

de Clippely, Geoffroy

Working Paper

Behavioral implementation

Working Paper, No. 2012-6

Provided in Cooperation with:

Department of Economics, Brown University

Suggested Citation: de Clippely, Geoffroy (2012) : Behavioral implementation, Working Paper, No. 2012-6, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/62630>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Behavioral Implementation: Supplemental Appendix

(Not for Publication)

Geoffroy de Clippel*

This Version: June 2012

1 Two Natural Extensions of Maskin Monotonicity, and Why They Are Inappropriate

A first natural idea to extend Maskin monotonicity into a property that would be applicable on any domain of individual choice functions is to use Bernheim and Rangel's (2009) extended revealed preference, as defined at the beginning of Section 4 of the main paper. For each individual i , each state θ , and each option x , the lower contour set $LC_i(x, \theta)$ of x given θ is now the set of options y such that x unambiguously dominates y given $C_i(\cdot, \theta)$. The property of Maskin monotonicity can then be reproduced as stated on page 11 of the main article. Unfortunately, a look at the SCR f_{ij} from Proposition 7 in the main text shows that Nash-implementable SCRs may violate this extension of Maskin monotonicity. To see this, consider $\theta \in \Theta$ such that $\theta_i \in A_i$ and $\theta_j \in \Theta_j \setminus A_j$ ranks x as the best element of Z . In that case, $f_{ij}(\theta) = x$, and the extended lower contour set of x for i is empty. Changing i 's type to $\theta'_i \in \Theta_i \setminus A_i$ such that x is second best expands that extended lower contour set to a non-empty set. The extension of Maskin monotonicity would then require f_{ij} to pick x at

*Brown University, Department of Economics, Providence, Rhode Island - declippel@brown.edu.

(θ'_i, θ_{-i}) if it was a necessary condition for Nash implementability, since f_{ij} is implementable. On the contrary, $f_{ij}(\theta'_i, \theta_{-i})$ is the element that is top ranked for i instead of x .

Maskin monotonicity means that the option that is selected at θ should remain selected if it becomes more appealing in the individual preference orderings. Here is a natural choice-based extension of this idea. An option $x \in X$ is *more appealing* according to C' than to C if $C'(S) = x$, for each S such that $C(S) = x$.

Choice-Based Monotonicity (CBM) *Let $\theta, \theta' \in \Theta$. If $f(\theta)$ is more appealing according to $C_i(\cdot, \theta'_i)$ than $C_i(\cdot, \theta_i)$, for each $i \in I$, then $f(\theta') = f(\theta)$.*

The property is in fact not new, as it first appeared in Aizerman and Aleskerov (1986) in their study of the aggregation of individual choice functions into social choice functions. It is not difficult to check that any social choice function that is Nash implementable must satisfy CBM (either directly, or as an immediate corollary to Proposition 1 in the main paper). While an interesting necessary condition, CBM turns out to be too weak, as it does not guarantee Nash implementability when combined with Property N. Here is a simple example to illustrate this fact.

Example 1 *Let $X = \{a, b, c\}$, let $C_1(\cdot, \theta_1)$ be the rational choice function associated to the ordering $a \succ b \succ c$, let $C_2(\cdot, \theta_2)$ be the rational choice function associated to the ordering $b \succ c \succ a$, let $C_3(\cdot, \theta_3)$ be the rational choice function associated to the ordering $c \succ b \succ a$, let $C_1(\cdot, \theta'_1)$ be the choice function that coincides with $C_1(\cdot, \theta_1)$ except that $C_1(\{a, b\}, \theta'_1) = b$, and let $C_1(\cdot, \theta''_1)$ be the choice function that coincides with $C_1(\cdot, \theta_1)$ except that $C_1(\{a, c\}, \theta''_1) = c$. It is then easy to see that the social choice function f , defined by $f(\theta_1, \theta_2, \theta_3) = a$, $f(\theta'_1, \theta_2, \theta_3) = b$, and $f(\theta''_1, \theta_2, \theta_3) = c$, satisfies CBM and N. Yet it is not Nash implementable, as a consequence of Proposition 1. Indeed I now show that f violates Property M. Given that $f(\theta) = a$, and given the definition of $C_1(\cdot, \theta_1)$, it must be that $X_1(\theta) = \{a, b, c\}$, or $\{a, b\}$, or $\{a, c\}$, or $\{a\}$. Notice that $C_1(\{a, b, c\}, \theta'_1) = a$, $C_1(\{a, b\}, \theta'_1) = a$, $C_1(\{a, c\}, \theta'_1) = a$, and $C_1(\{a\}, \theta''_1) = a$. For Property M to be satisfied, $f(\theta'_1, \theta_2, \theta_3)$ or $f(\theta''_1, \theta_2, \theta_3)$ should have been equal to a .*

2 Proof of Proposition 1

For necessity, let (M, μ) be a mechanism that implements f , let $\theta \in \Theta$, and let $m^*(\theta)$ be a Nash equilibrium of the game induced by (M, μ) at θ . Let then $X_i(\theta) = \{\mu(m_i, m_{-i}^*(\theta)) | m_i \in M_i\}$, for each $i \in I$. By definition of implementation and of Nash equilibrium, we have: $f(\theta) = \mu(m^*(\theta)) = C_i(X_i(\theta), \theta)$, for all i and all θ , which proves the first part of property M. Suppose now that there is a θ' such that $C_i(X_i(\theta), \theta') = f(\theta)$, for all $i \in I$. Then $m^*(\theta)$ forms a Nash equilibrium of the game induced by (M, μ) at θ' , and hence $f(\theta') = \mu(m^*(\theta)) = f(\theta)$, by definition of Nash-implementability, which proves the second part of property M.

As for sufficiency, consider the set of messages $M_i = X \times \Theta \times \mathbb{Z}_+$, for each $i \in I$, and $\mu : M \rightarrow X$ defined as follows:

1. If $m_i = (f(\theta), \theta, 0)$, for each i , then $\mu(m) = f(\theta)$.
2. If there is $j \in I$ such that $m_i = (f(\theta), \theta, 0)$, for each $i \in I \setminus \{j\}$, and $m_j = (x, \theta', \alpha) \neq (f(\theta), \theta, 0)$, then $\mu(m) = x$ if $x \in X_i(\theta)$, and $\mu(m) = f(\theta)$ otherwise.
3. In all other cases, $\mu(m) = x$, where x is the first component in the report of the individual with the lowest index among those who submit the highest integer.

First notice that the strategy profile $m^*(\theta) = (f(\theta), \theta, 0)$ forms a Nash equilibrium of the game induced by (M, μ) at θ , following the first condition in Property M.

Second, consider a strategy profile m that forms an equilibrium in θ , and let x be the resulting outcome. Suppose first that $m_i = (f(\theta'), \theta', 0)$, for each i . Then $x = f(\theta')$, and the condition of Nash equilibrium imply that $C_i(X_i(\theta'), \theta) = f(\theta')$, for each i . The second condition in Property M implies that $f(\theta') = f(\theta)$. Transitivity thus implies that $x = f(\theta)$, as desired. Suppose next that there exists $j \in I$ such that $m_i = (f(\theta'), \theta', 0)$, for each $i \in I \setminus \{j\}$, and $m_j = (x, \theta'', \alpha) \neq (f(\theta'), \theta', 0)$. The conditions for Nash equilibrium imply that $C_j(X_j(\theta'), \theta) = x$ and $C_i(X, \theta) = x$, for all $i \in I \setminus \{j\}$. The additional

condition that distinguishes \mathfrak{M} from M implies that $x = f(\theta)$, as desired. Finally, in all other cases, the conditions for Nash equilibrium imply that $C_i(X, \theta) = x$, for each $i \in I$, and the additional condition that distinguishes \mathfrak{M} from M implies that $x = f(\theta)$, as desired. ■

3 Additional Results on Nash Implementation

Here are two related characterization results of Nash implementability, one for social choice rules that need not be single-valued, and one for environments with private consumption.

Proposition 1' (NECESSARY, AND SUFFICIENT CONDITIONS FOR NASH-IMPLEMENTABILITY OF MULTI-VALUED SOCIAL CHOICE RULES) *If a social choice rule F is Nash-implementable, then there exists a function $X_i : \{(x, \theta) \in X \times \Theta | x \in F(\theta)\} \rightarrow P(X)$, for each $i \in I$, such that:*

1. $C_i(X_i(x, \theta), \theta) = x$, for each $x \in F(\theta)$ and each $\theta \in \Theta$.
2. For all θ, θ' , and $x \in F(\theta)$, if $C_i(X_i(x, \theta), \theta') = x$, for each $i \in I$, then $x \in F(\theta')$.

If, in addition, $x \in F(\theta)$ for any θ and any x for which there exists i such that $C_j(X, \theta) = x$ for all $j \in I \setminus \{i\}$, then F is Nash-implementable.

The proof is virtually identical to that of Proposition 1, and is left to the reader.

The next result applies to the house allocation problem (see Section 5 of the paper), where consumption is private.

Proposition 1'' (NECESSARY, AND SUFFICIENT CONDITIONS FOR NASH-IMPLEMENTABILITY OF SOCIAL CHOICE RULES IN THE HOUSE ALLOCATION PROBLEM) *If a social choice rule $F : \Theta \rightarrow O^I$ is Nash-implementable, then there exists a function $X_i : \{(x, \theta) \in X \times \Theta | x \in F(\theta)\} \rightarrow P(O)$, for each $i \in I$, such that:*

1. $C_i(X_i(o, \theta), \theta) = o_i$, for each $o \in F(\theta)$ and each $\theta \in \Theta$.
2. For all θ, θ' , and all $o \in F(\theta)$, if $C_i(X_i(o, \theta), \theta') = o_i$, for each $i \in I$, then $o \in F(\theta')$.

If, in addition, $o \in F(\theta)$ for any θ and any o for which there exists i and o' such that $C_j(O, \theta) = o_j$, for all $j \in I \setminus \{i\}$, and either $C_i(O, \theta) = o_i$ or $C_i(X_i(o', \theta), \theta) = o_i$, for some $o' \in F(\theta)$, then F is Nash-implementable.

Again, the proof is virtually identical to that of Proposition 1, and is left to the reader.

4 More on the Relation Between P and M

P implies M on domains that contain only individual choice functions that satisfy IIA (see Theorem 3.2.1 in Dasgupta et al. (1979)). The next example shows that this implication need not hold when considering richer domains that may contain individual choice functions that violate IIA.

Example 2 Let $X = \{a, b, c, d\}$, $I = \{1, 2\}$, and $\bar{\Theta}_i$ be the set of all strict rational preferences on X such that d is first or second-best. Suppose that $C_i(S, \theta_i) = \arg \max_{\succ(\theta_i)} A_i(S)$, where $A_i(S)$ contains the first two elements of S according to the alphabetical order.¹ These choice procedures are consistent with the model described in Section 7 of the main paper and, more specifically, are rationalizable in the sense of Cherepanov et al. (2009), Lleras et al. (2010), Manzini and Mariotti (2012), and Masatlioglu et al. (2012). Consider the social choice function $f : \bar{\Theta} \rightarrow X$ defined as follows:

$$f(\theta) = \begin{cases} a & \text{if } a \succ(\theta_1)x \text{ and } a \succ(\theta_2)x, \forall x \in \{b, c, d\} \\ b & \text{if } b \succ(\theta_1)x \text{ and } b \succ(\theta_2)x, \forall x \in \{a, c, d\} \\ c & \text{if } c \succ(\theta_1)x \text{ and } c \succ(\theta_2)x, \forall x \in \{a, b, c\} \\ d & \text{otherwise.} \end{cases}$$

In other words, f picks d except if an alternative is top best for both individuals. It is easy to check that f satisfies Property P (and is implementable in dominant strategies, thanks to the sufficient condition in Proposition 3 of the

¹Notice that $C_i(\cdot, \theta_i) \neq C_i(\cdot, \theta'_i)$, for all $\theta_i, \theta'_i \in \bar{\Theta}_i$ such that $\theta_i \neq \theta'_i$.

main paper) with

$$X_i(\theta_{-i}) = \begin{cases} \{a, d\} & \text{if } a \succ (\theta_{-i})x, \forall x \in \{b, c, d\} \\ \{b, d\} & \text{if } b \succ (\theta_{-i})x, \forall x \in \{a, c, d\} \\ \{c, d\} & \text{if } c \succ (\theta_{-i})x, \forall x \in \{a, b, c\} \\ \{d\} & \text{otherwise.} \end{cases}$$

On the other hand, f violates Property M (and hence is not Nash-implementable). Let θ be such that d is ranked top for both individuals. The first condition in Property M implies that $C_i(X_i(\theta), \theta_i) = d$, for both $i = 1, 2$. Hence i pays attention to d , and $X_i^*(\theta)$ contains at most two elements: $X_1(\theta) = \{x, d\}$ and $X_2(\theta) = \{y, d\}$, for some $x, y \in \{a, b, c, d\}$. If θ' is such that $z \in X \setminus \{d, x, y\}$ is top-ranked and d is second-best, for both individuals, then $C_i(X_i(\theta), \theta'_i) = d$, for both $i = 1, 2$, which would contradict the second condition in Property M , given that $f(\theta') = z$. Hence f violates Property M , and is not Nash-implementable.

5 Pareto Correspondence for Choices with Limited Consideration

In this section I focus on the model of choice with fixed consideration sets described in Section 7.1 of the main paper. This model imposes a structure on the violations of rationality. Indeed, individuals are assumed to have a standard preference, but appear irrational because they maximize it only over the subset of feasible options they actively consider. Irrationality is thus understood as a mistake, and a paternalistic mechanism designer would find the normative criterion of Pareto efficiency in terms of the true underlying preferences more appealing than the rationale behind F^{BRE} or F^{Eff} (arguments along these lines were first provided by Masatlioglu et al. (2012) and Rubinstein and Salant (2012)).

The definition of such a criterion is not necessarily unequivocally defined, because a same choice function may be associated to multiple underlying preferences. For each $\theta_i \in \Theta_i$, let $\rho(\theta_i)$ be the set of θ_i^* such that $C_i(S, \theta_i) =$

$\arg \max_{\succ_{(\theta_i^*)}} S$, for all $S \in P(X)$. Notice that ρ_i is non-empty valued by condition (2) from the main paper. Condition (3) from the main paper implies, in addition, that ρ_i is single-valued if and only if, for each $x, y \in X$, there exists $S \in P(X)$ such that $A_i(S) = \{x, y\}$. For the necessary condition, suppose that x, y are such that $A_i(S) \neq \{x, y\}$, for all $S \in P(X)$. Consider then two rational types θ_i^* and $\hat{\theta}_i^*$ such that x and y are ranked below any other element of X , y is below x in θ_i^* , and above it in $\hat{\theta}_i^*$. It is easy to check that the choice functions generated by θ_i^* and $\hat{\theta}_i^*$ given A_i are identical, and hence we have found a type in Θ_i for which ρ_i is multi-valued. For the sufficient condition, consider two rational types $\theta_i^* \neq \hat{\theta}_i^*$. Then there exist x and y such that $x \succ_{\theta_i^*} y$ and $y \succ_{\hat{\theta}_i^*} x$. Then the associated choice functions must take different values at any S such that $A_i(S) = \{x, y\}$, and ρ_i must be single-valued.

So, to avoid unnecessary conceptual issues associated to the possibility of multiple underlying preferences, I will assume until the end of this section that for each $x, y \in X$, there exists $S \in P(X)$ such that $A_i(S) = \{x, y\}$. Natural models of choice with limited attention verify this assumption. For instance, individuals may pay attention to all options when few are available, but consider only strict subsets of feasible alternatives when many are available. Under this assumption, the Pareto correspondence can unequivocally be defined as follows:

$$F^{Pareto}(\theta) = \{x \in X \mid x \text{ is Pareto efficient given } \rho(\theta)\},$$

where $\rho(\theta) = (\rho_i(\theta_i))_{i \in I}$, for each $\theta \in \Theta$. It is easy to check (see arguments along these lines in Salant and Rubinstein (2008) and Masatlioglu et al. (2012)) that there is no systematic inclusion between F^{Pareto} , on the one hand, and F^{BRE} or F^{Eff} , on the other hand, except of course that they all coincide with the classic notion of Pareto efficiency on the rational domain.

The next result is negative: F^{Pareto} is implementable if and only if all participants are rational.

Proposition 12 *F^{Pareto} is Nash-implementable if and only if $A_i(S) = S$, for each S, i (or $C_i(\cdot, \theta_i)$ is rational, for all i and all θ_i).*

Proof: It is well-known that the Pareto correspondence is Nash-implementable

on the domain of rational choice functions, and hence we will focus on the necessary condition. Suppose that there exist $S \in P(X)$ and $i \in I$ such that $A_i(S) \subsetneq S$. Let S^* be a set with that property for i , and such that no larger set has that property for i (i.e. $A_i(S) = S$ for any set S larger than S^*). Let then $x \in S^* \setminus A_i(S^*)$, and consider $\theta_i^* \in \Theta_i^*$ such that options in $X \setminus S^*$ are ranked above all others, and x is ranked top among remaining options. Suppose also that the ranking associated to θ_j^* is strictly opposite to the ranking associated to θ_i^* , for all $j \in I \setminus \{i\}$. Let $\theta \in \Theta$ be such that $\rho(\theta) = \theta^*$. Observe that $x \in F^{Pareto}(\theta)$. Consider now the set $X_i(x, \theta)$ as given in Proposition 1'. $C_{\theta_i}(X_i(x, \theta)) = x$ implies that $A_i(X_i(x, \theta)) \subseteq S^*$. If $A_i(X_i(x, \theta)) = S^*$, then $X_i(x, \theta)$ is strictly larger than S^* , thereby contradicting the maximality of S^* . Hence $A_i(X_i(x, \theta)) \subsetneq S^*$. Let then $x' \in S^* \setminus A_i(X_i(x, \theta))$, and consider the rational type $\hat{\theta}_i^* \in \Theta_i^*$ that differs from θ_i^* only in that x' is now ranked top among all options in X . Let $\hat{\theta} \in \Theta$ be such that $\rho(\hat{\theta}) = \hat{\theta}^*$. The second part of the necessary condition in Proposition 1' implies that x should also belong to $F^{Pareto}(\hat{\theta}_i, \theta_{-i})$ if it were Nash-implementable, but it doesn't since x' Pareto dominates x . Hence it must be that $A_i(S) = S$, for each $i \in I$ and each $S \in P(X)$, to avoid the contradiction, and we are done proving the result. ■

6 Proof of Proposition 5

I start by showing that F^{BRC} is not Nash-implementable. Consider a type profile θ where each individual $i \neq 1$'s choice function is derived by maximizing the ordering where o_1^* is most preferred, o_{i+1}^* is second most preferred, o_{i+2}^* is third most preferred, etc., with the convention that $I + 1 = 2$. The first individual's choice function at that state is derived by maximizing the ordering where o_2^* is most preferred, o_3^* is second most preferred, etc., and o_1^* is least preferred, except that $C_1(\{o_1^*, o_i^*\}, \theta_1) = o_1^*$, for each $i > 1$. The allocation \bar{o} where the first individual keeps his endowment, and each subsequent individual gets the endowment of his follower (with I getting 2's endowment) belongs to the $F^{BRC}(\theta)$. Indeed, no object unambiguously dominates the endowment for the first individual, and hence a blocking coalition cannot contain 1. Notice

then that \bar{o} is such that each individual other than the first receives his most preferred object in $O \setminus \{o_1\}$, and hence it is impossible to find a blocking coalition that does not contain 1 either, which shows indeed that $\bar{o} \in F^{BRC}(\theta)$. By definition of θ , the first part of the necessary condition in Proposition 1'' can be satisfied only if $\bar{o}_i \in X_i(\bar{o}, \theta)$ and $o_1^* \notin X_i(\bar{o}, \theta)$, for all $i \neq 1$, and $X_1(\bar{o}_1, \theta) = \{\bar{o}_1\}$, or $\{\bar{o}_1, o_i^*\}$, for some $i \geq 2$. If $X_1(\bar{o}_1, \theta) = \{\bar{o}_1\}$, then the second part of the necessary condition implies that $\bar{o} \in F^{BRC}(\theta'_1, \theta_{-1})$, where θ'_1 is the rational choice function derived from θ_1 by changing the first individual's choice over pairs that contain o_1^* . To see that is impossible, observe that $F^{BRC}(\theta'_1, \theta_{-1})$ coincides with the regular core since all individual choice functions are rational at that type profile, and hence the core coincides with the outcome of the top-trading cycle procedure, which will have 1 and 2 achieve their most-preferred choice by trading their objects. In the case where $X_1(\bar{o}_1, \theta) = \{\bar{o}_1, o_i\}$, for some $i \geq 2$, consider $j \in I \setminus \{1, i\}$, and the type θ'_1 that differs from θ_1 only in that $C_1(\{o_1^*, o_j^*\}, \theta'_1) = o_j^*$ instead of o_1^* . The second part of the necessary condition in Proposition 1'' implies that $\bar{o} \in F^{BRC}(\theta'_1, \theta_{-1})$ if it were Nash-implementable, but it doesn't, as $\{1, j\}$ can unambiguously improve upon \bar{o} by trading with each others.

I now show that F^C is Nash implementable on any Θ . By definition of F^C , one can associate to any $x \in F^C(\theta)$ a collection $(T_i^{x, \theta})_{i \in I}$ of subsets of O such that $C_i(T_i^{x, \theta}, \theta_i) = x_i$, for each $i \in I$, and for all coalition S and all allocation α that is feasible for S , there exists a member i of S such that $\alpha_i \in T_i^{x, \theta}$. It is easy to check that the conditions 1. and 2. given in Proposition 1'' are satisfied for $X_i(x, \theta) = T_i^{x, \theta}$. Suppose now that θ and $i \in I$ are such that $C_j(O, \theta_j) = o_j$, for each $j \in I \setminus \{i\}$. If $C_i(O, \theta_i) = o_i$ as well, then one concludes that $o \in F^C(\theta)$ by taking $T_j = O$, for all $j \in I$. If $C_i(X_i(o', \theta), \theta_i) = o_i$, for some $o' \in F^C(\theta)$, then one concludes that $o \in F^C(\theta)$ by taking $T_j = O$, for all $j \in I \setminus \{i\}$ and $T_i = X_i(o', \theta)$ (notice that $o_i^* \in T_i$, by definition of $X_i(o', \theta)$).

Consider a type profile θ , $x \in F^C(\theta)$, a group S , and an allocation $o' \in \mathcal{F}(S)$. Hence there exists $i \in S$ such that $o'_i \in T_i$ and $C_i(T_i, \theta_i) = x_i$. It is thus impossible to find $o' \in \mathcal{F}(S)$ such that o'_i unambiguously dominates x_i , for all $i \in S$, and $x \in F^{BRC}(\theta)$. This establishes that $F^C \subseteq F^{BRC}$.

Finally, if θ is such that $C_i(\cdot, \theta_i)$ is rational, for each $i \in I$, and x belongs to the core for the associated revealed preferences, then define T_i as the lower contour set of x according to the revealed preference, for each i . By definition, $C_i(T_i, \theta_i) = x$, for all i . If there is S and $x' \in \mathcal{F}(S)$ such that $x'_i \in O \setminus T_i$, for each $i \in S$, then x'_i is revealed preferred to x_i for all $i \in S$, thereby contradicting the fact that x belongs to the core. This shows that F^C contains the core. The other inclusion follows from the fact that $F^C \subseteq F^{BRC}$, and F^{BRC} coincides with the core when individual choice functions satisfy IIA. ■

7 Core vs. Top Trading Cycle

This section follows the discussion from Section 5 in the main paper. If individual choice functions satisfy IIA, then F^C coincides with the core, and hence the allocation resulting from the top trading cycle is the only element in the core (which also coincides with the unique competitive equilibrium). Here is an example to show that the outcome of the top trading cycle, $\alpha^*(\theta)$, need not be the only element of $F^C(\theta)$ for states associated to individual choice functions that may violate IIA.

Example 3 *Suppose that $I = \{1, 2, 3\}$, and consider a type profile θ so that $C_1(O, \theta_1) = o_3^*$, $C_2(O, \theta_2) = o_2^*$, $C_3(O, \theta_3) = o_3^*$, $C_2(\{o_1^*, o_2^*\}, \theta_2) = o_1^*$, and $C_3(\{o_2^*, o_3^*\}, \theta_3) = o_2^*$. The top trading cycle leaves individuals with their initial endowments. Consider the alternative allocation $\beta = (o_3^*, o_1^*, o_2^*)$. It also belongs to $F^C(\theta)$. This can be proved using $T_1 = O$, $T_2 = \{o_1^*, o_2^*\}$ and $T_3 = \{o_2^*, o_3^*\}$. Notice indeed that $C_i(T_i, \theta_i) = \beta_i$, by definition of T_i , θ and β . To conclude the argument, one must show that, for all $S \in P(I)$ and all $\alpha \in \mathcal{F}(S)$, there exists $i \in S$ such that $\alpha_i \in T_i$. This is trivially true if $1 \in S$, since $T_1 = O$. If $1 \notin S$, then the feasibility of α implies that $\alpha_3 \in O \setminus \{o_1^*\}$. Hence the property is true by choosing $i = 3$ if he is a member of S . Finally, if neither 1 nor 3 are members of S , then it must be that $S = \{2\}$ and $\mathcal{F}(S) = \{o_2^*\} \subseteq T_2$, which establishes the property and the fact that $\beta \in F^C(\theta)$, as desired.*

8 f_{ij} is Nash implementable

This section follows the discussion from Section 6 in the main paper. To save on notations, f_{ij} will be denoted by f in this section. Let $\sigma : Z \rightarrow Z$ be an isomorphism such that $\sigma(z) \neq z$, for all $z \in Z$. Define $X_i, X_j : \Theta \rightarrow P(X)$ as follows:

$$X_i(\theta) = \begin{cases} Z \cup \{\gamma(\sigma(f(\theta)))\} & \text{if } \theta_i \in \Theta_i \setminus A_i \\ Z \cup \{\gamma(f(\theta))\} & \text{if } \theta_i \in A_i, \end{cases}$$

$$X_j(\theta) = \begin{cases} \{f(\theta)\} & \text{if } \theta_i \in \Theta_i \setminus A_i \\ Z & \text{if } \theta_i \in A_i, \end{cases}$$

and let $X_k(\theta) = \{f(\theta)\}$ for all θ and all $k \neq i, j$.

As for the mechanism used in the proof of Proposition 1, consider the set of messages $M_k = X \times \Theta \times \mathbb{Z}_+$, for each $k \in I$. The outcome function, $\mu : M \rightarrow X$, is a bit different though:²

1. If $m_k = (f(\theta), \theta, 0)$, for each k , then $\mu(m) = f(\theta)$.
2. If there is $k \in I \setminus \{j\}$ such that $m_l = (f(\theta), \theta, 0)$, for each $l \in I \setminus \{k\}$, and $m_k = (x, \theta', \alpha) \neq (f(\theta), \theta, 0)$, then $\mu(m) = x$ if $x \in X_k(\theta)$, and $\mu(m) = f(\theta)$ otherwise. If $m_l = (f(\theta), \theta, 0)$, for each $l \in I \setminus \{j\}$, and $m_j = (x, \theta', \alpha) \neq (f(\theta), \theta, 0)$, then $\mu(m) = x$ if $x \in X_j(\theta)$, and $\mu(m) = c$ otherwise (cf. definition of c in the main paper).
3. In all other cases, let k be the individual between i and j who submit the highest integer (say $k = i$ if both integers are equal). Let z be an arbitrary fixed element of Z . If $k = i$ and the first component of i 's message, call it x , belongs to $Z \cup \{\gamma(\sigma(y))\}$ (resp. Z if $y \in X \setminus Z$), where y is the first component of j 's message, then $\mu(m) = x$. If x does not belong to $Z \cup \{\gamma(\sigma(y))\}$ (resp. Z if $y \in X \setminus Z$), then $\mu(m) = z$. If $k = j$ and the first component of j 's message, call it x again, belongs to Z , then $\mu(m) = x$. If x does not belong to Z , then $\mu(m) = z$.

²While Property M is satisfied for this family $(X_k)_{k \in I}$, the extra condition that distinguishes \mathfrak{M} from M is not always satisfied.

I start by checking that the strategy profile $m^*(\theta) = (f(\theta), \theta, 0)$ (whose associated outcome is $f(\theta)$) forms a Nash equilibrium of the game induced by (M, μ) at θ . The set of options that individual i can reach by changing his message is $X_i(\theta)$, and it is easy to check that $C_i(X_i(\theta), \theta_i) = f(\theta)$, by definition of X_i . The set of options that individual j can reach by changing his message is $X_j(\theta) \cup \{c\}$, and again it is easy to check that $C_j(X_j(\theta) \cup \{c\}, \theta_j) = f(\theta)$, by definition of X_j . For all $k \neq i, j$, $C_k(X_k(\theta), \theta_k) = f(\theta)$, since $X_k(\theta) = \{f(\theta)\}$.

Next, consider a strategy profile m that forms an equilibrium in θ' , and let x be the resulting outcome. We have to prove that $x = f(\theta')$. Notice that for any m , individual i faces a set of options S that contains Z , plus in some cases an option in $\gamma(Z)$. If $\theta'_i \in \Theta_i \setminus A_i$, then

$$f(\theta') = C_i(Z, \theta'_i) = C_i(S, \theta'_i) = x,$$

where the first equality follows from the definition of f , the second equality follows from the fact that S equals Z plus (in some cases) an element of $\gamma(Z)$, out of which a individual who is not subject to the attraction effect picks the same element as when choosing in Z , and the third equality follows from the Nash equilibrium condition for i at θ' . Hence $x = f(\theta')$, as desired, and from now on I assume that $\theta'_i \in A_i$.

Consider the case where $m_k = (f(\theta), \theta, 0)$, for each k . Then $x = f(\theta)$, and the condition of Nash equilibrium implies in particular that $C_i(X_i(\theta), \theta'_i) = f(\theta)$. This is impossible if $\theta'_i \in \Theta_i \setminus A_i$, as $C_i(X_i(\theta), \theta'_i) = \sigma(f(\theta)) \neq f(\theta)$ in that case. Hence $\theta'_i \in A_i$, which implies $X_j(\theta) = Z$. The Nash equilibrium condition for j at θ' is $C_j(Z \cup \{c\}, \theta'_j) = f(\theta)$. Given that $\theta'_i \in A_i$, it must be that $f(\theta') = C_j(Z, \theta'_j)$, and hence $x = f(\theta) = f(\theta')$, as desired.

Consider now the case where $m_l = (f(\theta), \theta, 0)$, for each $l \in I \setminus \{i\}$, and $m_i = (y, \hat{\theta}, \alpha) \neq (f(\theta), \theta, 0)$. Given that $\theta'_i \in A_i$, it must be $f(\theta') = C_j(Z, \theta'_j)$. Notice that the condition of Nash equilibrium for j at θ' requires $x = C_j(Z \cup \{c\}, \theta'_j)$. Hence $x = f(\theta')$, as desired.

Consider the case where $m_l = (f(\theta), \theta, 0)$, for each $l \in I \setminus \{j\}$, and $m_j = (y, \hat{\theta}, \alpha) \neq (f(\theta), \theta, 0)$. The condition of Nash equilibrium for j at θ' implies

that $y \in X_j(\theta)$. Hence $x = y \in Z$. The Nash equilibrium condition for i at θ' requires $x = \sigma(y)$ (remember that $\theta'_i \in A_i$). Hence there is no Nash equilibrium in this case.

Let $k \geq 3$, and consider the case where $m_l = (f(\theta), \theta, 0)$, for each $l \in I \setminus \{k\}$, and $m_k = (y, \hat{\theta}, \alpha) \neq (f(\theta), \theta, 0)$. The equilibrium outcome (x) in that case is equal to $f(\theta)$. Yet, the Nash equilibrium condition for i at θ' requires $x = \sigma(f(\theta))$ (remember that $\theta'_i \in A_i$). Hence there is no Nash equilibrium in this case either.

Finally, consider any combination of messages that would form a Nash equilibrium at θ' that we have not considered yet. The Nash equilibrium condition for j at θ' requires $x = y = C_j(Z, \theta'_j)$. The Nash equilibrium condition for i at θ' requires $x = \sigma(y)$ (remember that $\theta'_i \in A_i$). Hence it is not possible to find a Nash equilibrium in that case either. ■

9 Endogenous Frames and Backward Induction

This section follows the discussion from Section 8 in the main paper. The first individual is assumed to be “fully” rational (see footnote 16 in the main paper), while there is no restriction on the second participant’s behavior. I restrict attention to two-stage mechanisms of perfect information where the mechanism designer systematically delegates the choice of the frame to the first individual (instead of picking the frame himself). Formally, a *two-stage mechanism* is composed by a finite set M_1 of messages for the first individual, a finite set $M_2(m_1)$ of messages for the second individual, for each $m_1 \in M_1$, and a function μ that associates an outcome in X to each pair of messages. The extensive-form associated with this mechanism starts with the first individual choosing an action $(m_1, \phi) \in M_1 \times \Phi$, and ends with the second individual choosing afterwards an action in $m_2 \in M_2(m_1)$. The outcome is then $\mu(m_1, m_2)$. The *backward induction equilibrium* at θ of such a mechanism is a couple $(m_1^*(\theta), m_2^*(\theta))$ of messages and a frame $\phi^*(\theta)$ such that

$$\mu(m_1^*(\theta), m_2^*(\theta)) = C_2^{\phi^*(\theta)}(\{\mu(m_1^*(\theta), m_2) | m_2 \in M_2(m_1^*(\theta))\}, \theta)$$

$$\mu(m_1^*(\theta), m_2^*(\theta)) = C_1(C_2^\phi(\{\mu(m_1, m_2)|m_2 \in M_2(m_1)\}, \theta)|m_1 \in M_1, \phi \in \Phi), \theta),$$

for all $\theta \in \Theta$. The social choice function $f : \Theta_1 \times \Theta_2 \rightarrow X$ is *implementable by backward induction* if there exists a mechanism $(M_1, M_2(\cdot), \mu)$ such that, for each θ , the backward induction equilibrium $(m_1^*(\theta), m_2^*(\theta), \phi^*(\theta))$ is such that $f(\theta) = \mu(m_1^*(\theta), m_2^*(\theta))$. It is easy to characterize the set of social choice functions that are implementable in this sense.

Proposition 13 *The social choice function f is implementable by backward induction if and only if there exists a sequence $(S_k)_{k=1}^K$ of non-empty subsets of X such that $f(\theta) = C_1(X_1(\theta), \theta)$, where $X_1(\theta) = \{C_2^\phi(S_k, \theta)|k = 1, \dots, K, \phi \in \Phi\}$, for each $\theta \in \Theta$.*

Proof: To establish the sufficient condition, let $M_1 = \{m_1, \dots, m_K\}$. Let $M_2(m_k)$ be a set of $|S_k|$ messages, and μ be such that $\{\mu(m_k, m_2)|m_2 \in M_2(m_k)\} = S_k$, for each k . It is easy to check that the resulting two-stage extensive-form game implements f by backward induction.

As for the necessary condition, let f be a social choice function that is implementable by backward induction via some mechanism $(M_1, M_2(\cdot), \mu)$. Let K be the total number of messages in M_1 , and enumerate them so that $M_1 = \{m_1, \dots, m_K\}$. Let then $S_k = \{\mu(m_k, m')|m' \in M_2(m_k)\}$, for each k . The two conditions for $(M_1, M_2(\cdot), \mu)$ to implement f then imply that $f(\theta) = C_1(X_1(\theta), \theta)$, where $X_1(\theta) = \{C_2(S_k, \theta)|k = 1, \dots, K, \phi \in \Phi\}$, for each $\theta \in \Theta$, as desired. ■

The following example illustrates how implementation by backward induction allows for frames to vary with types, which was impossible when considering implementation in Nash equilibrium or in dominant strategies.

Example 4 *Consider the following variant of the problem studied by de Clippel et al. (2011) for the selection of arbitrators. There are five candidates: $X = \{a, b, c, d, e\}$. Types encode strict preference orderings for both parties, but while the first party chooses by maximizing that ordering, candidates are always presented in a list, and the second individual makes his choice by applying Rubinstein and Salant's (2006) "Stop When You Start To Decline" procedure introduced in Section 7.2. Here is a variant with frames of de Clippel et al.'s*

(2011) shortlisting mechanism. The first party can choose an order for the candidates, and any one subset of X that contains three candidates. Candidates in the subset selected by the first individual are then presented in a list that is consistent with the order chosen by the first individual, and the second individual is free to appoint any candidate from that list to rule the case. The mechanism implements the social choice function f , where $f(\theta) = C_1(X_1(\theta), \theta_1)$, with $X_1(\theta) = \{C_2^\phi(S, \theta_2) | S \in \Sigma, \phi \in \Phi\}$, for each $\theta \in \Theta$, where Σ is the set of subsets of X that contains three candidates, and Φ is the set of lists over X . It is easy to check that this social choice function picks the first individual's top candidate whenever that candidate is not the worse option for the second party, and the first individual's second-best candidate otherwise. At equilibrium, frames change with the second individual's type: the first individual's top candidate appears first on the shortlist and the worst candidate for the second individual appears second, when the first individual's top candidate is not the worst for the second party, while the first individual's second-best candidate appears first on the shortlist and his most preferred candidate appears second, when the first party's top candidate is the worst for the other individual. It is also easy to check that there is no frame that the mechanism designer can pick in order to implement this social choice function either in Nash equilibrium (because it violates Property M for any list), in dominant strategies (because it violates Property P for any list), or by backward induction (there is no way to satisfy the condition of Proposition 13 when ϕ is fixed for all θ in the definition of $X_1(\theta)$ instead of being a variable).

Additional References

- de Clippel, G., K. Eliaz, and B. Knight**, 2011. On the Selection of Arbitrators. Brown University, Mimeo.
- Rubinstein, A., and Y. Salant**, 2012. Eliciting Welfare Preferences from Behavioural Data Sets. *Review of Economic Studies* **79**, 375-387.