Behavioral Implementation

By Geoffroy de Clippel*

Implementation theory assumes that participants' choices are rational, in the sense of being consistent with the maximization of a context-independent preference. The paper investigates implementation under complete information when individuals' choices need not be rational.

Implementation under complete information is a classic problem in mechanism design. The designer would like to implement a rule that selects acceptable outcomes as a function of a problem's characteristics. Unfortunately, while commonly known among participants, these characteristics are unknown to him. He must thus rely on their reports to tailor his selection of outcomes. Taking into account the participants' incentives to misrepresent their information, what are the rules that the designer can effectively implement?

Characteristics encode participants' preferences in standard implementation models. However, there is ample evidence in marketing, psychology and behavioral economics that people's choices need not be consistent with the maximization of a preference relation. Classic examples, which have played a key role in recent developments in choice theory,[1] include status-quo biases, attraction, compromise and framing effects, temptation and self-control, consideration sets, and choice overload. This paper expands implementation theory so as to be applicable in these circumstances as well. The following examples illustrate the scope of the analysis.

(a) (Hiring with Attraction Effect) Members of a hiring committee are meeting to select a new colleague. Up to six candidates are considered: x, y, z, x^*, y^* and z^*. The first three candidates are above the bar, while the last three fall below. For each $a \in \{x, y, z\}$, a^* is similar to a, but dissimilar to other candi-

* Brown University, Department of Economics, 64 Waterman Street, Providence, RI 02912, email: declippel@brown.edu. The paper benefited from insightful conversations with Kfir Eliaz, Eric Maskin, Kareen Rozen, and Roberto Serrano. I also thank seminar participants at Columbia University, ESSET 2013, Northwestern University, NYU, UCLA, UCSD, USC, the 13^{th} SAET conference, and the University of Toronto, as well as Jeff Ely, George Mailath, Ariel Rubinstein, Yuval Salant, Rene Saran, Rani Spiegler, Rajiv Vohra, and three anonymous referees for useful comments. Financial support from the Deutsche Bank through the Institute for Advanced Study is gratefully acknowledged. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

[1]See e.g. Kalai, Rubinstein and Spiegler (2002), Manzini and Mariotti (2007), Ambrus and Rozen (2009), and de Clippel and Eliaz (2012) for choices resulting from the combination of multiple conflicting selves, see Masatlioglu and Ok (2005) for choices with a status-quo bias, see Rubinstein and Salant (2006) for choices with order effects, see Bernheim and Rangel (2009) and Salant and Rubinstein (2008) on framing, see Lleras et al. (2010), Manzini and Mariotti (2012), Masatlioglu, Nakajima and Ozbay (2012), or Cherepanov, Feddersen and Sandroni (2013) on limited attention, see Lipman and Pesendorfer (2013) for a survey on choices with temptation and self-control.

dates. In the spirit of the "attraction" effect,[2] the committee members' individual choices may be influenced by the availability of a similar inferior alternative, e.g. choosing y out of $\{x, y, z\}$, but choosing x out of $\{x, x^*, y, z\}$. Finding which outcomes can be selected as a function of committee members' characteristics thus falls beyond the scope of standard implementation theory.

(b) (Collective Choice with Limited Willpower) Individuals in a support group are committing to make joint decisions. They take part in this group to achieve a common long-term goal. The problem is that individual choices are also influenced by a conflicting short-term craving. In a stylized model, think of an individual's willpower as the number k of alternatives that he or she can overlook to better fulfill his or her long-term goal: chosen options are top-ranked according to the long-term goal among those that are dominated by at most k alternatives for the short-term craving. Such behavior need not be consistent with rationality: one may be able to resist eating a slice of pizza for lunch when the alternative is a salad, but unable to resist both the burger and the pizza slice, and go for the slice if these two options are available on the menu in addition to the salad.[3]

(c) (Groups as Participants) The president of a University is consulting the chairs of its various departments to implement a new policy. Departmental decisions may be reached by following some negotiation protocol. Other chairs may reach decisions by aggregating their colleagues' preferences according to some rule. Either way, the social choice literature teaches us that, in most cases, the chairs' decisions cannot be explained through the maximization of a context-independent preference. The scope of this paper is thus broader than accommodating people's possible behavioral biases: violations of rationality also occur when individuals in the model represent groups of rational agents.[4]

In order to capture all these examples, and many more applications, states in this paper will encode the participants' choice correspondences instead of preferences. A social choice rule (SCR) associates a set of outcomes to each state. It captures the goal that the mechanism designer would like to implement. Unfortunately he does not know the state. He must thus rely on a mechanism, which defines a set of messages for each individual, and a function that associates an outcome to each message profile. A mechanism implements a SCR if the set of equilibrium outcomes coincides with the set of outcomes prescribed by the SCR, at every state.

To make the definition of implementation complete, one must specify what is meant by an equilibrium. A strategy profile forms a Nash equilibrium when individuals maximize preferences if the resulting outcome is among each individual's most-preferred options, within the set of outcomes that he or she can generate

[2]First identified by Huber, Payne and Puto (1982), it has been documented since then in various empirical and experimental settings, see references in Ok, Ortoleva, and Riella (2011, footnotes 2 and 3).

[3]Such choice pattern is consistent with $k = 1$, a long-term goal that ranks the salad above the pizza slice, in turn above the burger, and an opposite short-term goal.

[4]This interpretation motivated Hurwicz's (1986) work, a discussion of which is available at the end of this Introduction.

through unilateral deviations. This equilibrium notion admits a straightforward behavioral extension, replacing 'most-preferred' by 'chosen.' An equilibrium in this sense will still be referred to as a Nash equilibrium. Formal definitions are provided in Section I, and are illustrated in Section II with two examples of implementation, one (as in (a) above) where individuals may be subject to an attraction effect, and one (as in (b) above) where individuals have limited willpower to exercise self-control.

Section III provides necessary, and sufficient conditions for Nash implementability, which extends Maskin's (1999) classic result[5] from domains containing only choice correspondences that are consistent with preference maximization to *any* domain of choice correspondences. These conditions prove useful to delineate the limits of implementation in any application, but as for Maskin monotonicity, more work is usually needed to identify some or all implementable SCRs in each given problem. This paper is dedicated to the study of implementable extensions of the Pareto SCR. Additional practical implications of the general necessary and sufficient conditions are provided in de Clippel (2012).[6]

Pareto efficiency provides a classic example of SCR that is Nash implementable in the standard framework. How does Pareto's concept extend beyond the rational domain? Which extensions, if any, remain Nash implementable? Section IV applies the general results derived in Section III to provide sharp answers. There exist multiple extensions of the Pareto SCR that are Nash implementable on *all* domains. Interestingly, there exists a maximal implementable extension of the Pareto SCR, which can be characterized in simple terms: an option x is efficient in this sense if one can find implicit opportunity sets, one for each individual, such that (a) all individuals would pick x out of their respective opportunity sets, and (b) all options are accounted for, in the sense that each alternative to x belongs to the opportunity set of at least one individual. Similar properties on opportunity sets were independently introduced by Sugden (2004) in economic environments with private consumption, under the name of "opportunity criterion." The maximal implementable extension always refines, sometimes strictly, Bernheim and Rangel's (2009) extended notion of efficiency. There also exists a minimal implementable extension, which happens to be simple variant of its maximal counterpart, where implicit opportunity sets are essentially disjoint. Finally, I show that a social planner may sometimes design a mechanism to guarantee Pareto efficiency when individuals maximize preferences, but imperfectly so, for instance when individuals may be prone to mistakes when being overwhelmed by too many options (see Iyengar and Kamenica (2010) and some references therein).

[5]The paper circulated as a working paper from 1977 and 1998. Surveys on the large literature on implementation theory include Maskin (1985), Moore (1992), Palfrey (1992, 2002), Corchón (1996), Jackson (2001), Maskin and Sjöström (2002), and Serrano (2004).

[6]In particular, the working paper touches upon an important direction for future research, namely to understand what the general necessary and sufficient conditions imply in applications where individuals are subject to specific forms of biases. New implementable SCRs are identified for instance for classes of order-dependant choices. A notion of rich domain is also introduced in the paper, which allows to extend classic impossibility results.

Section V is dedicated to Shapley and Scarf's (1974) housing market, which captures some key features of matching problems and exchange economies. I characterize both the maximal, and a nearly minimal implementable extension of the *strong* Pareto SCR. In exchange economies, aggregate demand functions emanate from individual choices from budget sets. The concept of competitive equilibrium can thus be defined without relying on preferences (see Sugden (2004) and Bernheim and Rangel (2009) among others). I show that a competitive equilibrium exists in housing markets, independently of the individuals' choice functions. We will see that implementable extensions of Pareto play a key role in extending the two fundamental theorems of welfare economics. Finally, taking property rights into account raises the question of coalitional deviations and in particular of individual rationality in the absence of preferences. I show that the strong core also admits a maximal implementable extension, which corresponds to a simple refinement of Sugden's (2004) opportunity criterion. More importantly, competitive allocations belong to this extended core, which refines the first fundamental theorem of welfare economics.

The analysis throughout the paper highlights the key role of "opportunity sets" when taking individual choices instead of preferences as exogenous variables. Opportunity sets capture what is available to an individual in various circumstances. They take the form of budget sets in the case of a competitive equilibrium, capture the set of outcomes an individual can generate through unilateral deviations in case of a Nash equilibrium, or represent the set of objects that remain available after higher priority individuals have made their pick in serial dictatorship procedures. The classic property of Maskin monotonicity extends into a general necessary condition for Nash implementability that requires finding opportunity sets that are consistent with the SCR. It then follows that implementable extensions of Pareto efficiency (or the core) must themselves be defined in terms of implicit opportunity sets, which is the route that Sugden (2004) already followed, but for independent, normative reasons.

Related Literature

The importance of taking behavioral biases into account when designing mechanisms is attracting attention in the popular press, especially since Thaler and Sunstein's (2008) book "Nudge." This interest is also apparent in the academic literature, with an effort to adapt models in industrial organization to determine the best contracts that a monopolist or competing firms can offer to maximize their profits when customers are subject to specific choice biases (see Spiegler's (2011) book for a synthesis). The present paper extends implementation theory to problems where individual choices may be inconsistent with the maximization of a context-independent preference, motivated by recent developments in the theories of choice and welfare.

Proposition 2, which offers necessary and sufficient conditions for Nash implementability, is the only result in this paper that has some precedence in the literature. Inspired by implementation problems where a mechanism designer tries to elicit information from groups, Hurwicz (1986) shows how Maskin's (1999) classic results extend to the case where individual choice correspondences select the set of options that are undominated for some binary relation. By contrast, my analysis accommodates any form of individual behavior, which is important in view of the many varied choice patterns that may occur under bounded rationality. To illustrate the limits of Hurwicz's approach, notice that a choice correspondence is single-valued in his framework only if it is rational. Korpela[7] (2012) (see also Ray (2010)) independently studied the question of characterizing Nash implementability with no restriction on individual choices. I will explain in Section III the similarities and differences of Proposition 2 with these other results.

A few other relevant references include Eliaz (2002), who studies full implementation in Nash equilibrium that is robust to the presence of any number of "faulty" individuals below a fixed threshold, where faulty individuals may behave in any possible way; Cabrales and Serrano (2011), who investigate implementation problems under the behavioral assumption that agents myopically adjust their actions in the direction of better-responses or best-responses; Saran (2011), who studies under which conditions over individual choice correspondences over Savage acts does the revelation principle hold for partial Nash implementation with incomplete information; Glazer and Rubinstein (2012), who introduce a mechanism design model in which both the content and framing of the mechanism affect the agent's ability to manipulate the information he provides; and Bierbrauer and Netzer (2012) who study a first model of partial Bayesian implementation with intentions.

I. Definitions

Consider a set $\mathcal{I}$ of *individuals* who must select an option from a set $\mathcal{X}$. It is assumed throughout the paper that there are at least three individuals, except when discussing the case of two individuals at the end of Section III. Each individual i's choice behavior is captured by a choice correspondence that may vary with a *state*. Formally, i's *choice correspondence* given a state θ selects a non-empty subset, denoted $C_i(\mathcal{S},\theta)$, of each choice problem $\mathcal{S} \subseteq \mathcal{X}$. The set of all relevant states is denoted Θ. Information is *complete*, meaning that the state is common knowledge among individuals (but of course unknown to the mechanism designer).

Individual i's choice correspondence is *rational* at θ if there exists a complete transitive preference $\succeq$ such that $C(\mathcal{S},\theta) = \arg\max_{\succeq} \mathcal{S}$, for each subset $\mathcal{S}$ of $\mathcal{X}$. As is well-known, a choice correspondence is rational if and only if it satisfies Sen's (1971) properties α ("Independence of Irrelevant Alternatives", or IIA) and β ("Expansion"). Rational choice correspondences are single-valued (in which case rationality is equivalent to IIA) when preferences are strict.

A *social choice rule* (SCR) is a correspondence $f : \Theta \to \mathcal{X}$ that selects a non-

[7]I thank Rene Saran for the reference.

empty subset of options for each state. A *mechanism* is a collection $((\mathcal{M}_i)_{i\in\mathcal{I}}, \mu)$, where $\mathcal{M}_i$ is the set of messages available to i, and $\mu : \mathcal{M} \to \mathcal{X}$ is the outcome function (with $\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_i$).

A *strategy* for individual i is the choice of a message in $\mathcal{M}_i$. Under rationality, a strategy profile forms a Nash equilibrium if the resulting outcome is most-preferred for each individual within the set of options he or she can generate through unilateral deviations, that is by varying his or her own strategy while others play their part of the equilibrium. This definition admits a straightforward extension to games involving individual choice correspondences that cannot be explained by the maximization of a well-behaved preference: for each individual, the equilibrium outcome is among the chosen options within the set of outcomes he or she can generate through unilateral deviations.

Formally, individual i's opportunity set given a strategy profile m_{-i} for the other individuals is given by $O_i(m_{-i}) = \{\mu(m_i, m_{-i}) \mid m_i \in \mathcal{M}_i\} \subseteq \mathcal{X}$. Then a strategy profile $(m_i^*(\theta))_{i\in\mathcal{I}}$ forms a *Nash equilibrium* of the game induced by the mechanism $(\mathcal{M}, \mu)$ at a state θ if $\mu(m^*(\theta)) \in C_i(O_i(m_{-i}^*(\theta)), \theta)$, for all $i \in \mathcal{I}$. The mechanism $(\mathcal{M}, \mu)$ *implements* the SCR f *in Nash equilibrium* if $f(\theta) = \{\mu(m^*(\theta)) \mid m^*(\theta)$ is a Nash equilibrium at $\theta\}$, for all $\theta \in \Theta$. The SCR f is then said to be *Nash implementable*.

Individuals' choices are thus assumed to depend only on the set of options available to them. This rules out other, more general forms of bounded rationality where individual choices may vary for instance with the number of times an option appears in an opportunity set, the order in which messages are presented in the definition of the mechanism, or the very phrasing of messages in a common language. Incorporating these features provides an interesting direction for future research.

II. Illustration

A. *Building Willpower in Groups*

Temptation is often understood in economics through the lens of commitment preferences (see Lipman and Pesendorfer (2013) for a survey): an individual who anticipates having to fight temptation at the time of making a choice, may want to commit to a smaller set of options. By contrast, the focus in the psychology literature is often on individuals who make choices while being tempted, and how they deplete their willpower when exercising self-control (see e.g. Baumeister and Tierney (2011)).[8]

Here is a stylized model of choice that captures temptation, self-control and willpower in a way that is closer to the psychology literature. There are n indi-

[8]First attempts to capture willpower in economics include Ozdenoren, Salant and Silverman (2012) who study the optimal management of willpower over time, and Masatlioglu, Nakajima and Ozdenoren (2011) who study commitment preferences for individuals who anticipate having to fight temptation with limited willpower when having to make a choice.

viduals with a common long-term goal. This long-term goal is difficult to achieve due to the presence of tempting alternatives: choices are also influenced by a short-term craving. Each individual has some limited willpower to exercise self-control. In this example, willpower is captured by the number of tempting options an individual can overlook to better fulfill his long-term goal. Formally, given an ordering $\succ_L$ on $\mathcal{X}$ capturing the long-term goal, an ordering $\succ_S$ on $\mathcal{X}$ capturing the short-term craving, and an integer k_i capturing i's willpower, i's choice out of any $\mathcal{T} \subseteq \mathcal{X}$ is the most-preferred element for $\succ_L$ among those that are dominated by at most k_i alternatives according to $\succ_S$. Such behavior typically leads to violations of IIA. For instance, one may be able to resist eating a slice of pizza for lunch when the alternative is a salad, but unable to resist the temptation of both the burger and the pizza slice, and go for the slice if these two options are available on the menu in addition to the salad. This choice pattern can be explained with $k_i = 1$, salad $\succ_L$ pizza slice $\succ_L$ burger, and burger $\succ_S$ pizza slice $\succ_S$ salad.

Consider now a situation where a state determines a common long-term goal, and the individuals' (possibly different) short-term cravings. Is there a way to combine the individuals' limited willpower to help them better fulfill their common long-term goal? One idea is to decentralize the burden of choice by letting each individual be 'in charge of' only a small number of alternatives. Here is a simple mechanism to achieve this. Let $\mathcal{A}_i \subset \mathcal{X}$ be the subset of k_i elements that individual i will be in charge of. Let's assume that $\sum_{i\in\mathcal{I}} k_i \geq \#\mathcal{X}$, so that these subsets can be chosen so as to cover $\mathcal{X}$, that is $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{A}_i$. Each individual picks a message in support of one of the elements he is in charge of, as well as a non-negative integer representing the intensity with which he makes that statement. Formally, $\mathcal{M}_i = \mathcal{A}_i \times \mathbb{Z}_+$. The outcome is the option supported by the individual with most intense report (using any fixed tie-breaking rule if multiple messages are reported with the highest intensity).

The mechanism has the property that, for any message profile, individual i can generate at most $k_i + 1$ options when varying his own message, out of which he picks the option that is best for his long-term goal. Clearly, the option that is top ranked in $\mathcal{X}$ for the long-term goal is an equilibrium outcome of this mechanism, e.g. with the individual in charge of it suggesting it with intensity 1 and all other individuals supporting other options with intensity 0. Conversely, any Nash equilibrium outcome must be the best element in $\mathcal{X}$ for the long-term goal. Otherwise, the individual in charge of it will deviate by supporting it in a message whose intensity is larger than the messages from all other individuals. We have thus proved the following result.

PROPOSITION 1: *If $\sum_{i\in\mathcal{I}} k_i \geq \#\mathcal{X}$, then the SCR that systematically selects the top-choice of the common long-term goal is Nash implementable.*

The result assumes that the mechanism designer knows each individual's willpower index. More generally, willpower indices could vary with the state. Clearly, Proposition 1 remains valid provided the assumption $\sum_{i\in\mathcal{I}} k_i \geq \#\mathcal{X}$ holds with k_i as the minimum of i''s willpower index over all states.

B. *Hiring with Attraction Effect*

Remember the problem of hiring with attraction effect described in (a) in the Introduction. Formally, the set $\mathcal{X}$ of options is $\{x, y, z, x^*, y^*, z^*, nh\}$, where "$nh$" stands for "not hiring," while the set of states is $\Theta = (\mathcal{P} \times \{0,1\})^{\mathcal{I}}$, where $\mathcal{P}$ is the set of strict preferences on $\mathcal{X}$ that rank any of x, y, z above nh, and nh above any of x^*, y^*, z^*. Indices 1 or 0 capture whether the corresponding individual is subject to the attraction effect, or not. Having $\theta_i = (\succ, 0)$ means that individual i is rational and maximizes $\succ$ at θ. If $\theta_i = (\succ, 1)$ instead, then individual i is subject to the attraction effect, and in particular, picks x out of $\{x, x^*, y, z, nh\}$, y out of $\{x, y, y^*, z, nh\}$, and z out of $\{x, y, z, z^*, nh\}$.

Now fix two individuals, say i and j, and consider the following mechanism: $\mathcal{M}_i = \{x, y, z, j, j^*\}$, $\mathcal{M}_j = \{x, y, z\}$, with the outcome function μ (which depends only on i and j's messages) defined as follows:

μ	x	y	z
x	nh	x	x
y	y	nh	y
z	z	z	nh
j	x	y	z
j^*	x^*	y^*	z^*

Individual i's opportunity set given any message from j contains x, y, z, nh and one of x^*, y^*, z^*. If $\theta_i = (\succ, 0)$, a case in which i is rational, then the equilibrium outcome must be $\succ$-maximal on $\mathcal{X}$. If $\theta_i = (\succ, 1)$ instead, a case in which individual i is subject to the attraction effect, then "j" is the only message for i that can be part of a Nash equilibrium, as it is the only message that gives i the option a in the column where a^* is available in his opportunity set. Individual j's opportunity set given this message is $\{x, y, z\}$. Hence the equilibrium outcome must be j's most-preferred option among $\{x, y, z\}$.

It is immediate to check, conversely, that the candidate equilibrium outcomes we have identified as a function of θ are indeed supported by equilibrium strategies. We have thus identified a mechanism that implements the social choice rule f_{ij} that picks i's most-preferred option when i is rational, and picks j's top choice within $\{x, y, z\}$ when i is subject to the attraction effect. Interestingly, the domain is rich to the extent that it contains all profiles of individual choice correspondences associated to the maximization of some strict preference over $\{x, y, z\}$, the range of f_{ij}. With such a rich domain, a SCR that is implementable via a mechanism that relies on messages from only two rational individuals must be dictatorial (see Jackson and Srivastava (1992)). We see that such a result does not extend beyond the rational domain, as the outcome picked by f_{ij} varies with the reports of *both* i and j.

III. Necessary and Sufficient Conditions

The previous section provides only two examples of SCRs that are Nash implementable in two specific applications. It would be useful to have general conditions that could help us identify more systematically SCRs that are Nash implementable in various applications.

Necessary Condition

I start by providing a necessary condition that extends Maskin monotonicity beyond the rational domain. Consider a SCR $f : \Theta \to \mathcal{X}$ that is Nash implementable by a mechanism $((\mathcal{M}_i)_{i\in\mathcal{I}}, \mu)$. Let then m^* be a Nash equilibrium at θ whose associated outcome $\mu(m^*)$ coincides with an element x of $f(\theta)$. This strategy profile defines an opportunity set for each individual, call it O_i, by varying his or her own strategy while others play m^*_{-i}. By definition of Nash equilibrium, it must be that $x \in C_i(O_i, \theta)$, for all i. In addition, if there is some alternative state θ' such that $x \in C_i(O_i, \theta')$ for all i, then clearly m^* forms a Nash equilibrium at θ' as well. Hence x is an equilibrium outcome at θ' and $x \in f(\theta')$ if f is Nash implementable.

A collection $\mathcal{O} = \{O_i(x, \theta) \mid i \in \mathcal{I},\ x \in f(\theta),\ \theta \in \Theta\}$ of opportunity sets is *consistent with* f if

1) $x \in C_i(O_i(x, \theta), \theta)$, for all $\theta \in \Theta$, all $x \in f(\theta)$, and all $i \in \mathcal{I}$,

2) For all θ, θ' and $x \in f(\theta)$, if $x \in C_i(O_i(x, \theta), \theta')$, for all i, then $x \in f(\theta')$.

The argument from the previous paragraph illustrates how the existence of a consistent collection of opportunity sets must be a necessary condition for Nash implementability.

PROPOSITION 2: ***a*** *If a SCR f is Nash implementable, then there exists a collection of opportunity sets that is consistent with f.*

Consistency with a SCR f requires x to be a chosen alternative for each individual out of his or her opportunity set at (x, θ) when the state is θ. In addition, if $x \in f(\theta) \setminus f(\theta')$, then one must find at least one individual who would not pick x out of his or her (x, θ)-opportunity set when the state is θ'. This is reminiscent of Maskin monotonicity for the rational case: if $x \in f(\theta) \setminus f(\theta')$, then there exists i and y in x's lower contour set for $\succeq_i(\theta)$ such that i strictly prefers y over x at θ'.

The key distinction is that checking consistency simplifies under rationality: there is a collection of opportunity sets that is consistent with f if and only if the collection of lower contour sets is consistent with f.[9] The added complexity

[9]Indeed, the premise of condition 2 in the definition of consistency is harder to meet with larger opportunity sets, and opportunity sets cannot be larger than lower contour sets to avoid a contradiction with condition 1.

of having to consider many collections of opportunity sets is the price to pay for the necessary condition to be valid independently of any theory of choice. The many applications presented in this paper and in de Clippel (2012) show how consistency remains nonetheless workable and can provide important insights.

To illustrate consistency, notice how it is satisfied in the examples from Section II: for the willpower example, take $O_i(x, \theta) = \mathcal{A}_i \cup \{x\}$, where $\mathcal{A}_i \subseteq \mathcal{X}$ is the set of options that i is "in charge charge of"; for the hiring example, i's opportunity set contains x, y, z at all states, in addition to a^*, where $a = f_{ij}(\theta)$, if i is subject to the attraction effect at θ, and a^*, for some $a \neq f_{ij}(\theta)$, if i is rational at θ, while j's opportunity set is $\{x, y, z\}$ at states where i is subject to the attraction effect (all other opportunity sets contain only the option picked by f_{ij}).

Sufficient Condition

The existence of a collection of opportunity sets that is consistent with a SCR does not guarantee its implementability. For instance, it is well-known under rationality that an implementable SCR with a full range must pick alternatives that are top-ranked by all individuals. In the absence of rationality, "top-ranked" can be replaced by "chosen within $\mathcal{X}$." The SCR f *respects unanimity* if $x \in f(\theta)$ for any x, θ such that $x \in \cap_{i \in \mathcal{I}} C_i(\mathcal{X}, \theta)$.

Combining unanimity in this sense with a strengthening of consistency provides a useful sufficient condition for implementability. The collection $\mathcal{O}$ of opportunity sets is *strongly consistent* with f if it is consistent with f, and $x \in f(\theta)$, for all x, θ for which there exists j, θ' and $x' \in f(\theta')$ such that $x \in C_i(\mathcal{X}, \theta)$, for all $i \neq j$, and $x \in C_j(O_j(x', \theta'), \theta)$. The condition distiguishing strong consistency from its plain counterpart shares some resemblance with the unanimity property: if a state θ is such that all individuals except j include x in their set of choices from $\mathcal{X}$, and j includes x in his set of choices when facing one of his opportunity sets in the collection $\mathcal{O}$, then the SCR must include x at θ. We are now ready to state a partial converse to Proposition $2.a$.

Proposition $2.b$ *If f respects unanimity and there exists a collection of opportunity sets that is strongly consistent with f, then f is Nash implementable.*

With rationality, this sufficient condition is also necessary for full-range SCRs (see Moore and Repullo (1990)). This ceases to be the case when accommodating bounded rationality. For instance, the full-range SCR that picks the top-choice of the common long-term goal in Section II is implementable while violating the sufficient condition from Proposition $2.b$.

While Proposition $2.b$ allows to prove most results in the rest of the paper, a limit to its applicability comes from the fact that strong consistency and unanimity involves restrictions on the SCR as a function of what individuals would pick when facing $\mathcal{X}$ in its entirety. The example from Section II illustrates how SCRs can sometimes be successfully implemented by decentralizing the burden of choice, making sure that individuals always face small option sets. The next proposition provides a simple alternative sufficient condition that builds on this insight.

Proposition 2.*b*' *If there exists a collection of opportunity sets that is consistent with a SCR f, and there exist* $(\mathcal{X}_i)_{i\in\mathcal{I}}$ *such that* $x \in f(\theta)$ *for each* x, θ *with* $|\{i \in \mathcal{I} \mid x \in C_i(\mathcal{X}_i \cup \{x\}, \theta)\}| \geq |\mathcal{I}| - 2$*, then f is Nash implementable.*

Elements in $\mathcal{X}_i$ can be interpreted as outside options. In addition to the necessary condition of consistency, $f(\theta)$ is required to contain any alternative that all but at most two individuals would pick at θ when it is available in addition to their respective outside options. One is free to decide on the content of the sets $\mathcal{X}_i$, which in particular may be small. This offers a useful alternative to Proposition 2.*b* (see Proposition 6 below, for instance). The sufficient conditions from Propositions 2.*b* and 2.*b'* are not necessary though.[10] While their proofs suggest ways to provide finer sufficient conditions, I conjecture that a condition that is both necessary and sufficient would be so intricate that it would be of limited help in applications.

Independently of my work, Korpela (2012) studies under which assumptions on individual behavior, Moore and Repullo's (1990) conditions remain both necessary and sufficient for implementability (for SCRs with or without a full range). While always sufficient, necessity is shown to require Sen's property α. Unfortunately, most choice models with bounded rationality violate Sen's property.[11] Korpela's sufficient condition is the same as that of Proposition 2.*b* when considering SCRs with a full range, as in most of this paper. Ray (2010) shows that implementability implies a choice-based version of Maskin monotonicity.[12] It is not difficult to check that Ray's property is satisfied whenever there exists a collection of opportunity sets that is consistent with a SCR, and that it is too restrictive to provide the basis for a widely applicable sufficient condition (see de Clippel (2012)).

Proposition 2.*a* holds also in the case of two individuals. As in Moore and Repullo (1990) (see the difference between their conditions μ and $\mu 2$; as well as Dutta and Sen (1991)), an additional necessary condition holds in that case: there exists a function e that associates an element of $\mathcal{X}$ to each x, θ, x', θ' with $x \in f(\theta)$ and $x' \in f(\theta')$, such that (a) $e(x, \theta, x', \theta') \in O_1(x, \theta) \cap O_2(x', \theta')$, and (b) $e(x, \theta, x', \theta') \in f(\theta'')$, for all θ'' such that $e(x, \theta, x', \theta') \in C_1(O_1(x, \theta), \theta'') \cap C_2(O_2(x', \theta'), \theta'')$. Remember indeed that $O_i(x, \theta)$ corresponds to outcomes that i can reach through unilateral deviations when the other individuals play their part of the Nash equilibrium at θ in the mechanism that implements f. With only two individuals, columns and rows of such outcomes must intersect, hence the existence of such a function e. It is not difficult to check that Proposition 2.*b*

[10] For instance, the SCR that picks the top-choice of the common long-term goal in the willpower example from Section II.A is implementable but does not satisfy either sufficient conditions when $k_i = 1$ for all i and $|\mathcal{I}| = |\mathcal{X}| = 3$.

[11] For instance, it implies rationality for single-valued choice correspondences.

[12] This property previously appeared in Aizerman and Aleskerov (1986), who extend Arrow's impossibility theorem when aggregating arbitrary individual choice correspondences.

applies with two individuals after adding this condition. Clearly, Proposition 2.*b*' applies only when there are three or more individuals.

IV. Efficiency

With rational individuals, an option is said to be (weakly) *Pareto efficient* if there is no alternative that is unanimously strictly preferred. The Pareto SCR provides a classic example of SCR that is Nash implementable. In this section, I discuss possible extensions of the Pareto principle when individual choices need not be consistent with the maximization of a well-behaved preference.

Maximal and Minimal Implementable Extensions

A SCR $f : \Theta \to \mathcal{X}$ is an *extension* of the Pareto SCR if Θ contains all profiles of rational choice correspondences and $f(\theta)$ coincides with the set of Pareto efficient options when choice correspondences are rational at θ. It is a *maximal implementable extension* if it is Nash implementable and contains any implementable extension of the Pareto SCR. It turns out that there exist multiple implementable extensions of the Pareto SCR, that there exists a maximal implementable extension, which also admits a simple expression and is closely related to Sugden's (2004) opportunity criterion.

Say that an option x is *efficient* given a profile of individual choice correspondences if one can find a collection of *implicit opportunity sets*, one for each individual, such that all individuals would pick x out of their own implicit opportunity sets, and all the options have been accounted for, in the sense that any option in $\mathcal{X}$ belongs to the implicit opportunity set of at least one individual. Formally,

$$f^{eff}(\theta) = \{x \in \mathcal{X} | (\exists (\mathcal{Y}_i)_{i\in\mathcal{I}} \text{ subsets of } \mathcal{X}) : x \in C_i(\mathcal{Y}_i, \theta), \text{ for all } i, \text{ and } \mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i\},$$

for each $\theta \in \Theta$.[13]

PROPOSITION 3: *f^{eff} is Nash implementable on all domains, and is the maximal implementable extension of the Pareto SCR.*

Sugden (2004, pages 1016-17) writes: "*In normative economics, there is an increasing interest in criteria of opportunity rather than of preference satisfaction. In opportunity-based theories, value is attached to the size and richness of an individual's opportunity set - that is, the set of options from which he is free to choose.*" The use of implicit opportunity sets in the definition of f^{eff} is in line with this point of view. In fact, the property imposed on implicit opportunity sets used to justify the efficiency of an option is closely related to the opportunity

[13]Notice that f^{eff} has non-empty values, since it contains any option x such that $x \in C_i(\mathcal{X}, \theta)$, for some $i \in \mathcal{I}$, in the same way that any most preferred alternative of any individual is always Pareto efficient when choices are rational.

criterion proposed by Sugden (2004) for economic environments with private consumption. This relation will be made precise in the context of housing markets studied in the next section.

Remarkably, there is also a non-empty minimal implementable extension of the Pareto SCR, which is a simple, intuitive variant of f^{eff}. The only difference is that implicit opportunity sets are required to be essentially disjoint. Formally, for each θ, let $\hat{f}^{eff}(\theta)$ be the set of options x for which one can find a collection $(\mathcal{Y}_i)_{i\in\mathcal{I}}$ of subsets of $\mathcal{X}$ such that $x \in C_i(\mathcal{Y}_i, \theta)$, for all i, $\mathcal{Y}_i \cap \mathcal{Y}_j = \{x\}$, for all $i \neq j$, and $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i$. The process of accounting for all alternatives is now fully "decentralized," with each individual facing a distinct subset of alternatives to test against the option being sustained. Clearly $\hat{f}^{eff} \subseteq f^{eff}$, and $\hat{f}^{eff}$ has non-empty values.[14] The SCR $\hat{f}^{eff}$ is the *minimal implementable extension* of the Pareto SCR in the sense that $\hat{f}^{eff} \subseteq f$, for each Nash implementable extension f of the Pareto SCR.

PROPOSITION 4: *$\hat{f}^{eff}$ is Nash implementable on all domains, and is the minimal implementable extension of the Pareto SCR.*

Bernheim and Rangel's (2009) Notion of Efficiency

The question of how to extend the Pareto principle beyond the rational domain is being debated in the recent literature. Bernheim and Rangel (2009) posit that an option a is (strictly) *unambiguously preferred* to an alternative b given a choice correspondence C if $b \notin C(\mathcal{S})$, for all $\mathcal{S}$ containing a. Though necessarily incomplete when Property α is violated, this revealed ordering and the associated Pareto ranking may allow to compare some options. An option is *BR-efficient* if there is no alternative that is unambiguously preferred by all individuals. Let f^{bre} be the SCR that associates to each state θ the set of BR-efficient options at θ.

I now show that f^{eff} (and hence any implementable extension of the Pareto SCR) is a selection of f^{bre}, and that BR-efficient options need not belong to f^{eff}. Say that a SCR is *weakly Nash implementable* if it contains a Nash implementable SCR. By Proposition 3, we can conclude that f^{bre} is weakly Nash implementable, but not Nash implementable.

PROPOSITION 5: *$f^{eff} \subseteq f^{bre}$, and there exist states θ for which $f^{eff}(\theta) \subsetneq f^{bre}(\theta)$. Hence, f^{bre} is weakly Nash implementable, but not Nash implementable.*

Weak implementation guarantees that equilibrium outcomes systematically belong to the SCR, but leaves open the possibility that there may be states at which outcomes picked by the SCR cannot be achieved at equilibrium. It is thus weaker than the notion of (full) implementation introduced in Section I and studied in

[14] Again, any option x such that $x \in C_i(\mathcal{X}, \theta)$, for some i, belongs to $\hat{f}^{eff}(\theta)$ (simply take $\mathcal{Y}_i = \mathcal{X}$ and $\mathcal{Y}_j = \{x\}$, for all $j \neq i$).

a majority of the implementation literature with rational individuals. There are reasons to prefer full over weak implementability, see e.g. Thomson (1996). Even so, learning that a SCR is weakly implementable is valuable, and may be sufficient to fulfill the goal of the mechanism designer in some circumstances.

Pareto Efficiency for 'True' Underlying Preferences

Some models of bounded rationality offer structured departures from rationality, with new primitives (e.g. consideration sets, or salience ordering) influencing choices in addition to preferences. In such circumstances, welfare analysis could be conducted in terms of the 'true' underlying preferences, as suggested for instance by Rubinstein and Salant (2008, 2012), Manzini and Mariotti (2012), and Masatlioglu, Nakajima and Ozbay (2012). Is Pareto efficiency in that sense Nash implementable?

To tackle this question, consider choice models where individuals are rational when facing small choice problems; let's say binary problems so as to encompass as many choice correspondences as possible. Formally, states are restricted to the set Θ^* with the following property: for each $\theta \in \Theta^*$, there exists a preference profile $(\succeq_i)_{i\in\mathcal{I}}$ such that $x \succeq_i y$ if and only if $x \in C_i(\{x, y\}, \theta)$.

The Pareto SCR f^{par} associates to each $\theta \in \Theta^*$ the set of options that are Pareto efficient for the unique preference profile defined by pairwise choices at θ. Thinking in terms of implicit opportunity sets is helpful to study the implementability of f^{par}. Consider for instance a state where all individual preferences coincide. Pareto efficiency requires picking the common top element, say x. Suppose however that individuals do not pick x when facing large choice problems (e.g., they opt for a more salient option when overwhelmed by too many options). With only few individuals compared to the number of options available, any collection of implicit opportunity sets that is consistent with f^{par} (see Proposition 2.a) must omit some alternative y. By consistency, the SCR must still pick x at states where y Pareto dominates x, which means that f^{par} is not implementable when there are only few individuals.

With sufficiently many individuals, however, it is possible to better decentralize the burden of choice.[15] Consider a selection of $\hat{f}^{eff}$, where each implicit opportunity set $\mathcal{Y}_i$ used to justify an option x is required to contain at most two elements. The condition $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i$ then implies that x is Pareto efficient, since individuals are rational over binary problems. With sufficiently many individuals, one can show that there always exists at least one Pareto efficient option x that can be justified via binary implicit opportunity sets, and that the variant of $\hat{f}^{eff}$ is Nash implementable, which means that f^{par} is weakly Nash implementable. Even with many individuals, not all Pareto efficient options can be obtained this way, and f^{par} as a whole is not implementable.

[15] A similar idea already proved useful to establish Proposition 1.

PROPOSITION 6: *If $|\mathcal{I}| \geq 3|\mathcal{X}|$, then f^{par} is weakly Nash implementable. On the other hand, f^{par} is not Nash implementable, whatever the number of individuals, and no selection of f^{par} is Nash implementable if $|\mathcal{I}| \leq |\mathcal{X}| - 2$.*

Illustration

To better grasp what differentiates the concepts discussed in this section, consider three individuals, $\mathcal{I} = \{1, 2, 3\}$, who pick options from $\mathcal{X} = \{a, b, \ldots, z\}$. In all scenarios below, each individual's choices are among the top two feasible elements for some underlying preference.

In a first scenario, all three individuals have the same alphabetical preference $a \succ b \succ \cdots \succ z$. While the first individual systematically maximizes that ordering, the other two systematically pick the second-best option (a choice procedure first studied by Baigent and Gaertner (1996), and further discussed in Kalai, Rubinstein and Spiegler (2002)). All options are BR-efficient: options other than a are not unambiguously dominated by any alternative for individuals 2 and 3, and a is not unambiguously dominated by any alternative for the first individual. The SCR f^{par} is applicable to the extent that pairwise choices of all three individuals appear transitive. However, the concept is meaningful only if one believes that individuals maximize preferences in pairwise choices, which contradicts the story of how choices emerge in this scenario. Applying f^{par} gives $\mathcal{X}$ in its entirety as for BR-efficiency, since the pairwise choices of the first individual are always different from those of the other two individuals. The SCRs f^{eff} and $\hat{f}^{eff}$ happen to coincide, both selecting $\{a, b, c\}$.[16]

In a second scenario, the first individual also picks the second-best option for the alphabetical order. BR-efficiency now selects $\mathcal{X} \setminus \{a\}$, f^{par} selects $\{z\}$, while f^{eff} and $\hat{f}^{eff}$ both select $\{b, c, d\}$.

More generally, if the three individuals' choices always belong to the top two elements of underlying preferences, then for each option selected by f^{eff} (and a fortiori $\hat{f}^{eff}$), it is impossible to find more than three alternatives that Pareto dominate it for these underlying preferences, thus never picking an option that is too inefficient.

To find a case where f^{eff} and $\hat{f}^{eff}$ differ, suppose that all three individuals maximize the common alphabetical preference in all problems containing less than twenty elements, but may happen to make a small mistake, picking at worst the second-best option, when facing larger sets. Suppose in particular that c is picked both out of $\{a, c, d, \ldots, z\}$ and out of $\{b, c, d, \ldots, z\}$. In this scenario, c belongs to f^{eff}, using the implicit opportunity sets $\mathcal{Y}_1 = \{c\}$, $\mathcal{Y}_2 = \{a, c, d, \ldots, z\}$, and $\mathcal{Y}_3 = \{b, c, d, \ldots, z\}$, but not to $\hat{f}^{eff}$.[17]

[16] Indeed, a is picked by the first individual out of $\mathcal{X}$, b is picked by the second individual out of $\mathcal{X}$, and implicit opportunity sets supporting c are $\mathcal{Y}_1 = \{c, d, \ldots, z\}$, $\mathcal{Y}_2 = \{a, c\}$ and $\mathcal{Y}_3 = \{b, c\}$. However, there is no way to find implicit opportunity sets supporting options d to z, as they are dominated in the common underlying ordering by more than two options.

[17] To support c, at least one implicit opportunity set must contain a, c, and at least 18 other elements

V. Housing Markets

Results from Section III are now applied to Shapley and Scarf's (1974) housing market, which captures key features of exchange economies and matching problems. Individuals each own one indivisible object h_i^*, their *initial endowment*. The set of all available objects is given by $\mathcal{H} = \{h_i^* \mid i \in \mathcal{I}\}$, while the set of allocations achievable through trade is given by $\mathcal{X} = \{x \in \mathcal{H}^{\mathcal{I}} \mid x_i \neq x_j,\ \forall i \neq j\}$.

In most matching models, preferences over objects are strict, and individuals care only about their own consumption. I follow the same methodology, but without requiring individuals to be rational. For each $\mathcal{S} \subseteq \mathcal{X}$, let $\mathcal{S}_i$ be the set of objects that accrue to i in the different elements of $\mathcal{S}$. At each θ, individual i's choices among objects are captured by a function $\gamma_i(\cdot, \theta) : P(\mathcal{H}) \to \mathcal{H}$, where $P(\mathcal{H})$ is the set of subsets of $\mathcal{H}$. His choice correspondence over allocations is then given by: $C_i(\mathcal{S}, \theta) = \{x \in \mathcal{S} \mid x_i = \gamma_i(\mathcal{S}_i, \theta)\}$, for each $\mathcal{S} \subseteq \mathcal{X}$.

Efficiency

The SCR f^{eff} identified in the previous section admits an equivalent formulation in terms of opportunity sets of *objects* instead of opportunity sets of *allocations*: an object allocation $x \in \mathcal{X}$ belongs to $f^{eff}(\theta)$ if and only if[18] there exists a collection $(\mathcal{Z}_i)_{i\in\mathcal{I}}$ of subsets of objects such that (a) $x_i = \gamma_i(\mathcal{Z}_i, \theta)$, for each $i \in \mathcal{I}$, and (b) for each allocation $y \in \mathcal{X}$ there exists at least one individual j for which $\mathcal{Z}_j$ contains y_j.

Any allocation x for which $\gamma_i(\mathcal{H}, \theta) = x_i$ for some i belongs to f^{eff}. Yet, there might be two rational individuals j and k who would be better off by consuming x_k and x_j respectively. Under rationality, an option is *strongly Pareto efficient* if there is no alternative that leaves everyone at least as well off, and makes someone strictly better off. Interestingly, Proposition 3 admits an analogue for strong Pareto in housing markets.[19] Say that an allocation $x \in \mathcal{X}$ belongs to $f^{s.eff}(\theta)$ if there exists a sequence $(\mathcal{Z}_i)_{i\in\mathcal{I}}$ of subsets of objects such that (a) $x_i = \gamma_i(\mathcal{Z}_i, \theta)$, for each $i \in \mathcal{I}$, and (b) for each allocation $y \in \mathcal{X} \setminus \{x\}$ there exists at least one individual j for which $\mathcal{Z}_j \setminus \{x_j\}$ contains y_j. Observe that condition (b), interpreted as a property of the collection $(\mathcal{Z}_i)_{i\in\mathcal{I}}$ of opportunity sets, is precisely Sugden's (2004) "Opportunity Criterion."

different from b, while another implicit opportunity set must contain b, c, and at least 18 other elements different from a. These two implicit opportunity sets will have to contain a common option other than c, which is incompatible with $c \in \hat{f}^{eff}$.

[18] If the original definition is satisfied with the collection $(\mathcal{Y}_i)_{i\in\mathcal{I}}$ of subsets of $\mathcal{X}$, then the new definition is satisfied with $\mathcal{Z}_i = \{y_i \mid y \in \mathcal{Y}_i\}$, for each $i \in \mathcal{I}$. Conversely, if the new definition is satisfied with the collection $(\mathcal{Z}_i)_{i\in\mathcal{I}}$ of subsets of $\mathcal{H}$, then the original definition is satisfied with $\mathcal{Y}_i = \{y \mid y_i \in \mathcal{Z}_i\}$, for each $i \in \mathcal{I}$.

[19] As is well-known, the strong Pareto SCR is not always Nash implementable. This is why its extensions were not discussed in the general framework of the previous section. However, implementability does prevail in housing markets if individuals have strict preferences (with individual i being indifferent between two allocations x and y if and only if $x_i = y_i$).

PROPOSITION 7: *$f^{s.eff}$ is the maximal implementable extension of the strong Pareto SCR.*

It remains unknown whether strong Pareto admits a minimal implementable extension. However, I can propose an implementable extension that is close to being minimal. Let $\pi : \mathcal{I} \to \{1, \ldots, n\}$ be an enumeration of the individuals, and Π be the set of all such enumerations. The *serial dictatorship* SCR associated to π, denoted f^{π}, is defined by induction: for each k, individual j with $\pi(j) = k$ receives the object $\gamma_j(\mathcal{H} \setminus \mathcal{S}_k, \theta)$, where $\mathcal{S}_k$ is the set of objects allocated to individuals who precede j in the enumeration. Hence, j is effectively free to choose from an opportunity set whose content is given by his rank and what individuals with lower rank pick. Let then f^{SD} be the SCR obtained by considering all possible enumerations π, $f^{SD}(\theta) = \cup_{\pi \in \Pi} f^{\pi}(\theta)$, and $\hat{f}^{s.eff}$ be the SCR defined by: $\hat{f}^{s.eff}(\theta) = f^{SD}(\theta) \cup \{x \in \mathcal{X} \mid \exists j, \forall i \neq j : \gamma_i(\mathcal{H}, \theta) = x_i\}$.

PROPOSITION 8: *$\hat{f}^{s.eff}$ is Nash implementable. If f is an implementable extension of the strong Pareto SCR, then $f^{SD} \subseteq f$.*

Fundamental Theorems of Welfare Economics

The notion of competitive equilibrium is another central concept in economic environments with property rights. Opportunity sets play a key role in its definition, making it easy to extend the concept beyond rational choice. Suppose that objects' prices are given by $p \in \mathbb{R}^{\mathcal{H}}_{+}$. By selling his or her endowment, i enjoys an income $\$p(h_i^*)$; he or she can afford any object whose price is lower or equal to $p(h_i^*)$. Denote by $B_i(p)$ the corresponding budget set: $B_i(p) = \{h \in \mathcal{H} \mid p(h) \leq p(h_i^*)\}$. If the state is θ, then he or she picks $\gamma_i(B_i(p), \theta)$. A price vector p is a *competitive equilibrium* if supply equals demand, that is, for each object h there exists a unique individual i who picks h out of his or her budget set. An allocation is competitive if there is a competitive equilibrium inducing it.

The same equilibrium notion was introduced and studied by Sugden (2004) and Bernheim and Rangel (2009) in economies with divisible goods. The focus on housing markets allows to provide a first existence result. Remarkably, existence holds independently of the individuals' choice behavior. With rational individuals, the (unique) competitive allocation can be found via Shapley and Scarf's (1974) top trading cycle procedure. Dropping rationality, I prove in the Appendix that competitive allocations coincide with the outcome of an extended trading cycle procedure. Existence of a competitive equilibrium will follow, but multiplicity is now a possibility.

PROPOSITION 9: *There is at least one competitive allocation, sometimes more.*

The next proposition provides an extension of the two fundamental theorems of welfare economics beyond the rational domain.

PROPOSITION 10: *(a) Competitive allocations at θ belong to $f^{s.eff}(\theta)$, but need not belong to other (smaller) implementable extensions of strong Pareto.*

(b) Any element of $\hat{f}^{s.eff}(\theta)$ is competitive at θ for some endowment. The result need not hold for other (larger) implementable extensions of strong Pareto.

If individuals are rational at θ, then $f^{s.eff}(\theta) = \hat{f}^{s.eff}(\theta)$, and Proposition 10 indeed boils down to the well-known theorems of welfare economics.

Some previous attempts at extending the first theorem of welfare economics to encompass bounded rationality can be found in the literature. Sugden (2004) proved with divisible goods that any collection of opportunity sets that is market compatible must satisfy his opportunity criterion. Proposition 10 (a) provides an analogue for housing markets, phrased in terms of allocations instead of opportunity sets. Bernheim and Rangel (2009) prove that any behavioral equilibrium is BR-efficient. Proposition 10 (a) refines this result for housing markets since BR efficiency is more permissive than $f^{s.eff}$.

Individual Rationality and the Core

Pareto efficiency is desirable, but satisfying it is not sufficient for a trading mechanism such as competitive markets to be judged appealing. For instance, serial dictatorship selects outcomes that are Pareto efficient, but need not respect property rights. As illustration, some rational individual may end up consuming an object that is worse than his endowment. Fortunately, competitive allocations are not only Pareto efficient, but also individually rational, and even belong to the strong core: it is impossible to make all the members of a coalition weakly better off and at least one of them strictly better off by reallocating their initial endowments.

The following proposition characterizes the largest implementable extension of the strong core, shows that it is always non-empty, and refines the first fundamental theorem of welfare economics from Proposition 10.(a). For each coalition $\mathcal{S}$, let $\mathcal{F}(\mathcal{S})$ be the set of object allocations that can be achieved by its members, that is, the set of $\alpha \in \mathcal{H}^{\mathcal{S}}$ such that $\{\alpha_i \mid i \in \mathcal{S}\} = \{h_i^* \mid i \in \mathcal{S}\}$ and $\alpha_i \neq \alpha_j$ for $i \neq j$. Say that an allocation x belongs to $f^{s.core}(\theta)$ if there exists a collection $(\mathcal{W}_i)_{i \in \mathcal{I}}$ of subsets of objects such that $a)$ $x_i = \gamma_i(\mathcal{W}_i, \theta)$, for each $i \in \mathcal{I}$, and $b)$ for each coalition $\mathcal{S}$ and each allocation $y \in \mathcal{F}(\mathcal{S}) \setminus \{x_{\mathcal{S}}\}$ there exists at least one individual j for which $\mathcal{W}_j \setminus \{x_j\}$ contains y_j.

PROPOSITION 11: *$f^{s.core}$ is the maximal implementable extension of the strong core, it contains competitive allocations, and has thus non-empty values.*

Condition $b)$ in the definition of $f^{s.core}$ boils down to condition $b)$ in the definition of $f^{s.eff}$ when $\mathcal{S} = \mathcal{I}$. Hence $f^{s.core} \subseteq f^{s.eff}$. Individual rationality constraints are captured by considering singleton $\mathcal{S}$'s. It is instructive to see what these conditions become in presence of general choice functions. Condition *b*) for $\mathcal{S} = \{i\}$ amounts to $h_i^* \in \mathcal{W}_i$. In other words, initial endowments must belong to the implicit opportunity sets used to justify an allocation, and property rights translate into the requirement that individuals must have the freedom to pick their endowment if they wish to.

References

Aizerman, Mark A., and Fuad T. Aleskerov. 1986. "Voting operators in the space of choice functions." *Mathematical Social Sciences* 11: 201-242.

Ambrus, Attila, and Kareen Rozen. 2009. "Rationalizing Choice with Multi-Self Models." *Economic Journal*: forthcoming.

Bade, Sophie. 2008. "Housing Problems with Non-Rationalizable Behavior." Unpublished.

Baigent, Nick, and Wulf Gaertner. 1996. "Never Choose the Uniquely Largest: A Characterization." *Economic Theory* 8: 239-249.

Baumeister, Roy F., and John Tierney. 2011. "Willpower: Rediscovering the Greatest Human Strength." Penguin Press.

Bernhein, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124: 51-104.

Bierbrauer, Felix and Nick Netzer. 2012. "Mechanism Design and Intentions." University of Cologne Working Paper Series in Economics 53.

Cabrales, Antonio, and Roberto Serrano. 2011. "Implementation in Adaptative Better-Response Dynamics: Towards a General Theory of Bounded Rationality in Mechanisms." *Games and Economic Behavior* 73: 360-374.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2013. "Rationalization." *Theoretical Economics* 8: 775-800.

Corchón, Luis. 1996. "The Theory of Implementation of Socially Optimal Decisions in Economics." St. Martin's Press.

de Clippel, Geoffroy. 2012. "Behavioral Implementation." Brown University Economics Working Paper 2012-6.

de Clippel, Geoffroy, and Kfir Eliaz. 2012. "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects." *Theoretical Economics* 7: 125-162.

Dutta, Bhaskar, and Arunava Sen. 1991. "A Necessary and Sufficient Condition for Two-Person Implementation." *Review of Economic Studies* 58: 121-128.

Eliaz, Kfir. 2002. "Fault-Tolerant Implementation." *Review of Economic Studies* 69: 589-610.

Glazer, Jacob, and Ariel Rubinstein. 2012. "A Model of Persuasion with a Boundedly Rational Agent." *Journal of Political Economy* 120: 1057-1082.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding

Asymmetrically Dominated Alternatives: Violations of Regularity and the Slimilarity Hypothesis." *Journal of Consumer Research* 9: 90-98.

Hurwicz, Leonid. 1986. "On the Implementation of Social Choice Rules in Irrational Societies." In: Essays in Honor of Kenneth J. Arrow (Vol I), ed. by Heller et al., Cambridge University Press.

Iyengar, Sheena, and Emir Kamenica. 2010. "Choice Proliferation, Simplicity Seeking, and Asset Allocation." *Journal of Public Economics* 94: 530-539.

Jackson, Matthew O.. 2001. "A Crash Course in Implementation Theory." *Social Choice and Welfare* 18: 655-708.

Jackson, Matthew O., and Sanjay Srivastava. 1992. "On Two-Person Nash Implementable Choice Functions." *Social Choice and Welfare* 9: 263-264.

Kalai, Gil, Ariel Rubinstein, and Ran Spiegler. 2002. "Rationalizing Choice Functions by Multiple Rationales." *Econometrica* 70: 2481-2488.

Korpela, Ville. 2012. "Implementation without Rationality Assumptions." *Theory and Decision* 72: 189-203.

Lipman, Barton L., and Wolfgang Pesendorfer. 2013. "Temptation." In "Advances in Economics and Econometrics: Tenth World Congress" (Vol. 1), Ed. by Acemoglu, Arellano, and Dekel, Cambridge University Press.

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Ozbay. 2010. "When More is Less: Choice by Limited Consideration." Unpublished.

Manzini, Paola, and Marco Mariotti. 2007. "Sequentially Rationalizable Choice." *American Economic Review* 97: 1824-1839.

Manzini, Paola, and Marco Mariotti. 2012. "Categorize Then Choose: Boundedly Rational Choice and Welfare." *Journal of the European Economic Association* 10: 1141-1165.

Masatlioglu, Yusufcan, and Efe A. Ok. 2005. "Rational Choice with Status Quo Bias." *Journal of Economic Theory* 121: 1-29.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Ozbay. 2012. "Revealed Attention." *American Economic Review* 102: 2183-2205.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Emre Ozdenoren. 2011. "Revealed Willpower." Unpublished.

Maskin, Eric. 1985. "The Theory of Implementation in Nash Equilibrium: A Survey." In Social Goals and Social Organization, L. Hurwicz, D. Schmeidler, and H. Sonnenschein, eds., Cambridge University Press, 173-204.

Maskin, Eric. 1999. "Nash Equilibrium and Welfare Optimality." *Review of Economic Studies* 66: 23-38.

Maskin, Eric, and Tomas Sjöström. 2002. "Implementation Theory." In Handbook of Social Choice and Welfare, Vol. I, K. J. Arrow, A. Sen, and K. Suzumura, eds., Elsevier Science, New York, 237-288.

Moore, John. 1992. "Implementation, Contracts and Renegotiation in Environments with Complete Information." In Advances in Economic Theory, 4th World Congress of the Econometric Society, Vol. I, J. J. Laffont, ed., Cambridge University Press, Cambridge, UK, 182-282.

Moore, John, and Rafael Repullo. 1990. "Nash Implementation: A Full Characterization." *Econometrica* 56: 1191-1220.

Ok, Efe A., Pietro Ortoleva, and Gil Riella. 2011. "Revealed (P)Reference Theory." Unpublished.

Ozdenoren, Emre, Stephen W. Salant, and Dan Silverman. 2012. "Willpower and the Optimal Control of Visceral Urges." *Journal of the European Economic Association* 10: 342-368.

Palfrey, Thomas R.. 1992. "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design." In Advances in Economic Theory, 4th World Congress of the Econometric Society (Vol. I), J. J. Laffont, ed., Cambridge University Press, 283-323.

Palfrey, Thomas R.. 2002. "Implementation Theory." In Handbook of Game Theory with Economic Applications, Vol. III, R. J. Aumann and S. Hart, eds., Elsevier Science, New York, 2271-2326.

Ray, Korok T.. 2010. "Nash Implementation under Irrational Preferences." Unpublished.

Rubinstein, Ariel, and Yuval Salant. 2006. "A Model of Choice from Lists." *Theoretical Economics* 1: 3-17.

Rubinstein, Ariel, and Yuval Salant. 2008. "Some Thoughts on the Principle of Revealed Preference." In: Foundations of Positive and Normative Economics, ed. by A. Caplin and A. Schotter. Oxford University Press.

Rubinstein, Ariel, and Yuval Salant. 2012. "Eliciting Welfare Preferences from Behavioural Data Sets.“ *Review of Economic Studies* 79: 375-387.

Salant, Yuval, and Ariel Rubinstein. 2008. "(A,f): Choice with Frames." *Review of Economic Studies* 75: 1287-1296.

Saran, Rene. 2011. "Menu-Dependent Preferences and Revelation Principle." *Journal of Economic Theory* 146: 1712-1720.

Sen, Amartya K.. 1971. "Choice Functions and Revealed Preference." *Review of Economic Studies* 38: 307-317.

Serrano, Roberto. 2004. "The Theory of Implementation of Social Choice Rules." *SIAM Review* 46: 377-414.

Shapley, Lloyd S., and Herbert E. Scarf. 1974. "On Cores and Indivisibility." *Journal of Mathematical Economics* 1: 23-37.

Simon, Herbert A.. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics* 69: 99-118.

Spiegler, Ran. 2011. "Bounded Rationality and Industrial Organization." *Oxford University Press.*

Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences." *American Economic Review* 94: 1014-1033.

Thaler Richard H., and Cass R. Sunstein. 2008. "Nudge: Improving Decisions About Health, Wealth, and Happiness." *Yale University Press.*

Thomson, William. 1996. "Concepts of Implementation." *Japanese Economic Review* 47: 133-143.

Mathematical Appendix

The proofs of Propositions 1 and 2.a are available in the main text.

Proof of Proposition 2.b

Consider a canonical mechanism as in Maskin (1999) or Moore and Repullo (1990): $\mathcal{M}_i = \mathcal{X} \times \Theta \times \mathbb{Z}_+$, for each $i \in \mathcal{I}$, with the outcome function μ defined as follows:

(1) If $x \in f(\theta)$ and $m_i = (x, \theta, 0)$ for each i, then $\mu(m) = x$.

(2) If there is $j \in \mathcal{I}$ and $x \in f(\theta)$ such that $m_i = (x, \theta, 0)$, for each $i \in \mathcal{I} \setminus \{j\}$, and $m_j = (x', \theta', \alpha)$ with $(x', \theta', \alpha) \neq (x, \theta, 0)$, then $\mu(m) = x'$ if $x' \in O_j(x, \theta)$, and $\mu(m) = x$ 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.

To show that this mechanism implements f, observe that the set of options that i can generate through unilateral deviations is $O_i(x, \theta)$ if $m_j = (x, \theta, 0)$ with $x \in f(\theta)$, for all $j \neq i$, and is $\mathcal{X}$ otherwise. The first condition of consistency of $\mathcal{O}$ with f implies that, for each $x \in f(\theta)$, the strategy profile $m^x(\theta) = (x, \theta, 0)$ forms a Nash equilibrium of the game induced by $(\mathcal{M}, \mu)$ at θ. The equilibrium outcome is x. Consider now an equilibrium m at θ'' whose associated outcome is y. It remains to show that $y \in f(\theta'')$. This follows from the second condition of consistency if m satisfies the requirements in part (1) in the definition of μ, from the definition of strong consistency if m satisfies the requirements in part (2) in the definition of μ, and from unanimity if m satisfies the requirements in part (3) in the definition of μ. $\square$

Proof of Proposition 2.b'

Consider a variant of mechanism from Proposition 2.b: $\mathcal{M}_i = \mathcal{X} \times \Theta \times \mathbb{Z}_+ \times \mathcal{X}_i$, for each $i \in \mathcal{I}$, with the outcome function defined as follows:

(1) If $x \in f(\theta)$ and $m_i = (x, \theta, 0, \cdot)$ for each i, then $\mu(m) = x$.

(2) If there is $j \in \mathcal{I}$ and $x \in f(\theta)$ such that $m_i = (x, \theta, 0, \cdot)$, for each $i \in \mathcal{I} \setminus \{j\}$, and $m_j = (x', \theta', \alpha, \cdot)$ with $(x', \theta', \alpha) \neq (x, \theta, 0)$, then $\mu(m) = x'$ if $x' \in O_j(x, \theta)$, and $\mu(m) = x$ otherwise.

(3) In all other cases, $\mu(m) = x_i$, where x_i is the fourth component in the report of the individual i with the lowest index among those who submit the highest integer.

The difference with the mechanism from the proof of Proposition 2.b is thus that each individual i includes an element of $\mathcal{X}_i$ in his message, which becomes relevant in part (3) of the mechanism. As in the proof of Proposition 2.b, the fact that $\mathcal{O}$ is consistent with f implies that x is a Nash equilibrium outcome at θ, for each $x \in f(\theta)$, and that any outcome associated to a Nash equilibrium at θ' that satisfies the requirements in part (1) in the definition of the modified mechanism belongs to $f(\theta')$. Suppose next that m is a Nash equilibrium at θ'' satisfying the requirements of parts (2) or (3), resulting in an outcome y. It is easy to check that there are at least $|\mathcal{I}| - 2$ individuals k such that the set of outcomes they can generate through unilateral deviations is $\mathcal{X}_k \cup \{y\}$, and $y \in C_i(\mathcal{X}_k \cup \{y\}, \theta'')$ by definition of Nash equilibrium. Hence $y \in f(\theta)$ and f is implementable, as desired. $\square$

PROOF OF PROPOSITION 3

By definition, one can associate to any $x \in f^{eff}(\theta)$ a collection $(\mathcal{Y}_i^{x,\theta})_{i \in \mathcal{I}}$ of subsets of $\mathcal{X}$ such that $x \in C_i(\mathcal{Y}_i^{x,\theta}, \theta)$, for each $i \in \mathcal{I}$, and $\mathcal{X} = \cup_{i \in \mathcal{I}} \mathcal{Y}_i^{x,\theta}$. Then the collection $\mathcal{O} = \{\mathcal{Y}_i^{x,\theta} \mid i \in \mathcal{I},\ x \in f^{eff}(\theta),\ \theta \in \Theta\}$ of opportunity sets is consistent with f^{eff}. Strong consistency of $\mathcal{O}$ with f^{eff} and the fact that f^{eff} respects unanimity follow from the fact that $f^{eff}(\theta)$ includes any x such that $x \in C_i(\mathcal{X}, \theta)$, for some $i \in \mathcal{I}$. Hence f^{eff} is Nash implementable, by Proposition 2.b.

To show that f^{eff} extends the Pareto SCR, suppose that individual choice correspondences are rational at θ. If x is Pareto efficient, then it is easy to check that $x \in f^{eff}(\theta)$ by taking $\mathcal{Y}_i$ as the lower contour set of x for i. Conversely, suppose that the collection $(\mathcal{Y}_i)_{i \in \mathcal{I}}$ of implicit opportunity sets justify why $x \in f^{eff}(\theta)$. For any element y of $\mathcal{X}$, there exists i such that $y \in \mathcal{Y}_i$, since $\mathcal{X} = \cup_i \mathcal{Y}_i$, and hence y is in the lower contour set of x for i, since $x \in C_i(\mathcal{Y}_i, \theta)$. This proves that x is Pareto efficient.

I can now show that f^{eff} is the maximal implementable extension of the Pareto SCR. Suppose instead that there exist a Nash implementable extension f of the Pareto SCR, a state θ, and an option $x \in f(\theta) \setminus f^{eff}(\theta)$. By Proposition 2.$a$, let $\mathcal{O}$ be a collection of opportunity sets that is consistent with f. Since $x \notin f^{eff}(\theta)$, there must exist $y \in \mathcal{X} \setminus \bigcup_{i \in \mathcal{I}} O_i(x, \theta)$. Consider then any state θ' where all individuals maximize a preference that ranks y as the unique top best option, and x as the unique second best option. Notice that $C_i(O_i(x, \theta), \theta') = x$, for all

i, and hence $x \in f(\theta')$. This contradicts the fact that f extends the Pareto SCR, since x is not Pareto efficient at θ'. $\square$

Proof of Proposition 4

The proof that $\hat{f}^{eff}$ is Nash implementable on all domains is similar to that of Proposition 3, and is left to the reader (remember in particular footnote 14 when checking unanimity and strong consistency)

I now check that $\hat{f}^{eff}$ extends the Pareto SCR. Suppose that individual choice correspondences are rational at θ. If x is Pareto efficient, then it is easy to check that $x \in \hat{f}^{eff}(\theta)$ by taking $\mathcal{Y}_i = \{x\} \cup \mathcal{Y}'_i$, where $\mathcal{Y}'_i$ is the lower contour set of x for i minus the union over $j < i$ of the lower contour sets of x for j. Conversely, if $x \in \hat{f}^{eff}(\theta) \subseteq f^{eff}(\theta)$, then it is Pareto efficient by Proposition 3.

Finally, let f be a Nash implementable extension of the Pareto SCR. I show that $\hat{f}^{eff} \subseteq f$. Let $(\mathcal{Y}_i)_{i\in\mathcal{I}}$ be a collection of implicit opportunity sets that justifies why $x \in \hat{f}^{eff}(\theta)$. Let then θ' be a state where each individual i is rational, ranking x strictly above all other elements of $\mathcal{Y}_i$, and strictly below all elements of $\mathcal{X} \setminus \mathcal{Y}_i$. Since $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i$, x is Pareto efficient at θ', and $x \in f(\theta')$. By Proposition 2.a, let $\mathcal{O}$ be a collection of opportunity sets that is consistent with f. In particular, $O_i(x, \theta') \subseteq \mathcal{Y}_i$, for each i. Suppose now that there exist an individual i and an option $z \in \mathcal{Y}_i \setminus O_i(x, \theta')$. Since $\mathcal{Y}_i \cap \mathcal{Y}_j = \{x\}$, there is no individual j such that $z \in O_j(x, \theta')$. Consider then a state θ'' that differ from θ' only in that z is the unique top-best option for all individuals. We have $C_j(O_j(x, \theta'), \theta'') = \{x\}$, for each j, and hence $x \in f(\theta'')$, a contradiction since f coincides with the Pareto SCR at θ'' while z Pareto dominates x at θ''. To avoid this contradiction, it must be that $O_i(x, \theta') = \mathcal{Y}_i$, for each i. Since $x \in C_i(\mathcal{Y}_i, \theta)$, for all i, it follows that $x \in f(\theta)$, as desired. $\square$

Proof of Proposition 5

To show that any option $x \in f^{eff}(\theta)$ is BR-efficient at θ, let $(\mathcal{Y}_i)_{i\in\mathcal{I}}$ be a collection of implicit opportunity sets justifying why $x \in f^{eff}(\theta)$. For each alternative y, there exists i such that $y \in \mathcal{Y}_i$, since $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i$. Hence y does not unambiguously dominates x for i, given that $x \in C_i(\mathcal{Y}_i, \theta)$, and x is thus BR-efficient.

Next I show that BR-efficient options need not belong to f^{eff}. Start by enumerating the elements of $\mathcal{X}$: $\mathcal{X} = \{x_1, \ldots, x_K\}$. Let C^* be the rational choice correspondence that systematically picks the element with smallest index: $C^*(\mathcal{S}) = \{x_{\min\{k|x_k\in\mathcal{S}\}}\}$, for each $\mathcal{S} \subseteq \mathcal{X}$. Consider a state θ at which $C_i(\cdot, \theta) = C^*(\cdot)$, for each $i \neq 1$, $C_1(\{x_k, x_K\}, \theta) = \{x_K\}$, for each k, and $C_1(\mathcal{S}, \theta) = C^*(\mathcal{S})$, for all other subsets $\mathcal{S}$ of $\mathcal{X}$. Option x_K is BR-efficient, since x_K is not unambiguously dominated by any alternative for the first individual (see pairwise choices). However, $x_K \notin f^{eff}(\theta)$. Indeed, implicit opportunity sets justifying why x might be efficient, must be the singleton $\{x_K\}$ for individuals other than 1 (x_K is the worst option for them), and must be $\{x_K\}$ or a pair that contains x_K for individual 1 (in order for him to pick x_K). Hence, $\mathcal{X} \neq \cup_{i\in\mathcal{I}}\mathcal{Y}_i$, and $x_K \notin f^{eff}(\theta)$, as claimed.

Weak implementability of f^{bre}, and the fact that f^{bre} is not Nash implementable then follows from Proposition 3. □

Proof of Proposition 6

The result holds trivially when $\mathcal{X}$ contains only two elements, since individuals are rational in that case. It is thus assumed throughout the proof that $\mathcal{X}$ contains at least three elements.

For each $\theta \in \Theta^*$, let $f(\theta)$ be the set of options x for which one can find a collection $(\mathcal{Y}_i)_{i\in\mathcal{I}}$ of subsets of $\mathcal{X}$ such that $|\mathcal{Y}_i| \leq 2$, $x \in C_i(\mathcal{Y}_i,\theta)$, for all i, $\mathcal{Y}_i \cap \mathcal{Y}_j = \{x\}$, for all $i \neq j$, and $\mathcal{X} = \cup_{i\in\mathcal{I}}\mathcal{Y}_i$. This is simply a selection of $\hat{f}^{eff}$, where each implicit opportunity set used to justify an option contains at most two elements. To check that $f \subseteq f^{par}$, let $x \in f(\theta)$ and suppose that there exists y such that $y \succ_i^\theta x$, for all i. This means that $x \notin C_i(\{x,y\},\theta)$, for all i, but this contradicts the definition of x, since $\mathcal{Y}_j = \{x,y\}$ for some j.

If $|\mathcal{I}| \geq |\mathcal{X}|-1$, then $f(\theta) \neq \emptyset$, for all $\theta \in \Theta^*$. To check this, define x_i as one of the $\succeq_i^\theta$-minimal elements in $\mathcal{X} \setminus \{x_1,\dots,x_{i-1}\}$, by induction on $i = 1,\dots,|\mathcal{X}|-1$, and let x be the remaining element in $\mathcal{X} \setminus \{x_1,\dots,x_{|\mathcal{X}|-1}\}$. Clearly, $x \in f(\theta)$ (using $\mathcal{Y}_i = \{x_i,x\}$, for all $i = 1,\dots,|\mathcal{X}|-1$, and $\mathcal{Y}_i = \{x\}$, for all $i \geq |\mathcal{X}|-1$).

For each $x \in f(\theta)$, let $(\mathcal{Y}_i^{x,\theta})_{i\in\mathcal{I}}$ be a collection of subsets of $\mathcal{X}$ justifying why $x \in f(\theta)$. Clearly, the collection $\mathcal{O} = \{\mathcal{Y}_i^{x,\theta} \mid i \in \mathcal{I},\ x \in f(\theta),\ \theta \in \Theta\}$ is consistent with f. Checking the other condition of Proposition 2.b' will prove that f is implementable.[20] Let $x_1,\dots,x_{|\mathcal{X}|}$ be an enumeration of $\mathcal{X}$, and $\mathcal{X}_i = \{x_{(i) mod|\mathcal{X}|}\}$, for each $i \in \mathcal{I}$. If x,θ are such that $|\{i \in \mathcal{I} \mid x \in C_i(\{x_{(i) mod|\mathcal{X}|},x\},\theta)\}| \geq |\mathcal{I}|-2$, then for each $y \in \mathcal{X}$ there exists i such that $x \in C_i(\{x,y\},\theta)$ (since $|\mathcal{I}| \geq 3|\mathcal{X}|$). Hence $x \in f(\theta)$ and f is implementable.

I now show, by contradiction, that f^{par} is not Nash implementable (independently of the number of individuals). Otherwise, let $\mathcal{O}$ be a collection of opportunity sets that is consistent with f^{par}. Consider a state $\theta \in \Theta^*$ at which the first individual ranks x strictly above all other options (according to pairwise choices), while other individuals have the opposite ranking. Suppose in addition that the first individual does not pick x out of sets that contain three or more options, while other individuals are rational. The first condition of consistency implies that $O_i(x,\theta) = \{x\}$, for all $i \geq 2$, while $O_1(x,\theta)$ contains x and at most one other element y. Consider then $z \neq x,y$, and a rational state $\theta' \in \Theta^*$ where all individuals rank z top, and x second-best. By consistency, x must belong to $f^{par}(\theta')$, a contradiction.

[20] Proposition 2.b does not apply, as $\mathcal{O}$ is not necessarily strongly consistent with f, and f need not respect unanimity.

Suppose now that $|\mathcal{I}| \leq |\mathcal{X}| - 2$. I prove, by contradiciton, that there is no selection of f^{par} that is Nash implementable. Otherwise, let f be an implementable selection, and let $\mathcal{O}$ be a collection of opportunity sets that is consistent with f. Consider a state $\theta \in \Theta^*$ where pairwise choices rank x as the top element for all individuals, while they do not pick x out of sets that contain three or more options. Hence $f(\theta) = x$, and $O_i(x, \theta)$ contains x plus at most one other option, for each i, by the first consistency condition. Since $|\mathcal{I}| \leq |\mathcal{X}| - 2$, there must exist $y \in \mathcal{X} \setminus (\bigcup_i O_i(x, \theta))$. Consider then a rational state $\theta' \in \Theta^*$ where all individuals rank y top, and x second-best. By consistency, x must belong to $f(\theta')$, while $f(\theta') = y$, a contradiction. $\square$

Proof of Proposition 7

To show that $f^{s.eff}$ extends strong Pareto, suppose that $\gamma_i(\cdot, \theta)$ is rational at θ, for each i, and let $\mathcal{Z}_i$ be the union of $\{x_i\}$ with the (strict) lower contour set (in $\mathcal{H}$) of x_i for i. By definition, x is strongly Pareto efficient at θ if and only if there is no allocation $y \in \mathcal{X} \setminus \{x\}$ such that for each j, $y_j = x_j$ or y_j belongs to the (strict) upper contour set of x_j for j. This, in turn, is equivalent to: if $y \in \mathcal{X} \setminus \{x\}$, then $y_j \in \mathcal{Z}_j \setminus \{x_j\}$, for some j. This is also equivalent to the existence of a collection $(\hat{\mathcal{Z}}_i)_{i \in \mathcal{I}}$ of subsets of $\mathcal{H}$ justifying why $x \in f^{s.eff}(\theta)$, as desired. The necessary condition is trivial - simply take $\hat{\mathcal{Z}}_i = \mathcal{Z}_i$ for each i. For the sufficient condition, simply notice that $y_j \in \mathcal{Z}_j \setminus \{x_j\}$ if $y_j \in \hat{\mathcal{Z}}_j \setminus \{x_j\}$, since $\hat{\mathcal{Z}}_j \subseteq \mathcal{Z}_j$ (given that $x_j = \gamma_j(\hat{\mathcal{Z}}_j, \theta)$).

Next I show that $f^{s.eff}$ is Nash implementable in all housing markets. By definition, one can associate to any $x \in f^{s.eff}(\theta)$ a collection $(\mathcal{Z}_i^{x,\theta})_{i \in \mathcal{I}}$ of subsets of $\mathcal{H}$ such that $\gamma_i(\mathcal{Z}_i^{x,\theta}, \theta) = x_i$, for each $i \in \mathcal{I}$, and for each $y \in \mathcal{X} \setminus \{x\}$ there exists j such that $y_j \in \mathcal{Z}_j^{x,\theta} \setminus \{x_j\}$. With $\mathcal{Y}_i^{x,\theta} = \{y \in \mathcal{X} \mid y_i \in \mathcal{Z}_i^{x,\theta}\}$, for each i, it is easy to check that $\mathcal{O} = \{\mathcal{Y}_i^{x,\theta} \mid i \in \mathcal{I},\ x \in f^{s.eff}(\theta),\ \theta \in \Theta\}$ is consistent with $f^{s.eff}$. Next, notice that if the allocation $x \in \mathcal{X}$ is such that $x_i = \gamma_i(\mathcal{H}, \theta)$ for all individuals i except at most one individual j, then the collection of opportunity sets $(\mathcal{Z}_i)_{i \in \mathcal{I}}$ where $\mathcal{Z}_i = \mathcal{H}$ for all $i \neq j$ and $\mathcal{Z}_j = \{x_j\}$ justifies why $x \in f^{s.eff}(\theta)$.[21] Hence $f^{s.eff}$ respects unanimity, $\mathcal{O}$ is strongly consisent with $f^{s.eff}$, and $f^{s.eff}$ is Nash implementable, by Proposition 2.b.

To conclude, I show that $f^{s.eff}$ is the maximal implementable extension of strong Pareto. Suppose, to the contrary, that there exist an implementable extension f, and a state θ such that some option x belongs to $f(\theta)$, but not to $f^{s.eff}(\theta)$. By Proposition 2.a, let $\mathcal{O}$ be a collection of opportunity sets (subsets of $\mathcal{X}$) that is consistent with f. Let $\mathcal{Z}_i = \{z_i \in \mathcal{H} \mid z \in O_i(x, \theta)\}$, for each $i \in \mathcal{I}$. Since $x \notin f^{s.eff}(\theta)$, there must exist $y \in \mathcal{X} \setminus \{x\}$ such that $y_i \notin \mathcal{Z}_i \setminus \{x_i\}$, for all $i \in \mathcal{I}$. Consider then any state θ' where each individual i is rational, ranking

[21] To check condition (b) in the definition of $f^{s.eff}$, notice that if $y \neq x$, then there are at least two individuals i and j such that $x_i \neq y_i$ and $x_j \neq y_j$.

y_i top best, and x_i second best (when $x_i \neq y_i$, which happens for at least two individuals since $y \neq x$). Notice that $x \in C_i(O_i(x,\theta),\theta')$, for all i, and hence $x \in f(\theta')$, a contradiciton since f extends strong Pareto and x is not strongly Pareto efficient at θ'. $\square$

Proof of Proposition 8

To show that $\hat{f}^{s.eff}$ is Nash implementable, consider θ, x with $x \in \hat{f}^{s.eff}(\theta)$. If $x \in f^{\pi}(\theta)$ for some enumeration π, then define $\mathcal{H}_i(x,\theta) = \{x_j \mid \pi(j) \geq \pi(i)\}$ and $O_i(x,\theta) = \{y \in \mathcal{X} \mid y_i \in \mathcal{H}_i(x,\theta)\}$. If $x \in \hat{f}^{s.eff}(\theta) \setminus f^{\pi}(\theta)$, then there exists j such that $x_i = \gamma_i(\mathcal{H},\theta)$, for all $i \neq j$. Define $O_i(x,\theta) = \mathcal{X}$, for all $i \neq j$, and $O_j(x,\theta) = \{x\}$. It is easy to check that the resulting collection $\mathcal{O}$ is consistent with $\hat{f}^{s.eff}$. Strong consistency and unanimity follow from the fact that any allocation y for which there exists j such that $y_i = \gamma_i(\mathcal{H},\theta)$, for all $i \neq j$, belongs to $\hat{f}^{s.eff}(\theta)$, by definition.

I now check that $\hat{f}^{s.eff}$ coincides with strong Pareto when individuals are rational. As is well-known, the set of allocations that are strongly Pareto efficient coincides with the set of outcomes from serial dictatorship in housing markets with rational individuals. It remains to show that any allocation x for which there exists j such that $x_i = \gamma_i(\mathcal{H},\theta)$, for all $i \neq j$, is strongly Pareto efficient. Suppose instead that there exists a feasible allocation y such that $y \neq x$ and y weakly Pareto improves upon x. Given that $y \neq x$ and y is feasible, there exists at least two individuals i and j such that $y_i \neq x_i$ and $y_j \neq x_j$. Hence both i strictly prefers y_i over x_i and j strictly prefers y_j over x_j, which contradicts the definition of x.

Finally, let f be an implementable extension of strong Pareto. I conclude the proof by showing that $f^{SD} \subseteq f$. Let $x = f^{\pi}(\theta)$ for some state θ and some enumeration π of $\mathcal{I}$. Consider a state θ' where individuals are rational with i strictly preferring object x_j over x_i, for all j such that $\pi(j) < \pi(i)$, and strictly preferring x_i over x_k, for all k such that $\pi(k) > \pi(i)$. The allocation x is strongly Pareto efficient at θ'. Suppose on the contrary that there exists a feasible allocation $y \neq x$ that Pareto improves upon x. Let then j be the individuals for which $\pi(i)$ is minimal among all i's who strictly prefers y_i over x_i at θ'. Notice that $y_k = x_k$ for all individual k who precedes j in π. Given j's preference at θ', him preferring y_j over x_i implies that there exists k with $\pi(k) < \pi(j)$ such that $y_j = x_k$. This contradicts the feasibility of y since both k and j consume the same object. This proves that x is strongly Pareto efficient at θ', and hence $x \in f(\theta')$, since f extends strong Pareto. Applying Proposition $2.a$, let $\mathcal{O}$ be a collection of implicit opportunity sets that is consistent with f. I now prove that $\{y_i \in \mathcal{H} \mid y \in O_i(x,\theta')\} = \{x_j \mid \pi(j) \geq \pi(i)\}$, for all $i \in \mathcal{I}$. The fact that the former set is contained in the latter follows at once from the fact that $x \in C_i(O_i(x,\theta'),\theta')$, for all $i \in \mathcal{I}$. To show the opposite inclusion, suppose on the contrary that there exist two individuals i and j such that $\pi(j) < \pi(i)$ and there is no $y \in O_i(x,\theta')$ for which $y_i = x_j$. Consider then θ'' where individuals are rational, with the same preferences as in θ', except that i ranks x_j top best. Clearly, $x \in C_k(O_k(x, \theta'), \theta'')$, for each $k \in \mathcal{I}$, and $x \in f(\theta'')$, by Proposition 2.a. Yet x is Pareto dominated by the allocation z derived from x by exchanging individuals i and j's objects: $z_i = x_j$, $z_j = x_i$, and $z_k = x_k$ for all $k \neq i, j$. This contradicts the fact that f extends strong Pareto. We have thus shown that $\{y_i \in \mathcal{H} \mid y \in O_i(x, \theta')\} = \{x_j \mid \pi(j) \geq \pi(i)\}$, for all $i \in \mathcal{I}$. By definition of x, $x_i = \gamma_i(\{x_j \mid \pi(j) \geq \pi(i)\}, \theta)$, for all $i \in \mathcal{I}$. Hence $x \in C_i(O_i(x, \theta'), \theta)$, for all i, and $x \in f(\theta)$, by Proposition 2.a. □

Extended Trading Cycle Procedure

Start by identifying a set $\mathcal{H}_1$ of objects which coincides with the set of objects picked by individuals whose endowment belongs to $\mathcal{H}_1$ and whose opportunity set is $\mathcal{H}$. To check that such a set exists, draw an arc from individual i to individual j if i's choice from $\mathcal{H}$ is j's endowment. Sets as $\mathcal{H}_1$ correspond to cycles of this graph, hence the reference to 'trading cycles.' The existence of a cycle follows from the fact that $\mathcal{I}$ is finite. Given such an $\mathcal{H}_1$, look for a set $\mathcal{H}_2$ of different objects which coincides with the set of objects picked by individuals whose endowment belongs to $\mathcal{H}_2$ and whose opportunity set is $\mathcal{H} \setminus \mathcal{H}_1$. By iteration, the extended trading cycle procedure delivers a partition $(\mathcal{H}_k)_{k=1}^K$ of $\mathcal{H}$ such that $\mathcal{H}_k = \{\gamma_i(\cup_{l=k}^K \mathcal{H}_l, \theta) \mid i \text{ such that } h_i^* \in \mathcal{H}_k\}$, for each k. Individual i consumes the object $\gamma_i(\cup_{l=k(i)}^K \mathcal{H}_l, \theta)$, where $k(i)$ is the index k such that $h_i^* \in \mathcal{H}_k$. When rational, individuals get to consume their most preferred object among those that remain at the time they exit in the procedure, hence the modifier 'top' used by Shapley and Scarf. If individual choices satisfy IIA, then the outcome of the procedure does not depend on which cycle is realized at each round of the procedure when multiple such cycles exist. In the absence of rationality, however, the order of elimination may matter, and the procedure may deliver multiple allocations.[22]

LEMMA 1: *An allocation is competitive if and only if it is an outcome of the extended trading cycle procedure.*

Proof: Let p be a competitive equilibrium whose associated outcome is x, i.e. $x_i = \gamma_i(B_i(p), \theta)$. Define $\mathcal{H}_1$ as the set of objects that are most expensive, $\mathcal{H}_2$ as the set of objects that are most expensive among remaining objects, etc. If an individual i's endowment belongs to $\mathcal{H}_k$, then $B_i(p) = \cup_{l=k}^K \mathcal{H}_l$ and $x_i = \gamma_i(\cup_{l=k}^K \mathcal{H}_l, \theta)$. For each $h \in \mathcal{H}_1$, there exists i such that $x_i = h$. Given that only individuals whose endowment belongs to $\mathcal{H}_1$ can afford objects in $\mathcal{H}_1$, it must be that $h_i^* \in \mathcal{H}_1$. Hence $\mathcal{H}_1 \subseteq \{x_i \mid h_i^* \in \mathcal{H}_1\}$. Given that both sets have the same cardinality, the inclusion must be an equality. Applying this argument inductively, it follows that $\mathcal{H}_k = \{x_i \mid h_i^* \in \mathcal{H}_k\}$, for each k, as desired.

[22] Bade (2008) offers a first analysis of this extended trading cycle procedure. She shows that resulting allocations (which she does not relate to competitive allocations) are efficient and belong to the core associated to Bernheim and Rangel's (2009) extended revealed preferences (which are larger sets than $f^{s.eff}$ and $f^{s.core}$ respectively).

I now prove that any allocation x obtained via the extended trading cycle procedure must be competitive. Let $(\mathcal{H}_k)_{k=1}^K$ be the partition of objects that emerges from the extended trading cycle procedure. Consider the price vector p, with $p(h) = K + 1 - k(h)$, where $k(h)$ is the index k such that $h \in \mathcal{H}_k$. Hence $B_i(p) = \cup_{l=k(h_i^*)}^K \mathcal{H}_l$, $x_i = \gamma_i(B(p), \theta)$, for each i, and $\mathcal{H}_k = \{x_i \mid i \text{ such that } h_i^* \in \mathcal{H}_k\}$, for each k. Hence x is a competitive allocation. □

Proof of Proposition 9

Existence follows from Lemma 1, since the extended trading cycle procedure delivers at least one allocation in each state.

As for multiplicity, consider the following housing market. Individuals suffer from a status-quo bias when they are overwhelmed by too many options: $\gamma_i(\mathcal{S}, \theta) = h_i^*$ when $\mathcal{S}$ contains m alternatives in addition to h_i^* (where m is a positive integer smaller than $|\mathcal{H}|$). Otherwise, they pick a choice by maximizing preferences. The allocation where all individuals keep their endowments is competitive, e.g. with all prices equal to one. Let now $\mathcal{J}$ be any set of at most m individuals, and let $x_{\mathcal{J}}$ be the unique competitive allocation that prevails when trade occurs only among members of $\mathcal{J}$. This allocation can be derived by applying Shapley and Scarf's top trading cycle procedure since individuals maximize their preferences when facing opportunity sets with fewer than m objects. The allocation y that coincides with $x_{\mathcal{J}}$ for members of $\mathcal{J}$, while other individuals consume their endowmenta, is competitive. Indeed, let $p_{\mathcal{J}}$ be prices supporting $x_{\mathcal{J}}$ for $\mathcal{J}$, and let $\bar{p}$ be a price that is larger than the maximal component of $p_{\mathcal{J}}$. Then y is supported by the price vector q, where $q_i = p_i$, for all $i \in \mathcal{J}$, and $q_i = \bar{p}$, for all $i \in \mathcal{I} \setminus \mathcal{J}$. Clearly, $x_{\mathcal{J}} \neq h_{\mathcal{J}}^*$ for many preference profiles, and hence it is possible to have multiple competitive allocations. □

Proof of Proposition 10

(a) To show that competitive allocations at θ belong to $f^{s.eff}(\theta)$, let x be such a competitive allocation, with associated prices p. Observe that $p(x_i) = p(h_i^*)$, for all i. Indeed, $p(x_i) \leq p(h_i^*)$, for all i, since $x_i \in B_i(p)$, from which the equality follows, since $\sum_{i\in\mathcal{I}} p(x_i) = \sum_{i\in\mathcal{I}} p(h_i^*)$ (given that $\{x_i \mid i \in \mathcal{I}\} = \{h_i^* \mid i \in \mathcal{I}\}$). Let then $\mathcal{Z}_i = B_i(p)$, for each i. Hence $x_i = \gamma_i(\mathcal{Z}_i, \theta)$, by definition. It remains to show that for each $y \in \mathcal{X} \setminus \{x\}$ there exists i such that $y_i \in \mathcal{Z}_i \setminus \{x_i\}$. Otherwise, there exists $y \in \mathcal{X} \setminus \{x\}$ such that for each i, either $y_i = x_i$, or $y_i \notin B_i(p)$. Let $\mathcal{J}$ be the set of individuals i such that $x_i \neq y_i$. Notice that $\mathcal{J}$ is non-empty since $x \neq y$. We get $\sum_{i\in\mathcal{I}} p(h_i^*) = \sum_{i\in\mathcal{I}\setminus\mathcal{J}} p(y_i) + \sum_{i\in\mathcal{J}} p(h_i^*) < \sum_{i\in\mathcal{I}} p(y_i) = \sum_{i\in\mathcal{I}} p(h_i^*)$, a contradiction. The first equality follows from $p(h_i^*) = p(x_i)$, as shown earlier, and $x_i = y_i$ for $i \in \mathcal{I} \setminus \mathcal{J}$. The inequality follows from $p(h_i^*) < p(y_i)$ for i such that $y_i \notin B_i(p)$. The last equality follows from $\{y_i \mid i \in \mathcal{I}\} = \{h_i^* \mid i \in \mathcal{I}\}$.

I now show that the extension of the first fundamental theorem of welfare economics does not hold with $\hat{f}^{s.eff}$ instead of $f^{s.eff}$. Consider a housing market with at least four individuals, and a state θ such that $\gamma_1(\mathcal{S},\theta) = h_2^*$ if and only if $\mathcal{S} = \mathcal{H}$ or $\{h_2^*\}$, $\gamma_2(\mathcal{S},\theta) = h_1^*$ if and only if $\mathcal{S} = \mathcal{H}$ or $\{h_1^*\}$, and each individual $i \geq 3$ is rational, ranking object h_1^* top best, h_2^* second best, and h_{i+1}^* third best (with the convention $I + 1 = 3$). By Lemma 1, there exists a unique competitive allocation x, with $x_1 = h_2^*$, $x_2 = h_1^*$, and $x_i = h_{i+1}^*$, for each $i \geq 3$. Yet $x \notin \hat{f}^{s.eff}(\theta)$. Suppose, by contradiction, first that $x = f^\pi(\theta)$, for some enumeration π of $\mathcal{I}$. No individual i with $i \geq 3$ can come first or second, as his or her preference would result in him or her consuming h_1^* or h_2^*, which is different from x_i. Then the second individual in the enumeration π (either 1 or 2) picks an object different than the one he or she is allocated under x. Next, there cannot be an individual j such that $x_i = \gamma_i(\mathcal{H},\theta)$ for all $i \neq j$, as at least one such i will be greater or equal to 3 and that individual picks h_1^* out of $\mathcal{H}$ instead of x_i. Hence, $x \notin \hat{f}^{s.eff}$, as desired.

(b) I start by proving that allocations in $\hat{f}^{s.eff}$ are competitive for some initial endowments. Consider a state θ and an allocation x such that $x = f^\pi(\theta)$, for some enumeration π of $\mathcal{I}$. Suppose then that i's endowment is object x_i. Then, x constitutes a competitive allocation which is supported by the price vector $p(x_i) = I + 1 - \pi(i)$, for each $i \in \mathcal{I}$. Suppose now that there exists j such that $x_i = \gamma_i(\mathcal{H},\theta)$, for all $i \neq j$. Then x constitutes a competitive allocation when j's endowment is x_j (other endowments can be set arbitrarily). Indeed, it is supported by the prices $p(x_j) = 1$ and $p(x_i) = 2$, for all $i \neq j$.

Next I show that there exist θ, x such that $x \in f^{s.eff}(\theta)$ and yet x is not competitive for any reallocation of endowments. Consider for instance a state where individuals $i = 1,\dots,I-1$ suffer from a status-quo bias when too many options are available: $\gamma_i(\mathcal{H},\theta) = h_i^*$, for all $i \leq I-1$. On the other hand, individuals' choices are derived from the maximization of a strict preference when opportunity sets are smaller, with individual i ranking h_{i+1}^* top best. To make the example as simple as possible, assume that any such individual is rational when facing no more than $I-1$ options. Individual I is fully rational, systematically maximizing a preference ordering that ranks h_1^* top best. The allocation x where $x_i = h_{i+1}^*$ for each i (with the convention $I+1 = 1$), belongs to $f^{s.eff}(\theta)$. To check this, take $\mathcal{Z}_i = \mathcal{H} \setminus \{h_1^*\}$, for all $i < I$, and $\mathcal{Z}_I = \mathcal{H}$. Notice that $\gamma_i(\mathcal{Z}_i,\theta) = h_{i+1}^*$, for each i. Take any $y \in \mathcal{X} \setminus \{x\}$. There must exist at least two individuals, i and k, such that $y_i \neq x_i$ and $y_k \neq x_k$. If either i or k equals I, then we have found an individual j such that $y_j \in \mathcal{Y}_j \setminus \{x_j\}$, since $\mathcal{Y}_I = \mathcal{H}$. If both $i \neq I$ and $k \neq I$, then at least one of them receives an object that is different from h_1^*. Again, we have found an individual j such that $y_j \in \mathcal{Y}_j \setminus \{x_j\}$, and we have thus proved that $x \in f^{s.eff}(\theta)$. On the other hand, there is no allocation of endowments that make x competitive. Suppose on the contrary that the allocation of endowments w and the price vector p makes x competitive. Some object(s) must be among those that are most expensive, and individual(s) owning it (or them) under w can afford any object. Given that $x_i \neq h_i^*$ for each $i < I$, there must be some objects that i cannot afford, and hence h_i^*'s price is not maximal. Hence $p(h_I^*) > p(h_i^*)$, for all $i < I$. This contradicts $x_{I-1} = h_I^*$, since $I-1$ cannot afford this object. □

PROOF OF PROPOSITION 11

To show that $f^{s.core}$ extends the strong core, suppose that each individual choice function $\gamma_i(\cdot,\theta)$ is rational at θ. For each i, let $\mathcal{W}_i$ be the union of $\{x_i\}$ with the (strict) lower contour set (in $\mathcal{H}$) of x_i for i. By definition, x belongs to the strong core at θ if and only if there is no coalition $\mathcal{S}$ and no allocation $y \in \mathcal{F}(\mathcal{S}) \setminus \{x_{\mathcal{S}}\}$ such that for each $j \in \mathcal{S}$, $y_j = x_j$ or y_j belongs to the (strict) upper contour set of x_j for j. This, in turn, is equivalent to: for all coalition $\mathcal{S}$, if $y \in \mathcal{F}(\mathcal{S}) \setminus \{x_{\mathcal{S}}\}$, then $y_j \in \mathcal{W}_j \setminus \{x_j\}$, for some j. This is also equivalent to the existence of a collection $(\hat{\mathcal{W}}_i)_{i\in\mathcal{I}}$ of subsets of $\mathcal{H}$ justifying why $x \in f^{s.core}(\theta)$, as desired. The necessary condition is trivial - simply take $\hat{\mathcal{W}}_i = \mathcal{W}_i$ for each i. For the sufficient condition, observe that $y_j \in \mathcal{W}_j \setminus \{x_j\}$ if $y_j \in \hat{\mathcal{W}}_j \setminus \{x_j\}$, since $\hat{\mathcal{W}}_j \subseteq \mathcal{W}_j$ (given that $x_j = \gamma_j(\hat{\mathcal{W}}_j,\theta)$).

Let's check that $f^{s.core}$ is Nash implementable in all housing markets. By definition, one can associate to any $x \in f^{s.core}(\theta)$ a collection $(\mathcal{W}_i^{x,\theta})_{i\in\mathcal{I}}$ of subsets of $\mathcal{H}$ such that $\gamma_i(\mathcal{W}_i^{x,\theta},\theta) = x_i$, for each $i \in \mathcal{I}$, and for each $\mathcal{S} \subseteq \mathcal{I}$ and each $y \in \mathcal{F}(\mathcal{S}) \setminus \{x_{\mathcal{S}}\}$, there exists $j \in \mathcal{S}$ for which $\mathcal{W}_i^{x,\theta} \setminus \{x_j\}$ contains y_j. Define $O_i(x,\theta)$ as $\{x\}$ union the set of $y \in \mathcal{X}$ such that $y_i \in \mathcal{W}_i^{x,\theta}$, $y_j = h_i^*$ where j is the individual such that $y_i = h_j^*$, and $y_k = h_k^*$, for all $k \neq i,j$. The set $O_i(x,\theta)$ is constructed so that the set of objects that i can get by picking elements of $O_i(x,\theta)$ is precisely $\mathcal{W}_i^{x,\theta}$. In addition, elements y of $O_i(x,\theta)$ that are different from x are constructed so as to minimize the number of trades needed to give y_i to individual i.

By construction, $\mathcal{O} = \{O_i(x,\theta) \mid i \in \mathcal{I},\ x \in f^{s.core}(\theta),\ \theta \in \Theta\}$ is consistent with $f^{s.core}$. Next, notice that if the allocation $x \in \mathcal{X}$ is such that $x_i = \gamma_i(\mathcal{H},\theta)$ for all i, then the collection of opportunity sets $(\mathcal{W}_i)_{i\in\mathcal{I}}$ where $\mathcal{W}_i = \mathcal{H}$ for all i justifies why $x \in f^{s.core}(\theta)$. Hence $f^{s.core}$ respects unanimity. To show that $\mathcal{O}$ is strongly consistent with $f^{s.core}$, consider an allocation $x \in \mathcal{X}$ for which $x_i = \gamma_i(\mathcal{H},\theta)$ for all individuals $i \neq j$, and $x \in C_j(O_j(x',\theta'),\theta_j)$ for some $x' \in f^{s.core}(\theta')$. I check that $x \in f^{s.core}(\theta)$, using the collection $(\mathcal{W}_i)_{i\in\mathcal{I}}$ of implicit opportunity sets defined by $W_i = \mathcal{H}$, for $i \neq j$, and $\mathcal{W}_j = \mathcal{W}_j^{x',\theta'}$. Otherwise, there exist a coalition $\mathcal{S}$ and $y \in \mathcal{F}(\mathcal{S}) \setminus \{x_{\mathcal{S}}\}$ such that $y_i \notin \mathcal{W}_i \setminus \{x_i\}$, for all $i \in \mathcal{S}$. This is possible only if $j \in \mathcal{S}$, $y_i = x_i$ for all $i \in \mathcal{S} \setminus \{j\}$, $y_j \neq x_j$, $y_j \notin \mathcal{W}_j^{x',\theta'}$, and $x \neq x'$. Then $x_l = h_j^*$ for the individual l such that $x_j = h_l^*$, and $x_k = h_k^*$, for all $k \in \mathcal{S} \setminus \{k,l\}$, since $x \in C_j(O_j(x',\theta'),\theta)$. Feasibility of y, which coincides with x for members of $\mathcal{S}$ different from j, implies that $y_j = h_l^* = x_j$, a contradiction. Hence $\mathcal{O}$ is strongly consisent with $f^{s.core}$, and it is Nash implementable, by Proposition 2.b.

Finally, I show that $f^{s.core}$ is the maximal implementable extension of the strong core. Suppose, to the contrary, that there exist an implementable extension f, and a state θ such that some option x belongs to $f(\theta)$, but not to $f^{s.core}(\theta)$. By Proposition 2.a, let $\mathcal{O}$ be a collection of opportunity sets (subsets of $\mathcal{X}$) that is consistent with f. Let $\mathcal{W}_i = \{z_i \in \mathcal{H} \mid z \in O_i(x,\theta)\}$, for each $i \in \mathcal{I}$. Since $x \notin f^{s.core}(\theta)$, there must exist $\mathcal{S} \subseteq \mathcal{I}$ and $y \in \mathcal{F}(\mathcal{S})\backslash\{x_{\mathcal{S}}\}$ such that $y_i \notin \mathcal{W}_i\backslash\{x_i\}$, for all $i \in \mathcal{S}$. Consider then any state θ' where each individual i is rational, with an underlying strict preference that ranks x_i top if $i \in \mathcal{I}\backslash\mathcal{S}$, and ranks y_i top best, and x_i second best (when $x_i \neq y_i$, which happens for at least one individual since $y \neq x$) if $i \in \mathcal{S}$. Notice that $x \in C_i(O_i(x,\theta),\theta')$, for all i, and hence $x \in f(\theta')$. Yet x does not belong to the strong core at θ' (since all the members of $\mathcal{S}$ weakly prefer y over x, some with a strict preference), thereby contradicting the fact that f extends strong Pareto. $\square$