Efficiency in face of externalities when binding hierarchical agreements are possible

Jobst Heitzig

Potsdam Institute for Climate Impact Research
Transdisciplinary Concepts and Methods
P. O. Box 60 12 03, 14412 Potsdam, Germany
heitzig@pik-potsdam.de

PIK-INTERNAL DRAFT — NOT FOR CIRCULATION!

February 17, 2011

Abstract

A formal framework for the treatment of hierarchical coalition formation and hierarchical agreements under both the bargaining and blocking approaches to coalition formation is introduced, and some first positive results on the possibility of full agreement and the efficiency of hierarchical agreements in face of externalities are given. In particular, it is shown that the possibility of hierarchical agreements can lead to efficient outcomes in the standard Cournot oligopoly example and a certain public good example that can be seen as being relevant in the study of International Environmental Agreements.

1 Introduction

1.1 Motivation

The game-theoretic literature on coalition formation contains many hints that in face of externalities, outcomes may be inefficient even when binding agreements are possible and agents are farsighted and have sufficient information [6, 7, 4, 5]. Often, this is due to the fact that a typical model of coalition formation predicts the formation of several disjoint coalitions or one coalition other than the grand coalition which are then expected to cooperate no further, so that the outcome is assumed to be a Nash equilibrium in a non-cooperative game in which the formed coalitions are the players.

These findings are in sharp contrast to what the Coase "theorem" [1] predicts, claiming that under the above circumstances agents will find a way of

reaching an efficient outcome. The purpose of this paper is to analyse the possibility of efficient outcomes when coalitions can reach additional binding agreements with each other, i.e., when agreements can be *hierarchical,* allowing them to avoid the non-cooperative and usually inefficient Nash equilibrium outcome. When players are assumed to be farsighted and hierarchical agreements are possible, they must take into account this possibility from the beginning on, not only anticipating the forming and splitting of coalitions that leads to a partition of the players into a "coalition structure", but also anticipating the forming of additional coalitions of higher level, leading to a whole coalition hierarchy.

With respect to the important application of International Environmental Agreements (IEAs), much of the existing literature on these does not assume either form of farsightednes, and the often pessimistic assessment of the possibilities for efficient IEAs might well be influenced by this shortcoming. I will present here some hints that when hierarchical agreements are possible, efficient IEAs might be much more likely than thought.

1.2 Outline

After introducing our main concepts, we proceed roughly analogous to Ray's monograph [5]. We first study bargaining models in the spirit of Rubinstein [8] in which agents, following some protocol, sequentially make and respond to proposals to form a coalition. Then we study blocking models in the spirit of the concept of the core, analysing the stability of coalition structures against blocking or deviations by groups of agents. For each of these two approaches, we first consider the formally simpler case in which agreements are assumed to be irreversible, and then the more general case in which any agreement can be terminated if all signatories agree.

Our models of *hierarchical coalition formation* are mostly straightforward modifications of common models of (non-hierarchical) coalition formation, usually treating coalitions already formed as players in subsequent negotiations. We find that not only do hierarchical agreements often lead to an efficient outcome but their sheer possibility sometimes even enables the *immediate* formation of the grand coalition without the need to actually use that possibility. In the special case of fully symmetric partition functions, payoffs are then not only efficient but also symmetric.

For the example of a symmetric Cournot oligopoly with a linear demand curve, Ray and Vohra [6, 7] find that both bargaining and blocking models predict the formation of a small number of coalitions (cartels) which despite the symmetry of the game must be of different size, so that payoffs are both inefficient and asymmetric. In our models, we will see that the grand coalition forms either immediately and payoffs are efficient and symmetric, or forms eventually and payoffs are efficient and asymmetric, but symmetric in expectation.

The same holds for the structurally equivalent example of a public good whose production costs are determined by an efficient market with linear marginal costs, a result that might have important implications in the provision of environmental public goods such as greenhouse gas emission reductions.

$\boldsymbol{a}$	agreement hierarchy
a_C	agreement between the subcoalitions of coalition C
$a_C(\nu, x)$	agreed payoff for ν if joint payoffs for C are x
$\boldsymbol{a}(\nu, \eta)$	expected payoff for ν given
	final coalition hierarchy η and agreements hierarchy $\boldsymbol{a}$
$A(x, y)$	set of negotiators affected by move $x \to y$
$C, C', \ldots$	coalitions (groups of individuals)
Γ	underlying non-cooperative game
Γ_η	derived non-cooperative game between final negotiators
η	coalition hierarchy
I	set of individuals (players of Γ)
$i, j, k, \ldots$	individuals
m	$\ldots$
$N(\eta)$	set of negotiators (players of Γ_η)
$\nu, \nu', \ldots$	negotiators (representing individuals or coalitions)
n, n_η	number of individuals and negotiators
p	process of hierarchical coalition formation
ϱ	bargaining protocol
$S(C)$	subcoalitions of C (signatories to an agreement a_C)
$\sigma_i \in \Sigma_i$	individual strategy
$\sigma_\nu \in \Sigma_\nu$	coalitional strategy
$\boldsymbol{\sigma} \in \Sigma$	strategy vector
$u_i(\boldsymbol{\sigma}), u_\nu(\boldsymbol{\sigma})$	individual and joint payoff
v	partition function
$v(\nu, \eta)$	expected joint payoff of ν given η
x, y	states in ongoing coalition formation
ξ	discounted infinite-horizon payoffs

Table 1: List of symbols used in this paper

These first positive results suggest that the possibility of hierarchical agreements is relevant in both the bargaining and blocking approaches, and in both the irreversible and reversible cases, and the framework presented here might be a valuable starting point for more detailed analyses in future research.

1.3 Preliminaries

1.3.1 Underlying non-cooperative game

We assume that a finite set $I = \{1, \ldots, n\}$ of *individuals* faces a one-shot non-cooperative game Γ with inefficient Nash equilibria, so that it is potentially profitable to cooperate. Γ has transferable utility, complete information, and is given in strategic form with a non-empty set Σ_i of *strategies* of individual $i \in I$ and *payoff functions* u_i mapping *strategy vectors* $\boldsymbol{\sigma} \in \Sigma = \prod_{i \in I} \Sigma_i$ to real-valued individual payoffs $u_i(\sigma) \in \mathbb{R}$. It is unimportant here whether Σ_i consists of pure or mixed strategies. Note that we use the term "individuals"

instead of "players" or "agents" to avoid confusion later on, and use letters I and m for individuals since we need the more common letters N and n below for negotiators.

Before playing Γ, individuals can negotiate either freely or following a certain protocol to be specified later. During negotiations, groups of individuals can sign binding agreements about how to share payoffs in Γ, and we assume that any agreement includes the provision to maximize joint expected payoffs. (If individuals are risk-averse, payoffs can be replaced by utilities in a suitable way.)

1.3.2 Negotiators and coalition hierarchies

A group $C \subseteq I$ of individuals that has signed an agreement is called a *coalition.* The main difference to common models of coalition formation is that we assume that a coalition can act like a new individual and negotiate further agreements with other individuals or coalitions, forming larger coalitions. Still, we require coalitions to be non-overlapping, so members of a coalition cannot sign additional agreements with outsiders individually. To avoid confusion, we therefore call a participant ν of negotiations a *negotiators,* whether ν is an individual or an already formed coalition that has not yet signed any (additional) agreement with other negotiators.

At any time during negotiations, we describe the current *coalition hierarchy* as a non-empty set η of *coalitions,* i.e., of non-empty subsets of N, fulfilling certain conditions: (i) For each $i \in I$, the singleton $\{i\}$ belongs to η. (ii) If $C, T \in \eta$, either C and T are disjoint, or one contains the other. The initial hierarchy is the smallest possible hierarchy η_0, consisting of all singletons. The set of *negotiators at* η consists of its maximal elements:

$$N(\eta) = \{C \in \eta : T \supset C \text{ for no } T \in \eta\}. \tag{1}$$

Note that, formally, $N(\eta)$ is a partition of the individuals into $n_\eta = |N(\eta)|$ (maximal) coalitions, which is usually called a "coalition structure" and denoted by π in the literature. We use the letter N to emphasize that its members can be interpreted as players in an ongoing negotiation game. If $n_\eta = 1$, we say there is *full agreement.*

1.3.3 Partition functions and subcoalitions

To evaluate the prospects of all individuals should negotiations end with a hierarchy η, we have to determine the expected outcome when the final negotiators become the players of the resulting non-cooperative game. For this, let Γ_η be the non-cooperative game induced by Γ in which the players are the negotiators $\nu \in N(\eta)$, the strategy set Σ_ν for ν contains combinations σ_ν of strategies $\sigma_i \in \Sigma_i$ for each $i \in \nu$, and payoffs are joint payoffs, $u_\nu(\boldsymbol{\sigma}) = \sum_{i\in\nu} u_i(\boldsymbol{\sigma})$. We use the symbol $\boldsymbol{\sigma}$ for both a strategy vector in Γ_η and the induced strategy vector in Γ. If suitable, one can also allow ν to use correlated strategies, i.e., probability distributions over combinations of strategies $\sigma_i \in \Sigma_i$ for each $i \in \nu$, in which case $u_\nu(\boldsymbol{\sigma})$ denotes the expected joint payoff of ν. Let us assume that

all individuals have common beliefs as to what the expected payoffs will be when Γ_η is played, and denote these expected payoffs by $v(\nu,\eta)$. If Γ_η possesses a unique Nash equilibrium $\boldsymbol{\sigma}$, usually $v(\nu,\eta) = u_\nu(\boldsymbol{\sigma})$. This function v is called the *(transferable utility) partition function* of Γ and builds the basis of all our analyses. If there are ν, η, η' such that $v(\nu,\eta) \neq v(\nu,\eta')$, we say that the game has *externalities,* and we assume that this is indeed the case in general. As the grand coalition I can achieve any possible expected payoff in Γ, v will be *grand-coalition superadditive,* i.e.,

$$V(\{I\}) \geqslant V(\eta) \tag{2}$$

for all η, where $V(\eta) = \sum_{\nu\in N(\eta)} v(\nu,\eta)$ is the *total expected payoff with* η. If even $V(\{I\}) > V(\eta)$ for all $\eta \neq \{I\}$, we say that v *needs full agreement.*

Call v *symmetric* iff $v(\nu,\eta)$ only depends on ν's size $|\nu|$ (the number of individuals represented by that negotiator) and on the size distribution of all $\nu' \in N(\eta)$ (i.e., the information how many negotiators represent coalitions of size 1,2,3,...). Call v *fully symmetric* iff $v(\nu,\eta)$ only depends on n_η (the number of negotiators). Note that the underlying game Γ need not be symmetric for v to be fully symmetric, as the public good example in the next section shows.

Cournot oligopoly. In a Cournot oligopoly with m symmetric firms and a linear demand curve, Γ_η is symmetric and has a unique Nash equilibrium, so that v is fully symmetric, and $v(\nu,\eta) = 1/(n_\eta+1)^2$ up to a multiplicative constant, where η represents a hierarchical cartel structure and n_η is the number of competing top-level cartels in it.

If a set $S \subseteq N(\eta)$ of negotiators forms a new coalition C_S, the resulting hierarchy $\eta + S$ has a new member C_S, and the negotiators $\nu \in S$ are replaced by a single negotiator ν_S. Formally, since negotiators $\nu \in S$ are identified with coalitions, both this new negotiator ν_S and the newly formed coalition C_S are equal to the union of the coalitions in S:

$$\nu_S = C_S = \bigcup S, \tag{3}$$

$$\eta + S = \eta \cup \{C_S\}, \tag{4}$$

$$N(\eta + S) = (N(\eta) - S) \cup \{\nu_S\}. \tag{5}$$

If $v(\nu_S, \eta/S) \geqslant \sum_{\nu\in S} v(\nu,\eta)$ for all η and all $S \subseteq N(\eta)$, we say that v is *(fully) superadditive,* but in general this will not be the case (e.g., in the Cournot oligopoly).

Conversely, given a coalition hierarchy η and a coalition $C \in \eta$, the *subcoalitions* of C are the coalitions in η which are maximal proper subsets of C. Let $S(C) \subset \eta$ designate the set of all subcoalitions of C, and note that $C = \bigcup S(C) = \nu_{S(C)}$.

1.3.4 Agreement hierarchies and preferences

Since the game has externalities and agreements can be hierarchical, a coalition C will usually not know what their expected joint payoff is at the time they reach an agreement. Hence agreements must specify payoff allocation rules rather than actual payoffs. Formally, an *agreement between the subcoalitions of* C will be treated as a function a_C that maps each value $x \in \mathbb{R}$ to an *allocation* $a_C(\cdot, x)$ of *signatory expected payoffs* $a_C(\nu, x)$ for all subcoalitions $\nu \in S(C)$, so that each $a_C(\nu, x)$ is non-decreasing in x and

$$\sum_{\nu \in S(C)} a_C(\nu, x) = x. \tag{6}$$

The interpretation is that the subcoalitions of C agree that, once the final hierarchy and the expected joint payoff x that C can expect from it are known, they will use a strategy vector in Γ for which the expected joint payoff of each $\nu \in S(C)$ is $a_C(\nu, x)$. Since Γ has transferable utility, x can indeed be redistributed in any way the individuals in C agree on, by chosing a suitable strategy from Σ_C. Because $a_C(\nu, x)$ is non-decreasing in x, maximizing a signatory's expected individual payoff $a_C(\nu, x)$ is equivalent to maximizing the coalition's expected joint payoff x once the agreement is signed.

Often, the agreement will specify that the payoff difference to some offset value is allocated in some agreed proportions, in which case a_C will be of the form $a_C(\nu, x) = \alpha_C(\nu) + (x - A_C)w_C(\nu)/W_C$ where $\alpha_C(\nu) \in \mathbb{R}$, $w_C(\nu) \geqslant 0$, $\sum_{\nu \in C} \alpha_C(\nu) = A_C$, and $\sum_{\nu \in C} w_C(\nu) = W_C$. E.g., the weights $w_C(\nu)$ could be all equal, or of the form $w_C(\nu) = \sum_{i \in C} w_i$ with individual weights w_i.

An *agreement hierarchy* for η is now a vector $\boldsymbol{a}$ of agreements a_C, one for each $C \in \eta$. Given an agreement hierarchy for η, a set $S \subseteq N(\eta)$ of negotiators, and a potential agreement a_C for $C = \nu_S \in \eta/S$, and a potential expected payoff x for C, one can recursively calculate the resulting expected payoff $a_C(\nu, x)$ of any coalition $\nu \in \eta$ that is part of C, $\nu \subset C$: If $a_C(\nu', x)$ is already calculated and $\nu \in \nu'$, we have

$$a_C(\nu, x) = a_{\nu'}(\nu, a_C(\nu', x)). \tag{7}$$

In particular, each potential agreement a_C induces an allocation rule for expected individual payoffs, $a_C(i, x) = a_C(\{i\}, x)$.

When negotiations end with coalition hierarchy η and agreement hierarchy $\boldsymbol{a}$ and $C \in N(\eta)$ is a maximal coalition in η, x will turn out to be $v(C, \eta)$ and each $\nu \subset C$ has an expected payoff of

$$\boldsymbol{a}(\nu, \eta) = a_C(\nu, v(C, \eta)). \tag{8}$$

Hence we say that ν *prefers* the potentially resulting hierarchy $(\eta, \boldsymbol{a})$ to the potentially resulting hierarchy $(\eta', \boldsymbol{a}')$ iff $\boldsymbol{a}(\nu, \eta) > \boldsymbol{a}'(\nu, \eta')$.

2 Public goods with efficient markets

An interesting example that turns out to be equivalent to the Cournot oligopoly in a special case is that of a public good whose production can be bought on an efficient market. Each individual $i \in I$ chooses to produce a non-negative amount q_i of the public good, leading to total production $Q = \sum_{i \in I} q_i$, individual benefits $f_i(Q) \geqslant 0$ and total costs $g(Q) \geqslant 0$ that are shared proportionally, giving individual costs $c_i(Q, q_i) = g(Q)q_i/Q$. Hence the sets of strategies in Γ_η are $\Sigma_\nu = [0, \infty)$ and the payoff functions are

$$u_\nu(\boldsymbol{q}) = f_\nu(Q) - g(Q)q_\nu/Q, \tag{9}$$

where $f_\nu = \sum_{i \in \nu} f_i$ and $q_\nu = \sum_{i \in \nu} q_i$. Note that our example differs from the public good examples often found in the literature, for which the payoff functions usually have the form $u_\nu(\boldsymbol{q}) = f_\nu(Q) - g_\nu(q_\nu)$ with costs that only depend on individual production and are independent from total production.

Assume that marginal costs are non-decreasing, marginal benefits are non-increasing, and denote average unit costs by $h(Q) = g(Q)/Q \geqslant 0$, with derivative $h'(Q) = (g'(Q) - h(Q))/Q \geqslant 0$. Then the unique pure-strategy equilibrium of Γ_η is given by

$$q_\nu = \frac{f'_\nu(Q_\eta) - h(Q_\eta)}{h'(Q_\eta)} \tag{10}$$

where Q_η is the unique solution of

$$f'(Q_\eta) = (n_\eta - 1)h(Q_\eta) + g'(Q_\eta) \tag{11}$$

and $f = \sum_{i \in I} f_i$. Note that Q_η only depends on η via the number of remaining negotiators, n_η. The resulting equilibrium payoffs are

$$v(\nu, \eta) = f_\nu(Q_\eta) + h(Q_\eta)\frac{h(Q_\eta) - f'_\nu(Q_\eta)}{h'(Q_\eta)}. \tag{12}$$

In the special case with linear benefits $f_i(Q) = \beta_i Q$ and quadratic costs $g(Q) = Q^2$, one can easily see that

$$v(\nu, \eta) = 1/(n_\eta + 1)^2 \tag{13}$$

up to a constant factor, just as in the Cournot oligopoly. Note that then v is fully symmetric even though the benefit factors β_i may differ! This is important since for fully symmetric v, we will see below that full agreement is probable when agreements can be hierarchical.

3 Bargaining models with hierarchical agreements

3.1 Irreversible agreements

3.1.1 Agreement in levels

One possibility to define a model of bargaining with hierarchical agreements is to apply an existing bargaining model without hierarchical agreements iteratively,

treating the coalitions formed in one round as players in the next round. We will see that under mild conditions such a process will end after finitely many rounds with full agreement. Let us start with the model of [7] that generalizes Rubinstein-Stahl bargaining and can be formulated in our framework as follows:

Simple bargaining protocol. Starting with a set N of negotiators, the protocol produces a partition $\pi(N)$ (a "coalition structure") of N and an agreement a_S for each $S \in \pi(N)$ by a process of successive proposals and responses. In our framework, we assume that $N = N(\eta)$ for some already established coalition hierarchy η, possibly the all-singletons hierarchy η_0. We start with an empty $\pi(N)$ and a full set of active negotiators $T = N$. In each step, a *proposer* ν from the set of active negotiators T proposes to a set of negotiators $S \subseteq T$ with $\nu \in S$ an agreement a_S, and then the other members of S sequentially respond by either accepting or rejecting the proposal until one negotiator rejects or all have accepted. If all have accepted, the set S is subtracted from the set of active negotiators T and becomes an element of the partition $\pi(N)$. If a negotiator j rejects, T and $\pi(N)$ remain unchanged, and a new proposer is selected after a *delay* of one time unit. The selection of a proposer and the order of responders is governed by (i) a probability $\varrho \in [0, 1]$ with which the rejector of a proposal gets to be the next proposer, (ii) a function ϱ^p that otherwise selects a proposer $\varrho^p(T)$ for each possible non-empty subset $T \subseteq N$ of active negotiators, and (iii) a function ϱ^r that selects a next responder $\varrho^r(S')$ for each possible non-empty subset $S' \subseteq N$ of remaining responders. Hence the initial proposer is $\varrho^p(N)$, the first responder is $\varrho^r(S')$ with $S' = S - \{\varrho^p(N)\}$, the second responder is $\varrho^r(S'')$ with $S'' = S - \{\varrho^p(N), \varrho^r(S')\}$, and so on. After a rejection by ν', the next proposer is ν' with probability ϱ and $\varrho^p(T)$ with probability $1 - \varrho$. The simple bargaining protocol ends when $T = \emptyset$ and $\pi(N)$ is a full partition of N, and does not end if from some step on all proposals get rejected.

The *simple bargaining game* consists in applying the simple bargaining protocol to $N = N(\eta_0)$ and results in the following payoffs: If there were k time units of delay, each coalition C_S with $S \in \pi$ gets a payoff of $v(C_S, \eta') \cdot \delta^k$, where $\delta \in (0, 1)$ is some common *discount factor,* η' is the coalition hierarchy η plus all newly formed coalitions, $\eta' = \eta \cup \{\nu_S : S \in \pi(N(\eta))\}$. Each C_S then distributes those payoffs according to its agreement a_S. If, on the other hand, from some step on all proposals get rejected, the game does not end and all individuals' payoffs are zero, where it is assumed that v is non-negative: $v(C, \eta) \geqslant 0$ for all C, η.

In [5] it is proved that under our assumption of transferable utility and with a mild additional condition "NAW", the simple bargaining game has always a perfect equilibrium consisting of stationary (Markovian) strategies where the only source of mixing is in the (possibly) probabilistic choice of a coalition by each proposer and where the game ends in finite time and with no delays.

Iterated bargaining game. Starting with $\eta = \eta_0$, repeatedly apply the simple bargaining protocol to find $\pi(N(\eta))$ and each time add the new coalitions to η, i.e., replace η by $\eta \cup \{\nu_S : S \in \pi(N(\eta))\}$. The iterated bargaining game ends as soon as no further coalitions form, i.e., as soon as $\pi(N(\eta))$ consists only of singletons. Then each $\nu \in N(\eta)$ gets a payoff of $v(\nu, \eta) \cdot \delta^k$ which is distributed according to the resulting agreement hierarchy $\boldsymbol{a}$, where k is the total number of delays.

Our first result shows that the sheer possibility of hierarchical agreements can lead to immediate full and fair agreement without actually using that possibility:

Theorem 1 *In the limit of fast negotiations (with vanishing delay costs, $\delta \to 1$), if v is fully symmetric and needs full agreement, and if a rejector always proposes next ($\varrho = 1$), then the iterated bargaining game will lead to the immediate formation of the grand coalition in the first round and to a fair agreement to split $V(\{I\})$ equally.*

Proof (sketch): We proceed inductively over the number n_η of remaining negotiators. For $n_\eta = 1$, the claim is trivial. Now assume $n_{\eta'} > 1$ and the claim has been proved for all η with $n_\eta < n_{\eta'}$. By symmetry, it is easy to see that a proposal by some proposer ν to any set $S \subseteq N(\eta')$ will be rejected if it promises some ν' less payoff than ν, since otherwise ν' could reject and then propose a similar proposal in which only the payoffs of ν and ν' are exchanged. Hence a negotiator will either propose a fair split to some set, or make an unacceptable proposal. If the initial proposer proposes an equal split of $V(\{I\})$ to the full set $N(\eta')$ and all responders accept this, he gets $V(\{I\})/n_{\eta'}$. If instead he proposes an equal split to a smaller set S, he will know by our induction assumption that like each $S' \in \pi(N(\eta'))$, S will finally get an equal share of $V(\{I\})$, i.e., S will get $V(\{I\})/|\pi(N(\eta'))|$ and he will get $V(\{I\})/|\pi(N(\eta'))||S|$. The latter is no larger than $V(\{I\})/n_{\eta'}$ because $n_{\eta'} \leqslant |\pi(N(\eta'))||S|$ if $S \neq N(\eta')$. Indeed, for $|S| > 1$, it is strictly smaller than $V(\{I\})/n_{\eta'}$ since the remaining negotiators will then all form singletons to get a payoff of $V(\{I\})/(n_{\eta'} - |S| + 1) > V(\{I\})/|\pi(N(\eta'))|$ each, which is the best they can do once S has formed. Hence proposing to a non-full non-singleton set S leads to strictly smaller payoffs than forming the grand coalition. Finally, we have to show that it does also not help to form a singleton in the hope that others will form a non-grand coalition. This is because unlike in the non-hierarchical case, the other negotiators will then simply also form singletons, after which all these singletons will come back to the table in the next round of the iterated bargaining game. Hence the best the initial proposer can do is to propose an equal split to the full set, and all will accept. QED.

Cournot oligopoly with $n = 5$. In the simple, non-iterated bargaining game, the initial proposer forms a singleton and the next proposer unites the remaining four by proposing an equal split, so that the first gets $1/(2+1)^2 = 1/9$ and each other gets $1/4 \cdot 1/(2+1)^2 = 1/36$, which is both asymmetric and inefficient.

In the iterated bargaining game, if the initial proposer forms a singleton, the others will just do the same, so that in the next round all five are back at the table. If the grand coalition agrees on an equal split, each gets $1/5 \cdot 1/(1+1)^2 = 1/20$. If the initial proposer would succeed in forming a pair, the other three will form singletons, so that in the next iteration the pair and the three singletons would get $1/(4+1)^2 = 1/25 < 1/20$ each. If the initial proposer would succeed in forming a triple, the other two will form singletons, so that in the next iteration the triple and the two singletons would get $1/(3+1)^2 = 1/16 > 1/20$ each, but the initial proposer would only get a third of the triple's share, which is $1/48 < 1/20$. Likewise, forming a four-player coalition would give the initial proposer only $1/4 \cdot 1/(2+1)^2 = 1/36$. So the best thing is to immediately propose a fair split to the grand coalition.

Because it has the same partition function v, the same holds for the public good example with linear benefits and quadratic costs.

Let us now turn to the non-symmetric case.

Theorem 2 *If v needs full agreement and condition "NAW" from [5] is fulfilled, the iterated bargaining game has a perfect equilibrium in which the game ends with full agreement in finite time and with no delays.*

Proof (sketch): Using a form of backward induction, we construct the equilibrium recursively over the number of remaining negotiators n_η and prove inductively that (i) each application of the simple bargaining protocol produces at least one new coalition, and (ii) the resulting expected payoffs $b(\nu, \eta)$ of all negotiators $\nu \in N(\eta)$ are efficient, $\sum_{\nu \in N(\eta)} b(\nu, \eta) = V(\{I\})$. For $n_\eta = 1$, the strategy of the only negotiator I trivially consists in proposing to himself the only possible agreement, leading to an efficient expected payoff of $b(I, \{I\}) = V(\{I\})$. Now assume that $n_{\eta'} > 1$ and for each η with $n_\eta < n_{\eta'}$, the equilibrium strategy vector has been constructed already, leading to efficient expected payoffs b_ν for all $\nu \in N(\eta)$. We construct the equilibrium strategy vector for η' as follows: Let v' be the partition function defined by $v'(\nu, \eta) = b(\nu, \eta)$ if $n_\eta < n_{\eta'}$, and $v'(\nu, \eta) = v(\nu, \eta)$ if $n_\eta \geqslant n_{\eta'}$. In other words, v' encodes the fact that when the current negotiators $N(\eta')$ build at least one more coalition, the remaining play along the equilibrium path will lead to the efficient payoffs $b(\nu, \eta)$. Now let $\boldsymbol{s}$ be a perfect equilibrium of the simple bargaining game with players $N(\eta')$ and partition function v', and $\pi(N(\eta'))$ the resulting partition of $N(\eta')$. By definition of v', this $\boldsymbol{s}$ is a perfect continuation equilibrium of the iterated bargaining game at the subgame starting at η'. We have to show that $\pi(N(\eta'))$ does not only consist of singletons.

Assume it does. Then the resulting expected payoffs are given by $v(\nu, \eta)$. Because v needs full agreement and $n_{\eta'} > 1$, these payoffs are not efficient, $V(\eta') = \sum_{\nu \in N(\eta')} v(\nu, \eta') < V(\{I\})$. Let $\varepsilon = (V(\{I\}) - V(\eta'))/n_{\eta'} > 0$, Then standard arguments show that any proposer $\nu \in N(\eta')$ could improve her payoff by proposing to the grand coalition $S = N(\eta')$ the agreement a_S with $a_S(\nu, x) = (v(\nu, \eta') + \varepsilon)x/V(\{I\})$ for all $\nu \in S$, since that agreement would be accepted by

all. So $\boldsymbol{s}$ would not be a perfect equilibrium of the simple bargaining game after all, a contradiction to our assumption. In other words, $\boldsymbol{s}$ produces at least one new coalition, $\pi(N(\eta'))$ does not only consist of singletons, and the resulting expected payoffs are efficient by definition of v'. QED.

3.1.2 Agglomerative coalition formation

Although conceptually simple, the above approach seems a little ad hoc since it is unclear why a newly formed coalition C_S only re-enters the negotiation process as a new negotiator ν_S after the remaining negotiators have finished the current round of bargaining. A different approach therefore does not iterate the simple bargaining protocol but rather modifies it so that after the negotiators S form a coalition and become inactive, a new negotiator ν_S immediately joins the set of active negotiators:

Agglomerative bargaining protocol. Starting with some already established coalition hierarchy η and agreement hierarchy $\boldsymbol{a}$, possibly the all-singletons hierarchy η_0, the protocol produces a coalition hierarchy $\eta' \supseteq \eta$ and a corresponding agreement hierarchy $\boldsymbol{a}'$ by a process of successive proposals and responses. We start with $(\eta', \boldsymbol{a}') = (\eta, \boldsymbol{a})$. The set of active negotiators always equals $N(\eta')$. In each step, a proposer ν from the set of active negotiators $N(\eta')$ proposes to a set of negotiators $S \subseteq N(\eta')$ with $\nu \in S$ an agreement a_S, and then the other members of S sequentially respond by either accepting or rejecting the proposal until one negotiator rejects or all have accepted. If all have accepted, the newly formed coalition C_S and the agreemens a_S are added to η' and $\boldsymbol{a}'$, hence the set S is subtracted from the set of active negotiators $N(\eta')$ and the new negotiator ν_S is added to it. If a negotiator ν' rejects, η' remains unchanged, and a new proposer is selected from $N(\eta')$ after a *delay* of one time unit. The selection of a proposer and the order of responders is governed by (i) a probability $\varrho \in [0, 1]$ with which the rejector of a proposal gets to be the next proposer, (ii) a function ϱ^p that otherwise selects a proposer $\varrho^p(T)$ for each possible non-empty set T of active negotiators, and (iii) a function ϱ^r that selects a next responder $\varrho^r(S')$ for each possible non-empty set S' of remaining responders. After a rejection by ν', the next proposer is ν' with probability ϱ and $\varrho^p(N(\eta'))$ with probability $1 - \varrho$. The simple bargaining protocol ends when there is full agreement in η', and does not end if from some step on all proposals get rejected.

The *agglomerative bargaining game* consists in applying the agglomerative bargaining protocol to $\eta = \eta_0$ and results in the following payoffs: If there were k time units of delay, each individual $i \in I$ gets a payoff of $\boldsymbol{a}'(\{i\}, \eta') \cdot \delta^k$. If, on the other hand, from some step on all proposals get rejected, the game does not end and all individuals' payoffs are zero, where it is assumed that v is non-negative: $v(C, \eta) \geqslant 0$ for all C, η.

Note that the process bears some obvious similarity to agglomerative clustering procedures used in statistics, and it might be interesting to further study this relationship, although those usually join only two clusters to get a new cluster, needing $n-1$ steps of cluster formation, while our procedure often uses only one step of coalition formation and directly forms the grand coalition. For example, this is so for fully symmetric v, just as it was the case with the iterated protocol:

Theorem 3 *In the limit of fast negotiations (with vanishing delay costs, $\delta \to 1$), if v is fully symmetric and needs full agreement, and if a rejector always proposes next ($\varrho = 1$), then the agglomerative bargaining game will lead to the immediate formation of the grand coalition in the first step and to a fair agreement to split $V(\{I\})$ equally.*

Proof (sketch): Very similar to theorem 1, we proceed inductively over the number n_η of remaining negotiators. For $n_\eta = 1$, the claim is trivial. Now assume $n_{\eta'} > 1$ and the claim has been proved for all η with $n_\eta < n_{\eta'}$. By symmetry, it is again easy to see that a proposal by some proposer ν to any set $S \subseteq N(\eta')$ will be rejected if it promises some ν' less payoff than ν, since otherwise ν' could reject and then propose a similar proposal in which only the payoffs of ν and ν' are exchanged. Hence a negotiator will either propose a fair split to some set, or make an unacceptable proposal. If the initial proposer ν proposes an equal split of $V(\{I\})$ to the full set $N(\eta')$ and all responders accept this, he gets $V(\{I\})/n_{\eta'}$. If instead he proposes an equal split to a smaller but non-singleton set S, he will know by our induction assumption that like each $\nu' \in N(\eta') - S$, S will finally get an equal share of $V(\{I\})$, i.e., S will get $V(\{I\})/(n_{\eta'} - |S| + 1)$ and he will get $V(\{I\})/(n_{\eta'} - |S| + 1)|S|$. The latter is smaller than $V(\{I\})/n_{\eta'}$ because $n_{\eta'} < (n_{\eta'} - |S| + 1)|S|$ if $S \neq \{\nu\}, N(\eta')$. Hence proposing to a non-full non-singleton set S again leads to strictly smaller payoffs than forming the grand coalition. Obviously, it does also not help to form a singleton since that leaves η' unchanged and only leads to a new proposer who will then propose an equal split to the full set. Hence the best the initial proposer can do is to propose an equal split to the full set, and all will accept. QED.

The non-symmetric case is much more complicated and probably requires techniques similar to those in [5], in particular the notion of *equilibrium response vectors*. Leaving a detailed study for future research, we conjecture that similar to the simple bargaining game, also the agglomerative bargaining game will have perfect equilibria in stationary (Markov) strategies.

We conclude this section by discussing a simple non-symmetric example without externalities in which hierarchical agreements enable efficient outcomes:

Simple example A special individual 1 can realize payoff 1 with the help of one of the other two individuals, or $1+\mu$ in the grand coalition, where $\mu < 1/2$. More precisely, $I = \{1,2,3\}$, $v(I,\eta) = 1+\mu$, $v(\{1,i\},\eta) = 1$ for $i \in \{2,3\}$

and all η, and $v(C,\eta) = 0$ for all other S. In [5], it is shown that with the simple bargaining protocol, $\varrho = 1$ and $\delta \to 1$, the result is either the partition $\{\{1,2\},\{3\}\}$ or the partition $\{\{1,3\},\{2\}\}$, with payoff vectors $(1/2, 1/2, 0)$ or $(1/2, 0, 1/2)$. This is because no initial proposer can make a proposal to the grand coalition in which she gets more than $1/2$, since then one responder would be promised less than $1/2$ and will thus reject and realize a payoff of $1/2$ by proposing an equal split to one other individual. So a pair must form first, after which the grand coalition can no longer form when hierarchical agreements are not possible. Hence payoffs are asymmetric and inefficient.

With hierarchical agreements, the payoffs will still be asymmetric but efficient. First study what happens when coalition $\{1,2\}$ has already formed. Then there are only two negotiators left, and because of $\delta \to 1$, they will share their surplus equally in equilibrium, so $\{1,2\}$ gets $1 + \mu/2$ and 3 gets $\mu/2$. Likewise, when coalition $\{2,3\}$ has already formed, it can expect to get $1/2 + \mu/2$ in the end, and 1 can expect the same. Now assume no pair has yet formed, and let m_i be the payoff individual i can expect when she is the proposer. Then in each acceptable proposal to some set S, she has to promise all other $j \in S$ at least m_j since otherwise j will reject and realize m_j as the next proposer. So the values m_i must fulfil the following relationships:

$$m_1 = \max(1 + \mu/2 - m_2, 1 + \mu - m_2 - m_3), \tag{14}$$

$$m_2 = \max(1 + \mu/2 - m_1, 1 + \mu - m_1 - m_2), \tag{15}$$

$$m_3 = \max(1 + \mu/2 - m_1, 1 + \mu - m_1 - m_3), \tag{16}$$

and, by symmetry, $m_2 = m_3$. The solution of this is $m_1 = 1/2 + \mu/2$ and $m_2 = m_3 = 1/2$. If individual 1 is the initial proposer, she proposes to either $\{1,2\}$ or $\{1,3\}$ a split in which she finally gets $1/2 + \mu/2$ of the $1 + \mu/2$ that coalition expects to get in the end after an additional agreement with the third player. If individual $i \neq 1$ is the initial proposer, she proposes to $\{1,i\}$ a split in which she finally gets $1/2$.

3.2 Reversible agreements

Up until now, we assumed that agreements can be complemented with additional agreements but are irreversible in that they cannot be reverted once established. This makes it possible to analyse the process of coalition formation using some kind of backward induction that is no longer available when we allow agreements to be terminated.

Let us now assume that a top-level agreement $a_{S(C)}$ with $C \in N(\eta)$ can be terminated if all its signatories in $S(C)$ agree to do so, and define the following bargaining protocol similar to [3]:

Ongoing hierarchical bargaining. A *state* is a pair $x = (\eta, \boldsymbol{a})$ where η is a coalition hierarchy and $\boldsymbol{a}$ is a agreement hierarchy for η. Each of infinitely many periods $t = 0, 1, \ldots$ begins in some *going state* $x = (\eta, \boldsymbol{a})$, and each individual $i \in I$ gets a period payoff of $\boldsymbol{a}(\{i\}, \eta)$. Then a proposer $\nu \in N(\eta)$

proposes a move $x \to y$ to a new state $y = (\eta', \boldsymbol{a}')$. Define the set of *affected negotiators* $A(x, y)$ as follows: If $x \to y$ is an *elementary move* that either (i) leaves all existing coalitions and agreements unchanged and only adds one new coalition C with $S(C) \in N(\eta)$ and an agreement $a_{S(C)}$, or (ii) leaves the coalition hierarchy unchanged and only replaces some individual agreement $a_{S(C)}$ with $C \in \eta$ by a new agreement $a'_{S(C)}$, leaving all other agreements unchanged, or (iii) leaves all existing coalitions and agreements unchanged except that it removes one top-level coalition $C \in N(\eta)$ and its agreement $a_{S(C)}$, then the affected negotiators are the signatories to that agreement: $A(x, y) = S(C)$. If $x \to y$ is a more complicated move that changes more of η and $\boldsymbol{a}$, it can be represented as a unique minimal concatenation $x = x_0 \to x_1 \to \cdots x_{k-1} \to x_k = y$ of partial moves $x_{i-1} \to x_i$ of the above types (i)–(iii), and we put $A(x, y) = \bigcup_{i=1}^{k} A(x_{i-1}, x_i)$, i.e., each negotiator affected by a partial move is affected by the whole move.

After the proposal is made, each affected negotiator $\nu' \in A(x, y)$ responds by either accepting or rejecting the proposal, in an order governed by a function ϱ^r as before, until one negotiator rejects or all have accepted. If the proposal was accepted, the going state in period $t + 1$ is $(\eta', \boldsymbol{a}')$, otherwise it is $(\eta, \boldsymbol{a})$.

If one assumes that individuals use exponential discounting to derive discounted infinite-horizon utilities from period payoffs, one can then search for perfect equilibria in history-dependent or stationary (Markov) strategies.

A first general result is this:

Theorem 4 *Assume that v needs full agreement and negotiators follow some perfect equilibrium in stationary (Markov) strategies. Let $\xi_i(x)$ be the discounted infinite-horizon utility of $i \in I$ when the initial state is $x = (\eta, \boldsymbol{a})$, and put $\xi_\nu(x) = \sum_{i \in \nu} \xi_i(x)$ and $\Xi(x) = \sum_{i \in I} \xi_i(x)$. Then:*

(1) If the initial state has full agreement, then payoffs will never change.

(2) Assume that the initial state x does not have full agreement and the proposer is always selected independently from who rejected and who accepted a proposal. Then the initial proposer ν will make a proposal to change to a state y with $\xi_\nu(y) \geqslant \xi_\nu(x) + (V(\{I\}) - \Xi(x))/|n_\eta|$, and that proposal will be accepted.

Proof (sketch): (1) Any state y has $\Xi(y) \leqslant V(\{I\})$. So if a move involving a changed payoff vector is proposed, at least one $\nu \in N(\eta)$ faces a loss in discounted payoffs and will reject.

(2) Let $S = N(\eta)$ and define a_S so that by $a_S(\nu', V(\{I\})) = \xi_{\nu'}(x) + (V(\{I\}) - \Xi(x))/|n_\eta|$ for all $\nu' \in S$. Then if ν proposes a_S to S, each $\nu' \in S$ will accept since that guarantees her because of (1) a discounted payoff of $a_S(\nu', V(\{I\}))$ that is strictly larger than her discounted payoff of $\xi_{\nu'}(x)$ which she expects upon rejection. So ν can get at least $a_S(\nu, V(\{I\}))$ by making this proposal, or even more by making a different proposal that gets accepted. QED.

Note, however, that this does not imply immediate or even eventual full agreement since a proposer might find an even better proposal to a smaller

coalition, as it is the case in the simple asymmetric 3-individuals example at the end of Section 3.1.2. I conjecture, however, that under quite general conditions, it will be possible to prove that full agreement must emerge eventually in the above model.

4 Blocking models with hierarchical agreements

Unlike the bargaining models of the previous section, the blocking approach avoids the specification of protocols and rather studies the stability of agreements against possible deviations by individuals or groups of individuals. The basic idea is that a group S of individuals may leave a coalition C if they expect to be better off in the setting that eventually arises from this. In the definition of the classical concept of the *core,* it is assumed that after S has left, the remaining individuals $C - S$ stay together, but recent *farsighted* models [6, 4, 2] assume that the latter may split further as a reaction, or other coalitions might form.

4.1 Irreversible agreements

We start again with the assumption that an agreement is not only binding but also irreversible once it has been signed, so that we can hope negotiations will end in finite time and the process can be analysed by backwards induction. Before describing a specific model, let us generalize the idea of agreement in levels from Section 3.1.1 so that it works with many existing procedures of non-hierarchical coalition formation.

Assume we have a procedure μ of non-hierarchical coalition formation that maps a partition function v to a probability distribution of coalition structures and payoff vectors. More precisely, for each set of negotiators N, let $p_0(v, \pi) \in [0, 1]$ be the probability that the model results in a partition π of N if it is applied to a partition function v for N that maps partitions π of N and coalitions $S \in \pi$ to payoffs $v(S, \pi)$. Moreover, let $\boldsymbol{e}(v, \pi) \in I\!R^N$ with $\sum_{\nu \in S} e_\nu(v, \pi) = v(S, \pi)$ for each $S \in \pi$ be the expected value of the payoff vector that results when the procedure results in the partition π. If the procedure μ could be a specific bargaining protocol, p and $\boldsymbol{e}$ can be derived from the protocol directly. If μ refers to some set-valued solution from the blocking approach, e.g., the core, p and $\boldsymbol{e}$ can be interpreted as representing the common beliefs of all negotiators about what particular member of the set-valued solution will arise with what probability under suitable assumptions of rationality.

Call a partition function v for N *non-singletons efficient* iff

$$\sum_{S \in \pi} v(S, \pi) = v(N, \{N\}), \tag{17}$$

$$\sum_{\nu \in N} v(\{\nu\}, \pi_0) < v(N, \{N\}) \tag{18}$$

for all partitions π containing a non-singleton coalition, where π_0 is the all-singletons partition.

Now assume that μ has the following property (*): If v is non-singletons efficient, μ results in a non-singletons partition with positive probability, $1 - p_0(v, \pi_0) > 0$. We will argue below that many common procedures μ indeed fulfil (*). From such a procedure of non-hierarchical coalition formation μ, we can now construct a procedure of hierarchical coalition formation in levels that will end with full cooperation after finitely many applications:

Iterative application of μ. Given a partition function v that needs full agreement and an already established coalition hierarchy η, the iterative application of μ to η is defined recursively over the number of negotiators n_η. For each larger coalition hierarchy $\eta' \supseteq \eta$, denote the probability that the result will be η' by $q(\eta, \eta') \in [0, 1]$, and let $\xi_\nu(\eta, \eta')$ denote the expected final payoff of $\nu \in N(\eta)$ if the result is η' with $q(\eta, \eta') > 0$. We derive $q(\eta, \eta')$ and $\xi_\nu(\eta, \eta')$ using backwards induction as follows, and prove at the same time inductively that $q(\eta, \eta') = 0$ if η' has no full agreement: If $n_\eta = 1$, η has already full agreement, no further application of μ is needed, and we have $q(\eta, \eta) = 1$, $\xi_\nu(I, \eta) = V(\{I\})$, and $q(\eta, \eta') = 0$ for all $\eta' \neq \eta$. For $n_\eta > 1$, define a non-singletons efficient partition function v_η on $N(\eta)$ as follows: Let π_0 be the all-singletons partition of $N(\eta)$. For each negotiator $\nu \in N(\eta)$, let $v_\eta(\{\nu\}, \pi_0)$ be the non-cooperative payoff to ν if no further agreements are made and negotiations were to end with hierarchy η:

$$v_\eta(\{\nu\}, \pi_0) = v(\nu, \eta). \tag{19}$$

For each other partition $\pi \neq \pi_0$ of the set $N(\eta)$ of current negotiators, let $\eta + \pi$ be the coalition hierarchy that results if this round of negotiations ends with partition π,

$$\eta + \pi = \eta \cup \{C_S : S \in \pi\}, \tag{20}$$

and for each $S \in \pi$, let $v_\eta(S, \pi)$ be the expected payoff the new coalition C_S can get if after this round μ is applied iteratively to $\eta + \pi$:

$$v_\eta(S, \pi) = \sum_{\eta' \supseteq \eta + \pi} q(\eta + \pi, \eta') \xi_{\nu_S}(\eta + \pi, \eta'). \tag{21}$$

Since v needs full agreement, $\sum_{\nu \in (\eta)} v(\{\nu\}, \pi_0) < V(\{I\}) = v(N(\eta), \{N(\eta)\})$. Since all η' with $q(\eta + \pi, \eta') > 0$ have full agreement, $\sum_{S \in \pi} v_\eta(S, \pi) = V(\{I\}) = v(N(\eta), \{N(\eta)\})$. Hence v_η is non-singletons efficient. Now apply the procedure μ of non-hierarchical coalition formation to the set $N(\eta)$ of negotiators and the non-singletons efficient partition function v_η, leading to probabilities $p_0(v_\eta, \pi)$ and expected payoffs $\boldsymbol{e}(v_\eta, \pi)$ with $1 - p_0(v_\eta, \pi_0) > 0$. In other words, the application of μ will convert the coalition hierarchy η into the coalition hierarchy $\eta + \pi$ with probability $p_0(v_\eta, \pi)$. Although the result might be $\eta + \pi_0 = \eta$ with

positive probability, the repeated application of μ to the same η will eventually give a properly larger hierarchy $\eta + \pi \supset \eta$, with probabilities

$$p(v_\eta, \pi) = p_0(v_\eta, \pi)/(1 - p_0(v_\eta, \pi_0)) \tag{22}$$

and an expected finite number of iterations,

$$t_\eta = 1/(1 - p_0(v_\eta, \pi_0)). \tag{23}$$

The probability that the iterative application of μ to η will result in η' is now

$$q(\eta, \eta') = \sum_{\pi \neq \pi_0} p(v_\eta, \pi) q(\eta + \pi, \eta') \tag{24}$$

which is still zero unless η' has full agreement. The expected final payoff of $\nu \in N(\eta)$ if the result is η' is

$$\xi_\nu(\eta, \eta') = \sum_{\pi \neq \pi_0} p(v_\eta, \pi) e_\nu(v_\eta, \pi). \tag{25}$$

This completes the recursive derivation of q and ξ and the inductive proof that the procedure ends in finite time with full agreement and hence with efficient payoffs.

Typical farsighted blocking approaches fulfil condition (*): Assume v is non-singletons efficient but still the all-singletons structure π_0 will form with certainty, $p_0(v, \pi_0) = 1$. Since $\sum_{\nu \in N} v(\{\nu\}, \pi_0) < v(N, \{N\})$, there is an allocation of the grand coalition's worth $v(N, \{N\})$ into individual payoffs $a_\nu > v(\{\nu\}, \pi_0)$ for all $\nu \in N$. Then a proposal to realize a in the grand coalition cannot credibly be blocked by any group S of players, since if they leave, they must expect (because of farsightedness) that the all-singletons structure π_0 will form with certainty (as assumed) and they would get only $\sum_{\nu \in S} v(\{\nu\}, \pi_0) < \sum_{\nu \in S} a_\nu$. Hence if no other non-singletons partition has a positive probability of forming, the grand coalition must have a positive probability of forming and agreeing on an allocation such as a, contradicting the assumption that the all-singletons structure will form with certainty. Hence farsighted blocking models must fulfil (*) and will thus lead to efficient final payoffs when applied iteratively to produce hierarchical agreements.

Still, we have to motivate why it might be reasonable to expect that once a blocking set of negotiators S has left the table, they will later on re-enter negotiations, but not until a partition of the remaining negotiators has been established. To this end, let us invoke the following image: Let us assume all individuals share the following assumptions as to how negotiations will proceed: Negotiations take place in a sequence of rounds with a non-increasing number of negotiators, starting with one negotiator for each individual $i \in I$, and only ending when there is only one negotiator left. At each point in time

during a round, negotiators will be distributed over meeting rooms, starting with the grand coalition in one room, and this distribution can be described by a partition π. The current partition π is known to all negotiators at all times. Groups of negotiators can leave a room and move to a new room, but may not join already occupied rooms, hence π can only get finer during a round. (Most blocking approaches assume this, as it allows for a recursive analysis from finer to coarser partitions. In principle, however, we might also assume groups can join a coalition in another room if that coalition agrees to negotiate with them). Each "negotiating group" $S \in \pi$ is farsighted and estimates the total payoff $a_S(\pi)$ that they can expect to get in the end if the current partition π becomes the final coalition structure of this round, by looking forward to what agreements they can expect the remaining rounds to bring about. Each $S \in \pi$ will then discuss how to distribute this expected final payoff $e_S(\pi)$ among its members. At each point in time, the expected payoffs $a_\nu(\pi)$ of all $\nu \in S$ reflect the currently discussed payoff distribution, with $\sum_\nu a_\nu(\pi) = e_S(\pi)$. At each point in time, S will either be in discussion, in temporary agreement, or will split in two parts G and $S \setminus G$. Since time is essentially continuous, at most one coalition S splits at any point in time, so that all negotiators can adapt their expectations to the new situation before considering another split. As soon as in some round k all coalitions are in temporary agreement, the current partition π and the temporary agreements become the final coalition structure π_k and agreements of that round, each $S \in \pi_k$ will appoint a negotiator ν_S for the next round, and these meet again in the central room to proceed with round $k+1$. If there are externalities, no coalition S has an incentive to finalize their temporary agreement before they know the coalition structure, i.e., before all other coalitions are in temporary agreement as well. Hence all coalitions in round k can indeed be expected to form essentially at the same time, so that the assumption that negotiations take place in rounds is at least consistent.

Let us now revisit the important example of Cournot oligopolies:

Cournot oligopoly with $n = 5$. Without hierarchical agreements, a farsighted blocking approach gives this result: An allocation of the grand coalitions payoff of $1/4$ must promise at least one of the five firms, say 1, at most $1/20$. If 1 can hope that when it forms a singleton coalition, the other four will stay together and form a second coalition 2345, both coalitions would get $1/9 > 1/20$ each in the resulting non-cooperative Cournot-Nash equilibrium. If those four would split further in two pairs, say 23 and 45, then 1, 23, and 45 would each get $1/16$. Because $2 \cdot 1/16 > 1/9$, 2345 cannot distribute their joint payoff of $1/9$ to give each pair at least $1/16$. Hence such a split in two pairs would indeed happen after 1 has left. Given the partition 1,23,45, no further split is profitable for the deviator, so this partition would be considered stable by a farsighted blocking approach, as would any other partition into a singleton and two pairs. Similar arguments show that no coarser partition would be stable, hence one would expect that each of those 1+2+2 partitions would occur

with the same probability.

With the possibility of hierarchical agreements, the analysis is different since a deviating coalition must look forward to payoffs resulting from the whole hierarchical process. Assume that in some round, there are two negotiators left. Then the grand coalition between them is stable since it can give both $1/2 \cdot 1/4 = 1/8$, while each would only get 1/9 if they stay apart. So, by symmetry, each of the two can expect to get 1/8.

Now assume that in some round, there are three negotiators left. Then the grand coalition can expect to get 1/4, a partition into two coalitions can expect to get 1/8 each (as just argued), and a partition into three singletons can expect to get 1/16 each. So the only stable agreement in a partition into two coalitions gives the lone negotiator 1/8 and each member of the pair 1/16, since otherwise one of the latter would walk away. The grand coalition is not stable here since it cannot give each member at least 1/8. I.e., with three negotiators, the result will be one of the payoff vectors $(1/8, 1/16, 1/16)$, $(1/16, 1/8, 1/16)$, and $(1/16, 1/16, 1/8)$, which is efficient but asymmetric. Still, because of symmetry, it is sensible to assign equal probabilities to these three outcomes, so that each of the three negotiators can expect to get 1/12.

Now assume that in some round, there are four negotiators 1,2,3,4 left.

- The all-singletons partition 1,2,3,4 is stable and each gets 1/25.
- The partition 1,2,34 is stable: each coalition can expect 1/12, so the agreement to give both members $1/24 > 1/25$ is stable.
- The partition 1,234 is unstable: the triple 234 can expect 1/8, so at least one of them, say 2, gets at most $1/24 < 1/12$ and wants to leave in order to get 1/12 in the stable partition 1,2,34.
- The partition 12,34 is unstable: the pair 12 expects 1/8, at least one of them gets at most $1/16 < 1/12$ and wants to leave in order to get 1/12 in the stable partition 1,2,34.
- Finally, the grand coalition 1234 is unstable: at least one of them, say 1, gets at most $1/16 < 1/12$ and wants to leave in order to get 1/12 in one of the stable partitions 1,2,34 or 1,3,24 or 1,4,23, expecting that in the unstable intermediate partition 1,234, one of 1,2,3 will leave.

The expected result for four negotiators is thus a partition into a pair and two singletons, with asymmetric payoffs $1/24, 1/24, 1/12$, and 1/12. Still, expected payoffs are 1/16 for each negotiator because each such partition must be expected with equal probability.

Finally, we can now analyse the five-firm situation:

- The partition 1,2,3,4,5 is stable and each gets 1/36.
- The partition 1,2,3,45 is stable: each coalition gets 1/16, so the pair can give each member $1/32 > 1/36$.

- The partitions 1,2,345 and 1,23,45 are unstable: each coalition gets 1/12, so at least one member of 345 or 45 gets at most $1/24 < 1/16$ and wants to leave.
- The partitions 1,2345 and 12,345 are unstable: each coalition gets 1/8, so at least one member of 2345 or 345 gets at most $1/24 < 1/16$ and wants to leave.
- Finally, the grand coalition 12345 is unstable: at least one member gets at most 1/20 and wants to leave to get 1/16 in a partition with a pair and two other singletons.

The expected result is thus a partition into a pair and three singletons, each of which gets 1/16.

Summing up, our analysis predicts the following:

- In the first round a two-member coalition forms, say 45, leading to the hierarchy 1,2,3,45.
- In the second round, two of the four negotiators form a new coalition, say either 23, leading to the hierarchy 1,23,45, or 345, leading to the hierarchy 1,2,3(45).
- In the third round, again two of the now three remaining negotiators form a new coalition, so that the resulting hierarchy has one of the forms 1,2(3(45)) or 1,(23)(45) or 12,3(45) up to permutations.
- In the final fourth round, the two remaining negotiators form the grand coalition, so that the final hierarchy has one of the forms 1(2(3(45))) or 1((23)(45)) or (12)(3(45)) up to permutations. The corresponding payoff vectors are $(1/8, 1/16, 1/32, 1/64, 1/64)$ or $(1/8, 1/32, 1/32, 1/32, 1/32)$ or $(1/16, 1/16, 1/16, 1/32, 1/32)$.

In other words, payoffs are asymmetric as in the model without hierarchical agreements, but are efficient unlike in that model. Two of the possible hierarchies, 1((23)(45)) and (12)(3(45)), have the partition into a singleton and two pairs as an intermediate step after which two of these coalitions sign a further agreemen. The third possible hierarchy 1(2(3(45))), however, is built in contradiction to the non-hierarchical model since only one initial pair forms which then succesively collects the remaining firms one by one. This is possible because although in the non-hierarchical model the move from 1,2,3,45 to 1,2,345 would imply a loss of $1/25 - 1/48$ for firm 3, in the hierarchical model the move from 1,2,3,45 to 1,2,3(45) raises 3's expected final payoff from the non-cooperative outcome of 1/25 it will get if 1,2,3,45 is the final hierarchy, to the cooperative outcome of 1/24 it can expect if negotiations continue after 1,2,3(45) has formed.

Again, the same analysis holds for the public good example with linear benefits and quadratic costs since it has the same partition function v.

4.2 Reversible agreements

Similar to the model in [4], and using the terminology and notation of Section 3.2, we could define the following model:

Process of hierarchical coalition formation. A process of hierarchical coalition formation is a discrete-time stochastic process p on the set of states $x = (\eta, \boldsymbol{a})$ such that $p(x, y) = 0$ for all pairs (x, y) outside a set M of *feasible moves*. It leads to discounted infinite horizon payoffs ξ_i fulfilling $\xi_i(x) = (1 - \delta)\boldsymbol{a}(i, \eta) + \delta \int \xi_i(y)p(x, dy)$ for all $i \in I$ and $x = (\eta, \boldsymbol{a})$. We call p an *equilibrium process of hierarchical coalition formation (EPHCF)* iff for each pair of states $x = (\eta, \boldsymbol{a}), y = (\eta', \boldsymbol{a}')$, the following holds:

(i) If $p(x, y) > 0$ with $y \neq x$, the move must be *strictly profitable*, i.e., all affected negotiators must strictly profit from the move: $\xi_\nu(y) > \xi_\nu(x)$ for all $\nu \in A(x, y)$.

(ii) If $p(x, y) > 0$ with $y \neq x$, no strictly profitable move $x \to z \in M$ with $\bigcup A(x, z) \subseteq \bigcup A(x, y)$ must strongly Pareto-dominate $x \to y$, i.e., $\xi_\nu(y) > \xi_\nu(z)$ for some $\nu \in A(x, y)$, or $\xi_\nu(y) \geqslant \xi_\nu(z)$ for all $\nu \in A(x, y)$.

(iii) If there is a strictly profitable and not strongly Pareto-dominated move $x \to y \in M$, then $p(x, x) = 0$.

The rationale of (ii) is that the negotiators in $A(x, y)$ will not initiate a move to y when another set of negotiators $A(x, z)$ who represent a subset of the individuals represented by $A(x, y)$ can initiate a move to a state z that is even better for all of them.

At this point, it remains unclear whether an EPHCF must always exist when $x \to y$ is a feasible move for all possible pairs of states, and whether it will lead eventually to full agreement. For a more restrictive choice of feasible moves M, in which only one agreement can be changed at a time, one can however prove the following:

Theorem 5 *Assume v needs full agreement.*

(1) If adding an agreement to form the grand coalition is always a feasible move, then the following is an EPHCF: For states x with full agreement, put $p(x, x) = 1$. For each state $x = (\eta, \boldsymbol{a})$ without full agreement, select one state y that is the outcome of adding to x a strictly profitable agreement for the grand coalition, and put $p(x, y) = 1$.

(2) Let M be the set of elementary moves of type (i), (ii), or (iii) defined in Section 3.2, and p an EPHCF with feasible moves M. Then the moves $x \to y$ with $p(x, y) > 0$ and $x \neq y$ build an acyclic graph whose terminal nodes all have full agreement, and the process will end in a state with full agreement after at most $2n - 3$ steps.

Proof: (1) When x has full agreement, no strictly profitable move is possible, hence putting $p(x,x) = 0$ fulfils conditions (i)–(iii). When $x = (\eta, \boldsymbol{a})$ does not have full agreement, it has inefficient payoffs since v needs full agreement. Hence there is a proposal a_S to the full set of negotiators $S = N(\eta)$ that gives each $\nu \in S$ part of the additional surplus possible in $\eta' = \eta + S$. Let $\boldsymbol{a}'$ be $\boldsymbol{a}$ together with a_S, and put $y = (\eta', \boldsymbol{a}')$. Then the move $x \to y$ is feasible, strictly profitable, and not strongly Pareto-dominated, hence putting $p(x,y) = 1$ and $p(x,z) = 0$ for all $z \neq y$ fulfils conditions (i)–(iii).

(2) Let G be the graph of all moves $x \to y$ with $p(x,y) > 0$ and $x \neq y$. First, note that no move of type (ii) that only changes the agreed payoffs can be profitable since it would need to make at least one signatory worse-off. Hence each move $x \to y$ with $p(x,y) > 0$ either adds or removes a top-level agreement.

Acyclicity: For each two states x, y, there is a unique shortest sequence $s(x,y)$ of feasible moves from x to y that first removes at most $n-1$ agreements and then adds at most $n-1$ other agreements. For any longer sequence of feasible moves from x to y there is at least one agreement that is both added and removed along that sequence, and at most one of these partial moves can be profitable, hence such a longer sequence cannot build a directed path in G. Thus G either contains $s(x,y)$ as the unique directed path from x to y or contains $s(y,x)$ (which is the reverse of $s(x,y)$) as the unique directed path from x to y, or contains no directed paths between x and y at all. It cannot contain both since if the moves on $s(x,y)$ are profitable, their reverse moves are not. Hence G is acyclic.

Terminal nodes: If x has full agreement, no strictly profitable move is possible, hence $p(x,y) = 0$ for all $x \neq y$ by condition (i) and hence $p(x,x) = 1$, so x is a terminal node of G. If $x = (\eta, \boldsymbol{a})$ does not have full agreement, define y as in the proof of (1) above. Then the move $x \to y$ is feasible, strictly profitable, and not strongly Pareto-dominated, hence $p(x,x) = 0$ by condition (iii), so x is not a terminal node of G.

End: If the starting node has full agreement, it is terminal, so the process has already ended after zero steps. Otherwise, the process removes at most $n-2$ agreements and then adds at most $n-1$ other agreements before ending in a state will full agreement after at most $n-2+n-1 = 2n-3$ steps. QED.

Note that the $2n-3$ bound is sharp as the following EPHCF for the example at the end of Section 3.1.2 shows: For each state x with full agreement, put $p(x,x) = 1$. Now let $x = (\eta, \boldsymbol{a})$ be a state without full agreement and put $p(x,y) = 1$ for y defined as follows: If no coalitions exist yet ($\eta = \eta_0$), let y be the outcome of any profitable agreement between individuals 1 and 2. If the coalition $\{1,2\}$ or the coalition $\{1,3\}$ is already in η, let y be the outcome of any profitable agreement between that coalition and the remaining individual. If, finally, the coalition $\{2,3\}$ is already in η, let y be the outcome of removing their agreement. The latter move is profitable since in both x and y, both individuals 2 and 3 get no period payoff, but in y they will both get positive payoff after less moves than in x. When play starts with a coalition $\{2,3\}$, it first removes this agreement, then builds the coalition $\{1,2\}$, and finally builds

the grand coalition, ending with full agreement after $3 = 2n - 3$ steps.

The above proof of (2) relies on the fact that only atomic moves are considered feasible. Otherwise the acyclicity argument would not hold. Still, even when complex moves are feasible which simultaneously terminate and initiate several agreements, it seems plausible that full agreement will be reached. Consider an EPHCF p and an initial state $x = (\eta, \boldsymbol{a})$ without full agreement. Then, by assumption, each individual $i \in I$ believes her discounted infinite-horizon payoffs are $\xi_i(x)$, and we have inefficient expected payoffs $\Xi(x) = V(\{I\}) - \varepsilon$ that allow for an additional surplus of $\varepsilon > 0$ when the grand coalition forms. Let y be any state that results in adding to $\boldsymbol{a}$ an additional agreement to form the grand coalition and give each $i \in I$ some positive share of ε. Then the move $x \to y$ is not only feasible and strictly profitable but will also guarantee each individual a payoffs that is *certainly* larger than the expected value of the usually stochastic payoff they would get if p were followed. It seems plausible to assume that then the move $x \to y$ has at least a positive probability of being made. Hence our small set of axioms for an EPHCF should

negotiators would rather sign an agreement that forms the grand coalition and gives each ν

5 Conclusion

We have seen that the possibility of hierarchical agreements can lead to efficient outcomes in situations in which the existing literature on (non-hierarchical) coalition formation predicts inefficiency. By presenting a formal framework of negotiators and coalition and agreement hierarchies, I hope to have paved some ground for future work on this important question. Furthermore, the promising results in the public good example might be valuable in the game-theoretic study of International Environmental Agreements.

References

[1] R H Coase. The problem of social cost. *Journal of Law and Economics*, 3:1–44, 1960.

[2] P Herings. Coalition formation among farsighted agents. *Games*, 2010.

[3] Kyle Hyndman and Debraj Ray. Coalition formation with binding agreements. *Review of Economic Studies*, 74:1125–1147, 2007.

[4] H Konishi and Debraj Ray. Coalition formation as a dynamic process. *Journal of Economic Theory*, 110(1):1–41, 2003.

[5] Debraj Ray. *A Game-Theoretic Perspective on Coalition Formation (The Lipsey Lectures)*. Oxford University Press, USA, 2007.

[6] Debraj Ray and Rajiv Vohra. Equilibrium Binding Agreements. *Journal of Economic Theory*, 73:30–78, 1997.

[7] Debraj Ray and Rajiv Vohra. A theory of endogenous coalition structures. *Games and Economic Behavior*, 26(2):286–336, 1999.

[8] A Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 1982.