Hypothetical Bargaining and Equilibrium Refinement in Non-Cooperative Games [*]

Mantas Radzvilas[†]

December 2016

Abstract

Virtual bargaining theory suggests that social agents aim to resolve non-cooperative games by identifying the strategy profile(s) which they would agree to play if they could openly bargain. The theory thus offers an explanation of how social agents resolve games with multiple Nash equilibria. One of the main questions pertaining to this theory is how the principles of the bargaining theory could be applied in the analysis of hypothetical bargaining in non-cooperative games. I propose a bargaining model based on the benefit-equilibrating bargaining solution (BES) concept for non-cooperative games, broadly in line with the principles underlying Conley and Wilkie's (2012) ordinal egalitarian solution for Pareto optimal point selection problems with finite choice sets. I provide formal characterizations of the ordinal and the cardinal versions of BES, discuss their application to n-player games, and compare model's theoretical predictions with the data available from several experiments involving 'pie games'.

1 Introduction

A central solution concept of the standard game theory is the Nash equilibrium – a pure or mixed strategy profile which is such that no rational player is motivated to unilaterally deviate from it by playing a different strategy. At least intuitivelly, however, some Nash equilibria are more convincing rational solutions of games than others: Even a simple game may have a Nash

[*] This is a work-in-progress paper, do not cite without permission.

[†] London School of Economics, Department of Philosophy, Logic and Scientific Method (e-mail: m.radzvilas@lse.ac.uk).

equilibrium which seems unlikely to be played by players who understand the structure of the game and believe each other to be intelligent decision-makers.

Consider the Hi-Lo game depicted in Figure 1, in which two players simultaneously and independently choose between the two pure strategies: *hi* and *lo*. The left and the right number in each cell represents row and column player's payoffs respectively[1].

	hi	*lo*
hi	$2, 2$	$0, 0$
lo	$0, 0$	$1, 1$

Figure 1: Hi-Lo game

There are two pure strategy Nash equilibria in this game: (hi, hi) and (lo, lo). There is a third Nash equilibrium in mixed strategies, in which both players randomize between the pure strategies *hi* and *lo* with probabilities 1/3 and 2/3 respectively. From the perspective of standard game theory, every Nash equilibrium is a rational solution of the game. For many people, the attainment of the Nash equilibrium (hi, hi) appears to be an intuitively 'obvious' definitive resolution of this game: It is the best outcome for both players and there is no conflict of interests in this game. Experimental results support this intuition by revealing that over 90% of the time people opt for strategy *hi* in this game[2]. The standard game theory cannot single out the Nash equilibrium (hi, hi) neither as *unique* rational solution, nor as a more likely outcome of this game.

This prompted the emergence of multiple theories which purport to explain how players resolve games with multiple rational solutions. One of the more recent approaches is the theory of virtual bargaining suggested by Misyak and Chater (2014) and Misyak et al. (2014). It is a hypothetical, or fictitious, bargaining model which aims to provide an *individualistic* explanation of how players may resolve a non-cooperative game by identifying a *feasible* and *mutually advantageous* solution – an outcome which can be

[1] Unless it is stated otherwise, the payoff numbers in the matrices are Von Neumann and Morgenstern utilities. The payoffs are assumed to represent all the relevant motivations of players, including pro-social preferences, such as inequity aversion, altruism, sensitivity to social norms, and so on.

[2] See Bardsley et al. (2010) who, among a number of other games, report results from experiments with two versions of the Hi-Lo game, where the outcome (hi, hi) yields each player a payoff of 10 while the outcome (lo, lo) yields either 9 or 1, depending on the gameâs version.

implemented via joint actions of self-oriented decision-makers and is individually advantageous for every interacting individual. The theory suggests that decision-makers choose their strategies on the basis of what strategy profile(s) they would agree to play if they could openly bargain – engage in real negotiations, in which each player can communicate his or her offers to the other players and receive their counteroffers.

The idea of hypothetical bargaining warrants further theoretical and empirical investigation for three reasons. First, every standard bargaining solution is, essentially, an equilibrium refinement. In bargaining games where players' agreements are not binding, the set of feasible agreements is the set of correlated equilibria. A bargaining solution is a correlated equilibrium which satisfies a number of *desirable properties*, which can be interpreted as a *expectation* of the outcome of an open bargaining process involving self-oriented individuals of roughly equal bargaining power (for extensive discussion, see Myerson (1991). It seems reasonable to believe that certain formal properties of bargaining solutions that decision-makers find desirable may also be deemed relevant by players searching for mutually advantageous solutions of non-cooperative games.

Second, bargaining theory is a branch of non-cooperative game theory: The bargaining solution concepts rely on the same basic principles of orthodox game theory as solution concepts of non-cooperative games. A bargainer is a self-oriented decision-maker – an individual who aims to maximally advance his/her personal interests, and only cares about the interests of the other interacting individuals insofar as their actions may promote or hinder the advancement of his or her own personal interests. Like a best-response reasoner, a hypothetical bargainer deviates from the agreement if a unilateral deviation is personally beneficial. For this reason, hypothetical bargaining solutions are compatible with the orthodox notion of individual rationality, and have conceptually appealing stability properties.

Third, the idea that people aim to resolve non-cooperative games by identifying mutually advantageous solutions seems to be supported by experimental results. The experiment of Colman and Stirk (1998) with coordination games suggests that a substantial proportion of people use some notion of mutual advantage when reasoning about their choice options in non-cooperative games.

One of the fundamental questions which requires further conceptual and empirical exploration is what properties a strategy profile must have in order to be identified by hypothetical bargainers as the hypothetical bargaining solution of a game. Without an adequate answer to this question, a rigorous empirical testing of this new theory is not possible. Misyak and Chater (2014) use the Nash (1950) bargaining solution as an approximation to what hypothetical bargainers would identify as the bargaining solution of a non-cooperative game.

In this paper, I argue that the use of Nash bargaining solution for the analysis of hypothetical bargaining in non-cooperative games is problematic. I will suggest an alternative *benefit-equilibrating hypothetical bargaining solution* (later abbreviated as BES) concept for non-cooperative games, broadly in line with the principles underlying the ordinal egalitarian solution for Pareto-optimal point selection problems with finite choice sets suggested by Conley and Wilkie (2012). I will argue that the proposed solution concept can be applied to cases where interpersonal comparisons of decision-makers' payoffs in the original game are assumed not to be meaningful. I offer both the ordinal and the cardinal version of this solution concept and discuss its properties, application to n-player games, and theoretical predictions using a number of experimentally relevant examples.

The rest of the paper is structured as follows. In section 2 I discuss the virtual bargaining theory and the reasons of why the application of the standard Nash bargaining solution to non-cooperative games is both conceptually and empirically problematic. In sections 3 and 4 I propose the ordinal and cardinal versions of BES and discuss their formal properties. I also discuss the application of BES to n-player games. In section 5 I discuss the BES predictions in several experimentally relevant games. With section 6 I conclude and discuss the explanatory scope of the proposed model.

2 Hypothetical Bargaining

According to the theory of virtual bargaining, a player who reasons as a hypothetical bargainer interprets all the pure and mixed strategy profiles of a game as *possible agreements* – outcomes which the players could attain via specific combinations of their actions. A hypothetical bargainer then identifies a set of *feasible agreements*. Each feasible agreement is a strategy profile, such that no player can exploit the other players by unilaterally deviating from it. A decision-maker who reasons as a hypothetical bargainer then identifies a feasible agreement (or agreements) which s/he believes the players would agree to play in open bargaining, and plays his/her part in realizing that agreement, provided that s/he has a reason to believe that the other players are hypothetical bargainers who will carry out their end of the agreement. Each agreement identified by hypothetical bargainers as the hypothetical bargaining solution of a game is assumed to be the mutually beneficial and agreeable solution of a game. Misyak and Chater suggest that the 'goodness of a feasible bargain is, following Nash's theory of bargaining, the product of the utility gains to each player (relative to a no-agreement baseline) of adhering to that agreement' (Misyak and Chater 2014: 4).

The Nash bargaining solution has been developed to resolve a specific type of game, known as the Nash bargaining problem. In the standard bargaining problem, each player's utility function is defined over the set of lotteries over all the possible distributions of some divisible resource. The Nash bargaining solution of the standard bargaining problem is a *unique* Pareto optimal distribution of the good. In other types of non-cooperative games, however, players' utility functions may represent all kinds of motivations which are relevant for player's evaluation of the possible outcomes. Because of this, a non-cooperative game may have multiple Pareto optimal feasible agreements, even multiple agreements which maximize the Nash product, and each agreement may be associated with a *different allocation* of personal payoff gains. Since hypothetical bargainers are assumed to be self-oriented decision-makers, it stands to reason to assume that they would not be indifferent between agreements associated with different allocations of personal payoff gains, and so the question of how a conflict over different alternative allocations of players' personal payoff gains would be resolved becomes important.

For example, consider the two player coordination game with three Pareto optimal outcomes[3] depicted in Figure 2a. The game has four pure strategy Nash equilibria: three Pareto optimal Nash equilibria $(s1, t1)$, $(s2, t2)$, $(s3, t3)$ and a Pareto inefficient Nash equilibrium $(s4, t4)$. The game also has eleven Nash equilibria in mixed strategies. Every mixed strategy Nash equilibrium yields each player a lower personal payoff that any of the pure strategy Nash equilibria. Notice that the Nash equilibrium $(s4, t4)$ is the profile of players' *maximin* strategies, which maximize the payoff that each player can guarantee to himself/herself, irrespective of what the other player does.

Suppose that both players identify the Nash equilibrium $(s4, t4)$ as the disagreement profile. Relative to the disagreement point, the product of players' payoff gains associated with the Nash equilibria $(s1, t1)$, $(s2, t2)$ and $(s3, t3)$ is 9. The players should identify all three Pareto optimal Nash equilibria as the Nash bargaining solutions of this game. Notice, however, that each of the three solutions is associated with a different allocation of players' personal payoff gains. The Nash equilibria $(s1, t1)$ and $(s3, t3)$ maximize the personal payoff of one of the players, but yield a payoff which is only

[3] An allocation of payoffs associated with an outcome is said to be Pareto optimal if there is no alternative outcome associated with an allocation of payoffs which makes at least one interacting player better off without making any other player worse off.

	t1	*t2*	*t3*	*t4*
s1	$10, 2$	$0, 0$	$0, 0$	$0, 1$
s2	$0, 0$	$4, 4$	$0, 0$	$0, 1$
s3	$0, 0$	$0, 0$	$2, 10$	$0, 1$
s4	$1, 0$	$1, 0$	$1, 0$	$1, 1$

a

	L	*R*	*B*
L	$5, 6$	$0, 0$	$0, 0$
R	$0, 0$	$6, 5$	$0, 0$
B	$0, 0$	$0, 0$	$5, 5$

b

	L	*R*	*B*
L	$5, 6$	$0, 0$	$0, 0$
R	$0, 0$	$6, 5$	$0, 0$
B	$0, 0$	$0, 0$	$7, 5$

c

Figure 2: Coordination games with multiple weakly Pareto optimal outcomes

slightly higher than the *maximin* payoff for the other. If one of the two Nash equilibria were implemented, one of the players would forego an opportunity to maximally advance his/her personal interests with an alternative available agreement, and would get a payoff gain which, relative to his/her disagreement payoff, yields only 1/9 of the maximum payoff gain attainable to him/her in this game[4]. The Nash equilibrium $(s2, t2)$ is the second-best solution of the game for both players, which yields each player 1/3 of the total maximum attainable payoff gain. Since hypothetical bargainers are assumed to be self-oriented decision-makers, they should not be indifferent between the three Pareto optimal Nash equilibria, even though each of them is associated with the same product of players' payoff gains. The Nash bargaining solution concept does not answer the question of how self-interested individuals would resolve such a conflict over allocations of their personal payoff gains.

Given the set of possible alternative allocations of payoff gains available in this game, a disadvantaged party could raise an objection that an offer

[4] Notice that each player's disagreement payoff is 1. The maximum personal payoff attainable in this game is 10. Relative to disagreement payoff, the maximum payoff *gain* that each player can attain in this game is 9. Relative to disagreement payoff of 1, the payoff gain of the disadvantaged player associated with the Nash equilibria $(s1, t1)$ and $(s3, t3)$ is 1. Therefore, the disadvantaged player gets 1/9 of his/her maximum payoff gain attainable in this game.

to implement either the $(s1, t1)$ or $(s2, t2)$ via joint actions is 'unfair', and propose a counter-offer. Notice that the disadvantaged party could force the other player to consider a counter-offer by threatening him/her to end the negotiations and play his/her *maximin* strategy, thus forcing that player to best respond by playing his/her *maximin* strategy as well. By doing this, a disadvantaged player would only loose a payoff gain of 1.

Given each player's ability to reject the 'unfair' offer of the other player by threatening to end the negotiations, rational players should settle on playing $(s2, t2)$: Relative to players' *maximin* payoffs, it yields each player the same share of the maximum attainable payoff gain and ensures each player a payoff which is higher than his/her disagreement payoff.

The model of hypothetical bargaining based on the Nash bargaining solution fails to account for people's strategy choices in a considerable number of experimentally relevant games. For example, coordination game 3b has three weakly Pareto optimal[5] pure strategy Nash equilibria (L, L), (R, R) and (B, B), as well as four Nash equilibria in mixed strategies. Every mixed strategy Nash equilibrium yields both players a lower personal payoff than any of the pure strategy Nash equilibria. Rational players should therefore prefer any pure strategy Nash equilibrium over any mixed strategy Nash equilibrium as the bargaining solution of this game[6]. Assuming that players' disagreement profile is the mixed strategy *maximin* Nash equilibrium $\left(\frac{5}{17}L, \frac{6}{17}R, \frac{6}{17}B; \frac{6}{17}L, \frac{5}{17}R, \frac{6}{17}B\right)$, they should identify the Nash equilibria (L, L) and (R, R) as the Nash bargaining solutions of this game. The players would also identify the Nash equilibria (L, L) and (R, R) as the Nash bargaining solutions if they were to use the mixed Nash equilibrium $\left(\frac{5}{11}L, \frac{6}{11}R; \frac{6}{11}L, \frac{5}{11}R\right)$ as the disagreement profile. Relative to mixed strategy Nash equilibria $\left(\frac{5}{11}L, \frac{6}{11}B; \frac{1}{2}L, \frac{1}{2}B\right)$ and $\left(\frac{1}{2}R, \frac{1}{2}B; \frac{5}{11}R, \frac{6}{11}B\right)$, they would identify the profile (L, L) and the profile (R, R) as the Nash bargaining solution of this game respectivelly. Notice that the Nash equilibrium (B, B) would never be identified the Nash bargaining solution, irrespective of which mixed strategy Nash equilibrium were chosen as the disagreement profile. Hypothetical bargainers should therefore not be observed choosing strategy B at all. Experimental results, however, reveal that approximately 90% of people opt for B in this game (see, for instance, Crawford et al. 2008).

It could be argued that people's behaviour observed in game 3b is determined by coordination success considerations, which become relevant in

[5] An allocation of payoffs associated with an outcome is said to be Pareto optimal in a weak sense if there is no alternative outcome associated with an allocation of payoffs which makes every interacting player strictly better off.

[6] The four mixed strategy Nash equilibria are: $(1)\left(\frac{5}{11}L, \frac{6}{11}R; \frac{6}{11}L, \frac{5}{11}R\right)$, $(2)\left(\frac{5}{17}L, \frac{6}{17}R, \frac{6}{17}B; \frac{6}{17}L, \frac{5}{17}R, \frac{6}{17}B\right)$, $(3)\left(\frac{5}{11}L, \frac{6}{11}B; \frac{1}{2}L, \frac{1}{2}B\right)$, $(4)\left(\frac{1}{2}R, \frac{1}{2}B; \frac{5}{11}R, \frac{6}{11}B\right)$.

non-cooperative games with multiple bargaining solutions: Since the game 3a has two indistinguishable Nash bargaining solutions, the probability of both players choosing the same solution is 1/2. The *ex ante* strictly Pareto-dominated Nash equilibrium (B, B) is unique. Given the coordination success rate, the *ex post* expected payoff associated with the Nash equilibrium (B, B) is higher than the one associated with the Nash bargaining solutions. The *ex ante* Pareto optimal bargaining solutions (L, L) and (R, R) are thus *ex post* Pareto dominated by the Nash equilibrium (B, B)[7].

However, the model also fails to account for decision-makers' observed choices game 3b, which has three weakly Pareto optimal pure strategy Nash equilibria: (L, L), (R, R) and (B, B). Each of the four mixed strategy Nash equilibria yields each player a lower personal payoff than any of the pure strategy Nash equilibria. Relative to players' payoffs associated with the *maximin* mixed strategy Nash equilibrium $\left(\frac{5}{17}L, \frac{6}{17}R, \frac{6}{17}B; \frac{42}{107}L, \frac{35}{107}R, \frac{30}{107}B\right)$, as well as payoffs associated with any other mixed strategy Nash equilibrium of this game[8], the Nash equilibrium (B, B) should be identified as the unique Nash bargaining solution of this game. Experiments reveal that only 1/3 of people opt for B in this game, while approximately 2/3 of people opt either for strategy L or strategy R, which indicates that some kind of benefit distribution considerations may be at play (for experimental results, see Crawford et al. 2008).

In the following sections, I will argue that a certain type of comparisons of foregone opportunities plays an important role in hypothetical bargaining, and that BES offers a plausible explanation of how such comparisons of foregone opportunities may influence players' choices in non-cooperative games.

3 The Ordinal BES

3.1 The Intuition Behind the Ordinal BES

In negotiations, every self-oriented individual wants to maximize the advancement of personal interests. S/he is therefore motivated to push the other bargaining party or parties to accept as many of his/her initial demands as possible. If bargainers have conflicting interests, an agreement can only be reached by at least one of them making a *concession* – giving up some of the initial demands. A self-oriented negotiator will seek to reach an

[7] This coordination aid has been considered by Bardsley et al. (2010) and Faillo et al. (2016).

[8] (1)$\left(\frac{5}{11}L, \frac{6}{11}R; \frac{6}{11}L, \frac{5}{11}R\right)$, (2)$\left(\frac{5}{11}L, \frac{6}{11}B; \frac{7}{12}L, \frac{5}{12}B\right)$, (3)$\left(\frac{1}{2}R, \frac{1}{2}B; \frac{7}{13}R, \frac{6}{13}B\right)$.

agreement which minimizes the number of his/her foregone initial demands. S/he can evaluate the 'goodness' of each feasible agreement on the basis of the number of initial demands that s/he would have to forego in order for that agreement to be reached: An agreement which could be reached with a smaller number of foregone initial demands should always be deemed personally more beneficial than the one which would require a larger sacrifice of initial demands (for a detailed discussion, see Zhang and Zhang 2008).

The bargainers can use another criterion for evaluating the feasible bargaining agreements. Assuming that each bargainer knows the set of each opponent's initial demands, s/he can evaluate the feasible agreements by comparing the number of initial demands that s/he would have to give up in order to reach a particular agreement with the number of initial demands that would have to be sacrificed by others: An agreement which, relative to the number of initial demands given up by other bargainers, requires the bargainer to give up less of the initial demands should be deemed more acceptable by him/her than any agreement which, relative to the numbers of foregone initial demands of others, requires him/her to give up more of the initial demands.

The ordinal BES is based on the principle that hypothetical bargainers evaluate the feasible agreements by comparing the distributions of foregone initial demands among the interacting bargainers associated with each agreement: An agreement with a more equitable distribution of foregone initial demands among the interacting bargainers is deemed more acceptable than the one with a less equitable distribution of foregone initial demands.

The principles underlying the BES can be applied to analysis of non-cooperative games where players only have ordinal information about each other's preferences over outcomes. For example, consider a simple ordinal coordination problem depicted in Figure 4. The left and the right number in each cell represents row and column player's *ordinal* preferences over outcomes respectively.

	t1	*t2*	*t3*	*t4*
s1	$100, 3$	$0, 0$	$0, 0$	$0, 0$
s2	$0, 0$	$60, 5$	$0, 0$	$0, 0$
s3	$0, 0$	$0, 0$	$40, 9$	$0, 0$
s4	$0, 0$	$0, 0$	$0, 0$	$20, 1$

Figure 3: Ordinal coordination game

Since mixed outcomes can only be defined with cardinal payoffs, the players who reason as hypothetical bargainers should identify the four pure strategy Nash equilibria as the feasible agreements of this ordinal game: $(s1, t1)$,$(s2, t2)$, $(s3, t3)$, $(s4, t4)$. Assuming that players' ordinal preferences are common knowledge, each player can determine the number of *preferred alternative agreements* that each player would forego if each of the feasible agreements were chosen for implementation. For example, if outcome $(s1, t1)$ were chosen, the row player's personal interests would be maximally advanced, which means that s/he would not forego any opportunities to advance his or her personal interests. The column player, on the other hand, prefers outcomes $(s2, t2)$ and $(s3, t3)$ over the outcome $(s1, t1)$. If the outcome $(s1, t1)$ were chosen, s/he would forego two preferred alternative agreements.

If players' preferences are common knowldge, each decision maker knows every other decision-maker's preferential rankings of feasible agrements based on the numbers of foregone preferred alternatives. Players' preferential rankings are shown in Figure 5:

$$\text{Row:} \begin{bmatrix} \text{Agreement} & \text{Foregone alternatives} \\ (s1, t1) & 0 \\ (s2, t2) & 1 \\ (s3, t3) & 2 \\ (s4, t4) & 3 \end{bmatrix} \qquad \text{Column:} \begin{bmatrix} \text{Agreement} & \text{Foregone alternatives} \\ (s3, t3) & 0 \\ (s2, t2) & 1 \\ (s1, t1) & 2 \\ (s4, t4) & 3 \end{bmatrix}$$

Figure 4: Players' foregone preferred alternatives

In explicit bargaining, rational bargainers should easily agree to restrict their negotiations to a subset of feasible agreements including outcomes $(s1, t1)$, $(s2, t2)$ and $(s3, t3)$. This restriction of the bargaining set is clearly mutually beneficial: For each bargainer, any agreement in the aforementioned subset guarantees a strictly lower number of foregone preferred alternatives than the agreement $(s4, t4)$. Among the agreements $(s1, t1)$, $(s2, t2)$ and $(s3, t3)$, however, no agreement is associated with strictly lower numbers of foregone preferred alternatives for both players. Hypothetical bargainers could evaluate the feasible agreements in this set by comparing, how the foregone preferred alternatives would be distributed among them if each of the agreements were chosen to be implemented. Notice that outcome $(s2, t2)$ minimizes the difference between numbers of players' foregone preferred alternatives. In other words, *among the three weakly Pareto optimal feasible agreements* $(s1, t1)$*,* $(s2, t2)$ *and* $(s3, t3)$, agreement $(s2, t2)$ ensures a maximally equitable distribution of foregone preferred alternatives. The Nash equilibrium $(s2, t2)$ is the BES of this game[9].

3.2 Formalization

Let $\Gamma^o = \left(\{1,2\}, \{S_i, \succeq_i\}_{i\in\{1,2\}}\right)$ be an ordinal two player game where $\mathbf{S} = S_1 \times S_2$ is the set of pure strategy outcomes and $\succeq_i$ is the complete and transitive preference ordering of strategy profiles in $\mathbf{S}$ of player $i \in \{1,2\}$. Since mixed outcomes can only be defined with cardinal payoffs, only pure strategy outcomes are considered.

Let $\mathcal{A} \in \mathcal{P}(\mathbf{S})$ be a set of feasible agreements in $\mathbf{S}$. It will be assumed that $\mathcal{A} \neq \emptyset$[10]. The set of feasible agreements is the set of the pure strategy Nash equilibria of Γ^o:

$$\mathcal{A} = \left\{\mathbf{s} \in \mathbf{S} : \mathbf{s} \in \mathbf{S}^{NE}\right\}, \text{where } \mathbf{S}^{NE} \in \mathcal{P}(\mathbf{S}). \tag{1}$$

For each feasible agreement $\mathbf{s} \in \mathcal{A}$, we can define the *cardinality of the preferred set of alternatives* for every player $i \in \{1,2\}$:

$$\mathcal{C}_i(\mathbf{s}, \mathcal{A}) \equiv \{|T|, \text{ where } \mathbf{s}' \in T \text{ if and only if } \mathbf{s}' \in \mathcal{A} \text{ and } \mathbf{s}' \succ_i \mathbf{s}\}. \tag{2}$$

For any two agreements $\mathbf{s} \in \mathcal{A}$ and $\mathbf{s}' \in \mathcal{A}$, it is assumed that $\mathbf{s} \succ_i \mathbf{s}'$ only if $\mathcal{C}_i(\mathbf{s}) < \mathcal{C}_i(\mathbf{s}')$. It is possible to define the *set of maximally mutually advantageous feasible agreements:*

$$\mathbf{s} \in \mathcal{A}^m \Rightarrow \mathbf{s}' \notin \mathcal{A} : \mathcal{C}_i(\mathbf{s}', \mathcal{A}) < \mathcal{C}_i(\mathbf{s}, \mathcal{A}) \, \forall i \in \{1,2\}. \tag{3}$$

For any agreement $\mathbf{s} \in \mathcal{A}^m$, a measure of the difference between hypothetical bargainers' cardinalities of the preferred sets of alternatives can be defined in the following way:

$$|\mathcal{C}_i(\mathbf{s}, \mathcal{A}) - \mathcal{C}_{j\neq i}(\mathbf{s}, \mathcal{A})|. \tag{4}$$

BES function $\phi^o(\cdot)$ satisfies, for every $\mathcal{A}$,

$$\phi^o(\mathcal{A}) \in arg\ min_{\mathbf{s}\in\mathcal{A}^m} \{|C_i(\mathbf{s}, \mathcal{A}) - C_{j\neq i}(\mathbf{s}, \mathcal{A})|\}. \tag{5}$$

[9] BES is based on the principles which are similar to the ones underlying Conley and Wilkie's (2012) ordinal egalitarian bargaining solution (OEBS) for finite sets of Pareto optimal points. OEBS is a Pareto optimal point associated with strictly equal numbers of foregone preferred alternatives. BES is based on a weaker equity requirement: It is any weakly Pareto optimal outcome which, given a particular set of weakly Pareto optimal outcomes, minimizes the difference between the cardinalities of players' preferred sets of alternatives. In some games, a benefit-equilibrating solution may not be strictly ordinally egalitarian. However, it is a maximally ordinally equitable outcome available in a particular set of feasible agreements. For an in-depth discussion of OEBS, axiomatic characterization and proofs, see Conley and Wilkie (2012).

[10] Some games may not have any Nash equilibria in pure strategies. In those cases $\mathcal{A} = \emptyset$.

3.3 The Properties of the Ordinal BES

Existence: Ordinal BES (not necessarily unique) exists in every finite two player ordinal game with at least one Nash equilibrium in pure strategies.

For any finite ordinal game, $\mathbf{S}^{NE} \in \mathcal{P}(\mathbf{S})$ is always finite. Since $\mathcal{A} = \mathbf{S}^{NE}$, the set $\mathcal{A}$ is finite as well. It is therefore always possible to define, for every feasible agreement $\mathbf{s} \in \mathcal{A}$, the cardinality of the preferred set of alternatives for every player $i \in \{1, 2\}$. In every finite set of feasible agreements $\mathcal{A} = \mathbf{S}^{NE}$, there exists $\mathbf{s} \in \mathcal{A}$, such that

$$\mathbf{s}' \notin \mathcal{A} : \mathcal{C}_i(\mathbf{s}', \mathcal{A}) < \mathcal{C}_i(\mathbf{s}, \mathcal{A}) \, \forall i \in \{1, 2\} . \tag{6}$$

It follows that $\mathbf{s} \in \mathcal{A}^m$, which means that $\mathcal{A}^m \neq \emptyset$. Therefore, a BES exists in every Γ^o, such that $\mathbf{S}^{NE} \neq \emptyset$.

Feasible weak Pareto optimality: Ordinal BES is a weakly Pareto optimal *feasible* agreement.

Let $\mathcal{A}^{wpo} \subseteq \mathcal{A}$ denote the set of weakly Pareto feasible agreements of Γ^o. A feasible agreement $\mathbf{s} \in \mathcal{A}$ belongs to a set of weakly Pareto optimal feasible agreements *only if* there is no alternative outcome $\mathbf{s}' \in \mathcal{A}$ such that $\mathbf{s}' \succ_i \mathbf{s}$ for all $i \in \{1, 2\}$. In terms of cardinalities of preferred sets, this condition can be defined as follows:

$$\mathbf{s} \in \mathcal{A}^{wpo} \Rightarrow \mathbf{s}' \notin \mathcal{A} : \mathcal{C}_i(\mathbf{s}', \mathcal{A}) < \mathcal{C}_i(\mathbf{s}, \mathcal{A}) \, \forall i \in \{1, 2\} . \tag{7}$$

Definition (7) is equivalent to definition (3), which implies that $\mathcal{A}^{wpo} = \mathcal{A}^m$. From characterization (5), it follows that $\phi^o(\mathcal{A}) \subseteq \mathcal{A}^{wpo}$.

Invariance under additions of Pareto irrelevant alternatives: For any two ordinal games Γ^o and $\Gamma^{o'}$, such that $\mathcal{A} = \mathcal{A}$', it is always the case that $\phi^o(\mathcal{A}) = \phi^o(\mathcal{A}')$.

Since $\mathcal{A}^m = \mathcal{A}^{wpo}$, from definition (3) it follows that, for every $i \in \{1, 2\}$, any $\mathbf{s} \notin \mathcal{A}^m$ is such that $\mathcal{C}_i(\mathbf{s}, \mathcal{A}) > \mathcal{C}_i(\mathbf{s}', \mathcal{A}) \ \forall \mathbf{s}' \in \mathcal{A}^m$. From definition 2, it follows that, for any $\mathcal{A}^m \subseteq \mathcal{A}$, it must be the case that every $\mathbf{s}' \in \mathcal{A}^m$ is such that $\mathcal{C}_i(\mathbf{s}', \mathcal{A}) = \mathcal{C}_i(\mathbf{s}', \mathcal{A}^m) \ \forall i \in \{1, 2\}$. For any two ordinal games Γ^o and Γ^o', such that $\mathcal{A}^m = \mathcal{A}'^m$, for every $\mathbf{s} \in \mathcal{A}^m$ it must be the case that $\mathbf{s} \in \mathcal{A}'^m$, and so $\mathcal{C}_i(\mathbf{s}, \mathcal{A}^m) = \mathcal{C}_i(\mathbf{s}, \mathcal{A}'^m) \ \forall i \in \{1, 2\}$. It follows that, for any $\mathbf{s} \in \mathcal{A}^m$, it must be the case that $\mathcal{C}_i(\mathbf{s}, \mathcal{A}) = \mathcal{C}_i(\mathbf{s}, \mathcal{A}') \ \forall i \in \{1, 2\}$. Therefore, $\phi^o(\mathcal{A}) = \phi^o(\mathcal{A}')$.

Ordinal invariance: Ordinal BES is invariant under order-preserving transformations of players' ordinal preference representations.

Notice that $\mathcal{A} = \mathbf{S}^{NE}$ of every Γ^o. It follows that $\phi^o(\mathcal{A}) \subseteq \mathbf{S}^{NE}$ of Γ^o. For any ordinal games Γ^o and $\Gamma^{o'}$, such that $\mathbf{S}^{NE} = \mathbf{S}'^{NE}$, it must be the case that $\mathcal{A} = \mathcal{A}'$, which implies that $\mathcal{A}^m = \mathcal{A}'^m$. It follows that $\phi^o(\mathcal{A}) = \phi^o(\mathcal{A}')$. Every pure strategy Nash equilibrium is invariant under order-preserving transformations of players' ordinal preference representations, and so is BES.

Independence of irrelevant strategies: Ordinal BES is invariant under additions of strictly dominated strategies.

In every finite ordinal game Γ^o, we can use ordinal preferences to define a best-response correspondence $\beta_i : \mathbf{S} \to \mathbf{S}_i$ of player $i \in \{1,2\}$, which maps each pure strategy profile $\mathbf{s} \in \mathbf{S}$ to the finite set of pure best responses of player i to profile $\mathbf{s} \in \mathbf{S}$:

$$\beta_i(\mathbf{s}) = \{s_i \in S_i : (s_i, s_{j\neq i}) \succeq_i (\tilde{s}_i, s_j)\, \forall \tilde{s}_i \in S_i\}. \tag{8}$$

Notice that $\beta_i(\mathbf{s}) \subseteq S_i$. From this we can define the set $S_i^{br}(\mathbf{S}) \subseteq S_i$ of pure best responses of player $i \in \{1,2\}$ to the finite set of pure strategy profiles $\mathbf{S} = \{\mathbf{s}^1, ..., \mathbf{s}^n\}$:

$$S_i^{br}(\mathbf{S}) = \left\{s_i \in S_i : s_i \in \beta_i\left(\mathbf{s}^i\right) \text{ for some } \mathbf{s}^i \in \mathbf{S}\right\}. \tag{9}$$

Every Nash equilibrium is a profile of best responses, and so $\mathbf{S}^{NE} \subseteq \left(S_1^{br}(\mathbf{S}) \times S_2^{br}(\mathbf{S})\right)$. Any strictly dominated strategy $s_i \in S_i$ is never a best response, which implies that $s_i \notin S_i^{br}(\mathbf{S})$. Since $\phi^o(\mathcal{A}) \subseteq \mathcal{S}^{NE}$, BES is invariant under addition of any strategy s_i, such that $s_i \notin S_i^{br}(\mathbf{S})$ for every $i \in \{1,2\}$.

Individual rationality: Ordinal BES is an outcome which, for $i \in \{1,2\}$, is at least as good as ordinal *maximin* outcome.

In terms of cardinalities of preferred sets, an ordinal *maximin* threshold of $i \in \{1,2\}$ can be defined as follows:

$$\mathcal{C}_i^{mnm}(\mathbf{S}) = min_{s_i \in S_i}\ \left\{max_{s_{j\neq i} \in S_j}\mathcal{C}_i(\mathbf{s}, \mathbf{S})\right\}. \tag{10}$$

Ordinal BES satisfies the individual rationality requirement if and only if, for every $i \in \{1,2\}$, the set of BES is always such that

$$\mathcal{C}_i^{mnm}(\mathbf{S}) \geq \mathcal{C}_i(\mathbf{s}, \mathcal{A})\, \forall \mathbf{s} \in \phi^{\mathbf{o}}(\mathcal{A}) \tag{11}$$

Since $\mathcal{A} = \mathbf{S}^{NE}$, it follows that $\phi^o(\mathcal{A}) \subseteq \mathbf{S}^{NE}$. If a strategy profile $\mathbf{s} \in \mathbf{S}$ is a Nash equilibrium, the preferences of player $i \in \{1,2\}$ are as follows:

$$(s_i, s_{j\neq i}) \in \mathbf{S}^{NE} \Rightarrow (s_i, s_j) \succeq_i (\tilde{s}_i, s_j)\, \forall \tilde{s}_i \in S_i. \tag{12}$$

In terms of cardinalities of preferred sets, property (15) can be defined as follows:

$$(s_i, s_{j\neq i}) \in \mathbf{S}^{NE} \Rightarrow \mathcal{C}_i\left((s_i, s_j), \mathbf{S}\right) \leq \mathcal{C}_i\left((\tilde{s}_i, s_j), \mathbf{S}\right) \forall \tilde{s}_i \in S_i. \tag{13}$$

Notice that the *maximin* strategy $s_i^{mxm} \in S_i$ of each player $i \in \{1,2\}$ is such that

$$\mathcal{C}_i\left((s_i^{mxm}, s_{j\neq i}), \mathbf{S}\right) \leq \mathcal{C}_i^{mnm}(\mathbf{S})\, \forall s_j \in S_j. \tag{14}$$

Every Nash equilibrium of Γ^o must have the following property:

$$(s_i, s_{j\neq i}) \in \mathbf{S}^{NE} \Rightarrow (s_i, s_j) \succeq_i (s_i^{mxm}, s_j)\, \forall i \in \{1,2\}. \tag{15}$$

In terms of cardinalities of preferred sets, property (15) can be characterized as follows:

$$(s_i, s_{j\neq i}) \in \mathbf{S}^{NE} \Rightarrow \mathcal{C}_i(s_i, s_j) \leq \mathcal{C}_i(s_i^{mxm}, s_j)\, \forall i \in \{1,2\}. \tag{16}$$

Since $\phi^o(\mathcal{A}) \subseteq \mathbf{S}^{NE}$, the individual rationality requirement is always satisfied.

4 The Cardinal BES

4.1 The Intuition Behind the Cardinal BES

To grasp the intuition behind the cardinal BES, consider the two player three strategy coordination game depicted in Figure 5. It has three pure strategy Nash equilibria in this game: $(s1, t1), (s2, t2)$ and $(s3, t3)$. There are also

	t1	*t2*	*t3*
s1	$100, 98$	$0, 0$	$0, 0$
s2	$0, 0$	$2, 99$	$0, 0$
s3	$0, 0$	$0, 0$	$1, 100$

Figure 5: Coordination game with three weakly Pareto optimal outcomes

four Nash equilibria in mixed strategies[11]. To simplify the presentation of the key principles, in this particular example only the pure strategy Nash equilibria will be considered as feasible agreements.

If the players were to treat this game as an ordinal bargaining problem, they would identify the Nash equilibrium $(s2, t2)$ as the ordinal BES. Given the available information about players' cardinal payoffs, intuitivelly this solution does not seem reasonable: The column player's loss of the maximum attainable utility seems to be insignificant compared to the loss of the row player. In real-world negotiations, the row player could be expected not to accept anything else but the agreement $(s1, t1)$. If the column player refused, the row player would suffer relatively insignificant payoff losses from playing his/her mixed *maximin* strategy[12] rather than playing a part in realizing the agreement $(s2, t2)$.

Although this intuition is compelling, the expected utility theory does imply the interpersonal comparability of players' cardinal utilities. In other words, the theory offers no answer to the question of how one player's utility units should be 'converted' into utility units of another player (for extensive discussion, see Luce and Raiffa 1957). However, the players could identify the Nash equilibrium $(s1, t1)$ as the BES of this game without being able to compare utility units in the aforementioned sense. This would happen if they were to evaluate the feasible outcomes by comparing their normalized losses of the maximum attainable individual advantage associated with the implementation of each feasible agreement.

Such comparisons can be performed on the basis of Raiffa (1953) normalization procedure, which can be used to measure the level of satisfaction of decision-maker's preferences. According to this procedure, the level of individual advantage gained from a particular outcome can be defined as the extent by which that outcome advances the player's personal payoff from his/her reference point relative to the largest advancement possible, where the latter is associated with the attainment of the outcome that s/he prefers the most.

For the purposes of BES, each hypothetical bargainer's most preferred outcome will be defined as his or her most preferred feasible agreement[13]:

$$u_i^{max} = max_{\mathbf{s} \in \mathbf{S}^{NE}} u_i \left(\mathbf{s} \right) \tag{17}$$

[11] The four mixed strategy Nash equilibria are: $(1)\left(\frac{99}{197}s1, \frac{98}{197}s2; \frac{1}{51}t1, \frac{50}{51}t2\right)$, $(2)\left(\frac{4950}{14701}s1, \frac{4900}{14701}s2, \frac{4851}{14701}s3; \frac{1}{151}t1, \frac{50}{151}t2, \frac{100}{151}t3\right)$,$(3)\left(\frac{50}{99}s1, \frac{49}{99}s3; \frac{1}{101}t1, \frac{100}{101}t3\right)$, $(4)\left(\frac{100}{199}s2, \frac{99}{199}s3; \frac{1}{3}t2, \frac{2}{3}t3\right)$.

[12] In this case, the *maximin* strategy of the row player is mixed strategy $\left(\frac{4950}{14701}s1, \frac{4900}{14701}s2, \frac{4851}{14701}s3\right)$.

[13] This definition of the best outcome is in line with the definition used in some of the standard bargaining solutions, such as the Kalai-Smorodinsky (1975) bargaining solution.

With respect to hypothetical bargainers' reference points, two definitions seem reasonable. One possibility is to set each hypothetical bargainer's reference point to be the worst personal payoff associated with a *rationalizable outcome* of a game:

$$u_i^{ref} = min_{\mathbf{s}\in\mathbf{S}^{br}} u_i\left(\mathbf{s}\right), \text{ where } \mathbf{S}^{br} = \left(S_1^{br} \times S_2^{br}\right) \subseteq \mathbf{S} \tag{18}$$

The intuition behind this definition is that hypothetical bargainers who were to fail to reach an agreement in open bargaining would have no joint plan on how to play the game. In such a situation of strategic uncertainty, the players could attempt to coordinate their actions by guessing each other's strategy choice. If rationality is common knowledge, the players should only consider rationalizable strategies.

Another possibility is to set each hypothetical bargainer's reference point to be his or her *maximin payoff level in rationalizable strategies*:

$$u_i^{ref} = max_{s_i\in S_i}\left\{min_{s_{-i}\in S_{-i}^{br}} u_i\left(\mathbf{s}\right)\right\} \tag{19}$$

The intuition behind this definition is that a hypothetical bargainers who were to fail to reach an agreement would respond by choosing a strategy which guarantees the highest personal payoff, irrespective of which one of the rationalizable strategies the opponent is going to choose.

The question of which reference point is the best approximation to how real-world hypothetical bargainers reason about their options in games cannot be answered on the basis of formal theoretical analysis alone. Further empirical research is required to answer this question. It is possible that decision-maker's choice of a reference point may depend on how high his/her personal stakes are in a particular game: A decision-maker may adopt a more cautious approach in a game where the personal stakes are high, while be more willing to risk in a game where the personal stakes are relatively insignificant. For the purposes of the following theoretical discussion, definition (17) will be used. This reference point seems reasonable for a model describing hypothetical bargainers' behaviour in experimental games with relatively low personal stakes.

Consider, again, the game depicted in Figure 5. For the row player, the most preferred feasible agreement is the Nash equilibrium $(s1, t1)$, while the least preferred rationalizable outcome is any outcome of this game associated with a payoff of 0. The levels of individual advantage associated with each of the feasible agreements can be established with the following transformation of row player's original payoffs:

$$u_r^{\iota}\left(\mathbf{s}\right) = \frac{u_r\left(\mathbf{s}\right) - u_r^{ref}}{u_r^{max} - u_r^{ref}} \text{ ,where } \mathbf{s} \text{ is a profile of } \textit{rationalizable} \text{ strategies.} \tag{20}$$

For example, the level of individual advantage associated with outcome $(s2, t2)$ is 0.02. Since the maximum attainable *level of individual advantage* is 1, the row player would loose 0.98 of the maximum attainable individual advantage if outcome $(s2, t2)$ were chosen.

Players' levels of individual advantage and individual advantage losses associated with each feasible agreement are shown in Figure 6.

$$\begin{bmatrix} Agr. & u_r^\iota & u_c^\iota & 1-u_r^\iota & 1-u_c^\iota \\ (s1,t1) & 1 & 0.98 & 0 & 0.02 \\ (s2,t2) & 0.02 & 0.99 & 0.98 & 0.01 \\ (s3,t3) & 0.01 & 1 & 0.99 & 0 \end{bmatrix}$$

Figure 6: Players' levels of individual advantage and losses of maximum individual advantage associated with each agreement

Notice that outcome $(s1, t1)$ *minimizes the difference between players' individual advantage losses*: In percentage terms, the row player would loose 0% of the individual advantage, while the column player would loose just 2%. The Nash equilibrium $(s1, t1)$ is the BES of this coordination game.

When hypothetical bargainers evaluate the feasible agreements, they equate units of measures of their individual advantage — the advancement of their personal interests relative to what they personally deem to be the best and the worst outcome of their interaction. In order to use this measure, hypothetical bargainers need to know each other's cardinal payoffs and reference points, but they need not be able to make interpersonal comparisons of their attained well-being. In other words, BES is a *formal arbitration scheme*: It operates purely on the basis of information about players' reference points and the cardinal payoffs represented by the numbers in the payoff matrix, and so can be used in cases where hypothetical bargainers have no clue as to what kind of personal motivations or welfare levels those utility numbers actually represent.

4.2 Formalization

Let $\Gamma = \left(\{1,2\}, \{S_i, u_i\}_{i \in \{1,2\}} \right)$ be a normal form two player game, where S_i is the set of pure strategies of $i \in \{1,2\}$ and $u_i : \mathcal{L}(\varSigma) \rightarrow \mathbb{R}$ is the cardinal utility function of player $i \in \{1,2\}$ that represents his/her preferences over the set of lotteries over the *set of possible agreements* – the set of mixed strategy profiles $\varSigma = (\varSigma_1 \times \varSigma_2)$ of Γ. Each mixed strategy $\sigma_i \in \varSigma_i$ should be interpreted as a randomized action of player $i \in \{1,2\}$, where $\sigma_i(s_i)$ denotes the probability of player $i \in \{1,2\}$ choosing pure strategy $s_i \in S_i$. Each mixed strategy profile $\sigma \in \Sigma$ should be interpreted as a profile of players' randomized actions.

Let $\Sigma_i^{br} \subseteq \Sigma_i$ denote the set of *rationalizable* strategies of $i \in \{1,2\}$ and $\Sigma^{br} = \left(\Sigma_1^{br} \times \Sigma_2^{br}\right)$ the set of rationalizable strategy profiles of Γ. Let $u_i^{ref} := min_{\sigma \in \Sigma^{br}} u_i(\sigma)$ denote the *reference point* of player $i \in \{1,2\}$.

Let $\Sigma^f \subseteq \Sigma^{br}$ denote the set of feasible agreements of Γ which is defined as follows:

$$\Sigma^f = \left\{\sigma \in \Sigma : \sigma \in \Sigma^{NE}\right\}. \tag{21}$$

Notice that definition (21) implies that $\Sigma^f = \Sigma^{NE}$. Let $u_i^{max} := max_{\sigma \in \Sigma^f} u_i(\sigma)$ denote the utility associated with i's most preferred feasible agreement. Subject to the constraint $u_i^{max} \neq u_i^{ref}$, the individual advantage of player $i \in \{1,2\}$ associated with any feasible agreement $\sigma \in \Sigma^f$ will be defined as follows:

$$u_i^{\iota}(\sigma) = \frac{u_i(\sigma) - u_i^{ref}}{u_i^{max} - u_i^{ref}}. \tag{22}$$

Notice that if i's utility function is such that $u_i^{max} = 1$ and $u_i^{ref} = 0$, then $u_i(\sigma) = u_i^{\iota}(\sigma)\ \forall \sigma \in \Sigma^{br}$.

Let $\Sigma^{fm} \subseteq \Sigma^f$ denote the set of *maximally mutually advantageous agreements*, which will be defined as follows:

$$\sigma \in \Sigma^{fm} \Rightarrow \sigma' \notin \Sigma^f : u_i^{\iota}(\sigma') > u_i^{\iota}(\sigma)\, \forall i \in \{1,2\}. \tag{23}$$

A measure of loss of maximum individual advantage $\varphi_i(\cdot\,,\cdot)$ of player $i \in \{1,2\}$ will be defined as follows:

$$\varphi_i\left(\sigma, \Sigma^{br}\right) = \left(\frac{u_i^{max} - u_i^{ref}}{u_i^{max} - u_i^{ref}}\right) - \left(\frac{u_i(\sigma) - u_i^{ref}}{u_i^{max} - u_i^{ref}}\right) = 1 - u_i^{\iota}(\sigma)\,, \text{ where } \sigma \in \Sigma^f. \tag{24}$$

The difference between players' losses of maximum attainable individual advantage associated with any $\sigma \in \Sigma^f$ can be determined as follows:

$$\left|\varphi_i\left(\sigma, \Sigma^{br}\right) - \varphi_{j \neq i}\left(\sigma, \Sigma^{br}\right)\right|. \tag{25}$$

The cardinal BES function $\phi^c(\cdot,\cdot)$ satisfies, for every $\Sigma^f \subseteq \Sigma^{br}$,

$$\phi^c\left(\Sigma^f, \Sigma^{br}\right) = arg\ min_{\sigma \in \Sigma^{fm}} \left\{\left|\varphi_i\left(\sigma, \Sigma^{br}\right) - \varphi_{j \neq i}\left(\sigma, \Sigma^{br}\right)\right|\right\}. \tag{26}$$

4.3 The Properties of the Cardinal BES

Existence: Cardinal BES exists in every finite two player game.

Nash (1951) proved that an equilibrium in mixed strategies exists in every finite game with a finite set of players. In every finite game, there exists $\sigma \in \varSigma^{NE}$, such that

$$\sigma' \notin \varSigma^{NE} : u_i\left(\sigma'\right) > u_i\left(\sigma\right) \forall i \in \{1,2\}. \tag{27}$$

From definition (22) and property (27), it follows that every finite Γ has at least one $\sigma \in \varSigma^{NE}$, such that

$$\sigma' \notin \varSigma^{NE} : u_i^{\iota}\left(\sigma'\right) > u_i^{\iota}\left(\sigma\right) \forall i \in \{1,2\}. \tag{28}$$

From definitions (23) and property (28), it follows that $\varSigma^{fm} \neq \emptyset$ in every finite Γ, and so $\phi^c\left(\varSigma^f, \varSigma^{br}\right) \neq \emptyset$ in every finite Γ.

Feasible weak Pareto optimality: Cardinal BES is a weakly Pareto optimal *feasible* agreement.

Let $\varSigma^{fwpo} \subseteq \varSigma^f$ denote the set of feasible weakly Pareto optimal agreements of Γ. A set of Pareto optimal feasible agreements can be characterized as follows:

$$\sigma \in \varSigma^{fwpo} \Rightarrow \sigma' \notin \varSigma^f : u_i\left(\sigma'\right) > u_i\left(\sigma\right) \forall i \in \{1,2\}. \tag{29}$$

From definitions (22) and (29), it follows that a set of weakly Pareto optimal feasible agreements can be characterized as follows:

$$\sigma \in \varSigma^{fwpo} \Rightarrow \sigma' \in \varSigma^f : u_i^{\iota}\left(\sigma'\right) > u_i^{\iota}\left(\sigma\right) \forall i \in \{1,2\}. \tag{30}$$

From definitions (23) and (30), it folows that $\varSigma^{fwpo} = \varSigma^{fm}$. Since it is the case that $\phi^c\left(\varSigma^f, \varSigma^{br}\right) \subseteq \varSigma^{fm}$, it follows that $\phi^c\left(\varSigma^f, \varSigma^{br}\right) \subseteq \varSigma^{fwpo}$.

Invariance under additions of irrelevant alternatives: Cardinal BES is invariant under additions of non-rationalizable outcomes.

Notice that u_i^{ref} and u_i^{max} are associated with some $\sigma \in \varSigma^{br}$ for every $i \in \{1,2\}$, and that $\varSigma^f \subseteq \varSigma^{br}$, where $\varSigma^{br} = \left(\varSigma_1^{br} \times \varSigma_2^{br}\right)$. For any two games Γ and Γ', such that $\varSigma^{br} = \varSigma^{br'}$, it must be the case that $\varSigma^{NE} = \varSigma^{NE'}$. From definition (21), it follows that $\varSigma^f = \varSigma^{f'}$, and so $\varSigma^{fm} = \varSigma^{fm'}$. From characterization (26), it follows that $\phi^c\left(\varSigma^f, \varSigma^{br}\right) = \phi^c\left(\varSigma^{f'}, \varSigma^{br'}\right)$. Every non-rationalizable outcome $\sigma \in \varSigma$ is such that $\sigma \notin \varSigma^{br}$. Therefore, $\varSigma^{br} \subseteq \varSigma$ is invariant to additions of σ to $\varSigma$, such that $\sigma \notin \varSigma^{br}$.

Individual rationality: Cardinal BES yields each player a payoff which is at least as high as the *maximin* payoff.

The cardinal *maximin* threshold of player $i \in \{1,2\}$ can be defined as follows:

$$u_i^{mxm} = max_{\sigma_i \in \Sigma_i} \left\{ min_{\sigma_{j \neq i} \in \Sigma_j} u_i(\sigma) \right\}. \tag{31}$$

The cardinal BES satisfies the individual rationality requirement if and only if, for every $i \in \{1,2\}$,

$$u_i(\sigma) \geq u_i^{mxm} \; \forall \sigma \in \phi^c\left(\Sigma^f, \Sigma^{br}\right) \tag{32}$$

The *maximin* strategy $\sigma_i^{mxm} \in \Sigma_i$ of player $i \in \{1,2\}$ is such that

$$u_i\left(\sigma_i^{mxm}, \sigma_{j \neq i}\right) \geq u_i^{mxm} \; \forall \sigma_j \in \Sigma_j. \tag{33}$$

If a strategy profile $\sigma \in \Sigma$ is a Nash equilibrium, the preferences of $i \in \{1,2\}$ are as follows:

$$\left(\sigma_i, \sigma_{j \neq i}\right) \in \Sigma^{NE} \Rightarrow u_i\left(\sigma_i, \sigma_j\right) \geq u_i\left(\tilde{\sigma}_i, \sigma_j\right) \forall \tilde{\sigma}_i \in \Sigma_i \tag{34}$$

Every Nash equilibrium must satisfy the following condition:

$$\left(\sigma_i, \sigma_{j \neq i}\right) \in \Sigma^{NE} \Rightarrow u_i\left(\sigma_i, \sigma_j\right) \geq u_i^{mxm} \; \forall i \in \{1,2\}. \tag{35}$$

Since $\phi^c\left(\Sigma^f, \Sigma^{br}\right) \subseteq \Sigma^{NE}$, the individual rationality requirement is satisfied.

Independence of irrelevant strategies: Cardinal BES is invariant under additions of strictly dominated strategies.

In every cardinal game Γ, we can use the cardinal preferences to define a best-response correspondance $\mathcal{B}_i : \Sigma \rightarrow \Sigma_i$ of player $i \in \{1,2\}$, which maps each mixed strategy profile $\sigma \in \Sigma$ to the finite set of mixed best responses of i to profile $\sigma \in \Sigma$:

$$\mathcal{B}_i(\sigma) = \left\{\sigma_i \in \Sigma_i : u_i\left(\sigma_i, \sigma_{j \neq i}\right) \geq u_i\left(\tilde{\sigma}_i, \sigma_j\right) \forall \tilde{\sigma}_i \in \Sigma_i\right\}. \tag{36}$$

Notice that $\mathcal{B}_i(\sigma) \subseteq \Sigma_i$. From this we can define the set $\Sigma_i^{br} \subseteq \Sigma_i$ of mixed best responses of $i \in \{1,2\}$ to the set of mixed strategy profiles $\Sigma = (\Sigma_1 \times \Sigma_2)$ of Γ:

$$\Sigma_i^{br} = \left\{\sigma_i \in \Sigma_i : \sigma_i \in \mathcal{B}_i(\sigma) \text{ for some } \sigma \in \Sigma\right\}. \tag{37}$$

Every Nash equilibrium is a profile of best responses, and so $\Sigma^{NE} \subseteq \left(\Sigma_1^{br} \times \Sigma_2^{br}\right)$. If strategy $\sigma_i \in \Sigma_i$ is strictly dominated, it must be the case that $\sigma_i \notin \Sigma_i^{br}$. Since $\phi^c\left(\Sigma^f, \Sigma^{br}\right) \subseteq \Sigma^{NE}$, the cardinal BES is invariant under additions of any strategy σ_i to Σ_i, such that $\sigma_i \notin \Sigma_i$.

Invariance under positive scalar transformations of payoffs: For any Γ', which involves a transformation of Γ only of the form $u'_i = au_i$, where $a > 0$, it is always the case that $\phi^c\left(\mathit{\Sigma}^f, \mathit{\Sigma}^{br}\right) = \phi^c\left(\mathit{\Sigma}^{f'}, \mathit{\Sigma}^{br'}\right)$.

Notice that $\mathit{\Sigma}^f = \mathit{\Sigma}^{NE}$ for every Γ, which implies that $\phi^c\left(\mathit{\Sigma}^f, \mathit{\Sigma}^{br}\right) \subseteq \mathit{\Sigma}^{NE}$ of every Γ. Each mixed strategy profile $\sigma \in \mathit{\Sigma}$ is a tuple (σ_1, σ_2), where $\sigma_i \in \mathit{\Sigma}_i$ is a mixed strategy of $i \in \{1, 2\}$, which in a finite game assigns a probability distribution over the finite set S_i of pure strategies of $i \in \{1, 2\}$. The support of every $\sigma_i \in \mathit{\Sigma}$ can be defined as follows:

$$\mathit{Supp}\left(\sigma_i\right) = \left\{s_i \in S_i : \sigma_i\left(s_i\right) > 0\right\}, \text{ where } \sigma_i\left(s_i\right) \text{ is the probability of } s_i \in S_i\ . \tag{38}$$

The support of each mixed strategy profile $\sigma \in \mathit{\Sigma}$ can be defined as follows:

$$\mathit{Supp}\left(\sigma\right) = \left(\mathit{Supp}\left(\sigma_1\right) \times \mathit{Supp}\left(\sigma_2\right)\right) \subseteq \mathbf{S}, \text{ where } \mathbf{S} = \left(S_1 \times S_2\right). \tag{39}$$

The probability of players playing any pure strategy profile $\mathbf{s} \in \mathit{Supp}\left(\sigma\right)$ is

$$\sigma\left(\mathbf{s}\right) = \left(\sigma_1\left(s_1\right) \times \sigma_2\left(s_2\right)\right) = \prod_{i \in \{1,2\}} \sigma_i\left(s_i\right). \tag{40}$$

The expected utility of $i \in \{1, 2\}$ associated with $\sigma \in \mathit{\Sigma}$ is

$$u_i\left(\sigma\right) = \sum_{\mathbf{s} \in \mathit{Supp}(\sigma)} \left(\prod_{i \in \{1,2\}} \sigma_i\left(s_i\right)\right) u_i\left(\mathbf{s}\right). \tag{41}$$

The expected utility of $i \in \{1, 2\}$ playing a pure strategy $s_i \in S_i$ against j's mixed strategy $\sigma_j \in \mathit{\Sigma}_j$ is

$$u_i\left(s_i, \sigma_j\right) = \sum_{s_j \in \mathit{Supp}(\sigma_j)} \sigma_j\left(s_j\right) u_i\left(s_i, s_j\right). \tag{42}$$

If $\left(\sigma_i, \sigma_j\right) \in \mathit{\Sigma}^{NE}$, any pair $s_i \in \mathit{Supp}\left(\sigma_i\right)$ and $\tilde{s}_i \in \mathit{Supp}\left(\sigma_i\right)$ of $i \in \{1, 2\}$ is such that

$$u_i\left(s_i, \sigma_j\right) = u_i\left(s_i, \sigma_j\right). \tag{43}$$

Which can be rewritten as follows:

$$\sum_{s_j \in \mathit{Supp}(\sigma_j)} \sigma_j\left(s_j\right) u_i\left(s_i, s_j\right) = \sum_{s_j \in \mathit{Supp}(\sigma_j)} \sigma_j\left(s_j\right) u_i\left(\tilde{s}_i, s_j\right). \tag{44}$$

Suppose that Γ' is a transformation of Γ, such that $u'_i = au_i$ for every $i \in \{1, 2\}$, where $a > 0$. The expected utility of $i \in \{1, 2\}$ from playing a pure strategy $s_i \in S_i$ against $\sigma_j \in \mathit{\Sigma}_j$ can be defined as follows:

$$u'_i\left(s_i, \sigma_j\right) = \sum_{s_j \in \mathit{Supp}(\sigma_j)} \sigma_j\left(s_j\right) au_i\left(s_i, s_j\right). \tag{45}$$

If $\left(\sigma'_i, \sigma'_j\right) \in \Sigma^{NE'}$ of Γ', any pair $s_i \in Supp\left(\sigma'_i\right)$ and $\tilde{s}_i \in Supp\left(\sigma'_i\right)$ of $i \in \{1, 2\}$ is such that

$$\sum_{s_j \in Supp\left(\sigma\prime_j\right)} \sigma'_j\left(s_j\right) a u_i\left(s_i, s_j\right) = \sum_{s_j \in Supp\left(\sigma'_j\right)} \sigma'_j\left(s_j\right) a u_i\left(\tilde{s}_i, s_j\right). \tag{46}$$

Which is equivalent to

$$\sum_{s_j \in Supp\left(\sigma\prime_j\right)} a\sigma'_j\left(s_j\right) u_i\left(s_i, s_j\right) = \sum_{s_j \in Supp\left(\sigma'_j\right)} a\sigma'_j\left(s_j\right) u_i\left(\tilde{s}_i, s_j\right). \tag{47}$$

Since $a > 0$ is constant, $\sigma'_j\left(s_j\right) = \sigma_j\left(s_j\right)$ for every $s_j \in Supp\left(\sigma'_j\right)$ of $i \in \{1, 2\}$.

4.4 Application to N-Player Games

Notice that a strictly egalitarian BES of a two player game is a strategy profile $\sigma \in \Sigma^{fm}$, such that

$$\left|1 - u_i^{\iota}(\sigma)\right| = \left|1 - u_{j \neq i}^{\iota}(\sigma)\right|. \tag{48}$$

It follows that BES is such that $u_i^{\iota}(\sigma) = u_j^{\iota}(\sigma)$. This property of the strictly egalitarian BES can be used in the analysis of n-player games. In any n-player game with a unique strictly egalitarian maximally mutually advantageous feasible agreement, the identification of the BES is unproblematic. In other games, hypothetical bargainers could distinguish the maximally individually advantageous feasible agreements associated with a *more equitable* distribution of individual advantage losses (foregone preferred alternatives in ordinal games) from those associated with a *less equitable* distribution of individual advantage losses.

Let $\Gamma = \left(I, \{S_i, u_i\}_{i \in I}\right)$ be any cardinal game, where $I = \{1, ..., n\}$ is the set of players, S_i is the set of strategies of $i \in \{1, 2\}$, and $u_i : \mathcal{L}\left(\times_{i \in I} \Sigma_i\right) \rightarrow \mathbb{R}$ is i's preferences over the set of lotteries over the set of possible agreements. The levels of individual advantage and the set of maximally mutually advantageous agreements are determined in the same way as in the two player case. Let $\sum_{i \in I} u_i^{\iota}(\sigma)$ denote the sum of players' individual advantage levels associated with some feasible agreement $\sigma \in \Sigma^{fm}$. A strictly egalitarian BES must be such that, for every $i \in I$,

$$\frac{u_i^{\iota}(\sigma)}{\sum_{i \in I} u_i^{\iota}(\sigma)} = \frac{1}{n}. \tag{49}$$

In many games, a strictly egalitarian BES will not exist, but the equity of any two feasible maximally mutually advantageous agreements can be compared by comparing the ratio of each player's individual advantage level to the sum of players' individual advantage levels associated with each of the feasible agreements with a ratio $1/m$ that represents the ratio of each player's individual advantage level to the sum of players' individual advantage levels associated with a *hypothetical* strictly egalitarian BES. That is, for any $\sigma \in \Sigma^{fm}$, we can determine the difference between the actual ratio of i's level of individual advantage to the sum of players' individual advantage levels and the ideal egalitarian ratio:

$$\left| \frac{u_i^\iota(\sigma)}{\sum_{i \in I} u_i^\iota(\sigma)} - \frac{1}{n} \right|. \tag{50}$$

For any two feasible agreements $\sigma \in \Sigma^{fm}$ and $\sigma' \in \Sigma^{fm}$, agreement $\sigma \in \Sigma^{fm}$ is more egalitarian than agreement $\sigma' \in \Sigma^{fm}$ if, *for every* $i \in I$,

$$\left| \frac{u_i^\iota(\sigma)}{\sum_{i \in I} u_i^\iota(\sigma)} - \frac{1}{n} \right| < \left| \frac{u_i^\iota(\sigma')}{\sum_{i \in I} u_i^\iota(\sigma')} - \frac{1}{n} \right|^{14}. \tag{51}$$

For example, consider a three player game depicted in Figure 7. The row player chooses between strategies $s1$ and $s2$, the column player chooses between strategies $t1$ and $t2$, and the matrix player chooses between matrices $m1$ and $m2$. To simplify the example, only pure strategy outcomes will be considered.

m1	*t1*	*t2*
s1	$10, 9, 9$	$0, 0, 0$
s2	$0, 0, 0$	$5, 5, 5$

m2	*t1*	*t2*
s1	$4, 4, 4$	$0, 0$
s2	$0, 0$	$6, 8, 10$

Figure 7: Three player coordination game played by three hypothetical bargainers

This game has two maximally mutually advantageous feasible agreements: $(s1, t1, m1)$ and $(s2, t2, m2)$. The worst rationalizable outcome for each player is any strategy profile associated with a payoff of 0. Let $U^\iota_{\{r,c,m\}} = u_r^\iota(\mathbf{s}) + u_c^\iota(\mathbf{s}) + u_m^\iota(\mathbf{s})$ denote the sum of row, column and matrix players' individual advantage levels associated with some $\mathbf{s} \in \mathbf{S}^{fm}$. If this game had an egalitarian BES, it would be some $\mathbf{s} = (s_r, s_c, s_m)$, such that

$$\frac{u_r^\iota(\mathbf{s})}{U^\iota_{\{r,c,m\}}(\mathbf{s})} = \frac{u_c^\iota(\mathbf{s})}{U^\iota_{\{r,c,m\}}(\mathbf{s})} = \frac{u_m^\iota(\mathbf{s})}{U^\iota_{\{r,c,m\}}(\mathbf{s})} = \frac{1}{3}. \tag{52}$$

[14] In any n-player ordinal game, a feasible agreement $\mathbf{s} \in \mathcal{A}^m$ is more egalitarian that $\mathbf{s}' \in \mathcal{A}^m$ if, for every $i \in \{1, ..., n\}$,
$\left| \frac{\mathcal{C}_i(\mathbf{s},\mathcal{A})}{\sum_{i \in \{1,...,n\}} \mathcal{C}_i(\mathbf{s},\mathcal{A})} - \frac{1}{n} \right| < \left| \frac{\mathcal{C}_i(\mathbf{s}',\mathcal{A})}{\sum_{i \in \{1,...,n\}} \mathcal{C}_i(\mathbf{s}',\mathcal{A})} - \frac{1}{n} \right|.$

The sum of players' individual advantage levels associated with agreement $(s1, t1, m1)$ is 2.9. The sum of players' individual advantage levels associated with agreement $(s2, t2, m2)$ is 2.48889. The ratio of each player's individual advantage level to the sum of players' individual advantage levels associated with each outcome, and the difference between each player's actual ratio and the hypothetical egalitarian BES ratio are shown in Figure 8 (the numbers are rounded off to 3 decimal places):

$$\begin{bmatrix} Agr. & \frac{u_r^\iota}{U^\iota_{\{r,c,m\}}} & \frac{u_c^\iota}{U^\iota_{\{r,c,m\}}} & \frac{u_m^\iota}{U^\iota_{\{r,c,m\}}} & \left|\frac{u_r^\iota}{U^\iota_{\{r,c,m\}}} - \frac{1}{3}\right| & \left|\frac{u_c^\iota}{U^\iota_{\{r,c,m\}}} - \frac{1}{3}\right| & \left|\frac{u_m^\iota}{U^\iota_{\{r,c,m\}}} - \frac{1}{3}\right| \\ (s1, t1, m1) & 0.345 & 0.345 & 0.310 & 0.011 & 0.011 & 0.023 \\ (s2, t2, m2) & 0.241 & 0.357 & 0.402 & 0.092 & 0.023 & 0.069 \end{bmatrix}$$

Figure 8: The ratios of each player's individual advantage level to the sum of players' individual advantage levels and the distance between each player's actual ratio and the egalitarian BES ratio.

The difference between the ratio of each player's level of individual advantage to the sum of players' individual advantage level and the egalitarian ratio 1/3 associated with the outcome $(s1, t1, m1)$ is smaller than the one associated with the outcome $(s2, t2, m2)$, which means that the Nash equilibrium $(s1, t1, m1)$ is the BES.

In a two player game, it seems reasonable to assume that a player will not search for the BES if s/he does not believe that the opponent will do that as well. More complicated problems arise in n-player games when some of the players are hypothetical bargainers while others are not. For example, suppose that it is common knowledge among the row and the column player that they are hypothetical bargainers, but they have no information about the matrix player's type. They could not attain the BES of a game depicted in Figure 7 without the matrix player choosing strategy $m1$. In this situation of strategic uncertainty, the row and the column player could resort to playing a combination of strategies $(s2, t2)$, since it guarantees each player a minimum payoff of 5, irrespective of what the matrix player does. This example shows that the BES may not be chosen to be implemented in strategic situations where some of the players are not hypothetical bargainers, or in situations where hypothetical bargainers are uncertain about each other's reasoning mode.

5 Explanatory relevance

One of the fundamental questions pertaining to the hypothetical bargaining theory is whether it can explain real-world decision-maker's behaviour in strategic interactions. Crawford et al. (2008) and Faillo et al. (2013) conducted experiments, in which the participants were presented with two-player 'pie games', in which they had to choose one of the three outcomes represented as segments of a pie. Each segment represents a specific pair of payoffs to the interacting players. If both participants chose the same outcome, they received positive payoffs. An, example of a normal form representation of a 'pie games' is provided in Figure 9.

	R1	*R2*	*R3*
R1	$9, 10$	$0, 0$	$0, 0$
R2	$0, 0$	$10, 9$	$0, 0$
R3	$0, 0$	$0, 0$	$9, 9$

Figure 9: A 3x3 pie game represented in normal form

The structure of a pie games is suitable for testing the theory of hypothetical bargaining, since they share certain structural similarities with the standard bargaining games: The players have to choose between several different allocations of payoffs and in case they do not choose the same allocation they receive nothing. In addition, each allocation of payoffs is a Nash equilibrium.

Tables 1 and 2 summarize the results of Faillo et al. (2013) and Crawford et al. (2008) respectivelly. The theoretical predictions of the BES model are indicated by ^{bes}.

	$G1$	$G2$	$G3$	$G4$	$G5$	$G6$	$G7$	$G8$	$G9$	$G10$	$G11$
$R1$	$9, 10$	$9, 10$	$9, 10$	$9, 10$	$10, 10$	$10, 10$	$10, 10$	$10, 10$	$9, 12$	$10, 10$	$9, 11$
$R2$	$10, 9$	$10, 9$	$10, 9$	$10, 9$	$10, 10$	$10, 10$	$10, 10$	$10, 10$	$12, 9$	$10, 10$	$11, 9$
$R3$	$9, 9$	$11, 11$	$9, 8$	$11, 10$	$9, 9$	$11, 11$	$9, 8$	$11, 10$	$10, 11$	$11, 9$	$10, 10$
$N\,(\%)\,R1$	14	0	51^{bes}	16	48^{bes}	1	51^{bes}	26	16	43^{bes}	6
$N\,(\%)\,R2$	11	1	45^{bes}	4	34^{bes}	3	31^{bes}	22	11	27^{bes}	7
$N\,(\%)\,R3$	74^{bes}	99^{bes}	4	80^{bes}	18	96^{bes}	18	52^{bes}	73^{bes}	30	86^{bes}

Table 1: A summary of Faillo et al. (2013) results. Choices predicted by the BES model are indicated by ^{bes}.

In Faillo et al. (2013) experiment, the BES model is a resonably good predictor of choices in 10 out of 11 games (does not account for 30% of people choosing $R3$ in $G10$). In Crawford et al. (2008) experiment, the BES is a reasonably good predictor in 4 out of 5 games (does not account for choices in $AM3$). These results are by no means conclusive, but they suggest that the BES concept offers an empirically relevant alternative explanation

	$AM1$	$AL1$	$AM2$	$AM3$	$AM4$
L	$5,6$	$5,10$	$5,6$	$5,6$	$6,7$
R	$6,5$	$10,5$	$6,5$	$6,5$	$7,6$
B	$5,5$	$5,5$	$6,5$	$7,5$	$7,5$
$N\,(\%)\,L\,(P1)\,;N\,(\%)\,L\,(P2)$	$6;7$	$0;13$	$53;21^{bes}$	$40;38$	$35;33^{bes}$
$N\,(\%)\,R\,(P1)\,;N\,(\%)\,R\,(P2)$	$6;0$	$7;13$	$16;33^{bes}$	$35;29^{bes}$	$40;33^{bes}$
$N\,(\%)\,B\,(P1)\,;N\,(\%)\,B\,(P2)$	$88;93^{bes}$	$93;73^{bes}$	$32;46^{bes}$	$25;33$	$0,14$

Table 2: A summary of Crawford et al. (2008) results. The choices of player 1 (P1) and player 2 (P2) are presented separately. Choices predicted by the BES model are indicated by ^{bes}.

of how people identify the solutions of games with multiple Nash equilibria, and so the BES model at least warrants further empirical testing.

6 Conclusion

In this paper I predominantly focused on discussing the BES concept as a possible representation of the properties of outcomes that hypothetical bargainers would identify as mutually beneficial and agreeable solutions of non-cooperative games. The proposed solution concept is an equilibrium concept, broadly in line with the traditional equilibrium refinements of non-cooperative games.

The theory of hypothetical bargaining is not a theory of how players coordinate their actions, only how they identify the desirable solutions of non-cooperative games. The game may have multiple bargaining solutions, and so decision-makers' ability to coordinate their actions may depend on factors that have nothing to do with how mutually advantageous it would be for the players to end up at that outcome in terms of their personal payoffs associated with it. For example, the game depicted in Figure 10 has two BES: $(hi1, hi1)$ and $(hi2, hi2)$. The probability of players coordinating their actions on one of the outcomes by choosing strategies $hi1$ and $hi2$ at random is 1/4. The players could coordinate their actions by taking into account the coordination success rate and choose an *ex ante* Pareto dominated outcome (lo, lo) which, due to its uniqueness, *ex post* Pareto dominates outcomes $(hi1, hi1)$ and $(hi2, hi2)$, and so is the *ex post* BES of this game. However, other coordination aids, such as label salience, could also be used (see Bacharach and Bernasconi 1997). The possibility of there being multiple coordination aids for players to choose from may leave them facing a coordination problem of a different type – one related to the choice of a coordination aid to resolve the game they play. The choice of coordination aid will likely depend on decision-maker's beliefs about which aids are most likely to be adopted by other players, which may in turn be determined by decision-makers' social and cultural background, social norms, conventions, and many other factors unrelated to the structure of the game itself.

	hi1	*hi2*	*lo*
hi1	$10, 10$	$0, 0$	$0, 0$
hi2	$0, 0$	$10, 10$	$0, 0$
lo	$0, 0$	$0, 0$	$9, 9$

a

Figure 10: Extended Hi-Lo game

The aim of this paper was not to provide a complete theory of hypothetical bargaining. Further empirical research is required to test the empirical validity of the model, as well as to determine the conditions under which social agents might engage in hypothetical bargaining. Since the model provides testable predictions in experimental games, its further empirical testing seems possible. However, since observed choices can often be explained in terms of multiple accounts of what decision-makers try to achieve in games, these studies may need to consider a broader evidence base than mere observations of choices.

References

Bacharach, M. and M. Bernasconi (1997). The variable frame theory of focal points: An experimental study. *Games and Economic Behaviour 19*, 1–45.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal 120*, 40–79.

Colman, A. M. and J. A. Stirk (1998). Stackelberg reasoning in mixed-motive games: An experimental investigation. *Journal of Economic Psychology 19*, 279–293.

Conley, J. P. and S. Wilkie (2012). The ordinal egalitarian bargaining solution for finite choice sets. *Social Choice and Welfare 38*, 23–42.

Crawford, V. P., U. Gneezy, and Y. Rottenstreich (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *Americal Economic Review 98*(4), 1443–1458.

Faillo, M., A. Smerilli, and R. Sugden (2013). The roles of level-k and team reasoning in solving coordination games. Paper provided by Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia in its series CEEL Working Papers with number 13-06.

Faillo, M., A. Smerilli, and R. Sugden (2016). Can a single theory explain coordination? An experiment on alternative modes of reasoning and the conditions under which they are used. CBES [Centre for Behavioural and Experimental Social Science] Working paper 16-01, University of East Anglia.

Kalai, E. and M. Smorodinsky (1975). Other solutions to the Nash's bargaining problem. *Econometrica 43*(3), 513–518.

Luce, D. and H. Raiffa (1957). *Games and Decisions: Introduction and Critical Survey.* Dover Publications, Inc.

Misyak, J. B. and N. Chater (2014). Virtual bargaining: A theory of social decision-making. *Philosophical Transactions of the Royal Society B 369*, 1–9.

Misyak, J. B., T. Melkonyan, H. Zeitoun, and N. Chater (2014). Unwritten rules: Virtual bargaining underpins social interaction, culture, and society. *Trands in Cognitive Sciences 18*(10), 512–519.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict.* Harvard University Press.

Nash, J. F. (1950). The bargaining problem. *Econometrica 18*(2), 155–162.

Nash, J. F. (1951). Non-cooperative games. *The Annals of Mathematics 54*(2), 286–295.

Raiffa, H. (1953). Arbitration schemes for generalized two person games. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the Theory of Games II*, pp. 361–387. Princeton University Press.

Zhang, D. and Y. Zhang (2008). An ordinal bargaining solution with fixed-point property. *Journal of Artificial Intelligence Research 33*, 433–464.