Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects*

Geoffroy de Clippel and Kfir Eliaz†

December 2010

Abstract

Among the most important and robust violations of rationality are the attraction and the compromise effects. The compromise effect refers to the tendency of individuals to choose an intermediate option in a choice set, while the attraction effect refers to the tendency to choose an option that dominates some other options in the choice set. This paper argues that both effects may result from an individual's attempt to overcome the difficulty of making a choice in the absence of a single criterion for ranking the options. Moreover, we propose to view the resolution of this choice problem as a cooperative solution to an *intra*-personal bargaining problem among different selves of an individual, where each self represents a different criterion for choosing. We first identify a set of properties that characterize those choice correspondences that coincide with our bargaining solution, for some pair of preference relations. Second, we provide a revealed-preference foundation to our bargaining solution and characterize the extent to which these two preference relations can be uniquely identified.

Alternatively, our analysis may be reinterpreted as a study of (*inter*-personal) bilateral bargaining over a finite set of options. In that case, our results provide a new characterization, as well as testable implications, of an ordinal bargaining solution that has been previously discussed in the literature under the various names of fallback bargaining, unanimity compromise, Rawlsian arbitration rule and Kant-Rawls social compromise.

*We thank Bart Lipman, Mihai Manea, Roberto Serrano, and two anonymous referees for helpful comments. We also thank Jim Campbell for his research assistance. Financial assistance from NSF grant SES-0851210 and from the C.V. Starr Program in Commerce, Organizations, and Entrepreneurship is gratefully acknowledged.

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

1. INTRODUCTION

Many of the decision problems we face are complicated by the fact that there is no single dimension or criterion for evaluating the available alternatives. For example, when searching for an apartment or a house, the ranking of the available options may be very different depending on whether the criterion we use is price, size, proximity to work or quality of schools. Similarly, when choosing a car, there are several different criteria or dimensions that one may use such as price, safety, gas efficiency, size, color or esthetics. Also, in deciding between academic job offers there is no one obvious criterion to use as one may consider the ranking of the department, the number of faculty members in one's field, the financial terms, the location, etc. Often there can be many different dimensions or criteria that one may use, making it difficult, if not impossible, to take all of them into account. This often leads us to focus only on a limited number of dimensions, which we deem most important. However, we are still faced with the difficult task of resolving the trade-off between these dimensions.

Evidence from numerous studies in economics, psychology and marketing suggests that individuals often find it difficult to resolve the conflict about how much of one dimension to trade off in favor of another, and typically tend to resort to simple heuristics that lead to systematic violations of rationality. One common heuristic is known as "reason-based choice" (see Simonson (1989), Tversky and Shafir (1992) and Shafir, Simonson and Tversky (1993)): in the absence of a single criterion for ranking available options (what is often referred to as "choice under conflict"), choices may be explained "in terms of the balance of reasons for and against the various alternatives" (see Shafir, Simonson and Tversky (1993)). According to this heuristic, "relations among alternatives in choice sets may influence choice by providing reasons for preferring certain alternatives over others" (Simonson (1989)). Consequently, reason-based choice may lead to systematic violations of the Weak Axiom of Revealed Preferences (WARP).

Among the most studied and robust violations are the *attraction* and the *compromise* effects. The attraction effect was first demonstrated by Huber, Payne and Puto (1982), while the compromise effect was introduced by Simonson (1989).[1] The attraction effect refers to the ability of an asymmetrically dominated or relatively inferior alternative, when added to a set, to increase the choice probability of the dominating alternative. The compromise effect refers to the ability of an "extreme" (but not inferior) alternative,

[1]These studies have sprung a whole literature devoted to replicating and extending these effects to various decision problems, including real, monetary choices. For references see Shafir, Simonson and Tversky (1993), Kivetz, Netzer and Srinivasan (2004) and Ariely (2008).

when added to a set, to increase the choice probability of an "intermediate" alternative. To illustrate these two effects, consider two options, A and B. Suppose there are two dimensions or criteria for evaluating these options such that B is better than A along the first dimension while A is better than B along the second dimension (see Figure 1). For example, suppose A and B are two equally priced apartments, but one is closer to work while the other has better schools.

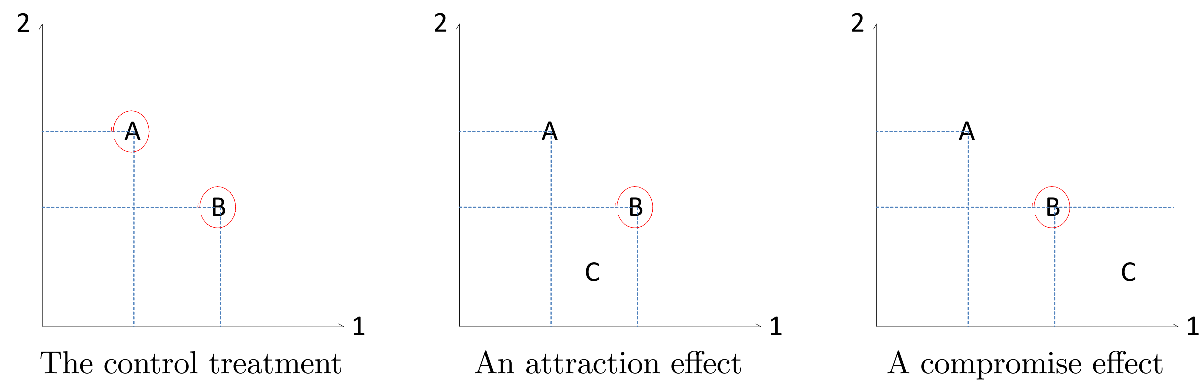

Figure 1

In a typical experimental study, both A and B are chosen - usually in equal proportions - by a control group of subjects. The attraction effect is observed when a third alternative, C, is added to the set such that it is dominated by only one of the other two options (say, B, as in Figure 1). When subjects are asked to choose from $\{A, B, C\}$, the vast majority of them tend to choose B. The compromise effect occurs when C is added such that it is even better than B along the first dimension but worse than it along the second dimension (i.e., according to the first dimension, C is better than B, which is better than A, while the opposite ranking is obtained according to the second dimension). In such a case, most subjects again tend to pick B. These findings may be interpreted as systematic violations of the Weak Axiom of Revealed Preferences (WARP) by considering a choice correspondence that selects both A and B from $\{A, B\}$, but chooses B alone from $\{A, B, C\}$.[2] The introduction of these two effects has generated a huge literature in marketing aimed at understanding the source of the effects and their implications for positioning, branding and advertising (see Kivetz, Netzer, and Srinivasan (2004)). The common consensus among psychologists is that the attraction and compromise effects are not two separate phenomena, but rather two manifestations of the same heuristic, mainly, reason-based choice. Indeed, in the seminal study of Simonson (1986), the *same*

[2] More specifically, this is a violation of the β-axiom proposed by Sen (1971).

group of subjects exhibited both effects in roughly the same magnitude.

This paper proposes and characterizes a model of reason-based choice that generates both the attraction and the compromise effects. We envision the decision-maker as trying to reach a compromise between conflicting "inner selves", representing the different attributes or dimensions of the available options. We then propose to view the final choice (i.e., the "balancing of reasons for and against") as a *cooperative* solution to a bargaining among the different selves. In the spirit of the literature on dual-selves (e.g., the $\beta - \delta$ models of present bias, Benhabib and Bisin (2004), Bernheim and Rangel (2004), Eliaz and Spiegler (2006), Fudenberg and Levine (2006)), most of our analysis is focused on behavior that may be explained with at most two selves.

We start by considering the two relevant criteria or dimensions, and their associated rankings (which may be identical), as primitives of the model (e.g., think of choosing among products with two attributes such as price and quality, price and size, shipping rate and date of arrival, sugar and fat content, etc). Formally, our first model consists of a finite set of options X and a pair of linear orderings on this set, $\succ = (\succ_1, \succ_2)$. Each ordering is interpreted as the (known) preference relation of one of the individual's dual selves. A bargaining problem is defined to be a non-empty subset of options S. For a given preference profile $\succ$, a bargaining solution is a correspondence $C_\succ$ that associates with every bargaining problem S a subset of S.

Which cooperative bargaining solution can capture our dual-self interpretation of reason-based choice? Our first main result (Theorem 1) establishes the existence of a unique bargaining solution that capture - and extend - the attraction and compromise effects (properties we call "Attraction" and "No Better Compromise"), in addition to a number of other properties that capture a notion of consistency across decision problems, the cooperative nature of the bargaining, immunity to framing and symmetry. To describe this solution, imagine that for every bargaining problem, each bargainer assigns each option a score equal to the number of elements in its lower contour set. Hence, each option is associated with a pair of scores. The bargaining solution selects the options whose minimal score is highest. This solution has been previously discussed in the literature under various names: "Rawlsian arbitration rule" (Sprumont (1993)), "Kant-Rawls Social Compromise" (Hurwicz and Sertel (1997)), "fallback bargaining" (Brams and Kilgour (2001)), as well as "unanimity compromise" (Kibris and Sertel (2007)). In contrast to the Nash or Kalai-Smordinsky solutions, this bargaining solution is purely ordinal and applies to any arbitrary finite set of options.[3]

[3] Mariotti (1998) proposes an extension of the Nash bargaining solution to finite environments. How-

Next we consider an environment in which there is no obvious way to rank the options along two dimensions. We interpret our focus on only two dimensions as an assumption that the decision-maker can process only a limited number of dimensions or attributes. Thus, if the options are characterized by a large number of attributes, it may not be clear which two dimensions the decision-maker focuses on. Hence, an outside observer may not be able to infer what rankings the decision-maker uses to evaluate the options. Alternatively, there may be only two salient dimensions or attributes, but it is not obvious how a decision-maker would rank the options along each dimension (consider, for example, attributes such as color, taste, smell). In such an environment the only observations we may have about the decision-maker are the choices he makes (i.e., his choice correspondence). We ask the following question: what are the necessary and sufficient conditions for representing the decision-maker *as if* he has two selves (each characterized by a linear ordering on X), which make a choice according to the fallback bargaining solution?

Our second main result (Theorem 2) identifies these conditions. This result relies on the notions of "revealed Pareto dominance" and "revealed compromises". An option x is revealed to be Pareto superior to y if it is chosen over y in a pairwise comparison. An option y is revealed to be a compromise between x and z if no option in this triplet is revealed to be Pareto superior over another, and y is chosen uniquely from $\{x, y, z\}$. The necessary and sufficient conditions identified in Theorem 2 include the revealed versions of the relevant properties characterized in Theorem 1, in addition to properties that capture the consistency of the revealed Pareto relation and the consistency of revealed compromises.

We next address the question of "identifiability": to what extent can we identify the set of preference profiles that are compatible with the observed choices? Clearly, exchanging the rankings between the two selves does not affect the bargaining solution. Theorem 3 argues that there is a sense in which any further multiplicity is with respect to "irrelevant alternatives". This means that for any given bargaining problem S, we can pin down the pair of preferences over the minimal set of options that Pareto dominate any option outside this set.

So far, we have interpreted our choice procedure as a solution to an *intra*-personal bargaining problem. Alternatively, we may interpret it as a solution to an *inter*-personal bargaining problem where two distinct individuals need to agree on an option. While most of the choice theoretic literature aims to characterize testable implications of models of *individual* decision-making, the same set of tools may be applied to models of

ever the extended solution still uses cardinal information as it is defined over sets of payoff vectors.

collective decision-making. Since many collective decisions are achieved through bargaining, it seems important to identify the necessary and sufficient conditions for inferring the bargainers' preferences and for modelling their decisions as an outcome of cooperative bargaining. This paper takes a first step in this direction by studying situations in which two individuals bargain over some finite, arbitrary set of alternatives. We, therefore, focus on *ordinal* bargaining solutions on finite domains. Among such solutions, the fallback bargaining solution has received much attention in the literature. Moreover, this solution has a simple non-cooperative foundation, which is similar in spirit to Rubinstein's (1982) alternative-offer game. Theorems 2 and 3 then provide testable implications of the fallback solution and characterize the extent to which the bargainers' preferences may be recovered from the data.

The rest of the paper is organized as follows. The related literature is discussed in the next section. Section 3 defines the basic concepts and notation. This is followed by an axiomatic characterization of the fallback solution for known preferences in Section 4. The revealed-preference analysis of this solution is presented in Section 5. Finally, Section 6 discusses possible extensions and provides some concluding remarks.

2. RELATION TO THE LITERATURE

In relation to the literature, our paper makes the following contributions. First, we propose a *single* model that "explains" both the attraction and the compromise effects and characterize its testable implications. Second, we provide a revealed-preference foundation for a dual-self model in which the selves strive to reach compromise rather than to behave non-cooperatively. Third, our axiomatic characterization also provides a revealed-preference foundation for a cooperative bargaining solution. To better assess these contributions, we discuss below some of the related papers in the literature.

Explaining attraction and compromise

A number of recent papers have proposed formal models that explain either the attraction effect or the compromise effect. However, there is no single model in this literature that generates both effects in a single-person decision problem (such as those encountered in the experiments that document these effects). Ok, Ortoleva and Riella (2008) relax the Weak Axiom of Revealed Preferences to allow for choice behavior that exhibits the attraction effect, but *not the compromise effect.* They propose a reference-dependent choice model in which given a choice problem S, the decision-maker maximizes a real function u

over those options that Pareto dominate a reference point $r(S)$ according to a sequence of utility functions $\mathbf{u}$. This choice procedure may be interpreted as a bargaining problem with a *continuum* of bargainers and a disagreement point $r(S)$, where the solution maximizes a social welfare function (SWF) u over the set of options that are "individually rational". The authors characterize necessary and sufficient conditions on choice data to be consistent with some bargaining model $(r, \mathbf{u}, u)$. One of these conditions, labeled "reference-dependent WARP", rules out the compromise effect.[4]

The attraction effect was also addressed in Lombardi (2007), which axiomatizes the following choice procedure. Given a set of options, the decision-maker first removes elements that are dominated according to a fixed, possibly incomplete, preference relation. From the remaining options, the decision-maker eliminates those alternatives, whose lower contour set is strictly contained in that of another remaining option. Unlike our model, this choice procedure does not explain the compromise effect and may end up choosing the entire choice set.

In the marketing literature, Kivetz, Netzer and Srinivasan (2004a,b) argue that individuals may exhibit the compromise effect when choosing among multi-attribute options because of the rule they use to aggregate the different subjective values they assign to the attributes. The authors propose several functional forms of aggregation rules that can generate the compromise effect and test the predictions of these functions on experimental data. Their framework, unlike ours, is cardinal in nature and assumes a particular structure on the set of options. Depending on parameter values, and on the distance between options, some of the functions proposed by Kivetz et al. (2004a,b) may also generate some instances of an attraction effect. In contrast, our decision-making model always generates both attraction and compromise effects for any arbitrary, finite set of options. Furthermore, we characterize the testable implications of our model and the necessary and sufficient conditions for identifying its primitives from choice data.

Kamenica (2008) provides a novel argument that in a monopolistic market with enough uninformed but *rational* consumers, there are some conditions that guarantee the existence of equilibria in which the uninformed consumers exhibit the compromise effect, or the attraction effect, with positive probability. While this argument suggests an original interpretation of why consumers in a market may exhibit compromise/attraction-like

[4] To see this, recall that the compromise effect means that whenever the choice out of any pair in $\{x, y, z\}$ is the pair itself, then only a single element will be chosen from the triplet. Suppose y is chosen. If the choice correspondence satisfies "reference-dependent WARP" then either x or z act as a "potential reference point" for y, meaning that y must be chosen uniquely from $\{x, y\}$ or from $\{y, z\}$, a contradiction.

behavior, there are many instances - such as the numerous experiments that document the compromise and attraction effects - in which individuals consistently exhibit these effects outside the market when they are not engaged in a non-cooperative game against some seller.

Rationalization via aggregation of multiple rationales

A few recent papers have proposed to model systematic violations of IIA as the result of a choice procedure that aggregates multiple orderings ("rationales") on the set of alternatives. Ambrus and Rozen (2009) investigate what choice functions are rationalizable with a given social welfare function and a given number of selves. Green and Hojman (2007) propose a welfare criterion for evaluating irrational choices, by modeling these choices as reflecting a weighted aggregation of all possible strict orderings on the set of options. In contrast to us, these studies are not concerned with deriving testable implications and focus on a different set of questions than we do.[5]

Testable implications of collective decision-making

Finally, our paper is related to a small but growing literature that aims to provide testable implications for models of collective decision-making. Among those papers that employ a revealed-preference methodology, the most closely related are Sprumont (2000) and Eliaz, Richter and Rubinstein (2009). The former provides a choice theoretic characterization of Nash equilibrium and the Pareto correspondence, while the latter characterizes the choice correspondence that selects the top element(s) of two preference orderings.

A number of other papers explore similar questions but without employing a revealed-preference methodology. Chiappori (1988) characterizes the conditions under which it is possible to recover the preferences and decision process of two individuals, who consume leisure and some Hicksian composite good, from observations on their labor supply functions. Chiappori and Ekeland (2006) extend this analysis and characterize the necessary and sufficient conditions for recovering the individual preferences of a group of individuals from observations on their aggregate consumption and the common budget constraint that they face. Chiappori and Donni (2005) analyze the testable implications

[5] There is also a number of choice theoretic papers that proposes to rationalize irrational behavior using procedures that rely - but do not aggregate - on multiple (not necessarily complete) binary relations. See, e.g., Kalai, Rubinstein and Spiegler (2002), Manzini and Mariotti (2007) and Cherepanov, Feddersen and Sandroni (2008).

of the Nash bargaining solution in an environment where two individuals need to agree on the allocation of a pie among themselves and where disagreement leads each to receive some reservation payment. In a similar vein, Chambers and Echenique (2008) study the testable implications of the standard model of two-sided markets with transfers and characterize the sets of matchings that may be generated by the model.

3. DEFINITIONS

X will denote the finite set of all potential options. A *bargaining problem* is a subset of X. A *bargaining solution* C associates to each bargaining problem S a nonempty subset $C(S)$ of S. A (strict) *linear ordering* on X is a relation defined on $X \times X$ that is complete, transitive, and anti-reflexive. The set of all possible linear orderings is denoted $L(X)$.

Let $\succ = (\succ_1, \succ_2) \in L(X)^2$, and let S be a bargaining problem. The *score* of x in S along dimension i ($i = 1$ or 2) is the number of feasible options that are (strictly) worse than x for $\succ_i$:

$$s_i(x, S, \succ) = |\{y \in S | x \succ_i y\}|$$

The *fallback bargaining solution* $C^f_\succ$ associated with $\succ$ assigns to each bargaining problem S the set of options in S that maximize the minimum (over $i = 1, 2$) of the scores:

$$C^f_\succ(S) = \arg\max_{x \in S} \min_{i=1,2} s_i(x, S, \succ).$$

Our first characterization of the fallback bargaining will involve regularity conditions relating various bargaining solutions which can be indexed by a pair of linear orderings as these two underlying preferences change. A *bargaining operator* is a function $\mathcal{C}$ that associates a bargaining solution $\mathcal{C}(\succ)$ to each pair $\succ = (\succ_1, \succ_2)$ of linear orderings on X. The image $\mathcal{C}(\succ)$ of the bargaining operator associated to $\succ$ will be denoted $C_\succ$ from now on, and the fallback bargaining operator will be denoted $\mathcal{C}^f$.

As pointed out in the introduction, the fallback bargaining solution has already appeared under various names in the literature on interpersonal bargaining (Sprumont (1993), Hurwicz and Sertel (1997), Brams and Kilgour (2001), Kibris and Sertel (2007)). The terminology of "fallback bargaining" is taken from Brams and Kilgour (2001), where they offer a nice reinterpretation of the solution. For each bargaining problem S, and each integer between 1 and $|S|$, let $E_i(S, k)$ be the set of k best options in S according to i's preferences. Let k^* be the smallest k such that $E_1(S,k) \cap E_2(S,k) \neq \emptyset$. Then $C^f_\succ(S) = E_1(S,k^*) \cap E_2(S,k^*)$.

In other words, if both criteria agree on what the best option is, then it is the solution. Otherwise, the decision-maker looks for option(s) that would be ranked either top or second-best by both criteria. If no option satisfies this property, then the decision-maker iterates the procedure by allowing for third-best alternatives, and so on so forth. This simple algorithm for deriving the elements in the solution illustrates the appeal of the fallback solution as a descriptive model of multi-criteria decision making.

Figure 2 illustrates how the fallback solution generates the attraction and compromise effects.

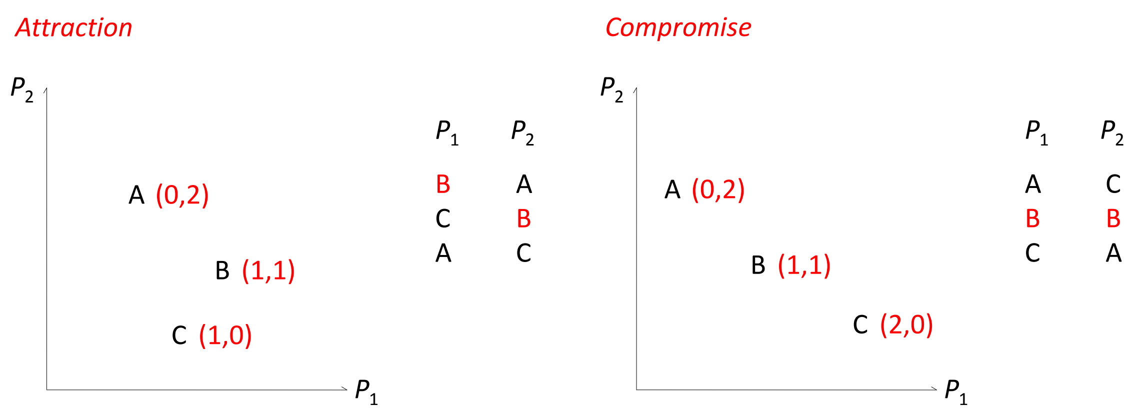

Figure 2

In both cases, both A and B would get a minimal score of 0 if C was not available. Adding C changes the scores, and B now gets the largest minimal score in both cases. It thus becomes selected uniquely by the fallback solution.

It is also interesting to note that in the spirit of the Nash program, fallback bargaining has a non-cooperative foundation. The two bargainers alternate in proposing one of the available options as a possible agreement. If the responder accepts, the game ends and the proposed option is adopted. If the responder rejects, the proposed option is removed from the set, and the responder now proposes one of the remaining options. The game continues until either an agreement is reached or there remains only a single option, which is then adopted. Anbarci (2006) shows that the unique subgame-perfect equilibrium of this game is an element in $C^f_\succ(S)$.

The fallback solution applies an egalitarian criterion to a canonical representation of the ordinal preferences. It is interesting to think about applying a utilitarian criterion, which selects the set of elements,

$$\arg\max_{x\in S}[s_1(x,S,\succ)+s_2(x,S,\succ)]$$

Note that this solution has two important shortcomings. First, it selects all the elements of S whenever they are all Pareto optimal for the pair of orderings $(\succ_1,\succ_2)$, and hence does not capture the compromise effect. Second, in contrast to the fallback solution, the Borda rule is not robust to common monotonic transformations of the bargainers' ordinal preferences in the sense that it is sensitive to how the score of an option changes as it moves up in the ranking (cf. "scoring rules").

4. PREFERENCE-BASED AXIOMATIC CHARACTERIZATION

The aim of this section is to establish that the fallback bargaining operator is the unique operator to captures a number of desiderata. First, its associated bargaining solutions should exhibit properties that capture a plausible interpretation of attraction and compromise. Second, we should be able to interpret those solutions as a "procedurally-rational" heuristic. Thus, they should exhibit some form of consistency across decision/bargaining problems. Third, the solutions should capture our idea that the bargaining among the selves is in some sense "cooperative". Finally, we wish to interpret all the options selected by the solutions (i.e., any "agreement" reached by the two selves) as being on "equal footing" in terms of their desirability and robustness to small changes in the bargaining prolem.

We focus our attention on a class of bargaining operators, which satisfy some basic properties from axiomatic bargaining and social choice. This would allow us to meaningfully interpret the associated correspondences $C_\succ$ as a bargaining solution.[6] Specifically, a bargaining operator $\mathcal{C}$ is *regular* if

1. it is *neutral* in the sense of not having an a priori bias in favor or against some elements of X. Let $g: X\to X$ be an isomorphism. Then $C_{g(\succ)}(g(S))=g(C_\succ(S))$, where $g(S)=\{g(x)|x\in S\}$ and $g(\succ)\in L(X)^2$ is such that $xg_i(\succ)y$ if and only if $g^{-1}(x)\succ_i g^{-1}(y)$, for all $x,y\in X$ and both $i\in\{1,2\}$.

2. it is *anonymous* in the sense that it treats both orderings with equal relevance: $C_{(\succ_2,\succ_1)}(S)=C_{(\succ_1,\succ_2)}(S)$.

[6] Recall that we denote the image of $\succ$ under the bargaining operator $\mathcal{C}$ by $C_\succ$ instead of $\mathcal{C}(\succ)$.

3. options are selected using only the parts of the two orderings that are relevant to the problem. If $\succ'$ is an alternative pair of linear orderings (defined on X) that coincide with $\succ$ on $S \times S$, then $C_{\succ'}(S) = C_{\succ}(S)$.[7]

It is certainly of interest to investigate how our theory would adapt if one eliminates some or all of these properties. Dropping neutrality would allow to accomodate some framing effects, where the label of the available options may influence the choice (e.g. options are presented in a list, or are offered by trademarks with varying impact, etc.). Dropping the second property would add the possibility of having one of the two criteria as being more relevant than the other (e.g. caring more about the size of the car than its color). Dropping the third property would allow to consider choice procedures where the decision maker is influenced by options he aspires to, but cannot afford. Yet, we believe that one must first understand the attraction and compromise effects in their purest form, in absence of all these additional features. The regularity property thus defines a benchmark which can be used to build more elaborate theories.

Contrary to the regularity conditions, our main axioms do not involve restrictions on the bargaining operator as the underlying pair of preferences change. Hence, they will be imposed for all the bargaining solutions in the image of the operator. They are thus assumed to be valid for each $\succ \in L(X)^2$, and each $S \subseteq X$.

Attraction (ATT) - *Let $x \in X \setminus S$ be such that $y \succ x$, for some $y \in C_{\succ}(S)$. Then $C_{\succ}(S \cup \{x\}) = \{y \in C_{\succ}(S) | y \succ x\}$.*[8]

ATT formalizes the idea that adding a dominated alternative reinforces the appeal of an option to the decision maker. This property is best understood by decomposing it into two parts. First, whenever option x is added to a set S, it seems reasonable to expect that the set of options that were previously selected, and which dominate x, should continue to be chosen, i.e., $\{y \in C_{\succ}(S) | y \succ x\} \subseteq C_{\succ}(S \cup \{x\})$. We view the attraction effect as the converse inclusion, $C_{\succ}(S \cup \{x\}) \subseteq \{y \in C_{\succ}(S) | y \succ x\}$, i.e., when choosing from the new set, one's attention is drawn to the previously selected options that dominate x. Thus, the solution to the enlarged problem obtained by adding x as a feasible option should be the intersection of the solution to the original problem with those options that Pareto

[7] Similar properties have been used repeatedly in the classical theories of bargaining and social choice (see, for example, Karni and Schmeidler (1975)).

[8] $\succ$ refers to the Pareto relation (incomplete ordering on $X \times X$) when comparing options, i.e. $x \succ y$ means $x \succ_1 y$ and $x \succ_2 y$. On the other hand, the symbol $\succ$ in $C_{\succ}$ refers to the pair $(\succ_1, \succ_2)$ of linear orderings on X. We do not introduce different symbols because the right meaning is always obvious when used in context.

dominate x whenever that set is nonempty.

No Better Compromise (NBC) - *If both x and y belong to $C_{\succ}(S)$, then there does not exist $z \in S$ and $i \in \{1,2\}$ such that $x \succ_i z \succ_i y$ and $y \succ_{-i} z \succ_{-i} x$.*

NBC captures the idea that the bargainers are trying to reach a compromise. If two bargainers were not able to agree on a single option - so that both x and y are identified as possible agreements - then it must be that there was no option z available that could serve as a compromise between x and y. By this we mean that there was no alternative z that falls "in between" x and y, in that it is better than x along the dimension where it is worse that y, and better than y along the dimension where it is worse than x.[9]

Removing an Alternative (RA) - *If $C_{\succ}(S) \neq \{x\}$, then $C_{\succ}(S \setminus \{x\}) \cap C_{\succ}(S) \neq \emptyset$.*

RA captures the sense in which the bargaining solutions associated to the bargaining operator may be interpreted as a procedurally-rational heuristic. Since both ATT and NBC are typically incompatible with WARP, we propose a weaker consistency property. If an option (that is not the unique choice of the decision maker) is dropped, then at least one of the options that were chosen in the original problem belongs to the solution of the reduced problem. Observe that RA is equivalent to IIA if the bargaining solution is single-valued, as RA can be applied iteratively if one needs to eliminate multiple irrelevant alternatives. Yet, moving to correspondences, the slight difference between the two properties when eliminating a single alternative can lead to major differences in terms of choices. In addition, RA also expresses some form of continuity in our discrete setting. Indeed, making a small change in the set of available options (i.e. dropping only one alternative) should not modify too much the set of selected elements (i.e. nonempty intersection) whenever this set is not a singleton.

Efficiency (EFF) - *If $x \in C_{\succ}(S)$, then there does not exist $y \in S$ such that $y \succ x$.*

EFF captures the cooperative nature of the bargaining. It is also a standard property in axiomatic bargaining and social choice.

Symmetry (SYM) - *If $x, y \in C_{\succ}(S)$ and there exists $z \in S \setminus \{x, y\}$ such that $x \notin C_{\succ}(S \setminus \{z\})$, then there exists $z' \in S \setminus \{x, y\}$ such that $y \notin C_{\succ}(S \setminus \{z'\})$.*

[9] Note that since we are using only ordinal information, any element z such that $x \succ_i z \succ_i y$ and $y \succ_j z \succ_j x$ is interpreted a "compromise", regardless of how it is ranked relative to other elements that are ranked in between x and y. One may question this interpretation if, for example, $x \succ_i z \succ_i w \succ_i v \succ_i y$ and $y \succ_i w \succ_i v \succ_i z \succ_j x$. In this case it may seem less reasonable to consider z a compromise between x and y, since in some sense, it is "closer" to x than to y. We return to this point in the concluding section, where we discuss possible extensions.

SYM formalizes a sense in which all the options selected by the solution are of equal "status". Suppose x and y are both in the solution. Imagine that one of the bargainers makes the following argument against the inclusion of x: "x is not selected when the option z is removed from the table; but since we did not choose z, we may consider it off the table, hence, we should not select x". Such an argument would not be convincing if the other bargainer could counter by observing that a similar claim can be made against y : if we remove z', which was not chosen, then y will no longer be selected. Observe that SYM is vacuous if the choice method is rational, but it does place a nontrivial restriction on irrational procedures. However, this property is satisfied by some well-known social welfare functions such as the Borda rule mentioned above.

Our main result in this section relies on the following inductive characterization of the fallback bargaining solution (we relegate its proof to the Appendix).

Lemma 1 *Let $\succ \in L(X)^2$, and let S be a bargaining problem with at least four elements. Then,*

1. *$C^f_\succ(S) = \{x\}$ if and only if*
   (a) *$x \in C^f_\succ(S \setminus \{w\})$, for each $w \in S \setminus \{x\}$, and*
   (b) *for each $y \in S \setminus \{x\}$, there exists $w \in S \setminus \{y\}$ such that $y \notin C^f_\succ(S \setminus \{w\})$.*
2. *$C^f_\succ(S) = \{x, y\}$ if and only if*
   (a) *$C^f_\succ(S \setminus \{w\}) \subseteq \{x, y\}$, for each $w \in S$, and*
   (b) *there exists $w \in S \setminus \{x, y\}$ such that $C^f_\succ(S \setminus \{w\}) = \{x\}$ if and only if there exists $w' \in S \setminus \{x, y\}$ such that $C^f_\succ(S \setminus \{w'\}) = \{y\}$.*
3. *If $C^f_\succ(S) = \{x\}$, $C^f_\succ(S \setminus \{y\}) = \{x, z\}$, and $C^f_\succ(S \setminus \{z\}) = \{x, y\}$, then there exists $i \in \{1, 2\}$ such that $y \succ_i x \succ_i z$ and $z \succ_{-i} x \succ_{-i} y$.*
4. *If $C^f_\succ(S) = C^f_\succ(S \setminus \{w\}) = \{x, y\}$, for all $w \in S \setminus \{x, y\}$, then $x \succ w$ and $y \succ w$, for all $w \in S \setminus \{x, y\}$.*

Theorem 1 *$\mathcal{C}^f$ is the only regular bargaining operator that satisfies EFF, ATT, NBC, RA, and SYM.*

Proof: We first check that $\mathcal{C}^f$ satisfies the axioms. EFF and regularity follow immediately from the definition. RA and SYM follow from Lemma 1 when starting from a set that contains at leats four options. As for sets with three options, notice that $x \in C^f_{\succ}(\{x,y,z\})$ implies that $x \in C^f_{\succ}(\{x,y\}) \cap C^f_{\succ}(\{x,z\})$, for each x,y,z, by EFF, thereby showing RA. Also, $x,y \in C^f_{\succ}(\{x,y,z\})$ implies that $x \succ z$ and $y \succ z$, which guarantees SYM. As for ATT, observe that $\min_{i=1,2} s_i(y, S \cup \{x\}, \succ) = \min_{i=1,2} s_i(y,S,\succ)+1$, for each $y \in C^f_{\succ}(S)$ such that $y \succ x$, while the minimal score of any other option cannot increase by more than one point. Hence any such y must belong to $C^f_{\succ}(S \cup \{x\})$, and any option that was not selected for S does not belong to $C^f_{\succ}(S \cup \{x\})$. Now we only have to show that $z \notin C^f_{\succ}(S \cup \{x\})$ when $C^f_{\succ}(S) = \{y,z\}$, $y \succ x$ and $z \not\succ x$. To fix the notation, suppose that $\arg\min_{i=1,2} s_i(y,S,\succ) = 1$ and $\arg\min_{i=1,2} s_i(z,S,\succ) = 2$. Hence $z \succ_1 y$, and transitivity implies that $z \succ_1 x$. In turn, this implies that $x \succ_2 z$. The minimal score of z thus remains constant when adding x, and $z \notin C^f_{\succ}(S \cup \{x\})$. Finally for NBC, suppose that $C^f_{\succ}(S) = \{x,y\}$ and that there exists $z \in S$ such that $x \succ_i z \succ_i y$ and $y \succ_{-i} z \succ_{-i} x$. Hence it must be that the minimal score for y is reached along dimension i, and it is equal to the minimal score of x that is reached along dimension $-i$. On the other hand, z scores at least one additional point than y (resp. x) along dimension i (resp. $-i$). Hence a contradiction with the fact that x and y have the largest minimal score among all the elements of S.

We now move to the more difficult part of the proof, showing the necessary condition. Thus let $\mathcal{C}_{\succ}$ be a regular bargaining operator that satisfies the five axioms. We prove that $C_{\succ} = C^f_{\succ}$ in two main steps.

Step 1 *Let $\succ \in L(X)^2$, let $C_{\succ}$ be a bargaining solution that satisfies ATT, NBC, RA, EFF, and SYM. If $C_{\succ}(T) = C^f_{\succ}(T)$, for all $T \subseteq X$ with two or three elements, then $C_{\succ}(S) = C^f_{\succ}(S)$, for all $S \subseteq X$.*

We prove that $C_{\succ}(S) = C^f_{\succ}(S)$, for all $S \subseteq X$, by induction on the number of elements in S. By assumption, the result is true when $|S| = 2$ or 3. We assume now that the result holds for any subset of X with at most $s-1$ elements, and we choose a set S with exactly s elements ($s \geq 4$). We have to prove that $C_{\succ}(S) = C^f_{\succ}(S)$.

First, note that $C_{\succ}(S)$ has at most two elements. Suppose on the contrary that $x,y,z \in C_{\succ}(S)$. EFF implies that there is no Pareto comparison between any pair of elements in $\{x,y,z\}$. Hence one of these three options must fall "in between" the other two, leading to a contradiction with NBC.

Suppose now that $C^f_\succ(S) = \{x, y\}$, for some $x, y \in S$. Lemma 1 and the induction hypothesis imply that $C_\succ(S \setminus \{w\}) = C^f_\succ(S \setminus \{w\}) \subseteq \{x, y\}$, for each $w \in S$. Notice that $C_\succ(S)$ cannot include an element different from x and y. Indeed, $|C_\succ(S)| \leq 2$ would then imply that $C_\succ(S)$ is either $\{z\}$, $\{x, z\}$, $\{y, z\}$, or $\{z, z'\}$, for some $z, z' \in S \setminus \{x, y\}$, and RA would lead to a contradiction with $C_\succ(S \setminus \{w\}) \subseteq \{x, y\}$, for all $w \in S$. So we'll be done after proving that $C_\succ(S)$ is equal to neither $\{x\}$, nor $\{y\}$. Suppose on the contrary that $C_\succ(S) = \{x\}$ (a similar reasoning applies for y). RA implies that $x \in C_\succ(S \setminus \{w\})$, for all $w \in S \setminus \{x\}$. If there exists $w \in S \setminus \{y\}$ such that $y \notin C_\succ(S \setminus \{w\})$, then $C^f_\succ(S \setminus \{w\}) = C_\succ(S \setminus \{w\}) = \{x\}$. Lemma 1 and the induction hypothesis imply that there exists $w' \in S \setminus \{x, y\}$ such that $C_\succ(S \setminus \{w'\}) = C^f_\succ(S \setminus \{w'\}) = \{y\}$, a contradiction with the fact that $x \in C_\succ(S \setminus \{w'\})$. We must conclude that $C^f_\succ(S \setminus \{w\}) = C_\succ(S \setminus \{w\}) = \{x, y\}$, for all $w \in S \setminus \{x, y\}$. By part (4) of Lemma 1, $x \succ w$ and $y \succ w$, for all $w \in S \setminus \{x, y\}$. Therefore $C_\succ(\{x, w\}) = \{x\}$ and $C_\succ(\{y, w\}) = \{y\}$ since $C_\succ = C^f_\succ$ on pairs. We also have $C_\succ(\{x, y\}) = C^f_\succ(\{x, y\}) = \{x, y\}$, and applying ATT iteratively (adding elements of $S \setminus \{x, y\}$ one at a time), we conclude that $C_\succ(S) = \{x, y\}$, contradicting the original assumption that $C_\succ(S) = \{x\}$.

To conclude the proof of Step 1, suppose that $C^f_\succ(S) = \{x\}$, for some $x \in S$. If $C_\succ(S) = \{y\}$, for some $y \neq x$, then $y \in C_\succ(S \setminus \{w\})$, for all $w \in S \setminus \{y\}$, by RA. This leads to a contradiction with Lemma 1, since there must exist $w \in S \setminus \{y\}$ such that $y \notin C^f_\succ(S \setminus \{w\}) = C_\succ(S \setminus \{w\})$. It is also impossible to have $C_\succ(S) = \{y, z\}$, for some y, z different from x. Indeed, RA applied to both $C_\succ$ and $C^f_\succ$ would then imply that $C^f_\succ(S \setminus \{y\}) = \{x, z\}$, and $C^f_\succ(\{S \setminus \{z\}) = \{x, y\}$. Part (3) of Lemma 1 implies that there exists $i \in \{1, 2\}$ such that $y \succ_i x \succ_i z$ and $z \succ_{-i} x \succ_{-i} y$, a contradiction with NBC. Suppose now that $C_\succ(S) = \{x, y\}$, for some y different from x. Lemma 1 implies that there exists $w \in S \setminus \{y\}$ such that $y \notin C^f_\succ(S \setminus \{w\}) = C_\succ(S \setminus \{w\})$. SYM implies that there exists $w' \in S \setminus \{x\}$ such that $x \notin C_\succ(S \setminus \{w'\}) = C^f_\succ(S \setminus \{w'\})$, which is impossible. Hence $C_\succ(S) = \{x\}$, as desired. This concludes the proof of Step 1.

Step 2 *Let $\mathcal{C}$ be a regular bargaining operator that satisfies ATT, NBC, RA and EFF. Then $C_\succ(T) = C^f_\succ(T)$ for all $T \subseteq X$ with two or three elements, and all $\succ \in L(X)^2$.*

Let $\succ \in L(X)^2$. Suppose first $T = \{x, y\}$. If $x \succ y$, then $C^f_\succ(T) = \{x\}$. By EFF, $y \notin C_\succ(\{x, y\})$, and hence $C_\succ(T) = \{x\}$, as desired. A similar reasoning applies if $y \succ x$. If $x \succ_1 y$ and $y \succ_2 x$, then $C^f_\succ(T) = \{x, y\}$. Suppose, on the other hand, that $C_\succ(T) = \{x\}$. Let $g : X \to X$ be the isomorphism defined by $g(x) = y$, $g(y) = x$, and $g(z) = z$, for all $z \in X \setminus \{x, y\}$. The first regularity property implies that $C_{g(\succ)}(g(T)) = \{y\}$. Notice though that $g(T) = T$, and $g(\succ)$ equals $(\succ_2, \succ_1)$ when restricted to T. The second and third regularity properties then imply that $C_\succ(T) = \{y\}$, a contradiction. Similarly, $C_\succ(T) = \{y\}$ would lead to a contradiction, and we conclude that $C_\succ(T) = \{x, y\}$, as desired. A similar reasoning applies if $y \succ_1 x$ and $x \succ_2 y$.

Let now $T = \{x, y, z\}$. If one of the elements, say x, Pareto dominates the other two, then by EFF, $C^f_\succ(T) = \{x\} = C_\succ(T)$. If two elements, say x and y, are not Pareto dominated, but both Pareto dominate z, then $C^f_\succ(T) = \{x, y\}$. The previous paragraph implies that $C_\succ(\{x, y\}) = \{x, y\}$, and ATT implies that $C_\succ(T) = \{x, y\}$, as desired. If two pairs of elements are not Pareto comparable, say (x, y) and (x, z), but the third one is, say $y \succ z$, then $C^f_\succ(T) = \{y\}$. The previous paragraph implies that $C_\succ(\{x, y\}) = \{x, y\}$, $C_\succ(\{x, z\}) = \{x, z\}$, and $C_\succ(\{y, z\}) = \{y\}$. ATT implies that $C_\succ(T) = \{y\}$ as well, as desired. There remains the last case, where there is no Pareto comparison out of any pair in T, let's say $x \succ_1 y \succ_1 z$ and $z \succ_2 y \succ_2 x$. Then $C^f_\succ(T) = \{y\}$. We already proved in Step 1 that $C_\succ(T)$ contains at most two elements. It cannot be $\{x, z\}$, because of NBC. If $C_\succ(T) = \{x, y\}$, then consider the isomorphism $g : X \to X$ defined by $g(x) = z$, $g(z) = x$, and $g(\xi) = \xi$, for all $\xi \in X \setminus \{x, z\}$. The first regularity property implies that $C_{g(\succ)}(g(T)) = \{y, z\}$. Notice though that $g(T) = T$, and $g(\succ)$ equals $(\succ_2, \succ_1)$ when restricted to T. The second and third regularity properties then imply that $C_\succ(T) = \{y, z\}$, a contradiction. A similar argument shows that it is impossible to have $C_\succ(T) = \{y, z\}$, $\{x\}$, or $\{z\}$. Hence $C_\succ(T) = \{y\}$. This concludes the proof of Step 2, and hence the proof of the theorem. ■

It can be shown that the axioms appearing in Theorem 1 are independent: dropping any one of them expands the set of compatible solutions. For example, the analogue of the Borda rule in our setting:

$$C_\succ(S) = \arg\max_{x \in S}[s_1(x, S, \succ) + s_2(x, S, \succ)],$$

for each S and $\succ$, generates a regular bargaining operator that satisfies all our axioms except NBC. Moreover, the Borda rule does not exhibit the classical compromise effect over triplets because it selects all three elements whenever they are not Pareto comparable. It does, however, exhibit the attraction effect. Another interesting example is the fallback solution applied only to the set of Pareto efficient alternatives, $C_\succ(S) = C^f_\succ[EFF_\succ(S)]$, where

$$EFF_\succ(S) = \{x \in S \mid \textit{for all } y \in S,\ x \succ_i y \textit{ for some } i \in \{1, 2\}\} \tag{1}$$

Note that the fallback solution is applied here to a subset of options, whose score is unaffected by dominated elements. It therefore violates ATT, and in addition, does not exhibit the attraction effect. It does, however, exhibit the compromise effect. Our third and final example is the lexicographic refinement of the fallback solution,

$$CL^f_{\succ}(S) = \{x \in C^f_{\succ}(S) | s_i(x, S, \succ) \geq s_i(y, S, \succ), \forall i, \forall y \in C^f_{\succ}(S)\}.$$

for each S and $\succ$, generates a bargaining operator that satisfies all our axioms except RA. Further details are available from the authors upon request.

Both Sprumont (1993) and Kibris and Sertel (2007) have already provided some axiomatic characterizations of the fallback solution in an *inter*-personal bargaining context. The main axioms in these previous papers restrict the behavior of the solution across problems that differ in the bargainers' *preferences*.[10] As we show in the next section, our axioms in the case of known preferences can be extended to the case of unknown preferences by replacing statements about preferences with statements about choices from pairs and triplets (where these choices have a natural interpretation of 'revealed Pareto dominance' and 'revealed compromises'). This is possible because we follow the approach of individual choice theory and express all the axioms (apart for regularity, which we drop in the next section) in terms of how the bargaining solution - for a *fixed* profile of preferences - changes when elements are added to or removed from the bargaining problem. In contrast, Sprumont (1993) and Kibris and Sertel (2007) take a social choice approach and express their axioms in terms of how the bargaining solution - over a fixed set of options - changes when the preferences of the individual bargainers change.

We view ATT and NBC as descriptive properties of a "representative decision-maker" in the sense that they capture the observed behavior of the *majority* of subjects in the relevant experiments.[11] Most of the choice theoretic literature so far has focused instead on relaxing the rationality postulate in order to characterize choice procedures that capture a wide variety of behavior. We, on the other hand, focus on specific violations of rationality and attempt to characterize the choice procedures that generate them. This is why we impose these violations as properties.[12] By not offering a general model of

[10] The only exception is the axiom of "minimal connectedness" in Kibris and Sertel (2007), which has the same implication as NBC.

[11] As we mentioned in the Introduction, the revealed-preference exercise of the next section can also be interpreted as providing testable implications for *inter*-personal fallback bargaining. Under this interpretation, the ATT and NBC properties have a normative appeal in the sense that they suggest how the bargainers can resolve an impasse (i.e., a situation where they cannot agree on a unique solution to the bargaining problem).

[12] By analogy, consider the well-studied model of rank dependent preferences with a concave decision-weight function. Segal (1987) showed that a decision-maker with these preferences will exhibit Allais-type

choice, our model gains in predictive power, and sheds more light on the regularities that may exist in the seemingly irrational behavior of the majority of subjects. In the next section, the preference orderings will not be a primitive of the model anymore. Our new model will thus accomodate rational choice behaviors, in which case the individual's choices reveal that there is no conflict between his two revealed preference orderings.

5. REVEALED PREFERENCES

The previous two sections suggest that the fallback bargaining procedure may potentially explain systematic violations of WARP in multi-criteria decision problems. One difficulty in testing this hypothesis is that in many situations we do not directly observe the criteria used by the decision-maker, nor do we observe how the options are ranked according to each criterion. All we may hope to observe are the final choices across different decision problems. A natural question that arises is, what properties of these choices are necessary and sufficient to represent the decision-maker as if he has two criteria in his mind for ranking the options, and he resolves the conflict between these criteria by applying the fallback bargaining procedure? Suppose the observed choices do satisfy the sufficient conditions of the representation. Can we identify (at least partially, and, if so, to what extent) the two underlying linear orderings? We answer both questions in this section.

Characterization

The approach we take is to try and adapt Theorem 1 to bargaining solutions that are not preference-based. Note first that the three regularity conditions of the previous section are no longer useful as they restricted the behavior of the solution across different preference profiles. However, the main properties of Theorem 1 can be suitably adapted to the current environment.[13]

RA and SYM can be rephrased directly:

Removing an Alternative (RA) - *If* $C(S) \neq \{x\}$*, then* $C(S \setminus \{x\}) \cap C(S) \neq \emptyset$.

behavior in a wide class of choice problems (what Segal refers to as "The Generalized Allais Paradox" or GAP for short). This suggests that an alternative axiomatization of these preferences can start by imposing GAP as a property.

[13]For notational simplicity, we keep the same names for the axioms as in the previous section. Of course, though their motivation is similar, their formulation is not since the models are different. We feel this would not create any confusion since the implied meaning is clear in each section given the relevant context.

Symmetry (SYM) - *If* $x, y \in C(S)$ *and there exists* $z \in S \setminus \{x, y\}$ *such that* $x \notin C(S \setminus \{z\})$*, then there exists* $z' \in S \setminus \{x, y\}$ *such that* $y \notin C(S \setminus \{z'\})$.

To adapt ATT and EFF, we introduce a notion of revealed Pareto comparison.

Definition 1 x *is "revealed Pareto superior" to* y *if* $C(\{x, y\}) = \{x\}$

That is, whatever dimensions or criteria the decision-maker uses to evaluate the two options, x is better than y according to all of them. On the other hand, $C(\{x, y\}) = \{x, y\}$ means that there is a negative correlation when comparing x and y across dimensions: x is preferred to y along one, while y is preferred to x along the other. EFF and ATT can now be rephrased using only observed choices:

Efficiency (EFF) - *If* $x \in C(S)$*, then there does not exist* $y \in S$ *such that* y *is revealed Pareto superior to* x.

Attraction (ATT) - *Let* $x \in X \setminus S$ *be such that* y *is revealed Pareto superior to* x*, for some* $y \in C(S)$*. Then* $C(S \cup \{x\}) = \{y \in C(S) | C(\{x, y\}) = \{y\}\}$.

To redefine NBC, we introduce the notion of revealed compromise.

Definition 2 *An option* z *is "revealed to be a compromise between* x *and* y*" if it is chosen uniquely from* $\{x, y, z\}$*, but no element in this triplet is revealed to be Pareto superior to another.*

No Better Compromise (NBC) - *If both* x *and* y *belong to* $C(S)$*, then there does not exist* $z \in S$ *such that* z *is revealed to be a compromise between* x *and* y.

The above properties, however, do not guarantee the existence of two linear orderings such that the decision maker's choices can be explained by applying the fallback solution. First, these properties (and in particular, RA, which weakens WARP) do not imply that the revealed Pareto relation is transitive. Thus, to have any hope of recovering a pair of preferences, the following condition must be met:

Pairwise Consistency (PC) - *If* x *is revealed Pareto superior to* y*, and* y *is revealed Pareto superior to* z*, then* x *is revealed Pareto superior to* z.

Second, none of the above properties imply the compromise effect. To see this, suppose the decision-maker has a pair of orderings in his mind (which are not observeable to us) such that $x \succ_1 z \succ_1 y$, while $y \succ_2 z \succ_2 x$. Then a choice rule that picks $\{x, y\}$ would satisfy NBC without exhibiting the compromise effect. We must, therefore, take into account a new testable implication: if there is no revealed Pareto comparison between any two elements of $\{x, y, z\}$, then there must be a revealed compromise.

Existence of a Compromise (EC) - *If the choice out of any pair in* $\{x, y, z\}$ *is the pair itself, then* $C(\{x, y, z\})$ *is a singleton.*

The next two examples motivate our final axiom. They demonstrate that none of our axioms thus far guarantee that revealed compromises, and their interaction with revealed Pareto dominance, are consistent with an underlying pair of preferences.

Example 1 *Let* $X = \{a, b, c, d\}$ *and let* C *be the bargaining solution that selects both elements out of any pair, and such that* $C(\{a, b, c\}) = \{b\}$, $C(\{a, b, d\}) = \{d\}$, $C(\{a, c, d\}) = \{d\}$, $C(\{b, c, d\}) = \{d\}$, *and* $C(\{a, b, c, d\}) = \{d\}$. *It is not difficult to check that* C *satisfies the seven axioms listed so far, but there is no pair* $(\succ_1, \succ_2)$ *of linear orderings such* $C = C^f_\succ$. *The inconsistency leading to this impossibility is easy to understand:* $C(\{a, b, c\}) = \{b\}$ *reveals that* b *is "in between"* a *and* c, *while* $C(\{a, b, d\}) = \{d\}$ *and* $C(\{b, c, d\}) = \{d\}$ *reveals that* d *is "in between" both* a *and* b, *and* b *and* c.

Example 2 *Let* $X = \{a, b, c, d\}$ *and let* $(\succ_1^*, \succ_2^*)$ *be the two linear orderings defined as follows:* $d \succ_1^* a \succ_1^* b \succ_1^* c$ *and* $d \succ_2^* c \succ_2^* b \succ_2^* a$. *Let* C *be the bargaining solution such that* $C(\{b, d\}) = \{b, d\}$ *and* $C(S) = C^f_{\succ^*}(S)$, *for all* $S \subseteq X$ *different from* $\{b, d\}$. *It is not difficult to check that* C *satisfies the seven axioms listed so far, but there is no pair* $(\succ_1, \succ_2)$ *of linear orderings such* $C = C^f_\succ$. *The inconsistency here is rooted in the way revealed Pareto comparisons should combine with revealed compromises:* b *is revealed to be "in between"* a *and* c, d *is revealed to be Pareto superior to both* a *and* c, *yet* b *is revealed non comparable to* d.

To rule out the inconsistencies illustrated in these examples, we introduce a property that captures another sense in which compromises have a special status. Suppose y is revealed to be a compromise between x and z. One way to interpret this is that after a long process of deliberation - where one party argued in favor of x, while the other argued in favor of z - the two parties agreed to settle on y. Thus, the choice of y may be viewed as internalizing all the considerations in favor of each of the alternatives. This suggests that if a new option, w, becomes available, the parties would compare w only with y, and would not ignore the previous arguments that led to the agreement on y by opening up the discussion on all available options. Furthermore, if reaching a compromise has special status to the bargainers, then they would require a good enough reason to abandon it completely in favor of a new option. In particular, the parties may replace a compromise with a new option only when the latter Pareto dominates the former.

Overcoming a Compromise (OC) - *Suppose that y is revealed to be a compromise between x and z. If $C(\{w, x, y, z\}) = \{w\}$, then $C(\{y, w\}) = \{w\}$.*

The fallback bargaining solution satisfies an axiom of this type for all bargaining problems, but we phrased it for bargaining problems with only four elements because this is all what is needed to establish our result, as hinted by the two previous examples.

Our second main result establishes that the testable implications we have identified are also sufficient to guarantee the existence of two linear orderings such that the decision maker's choices may be explained by the fallback solution.

Theorem 2 *A bargaining solution C satisfies EFF, ATT, NBC, RA, SYM, PC, EC and OC if and only if there exists $\succ \in L(X)^2$ such that $C = C^f_\succ$.*

We start by providing a sketch of the proof to show that any bargaining solution satisfying the axioms must be a fallback solution for some pair of strict preferences. The formal proof follows. The argument unfolds in two main steps. First, we show that a choice correspondence C satisfying EFF, ATT, NBC, RA and SYM exhibits the following property: if there exists a preference profile $\succ$ such that C coincides with $C^f_\succ$ on all pairs and triplets, then this remains true on all subsets of X. To prove this, we adapt the arguments from the first step of the proof of Theorem 1, which established a similar claim for preference-based bargaining solutions.

In the second step, the more challenging part of the proof, we construct a preference profile $\succ$ such that C coincides with $C^f_\succ$ on all pairs and triplets. The difficulty here lies in the requirement that two preference relations defined on one pair or triplet must be consistent with relations defined on different pairs and triplets. For example, when we are given $C(\{x, y\}) = \{x, y\}$, we conclude that one bargainer prefers x to y, while the other bargainer has the opposite ranking. Suppose we are also given that $C(\{y, z\}) = \{y, z\}$. Then, again, we conclude that the two bargainers have opposite rankings of y and z. The question is, how do we determine whether or not the bargainer who ranks x over y also ranks y over z?

To answer this question, we use the choice data from triplets, and construct the two linear orderings inductively. We begin with one pair of elements and construct two preference relations over them. We then add a third element and extend the previous pair of preferences to cover all three elements. We then continue adding one element at a time and extending the relations from the previous step to cover the newly added element until we have covered all of X.

However, for this construction to succeed, the elements must be added in a particular order. First, we partition the set of elements into "revealed Pareto layers". The highest Pareto layer, denoted EFF^1, consists of all the elements in X that are not revealed to be Pareto inferior to any other element. Similarly, the second-highest Pareto layer, EFF^2, is defined as the set of elements in $X \backslash EFF^1$ that are not revealed to be Pareto inferior to any element not in EFF^1. The next revealed Pareto layers are defined in a similar manner. Each Pareto layer EFF^k is further partitioned into "inner" layers defined as follows. The most extreme layer, denoted $\mathcal{E}^{k,1}$, contains the set of elements (at most two) that are never revealed to be compromises within the Pareto layer EFF^k. The next inner layer contains those elements that are never revealed to be compromises within $EFF^k \backslash \mathcal{E}^{k,1}$. Continuing this way we end with the most interior layer. Given these partitions, the construction of the two preference relations proceeds as follows: we begin with the highest Pareto layer from which we choose the most extreme points and move inward. Once we cover the entire Pareto layer, we move to the next Pareto layer and again, begin with the extreme points and move inwards. A series of lemmas in the proof of Theorem 2 establish that the above method leads to two preference relations that are well-defined and transitive.

We are now ready to present the proof of Theorem 2.

Proof: We have already proved in the previous section that $C^f_\succ$ satisfies RA, SYM, EFF, ATT, and NBC, for each $\succ \in L(X)^2$. PC and EC are straightforward to check, and hence only OC remains. The fallback solution generates the choice data on $\{x, y, z\}$ as in OC only if $x \succ_i y \succ_i z$ and $z \succ_{-i} y \succ_{-i} x$, for some $i \in \{1, 2\}$. Hence the minimal score of y in the quadruplet is at least 1. For w to be chosen alone, it must be better than at least two alternatives for each ordering, and hence $w \succ y$, or $C^f_\succ(\{w, y\}) = \{w\}$, as desired.

Let now C be a bargaining solution that satisfies SYM, RA, PC, EFF, ATT, NBC, EC and OC. It is not difficult to adapt the argument from the first step in the proof of Theorem 1 to show that $C = C^f_\succ$ if $\succ \in L(X)^2$ is such that $C(T) = C^f_\succ(T)$ for all $T \subseteq X$ with two or three elements. The difficult part is to show that there indeed exists a pair $(\succ_1, \succ_2)$ of linear orderings such that $C(T) = C^f_\succ(T)$ for all $T \subseteq X$ with two or three elements. We will proceed via an inductive argument. For each strictly positive integer k, let EFF^k be the following subset of X:

$$EFF^k = \{x \in X \setminus [\cup_{j=0}^{k-1} EFF^j] | \nexists y \in X \setminus [\cup_{j=0}^{k-1} EFF^j] : C(\{x, y\}) = \{y\}\}$$

(with the convention $EFF^0 = \emptyset$). EFF^1 is the set of elements that are C-Pareto efficient in X. EFF^2 is the set of alternatives that are C-Pareto efficient in $X \setminus EFF^1$. These are "second-best" options in X. Notice that EFF^k is nonempty for each k such that $X \setminus [\cup_{j=1}^{k-1} EFF^j]$ is nonempty, since X is finite and C satisfies PC. Let K be the smallest positive integer such that $EFF^{K+1} = \emptyset$. X is thus partitioned into a collection $(EFF^k)_{k=1}^K$ of layers of options that are constrained efficient at different levels k.

Each such Pareto layer needs itself to be partitioned into subsets of one or two elements, as follows:

$$\mathcal{E}^{k,l} = \{x \in EFF^k \setminus [\cup_{j=0}^{l-1} \mathcal{E}^{k,j}] |\ \nexists y, z \in EFF^k \setminus [\cup_{j=0}^{l-1} \mathcal{E}^{k,j}] : C(\{x, y, z\}) = \{x\}\},$$

for each $k \in \{1, \ldots, K\}$, and each strictly positive integer l (with the convention $\mathcal{E}^{k,0} = \emptyset$, for each k). EC implies that a single element must be chosen out of any triplet in EFF^k. $\mathcal{E}^{k,1}$ is the set of elements that are never chosen out of any such triplets. These can be interpreted as extreme elements of the layer EFF^k. $\mathcal{E}^{k,2}$ is the set of elements that are extreme in the sub-layer $EFF^k \setminus \mathcal{E}^{k,1}$, and so on so forth. The next lemma, whose proof is available in the Appendix, highlights the structure of these sets.

Lemma 2 *Let $k \in \{1, \ldots K\}$, and let l be a strictly positive integer. If $EFF^k \setminus [\cup_{j=0}^{l-1} \mathcal{E}^{k,j}]$ has at least two elements, then $\mathcal{E}^{k,l}$ is nonempty and contains exactly two elements.*

Let L_k be the smallest positive integer such that $\mathcal{E}^{k,L_k+1} = \emptyset$. EFF^k is thus partitioned into a collection $(\mathcal{E}^{k,l})_{l=1}^{L_k}$ of pairs of alternatives (and perhaps one singleton if $\mathcal{E}^{k,L_k}$ contains only one element). An element that belongs to a layer $\mathcal{E}^{k,l}$ for some large l can be interpreted as not too extreme, in that it is chosen as a compromise out of more triplets in EFF^k.

We are now ready to define $\succ$, and prove that $C(T) = C^f_\succ(T)$ for every $T \subseteq X$ with two or three elements, by induction. We start with a pair of elements in X, then add a third element, and so on up to the point all the elements of X have been considered. We have to be careful, though, to follow some special order for the argument to work. It follows from our previous definition that each element of X belongs to a unique atom $\mathcal{E}^{k,l}$, for some $l \in \{1, \ldots, L_k\}$ and some $k \in \{1, \ldots K\}$. This fact will help us determine the right order in which elements must be added. Indeed, let $(k(x), l(x))$ be these two positive integers associated to x. We will follow the convention that x is added before x' if $(k(x), l(x))$ is lexicographically inferior to $(k(x'), l(x'))$. As we know from Lemma 2, this rule does not uniquely specify the ordering, as an atom $\mathcal{E}^{k,l}$ usually contains two elements. We do not further specify how elements are added in the inductive argument, as this is inconsequential for the construction of $\succ$, and the proof that $C = C^f_{\succ}$ on pairs and triplets.[14]

Let x and y be the first two elements of X for which $\succ$ must be defined. If $C(\{x, y\}) = \{x\}$, then we impose that $x \succ_1 y$ and $x \succ_2 y$. Similarly, if $C(\{x, y\}) = \{y\}$, then we impose that $y \succ_1 x$ and $y \succ_2 x$. Finally, if $C(\{x, y\}) = \{x, y\}$, then we impose that $x \succ_1 y$ and $y \succ_2 x$, or $y \succ_1 x$ and $x \succ_2 y$. Either way works, and one may choose one of the two options arbitrarily. Of course, $C(\{x, y\}) = C^f_{\succ}(\{x, y\})$, by construction.

Suppose now that $\succ$ has been defined on a subset S of X, and that $C(T) = C^f_{\succ}(T)$ for each $T \subseteq S$ with two or three elements, while the next element to be added is $w \in X \setminus S$. We now define the extension $\succ^*$ over $S \cup \{w\}$. Of course, $\succ^*$ is defined so as to coincide with $\succ$ on S, i.e. $x \succ^*_i y$ if and only if $x \succ_i y$, for each $x, y \in S$ and each $i = 1, 2$. The important question to answer is how elements of S compare with w under $\succ^*$. For this, we partition S into two subsets:

$$A_w = \{x \in S | C(\{w, x\}) = \{x\}\}$$

$$B_w = \{x \in S | C(\{w, x\}) = \{w, x\}\}.$$

Notice that $A_w \cap B_w = \emptyset$, and $S = A_w \cup B_w$, because there is no $x \in S$ such that $C(\{w, x\}) = \{w\}$ (given the way we add elements in our inductive argument). For each $x \in A_w$, we impose $x \succ^*_1 w$ and $x \succ^*_2 w$. As for an element $x \in B_w$, we must distinguish two cases. In the first case, we assume that there exists $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$. Then we impose $x \succ^*_1 w$ and $w \succ^*_2 x$ when there exists $y \in B_w$ such that $x \succ_1 y$ and $C(\{x, w, y\}) = \{w\}$, and $w \succ^*_1 x$ and $x \succ^*_2 w$ when there exists $y \in B_w$ such that $y \succ_1 x$ and $C(\{x, w, y\}) = \{w\}$. We need to check that this is well-defined. This follows from the next lemma, whose proof is available in the appendix.

Lemma 3 *If there exists $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, then, for each $x \in B_w$, there exists $y \in B_w$ such that $C(\{x, w, y\}) = \{w\}$. In addition, if $y, y' \in B_w$ are such that $C(\{x, w, y\}) = C(\{x, w, y'\}) = \{w\}$, then $x \succ_i y$ if and only if $x \succ_i y'$, for both $i = 1, 2$.*

In the second case, namely when there does not exist $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, we impose $x \succ^*_1 w$ and $w \succ^*_2 x$ if there exists $\xi \in A_w$ and $y \in B_w$ such that $y \succ_1 \xi$, and $w \succ^*_1 x$ and $x \succ^*_2 w$ if there exists $\xi \in A_w$ and $y \in B_w$ such that $y \succ_2 \xi$. If there is

[14] Identifiability, i.e. the possibility of finding multiple pairs of ordering $\succ$ such that $C = C^f_{\succ}$, is the subject of the next theorem.

no $\xi \in A_w$ and no $y \in B_w$ such that either $y \succ_1^* \xi$ or $y \succ_2^* \xi$, then one is free to choose either definition, i.e. $x \succ_1^* w$ and $w \succ_2^* x$, for all $x \in B_w$, or $w \succ_1^* x$ and $x \succ_2^* w$, for all $x \in B_w$.[15] Here, too, we need to check that this is well defined. This follows from the next lemma, whose proof is available in the appendix.

Lemma 4 *If there do not exist $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, then there do not exist $\xi, \xi' \in A_w$ and $y, y' \in B_w$ such that $y \succ_2 \xi$ and $y' \succ_1 \xi'$.*

Now that the pair $\succ^*$ of linear orderings has been defined on $S \cup \{w\}$, we should check that they are transitive, i.e. for $i = 1, 2$, $x \succ_i^* w$ if $x \succ_i y$ and $y \succ_i^* w$, $x \succ_i y$ if $x \succ_i^* w$ and $w \succ_i^* y$, and the reverse rankings of both of these cases. We postpone the argument to the appendix.

We will be done with our inductive argument and the proof of Step 2 after proving that $C(T) = C^f_{\succ^*}(T)$, for all $T \subseteq S \cup \{w\}$ with two or three elements. When $w \notin T$ this follows directly from the inductive step. Consider some pair $\{x, w\}$, where $x \in S$. If $x \in A_w$, then $C(\{x, w\}) = \{x\}$ and $\succ^*$ satisfies: $x \succ_1^* w$ and $x \succ_2^* w$. Hence, $C^f_{\succ^*}(\{x, w\}) = \{x\}$ as well, as desired. If $x \in B_w$, then $C(\{x, w\}) = \{x, w\}$ and $\succ^*$ satisfies: $x \succ_i^* w$ and $w \succ_{-i}^* x$ for some $i \in \{1, 2\}$. Hence, $C^f_{\succ^*}(\{x, w\}) = \{x, w\}$ as well, as desired.

Consider next a triplet $\{x, y, w\}$. If $\{x, y\} \subseteq A_w$, then $x \succ^* w$ and $y \succ^* w$. The inductive step and ATT imply: $C(\{x, y, w\}) = C(\{x, y\}) = C^f_{\succ^*}(\{x, y\}) = C^f_{\succ^*}(\{x, y, w\})$, as desired.

Suppose next that only one of the alternatives in $\{x, y\}$, say x, belongs to A_w, in which case $y \in B_w$. PC implies that $C(\{x, y\}) = \{x\}$ or $\{x, y\}$. In the former case, x is the only C-efficient (resp. $\succ^*$-efficient) option in $\{x, y, w\}$, and hence $C(\{x, y, w\}) = \{x\} = C^f_{\succ^*}(\{x, y, w\})$, by EFF, as desired. If $C(\{x, y\}) = \{x, y\}$, then $C(\{x, y, w\}) = \{x\}$ by ATT. The constructed preference profile $\succ^*$ satisfies $x \succ_i^* w \succ_i^* y$ and $y \succ_{-i}^* x \succ_{-i}^* w$ (here we use the fact that $\succ_i^*$ is transitive, which is proven in the appendix), for some $i \in \{1, 2\}$. Hence $C^f_{\succ^*}(\{x, y, w\}) = \{x\}$ as well, as desired.

Finally, we consider the case in which neither x nor y belong to A_w. This means that $x, y \in B_w$. Suppose that $C(\{x, y\})$ is a singleton, say $\{x\}$. Then, $C(\{x, y, w\}) = \{x\}$, by ATT. The constructed preference profile $\succ^*$ satisfies $x \succ_i^* y \succ_i^* w$ and $w \succ_{-i}^* x \succ_{-i}^* y$ (again, remember that $\succ_i^*$ and $\succ_{-i}^*$ are transitive), for some $i \in \{1, 2\}$. Hence $C^f_{\succ^*}(\{x, y, w\}) = \{x\}$ as well, as desired.

[15] Note that in this case, every element in A_w is revealed to be Pareto superior to any element outside this set. As we show in the next subsection, this is the only case where we cannot uiquely identify the two preference relations that are consistent with the choice data.

Now comes the last, and most difficult, case where $C(\{x, y\}) = \{x, y\}$ and $x, y \in B_w$. By construction, $x \succ_i y$ and $y \succ_{-i} x$, for some $i \in \{1, 2\}$. Since the choice out of any pair in $\{x, y, w\}$ is the pair itself, EC implies that $C(\{x, y, w\})$ is a singleton. Assume w.l.o.g. that x has been added before y in the induction.

If $C(\{x, y, w\}) = \{w\}$, then by construction, $x \succ_i^* w \succ_i^* y$ and $y \succ_i^* w \succ_i^* x$. Therefore, $C^f_{\succ^*}(\{x, y, w\}) = \{w\}$ as well, as desired.

Assume $C(\{x, y, w\}) = \{x\}$. Observe that $k(x) \leq k(y) \leq k(w)$, since y is added after x, and w after y. In addition, x, y and w cannot all lie in the same C-Pareto layer, i.e. $k(x) < k(w)$. To see why, suppose on the contrary that $\{x, y, w\} \subseteq EFF^{k(x)}$. Then $l(x) \leq l(y) \leq l(w)$, since y is added after x and w is added after y. Hence, by the definition of $\mathcal{E}^{k(x),l(x)}$, $C(\{x, y, w\}) \neq \{x\}$, a contradiction. Since $k(w) > k(x)$, there must exist $w' \in S$ such that $k(w') = k(x)$ and $C(\{w, w'\}) = \{w'\}$. Lemma 9 from the Appendix implies that $C(\{x, y, w'\}) = \{x\}$. Hence $C^f_{\succ}(\{x, y, w'\}) = \{x\}$, by the induction hypothesis, and we must have: $w' \succ_i x \succ_i y$ and $y \succ_{-i} x \succ_{-i} w'$. Since $C(\{w, w'\}) = \{w'\}$, we know that $w' \succ^* w$. By transitivity, we get $x \succ_{-i}^* w$. Since $C(\{x, w\}) = \{x, w\}$, we have $w \succ_i^* x$. Hence $C^f_{\succ^*}(\{x, y, w\}) = \{x\}$, as desired.

Assume finally that $C(\{x, y, w\}) = \{y\}$. If $k(x) = k(y) = k(w)$, then $l(x) \leq l(y) \leq l(w)$, since y is added after x, and w after y. In order to have $C(\{x, y, w\}) = \{y\}$, it must be that $l(y) > l(x)$, by definition of $\mathcal{E}^{k(x),l(x)}$. Lemma 2 implies that there exists another element x' in $\mathcal{E}^{k(x),l(x)}$. Since $l(y) > l(x)$, it must be that $C(\{x, y, x'\}) = \{y\}$. In order to satisfy the induction hypothesis and the convention $x \succ_i y$, we must have $y \succ_i x'$. Since $l(w) > l(x)$, it must be that $C(\{x, w, x'\}) = \{w\}$. The second statement from Lemma 7 in the Appendix implies that $C(\{w, y, x'\}) \neq \{y\}$, since $C(\{x, y, w\}) = \{y\}$. On the other hand, $C(\{w, x', y\})$ must be a singleton by EC, and cannot be $\{x'\}$ either, since $l(x') < l(y) \leq l(w)$. Hence $C(\{w, x', y\}) = \{w\}$, and $y \succ_i^* w$, by definition. We conclude that $x \succ_i y \succ_i^* w$ and $w \succ_{-i}^* y \succ_{-i} x$, which implies $C^f_{\succ^*}(\{w, x, y\}) = \{y\}$, as desired.

To conclude, suppose that $k(x) < k(w)$. Since $C(\{x, w\}) = \{x, w\}$ and $C(\{y, w\}) = \{y, w\}$, we have three cases to consider:

Case 1) $x \succ_i y \succ_i^* w$ and $w \succ_{-i}^* y \succ_{-i} x$,

Case 2) $x \succ_i^* w \succ_i^* y$ and $y \succ_{-i}^* w \succ_{-i}^* x$, and

Case 3) $w \succ_i^* x \succ_i y$ and $y \succ_{-i} x \succ_{-i}^* w$.

If Case 1 prevails, then $C^f_{\succ^*}(\{x, y, w\}) = \{y\}$. So we will be done after proving that Cases 2 and 3 are impossible.

In Case 2 there are elements on both sides of w according to $\succ^*$, hence, we may apply Lemma 3. Thus, there exists $x' \in B_w$ such that $C(\{x, w, x'\}) = \{w\}$. It must be that

$C(\{x', y\}) = \{x', y\}$, as otherwise we get a contradiction with $C(\{w, x, y\}) = \{y\}$ via Lemma 9. Since $x \succ_i^* w$, it must be that $x \succ_i x'$. Transitivity of $\succ^*$ also implies that $x \succ_i y$. So we have two subcases to consider:

Case 2a: $x \succ_i y \succ_i x'$ and vice versa for $-i$ (because the choice out of both $\{x, y\}$ and $\{x', y\}$ is the pair itself), and

Case 2b: $x \succ_i x' \succ_i y$ and vice versa for $-i$.

Knowing that $C(\{x, w, x'\}) = \{w\}$ and $C(\{w, x, y\}) = \{y\}$, subcase 2b (leading to $C(\{x, x', y\}) = \{x'\}$, by the induction hypothesis) is incompatible with RA, given that $C(\{w, x, x', y\})$ contains at most two elements (see Lemma 5 in the Appendix). RA can be satisfied in case 2a only if $C(\{w, x, x', y\}) = \{y\}$ or $\{w, y\}$. The former leads to a contradiction with OC. In the second case, notice that a single option must be selected out of $\{w, x', y\}$ by EC, and it must be w by RA and SYM. Recall that $y \succ_i x'$ in case 2a, and hence, $y \succ_i^* w$ by definition of $\succ^*$, in contradiction to Case 2.

As for Case 3, let $w' \in EFF^{k(x)}$ be such that $C(\{w, w'\}) = \{w'\}$. Hence $w' \succ_i^* w$, by definition, and transitivity implies that $w' \succ_i x \succ_i y$. $C(\{w', x\}) = \{w', x\}$ then implies $y \succ_{-i} x \succ_{-i} w'$. On the one hand, we could conclude that $C(\{w', y\}) = \{w', y\}$, and hence $C(\{x, y, w'\}) = \{y\}$ by Lemma 9, or $C_{\succ}^f(\{x, y, w'\}) = \{y\}$, by the induction hypothesis. On the other hand, if one can compute $C_{\succ}^f(\{x, y, w'\})$ directly from $\succ$, in which case one gets $\{x\}$, hence the contradiction. ■

The axioms appearing in Theorem 2 are independent. Details are available from the authors upon request.

Identifiability

There is no hope to identify uniquely the underlying preference relations on both dimensions. Indeed, there is no way to tell which ordering should be associated to a specific self or dimension of choice: if $C = C_{\succ}^f$, for some pair $(\succ_1, \succ_2)$ of linear orderings on X, then we also have $C = C_{(\succ_2, \succ_1)}^f$ (cf. second regularity condition in the previous section). One may wonder whether this is the only source of multiplicity. The answer is not quite, but almost, as the following example and theorem illustrate.

Example 3 *Consider* $X = \{a, b, c, d\}$ *and* $C = C_{\succ}^f$, *where* $a \succ_1 b \succ_1 c \succ_1 d$ *and* $b \succ_2 a \succ_2 d \succ_2 c$. *It is not difficult to check that* C *is also equal to* $C_{\succ'}^f$, *where* $b \succ_1' a \succ_1' c \succ_1' d$ *and* $a \succ_2' b \succ_2' d \succ_2' c$. *The careful reader will notice that* $\succ'$ *is obtained from* $\succ$ *by exchanging the preferences of the two selves only as far as* a *and* b *are concerned. This* *change is irrelevant as far as the fallback bargaining solution is concerned, because both a and b Pareto dominate both c and d according to $\succ$, implying that c and d are irrelevant when it comes to determine the solution of any subset S of X that include either a, b, or both.*

A subset S of X is *C-dominant* if it is non-empty and $C(\{x,y\}) = \{x\}$, for all $x \in S$ and all $y \in X \setminus S$.[16] Observe that if S and S' are both C-dominant, then $S \subseteq S'$ or $S' \subseteq S$. Also X is trivially C-dominant. So there exists a unique minimal C-dominant set S_1^* in X. Similarly, a subset S of $X \setminus S_1^*$ is C-dominant in $X \setminus S_1^*$ if it is non-empty and $C(\{x,y\}) = \{x\}$, for all $x \in S$ and all $y \in X \setminus (S \cup S_1^*)$. Let S_2^* be the minimal C-dominant set in $X \setminus S_1^*$. Iterating the procedure, one obtains a partition of X into a finite sequence $\Pi = (S_1^*, \ldots, S_K^*)$ of sets with the property that S_k^* is the minimal C-dominant set in $X \setminus \cup_{j=1}^{k-1} S_j^*$.

Theorem 3 *Let $\succ, \succ'$ be two pairs of strict linear orderings. Then $C_{\succ}^f = C_{\succ'}^f$ if and only if $\succ'$ can be obtained from $\succ$ by permuting the two orderings over atoms of* Π *that contains at least two elements.*

Proof: The sufficient condition is easy to check, and we focus attention only on the necessary condition. Let C be the common bargaining solution. Since it coincides with the fallback bargaining solution for some pair of orderings, it satisfies the axioms listed in the previous section, and the induction we followed in the proof of Theorem 2 can be reproduced here as well.

Let S^* be an atom of the partition Π. We will prove that $\succ$ and $\succ'$, restricted to S^*, must coincide or be a permutation of each others. The result will follow, since there is only one way of patching together the orderings obtained on the different atoms of Π, so as to be consistent with C: $x \succ y$ if and only if x belongs to an atom that comes before the atom to which y belongs. For the sake of notational simplicity, we will assume that S^* is the first atom of Π with at least two elements, but the reasoning can easily be extended by induction to any subsequent atom (the argument is trivial if S^* is the first atom and it has only one element). Let x, y be the first two elements to be considered in the induction of Theorem 2. Notice that $C(\{x,y\}) = \{x,y\}$, as otherwise either $\{x\}$ or $\{y\}$ would be C-dominant, a contradiction with the fact that S^* is minimal. Let i, j be such that $x \succ_i y$, $y \succ_{-i} x$, $x \succ'_j y$, $y \succ'_{-j} x$. Let us now think about how $\succ$ and $\succ'$ would extend to larger sets by adding elements in an order that follows the induction

[16] If C satisfies EFF, then S is C-dominant if and only if $C(T) \subseteq S$, for each $T \subseteq X$ such that $S \cap T \neq \emptyset$.

from the proof of Theorem 2. Let w be the third element in the induction. Looking back at the proof, notice that there is only one possible such extension so that the associated fallback solution coincides with C. For instance, if $x \in A_w$, then imposing anything else than $x \succ w$ and $x \succ' w$ would lead to a contradiction with $C(\{w, x\}) = \{x\}$. If $x \in B_w$ and $C(\{x, w, y\}) = \{w\}$, then imposing anything else than $x \succ_i w \succ_i y$, $y \succ_{-i} w \succ_{-i} x$, $x \succ'_j w \succ'_j y$, and $y \succ'_{-j} w \succ'_{-j} x$ would contradicts C. Finally, suppose $C(\{w, x, y\}) \neq \{w\}$ and $x \in B_w$. If $y \in B_w$, then since w is the third element to be added, $l(y) < l(w)$ (recall the notation from our induction argument in Theorem 2), which means that $C(\{w, x, y\}) \neq \{w\}$, a contradiction. Therefore, $y \in A_w$, and one must choose $y \succ w$ and $y \succ' w$ which, together with transitivity, forces us to choose $x \succ_i w$, $w \succ_{-i} x$, $x \succ'_j w$, and $w \succ'_{-j} x$. More generally, it is easy to check that the inductive argument from Theorem 2 implies that the definition of $\succ$ on $\{x, y\}$ uniquely determines its definition on larger sets obtained by adding elements w, except if B_w is non-empty, there are no $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, and no $(\xi, y) \in A_w \times B_w$ such that either $y \succ_1 \xi$ or $y \succ_2 \xi$. The same is true for $\succ'$. The rest of the proof amounts to show that these conditions occur only if A_w is C-dominant. Since this would contradict the minimality of S^*, it will imply that $\succ = \succ'$ on S^* (if $i = j$) or are a permutation of each others (if $i = -j$), as desired.

Suppose thus that B_w is non-empty, there are no $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, and no $(\xi, y) \in A_w \times B_w$ such that either $y \succ_1 \xi$ or $y \succ_2 \xi$. The same is true for $\succ'$. Let $a \in A_w$ and $b \in X \setminus A_w$. We have to prove that $C(\{a, b\}) = \{a\}$. If b is added before w in the induction, then $b \in B_w$, and the result follows trivially from the conclusion that no element in B_w is ever chosen in a pair containing an element in A_w. Suppose now that b is added after w in the induction, i.e. $(k(b), l(b))$ lexicographically dominates $(k(w), l(w))$. Suppose first that $k(b) = k(w)$. Since there is no $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{z\}$, it must be that $l(w) = 1$. Since B_w is non-empty, it must be that there exists another element w' such that $k(w') = k(w)$ that has been added before w - this must be the other element of the atom $\mathcal{E}^{(k(w),1)}$ (remember that those atoms contain at most two elements, see Lemma 2). Hence $C(\{w, b, w'\}) = \{b\}$. Since there is no element in A_w and no element in B_w from which C picks both elements, $C(\{a, w'\}) = \{a\}$. Since $C = C_\succ$, we must have $a \succ w$, $a \succ w'$, and there exists $i \in \{1, 2\}$ such that $w \succ_i b \succ_i w'$ and $w' \succ_i b \succ_i w$. Hence $C(\{a, b\}) = \{a\}$, as desired. Finally, if $k(b) > k(w)$, then there exists x'' such that $k(x'') = k(w)$ and $C(\{x'', b\}) = \{x''\}$ (x'' could be w itself). By essentially the same argument as above, we may conclude that $C(\{a, x''\}) = \{a\}$, and hence, $C(\{a, b\}) = \{a\}$, by PC, as desired. ■

6. CONCLUDING REMARKS

This paper proposes to formalize the notion of "reason-based choice" as a cooperative solution to bargaining between the conflicting inner selves of a decision-maker. The specific bargaining solution we characterize, the fallback bargaining solution, has several noteworthy features. First, it can be implemented with a straightforward algorithm, which is reminiscent of usual methods weighing pros and cons. Second, it admits a simple non-cooperative foundation, and hence may be interpreted as the outcome of interpersonal as well as intra-personal bargaining. Third, it has been studied before in the social choice literature. Finally, it generates two well known violations of rationality - the attraction and the compromise effects - while retaining substantial testable implications. Moreover, when a choice problem involves both attraction and compromises, fallback bargaining makes a prediction as to what outcome would be chosen. In particular, when there is a single option x, which is ranked in between all others, and only one pair of elements (y, z) (such that $y, z \neq x$) that are Pareto comparable, then fallback bargaining would select x alone. In this sense, fallback bargaining may be viewed as "favoring" the compromise effect over the attraction effect.

In what follows, we discuss two interesting extensions of the fallback bargaining solution.

More than two bargainers

It is straightforward to adapt the definition of the fallback solution to any number of selves. Given n strict preference orderings on X, $\succ = (\succ_1, \ldots, \succ_n)$, define the fallback bargaining solution associated with $\succ$ as follows:

$$C^f_{\succ}(S) = \arg\max_{x \in S} \min_{i=1,\ldots,n} s_i(x, \succ, S),$$

for each $S \subseteq X$, where

$$s_i(x, S, \succ) = |\{y \in S | x \succ_i y\}|.$$

Moving from one to two selves allows to explain irrational choice patterns, accomodating both the attraction and the compromise effects, while retaining some significant predictive power. Restricting attention to the dual-self case is the natural place to start (see e.g. Gul and Pesendorfer (2001), Fudenberg and Levine (2006), and Manzini and Mariotti (2007)), but it is also important to get a sense of how much more permissive the model becomes, should one leave the number of selves unrestricted. Contrary to numerous other models (see e.g. Kalai et al. (2002) and especially Ambrus and Rozen (2010)), fallback bargaining does retain some predictive power independently of the number of selves involved.

Proposition 1 *Let C be a bargaining solution for which there exist n strict preference orderings defined on X, $\succ = (\succ_1, \ldots, \succ_n)$, such that $C = C^f_\succ$. Then C satisfies RA, EFF, NBC, and PC. It also satisfies the following three properties:*

1. *(one inclusion from ATT) If $x \notin S$, then $\{y \in C(S) | C(\{x, y\}) = \{y\}\} \subseteq C(S \cup \{x\})$.*

2. *(strengthening of RA) If $x \in C(S)$ and $C(S) \neq \{x\}$, then $C(S \setminus \{x\}) = C(S) \setminus \{x\}$. If $x \notin C(S)$, then either $C(S \setminus \{x\}) \subseteq C(S)$ or $C(S) \subseteq C(S \setminus \{x\})$.*

3. *(weaker version of OC) Suppose that C selects the pair out of every pair in $\{a, b, c\}$. If $C(\{a, b, c\}) = \{b\}$ and $C(\{a, b, c, d\}) = \{d\}$, then $C(S) = \{d\}$ for any triplet S that contains d and two other elements from $\{a, b, c\}$.*

The proof is available from the authors upon request. More work is needed to check whether the axioms in Proposition 1 are also sufficient to guarantee the existence of n strict preference orderings such that $C = C^f_\succ$. We note that the properties, ATT, EC, SYM and OC are not systematically true anymore when allowing for any number of selves.

When alternatives have more than two dimensions, one may further question our assumption that all dimensions are treated equally. A natural extension would be to allow the individual to put different weights on different dimensions, and to make a choice according to, say, a "weighted" fallback solution. When the weights of the dimensions and the ranking within each dimension are not observeable, the revealed exercise would be to try and infer both from observed choices. One potential concern with this is identifiability: the additional freedom to choose the weights on the dimensions may allow the same choice correspondence to be consistent with a wide variety of preferences.

Intensities

One has the intuition that the prevalence of the attraction and compromise effects in applications may depend on factors that cannot be captured in our ordinal model. More specifically, choices may be influenced by some trade-offs that involve a notion of "distance" or "intensity". An individual may exhibit a compromise effect when $x = (100, 1)$, $y = (50, 50)$ and $z = (1, 100)$, but (perhaps) not when $y = (2, 2)$. Similarly, he may be more likely to exhibit the attraction effect when $x = (60, 40)$, $y = (59, 39)$ and $z = (40, 60)$, but (perhaps) not when $y = (41, 39)$.

To better understand how fallback bargaining could be modified so as to accomodate these alternative choice patterns, notice first how robust the patterns studied in this paper can be. Indeed, they prevail for any preference-based aggregation method (without any restriction on the number of selves) that would pick x whenever there exists a unique option x which is not bottom ranked by any self. Even if one introduces "intensities" to compute scores, e.g. the sum of the distances with respect to options in the lower contour set along each relevant dimension, one would still retain the exact same patterns of choices when maximizing the minimum of the modified scores. An interesting variant to the maximization of the minimum, is the maximization of the weighted sum of the minimal and the maximal scores, i.e.,

$$W^{\alpha}(x, S, , \succ) = \alpha \min\{s_1(x, S, \succ), s_2(x, S, \succ)\} + (1 - \alpha) \max\{s_1(x, S, \succ), s_2(x, S, \succ)\}$$

where α is a parameter between $\frac{1}{2}$ and 1. With ordinal scores, as studied in the present paper, maximizing W^{α} coincides with the Borda rule when $\alpha = 1/2$, and with the fallback solution when $\alpha = 1$.

Introducting intensities to the way scores are computed allows to capture more subtle attraction and compromise effects. As suggested above, if each alternative x is a vector of two real numbers, (x_1, x_2), the score $s_1(x, S, \succ)$ could be defined as the sum of differences, $(x_1 - y_1)$, over all y in S with $y_1 \leq x_1$. Thus, under this scoring method, if $x = (100, 1)$ and $z = (1, 100)$, an element (y, y) would be chosen as a compromise only if $y > (200 - 199\alpha)/(2 - \alpha)$. An even richer way to define scores would be to make them a function of the *distribution* of elements in the lower contour sets. These extensions are left for future research.

APPENDIX

Proof of Lemma 1

Proof of (1). *Necessity.* Suppose that $C^f_\succ(S) = \{x\}$. For each $w \in S \setminus \{x\}$ and each $y \in S \setminus \{x, w\}$, we have:

$$\min_{i=1,2} s_i(x, S \setminus \{w\}, \succ) \geq \min_{i=1,2} s_i(x, S, \succ) - 1 \geq \min_{i=1,2} s_i(y, S, \succ) \geq \min_{i=1,2} s_i(y, S \setminus \{w\}, \succ)$$

and hence $x \in C^f_\succ(S \setminus \{w\})$, as desired.

Let now $y \in S \setminus \{x\}$. Suppose that $j \in \arg\min_{i=1,2} s_i(y, S, \succ)$. If there exists $w \in S$ such that $y \succ_j w$, then we have:

$$\min_{i=1,2} s_i(x, S \setminus \{w\}, \succ) \geq \min_{i=1,2} s_i(x, S, \succ) - 1 > \min_{i=1,2} s_i(y, S, \succ) - 1 = \min_{i=1,2} s_i(y, S \setminus \{w\}, \succ)$$

and hence $y \notin C^f_\succ(S \setminus \{w\})$. If there does not exist $w \in S$ such that $y \succ_j w$, then $\min_{i=1,2} s_i(y, S \setminus \{w\}, \succ) = 0$, and $y \notin C^f_\succ(S \setminus \{w\})$, for each $w \in S \setminus \{y\}$, since $|S \setminus \{w\}| \geq 3$, and the minimal score attained at the chosen element(s) is always larger or equal to the first integer below half the number of elements in the choice set.

Sufficiency. Assuming that conditions 1(a) and (b) are true, we need to prove that $C^f_\succ(S) = \{x\}$. If $C^f_\succ(S) = \{y\}$ for some $y \in S \setminus \{x\}$, then the necessary condition for subcase 1 implies that $y \in C^f_\succ(S \setminus \{w\})$, for all $w \in S \setminus \{y\}$, thereby contradicting 1(b). If $C^f_\succ(S) = \{y, z\}$ for some $y, z \in S \setminus \{x\}$, then condition 2(a) implies that $C^f_\succ(S \setminus \{w\}) \subseteq \{y, z\}$, for all $w \in S$, thereby contradicting 1(a). Finally, suppose that $C^f_\succ(S) = \{x, y\}$ for some $y \in S \setminus \{x\}$. Condition 1(b) implies that there exists $w \in S \setminus \{y\}$ such that $y \notin C^f_\succ(S \setminus \{w\})$. Condition 2(b) implies that there exists $w' \in S \setminus \{x\}$ such that $C^f_\succ(S \setminus \{w'\}) = \{y\}$, thereby contradicting 1(a). We must conclude that $C^f_\succ(S) = \{x\}$, as desired.

Proof of (2). *Necessity of 2(a).* Suppose that $C^f_\succ(S) = \{x, y\}$. Then

$$\min_{i=1,2} s_i(x, S, \succ) = \min_{i=1,2} s_i(y, S, \succ)$$

Assume that $\arg\min_{i=1,2} s_i(x, S, \succ) = 1$ and $\arg\min_{i=1,2} s_i(y, S, \succ) = 2$ (a similar reasoning applies if 1 and 2 are exchanged). Let $k = s_1(x, S, \succ) = s_2(y, S, \succ)$. First, note that the minimal score of x (resp. y) does not change in $S \setminus \{x\}$ (resp. $S \setminus \{y\}$), while the minimal score of any other element $w \in S \setminus \{x, y\}$ does not increase. Since $\min_{i=1,2} s_i(w, S, \succ) < k$, we have that $C^f_\succ(S \setminus \{x\}) = \{y\}$ (resp. $C^f_\succ(S \setminus \{y\}) = \{x\}$).

Second, consider some $w \in S\backslash\{x,y\}$. Observe that it is impossible to have $w \succ_1 x$ and $w \succ_2 y$, since the minimal score of w in S would then be larger than the minimal score of both x and y. If $w \succ_1 x$ (resp. if $w \succ_2 y$), then the minimal score of x (resp. y) is the same in both S and $S \setminus \{w\}$, and therefore remains strictly larger than the minimal score of any element in $S \setminus \{w,x,y\}$ (since it does not increase by deleting w). Hence $C^f_\succ(S \setminus \{w\}) \subseteq \{x,y\}$, as desired. If $x \succ_1 w$ and $y \succ_2 w$ (implying that $x \succ w$ and $y \succ w$), then the minimal scores of x and y are the same in $S\backslash\{w\}$ and equal to $k-1$. Let $z \in S\backslash\{x,y,w\}$. Since $C^f_\succ(S) = \{x,y\}$, and since the minimal score of any element does not increase when a set is shrunk, $\min_{i=1,2} s_i(z, S\backslash\{w\}, \succ) \leq k-1$. If $\min_{i=1,2} s_i(z, S\backslash\{w\}, \succ) = k-1$, then $\min_{i=1,2} s_i(z, S, \succ) = k-1$ and w is ranked above z according to $\arg\min_{i=1,2} s_i(z, S, \succ)$. But since $x \succ w$ and $y \succ w$, and the minimal scores of x and y in S equal k, it follows that the minimal score of z in S must be at most $k-2$, a contradiction. It follows that $\min_{i=1,2} s_i(z, S\backslash\{w\}, \succ) < k-1$ for all $z \in S\backslash\{x,y,w\}$, and hence, $C^f_\succ(S\backslash\{w\}) = \{x,y\}$.

Necessity of 2(b). Suppose now that $C^f_\succ(S\backslash\{w\}) = \{x\}$. From the previous paragraph it follows that this is true if and only if $w \succ_1 x$ and $y \succ_2 w$. Hence there exists $w' \in S$ such that $x \succ_1 w'$ and $w' \succ_2 y$, as otherwise the minimal score of y is strictly larger than the minimal score of x, and $C^f_\succ(S \setminus \{w'\}) = \{y\}$, as desired.

Sufficiency. Assuming that conditions 2(a) and (b) are true, we need to prove that $C^f_\succ(S) = \{x,y\}$. If $z \in C^f_\succ(S)$, for some $z \in S \setminus \{x,y\}$, then the necessary condition for subcases 1 and 2 implies that $z \in C^f_\succ(S \setminus \{w\})$, for some $w \in S$, thereby contradicting 2(a). If $C^f_\succ(S) = \{x\}$, then 1(b) and 2(a) imply that $C^f_\succ(S \setminus \{w\}) = \{x\}$, for some $w \in S \setminus \{x\}$. On the other hand, 1(a) implies that $x \in C^f_\succ(S \setminus \{w'\})$, for all $w' \in S \setminus \{x\}$, and this leads to a contradiction with condition 2(b). A similar reasoning shows that $C^f_\succ(S) \neq \{y\}$, and hence $C^f_\succ(S) = \{x,y\}$, as desired.

Proof of (3). Suppose $C^f_\succ(S) = \{x\}$, $C^f_\succ(S \setminus \{y\}) = \{x,z\}$, and $C^f_\succ(S \setminus \{z\}) = \{x,y\}$. If $y \succ z$, then y looses one point along both dimensions when dropping z, and the minimal score of x remains strictly larger than that of y in $S \setminus \{z\}$, hence a contradiction with $C^f_\succ(S \setminus \{z\}) = \{x,y\}$. Similarly, it cannot be that $z \succ y$. There is no Pareto relation between x and z, and x and y either, since $C^f_\succ(S \setminus \{y\}) = \{x,z\}$ and $C^f_\succ(S \setminus \{z\}) = \{x,y\}$. Let $i \in \{1,2\}$ be such that $y \succ_i z$. Three cases remain possible: 1) $x \succ_i y \succ_i z$ and $z \succ_{-i} y \succ_{-i} x$; 2) $y \succ_i x \succ_i z$ and $z \succ_{-i} x \succ_{-i} y$; or 3) $y \succ_i z \succ_i x$ and $x \succ_{-i} z \succ_{-i} y$. Consider case 1). Since y is above x along $-i$ and $C^f_\succ(S \setminus \{z\}) = \{x,y\}$, it must be that the minimal score of y in $S \setminus \{z\}$ is attained along the i-dimension, and is equal to the minimal score of x in $S \setminus \{z\}$ which is attained along the $-i$-dimension. Adding z, the minimal score of y increases by one point, while that of x remains unchanged, hence a contradiction with $C^f_{\succ}(S) = \{x\}$. Case 3) leads to a similar contradiction. Hence only case 2) remains, as desired.

Proof of (4). Follows from the proof of (2). ■

Proof of Lemma 2

Lemma 5 *Let C be a bargaining solution that satisfies EFF, NBC, and EC. Then $|C(S)| \leq 2$, for all $S \subseteq X$.*

Proof: Suppose that one can find three elements x, y, z in $C(S)$, for some $S \subseteq X$. EFF implies that the choice out of any pair in $\{x, y, z\}$ is the pair itself, and EC implies that a single element must be chosen out of the triplet. This contradicts NBC. ■

Lemma 6 *Let C be a bargaining solution that satisfies SYM, RA, PC, EFF, ATT, NBC, and EC. Let w, x, y, z be four distinct elements of X. If $C(\{w, x, y, z\}) = \{x, y\}$, then $C(\{w, x, z\}) = \{x\}$.*

Proof: RA implies that $x \in C(\{w, x, z\})$. Lemma 5 implies that we will be done after proving that $C(\{w, x, z\})$ is not equal to $\{w, x\}$, nor $\{x, z\}$. Since the argument is similar in both cases, we will only show how to rule out the first one. Suppose on the contrary that $C(\{w, x, z\}) = \{w, x\}$. EFF implies that $C(\{w, x\}) = \{w, x\}$, $C(\{w, z\}) \neq \{z\}$, and $C(\{x, z\}) \neq \{z\}$. EC implies that it is impossible to have $C(\{w, z\}) = \{w, z\}$ and $C(\{x, z\}) = \{x, z\}$. ATT also implies that it is impossible to have $C(\{w, z\}) = \{w\}$ and $C(\{x, z\}) = \{x, z\}$, or $C(\{w, z\}) = \{w, z\}$ and $C(\{x, z\}) = \{x\}$. Hence $C(\{x, z\}) = \{x\}$ and $C(\{w, z\}) = \{w\}$. Also, $C(\{w, x, y, z\}) = \{x, y\}$ implies, by EFF, that $C(\{x, y\}) = \{x, y\}$, $C(\{y, z\}) \neq \{z\}$, and $C(\{w, y\}) \neq \{w\}$.

Notice that $C(\{w, x, y\})$ must be a singleton - because of EC if $C(\{w, y\}) = \{w, y\}$, and because of ATT if $C(\{w, y\}) = \{y\}$. Suppose $C(\{w, x, y\}) = \{y\}$. If $C(\{y, z\}) = \{y\}$, then $C(\{x, y, z\}) = \{x, y\}$, by ATT, and we get a contradiction with SYM. If $C(\{y, z\}) = \{y, z\}$, then it must be that $C(\{w, y\}) = \{w, y\}$ to avoid a contradiction with PC. ATT thus implies that $C(\{w, y, z\}) = \{w\}$, which contradicts RA. Suppose $C(\{w, x, y\}) = \{x\}$. Then by SYM, $C(\{x, y, z\}) = \{y\}$. But this contradicts ATT because $C(\{x, y\}) = \{x, y\}$ and $C(\{x, z\}) = \{x\}$. Hence, the original hypothesis that $C(\{w, x, z\}) = \{w, x\}$ is false, and we are done with the proof. ■

Lemma 7 *Let C be a bargaining solution that satisfies SYM, RA, PC, EFF, ATT, NBC, EC, and OC, and let w, x, y, z be four distinct elements of X such that the choice out of any pair is the pair itself. Then the three following statements are true:*

1. *If $C(\{w, x, y\}) = \{x\}$ and $C(\{x, y, z\}) = \{y\}$, then $C(\{w, x, z\}) = \{x\}$.*

2. *It is impossible to have $C(\{x, y, z\}) = \{y\}$, $C(\{x, w, y\}) = \{w\}$, and $C(\{y, w, z\}) = \{w\}$.*

3. *If $C(\{w, x, z\}) = \{x\}$ and $C(\{x, y, z\}) = \{y\}$, then $C(\{w, x, y\}) = \{x\}$.*

Proof: For the first statement, assume that $C(\{w, x, y\}) = \{x\}$ and $C(\{x, y, z\}) = \{y\}$. RA implies that $C(\{w, x, y, z\})$ cannot be w nor y since $w, y \in \{w, x, y\}$ and $C(\{w, x, y\}) = \{x\}$, and cannot be $\{x\}$, nor $\{z\}$, since $x, z \in \{x, y, z\}$ and $C(\{x, y, z\}) = \{y\}$. Lemma 5 implies that $C(\{w, x, y, z\})$ must contain two elements. RA rules out $\{w, y\}$, $\{x, z\}$, $\{w, x\}$, $\{y, z\}$, and $\{w, z\}$. Hence it must be $\{x, y\}$. Applying Lemma 6, we conclude that $C(\{w, x, z\}) = \{x\}$, as desired.

For the second statement, assume that $C(\{x, y, z\}) = \{y\}$, $C(\{x, w, y\}) = \{w\}$, and $C(\{y, w, z\}) = \{w\}$. It is not difficult to check that RA and Lemma 5 imply that $C(\{w, x, y, z\})$ must equal $\{w\}$ or $\{w, y\}$. The former case leads to a contradiction with OC, while the other leads to a contradiction with SYM.

For the third statement, assume that $C(\{w, x, z\}) = \{x\}$ and $C(\{x, y, z\}) = \{y\}$. EC implies that $C(\{w, x, y\})$ must be a singleton. Suppose that $C(\{w, x, y\}) = \{w\}$. Thanks to the first statement, we can combine this with $C(\{w, x, z\}) = \{x\}$, to conclude that $C(\{w, y, z\}) = \{w\}$. Hence a contradiction with the second statement (w is "in between" both x and y, and y and z, while y is "in between" x and z). If $C(\{w, x, y\}) = \{y\}$, then one gets again a contradiction with the second statement (y is "in between" both w and x, and x and z, while x is "in between" w and z). ■

Proof of Lemma 2: We want to prove that, for each set $Y \subseteq X$ with at least two elements and such that the choice out of any pair in Y is the pair itself, there exist exactly two elements in Y that are not chosen out of any triplet in Y. This is done by induction on the number of elements in Y. The result is trivial if $|Y| = 2$ or 3. Let α be a positive integer larger or equal to 3, and suppose that the result holds for all set with no more than α elements. Consider now a set Y with $\alpha + 1$ elements.

First notice that there cannot be more than two elements in Y that are not chosen out of any triplet, since the choice out of any triplet in Y is a singleton, by EC. Since Y

has more than three elements, we can choose $y, x, x' \in Y$ such that $C(\{x, y, x'\}) = \{y\}$. Let ξ, ξ' be the two elements in $Y \setminus \{y\}$ that are not chosen out of any triplet in $Y \setminus \{y\}$ (using the induction hypothesis). We will be done with the proof after showing that these two elements are not chosen out of any triplet in Y. This amounts to show that $C(\{\xi, y, z\}) \neq \{\xi\}$, for all $z \in Y \setminus \{\xi, y\}$, and $C(\{\xi', y, z\}) \neq \{\xi'\}$, for all $z \in Y \setminus \{\xi', y\}$ (since we already know that ξ and ξ' are not chosen out of any triplet in $Y \setminus \{y\}$). We prove the first statement only, the argument with ξ' instead of ξ being similar. We proceed by considering three cases.

Case 1: $\{x, x'\} = \{\xi, \xi'\}$. In that case, we know that $C(\{\xi, y, \xi'\}) = \{y\}$. Suppose to the contrary of what we want to prove that $C(\{\xi, y, z\}) = \{\xi\}$, for some $z \in Y \setminus \{\xi, y\}$. It must be that $z \neq \xi'$, and hence $C(\{\xi, z, \xi'\}) = \{z\}$, by definition of ξ, ξ'. On the other hand, the first statement of Lemma 7 implies that $C(\{\xi, z, \xi'\}) = \{\xi\}$, hence the desired contradiction.

Case 2: $\{x, x'\} \cap \{\xi, \xi'\} \neq \emptyset$, but $\{x, x'\} \neq \{\xi, \xi'\}$. Suppose for instance that $x = \xi$ (the argument for the three other cases $x = \xi'$, $x' = \xi'$, and $x' = \xi$ is similar). We know that $C(\{\xi, y, x'\}) = \{y\}$ and $C(\{\xi, x', \xi'\}) = \{x'\}$ (by definition of ξ, ξ'). Suppose to the contrary of what we want to prove that $C(\{\xi, y, z\}) = \{\xi\}$, for some $z \in Y \setminus \{\xi, y\}$. Observe that $C(\{y, x', \xi'\})$ cannot be $\{y\}$ because of the second statement of Lemma 7, and it cannot be $\{\xi'\}$ to avoid a contradiction with the first statement of Lemma 7. EC implies that $C(\{y, x', \xi'\}) = \{x'\}$. The first statement of Lemma 7 now implies that $C(\{\xi, y, \xi'\}) = \{y\}$. Hence we can assume that z is different from ξ', and we know that $C(\{\xi, z, \xi'\}) = \{z\}$, by definition of ξ, ξ'. This leads to a contradiction with the first statement of Lemma 7, since $C(\{\xi, y, z\}) = \{\xi\}$.

Case 3: $\{x, x'\} \cap \{\xi, \xi'\} = \emptyset$. Suppose to the contrary of what we want to prove that $C(\{\xi, y, z\}) = \{\xi\}$, for some $z \in Y \setminus \{\xi, y\}$. If $C(\{x, x', \xi\}) = \{\xi\}$, then we reach a contradiction with $C(\{\xi, x, \xi'\}) = \{x\}$ and $C(\{\xi, x', \xi'\}) = \{x'\}$, via the first statement of Lemma 7. Hence $C(\{x, x', \xi\}) = \{x\}$ or $\{x'\}$. We consider only the first case, the argument for the second case being similar. The third statement of Lemma 7 implies $C(\{x, y, \xi\}) = \{x\}$, since $C(\{x, y, x'\}) = \{y\}$. Hence $C(\{\xi, y, \xi'\}) \neq \{\xi\}$, as otherwise one would get a contradiction with the second statement of Lemma 7 (with x being "in between" both y and ξ, and ξ and ξ', while ξ is "in between" y and ξ'). So $z = \xi'$ is impossible. If $z \neq \xi'$, then $C(\{\xi, z, \xi'\}) = \{z\}$. Once combined with $C(\{\xi, y, z\}) = \{\xi\}$, the first statement of Lemma 7 implies that $C(\{\xi, y, \xi'\}) = \{\xi\}$, a contradiction again. ■

Proof of Lemma 3

Let $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$, and let $x \in B_w$. We will be done with the first part of the statement after proving that either $C(\{x, w, z\}) = \{w\}$ or $C(\{x, w, z'\}) = \{w\}$ (meaning that one can actually choose y in $\{z, z'\}$). Notice first that $C(\{x, w, z\})$ must be a singleton, by EC if $C(\{x, z\}) = \{x, z\}$, or by ATT if $C(\{x, z\})$ is a singleton. A similar argument implies that $C(\{x, w, z'\})$ is a singleton as well. Suppose now, on the contrary to what we want to prove, that $C(\{x, w, z\}) \in \{x, z\}$ and $C(\{x, w, z'\}) \in \{x, z'\}$. Notice that we must have $C(\{x, w, z\}) = C(\{x, w, z'\})$, as otherwise we would have a contradiction with Lemma 5 and RA (there is no way to select at most two elements out of $\{w, x, z, z'\}$, that lead to a nonempty intersection with three different singleton choices in three subsets of cardinality three). Hence it must be that both $C(\{x, w, z\})$ and $C(\{x, w, z'\})$ equal $\{x\}$. It is not difficult to check that this, combined $C(\{z, w, z'\}) = \{w\}$, implies that $C(\{w, x, z, z'\}) = \{x\}$ or $\{w, x\}$, again as a consequence of Lemma 5 and RA. SYM makes the second case impossible. Indeed, w does not belong to neither $C(\{x, w, z\})$, nor $C(\{x, w, z'\})$. So we are forced to conclude that $C(\{w, x, z, z'\}) = \{x\}$, but then we get a contradiction with OC since $x, z, z' \in B_w$. We are thus done with the proof of the first part of the statement.

As for the second part, let $y, y' \in B_w$ be such that $C(\{x, w, y\}) = \{w\}$ and $C(\{x, w, y'\}) = \{w\}$. Suppose, to the contrary of what we want to prove, that $x \succ_1 y$ and $y' \succ_1 x$. Notice that $C(\{x, y\}) = \{x, y\}$, as otherwise $C(\{x, w, y\}) = \{x\}$ or $\{y\}$, by ATT. Similarly, $C(\{x, y'\}) = \{x, y'\}$. Hence $y \succ_2 x$ and $x \succ_2 y'$. By the induction hypothesis, $C(\{x, y, y'\}) = C^f_{\succ}(\{x, y, y'\})$. Hence $C(\{x, y, y'\}) = \{x\}$. Combining this with $C(\{x, w, y\}) = \{w\}$ and $C(\{x, w, y'\}) = \{w\}$, Lemma 5 and RA imply that $C(\{w, x, y, y'\}) = \{w\}$ or $\{w, x\}$. The second case would lead to a contradiction with SYM, and hence $C(\{w, x, y, y'\}) = \{w\}$, but this leads to a contradiction with OC, since $C(\{w, x\}) = \{w, x\}$, $C(\{x, y\}) = \{x, y\}$, $C(\{x, y'\}) = \{x, y'\}$, and $C(\{y, y'\}) = \{y, y'\}$. We are thus done with the proof of the second and last part of the statement. ■

Proof of Lemma 4

Lemma 8 *Let C be a bargaining solution that satisfies SYM, RA, EFF, NBC, EC, and OC. Suppose that the choice out of any pair in $\{x, y, y'\}$ is the pair itself, and that $C(\{x, y, y'\}) = \{x\}$. If $x \in A_w$ and $y, y' \in B_w$, then $C(\{y, y', w\}) = \{w\}$.*

<u>Proof</u>: ATT implies that $C(\{x, y, w\}) = C(\{x, y', w\}) = \{x\}$. Since $C(\{x, y, y'\}) = \{x\}$, it follows from Lemma 5, RA and SYM that $C(\{x, y, y', w\}) = \{x\}$. EC implies that $C(\{y, y', w\})$ is a singleton. If $C(\{y, y', w\}) = \{y\}$, then we get a contradiction with

OC, since $C(\{x,y\}) = \{x,y\}$. By a similar argument, $C(\{y,y',w\}) \neq \{y'\}$, and hence $C(\{y,y',w\}) = \{w\}$. ■

Proof of Lemma 4: Assume, by contradiction, that there exist $\xi, \xi' \in A_w$ and $y, y' \in B_w$ such that $y \succ_2 \xi$ and $y' \succ_1 \xi'$. Hence $C(\{\xi,y\}) \neq \{\xi\}$, by definition of $\succ$ on S. Also, $C(\{\xi,y\}) \neq \{y\}$, as otherwise we would get a contradiction with $y \in B_w$ via PC, since $\xi \in A_w$. Hence $C(\{\xi,y\}) = \{\xi,y\}$. A similar argument implies that $C(\{\xi',y'\}) = \{\xi',y'\}$. By definition of $\succ$ on S, we have:

$$\begin{array}{cc} \xi \succ_1 y & y \succ_2 \xi \\ y' \succ_1 \xi' & \xi' \succ_2 y' \end{array} \tag{2}$$

The proof proceeds by considering two cases.

Case 1 $C(\{\xi,y'\}) = \{\xi\}$ *and* $C(\{\xi',y\}) = \{\xi'\}$

By definition of $\succ$, we have: $\xi \succ y'$ and $\xi' \succ y$. Combining this with (2), it follows that $\xi \succ_1 y' \succ_1 \xi' \succ_1 y$ and $\xi' \succ_2 y \succ_2 \xi \succ_2 y'$ (by the induction hypothesis, the two relations in $\succ$ are transitive). Since $C = C^f_\succ$ on triplets in S, we conclude that $C(\{\xi,y,y'\}) = \{\xi\}$ and $C(\{\xi',y,y'\}) = \{\xi'\}$. ATT implies that $C(\{w,\xi,y\}) = \{\xi\}$ and $C(\{w,\xi',y'\}) = \{\xi'\}$, whereas EFF implies that $C(\{w,\xi,y'\}) = \{\xi\}$ and $C(\{w,\xi',y\}) = \{\xi'\}$. SYM, Lemma 5, and RA imply that $C(\{w,\xi,y,y'\}) = \{\xi\}$ and $C(\{w,\xi',y,y'\}) = \{\xi'\}$. This leads to a contradiction with OC if $C(\{w,y,y'\}) = \{y\}$ or $\{y'\}$, since $y,y' \in B_w$, $C(\{y,y'\}) = \{y,y'\}$, $C(\{\xi,y\}) = \{\xi,y\}$, and $C(\{\xi',y'\}) = \{\xi',y'\}$. EC implies that $C(\{w,y,y'\})$ is a singleton, and hence $C(\{w,y,y'\}) = \{w\}$, but this contradicts the assumption of Lemma 4. Hence this first case is impossible, and we have to look into the second case.

Case 2 $C(\{\xi,y'\}) \neq \{\xi\}$ *and/or* $C(\{\xi',y\}) \neq \{\xi'\}$.

We consider the case where $C(\{\xi,y'\}) \neq \{\xi\}$. A similar reasoning applies if $C(\{\xi',y\}) \neq \{\xi'\}$. $C(\{\xi,y'\}) = \{y'\}$ would lead to a contradiction with $y' \in B_w$ via PC, since $\xi \in A_w$. Hence $C(\{\xi,y'\}) = \{\xi,y'\}$. If $\xi \succ_1 y'$, then by the induction hypothesis that $C = C^f_\succ$ on pairs in S, it follows that $y' \succ_2 \xi$. From (2) and the transitivity of the relations in $\succ$, it follows that $\xi \succ_1 y' \succ_1 \xi'$ and $\xi' \succ_2 y' \succ_2 \xi$. The induction hypothesis also implies that $C = C^f_\succ$ on triplets in S, and hence, $C(\{\xi,\xi',y'\}) = \{y'\}$. On the other hand, ATT implies that $C(\{w,\xi,y'\}) = \{\xi\}$ and $C(\{w,\xi',y'\}) = \{\xi'\}$. There is no way of defining

$C(\{w,\xi,\xi',y'\})$ so as to satisfy Lemma 5 and RA. Hence it must be that $y' \succ_1 \xi$. Since $C = C^f_\succ$ on pairs in S, we have that $\xi \succ_2 y'$. Hence, $y' \succ_1 \xi \succ_1 y$ and $y \succ_2 \xi \succ_2 y'$, by (2) and the transitivity of the relations in $\succ$. Also, $C = C^f_\succ$ on triplets in S, and hence $C(\{\xi,y,y'\}) = \{\xi\}$. Lemma 8 implies $C(\{y,y',w\}) = \{w\}$, a contradiction with the assumption of Lemma 4. Case 2 is thus impossible as well. ■

$\succ_1^*$ and $\succ_2^*$ are transitive

Transitivity is the subject of Lemmas 10 and 11. Before stating and proving these lemmas, we need to establish a useful property.

Lemma 9 *Let C be a bargaining solution that satisfies ATT, NBC, RA, EFF, SYM, EC, PC and OC. Let x,y,z,z' be four elements of X such that the solution out of any pair in $\{x,y,z\}$ is the pair itself, $C(\{y,z'\}) = \{y,z'\}$, and $C(\{z,z'\}) = \{z'\}$. Then $C(\{x,y,z\}) = \{y\}$ if and only if $C(\{x,y,z'\}) = \{y\}$.*

<u>Proof</u>: Notice that $C(\{x,z'\}) \neq \{x\}$, as otherwise we would get a contradiction with $C(\{x,z\}) = \{x,z\}$ via PC, since $C(\{z,z'\}) = \{z'\}$. Independently of whether $C(\{x,z'\}) = \{z'\}$ or $\{x,z'\}$, ATT implies that $C(\{x,z,z'\}) = C(\{y,z,z'\}) = \{z'\}$.

If $C(\{x,y,z\}) = \{y\}$, then Lemma 5 and RA imply that $C(\{x,y,z,z'\}) = \{z'\}$ or $\{y,z'\}$. The former case leads to a contradiction with OC. In the latter case, SYM implies that $z' \notin C(\{x,y,z'\})$, since $C(\{y,z,z'\}) = \{z'\}$. $C(\{x,z'\}) = \{z'\}$ would imply $C(\{x,y,z'\}) = \{z'\}$, by ATT, a contradiction. Hence $C(\{x,z'\}) = \{x,z'\}$, and EC implies that $C(\{x,y,z'\})$ must be a singleton, or $C(\{x,y,z'\}) = \{y\}$ given RA, as desired.

If $C(\{x,y,z'\}) = \{y\}$, then Lemma 5 and RA imply that $C(\{x,y,z,z'\}) = \{y,z'\}$. Lemma 6 implies in turn that $C(\{x,y,z\}) = \{y\}$, as desired. ■

Lemma 10 *Let $(\succ_1,\succ_2)$ be two complete, transitive and anti-reflexive orderings defined over $S \subseteq X$ such that $C = C^f_\succ$ on pairs and triplets in S, let $w \in X \setminus S$, let $(\succ_1^*,\succ_2^*)$ be the extensions of $(\succ_1,\succ_2)$, as defined in the main text, let x,y be two elements of S, and let $i \in \{1,2\}$. If $x \succ_i y$ and $y \succ_i^* w$, then $x \succ_i^* w$. Similarly, if $w \succ_i^* y$ and $y \succ_i x$, then $w \succ_i^* x$.*

<u>Proof</u>: The second statement being symmetric to the first, its proof is very similar and is therefore omitted. We are thus assuming that $x \succ_i y$ and $y \succ_i^* w$, and we want to prove that $x \succ_i^* w$. If $x \in A_w$, then we are done. So we'll assume $x \in B_w$.

Suppose that there is no $z,z' \in B_w$ such that $C(\{z,w,z'\}) = \{w\}$. If $y \in A_w$, then $x \succ_i^* w$, by definition of $\succ_i^*$. Suppose now that $y \in B_w$. Our construction of $\succ^*$ is such that either $z \succ_i^* w$ for all $z \in B_w$, or $w \succ_i^* z$ for all $z \in B_w$. Hence $x \succ_i^* w$, as desired. So from now on we assume that there exist $z, z' \in B_w$ such that $C(\{z, w, z'\}) = \{w\}$.

By Lemma 3, there exists $x' \in B_w$ such that $C(\{x, w, x'\}) = \{w\}$. If $x \succ_i x'$, then $x \succ_i^* w$, by construction, and we are done. So we prove in the remainder that $x' \succ_i x$ is impossible. So we will assume, on the contrary, that $x' \succ_i x$ and $x \succ_{-i} x'$.

Suppose first that $y \in A_w$. In that case, $C(\{x, y\})$ is different from $\{x\}$, as otherwise we get a contradiction with $x \in B_w$ via PC. $C(\{x, y\})$ is also different from $\{y\}$, since $x \succ_i y$, and $C = C^f_\succ$ on all pairs in S. Hence $C(\{x, y\}) = \{x, y\}$. Since $C = C^f_\succ$ on all pairs in S, we conclude $y \succ_{-i} x$. Given that $x' \succ_i x$ and $x \succ_{-i} x'$, the transitivity of $\succ$ implies that $x' \succ_i y$ and $y \succ_{-i} x'$. Since $C = C^f_\succ$ on all pairs in S, $C(\{x', y\}) = \{x', y\}$. Given that $y \in A_w$, ATT now implies that $C(\{x, y, w\}) = C(\{x', y, w\}) = \{y\}$. Since $C = C^f_\succ$ on all triplets in S, it follows that $C(\{x', x, y\}) = \{x\}$. But because $C(\{x, w, x'\}) = \{w\}$, there is no way to define $C(\{x, x', y, w\})$ so as to satisfy RA, given Lemma 5, and we get the desired contradiction.

Suppose next that $y \in B_w$. Then, it follows from $y \succ_i^* w$ that $w \succ_{-i}^* y$, by construction. If $C(\{x', y\}) = \{x'\}$, then $x' \succ y$, by construction, and hence $x \succ y$ (by assumption for i and by transitivity for $-i$). Since $C = C^f_\succ$ on pairs in S, we conclude that $C(\{x, y\}) = \{x\}$. ATT implies that $C(\{x, y, w\}) = \{x\}$ and $C(\{x', y, w\}) = \{x'\}$. It becomes impossible to define $C(\{x, x', y, w\})$ so as to satisfy RA and Lemma 5, given that $C(\{x, x', w\}) = \{w\}$. So we must conclude that $C(\{x', y\}) \neq \{x'\}$, and hence $C(\{x', y\}) = \{x', y\}$ since $x' \succ_i y$ (this follows from our assumptions that $x \succ_i y$ and $x' \succ_i x$, and from the transitivity of $\succ$). Suppose $C(\{x, y\}) = \{x\}$. Note that $C(\{w, x\}) = \{w, x\}$, the choice from every pair in $\{x', w, y\}$ is the pair itself, and the same is true for $\{x', w, x\}$. It then follows from Lemma 9 that $C(\{x', y, w\}) = \{w\}$, and we get a contradiction with $y \succ_i^* w$, since $x' \succ_i y$ (see Lemma 3). As in the previous paragraph, we cannot have $C(\{x, y\}) = \{y\}$ either, because $x \succ_i y$. Hence $C(\{x, y\}) = \{x, y\}$. So $x' \succ_i x \succ_i y$ and $y \succ_{-i} x \succ_{-i} x'$, and $C(\{x, x', y\}) = \{x\}$ since $C = C^f_\succ$ on triplets in S. In addition, we also know that $C(\{x', w, x\}) = \{w\}$. Since $x', y \in B_w$ and $C(\{x', y\}) = \{x', y\}$, then $C(\{x', w, y\})$ must be a singleton, by EC. If $C(\{x', w, y\}) \in \{x', y\}$, then there is no way of defining $C(\{x, x', y, w\})$ so as to satisfy Lemma 5 and RA. Hence, $C(\{x', w, y\}) = \{w\}$, and we get a contradiction with $y \succ_i^* w$, since $x' \succ_i y$ (see Lemma 3). ■

Lemma 11 *Let $(\succ_1, \succ_2)$ be two complete, transitive and anti-reflexive orderings defined over $S \subseteq X$ such that $C = C^f_\succ$ on pairs and triplets in S, let $w \in X \setminus S$, let $(\succ_1^*, \succ_2^*)$ be* *the extensions of $(\succ_1, \succ_2)$, as defined in the main text, let x, y be two elements of S, and let $i \in \{1, 2\}$. If $x \succ_i^* w$ and $w \succ_i^* y$, then $x \succ_i y$.*

Proof: We wish to show that $x \succ_i y$. If $C(\{x,y\}) = \{x\}$, then we are done. Assume $C(\{x,y\}) \neq \{x\}$.

We first consider the case where $x \in A_w$. Hence $C(\{x,y\}) \neq \{y\}$, or $C(\{x,y\}) = \{x,y\}$, since otherwise we get a contradiction with $w \succ_i^* y$ via PC. Now assume that the conclusion of the lemma is wrong, i.e. $y \succ_i x$. Notice that there must exist $y' \in B_w$ such that $C(\{y,w,y'\}) = \{w\}$, as otherwise $y \succ_i^* w$, by definition of $\succ^*$, a contradiction. Since $w \succ_i^* y$, it must be that $y' \succ_i y$ and $y \succ_{-i} y'$, again by definition of $\succ^*$. Since $y \succ_i x$, $x \succ_{-i} y$, and $C = C^f_\succ$ on triplets in S, it follows that $C(\{x,y,y'\}) = \{y\}$. Given that w is added after y in our induction, it cannot be that $C(\{w,y\}) = \{w\}$. Since $w \succ_i^* y$, it cannot be that $C(\{w,y\}) = \{y\}$ either. Hence $y \in B_w$. ATT implies that $C(\{x,w,y\}) = \{x\}$, but then there is no way of defining $C(\{x,y,y',w\})$ so as to satisfy Lemma 5 and RA. We, therefore, conclude that $x \succ_i y$, as desired.

Consider next the case where $x \in B_w$. As in the previous paragraph, $y \in B_w$. By our construction of $\succ^*$, there must exist $x', y' \in B_w$ such that $C(\{x,w,x'\}) = \{w\}$ and $C(\{y,w,y'\}) = \{w\}$. If this was not true, then w would be ranked above or below both x and y according to $\succ_i^*$, thereby contradicting our assumption that $x \succ_i^* w$ and $w \succ_i^* y$.

Suppose that $C(\{x,y\}) = \{y\}$. Lemma 9 implies that $C(\{x',y,w\}) = \{w\}$. Since $w \succ_i^* y$, we must have $x' \succ_i y$. We must also have $x \succ_i x'$, since $C(\{x,x',w\}) = \{w\}$ and $x \succ_i^* w$. Transitivity of $\succ_i$ implies that $x \succ_i y$, as desired.

Suppose now that $C(\{x,y\}) = \{x,y\}$, and that $y \succ_i x$, contrarily to what we want to prove. Then $y' \succ_i y \succ_i x \succ_i x'$ and $x' \succ_{-i} x \succ_{-i} y \succ_{-i} y'$ in order to have $x \succ_i^* w$ and $w \succ_i^* y$. The solution out of any pair in $\{x,y,w\}$ is the pair itself. So $C(\{x,y,w\})$ is a singleton, by EC. It cannot be w, as this would imply $w \succ_i^* x$. Suppose that $C(\{x,y,w\}) = \{y\}$. Since $C(\{y,w,y'\} = \{w\}$, the first statement of Lemma 7 implies that $C(x,w,y') = \{w\}$, hence a contradiction with $x \succ_i^* w$, since $y' \succ_i x$. Suppose now that $C(\{x,y,w\}) = \{x\}$. Since $C(\{x,w,x'\} = \{w\}$, the first statement of Lemma 7 implies that $C(\{x',y,w\}) = \{w\}$, hence a contradiction with $w \succ_i^* y$, since $y \succ_i x'$. ■

References

Ariely, Dan. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* Harper Collins, NY.

Ambrus, Attila and Kareen Rozen. 2009. "Rationalizing Choice with Multi-Self Models." Mimeo.

Anbarci, Nejat. 2006. "Finite Alternating-Move Arbitration Schemes and the Equal Area Solution." *Theory and Decision*, 61: 21-50.

Benhabib, Jess and Alberto Bisin. 2004. "Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions." *Games and Economic Behavior*, 52(2): 460-92.

Bernheim, Douglas and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, 94(5): 1558-1590.

Brams, Steven J. and D. Marc Kilgour. 2001. "Fallback Bargaining." *Group Decision and Negotiation*, 10(4): 287-316.

Chambers, Christopher P. and Federico Echenique. 2008. "The Core Matchings of Markets with Transfers." Mimeo.

Cherepanov, Vadim, Timothy Feddersen and Alvaro Sandroni. 2008. "Rationalization." Mimeo.

Chiappori, Pierre-André. 1988. "Rational Household Labor Supply." *Econometrica,* 56: 63-89.

Chiappori, Pierre-André and Olivier Donni. 2005. "Learning From a Piece of Pie : The Empirical Content of Nash Bargaining." Mimeo.

Chiappori, Pierre-André and Ivar Ekeland. Forthcoming. "The Micro Economics of Efficient Group Behavior: Identification." *Econometrica.*

Eliaz, Kfir, Michael Richter and Ariel Rubinstein. 2009. "An Étude in Choice Theory: Choosing the Two Finalists." Mimeo.

Eliaz, Kfir and Ran Spiegler. 2006. "Contracting with Diversely Naïve Agents." *Review of Economic Studies*, 73(3): 689-714.

Fudenberg, Drew and David K. Levine. 2006. "A Dual-Self Model of Impulse Control," *American Economic Review,* 96: 1449-76.

Green, Jerry R. and Daniel A. Hojman. 2007. "Choice, Rationality and Welfare Measurement." Mimeo.

Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica,* 69(6): 1403–1435.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research,* 9(1): 90-98.

Hurwicz, Leonid and Murat R. Sertel. 1997. "Designing Mechanisms, in particular for Electoral Systems: The Majoritarian Compromise." Mimeo.

Kalai, Gil, Ariel Rubinstein and Ran Spiegler. 2002. "Rationalizing Choice Functions By Multiple Rationales." *Econometrica,* 70(6): 2481-2488.

Kamenica, Emir. 2008. "Contextual Inference in Markets: On the Informational Content of Product Lines." *American Economic Review*, 98(5): 2127-2149.

Karni, Edi and David Schmeidler. 1975. "Independence of Nonfeasible Alternatives, and Independence of Nonoptimal Alternatives." *Journal of Economic Theory*, 12: 488-493.

Kibris, Özgur and Murat R. Sertel. 2007. "Bargaining Over a Finite Set of Alternatives." *Social Choice and Welfare*, 28: 421-437.

Kivetz, Ran, Oded Netzer, and V. Srinivasan. 2004a. "Alternative Models for Capturing the Compromise Effect." *Journal of Marketing Research,* 41(3): 237-57.

Kivetz, Ran, Oded Netzer, and V. Srinivasan. 2004b. "Extending Compromise Effect Models to Complex Buying Situations and Other Context Effects." *Journal of Marketing Research,* 41(3): 262-268.

Lombardi, Michele. 2007. "Reason-Based Choice Correspondence." Mimeo.

Manzini, Paula and Marco Mariotti. 2007. "Sequentially Rationalizable Choice." *American Economic Review,* 95(5): 1824-39.

Manzini, Paula and Marco Mariotti. 2008. "Categorize Then Choose: Boundedly Rational Choice and Welfare." Mimeo.

Mariotti, Marco. 1998. "Nash Bargaining Theory Theory When the Number of Alternatives Can Be Finite." *Social Choice and Welfare,* 15: 413-421.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay. 2009. "Revealed Attention." Mimeo.

Ok, Efe A., Pietro Ortoleva, and Gil Riella. 2008. "Rational Choice with Endogenous Reference Points." Mimeo.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica*, 50(1): 97-109.

Segal, Uzi. 1987. "Some Remarks on Quiggin's Anticipated Utility [A Theory of Anticipated Utility]." *Journal of Economic Behavior and Organization*, 8(1): 145–54.

Sen, Amartya K. 1971. "Choice Functions and Revealed Preference." *Review of*

Economic Studies, 38(3): 307–317.

Shafir, Eldar, Itamar Simonson, and Amos Tversky. 1993. "Reason-based Choice." *Cognition,* 49: 11-36.

Simonson, Itamar. 1989. "Choice Based on Reasons: The Case for the Attraction and Compromise Effects." *Journal of Consumer Research,* 16(2): 158-74.

Sprumont, Yves. 1993. "Intermediate Preferences and Rawlsian Arbitration Rules." *Social Choice and Welfare,* 10(2): 1-15.

Sprumont, Yves. 2000. "On the Testable Implications of Collective Choice Theories." *Journal of Economic Theory,* 93(2): 205-32.

Tversky, Amos and Eldar Shafir. 1992. "Choice Under conflict: The Dynamics of Deferred Decision." *Psychological Science,* 3(6): 358-61.