האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

TWO-PARTY AGREEMENTS AS CIRCULAR SETS

By

BARRY O'NEILL

Discussion Paper # 722 (September 2018)

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

TWO-PARTY AGREEMENTS AS CIRCULAR SETS

Barry O'Neill*

Abstract: In making an agreement with someone, I conditionally promise to perform a certain action, conditioning my obligation on their both making a corresponding promise and performing their action. What promise should I require? That they simply commit to perform is not enough. I should demand the kind of promise I am making myself, and they should demand the same of me. This makes our promises indirectly self-referential. Assuming the performance actions are specified, my promise can be characterized as a set of available promises, all those the other could make to activate my obligation. We have an agreement if each one's promise is a member of the other's promise. Assume that the set P of *available promises* satisfies (1) Aczel's axiom for circular sets; (2) transitivity: if the obligation of $p \in P$ is activated by p', then $p' \in P$; and (3) superset closure: if $p \in P$ is activated by p', p is activated by any promise that implies (is a superset of) p'. The focus is on bargaining procedures that treat the parties symmetrically (e.g., no specified offerer or accepter.) Each party chooses an *agreement promise* p^* such that (4) if both make p^* and one performs, the other is obligated to perform; (5) if one makes p^* and the other does not, the former is not unilaterally obligated. It is shown that among available promise sets of a given size, exactly one contains an agreement promise and contains exactly one of them.

Date: September 12, 2018

*Department of Political Science, University of California, Los Angeles

Acknowledgement: I would like to thank Dov Samet, Yiannis Moschovakis, Arseniy Samsonov, Ellen Cohn, Liam Comerford, and Tom Schwartz for good advice.

1. INTRODUCTION

Why do we have a duty to keep our agreements? Rough intuition might say that agreements are based on suitably connected promises so the question is really: Why do we have a duty to keep our promises? Equating an agreement to mutual promises is attractive since a great deal has been written about when and why promises are binding, and it could be applied to agreements.

Several philosophers have argued the contrary, that agreements are not made up of promises. Gilbert (1993) considers some promise-based definitions and concludes that none of them work. She sees a difficulty in specifying the time when the offerer's obligation arises. Surely an obligation appears when the offer is made, but how can the offerer be obligated when the other has not yet accepted and might never accept? She views an agreement as a jointly adopted decision, a multi-person speech act, which is a separate linguistic category from promising. It is more than each individual promising in the same way that two people taking a walk together is more than their walking along next to each other. Other philosophers have proposed other bases for agreements; according to Mintoff (2004) they are "exchanges" of intentions.

These interpretations have their own difficulties. Defining agreements as exchanges begs the question. Sometimes people exchange insults. If two people find each other's pens on the street they are said to "exchange" them. People exchange views, glances, gifts or insults, but the exchanges that happen in agreements have a necessary element that these other cases do not: the parties are agreeing to it (O'Neill, 2018), and this brings us back to the meaning of agreeing. Also, if agreements are intentions or joint decisions, why are they obligatory? When I tell you that I intend to go to a movie tonight, I am making no moral commitment. Treating agreements as promises, as is done here, immediately identifies the source of the obligation.

The analysis here fits a negotiation procedure that gives neither party a special role. It idealizes the symmetry by having them choose from the same set of possible promises and make their promises simultaneously. A separate paper (O'Neill, 2018) treats the case of one party offering and the other accepting, and addresses Gilbert's objection about when their obligations arise.

Possible applications

An understanding of the foundations of agreements might clarify current debates over contractarian theories of society, the economic theory of contracts, or strategic game models of bargaining. Studies of the latter have focused on the rules of procedure and parties' beliefs about each other's beliefs and goals, but a more detailed notion of the agreement itself might suggest new modeling assumptions.

The analysis might be relevant to tort law. In finding a breach of contract, should the court have the defendant restore the plaintiff's welfare to where it would have been if the contract had been kept, or should it penalize the defendant for doing wrong? The former position fits the view that agreements are mainly matters of efficient planning, to be broken if one party finds it is worth the price. The latter position sees them as moral commitments. In support of the moral viewpoint, Shiffrin (2012) argues that the expediency view of contracts does wide social harm by undermining the general norm supporting promise-keeping. The analysis here bolsters her view by showing that contracts can be seen as promises.

Another application might be to international treaties. These differ from contracts in that they generally cannot be enforced. Traditionally, they have been called "treaties", "conventions", or "agreements", but more recently the United States in particular has avoided words that could be seen as limiting its freedom in security matters, and has tended to use "regimes," "frameworks" or "memoranda of understanding". The 2015 Iranian nuclear agreement was a "Joint Comprehensive Plan of Action." How does a mere plan have binding force? The text stated that each party "envisions" itself taking a certain action, that the action "will" be taken, and that the parties will take such and such "voluntary measures." Given that an agent is free to change an expectation or intention, is this a treaty at all? A reasonable answer would be yes, and the terms it uses are accurate. An international agreement typically works by resetting a poor equilibrium to a mutually beneficial one. It is in the interest of states to follow it, given they believe the other is following it. It typically includes a verification system to justify their beliefs. Its language should not pretend that it is based on morality, since a country that wants to violate it will come up with some justification anyway. By revealing the components of agreements the present analysis might show how far countries can avoid the traditional language and still generate the necessary expectations among their treaty partners.

One current dispute turns on whether the parties made an agreement at all. Russia has insisted that in 1990 Premier Gorbachev accepted the reunification of Germany only if NATO would not expand its membership up to Russia's borders. NATO went on to do that, however, and Russian officials have cited a broken agreement to justify some contentious policies, including intervention in the Ukraine. The NATO states deny that there was any agreement, even an informal one. Records of the 1990 high-level meeting suggest that the elements were expressed as somewhere between expectations and intentions. Shifrinson (2016, 17) writes,
". . . even Russian leaders claiming a broken promise do not argue that the Soviet Union received a formal deal. Instead the question of a non-expansion pledge involves whether various informal, even implicit, statements of US policy in 1990 can be viewed as promises or assurances . . ." The analysis here is relevant to when statements of intentions become promises and when promises link to produce agreements.

Outline

Section 2 shows why it is necessary to condition one's obligation not only on the other's performance but on the other's promise. There are many ways to do this, and, following the general method of Gilbert, the section shows why simple proposals do not work.

To define the idea of a promise within an agreement context, I take as given the actions that each party is to perform, and take as understood that each one's obligation is conditional on the other's performance. All that is left to define a promise is the set of promises from the other that would activate its obligation. The present framework is that the parties have a common set P of available promises, and they simultaneously choose a promise from it. They have an agreement if and only if their promises are members of each other. The circularity in the definition is treated by Aczel's non-standard set theory (1988), and Section 3 outlines its distinctive element, the Anti-Foundation Axiom (AFA). Its essential content is that a set is defined by the membership relationships among its subsets, so that two sets with the same membership structure are identical.

Section 4 puts requirements on the set P of available promises and on the particular promise p^* that both parties choose from it. It states the main result, that given the size of P, p^* is unique. One requirement is that a promise in P conditions its obligation only on promises that are in P, i.e., that the other can make. In set terms, P is *transitive*. Also P satisfies *superset closure*: if an obligation is activated by a certain promise, it is activated by any stronger promise. One promise is *stronger* than another if the former implies the latter. If you commit to paying me $10 if I promise to mow your lawn by Wednesday, and I promise to do it by Tuesday, you cannot object that I am promising too much. Since the actions our fixed, the focus is on another way a promise is stronger, when it is conditioned on less.

According to the theorem, if the number of available promises is fixed, each party knows what to say since among all systems there is exactly one such promise. It has the same form for any size, roughly, "I hereby obligate myself to perform conditional on your performing, and on your making either the promise to me that I am now making to you, or any promise that would produce an agreement if we both made it." Of course it would be easier to say, "I agree to the deal," but the goal is not to replace the normal language of agreeing, but to show its basis.

The theory idealizes the negotiation as the parties making their promises simultaneously. The word is meant in the informational rather than temporal sense, where each promises without knowing what the other is saying. One can think of them holding cards, each one named by a number and with a list of numbers underneath. They talk, then each chooses a card from the pack and places it face down on the table. Together they turn their cards over, and they have an agreement if each card's name is on the other's list. Section 5 argues that this is a reasonable representation of bargaining where neither player has a special role. It is consistent with the assumptions that they have the same set P of available promises and that they choose the same promise p^*. The latter is in fact a corollary of the main theorem.

Section 6 suggests that P's size is a social convention in the sense of Lewis (1969), where each has an incentive to choose use that size set because it expects the other to do so.

Section 7 discusses why the AFA is not a purely technical assumption in the agreement model. The AFA suggests that an agreement is based on mental events, the parties holding intentions to commit themselves and communicating those intentions. It is not a matter of physical gestures or words, like shaking hands or saying "I agree." The final section discusses further research, in particular the possibility of weakening the AFA.

2. PROBLEMS IN DEFINING AGREEMENTS AS PROMISES

Some simple definitions do not work. Suppose Alice wants her lawn mowed and Bob is ready to do it for \$10. Following the premise that the negotiation procedure is symmetrical, they are to make their promises simultaneously. An important point is that each must condition his or her obligation to act not just on the other performing, but on their making a certain promise. Failing to require that puts them at risk. Consider Version 1 as a candidate for defining an agreement. (The details of their performances, e.g., when and how the actions are to be done, are taken as understood, so that "B mows" means B does such and such a job within such and such a time frame. For clarity the parties refer to themselves in the third person, and they use "hereby" to distinguish their performative act of promising from an assertion that someone is promising.)

Version 1. Alice: "A hereby promises to pay if (B mows.)"
Bob: "B hereby promises to mow if (A pays.)"

These are the promises they are supposed to make, but suppose Bob stays silent. Then Alice is conditionally obliged to perform. Bob can decide later whether to mow, and if he does, she must pay him, but otherwise she should have recruited someone else. Alice has a right to complain since she wants to get his commitment now and make plans around it. Bob, of course, has the corresponding worry that Alice might stay silent. Agreements should not be liable to this kind of trick.

Version 2 tries to fix this by augmenting each condition by the underlined parts. The other must perform and must also promise to perform. Each requires the other to say what they were supposed to say in Version 1.

Version 2. Alice: "A hereby promises to pay if (B mows and <u>B promises to mow if A pays.)</u>"
Bob: "B hereby promises to mow if (A pays and <u>A promises to pay if B mows.)</u>"

Version 2's problem is that each person is delivering a weaker promise than the other requires. Bob is putting more conditions on his obligation to perform than Alice told him to. She wants him to promise that he will mow on condition simply that she pay, but his promise also requires her to make a certain promise. Since neither promise satisfies the other's condition, they have not made an agreement. A third version might try to patch this by augmenting each condition with the promise the other was supposed to make in Version 2. Then they would have an agreement but be back to the difficulty of Version 1, that one party could say less than prescribed and leave the other unilaterally committed (O'Neill, 2018).

Whatever the right promise is, the symmetry of the situation suggests it should be the same for both of them, and indeed the result of Section 4 will lead to this. ("Same" is abstracted from their identities and from the deeds they are supposed to perform. The promises in Version 2, for example, are the same in this sense: "I promise to perform if you both perform and promise to perform if I do.") This section's argument suggests that each should say, "I hereby promise to perform if you perform, and you make this same promise to me." This almost works but not quite because it violates the persuasive condition of superset closure, described in Section 4.

3. NON-WELL-FOUNDED SETS

If my promise is conditional on yours and vice versa, we are in a circle so that analyzing the situation in English will be confusing. The concept of a promise will be formalized. It can be characterized as a set: Alice's set contains those promises any of which from Bob would activate her obligation, and vice versa.

The circularity suggests the use of non-well-founded (NWF) set theory. After Bertrand Russell came upon his paradox, most axiomatic set theories have prohibited sets that are circular, either directly, like $p \in p$, or indirectly, like $p \in q \in p$. However, the regular axioms can be altered to allow self-membership. Aczel's approach (1988) has been the most studied, probably because it is simple to explain and, in a definable sense, adds the fewest new sets to the standard universe. His system is consistent if and only if the standard one is consistent (Aczel, 1988). Full statements and proofs appear in Devlin (1993), Barwise and Etchemendy (1987), Barwise and Moss (1996), Moschovakis (2006, Ch. 11 and App. B), and Aczel's original book (1988).

The standard way to avoid Russell's Paradox is restrict the universe of sets by means of the *foundation axiom* (FA). It forbids any p that allows an infinite sequence

$$. . . \in p_3 \in p_2 \in p_1 \in p. \tag{1}$$

If we rotate (1) vertically with p on top, the FA is saying that starting with p and proceeding downwards, we must eventually reach "foundations." In a metaphor of boxes (Barwise and Moss, 1988), someone gives me a present of p, and I open it to find other boxes, including p_1, which I open to find p_2, and so on. The FA says that I will not open boxes forever, that I will eventually come to ones with either memberless objects or nothing at all. The FA blocks the derivation of Russell's Paradox since it involves a set that belongs to itself:

$$. . . \in p \in p \in p \in p. \tag{2}$$

The FA also excludes the present concept of an agreement by which Alice's and Bob's promises satisfy $p_A \in p_B$ and $p_B \in p_A$. Each opens their box to find one that contains the original box:

$$. . . \in p_A \in p_B \in p_A \in p_B. \tag{3}$$

The central issue for an NWF system is: When do different descriptions represent the same set? In the regular theory, $p = \{r, s\}$ and $q = \{r, s\}$ are identical because their memberships are identical. This principle holds in NWF theory but is harder to apply. Are $p = \{p\}$ and $q = \{q\}$ identical? To go by their memberships would beg the question: p and q are the same if and only if $p = q$. The AFA regards two sets as identical whenever it can take them as such without contradiction. It will turn out that $p = \{p\}$ and $q = \{q\}$ are indeed the same, the set uniquely defined by the equation $\Omega = \{\Omega\}$. Other equations and systems of equations define sets. For example, there is exactly one Ω^* such that $\Omega^* = \{\varnothing, \Omega^*\}$.

Aczel's sets can be stated either as equations or as directed graphs, and we start with the graph approach. Let the standard Zermelo-Frankel universe of sets be U and let Azcel's universe be V. In this paper these will include only *pure* sets in the sense that the only memberless entity in U or V is the empty set $\varnothing$. Sets in either universe are designated p, q, r . . . . Define a *directed graph* as a pair $<G, \rightarrow>$ where G is a finite set of nodes and $\rightarrow \subseteq G \times G$ is a set of directed edges. Our application can drop finiteness without much change. Sometimes *"G"* is used for the graph itself. Nodes are designated $a, b, c, \ldots,$ and instead of *"$(a, b) \in \rightarrow$"*, we write *"$a \rightarrow b$"*. If $a \rightarrow b$, then b is called a *child* of a and a is *parent* of b. Node 3 is a parent of all the nodes in Figure 1. Thus it is a parent and child of itself. (The subscripts of the sets indicate their sizes: $|p_i| = i$. This convention will be followed when possible, i.e., when the sizes are different.)

An *accessible pointed graph (apg), $<G, \rightarrow, a^*>$,* is a directed graph with a distinguished node a^* such that there is a directed path from a^* to every other node. The distinguished node is called the *point*. In Figure 1 the point is taken to be node 3, although the same non-pointed graph would yield an apg whose point is 1. Node 0 could not be a point because it fails the accessibility requirement.

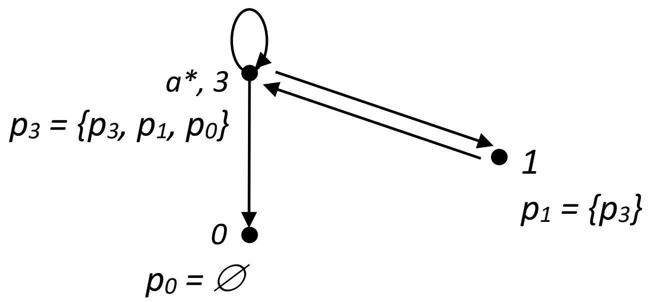

Figure 1. An apg with nodes 3, 1, and 0, and point 3. The point is typically drawn as the highest node. Sets p_0, p_1 and p_3 constitute a decoration.

The Anti-Foundation Axiom

A *decoration d* of a graph is a mapping from its nodes to sets, $d: G \rightarrow V$, such that for $a \in G$, $d(a) = \{d(b) : a \rightarrow b\}$. That is, a node is *decorated* by the set comprising the sets decorating its children. An apg *picture*s a set p if and only if $p = d(a^*)$ for some decoration d. It can be checked that the assignment in Figure 1 is a decoration. By substitution we can express all unknowns in p_3 and conclude that Figure 1 pictures any set satisfying $p = \{p, \{p\}, \varnothing\}$. So far nothing has been said about whether this equation has no solution in V, or one, or many.

The apg's in Figure 2 have no cycles, so they picture well-founded sets. A finite cycle-free graph has a unique decoration since we can assign $\varnothing$ to the childless nodes, then work from children to parents, finishing in a finite number of steps.

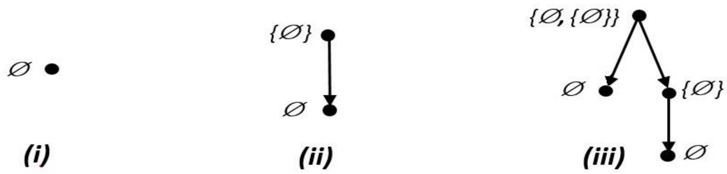

Figure 2. Graphs decorated by well-founded sets.

The apg's of Figure 3, in contrast, have cycles, and this fact raises the issues of existence and uniqueness of the sets they picture. If sets pictured by Figure 3(i) and (ii) existed, they would satisfy $\Omega = \{\Omega\}$, just as any sets pictured by (iii) and (iv) would satisfy $\Omega^* = \{\varnothing, \Omega^*\}$.

The *Anti-Foundation Axiom* asserts:

Every decorated apg pictures exactly one set.

It is claiming existence and uniqueness of the pictured sets, or, more accurately, declaring them. Accordingly, exactly one set contains only itself (Figure 3(i)), and exactly one set contains only the empty set and itself (Figure 3(ii)).

Since an apg can be constructed to picture any well-founded set, Aczel's universe *V* contains the standard universe *U*, and since NWF sets exist (Figures 1 and 3), *U* is a proper subclass of *V*. The system comprising the regular Zermelo-Frankel axioms for sets with the axiom of choice is labeled ZFC, and the system produced by substituting AFA for FA is labeled ZFA.

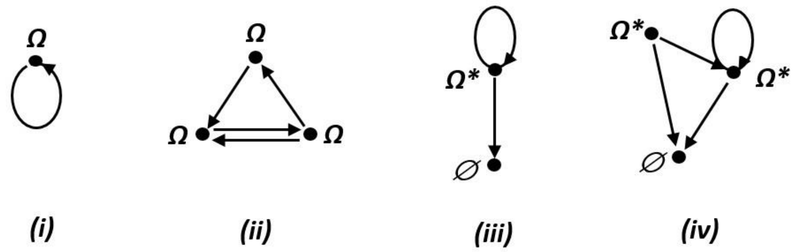

Figure 3. Apg's (i) and (ii) picture $\Omega = \{\Omega\}$; (ii) and (iv) picture $\Omega^* = \{\varnothing, \Omega^*\}$.

The AFA can also be formulated in terms of systems of set equations, according to what is called the *solution lemma*. Consider *n* set-valued variables connected by *n* equations whose left hand side involves a single variable and where each variable appears there exactly once, and where a subset of the variables appears on each right hand side. Exactly one *p* satisfies the system $p = \{p, q, \varnothing\}$, $q = \{p\}$, and $\varnothing = \{\}$, which corresponds to Figure 3. According to the solution lemma, the AFA implies that the *n* equations have a unique solution (Devlin, 1994). The rough way to see this is to construct a graph with *n* nodes labeled by the *n* variables and with the edges leading into a node determined by the corresponding equation.

Set identity and the maximum bisimulation

A given apg pictures exactly one set according to the AFA, but a given set is typically pictured by many apg's, which may not be isomorphic (Figure 3). To develop a graphical criterion for set equivalence, Aczel uses the concept of the maximum bisimulation. He defines equivalence in a circular way: two nodes are the same if and only if they have identical relationships to other nodes that are the same.

Definition: Let $(G,\rightarrow)$ be a graph and *R* a relation on $G \times G$. Then *R* is a *bisimulation* on *G* if and only if for all $a, b \in G$,

(B1) if aRb and $a \rightarrow a'$, then $\exists b'$ such that $a' R b'$ and $b \rightarrow b'$, and

(B2) if aRb and $b \rightarrow b'$, then $\exists a'$ such that $a' R b'$ and $a \rightarrow a'$.

A graph may have many bisimulations - the identity relation $I = \{(a,a), (b,b), . . . \}$ is always one – but set equivalence is determined by the *maximum bisimulation*, the one containing every relation that appears in any bisimulation. There is exactly one maximum bisimulation and it induces an equivalence relation on the nodes (Moss, 2008). The maximum bisimulation of Figure 4(i) is the relation $\{(a,b), (b,a), (c,d), (d,c)\} \cup I$. To show this we note that it must include (a,b): a has a child c, and since b is in a's equivalence block, by B1 b must have a child in c's block. Indeed it has d. Another child of a is a itself and it is in a's block, so b must have a child in a's block, and it has a. The same check must be carried out for b's children as well as for c and d. Given that the maximum bisimulation is an equivalence, checking B2 is redundant, but we must verify that the partition is maximum, e.g., that $\{a,c\}$ etc., are not in it. Dovier *et al.* (2004) give an algorithm.

To extend the definition to equivalence across two apg's, define the *disjoint union* of two graphs with disjoint node sets as the graph whose node set is the union of the individual node sets and whose edge set is the union of their edge sets. The NWF criterion for set identity is: if the points of two apg's are equivalent in their union's maximum bisimulation, the apg's picture the same set. We can apply the definition to show that Figures 4(ii) and (iii) picture the same set. Their union has eight nodes, and it can be checked that the maximal bisimulation puts three of them in one block decorated by p, three in a block decorated by r, and the final two in a block decorated by $\varnothing$. Both apg (ii) and apg (iii) picture $p = \{p, \{p, \varnothing\}\}$.

The canonical picture of a set

Figure 3(ii) pictures the set Ω but does it inefficiently. It has three nodes when a single one will do (Figure 3(i)). Also, (ii) is an incomplete picture since it omits some membership relations. Since every node's decoration is a subset of every other's, the full representation would have all possible edges, including loops.

Every set p has a unique *canonical picture*, an apg that pictures it fully with the fewest nodes. It is constructed by assigning exactly one node to p and to every p' such that $p' \in \ldots \in p$. (The rule of one node per set presumes a notion of set identity, which is, of course, the maximum bisimulation.) The edge $a \rightarrow b$ is included if and only if the set decorating node b is a member of the set decorating node a. Since the assignment of sets and nodes is 1-1, the apg has no redundant nodes. Each p has a unique canonical picture since the definition tells us how to construct it. For example, Figures 4(i) and (ii) are not canonical, but (iii) canonically pictures $p = \{p, \{p, \varnothing\}\}$.

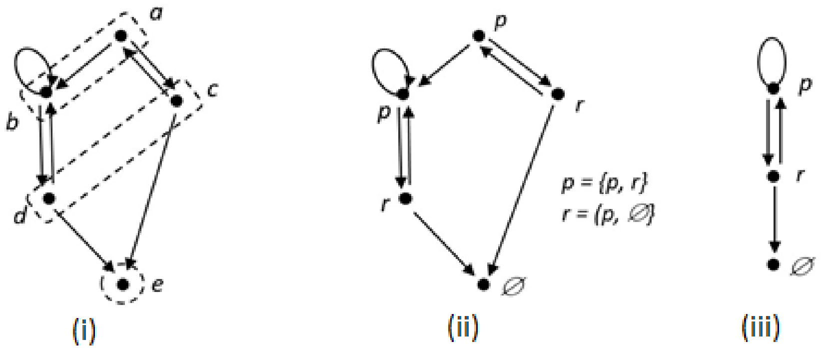

Figure 4. Equivalence within a graph: a and b in (i) are decoratable by p, as (ii) shows, and c and d are decoratable by r. Between graphs: the maximum bisimulation on (ii) and (iii) shows that both picture set p.

We can now say more about how moving from the FA to the AFA increases the standard universe. Table 1 shows the count of sets for each size of canonical picture. The WF and NWF values are, respectively, from Peddicord (1962) and Miloto and Zhang (1998). For $n = 1$ the only WF set is $\varnothing$. Allowing NWF sets adds Ω. For $n = 2$ the WF group has only $\{\varnothing\}$, and the NWF case adds $\Omega^* = (\varnothing, \Omega^*\}$. After that the NWF numbers grow quickly. The superset closure axiom of Section 4 will limit the NWF sets that are possible promises: for n = 1 to 6 the numbers are 2, 1, 4, 6, 8,10, . . . and will limit the sets that are possible promises systems to 2, 2, 2, 2, 2. . . (Figure 5).

n	**1**	**2**	**3**	**4**	**5**	**6**
WF	1	1	2	9	88	1,802
WF & NWF	2	2	16	504	52,944	17,294,923

Table 1. Numbers of sets with an n-node canonical picture.

4. PROMISE SYSTEMS AND THE AGREEMENT PROMISE

Example 1: Suppose each of two parties chooses from:

p_3 – "I promise to perform."
p_1 – "I promise to perform if you make promise p_3."
p_0 – "I promise nothing."

Given the earlier assumption that every promise obligation is implicitly conditional on the other's performing, the explicit version of p_3 would be "I promise to perform if you perform."

As sets these are

$p_3 = \{p_3, p_1, \varnothing\}$,
$p_1 = \{p_3\}$,
$p_0 = \varnothing$.

The set of the three available promises is designated P_3, which is identical to p_3. Figure 1 is P_3's canonical picture.

A *promise system* is a triple $<P, \in, \equiv>$, where $P \subset V$ is interpreted as the set of available promises, where these satisfy ZFA, where $\in$ is the membership relation on $P \times P$, and where $\equiv$ is an equivalence relationship on P x P representing set identity following the AFA. We will often refer to a promise system simply as P, and $|P|$ will be designated n. We require

(Transitivity) for all $p \in P$, $p \subseteq P$;

(Superset Closure) for all $p, q, q' \in P$, if $q \subset q'$ and $q \in p$, then $q' \in p$.

Transitivity means that a promise in P can require from the other party only promises that are available. Superset closure means that if p is activated by the other party making q, and if q' is a stronger promise than q, then p is activated by q'. Here promise q' is stronger than q if and only if its condition is weaker, i.e., if $q \subset q'$. The promises of Version 1 in Section 2, for example, are stronger than those of Version 2. Example 1 satisfies superset closure, but it would violate it if p_3 were changed to $p_3' = \{p_3', \varnothing\}$, since if $\varnothing \subset p_1$ and p_3' accepted $\varnothing$, then p_3' should accept p_1, which it does not. Superset closure does not seem to have been used in NWF applications but it is convincing in this context.

Turning to their choice of an appropriate promise, define $p^* \in P$ to be an *agreement promise* for P if and only if

(Self-effectiveness) $p^* \in p^*$, and

(Invulnerability) for all $p \in P$, if $p \in p^*$ then $p^* \in p$.

Self-effectiveness means that if each party makes p^*, they have an agreement. Invulnerability can be understood by its violation: if one party is expected to make a vulnerable p^*, the other can choose another p that leaves the first party unilaterally obligated. An invulnerable pair of promises is analogous to a Nash equilibrium pair of strategies in a game. These requirements represent the two arguments we used against the unsuccessful approaches in Section 2.

The two families of promise systems

Two families of promise systems are now constructed and it is shown that for all $n \geq 1$, each contains exactly one system of size n. They are designated N_n and A_n to suggest "no agreement promise" and "agreement promise", since, according to the theorem, N_n has none and A_n has exactly one. Let P designate the promise set of A_n or N_n. Given the convention $i = |p_i|$, with $n + 1$ possible sizes and n promises, one subscript will be missing, and it is designated j^*.

For N_n and A_n, define P as $\{p_0, p_1, \ldots p_{j^*-1}, p_{j^*+1}, \ldots p_n\}$, where each p_i comprises P's final i promises:

$$\begin{aligned} p_i &= \varnothing && \text{for } i = 0, \\ &= \{p_{n-i+1}, \ldots p_n\} && \text{for } i = 1 \text{ to } j^* - 1, \\ &= \{p_{n-i}, \ldots p_{j^*-1}, p_{j^*+1}, \ldots p_n\} && \text{for } i = j^*+1 \text{ to } n. \end{aligned}$$

The missing index is

$$\begin{aligned} j^* &= n/2 && \text{for } N_n,\ n \text{ even;} \\ &= n/2 + 1/2 && \text{for } N_n,\ n \text{ odd;} \\ &= n/2 + 1 && \text{for } A_n,\ n \text{ even;} \\ &= n/2 - 1/2 && \text{for } A_n,\ n \text{ odd.} \end{aligned}$$

The sets in N_n or A_n are totally ordered by set inclusion, and accordingly by strength of promise. The stronger promises, those with higher subscripts, accept a certain weak promise and everything higher in the order, including themselves. The weaker ones accept a strong promise and everything higher.

Theorem. Exactly two promise systems of size n exist, N_n and A_n. System N_n has no agreement promise and A_n has exactly one, its weakest self-effective promise.

The Appendix gives the proof. A computer routine checked its validity up to $n = 8$. Figure 5 shows N_n and A_n up to $n = 6$. Except for A_2, each graph can be used to generate the canonical pictures of its sets, except for $\varnothing$, by changing its point. The agreement promise in A_n is $p_{n/2}$ for n even and $p_{n/2+1/2}$ for n odd. The promises of larger systems can be generated in sequence by conditioning promises on the simplest promises (the unconditional one and the null one), then on promises conditioned on promises, etc., just as Version 1 led to Version 2.

5. ARE REAL AGREEMENTS FORMED SYMMETRICALLY?

The assumption that the parties choose from the same set of promises and choose the same one was justified by the similarity of their bargaining roles. Some procedures violate this by allowing one negotiator to make a credible take-it-or-leave-it offer. Other procedures build in symmetry. In *filtered mediation* a mediator asks each party separately for an offer and reveals what they said only if the offers are compatible (Jarque, *et al*., 2003; O'Neill 2003). Ultimatum bargaining and filtered mediation are asymmetrical or symmetrical because of their rules, but most negotiations are free-form, and in these cases the symmetry assumption is plausible. Negotiation is like a relationship where the couple's commitment cannot be seen as happening at any identifiable time. Little by little they come to a common belief about what it is. In the

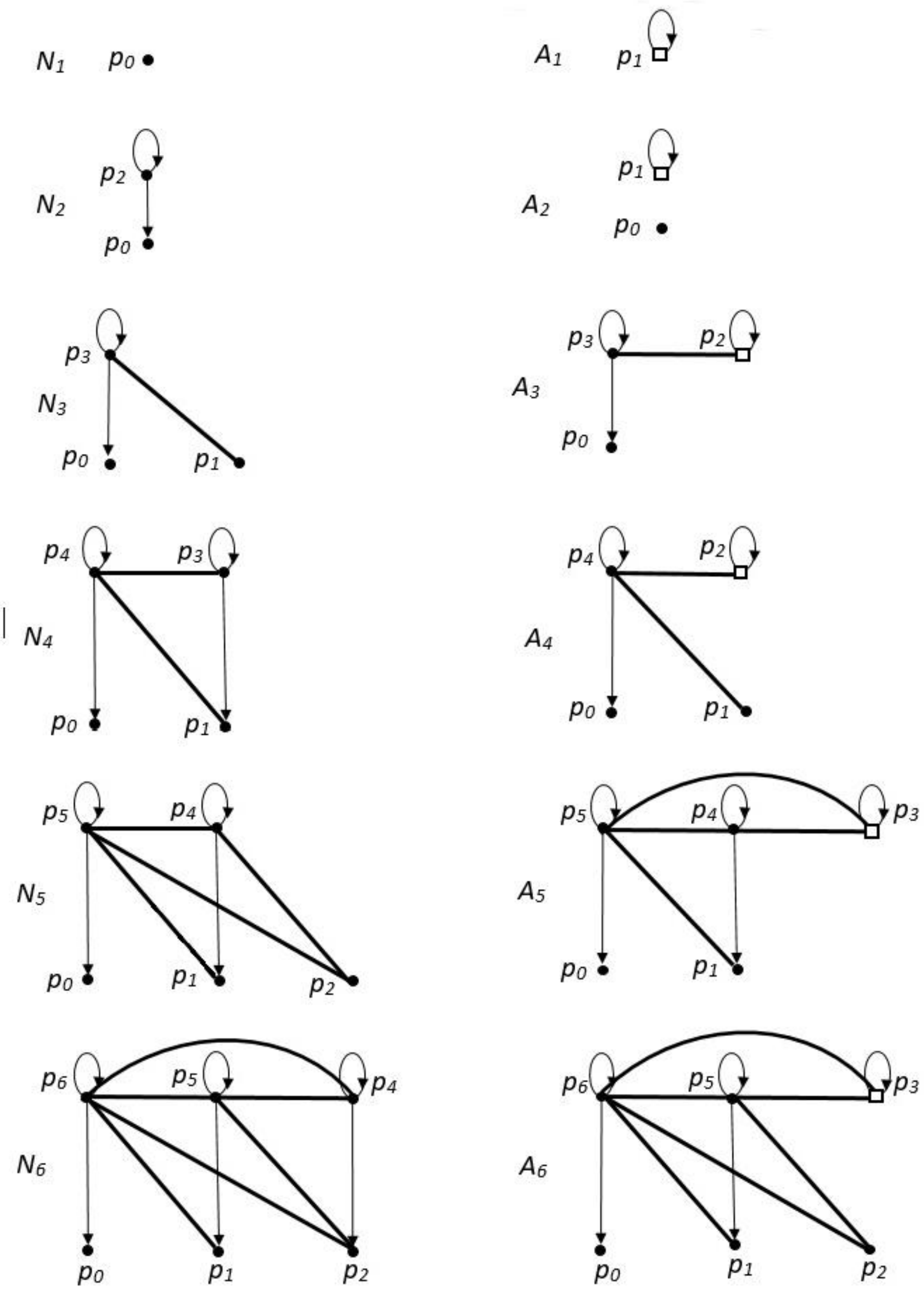

Figure 5. Promise systems for n = 1 to 6. Each graph generates an apg for each node except p_0. Thick lines show membership in both directions, and square nodes are agreement promises.

same way at the end of negotiation one might say, "so we understand each other that . . . ", but that is not the formation of the agreement. They already share an understanding and now are putting it on the record. Many negotiations have a final stage where one party and then the other sign a document, but, again, they are ratifying what they have already agreed on. If after the negotiation one of them balked at signing, that could be objected to as bad faith. Sometimes a party will try to introduce asymmetry by declaring their offer "final", but announcing that is short of announcing it credibly. Usually the other can respond with their own "final" offer. Symmetry fits the procedure of many negotiations in that they have no procedure.

In fact the first symmetry claim, that they have the same choices available, implies the second, that they choose the same one. Call p and p' *mutually effective* if and only if $p \in p'$ and $p' \in p$.

> **Corollary.** No promise system contains two promises that are invulnerable and mutually effective.

To show the corollary for N_n, n even, for example: if p_i, $i < j^*$, is mutually effective with p_k, then $k > j^*$, and all such p_k are vulnerable.

6. DETERMINING THE SIZE OF THE AVAILABLE SET

To choose their agreement promise the parties need a shared understanding of the size of the promise system, but there seems to be no persuasive way to determine that. The agreement promise in a size-n system is the unique one that accepts itself and all self-effective promises, so a tempting move might be to define the "universal" agreement promise, call it p^+, as the one that accepts itself and self-effective promises in systems of any size. However, this would violate closure: for A_2 (Figure 5) the agreement promise is Ω defined as $\{\Omega\}$, and if p^+ included Ω, by closure Ω would include p^+, contradicting Ω's definition. Also, The rule for moving from one A_n to the next larger one makes the agreement promises formally similar to the ordinal numbers, and with no maximum size on the systems, p^+ would correspond to "the largest ordinal", which is a contradiction as Burali-Forti showed (Copi, 1952).

It seems better to choose some reasonable finite n and take the choice as a social convention in the sense of Lewis (1969). A *social convention* is a rule telling members of a group what to do in a certain kind of situation; it is arbitrary in that other rules might do but it motivates members to follow it given each one's belief that the others will do the same. In this application, the incentive condition holds, since for $n \neq n'$ no promise in A_n contains any in $A_{n'}$, and if a party chose a promise from a system of an non-conventional size, there would be no agreement.

7. THE FACTUAL CONTENT OF THE ANTI-FOUNDATION AXIOM

Does the AFA say anything substantial about agreements, and if so, is it correct? Some formal systems include assumptions with no factual content. Length, mass, time, *et al.*, are taken as real rather than rational numbers, but for technical convenience, and no empirical test could confirm or disconfirm the assumption. The AFA, however, suggests that a certain empirical situation holds and it seems consistent with a reasonable conception of promising and agreeing. As shown below, it fits with the idea that promises are mental acts rather than physical words or gestures. If parties communicate their intentions to make an agreement, they have one. The AFA can deal with situations outside this assumption but only in an awkward and artificial way.

Suppose that two parties can form an agreement by either both raising their hands (p_r) or crossing their hearts (p_c). They make their gestures simultaneously and if they fail to choose the same one, they have no agreement. We give them two further options, making an unconditional promise (p_4) by making both gestures at the same time, or making no promise (p_0) by doing nothing.

$$p_4 = \{p_4, p_c, p_r, p_0\};$$
$$p_r = \{p_r, p_4\};$$
$$p_c = \{p_c, p_4\};$$
$$p_0 = \varnothing.$$

This set of available promises, labeled P_{rc}, satisfies transitivity and superset closure. Figure 6(i) pictures it and shows that exactly two promises satisfy self-effectiveness and invulnerability: raising one's hand and crossing one's heart.

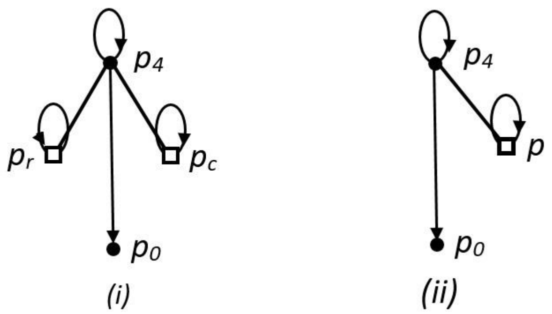

Figure 6. (i) To make an agreement, both raise their hands (p_c) or both cross their hearts (p_r). Apg (ii) with p as its point is the canonical picture of (i) and identifies p_c and p_r.

The system differs from those discussed above since it allows two different agreement promises, so that parties might want a deal but fail to coordinate. The AFA, however, regards the two promises as identical. Both can be decorated by $p = \{p_4, p\}$, with $p_4 = \{p_4, p, \varnothing\}$ and $p_0 = \varnothing$. (Their bisimilarity is not surprising since the graph has an automorphism mapping the two nodes into each other.) The canonical picture of p_4, Figure 6(ii), has only three nodes, and it is A_3 of Figure 5.

In declaring the heart and hand moves as the same, the AFA is claiming that the parties are making an agreement even when their gestures do not match. I believe this is reasonable given an appropriate conception of agreeing, that it is a mental event involving the communication of beliefs and intentions. It does not depend on uttering certain words like "I agree" or "I promise", or taking visible actions like signing a document or shaking hands. Even when their gestures do not match, the parties are telling each other that they hold the right intention, so by the mental conception of an agreement they are making one. (Scanlon, 1990, discusses this idea in the context of individual promises, concluding that one individual can make a promise to another even without a shared language or mutually understood gestures.)

A counterargument might be that one can imagine a real social custom that requires matching outward gestures as P_{rc} does. The present model can handle this by allowing conditionality in the performances. The model would interpret raising one's hand as, "I promise to (mow your lawn <u>if you raise your hand during your promise</u>), on condition that you pay me, etc." (The approach up to now has had conditionality only in the obligations.) Agreement promises with conditionality in their performances happen - we use them when we make a bet – but this way of formulating an generic agreement is unnecessarily complicated. One can often maintain an odd theory by inventing complicated rules for connecting it to observables, but under the AFA the purely mental conception is the most straightforward.

8. CONCLUSION

Non-well-founded sets have been used in logic, philosophy, computer science, and to a lesser extent in the theory of common knowledge (Lismont, 1994; Heifetz, 1996). The present application seems to be their furthest entry into the social sciences, but its purpose is not just to add a new method. Treating the problem in regular set theory would be cumbersome, and many of the results proven for NWF sets would not be available.

Non-well-founded sets tend to prompt suspicion among mathematicians, but that may be because they go against familiar ways of thinking. According to Azcel (1988, xvii) we are used that we "construct" sets, starting at the bottom, but this is just a metaphor to make an abstract object seem tangible. The issue is not whether an entity seems usual and can be treated as a physical object, but whether it leads to interesting results and applications, and NWF sets seem to fit many human phenomena. If NWF sets like $\Omega = \{\Omega\}$ seem mysterious, they is no more so than $\{\varnothing\}$ or $\{\varnothing, \{\varnothing\}$ of the regular theory. At its beginnings, formal set theory was open to them and adopted the Foundation Axiom mainly to avoid inconsistency.

The analysis shows that, contrary to some philosophers, agreements are reducible to promises. "I agree to the deal" means something like, "I hereby obligate myself to perform conditional on you performing and on you making either the promise to me that I am now making to you, or any promise that would produce an agreement if we both made it." The fact that such an explication is possible clarifies the debates about the contracts and the role of implied intentions in international treaties. By showing the mental approach is coherent, the analysis reduces the force of objections that the NATO states were not bound because they avoided being explicit, or that the Iranian deal was not a deal because it avoided certain words.

An interesting abstract result is that adding superset closure to the AFA drastically reduces the possible sets. A further project would add the superset closure axiom to an NWF system other than Aczel's. Some systems would preserve some of the consequences presented here; for example, Finsler's NWF system (1926; Aczel, 1988) also treats the hand and heart moves (Figure 6) as identical and fits the mental approach. Since the other systems have larger universes, more promise sets and more agreement promises might appear. The results might suggest further requirements for an appropriate promise.

REFERENCES

Aczel, Peter. *Non-Well-Founded Sets.* (1988). CSLI Lecture Notes, 14. Stanford: Stanford University, Center for the Study of Language and Information.

Barwise, Jon, and John Etchemendy. (1987). *The Liar*. Oxford: Oxford University Press.

Barwise, Jon, and Lawrence Moss. (1996). *Vicious Circles, On the Mathematics of Non-Well-Founded Phenomena.* Chicago: University of Chicago Press.

Black, Oliver. (2012). *Agreements, A Philosophical and Legal Study.* Cambridge: Cambridge University Press.

Copi, Irving. (1958). The Burali-Forti Paradox. *Philosophy of Science.* 25(4): 281-86.

Devlin, Keith. (1994). *The Joy of Sets: Fundamentals of Contemporary Set Theory.* 2nd Edition. Springer.

Dovier, Agostino, Carla Piazza and Alberto Policriti. (2004). An efficient algorithm for computing bisimulation equivalence. *Journal of Theoretical Computer Science.* 311(1-3): 221-56.

Finsler, Paul. (1926). Über die Grundlagen der Mengenlehre, I. *Math. Zeitschrift*, *25*: 683–713; translation in Booth, David (1996). *Finsler Set Theory: Platonism and Circularity*. New York: Springer.

Gilbert, Margaret. (1993). Is an agreement an exchange of promises? *Journal of Philosophy*. 90(12): 627-49.

Heifetz, Aviad. (1996). Non-well-founded type spaces. *Games and Economic Behavior.* 16(2): 202–17.

Jarque, Xavier, Clara Ponsati, and József Sákovics. (2003). Mediation: incomplete information bargaining with filtered communication. *Journal of Mathematical Economics.* 39(7): 803–30.

Lewis, David (1969). *Convention.* Cambridge: Harvard University Press.

Lismont, Luc. (1994). Common knowledge: relating anti-founded situation semantics to modal logic neighbourhood semantics. *Journal of Logic, Language and Information.* 3(4): 285-302.

Milito, Elaine, and Li Zhang. (1998). Classification of minimal finite hypersets. *Congressus Numerantium.* 133: 85-93.

Mintoff, Joseph. (2004). Is an agreement an exchange of intentions? *Pacific Philosophical Quarterly*. 85(1): 44–67.

Moschovakis, Yiannis. (2006). *Notes on Set Theory, 2nd ed*. New York: Springer.

Moss, Lawrence. (2008). Non-well-founded set theory. *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/spr2017/entries/nonwellfounded-set-theory/ .

O'Neill, Barry. (2003). Mediating national honor: lessons from the era of dueling. *Journal of Institutional and Theoretical Economics.* 159(1): 229-47.

O'Neill, Barry. (2018). Promises, Agreements, and Good Faith. Manuscript, UCLA Department of Political Science.

Peddicord, Richard. (1962). The number of full sets with *n* elements. *Proceedings of the American Mathematical Society.* 13(5): 825-28.

Russell, Bertrand. (1902). Letter to Frege. Pp. 124-25 in Jean van Heijenoort, ed. *From Frege to Gödel.* Cambridge, Mass: Harvard University Press, 1967.

Scanlon, Thomas. (1990). Promises and practices. *Philosophy & Public Affairs.* 19(3): 199-226.

Sheinman, Hanoch. (2012). Agreement as joint promise. Ch.16 in Hanoch Sheinman, ed., *Promises and Agreements: Philosophical Essays*. Oxford: Oxford University Press.

Shiffrin, Seana. (2011). Immoral, conflicting and redundant promises. In Jay Wallace, Rahul Kumar and Samuel Freeman, eds. *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon.* New York: Oxford University Press.

Shiffrin, Seana. (2012). Is a contract a promise? Ch.16 in Andrei Marmor, ed. *Routledge Companion to Philosophy of Law.* New York: Routledge.

Shifrinson, Joel. (2016) Deal or no deal? The end of the Cold War and the US offer to limit NATO expansion. *International Security*. 40: 7-44.

Takashi, Nitta, Okada Tomoko, and Athanassios Tzouvaras. (2003). Classification of non-well-founded sets and an application. *Mathematical Logic Quarterly*. 49(2): 187–200.

APPENDIX

Theorem. Exactly two size-n promise systems exist, N_n and A_n as defined in Section 4. System N_n has no agreement promise and A_n has exactly one, its weakest self-effective promise.

Proof: The first part of the proof will establish that N_n and A_n are promise systems and have n different members. The second part will show the converse, that any n-size promise system is one of these or the other. It will follow that the promise systems of size n are exactly these two.

A set's property of having n different members will be called *distinctness* where the value *of* n will be clear from the context. The following schema for superset closure will be used repeatedly:

$$\text{if } x \subset y \text{ and } x \in z, \text{ then } y \in z.$$

The first part of the proof will be stated only for N_n with n even, since the changes for the other cases are obvious. To show that N_n is a promise system: A promise system requires transitivity, superset closure, and the AFA. The definition of N_n implies transitivity. Superset closure means that (i) $p_i \subset p_k$ and (ii) $p_i \in p_j$ imply (iii) $p_k \in p_j$. The definitions of p_i imply both that condition (i) holds if and only if $i < k$, and that $i < k$ and (ii) together imply (iii). Therefore N_n satisfies superset closure.

Regarding the third criterion: Given the construction of N_n, the only implication of the AFA needing proof is that the n members are distinct. Construct a graph G with n nodes labeled $a_0, \ldots\ a_{n/2-1}, a_{n/2+1}, \ldots a_n$, (note that the missing index is $n/2$) and with edges such that $a_i \rightarrow a_j$ for nodes a_i, $a_j \in G$ if and only if $p_j \in p_i$ for sets p_i, $p_j \in N_n$. Clearly $d(a_i) = p_i$ is a decoration of G. Let $E(a_i)$ be the set of nodes equivalent to a_i by the maximum bisimulation. The goal is to show that $E(a_i) = \{a_i\}$ for every a_i, i.e., that the maximum bisimulation is identity. The induction argument will be applied for a fixed n and proceed from the extreme values of the indices towards the central ones. For the base cases a_0 and a_n: all nodes other than a_0 have children in some equivalence block but a_0 has no child in any block, so by the bisimulation condition B1, $E(a_0) = \{a_0\}$. Also by B1, $E(a_n) = \{a_n\}$ since only a_n has a child in $\{a_0\}$.

To show the induction steps: For a given $i \in [0,\ n/2 - 2]$ assume that each of $a_0, \ldots a_i$ and $a_{n-i}, \ldots a_n$ is a singleton in its block. The first claim to be proved, that $E(a_{i+1}) = \{a_{i+1}\}$, requires us to show that a_{i+1} is not in a block with any of $a_{i+2}, \ldots a_{n/2-1}, a_{n/2+1}, \ldots a_{n-i-1}$. This follows from B1, since by the definition of N_n, a_{i+1} has no child in $E(a_{n-i-1})$ while the others have a_{n-i-1}.

To show the second claim, that $E(a_{n-i-1}) = \{a_{n-i-1}\}$: By B1 a_{n-i-1} is not equivalent to any of $a_{i+2}, \ldots a_{n/2-1}, a_{n/2+1}, \ldots a_{n-i-2}$ because it has a child in $\{a_{i+1}\}$ and the others do not. Therefore the n sets in N_n are distinct by the AFA, and the first part of the theorem is proved.

The second part, the converse, is that an n-size promise system must be A_n or N_n. It will be assumed from now on that n is odd; proving the even case is parallel. As before, the induction proceeds within a promise system, moving from the extreme members to the middle ones. In contrast to what one might expect in an induction proof, each step does not define a further set in A_n or N_n since in the NWF framework none of these except $\varnothing$ are defined until all are defined. Instead we introduce n set-valued variables and each step produces an equation relating them.

It is convenient to treat $n = 1$ separately. There are two NWF sets of size 1, $p_0 = \varnothing$ which yields promise system N_1, and $p_1 = \{p_1\} = \Omega$ which yields A_1 (Figure 5). It will be assumed henceforth that $n \geq 3$.

An equation *assigns* variable $p \in P_n$ *within* P_n when p appears alone on the left side and a subset of P_n appears on the right. Let $P_n = \{p_0, \ldots p_{n/2-3/2}, p_\bullet, p_{n/2+3/2}, \ldots p_n\}$ comprise n variables that take values in V. Define $n - 1$ equations:

E_0 : $p_0 = \varnothing$,
E_i : $p_i = \{p_{n-i+1}, \ldots p_n\}$ for $i = 1$ to $n/2 - 3/2$;
$p_i = P_n - \{p_0, \ldots p_{n-i-1}\}$ for $i = n/2 + 3/2$ to $n - 1$;
E_n : $p_n = P_n$.

After these equations are derived, an n'th one will assign $p_\bullet$ and the solution lemma will be applied.

The induction steps go from $i = 0$ to $i = n/2 - 3/2$. Sets W_i and S_i include those variables that have been assigned by the start of step i. The notation is meant to suggest "weak" and "strong", the latter sets having members that are more numerous that contain themselves. The U_i are the variables still unassigned by the start of step i. With $W_0 = \varnothing$, $S_0 = \varnothing$, define, for $i = 1$ to $n/2 - 3/2$,

$W_i = \{p_0, \ldots p_{i-1}\}$;
$S_i = \{p_{n-i+1}, \ldots p_n\}$;
$U_i = P_n - W_i - S_i$.

Each induction step, $i = 0$ to $n/2 - 3/2$, involves four propositions. The first two specify the children of the two new variables, and the second two specify their parents.

A_i : there exists a unique $p_i \in U_i$ such that $u \not\in p_i$ for all $u \in U_i$;
B_i : there exists a unique $p_{n-i} \in U_i$ such that $u \in p_{n-i}$ for all $u \in U_i$;
C_i : $p_{n-i} \in u$ for all $u \in U_i - \{p_i, p_{n-i}\}$;
D_i : $p_i \not\in u$ for all $u \in U_i - \{p_i, p_{n-i}\}$.

To establish the base cases: A_0 states that $\varnothing \in P_n$. If it were otherwise, Ω would decorate all the members in violation of distinctness. Uniqueness is immediate, and defining p_0 as $\varnothing$ gives E_0.

Base case B_0 states that $P_n \in P_n$. Some set in P_n must include p_0, otherwise either some member of $P_n - \{p_0\}$ is $\varnothing$, or more than one member is Ω, either possibility violating distinctness. By closure any set containing p_0 contains all members of P_n. (In the closure schema take x as p_0, y as any set in P_n, and z as any set containing p_0.) The set containing p_0 is therefore unique, and designating it p_n gives E_n.

Next C_0 states that $p_0 \not\in u$ for all $u \in P_n - \{p_0, p_n\}$. As shown for B_0, such a set containing $\varnothing$ would duplicate p_n.

Finally D_0 states that $p_n \in u$ for all $u \in P_n - \{p_0, p_n\}$. Every $u \in P_n - \{p_0, p_n\}$ is non-empty as none should duplicate p_0, and by closure this fact implies $p_n \in u$. (Take x as any member of u, y as p_n, and z as u.)

For the induction steps, $i = 1$ to $n/2 - 3/2$: First we assume that propositions A_j, B_j, C_j, and D_j hold for $j = 0$ to $i - 1$ and deduce A_i, i.e., the existence of a unique $p_i \in U_i$ such that $u \not\in p_i$ for all $u \in U_i$, and also deduce E_i. Suppose the contrary, that for every $u \in U_i$, there exists $u' \in U_i$ such that $u' \in u$. By C_{i-1} and D_{i-1} all members of U_i could be decorated either by $p = \{p_{n-i}, \ldots p_n\}$ or by $p' = \{p', p_{n-i}, \ldots p_n\}$, violating distinctness. Propositions A_i, C_j and D_j for $j = 0$ to $i - 1$ show that the membership of p_i is uniquely determined and follows E_i.

Next, assume A_j holds for $j = 0$ to i, and that B_j, C_j, D_j hold or $j = 0$ to $i - 1$. Then we deduce B_i, that there is a unique $p_{n-i} \in U_i$ such that $u \in p_{n-i}$ for all $u \in U_i$, and we also deduce E_{n-i}. Note that $p_i \in u$ for some $u \in U_i$, since otherwise all sets in $U_i - \{p_i\}$ could be decorated by either $p = \{p_{n-i}, \ldots p_n\}$ or $p' = \{p', p_{n-i}, \ldots p_n\}$, violating distinctness. By closure $(S_i \cup U_i) \subset u$, and by D_0 to D_{i-1}, $W_{i-1} \cap u = \varnothing$. This specifies p_{n-i}'s membership, and it follows E_{n-i}.

Next, assume A_j and B_j hold for $j = 0$ to i, and C_j and D_j hold for $j = 0$ to $i - 1$ and deduce C_i, i.e., that $p_{n-i} \in u$ for all $u \in U_i - \{p_i, p_{n-i}\}$. For all $u \in U_i$, either $p_{n-i} \in u$ or there exist $u_1, \ldots u_k \in U_0$, with $k \geq 1$, such that $p_n \in u_1 \in u_2 \in \ldots \in u_k \in u$, otherwise the system would violate distinctness since every $u \in U_i$ would be either $u = \{p_{n-i}, \ldots p_n\}$ or $u = \{u, p_{n-i}, \ldots p_n\}$. Applying closure to p_{n-i}, u_1 and u_2 yields $p_{n-i} \in u_2$. (In the schema take x as u_1, y as p_{n-i}, and z as u_2). Repeating the argument a finite number of times gives $p_{n-i} \in u$.

Finally assume that A_j, B_j, and C_j hold for $j = 0$ to i and D_j holds for $j = 0$ to $i - 1$, in order to deduce D_i, that $p_i \not\in u$ for all $u \in U_i - \{p_i, p_{n-1}\}$. Suppose that $p_i \in u$ for some $u \in U_i - \{p_i, p_{n-1}\}$. As shown previously, closure implies that u must include all sets in U_i, so it would duplicate p_{n-i} contrary to distinctness.

For i higher than $n/2 - 3/2$ the distinctness arguments used above fail because they require at least two and sometimes three unassigned variables. That leaves one variable $p_\bullet$ unassigned. We know from the C_j and D_j propositions that $p_\bullet$ is disjoint with $\{p_0, \ldots p_{n/2-3/2}\}$ and a superset of $\{p_{n/2+3/2}, \ldots p_n\}$ so the only question is whether it contains itself. If it does not, the n equations reproduce the definition of N_n, and otherwise they give A_n.

The theorem's claim that N_n has no agreement promise and A_n has the one described follow immediately from the two systems' definitions. □