction assumption.

Assume then that some student i is not assigned to a school under μ^*. Because there are three schools and three students, this implies that student i finds at most two schools acceptable. In turn, this implies that any student i' who is assigned to a school under μ^* is assigned either to her most-preferred school or to her second most-preferred school. The reason is that i is rejected from all of her acceptable schools. Hence, any school s that is acceptable for i is assigned under μ^* to another student i'. This is only possible if i' prefers s to the school that has a vacant seat under μ^*. Hence, any such student i' is assigned to a school she prefers to at least one other school.

Consider any strategy profile $Q = (Q_{i_1}, Q_{i_2}, Q_{i_3})$ under DA^2 for which each student reports either her only acceptable school, or her two most-preferred acceptable schools in the same order as the order of her true preference. All strategies in Q are undominated under DA^2 (Lemma 5). The contradiction assumption is violated because we have $DA^2(Q) = \mu^*$. Indeed, under μ^*, unassigned students find at most two schools acceptable and other students are assigned either to their most-preferred or their second most-preferred acceptable school. Therefore, on this type profile, the Deferred Acceptance mechanism stops before reaching the acceptable schools not reported in Q (if any). Hence, when the profile is Q, mechanism DA^2 follows the same steps as the Deferred Acceptance, and thus yields the most efficient stable assignment.

There remains to show that, for all $y \in Y$ for which all undominated strategy profiles under BOS^2 lead to a stable assignment, all undominated strategy profiles under DA^2 also lead to a stable assignment. The proof is based on Lemma 1, which shows that the set of assignments obtained by undominated strategy profiles under DA^2 are nested in the set of assignments obtained by undominated strategy profiles under BOS^2.

Lemma 1 *For any undominated strategy profile Q of DA^2, there exists an undominated strategy profile Q' of BOS^2 such that $DA^2(Q) = BOS^2(Q')$.*

Proof Take any profile Q that is undominated under DA^2. Let assignment $\mu = DA^2(Q)$.

We construct a strategy profile Q' that is undominated under BOS^2 and such that $BOS^2(Q') = \mu$. For any student i who is *un*assigned under μ we let $Q'_i = Q_i$. For any student i who is assigned to a school under μ,

- we let $Q'_i = Q_i$ if $\mu(i)$ is reported first in Q_i,

- else Q'_i reports $\mu(i)$ first and also reports her most-preferred acceptable school different from $\mu(i)$ (if any).

First, we show that Q'_i is undominated under BOS^2. If i finds only one school acceptable, then reporting this school only is a dominant strategy under both BOS^2 and DA^2 (Lemma 2) and by construction this case is such that $Q'_i = Q_i$. Assume then that i finds at least two schools acceptable.

If the most-preferred school of student i is a top-priority school for i, then it is a dominant strategy to report this school first under DA^2 (Lemma 3) and i must be assigned to this school under μ, i.e., this school is $\mu(i)$. By construction, Q'_i reports $\mu(i)$ first, and therefore Q'_i is a dominant strategy under BOS^2 (Lemma 3). Assume then that i finds at least two schools acceptable and her most-preferred school is not a top-priority school for i.

- Case 1: $Q'_i = Q_i$.
  Since Q_i is undominated under DA^2, we have by Lemma 5 that Q_i reports two schools, ranks these two schools according to i's true preference and i weakly prefers these two schools over her most-preferred top-priority school. As $Q'_i = Q_i$, we then have that Q'_i is undominated under BOS^2 (Lemma 6).
- Case 2: $Q'_i \neq Q_i$.
  By construction of Q'_i, this case is such that i is assigned under μ and $\mu(i)$ is reported second in Q_i. Then, since Q_i is undominated under DA^2, by Lemma 5, i weakly prefers the two schools reported in Q_i over her most-preferred top-priority school. If $\mu(i)$ is i's most-preferred top-priority school, then Q'_i is undominated under BOS^2 (Lemma 6), because by construction of Q'_i this school is reported first in Q'_i. If $\mu(i)$ is not i's most-preferred top-priority school, then Q'_i is undominated under BOS^2 (Lemma 6) because by construction of Q'_i this strategy reports two schools, one of them being preferred to i's most-preferred top-priority school. (The school reported first in Q_i is strictly preferred to $\mu(i)$.)

Second, we show that $BOS^2(Q') = \mu$. Consider the subset I' of students who are unassigned under μ. Since $DA^2(Q) = \mu$, this implies that no student $i \in I'$ can be blocking in matching μ at a school she reports in Q_i. (Indeed, if such student i was blocking at a school s, then the student j for whom $\mu(j) = s$ should have been rejected from s in the course of DA^2 under Q, a contradiction.) In other words, the seat at the schools that i reports in Q_i are assigned under μ to competitors of i at these schools. By construction, for any student $i \in I'$ we have $Q'_i = Q_i$. Since any student $j \notin I'$ reports $\mu(j)$ first in Q'_j, this implies that all seats at all schools reported by any student $i \in I$ are assigned to competitors of i in the first round of BOS^2 under Q'. As a result, all students in I' are also unassigned under $BOS^2(Q')$. Finally, since any student $j \notin I'$ reports $\mu(j)$ first in Q'_j, student j is also assigned to $\mu(j)$ under $BOS^2(Q')$. Together, we have $BOS^2(Q') = \mu$. □

Consider any $y \in Y$ for which all undominated strategy profiles under BOS^2 lead to a stable assignment. By Lemma 1, for any undominated strategy profiles under DA^2, there is an undominated strategy profiles under BOS^2 that leads to the

same assignment. As a result, any undominated strategy profiles under DA^2 leads to a stable assignment.

Part 2. For some $y^* \in Y$, all undominated strategy profiles under DA^2 lead to stable assignments, while some undominated strategy profiles under BOS^2 leads to unstable assignments.

As shown in the proof of Proposition 1, type profile y presented in Table 1 has the required properties. Together, Part 1 and Part 2 imply that $F^{DA^2} \succ_{PHO} F^{BOS^2}$. □

The proof of Proposition 2 illustrates another reason why the partial order $\succeq_{PHO}$ is more workable than $\succeq_{PROP}$. Affirmative comparisons of $\succeq_{PHO}$ are only based on type profiles for which all equilibria are equivalent in terms of X. Importantly, it is sometimes easier to compare mechanisms on these particular type profiles. For instance, the proof of Proposition 2 takes advantage of the focus on these type profiles. A key step in the proof of Proposition 2 is that the US-assignments under DA^2 are nested in the US-assignments under BOS^2. (For short terminology, we refer to an assignment sustained by an undominated strategy profile under mechanism M simply as an US-assignment under M.) This directly implies that, if all US-assignments are stable under BOS^2, then all US-assignments are stable under DA^2. The weak comparison $F^{DA^2} \succeq_{PHO} F^{BOS^2}$ then follows from the fact that there is no type profile in our domain for which all US-assignments under DA^2 are unstable.

Given the relationships between $\succeq_{PHO^*}$ and $\succeq_{PHO}$, we can deduce from Proposition 2 that, according to $\succeq_{PHO^*}$, DA^2 performs weakly better than BOS^2 in terms of stability. The reason is that Proposition 2 shows that, according to $\succeq_{PHO}$, DA^2 performs strictly better than BOS^2 in terms of stability, and the preconditions for weak comparisons are the same for both partial orders. Also, we can deduce from Proposition 2 that, according to $\succeq_{PHO^*}$, BOS^2 does not perform weakly better than DA^2 in terms of stability. The reason is that the precondition for a strict comparison according to $\succeq_{PHO}$ precludes a reversed weak comparison according to $\succeq_{PHO^*}$. These two implications are recorded in Corollary 1.

Corollary 1 *Let the solution concept C be undominated strategy profiles. Let X denote the stable assignments correspondence. Letting F^{DA^2} and F^{BOS^2} be the functions respectively associated to DA^2 and BOS^2 by C and X, we have $F^{DA^2} \succeq_{PHO^*} F^{BOS^2}$ and $F^{BOS^2} \not\succeq_{PHO^*} F^{DA^2}$.*

Proof **Part 1.** $F^{DA^2} \succeq_{PHO^*} F^{BOS^2}$.

By definition of $\succeq_{PHO}$ and $\succeq_{PHO^*}$, this is a direct implication of $F^{DA^2} \succeq_{PHO} F^{BOS^2}$ (Proposition 2).

Part 2. $F^{BOS^2} \not\succeq_{PHO^*} F^{DA^2}$.

By definition of $\succeq_{PHO}$ and $\succeq_{PHO^*}$, this is a direct implication of $F^{DA^2} \succ_{PHO} F^{BOS^2}$ (Proposition 2). More precisely, this follows from the fact that there exists a type profile y (given in Table 1) for which all undominated strategy profiles of DA^2 leads to a stable assignment whereas it is not the case of all undominated strategy profiles of BOS^2. □

6 Concluding remark

The strength of the comparison between two mechanisms depends on the partial order used. One can be more confident that a mechanism will perform better than another when they can be ranked when using $\succeq_{PROP}$ than when this can only be done when using $\succeq_{PHO}$ or $\succeq_{PHO^*}$. However, given its two limitations, affirmative comparisons obtained with $\succeq_{PROP}$ are bound to be scarce and hard to obtain. In their absence, affirmative comparisons obtained with $\succeq_{PHO}$ or $\succeq_{PHO^*}$ may provide interesting indications about the respective performance to expect from two alternative mechanisms.

Appendix on school choice application

The school choice model

The model and notation are inspired from Haeringer and Klijn (2009). There are three students i_1, i_2 and i_3 and three schools s_1, s_2 and s_3, each endowed with one seat. Each student can be assigned to at most one school. Students have preferences over the schools they could be assigned to as well as the possibility of remaining unassigned (i.e., being self-matched). Each school has a strict priority ordering over the students. In this setting, a **(school choice) problem** is a pair $\pi = (R, \trianglerighteq)$ where

1. $R := (R_{i_1}, R_{i_2}, R_{i_3})$ is the (strict) **preference profile** of students over the three schools, and
2. $\trianglerighteq := (\trianglerighteq_{s_1}, \trianglerighteq_{s_2}, \trianglerighteq_{s_3})$ is the (strict) **priority profile** of schools over the three students.

The preference R_i of student i is a linear order over $S \cup \{i\}$. If student i strictly prefers school s over school s', we write $s \, P_i \, s'$. As usual, $s \, R_i \, s'$ denotes a weak preference, allowing for $s = s'$. We say that a school s is **acceptable** for a student i if $s \, P_i \, i$ and **unacceptable** if $i \, P_i \, s$. To avoid trivialities, we assume that all students find at least one school acceptable.

The priority $\trianglerighteq_s$ of school s is a linear order over the three students. If student i has a higher priority than student j at school s, then $i \triangleright_s j$ and we say that i is a **competitor** of j at school s. School s is a **top-priority** school for student i if i has no competitor at school s.

We denote by Π the domain of problems satisfying these assumptions.

An **assignment** is a function $\mu : \{i_1, i_2, i_3\} \to \{s_1, s_2, s_3\} \cup \{i_1, i_2, i_3\}$ that matches every student with a school or with herself. We say that student i is **assigned** in the former case, and **unassigned** in the latter case.

An assignment is **feasible** if no two students are assigned to the same school.

Given any problem π, an assignment μ is **stable** if it satisfies each of the three following properties.

Individual rationality: For any student i, we have $\mu(i)\ R_i\ i$.

Non-wastefulness: For any student i and any school s, if $s\ P_i\ \mu(i)$, then $\#\{j \in I \mid \mu(j) = s\} = 1$.

No justified-envy: For any two students i and j, if $\mu(j)\ P_i\ \mu(i)$, then j is a competitor of i at school $\mu(j)$.

A **(school choice) mechanism** M is a function that associates every problem π in some domain $\Pi^M \subseteq \Pi$ of problems with a feasible assignment. We say that a mechanism is individually rational, non-wasteful or stable, if $M(\pi)$ is individually rational, non-wasteful or stable for all $\pi \in \Pi^M$. As is common, when there is no ambiguity about $\trianglerighteq$, we often use $M(R)$ to denote the assignment selected by mechanism M.

We assume that the three schools report their priority ordering truthfully to the mechanism. A **type profile** y is a school choice problem $\pi = (R, \trianglerighteq)$ (and thus $Y = \Pi$), and the **players** of mechanism M are the three students. For the two mechanisms that we consider, the **strategy space** S_i of each student i consists in the set of reported preference Q_i for which at least one school is unacceptable and at least one school is acceptable.

For any type profile y, the pair (M, y) defines a strategic form game for which students report a preference and the outcome is the assignment selected by M under the profile of reported preferences. Given (M, y), the strategy-space of student i is the set of all the preferences of i that are featured in at least one problem of Π^M. We call these strategies **reported preferences**. A **reported profile** is a list $Q := (Q_{i_1}, Q_{i_2}, Q_{i_3})$ of the reported preferences of all students.

The outcome of the game when students report Q is assignment $M(Q)$. Student i evaluates this assignment according to her true preference R_i. In particular, strategy Q_i is a **(weakly) dominant strategy** for student i if

$$M_i(Q_i, Q_{-i})\ R_i\ M_i(Q'_i, Q_{-i}), \qquad \text{for any } Q_{-i} \text{ and any } Q'_i.$$

In turn, strategy Q_i is a **dominated strategy** for student i if

$$M_i(Q'_i, Q_{-i})\ R_i\ M_i(Q_i, Q_{-i}), \qquad \text{for any } Q_{-i} \text{ and some } Q'_i$$

and $M_i(Q'_i, Q'_{-i})\ P_i\ M_i(Q_i, Q'_{-i})$ for some Q'_{-i}. A strategy is **undominated** if it is not dominated.

Two mechanisms

In this section we describe the two school choice mechanisms we compare, which are members of the class considered in Haeringer and Klijn (2009). We first describe BOS^2, a constrained version of the Boston mechanism for which students are allowed to report preferences on two schools only.

Input : A (reported) school choice profile.

Round 1: Students apply to the school they reported as their favorite school. Every school that receives more applications than its capacity starts rejecting the lowest applicant in its priority ranking, up to the point where it meets its capacity. All other applicants are *definitively accepted* at the schools they applied to, and capacities are adjusted accordingly.

Round 2: Students who are not yet assigned apply to the school they reported as their second favorite school. Every school that receives more *new* applications in round 2 than its *remaining* capacity starts rejecting the lowest *new* applicants in its priority ranking, up to the point where it meets its capacity. All other applicants are definitively accepted at the schools they applied to. The algorithm terminates and all students not yet assigned remain unassigned.

We now turn to DA^2, a constrained version of the Deferred Acceptance mechanism for which students are allowed to report preferences on two schools only.

Input : A (reported) school choice profile.

Round 1: Students apply to the school they reported as their favorite school. Every school that receives more applications than its capacity *definitively rejects* the lowest applicant in its priority ranking, up to the point where it meets its capacity. All other applicants are *temporarily* accepted at the schools they applied to (this means they could be rejected at a later point).

Round 2: Students who were rejected in round 1 apply to the school they reported as their second favorite school. Every school considers the new applicants of round 2 *together with* the students it temporarily accepted. If needed, each school *definitely rejects* the lowest students in its priority ranking, up to the point where it meets its capacity. The algorithm terminates and all students not yet assigned remain unassigned.

Preliminary results on undominated strategies under DA^2 and BOS^2

Propositions 1 and 2 require identifying undominated strategies under DA^2 and BOS^2. The following lemmas provide the necessary results for such identification. They are direct implications of characterization results taken from Haeringer and Klijn (2009) and Decerf and Van der Linden (2018a).

Lemma 2 *If student i finds only one school acceptable, then reporting only this school is a dominant strategy under both BOS^2 and DA^2.*

Proof This is a straightforward implication of the characterization of dominant strategies in constrained *BOS* and constrained *DA* in Decerf and Van der Linden (2018b). □

Lemma 3 *Assume that the most-preferred school of student i is a top-priority school for i. Under both* BOS^2 *and* DA^2, (1) *i has a dominant strategy and* (2) *i is assigned to her most-preferred school when she plays her dominant strategy*.

Proof This is a straightforward implication of the characterization of undominated strategies in constrained *BOS* and dominant strategies in constrained *DA* in Decerf and Van der Linden (2018a) (Propositions 2 and 4). □

Lemma 4 *Assume that the second most-preferred school of student i is a top-priority school for i. Student i has a dominant strategy under* DA^2, *which consists in reporting these two schools truthfully*.

Proof This is a straightforward implication of the characterization of dominant strategies in constrained *DA* in Decerf and Van der Linden (2018a) (Proposition 2). □

Let student i's **most-preferred top-priority school** be the school that i prefers among the schools that are top-priority for i (if any).

Lemma 5 *Assume that the most-preferred school of student i is not a top-priority school for i. Strategy* Q_i *is undominated under* DA^2 *only if* Q_i *reports two schools*, Q_i *ranks these two schools according to i's true preference and i weakly prefers these two schools over her most-preferred top-priority school*.

Proof Haeringer and Klijn (2009) (Proposition 4.2) show that a necessary condition for Q_i to be undominated under DA^2 is that Q_i reports two schools and Q_i ranks these two schools according to i's true preference. Decerf and Van der Linden (2018a) (Proposition 3) show that another necessary condition for Q_i to be undominated under DA^2 is that i weakly prefers these two schools over her most-preferred top-priority school. □

Lemma 6 *Strategy* Q_i *is undominated under* BOS^2 *if and only if* (*i*) *the school reported first is i's most-preferred top-priority school or* (*ii*) *the school reported first is not top-priority for i and* Q_i *reports two schools, one of which is strictly preferred to i's most-preferred top-priority school*.

Proof This is a straightforward implication of the characterization of undominated strategies in constrained *BOS* in Decerf and Van der Linden (2018a) (Proposition 4). □

Acknowledgements We are very grateful to Martin Van der Linden for helpful comments and suggestions. We thank John Weymark who commented on a preliminary version of this work. We are grateful to one anonymous referee, one anonymous co-editor and the editor for suggestions that greatly helped improve the paper. We thank all the participants to the DEFIPP workshop and the 13th Meeting of the Society for Social Choice and Welfare for valuable comments and discussions. All remaining mistakes are of course ours.

References

Abdulkadiroğlu A, Che Y, Yasuda Y (2011) Resolving conflicting preferences in school choice: the "Boston mechanics" reconsidered. Am Econ Rev 101(1):399–410

Abdulkadiroglu A, Angrist J, Narita Y, Pathak PA (2019) Breaking ties: regression discontinuity design meets market design. Discussion paper, (2170)

Andersson T, Ehlers L, Svensson L-G (2014) Least manipulable envy-free rules in economies with indivisibilities. Math Soc Sci 69:43–49

Arribillaga RP, Massó J (2015) Comparing generalized median voter schemes according to their manipulability. Theoret Econ 11(2):547–586

Bonkoungou S, Nesterov A (2020) Reforms meet fairness concerns in school and college admissions. Working paper

Calsamiglia C, Haeringer G, Klijn F (2010) Constrained school choice: an experimental study. Am Econ Rev 100(4):1860–74

Chen Y, Kesten O (2017) Chinese college admissions and school choice reforms: a theoretical analysis. J Polit Econ 125(1):000–000

Combe J, Tercieux O, Terrier C (2017) The design of teacher assignment: theory and evidence. Working paper

Dasgupta P, Maskin E (2008) On the robustness of majority rule. J Eur Econ Assoc 6(5):949–973

Decerf B, Van der Linden M (2018a) In search of advice for participants in constrained school choice. SSRN Working Paper No. 3100311

Decerf B, Van der Linden M (2018b) Manipulability in constrained school choice. Available at SSRN 2809566

Dogan B, Ehlers L (2020a) Minimally unstable pareto improvements over deferred acceptance. SSRN working paper

Dogan B, Ehlers L (2020b) Robust minimal instability of the top trading cycles mechanism. SSRN working paper

Ergin H, Sönmez T (2006) Games of school choice under the Boston mechanism. J Public Econ 90(1–2):215–237

Fleurbaey M (2012) Social preferences for the evaluation of procedures. Soc Choice Welf 39:599–614

Gerber A, Barberà S (2016) Sequential voting and agenda manipulation. Theoret Econ 12:211–247 (**Forthcoming**)

Haeringer G, Klijn F (2009) Constrained school choice. J Econ Theory 144(5):1921–1947

Klijn F, Pais J, Vorsatz M (2013) Preference intensities and risk aversion in school choice: a laboratory experiment. Exp Econ 16:1–22

Pathak PA, Sönmez T (2013) School admissions reform in Chicago and England: comparing mechanisms by their vulnerability to manipulation. Am Econ Rev 103(1):80–106

Selten R (1991) Properties of a measure of predictive success. Math Soc Sci 21(2):153–167