Self-implementation of social choice correspondences 
in Nash equilibrium

Saptarshi Mukherjee\textsuperscript{1,3} · Hans Peters\textsuperscript{2}

Received: 5 March 2021 / Accepted: 24 June 2022 / Published online: 18 July 2022 © The Author(s) 2022

Abstract
A social choice correspondence is Nash self-implementable if it can be implemented in Nash equilibrium by a social choice function that selects from it as the game form. We provide a complete characterization of all unanimous and anonymous Nash self-implementable social choice correspondences when there are two agents or two alternatives. For the case of three agents and three alternatives, only the top correspondence is Nash self-implementable. In all other cases, every Nash self-implementable social choice correspondence contains the top correspondence and is contained in the Pareto correspondence. In particular, when the number of alternatives is at least four, every social choice correspondence containing the top correspondence plus the intersection of the Pareto correspondence with a fixed set of alternatives, is self-implementable.

1 Introduction

A social choice correspondence assigns a set of alternatives to a profile of agents’ preferences. A specific social choice correspondence—for instance, a particular political election system—may have been adopted since it has some properties that are regarded as attractive, but these properties may be lost if agents behave strategically—do not vote according to their true preferences. Thus, agents should be
provided with incentives in order to reach a truthful alternative, that is, an alternative assigned by the social choice correspondence to the true preference profile. This leads to the classical question of implementation, which dates back to Hurwicz (1972) and Maskin (1999): given a social choice correspondence, is there a game form such that for every preference profile the equilibrium outcomes of the resulting game coincide with the outcomes of the social choice correspondence? Here, ‘equilibrium’ may refer to Nash equilibrium or to other equilibrium concepts, for instance strong equilibrium.

Game forms in the literature used for implementing social choice correspondences can be quite complicated. Typically, they may ask agents to report complete preference profiles, and they often use integers or similar devices as part of strategies in order to endow single agents or coalitions—depending on the equilibrium concept—with sufficient possibilities to deviate, thus avoiding equilibria with outcomes that are not contained in the set assigned by the social choice correspondence. Although in the complete information context—also assumed in the present paper—agents are supposed to know all preferences, it is a demanding task for them to report complete preference profiles, and this is also not what one would ask from them in practical applications. A similar observation holds for other common ingredients of strategies, such as reporting an integer or an alternative.

Indeed, the complexity of the message space or space of strategies is an issue that has gathered much attention in the literature (see already Mount and Reiter 1974). Saijo (1988) considers implementation where agents are ordered, and only report their own preference and the preference of the next agent, besides an alternative and an integer. Tatamitani (2001) further reduces this to reporting only one’s own preference, plus an outcome and an integer. Recently, Koray and Yildiz (2018) consider implementation via rights structures, based on an idea by Sertel (2001), which also avoids the use of integer devices. For a survey on the basics of implementation see Jackson (2001).

In the present paper we restrict our attention to direct game forms which are single-valued selections (social choice functions) from the social choice correspondence under consideration. This means, simply, that the strategy space of each agent consists of the set of all preferences—agents do not report complete profiles, integers, or any similar or other information—and the social choice function selects (at every preference profile) from the social choice correspondence to be implemented. We consider full implementation in Nash equilibrium: the set of Nash equilibrium alternatives coincides with the set of alternatives assigned by the social choice correspondence. We do not require truthful implementation, that is, we do not require that reporting the true preferences is always a Nash equilibrium, as in Dasgupta et al. (1979).

Our main motivation for studying this case is twofold. First, considering that a social choice correspondence in vigor in a group or society is usually regarded as appropriate or desirable by its members, choosing according to a selection from it seems to be natural, both for simplicity and in order to convince agents to accept the game form (the rules of the game). In other words, since the direct mechanism used for implementing the social choice correspondence is associated with a selection
from it, the mechanism carries the same ‘spirit’ as that of the correspondence.\(^1\) Several papers have discussed the advantages of direct game forms. For example, Hammond (1996) shows that such game forms, in which individual strategies are direct reports of their preferences, are implicit in Sen’s model of rights (Sen 1970).

Second, but related, it seems to be an obvious and natural question to establish which social choice correspondences can be implemented this way, i.e., by a direct game form selecting from the correspondence. One could call this kind of implementation ‘natural’, but this word already has different meanings in this context (e.g., Saijo et al. 1996). Therefore, we use the term ‘self-implementation’. Also this expression, however, is not entirely new. Abdou and Keiding (1991, Sect. 6.1) informally suggest the term for partial implementation of a social choice correspondence by a selection from it. If this implementation is in strong equilibrium, then it coincides with exact and strong consistency as first studied in Peleg (1978). See also Remark 4.2.7 in Peleg (2002), where the term self-implementation is used with the same meaning as in Abdou and Keiding (1991).

As mentioned, we will focus attention on self-implementation in Nash equilibrium.\(^2\) Moreover, we consider only social choice correspondences which are unanimous and anonymous, and only anonymous selections (social choice functions). Under these restrictions our results are as follows. Let the ‘top correspondence’ $TC$ assign to each preference profile all alternatives that occur at top for at least one agent, and let the ‘Pareto correspondence’ $PC$ assign all Pareto optimal alternatives. Then every anonymous and unanimous Nash self-implementable social choice correspondence contains $TC$ and is contained in $PC$.\(^3\)\(^4\)\(^5\) If there are exactly two alternatives, then $TC$ and $PC$ of course coincide; and we show that $TC$ is Nash self-implementable if and only if the number of agents is not equal to 2, 3, 4, or 6; moreover, we describe all implementing selections (Theorem 3.3). If there are exactly two agents, then no anonymous and unanimous social choice correspondence is Nash self-implementable. These results, both for two alternatives and for two agents, can also be derived from results of Hurwicz and Schmeidler (1978), but we provide self-contained proofs. If there are three agents and three alternatives then only $TC$

\(^1\) In particular also, the effectivity function of the social choice correspondence is contained in the effectivity function of the selection, which can be interpreted as coalitions not losing any power. The two effectivity functions are not equal: if they were, then the social choice correspondence would be ‘constitutionally implementable’ (Peleg and Winter 2002), but since—as it turns out—a social choice correspondence implementable by a selection from it contains the top correspondence, this means that every coalition different from the grand coalition is only effective for the set of all alternatives; in turn this implies (by Theorem 6.2 in Peleg and Winter 2002) that the social choice correspondence cannot be constitutionally implementable. See also Peleg et al. (2005).

\(^2\) For self-implementation in strong equilibrium (Aumann 1959) see Peleg and Peters (2019), and also Sect. 6.3 in the present paper.

\(^3\) In particular, this means that Nash self-implementability implies that the implementing game form (social choice function) is ‘acceptable’—see Hurwicz and Schmeidler (1978) and Dutta (1984).

\(^4\) The latter is not a new observation. E.g., Peleg et al. (2005) show that Maskin monotonicity (which is implied by Nash implementability) and unanimity imply Pareto optimality.

\(^5\) Since any Nash self-implementable social choice correspondence contains the top correspondence, our game forms do not ‘enforce compromises’ in the sense of Börgers (1991), where Nash implementation in undominated strategies is considered.
is Nash self-implementable. If there are at least three agents and at least four alternatives, then any correspondence containing $TC$ plus the intersection of $PC$ with a fixed set of alternatives is Nash self-implementable; if this fixed set is empty then we obtain $TC$, and if this fixed set is the set of all alternatives then we obtain $PC$—it is an open question if there are still other social choice correspondences in between $TC$ and $PC$ that are Nash self-implementable. The remaining case is the case with three alternatives and at least four agents: then $TC$ is Nash self-implementable but it is an open question whether $TC$ is uniquely Nash self-implementable.

To the best of our knowledge, this concept of Nash self-implementation has not been explicitly studied in the literature, although some of our results can be derived from existing results, for instance the two-agent or two-alternatives case mentioned above. Nevertheless, as mentioned the proofs below are self-contained. Also, the social choice correspondences prominent in this paper have occurred frequently in the literature, for instance the Pareto Correspondence only recently in Mukherjee et al. (2019).

Throughout the paper we assume that the domain of preferences is the collection of all linear orders, i.e., strict preferences, but we show in Sect. 6 that our results can easily be extended to preference domains containing also weak preferences—the main difference is that in that case self-implementable correspondences may include weakly Pareto optimal alternatives. See Sect. 6.4 for details.

The outline of the paper is as follows. Section 2 presents the main definitions and a few basic results. Section 3 deals with the two agents or two alternatives cases. The case with at least four alternatives is presented in Sect. 4, while Sect. 5 treats the three alternatives case. Section 6 concludes with a summary and further discussion.

### 2 Definitions and preliminary results

Let $A$ be the set of $m \in \mathbb{N}$ alternatives, $m \geq 2$, and let $N = \{1, \ldots, n\}, n \in \mathbb{N}, n \geq 2$, be the set of agents. Let $L$ be the set of preferences, i.e., reflexive, complete, antisymmetric and transitive binary relations, on $A$. Then $L^N$ is the set of (preference) profiles. A social choice correspondence is a function $H : L^N \to 2^A \setminus \{\emptyset\}$. A social choice function is a function $F : L^N \to A$. For a social choice function $F$ and a profile $R^N$ we denote by $(F, R^N)$ the (ordinal) game with the agents as players who each have strategy set $L$, and for a (strategy) profile $Q^N \in L^N$ the outcome (alternative) $F(Q^N)$ is evaluated by $R^i$ for each $i \in N$. Profile $Q^N$ is a Nash equilibrium of the game $(F, R^N)$ if $F(Q^N)R^iF(Q^N\setminus \{i\}, \bar{Q}^i)$ for every $i \in N$ and $\bar{Q}^i \in L$, where $(Q^N\setminus \{i\}, \bar{Q}^i)$ is obtained from $Q^N$ by replacing $Q^i$ by $\bar{Q}^i$.

We say that a social choice function $F$ implements $H$ in Nash equilibrium if for every $R^N \in L^N$ we have

$$H(R^N) = \{x \in A \mid x = F(Q^N) \text{ for some Nash equilibrium } Q^N \text{ of } (F, R^N)\}. \quad (1)$$

A social choice function $F$ is a selection from a social choice correspondence $H$ if $F(R^N) \in H(R^N)$ for every $R^N \in L^N$. We call social choice correspondence $H$ (Nash) self-implementable if (1) holds for some selection $F$ from $H$. 
For a preference $R \in L$ the top alternative $t(R) \in A$ is the alternative such that $t(R)x$ for all $x \in A$. The social choice correspondence $TC$, defined by

$$TC(R^N) = \{ t(R^i) \mid i \in N \}$$

for all $R^N \in L^N$ is called the top correspondence. An alternative $x$ is Pareto optimal in $R^N$ if there is no $y \in A \setminus \{x\}$ such that $yRx$ for all $i \in N$. The correspondence $PC$, defined by

$$PC(R^N) = \{ x \in A \mid x \text{ is Pareto optimal in } R^N \}$$

for all $R^N \in L^N$ is called the Pareto correspondence. Clearly, $TC(R^N) \subseteq PC(R^N)$ for all $R^N \in L^N$.

A social choice correspondence $H$ is unanimous if for all $R^N \in L^N$ and $x \in A$ such that $t(R^i) = x$ for all $i \in N$ we have $H(R^N) = \{ x \}$. It is anonymous if for every permutation $\pi$ of $N$ and every $R^N \in L^N$, we have $H(R^N) = H(Q^N)$ where $Q^i = R^{\pi(i)}$ for every $i \in N$. Throughout, we will focus on unanimous and anonymous social choice correspondences. Clearly, $TC$ and $PC$ are unanimous and anonymous.

If a social choice correspondence $H$ is unanimous and anonymous and $F$ is a selection from it, then also $F$ is unanimous, but $F$ is not necessarily anonymous. In what follows we will always consider implementation by anonymous social choice functions. For convenience, we therefore make the following assumption throughout.

**Assumption** Throughout this paper every social choice correspondence and every social choice function is unanimous and anonymous.

The following lemmas collect some basic results, which will be useful throughout. The first lemma implies that, if a selection $F$ implements social choice correspondence $H$, then for every preference profile and every alternative there is an agent who can make sure that this alternative will be chosen. Furthermore, $H$ contains all top alternatives and is contained in the set of Pareto optimal alternatives.

**Lemma 2.1** Let selection $F$ implement social choice correspondence $H$. Let $R^N \in L^N$ and $x \in A$.

(a) There is an $i \in N$ and $Q^i \in L$ such that $F(R^N \setminus \{i\}, Q^i) = x$.

(b) If $R^i = R^j$ for all $i, j \in N$, then there is a $Q \in L$ such that $F(R^N \setminus \{k\}, Q) = x$ for all $k \in N$.

(c) $TC(R^N) \subseteq H(R^N) \subseteq PC(R^N)$.

**Proof**

(a) Let $y = F(R^N)$. If $y = x$ we are done. Otherwise, consider a profile $Z^N \in L^N$ with $t(Z^i) = x$ for all $i \in N$ and $yZ^iz$ for all $z \in A \setminus \{x\}$ (i.e., $x$ is ranked first and $y$ second by every agent). Then $H(Z^N) = \{ x \}$ by unanimity, and therefore $R^N$ is
not a Nash equilibrium of \((F, Z^N)\). Hence, there is an \(i \in N\) and a \(Q' \in L\) such that \(F(R^{N\setminus\{i\}}, Q')Z_i^y\), which implies that \(F(R^{N\setminus\{i\}}, Q') = x\).

(b) This follows directly from (a) and anonymity of \(F\).

(c) Suppose that \(z \in A \setminus PC(R^N)\). Then there is a \(\tilde{z} \neq z\) such that \(\tilde{z}R^i z\) for all \(i \in N\). If \(Z^N \in L^N\) such that \(F(Z^N) = z\), then by (a) there are \(i \in N\) and \(Q' \in L\) such that \(F(Z^N \setminus \{i\}, Q') = \tilde{z}\), so that \(Z^N\) is not a Nash equilibrium in \((F, Z^N)\). Hence, \(z \not\in H(R^N)\), and thus \(H(R^N) \subseteq PC(R^N)\).

Finally, suppose that \(v \in TC(R^N)\), without loss of generality \(t(R^1) = v\). Consider now a profile \(Z^N \in L^N\) such that \(Z_i^i = Z_i^j\) and \(\tilde{v}Z_i^y\) for all \(i, j \in N\) and \(\tilde{v} \in A\). By part (b) there is a \(\tilde{Z}^1 \in L\) such that \(F(Z^N \setminus \{1\}, \tilde{Z}^1) = v\), and since \(F(Z^N \setminus \{1\}, \tilde{Z}^1) \in H(Z^N \setminus \{1\}, \tilde{Z}^1) \subseteq PC(Z^N \setminus \{1\}, \tilde{Z}^1)\) by the first part of (c) we have that \(v\tilde{Z}^1\tilde{v}\) for all \(\tilde{v} \in A\). Let \(Q^N\) be a Nash equilibrium in \((F, Z^N \setminus \{1\}, \tilde{Z}^1)\) such that \(F(Q^N) = v\). Then we have \(F(Q^N \setminus \{1\}, V^i) = v\) for all \(i \in N \setminus \{1\}\) and \(V^i \in L\) since \(\tilde{v}Z^i\tilde{v}\) for all \(i \in N \setminus \{1\}\) and \(\tilde{v} \in A\). Since, moreover, \(v = t(R^1)\), it follows that \(Q^N\) is a Nash equilibrium in \((F, R^N)\). Therefore, \(v = F(Q^N) \in H(R^N)\), and thus \(TC(R^N) \subseteq H(R^N)\). 

\(\square\)

**Remark 2.2** As is well known, a necessary condition for Nash implementability of a social choice correspondence is Maskin monotonicity. Together with unanimity, Maskin monotonicity implies Pareto optimality: see Lemma 3.1 in Peleg et al. (2005). Thus, this argument provides another proof of part of the statement in Lemma 2.1(c).

The second lemma says that for every alternative \(x\) there is a collection of \(n\) preferences with one of them different from the \(n-1\) other preferences, such that \(x\) is assigned by \(F\) to every profile containing that particular preference and at least \(n-2\) of the other preferences.

**Lemma 2.3** Let selection \(F\) implement social choice correspondence \(H\). Let \(x \in A\). Then there are \(n\) preferences, denoted by \(P_{x,1}, \ldots, P_{x,n-1}, P_{x,n} \in L\), with \(P_{x,n} \neq P_{x,j}\) for all \(j \in N \setminus \{n\}\), such that \(F(R^N) = x\) for all \(R^N \in L^N\) for which there is a permutation \(\pi\) of \(N\) and a \(k \in N \setminus \{n\}\) with \(R^{\pi(n)} = P_{x,n}\) and \(R^{\pi(j)} = P_{x,j}\) for all \(j \in \{1, \ldots, n-1\}\setminus\{k\}\).

**Proof** Let \(Z^N \in L^N\) be a profile with \(Z_i^j = Z_i^j\) and \(yZ_i^y\) for all \(i, j \in N\) and \(y \in A\). By Lemma 2.1(b), (c) we have \(F(Z^N \setminus \{n\}, V^j) = x\) for some \(V^j \in L\) with \(t(V^j) = x\). Let \(Q^N\) be a Nash equilibrium in \((F, (Z^N \setminus \{n\}, V^j))\). Then, clearly, \(F(Q^N \setminus \{1\}, V^j) = x\) for every \(i \in N \setminus \{n\}\) and \(V^j \in L\). Also, \(Q^N \neq Q^j\) for all \(i \in N \setminus \{n\}\) since otherwise by anonymity \(F(Q^N \setminus \{i\}, V^j) = x\) for every \(i \in N\) and \(V^j \in L\), contradicting Lemma 2.1(a). The lemma follows by taking \(P_{x,j} = Q^j\) for every \(i \in N\) and anonymity. 

\(\square\)

Typically, for a strategy profile consisting of reported preferences \(P_{x,1}, \ldots, P_{x,n}\) as in Lemma 2.3, only possibly the agent who reports \(P_{x,n}\) can change the outcome from \(x\) to some other alternative, implying that this strategy profile is a Nash equilibrium whenever for that agent \(x\) is the (true) top alternative.
3 Two agents or two alternatives

3.1 Two agents

For the case of two agents we have the following impossibility result. We provide a proof, but the result also follows from a more general result in Hurwicz and Schmeidler (1978), see below.

Theorem 3.1 Let \( n = 2 \) and let \( H \) be a social choice correspondence. Then \( H \) is not self-implementable.

Proof Let \( F \) be a selection from \( H \), and let \((R^1, R^2) \in L^N\) such that \( t(R^1) \neq t(R^2) \). Write \( x = F(R^1, R^2) \in H(R^1, R^2) \), and suppose that \((Q^1, Q^2) \in L^N\) is a Nash equilibrium in \((F, (R^1, R^2))\) such that \( F(Q^1, Q^2) = x \). We derive a contradiction, which completes the proof. By applying Lemma 2.1(b) to the profile \((Q^1, Q^1)\) it follows that \( F(Q^1, R) = t(R^2) \) for some \( R \in L \). By applying Lemma 2.1(b) to the profile \((Q^2, Q^2)\) it follows that \( F(R', Q^2) = t(R^1) \) for some \( R' \in L \). Since \((Q^1, Q^2)\) is a Nash equilibrium in \((F, (R^1, R^2))\) it follows that \( t(R^1) = x = t(R^2) \), which is the desired contradiction. ☐

Remark 3.2 Hurwicz and Schmeidler (1978) require, in our terminology, the game \((F, R^N)\) to have a Nash equilibrium and moreover, they require every Nash equilibrium of \((F, R^N)\) to result in an alternative in \( PC(R^N) \), for every \( R^N \in L^N \). They show that if \( n = 2 \) then \( F \) has to be ‘strongly dictatorial’: this means that there is an agent who can achieve any alternative \( x \), independent of the strategies (reported preferences) of the other agents. This is Theorem 1 in their paper. Clearly, this implies our Theorem 3.1, since strong dictatorship is ruled out by anonymity. The result of Hurwicz and Schmeidler (1978) extends to the larger class of non-strict preferences—for this, see also the discussion in Sect. 6.

3.2 Two alternatives

In this subsection we assume \( A = \{a, b\} \), and in view of Theorem 3.1 we assume \( n \geq 3 \). We say that a profile \( R^N \) is of the form \( a^k b^{n-k} \) if \( a \) occurs exactly \( k \) times on top at \( R^N \), where \( 0 \leq k \leq n \). We also use this expression to denote a(ny) profile of that form. This notation is convenient since we only consider anonymous social choice functions and correspondences. We use the expressions \( ab \) and \( ba \) for the preferences with \( a \) and \( b \) on top, respectively.

The main result in this subsection is the following theorem, which completely characterizes all self-implementable social choice correspondences. This result can also be derived from Hurwicz and Schmeidler (1978), more precisely, from Lemmas 1 and 2 in their Sect. 5. We provide a self-contained proof in Appendix A.
Theorem 3.3 Let $A = \{a, b\}$.

(a) If $n \in \{2, 3, 4, 6\}$ then there exists no self-implementable social choice correspondence.

(b) If $n \notin \{2, 3, 4, 6\}$ then the only self-implementable social choice correspondence is the Pareto correspondence $PC$. Moreover, a selection $F$ implements $PC$ if and only if it satisfies the following conditions:

(i) $F(a^n) = F(a^1 b^{n-1}) = F(a^2 b^{n-2}) = a$,

(ii) $F(b^n) = F(a^{n-1} b^1) = F(a^{n-2} b^2) = b$,

(iii) there is no $k \in \{2, \ldots, n - 2\}$ such that $F(a^{k-1}, b^{n-k+1}) = F(a^k b^{n-k}) = F(a^{k+1}, b^{n-k-1})$.

Observe that the Pareto and top correspondences coincide in this case. Also, the implementing selection $F$ in Theorem 3.3 is not necessarily unique. The conditions (i)–(iii) uniquely determine $F(a^k b^{n-k})$ for $k = n - 3, \ldots, n$ and for $k = 0, \ldots, 3$, but for other values of $k$ there is freedom of choice subject to condition (iii). For $n = 5$ and $n = 7$ we have uniqueness of $F$, but for $n = 8$ and $n = 9$ there are each time two possibilities. For $n = 8$ we may have $F(a^5 b^4) = a$ or $F(a^5 b^4) = b$. For $n = 9$ we may have $F(a^5 b^4) = a$ and $F(a^6 b^3) = b$ or $F(a^5 b^4) = b$ and $F(a^6 b^3) = a$. The number of possible selections expands rapidly as $n$ increases.\footnote{Nash implementation for the case of two alternatives is also studied recently in Xiong (2021). That paper, however, considers implementation of social choice functions, and establishes an impossibility result.}

4 At least three agents and at least four alternatives

In game forms (mechanisms) that are used to implement social choice correspondences often strategies (messages) include mentioning an integer number, or sending a comparable message, in order to provide the agents with additional strategic power. Since in the case of self-implementation agents can exclusively report a preference, in Sect. 4.1 we will label these in a convenient way in order to replace the standard integer messaging. If there are at least four alternatives then there are many different preferences and therefore many different strategies or messages. This case is treated in Sect. 4.2. The case of three alternatives requires special attention and is treated in Sect. 5.

Throughout this section we write $A = \{x_1, \ldots, x_m\}$ and $M = \{1, \ldots, m\}$. Also, for a preference $R \in L$ we use notations such as $R = xyz \ldots; R = x_{i_1} \ldots x_{i_m};$ etc., meaning that $xRyRzR \ldots; x_{i_1} R \ldots x_{i_m};$ etc.

4.1 Labelling preferences

In this section we describe a labelling of the preferences. In fact, this will just be alphabetic labelling, with $x_1, \ldots, x_m$ as the ‘alphabet’. Just for completeness, we include the formula.
For $\emptyset \neq M' \subseteq M$ and $k \in M'$ let $\pi(k, M')$ be the number of elements of $M'$ smaller than or equal to $k$, formally: $\pi(k, M') = |\{\ell \in M' \mid \ell \leq k\}|$. Observe that $\pi(k, M) = k$ for every $k \in M$. We define
\[
\nu(x_{i_1} \ldots x_{i_m}) = \sum_{\ell = 1}^{m} (\pi(i_{\ell}, \{i_{\ell}, \ldots, i_m\}) - 1)(m - \ell)! + \pi(i_{m-1}, \{i_{m-1}, i_m\})
\]
as the label of preference $R = x_{i_1} \ldots x_{i_m}$.

**Example 4.1**

(a) Suppose $m = 3$ and let $A = \{a, b, c\}$ with $a = x_1$, $b = x_2$, $c = x_3$. The following table gives the six different preferences as columns with their top alternatives in the first row and their labels in the last row.

| $R$  | $a\ a\ b\ b\ c\ c$ |
|------|------------------|
|      | $b\ c\ a\ a\ b$ |
|      | $c\ b\ c\ a\ b\ a$ |
| $\nu(R)$ | 1 2 3 4 5 6 |

(b) Let $A = \{a, b, c, d, e\}$ with $a = x_1$, etc. Then preference $R = cbaed$ has the label
\[
\nu(R) = (3 - 1)(5 - 1)! + (2 - 1)(5 - 2)! + (1 - 1)(5 - 3)! + 2 = 56.
\]

Preferences $c \ldots$ have labels 49–72, preferences $cb \ldots$ have labels 55–60, and preferences $cba \ldots$ have the labels 55 and 56.

As illustrated in Example 4.1, this way of labelling has the property that all preferences that share the same top alternative(s), are labelled consecutively with no gaps. This is important later on.

For a profile $R^N \in L^N$, we denote by $\nu(R^N) = \sum_{i \in N} \nu(R^i)$ the sum of the labels of the preferences in $R^N$.

### 4.2 At least four alternatives

Throughout this subsection, $m \geq 4$. We will exhibit a large class of self-implementable social choice correspondences, containing both the top and the Pareto correspondences.

For $B \subseteq A$ we define the social choice correspondence $H_B$ by $H_B(R^N) = TC(R^N) \cup (PC(R^N) \cap B)$ for every $R^N \in L^N$. In words, $H_B$ assigns to each preference profile all top alternatives, and additionally all Pareto optimal alternatives that are in $B$. Then $H_B$ is unanimous and anonymous. Note that $H_B = TC$ if $B = \emptyset$, and $H_B = PC$ if $B = A$. The main result of this subsection will be that $H_B$ is self-implementable for every $B \subseteq A$.

We first outline the main features of the proof of this result. If $x$ is a top alternative of a profile $R^N$, then pick an agent, say $i$, with that alternative at top, and
consider the profile with \( i \) having \( x \) at top and the remaining alternatives in the order \( x_1, \ldots, x_m \), and the other agents having \( x \) at top and the remaining alternatives in the order \( x_m, \ldots, x_1 \). To this profile \( x \) is assigned, also still if one of the agents other than \( i \) reports a different preference. Hence, this is a Nash equilibrium. Next, if \( x \in B \) is Pareto optimal but not a top alternative at \( R^N \), then we partition the remaining alternatives in \( k \) subsets for \( k \) different agents, and let each of these \( k \) agents have \( x \) at top, the alternatives in that agent’s part of the partition at bottom in the order \( x_1, \ldots, x_m \), and the remaining alternatives in between, in the order \( x_1, \ldots, x_m \). All the other agents report \( x \) at top and the other alternatives in the order \( x_m, \ldots, x_1 \). To this profile \( x \) is assigned; an agent in the first category may deviate but can only obtain an alternative from that agent’s bottom part or \( x \), while an agent of the second category can deviate but the chosen outcome will still be \( x \). This is going to be a Nash equilibrium if the partition is chosen in the right way, ensuring that each agent in the first category prefers \( x \) to the bottom set of alternatives—this is possible since \( x \) is Pareto optimal, hence for each alternative there is an agent preferring \( x \) to that alternative. All other cases are taken care of by picking from the reported top alternatives by modulo counting using the numbers \( v(R^N) \), and will never result in a Nash equilibrium.

We now proceed with the formal treatment. If a subset \( B \) of \( A \) is used in the notation of a preference \( R \), e.g., \( R = \ldots B \ldots \), this means that each alternative of \( A \) is ranked either above or below the alternatives of \( B \), and that the alternatives of \( B \) are ranked consistently with the order \( x_1 \ldots x_m \); that is, if \( x_j, x_k \in B \) and \( j < k \) then \( x_jR x_k \). The notation \( R = \ldots \bar{B} \ldots \) has a similar meaning except that now the alternatives of \( B \) are ranked in reverse order: if \( x_j, x_k \in B \) and \( j < k \) then \( x_kR x_j \).

For each \( x \in A \) let \( P_x = xA \setminus \{ x \} \in L \) and \( \bar{P}_x = xA \setminus \{ x \} \in L \). Thus, \( P_x \) has \( x \) as top alternative and ranks the remaining alternatives in the order \( x_1, \ldots, x_m \). Also \( \bar{P}_x \) has \( x \) as top alternative, but ranks the remaining alternatives in reverse order.

Let \( x \in A \), \( \emptyset \neq K \subseteq N \), and let \( (B_i)_{i \in K} \) be a partition of \( A \setminus \{ x \} \), i.e., \( B_i \neq \emptyset \) for every \( i \in K \), \( \cup_{i \in K} B_i = A \setminus \{ x \} \) and \( B_i \cap B_j = \emptyset \) for all distinct \( i, j \in K \). We call \( R^N \in L^N \) an \((x, (B_i)_{i \in K})\)-profile if

- \( R^i = x(A \setminus (B_i \cup \{ x \}))B_i \) for every \( i \in K \), and
- \( R^i = \bar{P}_x \) for every \( i \in N \setminus K \).

In words, for an agent \( i \in K \), in \( R^i \) the top alternative is \( x \), next the alternatives that are not in \( B_i \) are ranked in the order \( x_1, \ldots, x_m \), and finally the alternatives in \( B_i \) are ranked in the same order. Any agent not in \( K \) ranks \( x \) again at top, but all remaining alternatives in reverse order.

We first show that in an \((x, (B_i)_{i \in K})\)-profile the set \( K \) and the sets \( B_i \) are uniquely determined.

**Lemma 4.2** Let \( R^N \) be an \((x, (B_i)_{i \in K})\)-profile and also an \((x, (B'_i)_{i \in K'})\)-profile. Then \( K = K' \), and \( B_i = B'_i \) for each \( i \in K \).
Proof First observe that, since \( m \geq 4 \), by definition of the preferences we have \( N \setminus K = N \setminus K' \), and thus also \( K = K' \). Next, the only possible preference of an agent \( i \in K \) where \( B_i \) and \( B'_i \) could be different would be a preference \( R' = x(A \setminus \{x\}) \), but then there can be only one such preference in \( R^K \) (since such preferences share the same bottom alternative), so that \( B_i = (A \setminus \{x\}) \setminus \cup_{\ell \in K \setminus \{i\}} B_{\ell} = (A \setminus \{x\}) \setminus \cup_{\ell \in K \setminus \{i\}} B'_{\ell} = B'_i \).

In view of Lemma 4.2, if \( R^N \) is an \((x, (B_i)_{i \in K})\)-profile, we can call it the \((x, (B_i)_{i \in K})\)-profile.

Fix \( B \subseteq A \). We define the anonymous selection \( F_B \) from \( H_B \) (actually, from the top correspondence) as follows. Let \( R^N \in L^N \).

1. If there are \( x \in A \) and distinct \( i, j \in N \) such that \( R^j = P_x \) and \( R^k = \overline{P_x} \) for all \( k \in N \setminus \{i, j\} \), then \( F_B(R^N) = x \).
2. If there are \( x \in B_i, K \subseteq N \) with \( |K| \geq 2 \), a partition \((B_i)_{i \in K}\) of \( A \setminus \{x\} \), and \( j \not\in K \) such that \((R^N \setminus \{j\}, \overline{P_x})\) is the \((x, (B_i)_{i \in K})\)-profile, then \( F_B(R^N) = x \).
3. If there are \( x \in B_i, K \subseteq N \) with \( |K| \geq 2 \), a partition \((B_i)_{i \in K}\) of \( A \setminus \{x\} \), and \( j \in K \) such that \((R^N \setminus \{j\}, x(A \setminus (B_j \cup \{x\})) B_j)\) is the \((x, (B_i)_{i \in K})\)-profile, then
\[
F_B(R^N) = \begin{cases} t(R^j) & \text{if } t(R^j) \in B_j \\ x & \text{otherwise.} \end{cases}
\]

Otherwise, let \( TC(R^N) = \{x_{j_1}, \ldots, x_{j_s}\} \) with \( j_1 < \ldots < j_s \), then \( F_B(R^N) = x_{j_s} \) where \( s = 1 + \lfloor v(R^N) \mod t \rfloor \).

Observe that the four cases in the definition of \( F_B \) are mutually exclusive and that \( F_B \) is a unanimous and anonymous selection from the top correspondence and therefore also from \( H_B \). Also observe that (1.1) treats the case of an \((x, (B_i)_{i \in K})\)-profile for \( |K| = 1 \), but has been defined separately both for clarity and since it applies also for \( x \not\in B \). Note that if, in the situation of (1.1), player \( i \) deviates, then (1.4) applies, so that player \( i \) can achieve any alternative other than \( x \) as well.

It is also important to observe that, if (1.4) applies, then any player \( i \) can achieve any alternative by putting it on top and choosing the ‘right’ preference, as explained in Sect. 5.1: there are \((m - 1)! \) possibilities, and \((m - 1)! > m \) since \( m \geq 4 \). To illustrate in more detail how this works, we consider an (arbitrary) example with four agents and four alternatives \( x_1 = a, x_2 = b, x_3 = c, \) and \( x_4 = d \). Let the preference profile be
\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
c & a & b & b \\
\cdot & d & a & d \\
\cdot & c & d & a \\
\cdot & b & c & c \\
\end{array}
\]
and suppose that agent 1 wants \( c \) to be chosen from the top elements \( a, b, c \). The preferences of agents 2, 3, and 4 have labels 6, 8, and 11, respectively, according to the formula in Sect. 4.1, hence in total 25. In order to achieve \( c \), the third top element of

\[\text{Springer}\]
the profile according to the order $abcd$, we need in (1.4) that $3 = 1 + [\nu(R^N) \mod 3]$. By choosing the preference $cabd$, which has label 13, we obtain $1 + [38 \mod 3] = 3$, as desired. In fact, $cbda$, which has label 16, would also work.

Our main result for this case is as follows.

**Theorem 4.3** $F_B$ implements $H_B$.

**Proof**

(a) Let $R^N \in L^N$ and $x \in H_B(R^N)$. We show that there is a Nash equilibrium in $(F_B, R^N)$ resulting in $x$.

If $x \in TC(R^N)$, then take $i \in N$ such that $t(R^i) = x$, and consider the profile $Q^N$ with $Q^i = P_x$ and $Q^j = \overline{P}_x$ for all $j \in N \setminus \{i\}$. Then, by (1.1), $F_B(Q^N) = x$, and again by (1.1) and the fact that $t(R^i) = x$, we have that $Q^N$ is a Nash equilibrium in $(F_B, R^N)$.

If $x \notin TC(R^N)$, then $x \in PC(R^N) \cap B$. For each agent $i \in N$ let $L(x, R^i)$ denote the strict lower contour set of $x$ at $R^i$, that is, $L(x, R^i) = \{y \in A \setminus \{x\} \mid x R^i y\}$. Define $B_1 = L(x, R^1), B_2 = L(x, R^2) \setminus B_1, \ldots$ and in general $B_{i-1} = L(x, R^i) \setminus (B_1 \cup \ldots \cup B_{i-1})$ for every $i \in N$. Let $K = \{i \in N \mid B_i \neq \emptyset\}$. Since $x \in PC(R^N)$, we have $\cup_{i \in K} L(x, R^i) = A \setminus \{x\}$, and in turn this implies that $(B_i)_{i \in K}$ is a partition of $A \setminus \{x\}$. Now let $Q^N$ be the $(x, (B_i)_{i \in K})$-profile. Observe that $|K| \geq 2$. Then $F_B(Q^N) = x$ by (1.2) or (1.3). By (1.2), $F_B(Q^N \setminus \{i\}, V^i) = x$ for every $i \notin K$ and $V^i \in L$. By (1.3), $F_B(Q^N \setminus \{i\}, V^i) \in \{x\} \cup B_i \subseteq \{x\} \cup L(x, R^i)$ for every $i \in K$ and $V^i \in L$. Hence, $Q^N$ is a Nash equilibrium in $(F_B, R^N)$.

(b) Let $x \in A$ and $Q^N, R^N \in L^N$ such that $x \in F_B(Q^N)$ and $x \notin H_B(R^N)$. We show that $Q^N$ is not a Nash equilibrium in $(F_B, R^N)$. Note that $x \notin TC(R^N)$.

If (1.1) applies to $Q^N$, then for the agent $i$ in (1.1) there is a $V^i \in L$ such that $t(V^i) = t(R^i)$ and $F_B(Q^N \setminus \{i\}, V^i) = t(V^i)$ by (1.4). Since $t(R^i) R^i x$ and $t(R^i) \neq x$, this implies that $Q^N$ is not a Nash equilibrium in $(F_B, R^N)$.

If (1.1) does not apply to $Q^N$ and $x \notin B$, then similarly every agent $i \in N$ can achieve $t(R^i)$ by (1.4), so that again $Q^N$ is not a Nash equilibrium in $(F_B, R^N)$.

Finally, suppose that (1.1) does not apply to $Q^N$ and $x \in B$. Then $x \notin PC(R^N)$, which implies that there exists $y \in A \setminus \{x\}$ such that $y R^i x$ for all $i \in N$. If (1.2) or (1.3) applies, then there is an agent $i \in K$ such that $y \in B_i$. If (1.2) applies, then agent $i$ can achieve $y$ by (1.3) or by (1.4). If (1.3) applies, then $i$ can achieve $y$ by (1.3). Hence, in each case, $Q^N$ is not a Nash equilibrium in $(F_B, R^N)$.

As already announced, by taking $B = \emptyset$ or $B = A$ we obtain the following corollary of Theorem 4.3.

**Corollary 4.4** $TC$ and $PC$ are self-implementable.

It is an open problem whether there are self-implementable social choice correspondences (necessarily between Top and Pareto) which are not of the form $H_B$. 

© Springer
5 Three alternatives

The case of three alternatives is more subtle, due to the fact that there are less different preferences and, thus, less different labels available for implicit messaging.

As in the previous section, \( A = \{x_1, \ldots, x_m\} \), with now \( m = 3 \). Throughout, however, we write \( a = x_1 \), \( b = x_2 \), and \( c = x_3 \). We repeat the table of Example 4.1:

| \( P_a \) | \( \overline{P}_a \) | \( P_b \) | \( \overline{P}_b \) | \( P_c \) | \( \overline{P}_c \) |
|---|---|---|---|---|---|
| \( a \) | \( a \) | \( b \) | \( b \) | \( c \) | \( c \) |
| \( b \) | \( c \) | \( a \) | \( a \) | \( b \) |
| \( c \) | \( b \) | \( c \) | \( a \) | \( b \) |

We first consider the top correspondence and next the Pareto correspondence.

5.1 The top correspondence

Theorem 4.3 still holds for \( m = 3 \) and the top correspondence, i.e., \( H_{\emptyset} \), but needs a different proof. The implementing selection is still \( F_B \) for \( B = \emptyset \). Its definition simplifies as follows. Let \( RN \in LN \).

1. If there are \( x \in A \) and distinct \( i, j \in N \) such that \( R^i = P_x \) and \( R^k = \overline{P}_x \) for all \( k \in N \setminus \{i, j\} \), then \( F_{\emptyset}(RN) = x \).
2. Otherwise, let \( TC(RN) = \{x_{j_1}, \ldots, x_{j_t}\} \) with \( j_1 < \ldots < j_t \), then \( F_{\emptyset}(RN) = x_{j_s} \) where \( s = 1 + \lfloor \nu(RN) \mod t \rfloor \).

Note that (2.1) corresponds to (1.1) in the general definition of \( F_B \), and (2.2) corresponds to (1.4).

We now have:

**Theorem 5.1** Let \( m = 3 \) and \( n \geq 3 \). Then \( F_{\emptyset} \) implements \( TC \).

**Proof**

(i) Let \( RN \in LN \) and let \( x \in TC(RN) \). Take an agent \( i \in N \) such that \( t(R^i) = x \), and let \( Q^N \) be the profile with \( Q^i = P_x \) and \( Q^j = \overline{P}_x \) for all \( j \in N \setminus \{i\} \). As in the proof of Theorem 4.3 it follows that \( Q^N \) is a Nash equilibrium in \((F_{\emptyset}, RN)\) with \( F_{\emptyset}(Q^N) = x \).

(ii) For the converse, suppose that \( Q^N \) is a Nash equilibrium in \((F_{\emptyset}, RN)\) for some \( Q^N \). Without loss of generality let \( F_{\emptyset}(Q^N) = a \). We wish to show that \( a \in TC(RN) \). Suppose not. Then without loss of generality \( a \) is ranked second everywhere at \( RN \), hence for every agent \( i \), \( R^i = P_b \) or \( R^i = P_c \). We distinguish cases according to \( Q^N \) and show that in each case we have a contradiction. In these cases we assign specific roles to agents, which is without loss of generality in view of Anonymity. As before, we denote \( \nu(Q^N) = \sum_{i \in N} \nu(Q^i) \), i.e., the sum of the labels of the preferences in \( Q^N \).
(1) $Q^1 = P_a$, $Q^2 = \ldots = Q^{n-1} = \overline{P_a}$. We consider the following subcases.

(1a) $Q^n = P_a$. Then for each agent $i \in \{2, \ldots, n-1\}$ we have $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = b$ for $V^i = \overline{P_b}$ and $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = c$ for $V^i = P_c$.

(1b) $Q^n = \overline{P_a}$. Similar to case (1a) but now for agent 1.

(1c) $Q^n \in \{P_b, P_c\}$. Then for each agent $i \in \{1, \ldots, n-1\}$ we have $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = b$ for $V^i = \overline{P_b}$ or $V^i = P_b$, now suppose that $R^1 = \ldots = R^{n-1} = P_c$ and $R^n \in \{P_b, P_c\}$, and suppose that $F_\emptyset(P_c, Q^{N\setminus\{1\}}) \neq c$ and $F_\emptyset(P_c, Q^{N\setminus\{1\}}) \neq b$. This implies that $(v(Q^N) - 1 + 5) \mod 3 = 0$ and $(v(Q^N) - 1 + 6) \mod 3 = 1$. In this case, let agent 2 deviate to $P_c$. Then $(v(Q^N) - 2 + 5) \mod 3 = 2$, implying that $F_\emptyset(Q^{N\setminus\{2\}}, P_c) = c$.

(1d) $Q^n \in \{P_c, P_c\}$. Then for each agent $i \in \{1, \ldots, n-1\}$ we have $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = c$ for $V^i = \overline{P_c}$ or $V^i = P_c$. Now suppose that $R^1 = \ldots = R^{n-1} = P_b$ and $R^n \in \{P_b, P_c\}$, and suppose that $F_\emptyset(P_b, Q^{N\setminus\{1\}}) \neq b$ and $F_\emptyset(P_b, Q^{N\setminus\{1\}}) \neq b$. This implies that $(v(Q^N) - 1 + 3) \mod 3 = 2$ and $(v(Q^N) - 1 + 4) \mod 3 = 0$. In this case, let agent 2 deviate to $P_c$. Then $(v(Q^N) - 2 + 3) \mod 3 = 1$, implying that $F_\emptyset(Q^{N\setminus\{2\}}, P_b) = b$.

(2) $Q^N$ is not of the form as in (1). We consider the following subcases.

(2a) $TC(Q^N) = \{a\}$. Then, if $q^i = \overline{P_a}$ for some $i \in N$ then $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = b$ for either $V^i = \overline{P_b}$ or $V^i = P_b$, and $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = c$ for either $V^i = \overline{P_c}$ or $V^i = P_c$. Otherwise, the same holds for every $i \in N$.

(2b) $TC(Q^N) = \{a, b\}$. Then for every $i \in N$ there is $V^i \in L$ such that $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = b$. Now suppose that $R^i = P_c$ for every $i \in N$. If, at $Q^N$, there is a unique agent $i$ with $q^i \in \{P_b, P_c\}$, then that agent has $V^i \in L$ such that $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = c$. Suppose there are at least two agents at $Q^N$ who have $b$ on top. Suppose $q^i = P_b$ for some $i \in N$. Then $v(Q^{N\setminus\{i\}}, P_c) \mod 3 = (v(Q^N) - 3 + 5) \mod 3 = 2$, so that $F_\emptyset(Q^{N\setminus\{i\}}, P_c) = c$. Otherwise, there is an agent $i \in N$ with $q^i = \overline{P_b}$. In that case, $v(Q^{N\setminus\{i\}}, P_c) \mod 3 = (v(Q^N) - 4 + 6) \mod 3 = 2$, so that $F_\emptyset(Q^{N\setminus\{i\}}, \overline{P_c}) = c$.

(2c) $TC(Q^N) = \{a, c\}$. This case is almost identical to case (2b) by switching the roles of $b$ and $c$.

(2d) $TC(Q^N) = \{a, b, c\}$. Then each for agent $i \in N$ with $t(q^i) = a$ there are $V^i, W^i$ such that $F_\emptyset(Q^{N\setminus\{i\}}, V^i) = b$ and $F_\emptyset(Q^{N\setminus\{i\}}, W^i) = c$. \hfill \Box

### 5.2 The Pareto correspondence

The following result implies that the Pareto correspondence is not self-implementable if, besides $m = 3$, also $n = 3$.

**Theorem 5.2** Let $m = n = 3$ and let the social choice correspondence $H$ be self-implementable. Then $H$ is the top correspondence.

**Proof** Let the selection $F$ implement $H$ in Nash equilibrium. Let $Q^N, R^N \in L^N$ such that, without loss of generality, $F(Q^N) = c \notin TC(R^N)$. In view of Lemma 2.1 it is sufficient to show that $Q^N$ is not a Nash equilibrium in $(F, R^N)$ for $R^N = (\overline{P_a}, \overline{P_a}, P_b)$. Let $\{P_{x,i} | x \in \{a, b\}, i \in N\}$ be a collection of preferences as in Lemma 2.3. If
$Q^1 \in \{P_{a,i} \mid i \in N\}$ then $F(Q^1, V^2, Q^3) = a$ for some $V^2 \in \{P_{a,i} \mid i \in N\}$, so that $Q^N$ is not a Nash equilibrium in $(F, R^N)$. A similar argument holds if $Q^1 \in \{P_{b,i} \mid i \in N\}$ and if $Q^2 \in \{P_{x,i} \mid i \in N, x \in \{a, b\}\}$. If $Q^1 = Q^2$ then by Lemma 2.1(b), $F(Q^1, Q^2, V^3) = b$ for some $V^3 \in L$, so again $Q^N$ is not a Nash equilibrium in $(F, R^N)$. Hence, $Q^1 \neq Q^2$. If $Q^1 = P_{c,3}$ or $Q^2 = P_{c,3}$ then $F(Q^1, Q^2, V^3) = c$ for every $V^3 \in L$. By Lemma 2.1(a) this implies that there is a $V^1 \in L$ with $F(V^1, Q^2, Q^3) = a$ or a $V^2 \in L$ with $F(Q^1, V^2, Q^3) = a$, and in both cases $Q^N$ is not a Nash equilibrium in $(F, R^N)$. If both $Q^1 \neq P_{c,3}$ and $Q^2 \neq P_{c,3}$ then $F(V^1, Q^2, P_{c,3}) = F(Q^1, V^2, P_{c,3}) = c$ for all $V^1, V^2 \in L$. By Lemma 2.1(a) this implies that there is a $V^3 \in L$ such that $F(Q^1, Q^2, V^3) = b$, so that also in this case $Q^N$ is not a Nash equilibrium in $(F, R^N)$.

The case of more than three agents is still open. In other words, for $m = 3$ and $n \geq 4$, does there exist a self-implementable social choice correspondence $H$ with $H \neq TC$ and $TC(R^N) \subseteq H(R^N) \subseteq PC(R^N)$ for all $R^N \in L^N$?

### 6 Summary and further discussion

The following table summarizes the results of the paper. A $\Theta$ sign means nonexistence of a unanimous and anonymous social choice correspondence Nash implementable by an anonymous selection. The question marks indicate the open problems mentioned earlier.

| $m$ | $n = 2$ | $n = 3$ | $n \geq 4$ |
|-----|---------|---------|------------|
| 2   | $\Theta$| $\Theta$| $\Theta$ for $n = 4, 6$; $PC = TC$ otherwise |
| 3   | $\Theta$| Only $TC$| $TC$, more? |
| $\geq 4$ | $\Theta$| $F_B$ (including $TC, PC$), more?| $F_B$ (including $TC, PC$), more? |

In the following subsections we discuss, respectively, the assumption of self-implementation, the roles of unanimity and anonymity, self-implementation in strong equilibrium, and the role of the preference domain.

#### 6.1 The self-implementation assumption

In this subsection we further discuss the requirement that the implementing social choice function be a selection from the social choice correspondence that is to be implemented, in other words, the self-implementation assumption. We do this by providing two examples. The first example is an implementation of $TC$ by a social choice function that does not always select from $TC$, and thus shows the (not surprising) fact that a self-implementable social choice correspondence may be implementable by a direct mechanism that is not a selection. The second example shows that a subcorrespondence of $TC$ may be implementable by a direct mechanism that is not a selection.
In both examples, as before, $A = \{x_1, \ldots, x_m\}$.

\textbf{Example 6.1} Let $n \geq 3$, $m \geq 4$, and consider the social choice function $\hat{F}$ defined as follows.

(i) $\hat{F}(R^N) = x_2$ if $R^N$ is a profile such that there is one agent with preference $x_1 x_2 x_3 x_4 \ldots x_m$ and all other agents have preference $x_1 x_3 x_2 x_4 \ldots x_m$.

(ii) $\hat{F}(R^N) = F_y(R^N)$ otherwise, where $F_y$ is defined by (2.1) and (2.2).

Then $\hat{F}$ Nash-implements $TC$, which can be seen by adding to the proof of Theorem 4.3 the consideration that a profile $R^N$ as in (i) cannot be a Nash equilibrium in the game $(F, Q^N)$ for any $Q^N \in L^N$. However, $\hat{F}$ is not a selection from $TC$.

\textbf{Example 6.2} Let $n \geq 3$, $m \geq 4$, and let the social choice correspondence $H$ be defined by $H(R^N) = \{x\}$ for all $R^N \in L^N$. We define the social choice function $\tilde{F}$ as follows. Let $R^N \in L^N$, and denote $v(R^N) = \sum_{i=1}^n v(R^i)$.

(1) If there is $\ell \in N$ such that $R^\ell = P_{x_1}$ and $R^i = \overline{P}_{x_1}$ for all $i \in N \setminus \{\ell\}$, then $\tilde{F}(R^N) = x_1$.

(2) If there is $j \in \{2, \ldots, m\}$ and distinct $\ell, \ell' \in N$ such that $R^\ell = P_{x_j}$ and $R^i = \overline{P}_{x_j}$ for all $i \in N \setminus \{\ell, \ell'\}$, then $\tilde{F}(R^N) = x_j$.

(3) Otherwise, if $TC(R^N) = \{x_{j_1}, \ldots, x_{j_t}\}$ with $j_1 < \ldots < j_t$, then $\tilde{F}(R^N) = x_{j_s}$ where $s = 1 + \left(v(R^N) \mod t\right)$.

By adding to the proof of Theorem 4.3 the consideration that a profile as in case (1a) can only be a Nash equilibrium if every agent has $x_1$ as top alternative in the true profile, we see that $\tilde{F}$ implements $H$. However, (2) implies that $\tilde{F}$ is not a selection from $H$.

\textbf{6.2 Unanimity and anonymity}

Throughout we have assumed that every social choice function and every social choice correspondence in this paper is unanimous and anonymous. If we drop unanimity, then (as an extreme example) the social choice correspondence $H$, defined by $H(R^N) = \{x\}$ for every $R^N \in L^N$, where $x \in A$ is fixed, is anonymous and self-implementable by the (unique) selection $F$ with $F(R^N) = x$. If we drop anonymity, then the (dictatorial) social choice correspondence $D$, defined by $D(R^N) = \{t(R^1)\}$ for every $R^N \in L^N$ is unanimous and self-implementable by the (unique) selection $F$ with $F(R^N) = t(R^1)$.
6.3 Self-implementation in strong equilibrium

Self-implementation in strong equilibrium requires that for a selection $F$ from a social choice correspondence $H$, the set of strong equilibrium alternatives in the game $(F, R^N)$ coincides with $H(R^N)$ for every $R^N \in L^N$, where strong equilibrium means that no subset of agents can profitably deviate. This condition is much more demanding, and has been studied in Peleg and Peters (2019). Under similar conditions as in the present paper, and if additionally the number of agents is strictly higher than the number of alternatives, this results in a (Pareto optimal) social choice correspondence based on so-called feasible elimination procedures, which were introduced in the seminal paper by Peleg (1978). Roughly, these elimination procedures consist of successive vetoing of bottom alternatives in a preference profile. For details see Peleg and Peters (2019).

6.4 Preference restrictions

Throughout the paper we have assumed that the domain of preferences is the set of all linear orders $L$ (reflexive, complete, antisymmetric, and transitive binary relations). In this subsection we discuss how this domain can be varied. Let $W$ denote the set of all weak orders, that is, reflexive, complete and transitive binary relations.

First, consider any domain $\tilde{W} \subseteq W$ satisfying the following two assumptions: (i) for all distinct $x, y \in A$ there is an $R \in \tilde{W}$ such that $xPyRz$ for all $z \in A \setminus \{x, y\}$, where $P$ denotes the asymmetric part of $R$, and (ii) for every $x \in A$ there is an $R \in \tilde{W}$ such that $zPx$ for all $z \in A \setminus \{x\}$. On such a domain, the basic results in Sect. 2 remain valid, provided that in Lemma 2.1(c) we replace $PC(R^N)$ by $WPC(R^N)$, where $x \in WPC(R^N)$ if and only if there is no alternative $y$ satisfying $yP^i x$ for every $i \in N$; thus, provided we weaken Pareto optimality of the social choice correspondence $H$ to weak Pareto optimality. This claim can be checked by going over the proofs. A special case is the original case $\tilde{W} = L$: then, of course, $WPC(R^N) = PC(R^N)$. Another special case is the case $W = W$, the set of all weak preferences.

The main result of the paper, Theorem 4.3, holds (at least) for any domain $\tilde{W}$ satisfying $L \subseteq \tilde{W} \subseteq W$, again provided that we replace $PC(R^N)$ by $WPC(R^N)$: hence, $H_B(R^N) = TC(R^N) \cup (WPC(R^N) \cap B)$ for every $R^N \in \tilde{W}^N$. In order to achieve this, the following modification to the definition of $F_B$ has to be made. In (1.3) and (1.4) of this definition we now choose a top alternative of an agent according to a fixed order, say $x_1, \ldots, x_m$, since a preference can have more than one top alternative. As to defining the number $\nu(R^N)$ for the modulo game, this can be done in the same way, but by only considering the strict preferences in a profile. Again, our claim can be checked by going over the proof.

The impossibility result for $n = 2$ in Sect. 3 also holds for $W$ instead of $L$. The result for $m = 2$ still holds for $W$ – then the top and weak Pareto correspondences coincide. Finally, Theorem 5.1 still holds for $W$ (with a simpler proof since additional labels are now available, also using preferences with indifference).
A Proof of Theorem 3.3

As in Sect. 3.2 we assume $A = \{a, b\}$ and $n \geq 3$. In Lemmas A.1–A.6 we assume that $H$ is a social choice correspondence and that selection $F$ implements $H$.

**Lemma A.1** $H = TC = PC$.

**Proof** Follows directly from Lemma 2.1(c). □

**Lemma A.2** $F(a^{1}b^{n-1}) = a$ and $F(a^{n-1}b^{1}) = b$.

**Proof** Since $F(b^{n}) = b$ by unanimity of $H$, Lemma 2.1(b) implies $F(a^{1}b^{n-1}) = a$. Since $F(a^{n}) = a$ by unanimity of $H$, Lemma 2.1(b) implies $F(a^{n-1}b^{1}) = b$. □

**Lemma A.3** $F(a^{2}b^{n-2}) = a$ and $F(a^{n-2}b^{2}) = b$.

**Proof** We only prove the first claim, the proof of the other claim is similar. By Lemma A.2 we have $F(R^{N}) = a$, where $R^{N} = a^{1}b^{n-1}$. Let $Q^{N}$ be a Nash equilibrium in $(F, a^{1}b^{n-1})$ with $F(Q^{N}) = a$, say $Q^{N} = a^{k}b^{n-k}$ for some $k \in \{1, \ldots, n-2, n\}$. By Lemma 2.1(a) and anonymity of $F$ we must have

$$F(a^{k-1}b^{n-k+1}) = b \quad (2)$$

or

$$F(a^{k+1}b^{n-k-1}) = b. \quad (3)$$

Without loss of generality in our profile $R^{N} = a^{1}b^{n-1}$ let $R^{1} = ab$. If both (2) and (3) are true then there is $j \in N \setminus \{1\}$ such that $F(Q^{N \setminus \{j\}}, \tilde{Q}^{j}) = b$ for $\tilde{Q}^{j}$ with $t(\tilde{Q}^{j}) \neq t(Q^{j})$. This contradicts that $Q^{N}$ is a Nash equilibrium in $(F, a^{1}b^{n-1})$. If only (3) is true then, since $Q^{N}$ is a Nash equilibrium in $(F, a^{1}b^{n-1})$, $Q^{j} = ab$ for all $j \in N \setminus \{1\}$. Then by Lemma A.2, also $Q^{1} = ab$, but $Q^{N} = a^{n}$ is not a Nash equilibrium in $(F, a^{1}b^{n-1})$ again by Lemma A.2. So we have that only (2) is true, which implies, again by the fact that $Q^{N}$ is a Nash equilibrium in $(F, a^{1}b^{n-1})$, that $Q^{j} = ba$ for all $j \in N \setminus \{1\}$. Since $F(Q^{N}) = a$, this implies that $Q^{1} = ab$, hence $Q^{N} = a^{1}b^{n-1}$, and since (3) does not hold, $F(a^{2}, b^{n-2}) = a$. □

**Lemma A.4** $n \neq 3$ and $n \neq 4$.

**Proof** If $n = 3$, then $F(a^{2}b) = b$ by Lemma A.2, and $F(a^{2}b) = a$ by Lemma A.3, contradiction. If $n = 4$, then both $F(a^{2}b^{2}) = a$ and $F(a^{2}b^{2}) = b$ by Lemma A.3, contradiction. □

**Lemma A.5** There is no $k \in \{2, \ldots, n-2\}$ such that $F(a^{k-1}, b^{n-k+1}) = F(a^{k}b^{n-k}) = F(a^{k+1}, b^{n-k-1})$. 

© Springer
Proof This follows from Lemma 2.1(a) by considering a profile $a^k b^{n-k}$.

Lemma A.6 $n \neq 6$.

Proof For $n = 6$, by Lemmas A.2 and A.3 we have $F(a^1 b^5) = F(a^2 b^4) = a$, hence by Lemma A.5, $F(a^3 b^3) \neq a$. Again by Lemmas A.2 and A.3, $F(a^4 b^2) = F(a^5 b^1) = b$, hence by Lemma A.5, $F(a^3 b^3) \neq b$. This contradiction completes the proof.

Lemmas A.1–A.6 provide necessary conditions for a social choice correspondence to be self-implementable. We now drop the assumption that $F$ implements $H$, and proceed with the converse.

Lemma A.7 Let social choice function $F$ be selection from PC satisfying

(i) $F(a^n) = F(a^{1}b^{n-1}) = F(a^2b^{n-2}) = a$,
(ii) $F(b^n) = F(a^{n-1}b^1) = F(a^{n-2}b^2) = b$,
(iii) there is no Nash equilibrium in $(F, R^n)$ with $k \in \{2, \ldots, n - 2\}$.

Then $F$ implements $PC$.

Proof Define the correspondence $H : L^N \to 2^{\{a, b\}}$ by

$$H(R^N) = \{ x \in \{a, b\} \mid x = F(Q^N) \text{ for some Nash equilibrium } Q^N \text{ in } (F, R^N) \}$$

for every $R^N \in L^N$. It is sufficient to prove that $H = PC$.

If $R^N = a^n$ then $F(R^N) = a$ by (i), and $R^N$ is a Nash equilibrium in $(F, R^N)$, so that $a \in H(R^N)$. Suppose that $Q^N \in L^N$ such that $F(Q^N) = b$. By (i), (ii), and (iii) it follows that there is some $i \in N$ and $Q^i \in L$ such that $F(Q^N \setminus \{i\}, Q^i) = a$. Hence $Q^N$ is not a Nash equilibrium in $(F, R^N)$. Therefore, $H(a^n) = \{a\} = PC(a^n)$.

Similarly one proves $H(b^n) = \{b\} = PC(b^n)$.

Now let $1 \leq k \leq n - 1$ and consider a profile $R^N = a^k b^{n-k}$. Let, without loss of generality, $R^i = ab$ for $i = 1, \ldots, k$. Let $Q^N = a^{n-1} b^1$ with $Q^n = ba$. Then $F(Q^N) = b$ by (ii). To show that $Q^N$ is a Nash equilibrium in $(F, R^N)$ it is sufficient to observe that $F(ba, Q^N \setminus \{i\}) = b$ for every $i \in \{1, \ldots, k\}$ by (ii). Analogously, $V^N = a^i b^{n-1}$ with $t(V^1) = a$ is a Nash equilibrium in $(F, R^N)$ with $F(V^N) = a$. Hence, $H(a^k b^{n-k}) = \{a, b\} = PC(a^k b^{n-k})$ for $1 \leq k \leq n - 1$, which completes the proof of the lemma.

Theorem 3.3 now follows from the above results.
material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abdou J, Keiding H (1991) Effectivity functions in social choice. Kluwer Academic Publishers, Dordrecht

Aumann RJ (1959) Acceptable points in general cooperative n-person games. In: Contributions to the theory of games, Vol. IV, Annals of mathematic studies no. 40. Princeton University Press, Princeton

Börgers T (1991) Undominated strategies and coordination in normal form games. Soc Choice Welf 8:65–78

Dasgupta P, Hammond P, Maskin E (1979) The implementation of social choice rules: some general results on incentive compatibility. Rev Econ Stud 46:185–216

Dutta B (1984) Effectivity functions and acceptable game forms. Econometrica 52:1151–1166

Hammond P (1996) Game forms versus social choice rules as models of rights. International Economic Association Series. In: Arrow KJ, Sen A, Suzumura K (eds) Social choice re-examined. Palgrave Macmillan, London

Hurwicz L (1972) On informationally decentralized systems. In: In Mcguire CB, Radner R (eds) Decision and organization. North-Holland, Amsterdam

Hurwicz L, Schmeidler D (1978) Construction of outcome functions guaranteeing existence and Pareto optimality of Nash equilibria. Econometrica 46:1447–1474

Jackson MO (2001) A crash course in implementation theory. Soc Choice Welf 18:655–708

Koray S, Yildiz K (2018) Implementation via rights structures. J Econ Theory 8:161–192

Maskin E (1999) Nash equilibrium and welfare optimality. Rev Econ Stud 66:23–38

Mount K, Reiter S (1974) The informational size of message spaces. J Econ Theory 8:161–192

Mukherjee S, Muto N, Ramaekers E, Sen A (2019) Implementation in undominated strategies by bounded mechanisms: the Pareto correspondence. J Econ Theory 180:229–243

Peleg B (1978) Consistent voting systems. Econometrica 46:153–161

Peleg B (2002) Game-theoretic analysis of voting in committees. Chapter 8. In: Arrow KJ, Sen AK, Suzumura K (eds) Handbook of social choice and welfare, Vol. 1

Peleg B, Peters H (2019) Self-implementation of social choice correspondences in strong equilibrium. In: Trockel W (ed) Social design: essays in memory of Leonid Hurwicz. Studies in economic design. Springer, Berlin Heidelberg, pp 277–291

Peleg B, Winter E (2002) Constitutional implementation. Rev Econ Des 7:187–204

Peleg B, Peters H, Storcken T (2005) Constitutional implementation of social choice correspondences. Int J Game Theory 33:381–396

Saijo T (1988) Strategy space reduction in Maskin’s theorem: sufficient conditions for Nash implementation. Econometrica 56:693–700

Saijo T, Tatamitani I, Yamato T (1996) Toward natural implementation. Int Econ Rev 37:949–980

Sen AK (1970) The impossibility of a Paretian liberal. J Political Econ 78:152–157

Sertel RM (2001) Designing rights: invisible hand theorems, covering and membership. Bogazici University, Mimeo

Tatamitani Y (2001) Implementation by self-relevant mechanisms. J Math Econ 35:427–444

Xiong S (2021) Designing referenda: an economist’s pessimistic perspective. J Econ Theory 191:105–133

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.