Preference Reversal and Group Strategy-Proofness

by

Dolors Berga* and Bernardo Moreno†

August 15th, 2020

* *Corresponding author.* Departament d'Economia, C/ Universitat de Girona, 10; Universitat de Girona, 17071 Girona, Spain. E-mail: dolors.berga@udg.edu

†Departamento de Teoría e Historia Económica, Facultad de Ciencias Económicas y Empresariales, Campus de El Ejido, 29071 Málaga, Spain. E-mail: bernardo@uma.es

Abstract: We study the problem of choosing one alternative given agent's strict preferences. We show that preference reversal (Eliaz, 2004) implies group strategy-proofness. Moreover, they are equivalent for the special cases where only two or three alternatives may be chosen.

Journal of Economic Literature Classification Numbers: D70, D71, C78.
Keywords: Social Choice Function, Group Strategy-Proofness, Preference Reversal.

1 Introduction

A social choice function f assigns one alternative to each preference profile P. Throughout the paper we assume strict preferences. Eliaz (2004) proposed the following property called preference reversal:

"If a social choice function chooses x at profile P and y at profile P', then there must be at least one agent such that she prefers x to y at P and y to x at P'."

Requiring a social choice function to satisfy preference reversal could seem to be a not too demanding property, as it just says that, when going from P to P', the position of y relative to x must have changed in favor of the new alternative for at least some agent. Indeed, Eliaz (2004) shows, in the universal domain of preferences, that any strategy-proof and onto social choice function satisfies preference reversal.[1]

Group strategy-proofness is a hard to achieve property of social choice functions, yet a very attractive one, since it frees the rule from manipulation by groups and requires the chosen outcome to be weakly efficient. Barberà, Berga, and Moreno (2010) study the domains on which it can be attained by some rules, and the relationship between those domains and the ones admitting rules that can avoid manipulation by individuals. As a particular case, when at most three alternatives can be selected by the rules, individual and group strategy-proofness turn out to be equivalent.

In this short note, we investigate the relationship between preference reversal and group strategy-proofness and show that the former has strong implications: it is a sufficient condition for social choice functions to be group strategy-proof, whatever its domain of definition might be. It is also necessary for group strategy-proofness, hence equivalent, for the special cases where only two or three alternatives are at stake.

While Section 2 introduces the model, Section 3 presents the main results and their proofs.

2 The Model

Let $N = \{1, 2, ..., n\}$ be a finite set of *agents* with $n \geqslant 2$ and A be a set of *alternatives.*

Let $\mathcal{P}$ be the set of all complete, reflexive, transitive, and antisymmetric binary relations on A. Let $P_i \in \mathcal{P}$ denote agent i's *preferences* and $P \in \mathcal{P}^n$ a preference profile written as $P = (P_C, P_{N\setminus C}) \in \mathcal{P}^n$ when we want to stress the role of coalition C in N. Let $\mathcal{D} = \prod_{i\in N} \mathcal{D}_i$ be the Cartesian product of the *set of admissible individual preferences* $\mathcal{D}_i \subseteq \mathcal{P}$.

A *social choice function* (or a rule) on $\mathcal{D}$ is a function $f : \mathcal{D} \to A$. We will omit mentioning the domain of f when no confusion arises.

One of the best known incentive properties is that of strategy-proofness, requiring the truth to be a dominant strategy for all agents. A more demanding

[1] See an unpublished work by Barberà (1981) for a related work.

form of nonmanipulations is obtained by requiring that no group of individuals, of any size, could benefit from joint departures from truthful preference revelation. This is the general idea underlying the notion of group strategy-proofness.

Definition 1 *A social choice function f on $\mathcal{D}$ is manipulable at $P \in \mathcal{D}$ by coalition $C \subseteq N$ if there exists $P'_C \in \prod_{i \in C} \mathcal{D}_i$ such that $f(P'_C, P_{N \setminus C}) P_i f(P)$ for all $i \in C$. A social choice function is **group strategy-proof** if it is not manipulable at any profile P by any coalition C.*

Notice that when the coalition C is a singleton we have strategy-proofness. In the proofs we will be explicit about when and how a coalition manipulates by writing that "a coalition C manipulates a social choice function f at a profile P via a subprofile of this coalition, P'_C ".

We now define the property introduced by Eliaz (2004), called preference reversal.

Definition 2 *A social choice function f on $\mathcal{D}$ satisfies **preference reversal** if for any preference profiles $P, P' \in \mathcal{D}$ and alternatives $x, y \in A$, $x \neq y$ such that $f(P) = x$ and $f(P') = y$, then there exists an agent $i \in N$ for which xP_iy and yP'_ix.*

3 Results and discussion

In this section we state and prove our results. First, whatever the domain of definition, any social choice function satisfying preference reversal does also satisfy group strategy-proofness. Second, both properties are equivalent when there are at most three alternatives, however this equivalence can not be generalized as we show by means of Example 5.

Theorem 3 *Any social choice function f satisfying preference reversal is group strategy-proof.*

Proof. By contradiction, suppose that f is not group strategy-proof. That is, there exist $C \subseteq N$, $P \in \mathcal{D}$, and $P'_C \in \mathcal{D}$, such that for any agent $i \in C$, $f(P'_C, P_{N \setminus C}) P_i f(P)$. Let $y = f(P'_C, P_{N \setminus C})$ and $x = f(P) \neq y$. Define $P' = (P'_C, P_{N \setminus C})$. By preference reversal, there exists an agent $k \in N$ for which xP_ky and yP'_kx. Since preferences of agents in $N \setminus C$ do not change, then $k \in C$. However, for each agent in C, yP_kx, which is the desired contradiction. ■

Although the converse does not hold in general, it is satisfied when there are at most three alternatives.

Theorem 4 *If there are at most three alternatives, any group strategy-proof social choice function f satisfies preference reversal.*[2]

[2] The result also holds for any set of alternatives when the rule selects at most three of them. The same proof works.

Proof. By contradiction, suppose that f is group strategy-proof but it violates preference reversal. That is, there exist P, $P' \in \mathcal{D}$ such that $f(P) = x$, $f(P') = y$, $y \neq x$, and for all i either (1) xP_iy and xP'_iy (2) yP_ix and yP'_ix, or (3) yP_ix and xP'_iy. Define as S_1, S_2, and S_3 the sets of agents satisfying each one of the these options, respectively. Consider the following cases:
Case 1: $S_1 = \varnothing$. Agents $S_2 \cup S_3$ would manipulate f at P via (P'_{S_2}, P'_{S_3}) since $f(P'_{S_2}, P'_{S_3}) = yP_if(P) = x$ for any $i \in S_2 \cup S_3$, which is a contradiction.
Case 2: $S_2 = \varnothing$. Agents $S_1 \cup S_3$ would manipulate f at P' via (P_{S_1}, P_{S_3}) since $f(P_{S_1}, P_{S_3}) = xP'_if(P') = y$ for any $i \in S_1 \cup S_3$, which is a contradiction.
Case 3: $S_1 \neq \varnothing$, $S_2 \neq \varnothing$. Suppose first that $S_3 \neq \varnothing$.
By group strategy-proofness, f must have the outcomes set as in Table 1 because: (1) $f(P'_{S_1}, P_{S_2}, P_{S_3}) \neq y$ since otherwise S_1 would manipulate f at $(P'_{S_1}, P_{S_2}, P_{S_3})$ via P_{S_1}; (2) $f(P_{S_1}, P'_{S_2}, P_{S_3}) \neq y$ since otherwise S_2 would manipulate f at P via P'_{S_2}; (3) $f(P_{S_1}, P_{S_2}, P'_{S_3}) \neq y$ since otherwise S_3 would manipulate f at P via P'_{S_3}; (4) $f(P_{S_1}, P'_{S_2}, P'_{S_3}) \neq x$ since otherwise S_1 would manipulate f at P' via P_{S_1}; (5) $f(P'_{S_1}, P_{S_2}, P'_{S_3}) \neq x$ since otherwise S_2 would manipulate f at $(P'_{S_1}, P_{S_2}, P'_{S_3})$ via P'_{S_2}; (6) $f(P'_{S_1}, P'_{S_2}, P_{S_3}) \neq x$ since otherwise S_3 would manipulate f at $(P'_{S_1}, P'_{S_2}, P_{S_3})$ via P'_{S_3}; (7) $f(P_{S_1}, P'_{S_2}, P_{S_3}) \neq x$ since otherwise $S_1 \cup S_3$ would manipulate f at P' via (P_{S_1}, P_{S_3}); (8) $f(P'_{S_1}, P_{S_2}, P'_{S_3}) \neq y$ since otherwise $S_1 \cup S_3$ would manipulate f at $(P'_{S_1}, P_{S_2}, P'_{S_3})$ via (P_{S_1}, P_{S_3}); (9) $f(P_{S_1}, P'_{S_2}, P'_{S_3}) \neq y$ since otherwise $S_2 \cup S_3$ would manipulate f at P via (P'_{S_2}, P'_{S_3}); (10) $f(P'_{S_1}, P_{S_2}, P_{S_3}) \neq x$ since otherwise $S_2 \cup S_3$ would manipulate f at $(P'_{S_1}, P_{S_2}, P_{S_3})$ via (P'_{S_2}, P'_{S_3}). Thus, f is defined as in Table 1 where z is neither x nor y:[3]

f	P_{S_3}	
	P_{S_2}	P'_{S_2}
P_{S_1}	x	z
P'_{S_1}	z	y/z

f	P'_{S_3}	
	P_{S_2}	P'_{S_2}
P_{S_1}	x/z	z
P'_{S_1}	z	y

Table 1. Outcomes of f when $S_3 \neq \varnothing$.

Suppose now that $S_3 = \varnothing$. Then, applying group strategy-proofness four times as follows we obtain that f is defined as in Table 2. First, $f(P'_{S_1}, P_{S_2}) \neq y$ since otherwise S_1 would manipulate f at (P'_{S_1}, P_{S_2}) via P_{S_1}. Second, $f(P_{S_1}, P'_{S_2}) \neq y$ since otherwise S_2 would manipulate f at (P_{S_1}, P_{S_2}) via P'_{S_2}. Third, $f(P'_{S_1}, P_{S_2}) \neq x$ since otherwise S_2 would manipulate f at (P'_{S_1}, P_{S_2}) via P'_{S_2}. Fourth, $f(P_{S_1}, P'_{S_2}) \neq x$ since otherwise S_1 would manipulate f at P' via P_{S_1}.

f	P_{S_2}	P'_{S_2}
P_{S_1}	x	z
P'_{S_1}	z	y

Table 2. Outcomes of f when $S_3 = \varnothing$.

[3] Note that in the case of two alternatives, we would get a contradiction in Table 1 since the social choice function would not be well-defined.

The following observation applies for S_3 being empty or not. Observe that $S_1 = T_1 \cup (S_1 \backslash T_1)$ such that for each agent $i \in T_1$, $xP'_i yP'_i z$ and for each agent i in $S_1 \backslash T_1$, $zP'_i y$ (and either $xP'_i z$ or $zP'_i x$). If $T_1 = \varnothing$, then $S_1 \backslash T_1 = S_1$ would manipulate f at P' via P_{S_1}. If $S_1 \backslash T_1 = \varnothing$, then $T_1 = S_1$ would manipulate f at $(P'_{S_1}, P_{N \backslash S_1})$ via P_{S_1}. Therefore, neither T_1 nor $(S_1 \backslash T_1)$ are empty. Similarly, $S_2 = T_2 \cup (S_2 \backslash T_2)$ such that for each agent $j \in T_2$, $zP_j yP_j x$ and for each agent j in $S_2 \backslash T_2$, $yP_j z$ (and either $xP_j z$ or $zP_j x$). If $T_2 = \varnothing$, then S_2 would manipulate f at $(P'_{S_1}, P_{S_2}, P'_{S_3})$ via P'_{S_2}. If $S_2 \backslash T_2 = \varnothing$, then S_2 would manipulate f at P via P'_{S_2}. Therefore, neither T_2 nor $(S_2 \backslash T_2)$ are empty.
Let $S_3 \neq \varnothing$. By group stratregy-proofness, $f(P'_{T_1}, P_{N \backslash T_1}) = x$, otherwise T_1 would manipulate f at $(P'_{T_1}, P_{N \backslash T_1})$ via P_{T_1} (1 in Table 3). Then, $f(P'_{T_1}, P'_{T_2}, P_{N \backslash (T_1 \cup T_2)}) = x$, otherwise T_2 would manipulate f at $(P'_{T_1}, P_{N \backslash T_1})$ via P'_{T_2} (2 in Table 3).
By group stratregy-proofness, $f(P_{S_1 \backslash T_1}, P'_{N \backslash (S_1 \backslash T_1)}) = y$, otherwise $S_1 \backslash T_1$ would manipulate f at P' via $P_{S_1 \backslash T_1}$ (3 in Table 4). Then, $f(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2}, P'_{S_3}) = y$, otherwise $S_2 \backslash T_2$ would manipulate f at $(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2}, P'_{S_3})$ via $P'_{S_2 \backslash T_2}$ (4 in Table 4). Finally, since $f(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2}, P_{S_3}) = x$ (by 2 in Table 3), $f(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2}, P'_{S_3}) = y$ (by 4 in Table 4), and agents in S_3 strictly prefer x to y under P', S_3 would manipulate f at $(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2}, P'_{S_3})$ via P_{S_3}. Thus, we get a contradiction when $S_3 \neq \varnothing$.

f for $P_{(S_1 \backslash T_1) \cup (S_2 \backslash T_2) \cup S_3}$		
	P_{T_2}	P'_{T_2}
P_{T_1}	x	
P'_{T_1}	x (1)	x (2)

Table 3. Outcomes of f for $P_{(S_1 \backslash T_1) \cup (S_2 \backslash T_2) \cup S_3}$ departing from Table 1.

f for $P'_{T_1 \cup T_2 \cup S_3}$		
	$P_{S_2 \backslash T_2}$	$P'_{S_2 \backslash T_2}$
$P_{S_1 \backslash T_1}$	y (4)	y (3)
$P'_{S_1 \backslash T_1}$		y

Table 4. Outcomes of f for $P'_{T_1 \cup T_2 \cup S_3}$ departing from Table 1.

Suppose now that $S_3 = \varnothing$. By group strategy-proofness, $f(P'_{T_1}, P_{N \backslash T_1}) = x$, otherwise T_1 would manipulate f at $(P'_{T_1}, P_{N \backslash T_1})$ via P_{T_1} (1' in Table 5). Then, $f(P'_{T_1}, P'_{T_2}, P_{N \backslash (T_1 \cup T_2)}) = x$, otherwise T_2 would manipulate f at $(P'_{T_1}, P_{N \backslash T_1})$ via P'_{T_2} (2' in Table 5). By group stratregy-proofness, $f(P_{S_1 \backslash T_1}, P'_{T_1}, P'_{S_2}) = y$, otherwise $S_1 \backslash T_1$ would manipulate f at P' via $P_{S_1 \backslash T_1}$ (3' in Table 5). On the other hand, $f(P_{S_1 \backslash T_1}, P'_{T_1}, P'_{S_2}) \neq y$, otherwise $S_2 \backslash T_2$ would manipulate f at $(P_{S_1 \backslash T_1}, P_{S_2 \backslash T_2}, P'_{T_1}, P'_{T_2})$ via $P'_{S_2 \backslash T_2}$ (4' in Table 5). This is a contradiction.

$P_{S_2\setminus T_2}$					
P_{T_2}			P'_{T_2}		
	P_{T_1}	P'_{T_1}		P_{T_1}	P'_{T_1}
$P_{S_1\setminus T_1}$	x	$x(1')$	$P_{S_1\setminus T_1}$		$x(2')$
$P'_{S_1\setminus T_1}$			$P'_{S_1\setminus T_1}$		
$P'_{S_2\setminus T_2}$					
P_{T_2}			P'_{T_2}		
	P_{T_1}	P'_{T_1}		P_{T_1}	P'_{T_1}
$P_{S_1\setminus T_1}$			$P_{S_1\setminus T_1}$		$?(3')(4')$
$P'_{S_1\setminus T_1}$			$P'_{S_1\setminus T_1}$		y

Table 5. Outcomes of f for $S_3 = \varnothing$ departing from Table 2.

This ends the proof. ■

Example 5 shows that with four alternatives in the range, there exist group strategy-proof social choice functions violating preference reversal.

Example 5 *Let $A = \{x, y, z, w\}$, $N = \{1, 2\}$, and the set of preferences profiles are $\mathcal{D}$=$\{(P_1, P_2)$,(P_1, P_2'),(P_1', P_2),$(P_1', P_2')\}$ where $xP_1wP_1zP_1y$, $zP_1'xP_1yP_1w$, $yP_2zP_2wP_2x$ and $zP_2'yP_2'xP_2'w$. Let f be a social choice function defined as follows:*

f	P_2	P_2'
P_1	x	w
P_1'	z	y

Observer that f is group strategy-proof but it violates preference reversal: let P, P', x and y and observe that xP_1y and $xP_1'y$ and yP_2x and $yP_2'x$.

We finally discuss two possible directions for further research. In problems of choice among multi-dimensional alternatives when preferences are multidimensional single-peaked, there exists a large class of rules that are strategy-proof,[4] however, it is well-known that there are members of this class that violate group strategy-proofness, hence preference reversal according to Theorem 3. Barberà, Berga, and Moreno (2010) show that under sequential inclusion, group and individual strategy-proofness are equivalent. It would be interesting, and left for future research, to explore under what circumstances group strategy-proofness and preference reversal coincide. For the case of two alternatives, there are several papers characterizing the set of rules satisfying (group) strategy-proofness.[5] Since our result states the equivalence between strategy-proofness and preference reversal, the latter can be used in all those characterization when it applies. For the case of three or more alternatives Eliaz (2004) obtains an impossibility

[4]See, for example, Border and Jordan (1983), Barberà, Sonnenschein, and Zhou (1991), Barberà, Gul, and Stacchetti (1993).

[5]See Larsson and Svensson (2006), Manjunath (2012), and recently Basile, Rao, and Rao (2020).

result of rules satisfying preference reversal when all strict preferences are admissible. Another question for further research would be to characterize the class of social choice functions satisfying preference reversal on well-known domains.

Acknowledgements

We thank the anonymous referee for the detailed reading of the paper. We also appreciate comments of S. Barberà and A. Nicolò. D. Berga acknowledges the support from the Spanish Ministry of Economy, Industry and Competitiveness through grant ECO2016-76255-P. B. Moreno thanks the support from Junta de Andalucía through grant UMA18-FEDERJA-130. Both authors thank the MOMA network. The usual disclaimer applies.

References

Barberà, S. (2001). A theorem on preference aggregation. *Mimeo*, Universitat Autònoma de Barcelona.

Barberà, S., Berga, D., Moreno, B. (2010). Individual versus group strategy-proofness: When do they coincide?. *Journal of Economic Theory* 145: 1648–1674.

Barberà, S., Berga, D., Moreno, B. (2012). Group strategy-proof social choice functions with binary ranges and arbitrary domains: characterization results. *International Journal of Game Theory* 41 (4): 791–808.

Barberà, S., Gul, F., Stacchetti, E. (1993). Generalized median voter schemes and committees. *Journal of Economic Theory* 61: 262–289.

Barberà, S., Sonnenschein, H., Zhou, L. (1991). Voting by committees. *Econometrica* 59: 595-609.

Basile, A., Rao, S., Rao, K.P.S.B. (2020). Binary strategy-proof social choice functions with indifference. *Economic Theory*. https://doi.org/10.1007/s00199-020-01273-1

Border, K., Jordan, J. (1983). Straightforward elections, unanimity and phantom voters. *Review of Economic Studies* 50: 153–170.

Eliaz, K. (2004). Social Aggregators. *Social Choice and Welfare* 22, 2: 317–330.

Larsson, B., Svensson, L.-G. (2006). Strategy-proof voting on the full preference domain. *Mathematical Social Sciences* 52:272–287.

Manjunath, V. (2012). Group strategy-proofness and social choice between two alternatives. *Mathematical Social Sciences* 63: 239–242.