A problem with strategy-proofness for abstract aggregation rules with metrically consistent preferences

Daniel Eckert and Christian Klamler
E-Mail: {daniel.eckert,christian.klamler}@uni-graz.at

Institute of Public Economics, University of Graz

Abstract. While distance functions are a natural way to derive preferences over outcomes for the aggregation of arbitrary characteristics, they also open the door to impossibility results familiar from the literature on proximity preservation in the framework of classical Arrovian social choice theory. We establish an impossibility result involving intuitively plausible properties of abstract aggregation rules all formulated in terms of distances.

1 Introduction

Strategy-proofness is a central criterion for collective decision processes such as voting or opinion pooling in expert committees. It was first formalized in the framework of social choice theory. The formulation of this condition however requires individual preferences over outcomes, which makes it difficult to apply to aggregation problems (like the aggregation of individual judgments) which are not formulated in the classical framework of preference aggregation. As distances play a well-established role in the literature on belief revision and belief merging it seems natural to circumvent this problem by deriving preferences over outcomes from arbitrary individual characteristics with the help of distance functions.[1] This however opens the door to impossibility results involving intuitively plausible properties formulated in terms of distances which are reminiscent of the impossibility results involving the condition of proximity preservation introduced in [1] into the framework of classical Arrovian social choice theory.

2 Formal framework and result

Given a set I of individuals and an arbitrary space of individual characteristics Θ (e.g. judgment sets in the context of judgement aggregation), we consider abstract aggregation rules f which assign to each profile $u = (\theta_1, ..., \theta_I) \in \Theta^I$ of individual characteristics a social characteristic $f(u) \in \Theta$. Profiles will be

[1] For a first discussion of strategy-proofness in the case of judgment aggregation which uses a similar intuition see [3].

considered as mappings $u : I \to \Theta$ and the set of all profiles Θ^I will be denoted by $\mathcal{U}$. Thus an abstract aggregation rule is a mapping $f : \mathcal{U} \to \Theta$.

Observe that by this definition of an aggregation rule, a universal domain condition is trivially fulfilled. Furthermore, the space of individual and of social characteristics is identical, as in the case of social welfare functions as opposed to social choice functions.

We make use of the following concepts:

The set of all ***i*-variants** $(\theta_1, ..., \theta_i', ..., \theta_I)$ of any profile $u = (\theta_1, ..., \theta_i, ..., \theta_I) \in \mathcal{U}$ will be denoted by $[u]_{-i}$ and is defined for any $u \in \mathcal{U}$ and $i \in I$ by

$$[u]_{-i} := \{u' \in \mathcal{U} | (\forall j \in I \backslash \{i\}) \; u'(i) \neq u(i) \wedge u'(j) = u(j)\}.$$

The concept of an i-variant can be extended from single individuals to subsets of individuals by defining for any $u \in \mathcal{U}$ and $S \subset I$

$$[u]_{-S} := \{u' \in \mathcal{U} | (\forall i \in S, \forall j \in I \backslash S) \; u'(i) \neq u(i) \wedge u'(j) = u(j)\}.$$

Conversely, $[u]_i$ will denote the set of all profiles where the characteristic of individual i is held constant and is defined for any $u \in \mathcal{U}$ and $i \in I$ by

$$[u]_i := \{u' \in \mathcal{U} : u'(i) = u(i)\}.$$

Similarly, for any profile $u \in \mathcal{U}$ the preimage of $f(u)$ will be denoted by $[u]_f$ and is defined for any $u \in \mathcal{U}$ by

$$[u]_f := \{u' \in \mathcal{U} : f(u') = f(u)\}.$$

Finally, an individual $i \in I$ is said to be **pivotal** at $u \in \mathcal{U}$ if there exists a profile $u' \in [u]_{-i} \backslash [u]_f$, i.e. if the individual can change the outcome for a given profile by changing her characteristic. (By a slight abuse of terminology, a pair of profiles $u \in \mathcal{U}$ and $u' \in [u]_{-i} \backslash [u]_f$ will also be called a pivotal pair.)

If preferences over outcomes are to be derived from distances[2] between characteristics the following consistency criterion seems intuitively plausible:

Definition 1. *A preference relation* ${}^d\succsim^\theta$ *on a set* Θ *of characteristics is* ***metrically consistent*** *with a given characteristic* $\theta \in \Theta$ *if, for the semimetric* $d : \Theta^2 \to \mathbb{R}$ *and all characteristics* $\theta', \theta'' \in \Theta$, $\theta' \succsim^\theta \theta''$ *whenever* $d(\theta, \theta') \leqslant d(\theta, \theta'')$.

With the help of these induced preference relations[3], a condition of strategy-proofness can now be formulated in the following way:

---

[2] A distance function $\Theta^2 \to \mathbb{R}$ is a semimetric if for all characteristics $\theta, \theta' \in \Theta$,

1. $d(\theta, \theta') \geqslant 0$
2. $d(\theta, \theta') = 0$ if and only if $\theta = \theta'$.
3. $d(\theta, \theta') = d(\theta', \theta)$

A metric is a semimetric which also satisfies the triangle inequality.

[3] As there is no danger of confusion we omit in the notation of the induced preference relation ${}^d\succ^\theta$ the superscript referring to the distance function.

Definition 2. *An aggregation rule $f : \mathcal{U} \to \Theta$ is **strategy-proof** if there does not exist a profile $u \in \mathcal{U}$ and an individual $i \in I$ with characteristic $u(i) = \theta$ such that $f(u') \succ^\theta f(u)$ for some i-variant $u' \in [u]_{-i}$, where $\succ^\theta$ denotes the asymmetric part of the preference relation $\succsim^\theta$.*

By the construction of the preference relation $\succsim^\theta$ from a distance function on characteristics the former definition of strategy-proofness is equivalent to the following in terms of distances:

Definition 3. *An aggregation rule $f : \mathcal{U} \to \Theta$ is **strategy-proof** with respect to a metrically consistent preference relation $\succsim^\theta$ on Θ if, for all individuals $i \in I$, and all pairs of profiles $u \in \mathcal{U}$ and $u' \in [u]_{-i}$ it holds that $d(u(i), f(u)) \leqslant d(u(i), f(u'))$.*

In the following we show that strategy-proofness is inconsistent with two intuitively plausible conditions on abstract aggregation rules equally formulated in terms of distances: an extension of the familiar Pareto condition and a new condition of minimal compensation, intended to limit the influence that any single individual can have.

Definition 4. *An abstract aggregation rule $f : \mathcal{U} \to \Theta$ is metrically **paretian** if for all characteristics $\theta, \theta' \in \Theta$ and all profiles $u \in \mathcal{U}$ it holds that $d(f(u), \theta) \leqslant d(f(u), \theta')$ whenever $d(u(i), \theta) \leqslant d(u(i), \theta')$ for all $i \in I$.*

Observe however that this property, while intuitively plausible, is stronger than the usual Pareto property, which only requires that for all characteristics $\theta \in \Theta$ and all profiles $u \in \mathcal{U}$ it holds that $d(f(u), \theta) = 0$ whenever $\theta = u(i)$ for all $i \in I$.

Our compensation property expresses in terms of distances the requirement that the influence of an individual should not be unlimited by all other individuals together:

Definition 5. *An abstract aggregation rule $f : \mathcal{U} \to \Theta$ is minimally **compensatory** if it is not the case that for any pair of profiles $u \in \mathcal{U}$ and $u' \in [u]_{-i} \backslash [u]_f$ at which some individual $i \in I$ is pivotal $d(f(u), f(u')) \leqslant d(f(u), f(u''))$ for all profiles $u'' \in [u']_{-I \backslash \{i\}}$.*

The inconsistency of these two conditions with the strategy-proofness of an abstract aggregation rule is now given by the following theorem:

Theorem 1. *A metrically paretian abstract aggregation rule $f : \mathcal{U} \to \Theta$ which is minimally compensatory cannot be strategy-proof for metrically consistent preferences.*

Proof. By minimal compensation there exist for each $i \in I$ who is pivotal at some pair of profiles, three profiles $u \in \mathcal{U}$, $u' \in [u]_{-i} \backslash [u]_f$ and $u'' \in [u']_{I \backslash \{i\}}$ such that $d(f(u), f(u'')) < d(f(u), f(u'))$. From the Pareto property it follows that it can thus not be the case that $d(u(j), f(u')) \leqslant d(u(j), f(u''))$ for all $j \in I$. Thus there exists an individual $j \in I$ such that $d(u(j), f(u'')) < d(u(j), f(u'))$. If it is not always the case that $j = i$ there must exist an individual $j \in I \backslash \{i\}$ and a profile $\widehat{u} \in [u']_i \cap [u']_f$ such that j is pivotal for the pair of profiles $\widehat{u}$ and u'' , whereby strategy-proofness is violated.

3 Discussion

From the literature on proximity preservation it is notorious that seemingly natural properties of aggregation rules formulated in terms of distances easily generate impossibility results (see [2]). The extreme weakness of the property of minimal compensation, which is introduced here to limit the influence that any single individual can have, however suggests that the lack of strategy-proofness might well be the price to pay for the avoidance of a totally uneven distribution of influence.

References

1. Baigent, N.: Preference proximity and anonymous social choice. Quarterly Journal of Economics, 102, 161–169 (1987)
2. Baigent, N., Eckert, D.: Abstract Aggregations and Proximity Preservation: An Impossibility Result. Theory and Decision, 56, 359–366 (2004)
3. Dietrich, F., List, C.: Strategy-proof judgment aggregation. Economics and Philosophy, 23, 269-300 (2007)