Formal Properties of Argumentation and Negotiation Frameworks

Paul E. Dunne

DSc The University of Edinburgh 2009

Table of Contents

THE UNIVERSITY OF EDINBURGH

ABSTRACT OF SUBMISSION FOR HIGHER DEGREE

(Regulation 1.4)

Name of Candidate: DR. PAUL E. DUNNE

Address : UNIVERSITY OF LIVERPOOL, DEPT. OF COMPUTER SCIENCE,
ASHTON BUILDING, ASHTON STREET, LIVERPOOL, UK

Postal Code: L69 3BX

Degree in view: DSc *Date:* February 2009

Title of Submission: Formal properties of argumentation and negotiation frameworks

Argumentation deals with the study of reasoning in "real world" settings, an important feature of which is that conclusions are rarely final: in contrast to the process of mathematical proof, positions arrived at via argumentation may be reversed in the light of further data becoming available or on account of changes in attitude. Computational models of argumentation are one important foundation of multiagent systems supporting tasks involving negotiation, e.g. when a number of agents wish to agree upon a division of resources amongst themselves.

The papers contributing to this thesis address algorithmic, combinatorial and computational complexity issues arising in two distinct but closely related contexts. The first considers these properties with reference to a widely used and influential computational model of argumentation. In this model, "arguments" are viewed as atomic elements and the principal relation of interest concerns whether an argument attacks another. The formal structure is, therefore, a directed graph (called an *argumentation framework)* $\langle X,A \rangle$ with X the set of arguments and $A \subseteq X \times X$, so that $\langle p,q \rangle \in A$ is read as "the argument p attacks the argument q". Argumentation frameworks provide a basis for defining collections of acceptable arguments as subsets of X meeting particular criteria. The analysis of such *extension-based semantics* in this thesis concentrates on algorithmic and complexity properties.

The second area considered deals with one model of distributed negotiation in multiagent systems in which agents attempt to agree upon an allocation of resources. In this model, one has a set of n agents (Ag), and a collection of m resources (R). These agents seek to agree a partition of R (starting from some initial allocation) that takes into account the value each agent assigns to distinct subsets of R. The negotiation model allows agents to propose exchanges and accept (or reject) offers made by other agents. The effects of limiting this general scheme, e.g. by restricting the number of agents or resources that may feature in a given offer, are examined in a series of papers analysing complexity properties related to deciding the existence of appropriate "contract paths" together with extremal results on the length of such negotiations.

Declaration

No part of this dissertation has been submitted or is currently being considered for the award of any other degree or postgraduate diploma. The author's involvement and contribution to the articles comprising the body of this thesis is summarised below. The ordering in which papers are given is that of the Table of Contents above.

Argumentation Frameworks

Coherence in Finite Argument Systems.
P.E. Dunne and T.J.M. Bench-Capon.
Artificial Intelligence, 141:187–203, October 2002

I began working on properties of argumentation frameworks after becoming aware of the graph-theoretic model of these. My colleague Trevor Bench-Capon and I have collaborated on a number of research papers in this field. Given our different areas of expertise, those whose contribution was of a highly technical and theoretical nature were formulated and developed by myself, whereas those whose perspective was rather more speculative and philosophically slanted were principally developed by him. This article (and the one following) comes within the former category. Its central topic concerns the precise complexity classification of a (then) open problem in extension-based argumentation semantics. The relationship between the technical properties of coherence and sceptical acceptance, the translation from Boolean propositional formulae to argument systems, and the subsequent exploitation of this translation in establishing complexity classifications were my contribution.

Two Party Immediate Response Disputes: Properties and Efficiency.
P.E. Dunne and T.J.M. Bench-Capon.
Artificial Intelligence, 149:221–250, October 2003

The important contributions of this article are the following: presenting a formal operational description of a dialogue process for determining the acceptability status of a given argument (the *Two Party Immediate Response* (or TPI) dispute of the title); the introduction of concepts of "dispute length" and "dispute complexity" as formal mechanisms for performing quantitative comparisons of dialogue protocols; establishing that dispute complexity can be analysed in terms of pre-existing models of propositional proof systems; and in presenting a simulation of TPI-derivations by sequent (Gentzen) calculus derivations. This simulation allows the construction of explicit examples of argument processes whose shortest resolution is exponential in terms of the number of arguments. All of these developments and the associated analysis were produced by me.

Signed: Paul E. Dunne; 12/02/2009

Declaration

On the complexity of linking deductive and abstract argument systems.
M. J. Wooldridge, P. E. Dunne, and S. Parsons.
Proc. AAAI-06, July 2006, pages 299–304

The aim of this paper was to examine the relationship between two different computational models of argumentation: the deductive approach – to which my coauthors had made several contributions – and the graph-theoretic model of abstract argument systems, which had been the main focus of work in the two articles discussed above. All of the results presented within the section "Argument sets, Distinctness, and Maximality" (with the exception of Theorem 4) were derived by me, as well as the preliminary analysis that led to the properties proven in the section entitled "Argument Graphs".

Computational properties of argument systems satisfying graph-theoretic constraints.
P. E. Dunne
Artificial Intelligence, 171:701–729, August 2007

Sole authorship.

Negotiation Frameworks in Multiagent Systems

Extremal behaviour in multiagent contract negotiation.
P. E. Dunne
Jnl. of Artificial Intelligence Research, 23:41–78, January 2005

Sole authorship.

Optimal Utterances in Dialogue Protocols.
P. E. Dunne and P. McBurney.
Proc. AAMAS'03, July 2003, ACM Press, pages 608–615.

This paper resulted from interest in importing the analysis of argumentation protocols, in particular the work on TPI–disputes, to more general agent negotiation protocols (a specialist interest of my coauthor Peter McBurney). The core research elements gave a precise formulation of the concept of an "optimal contribution" to a dialogue and an analysis of the associated "Optimal Utterance Problem" (Δ-OUP): this formulation and the complexity analysis relating Δ-OUP to the identification of optimal branching literals in a standard satisfiability algorithm were contributed by me.

Signed: Paul E. Dunne; 12/02/2009

Declaration

The complexity of contract negotiation.
P. E. Dunne, M. J. Wooldridge, and M. R. Laurence.
Artificial Intelligence, 16:23–46, May 2005

Between January 2002 and December 2004, I was Principal Investigator (with Mike Wooldridge involved as co-investigator and Mike Laurence employed as an RA) on an EPSRC funded project entitled "Algorithmics for Agent Design & Verification" (EPSRC Refce: GR/R60836/01). This article came about after identifying a number of omissions in treatments of resource allocation and negotiation models within multiagent systems. The important contributions made by this work were: highlighting the importance of concise representations of "utility" functions; formally defining one such representation (the straight-line program – SLP – approach); introducing, motivating, and analysing a number of decision problems relating to the realisability of allocations from a given initial allocation when the negotiation protocol restricts the form particular exchanges can have. The detailed work on all of these was initiated and carried out by me.

The complexity of deciding reachability properties of distributed negotiation schemes.
P. E. Dunne and Y. Chevaleyre.
Theoretical Computer Science, 396:113–144, May 2008

While the article above was still in press, I was invited to participate in a Technical Forum Group (TFG) on "Multiagent Resource Allocation" organised under the AgentLink programme, and which took place in Ljubljana between 28th February and 1st March 2005, Following my presentation of work reported in this article, there was some discussion concerning the gap between lower (NP–hard) and upper bounds (PSPACE) on the complexity of one group of problems that had been considered. Subsequently I was able to obtain an exact bound (PSPACE–completeness) for this group of problems, and gave an overview presentation at the follow-up TFG meeting in Budapest, in September 2005. My coauthor on the published article, Yann Chevaleyre, identified one way in which the argument could be simplified, thus allowing the original proof (which had used a seven agent construction) to be recast as a five agent analysis.

Signed: Paul E. Dunne; 12/02/2009

Coherence in Finite Argument Systems

ELSEVIER

Artificial Intelligence 141 (2002) 187–203

Artificial Intelligence

www.elsevier.com/locate/artint

Research Note

Coherence in finite argument systems

Paul E. Dunne *, T.J.M. Bench-Capon

Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

Received 7 December 2001

Abstract

Argument Systems provide a rich abstraction within which divers concepts of reasoning, acceptability and defeasibility of arguments, etc., may be studied using a unified framework. Two important concepts of the acceptability of an argument p in such systems are *credulous acceptance* to capture the notion that p *can* be 'believed'; and *sceptical acceptance* capturing the idea that if *anything* is believed, then p *must* be. One important aspect affecting the computational complexity of these problems concerns whether the admissibility of an argument is defined with respect to '*preferred*' or '*stable*' semantics. One benefit of so-called '*coherent*' argument systems being that the preferred extensions coincide with stable extensions. In this note we consider complexity-theoretic issues regarding deciding if finitely presented argument systems modelled as directed graphs are coherent. Our main result shows that the related decision problem is $\Pi_2^{(p)}$-complete and is obtained solely via the graph-theoretic representation of an argument system, thus independent of the specific logic underpinning the reasoning theory.

Keywords: Argument Systems; Coherence; Credulous reasoning; Sceptical reasoning; Computational complexity

1. Introduction

Since they were introduced by Dung [8], Argument Systems have provided a fruitful mechanism for studying reasoning in defeasible contexts. They have proved useful both to theorists who can use them as an abstract framework for the study and comparison of non-monotonic logics, e.g., [2,5,6], and for those who wish to explore more concrete contexts where defeasibility is central. In the study of reasoning in law, for example, they have

* Corresponding author.
E-mail address: ped@csc.liv.ac.uk (P.E. Dunne).

0004-3702/02/$ – see front matter
PII: S0004-3702(02)00261-8

been used to examine the resolution of conflicting norms, e.g., [12], especially where this is studied through the mechanism of a dispute between two parties, e.g., [11]. The basic definition below is derived from that given in [8].

Definition 1. An *argument system* is a pair $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$, in which $\mathcal{X}$ is a set of *arguments* and $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is the *attack relationship* for $\mathcal{H}$. Unless otherwise stated, $\mathcal{X}$ is assumed to be *finite*, and $\mathcal{A}$ comprises a set of ordered pairs of *distinct* arguments. A pair $\langle x, y \rangle \in \mathcal{A}$ is referred to as 'x *attacks* (or *is an attacker of*) y' or 'y *is attacked by* x'.

For R, S subsets of arguments in the system $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$, we say that

(a) $s \in S$ is *attacked* by R if there is some $r \in R$ such that $\langle r, s \rangle \in \mathcal{A}$.
(b) $x \in \mathcal{X}$ is *acceptable with respect to* S if for every $y \in \mathcal{X}$ that attacks x there is some $z \in S$ that attacks y.
(c) S is *conflict-free* if no argument in S is attacked by any other argument in S.
(d) A conflict-free set S is *admissible* if every argument in S is acceptable with respect to S.
(e) S is a *preferred extension* if it is a maximal (with respect to $\subseteq$) admissible set.
(f) S is a *stable extension* if S is conflict free and every argument $y \notin S$ is attacked by S.
(g) $\mathcal{H}$ is *coherent* if every preferred extension in $\mathcal{H}$ is also a stable extension.

An argument x is *credulously accepted* if there is *some* preferred extension containing it; x is *sceptically accepted* if it is a member of *every* preferred extension.

The graph-theoretic representation employed by finite argument systems, naturally suggests a unifying formalism in which to consider various decision problems. To place our main results in a more general context we start from the basis of the decision problems described by Table 1 in which: $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is an argument system as in Definition 1; x an argument in $\mathcal{X}$; and S a subset of arguments in $\mathcal{X}$.

Polynomial-time decision algorithms for problems (1) and (2) are fairly obvious. The results regarding problems (3–7) are discussed below. In this article we are primarily concerned with the result stated in the final line of Table 1: our proof of this yields (8) as an easy corollary.

Table 1
Decision problems in finite argument systems and their complexity

	Problem	Decision question	Complexity
1	ADM($\mathcal{H}, S$)	Is S admissible?	P
2	STAB-EXT($\mathcal{H}, S$)	Is S a *stable* extension?	P
3	PREF-EXT($\mathcal{H}, S$)	Is S a *preferred* extension?	CO-NP-complete
4	HAS-STAB($\mathcal{H}$)	Does $\mathcal{H}$ have any stable extension?	NP-complete
5	CA($\mathcal{H}, x$)	Is x in some *preferred* extension?	NP-complete
6	IN-STAB($\mathcal{H}, x$)	Is x in some *stable* extension?	NP-complete
7	ALL-STAB($\mathcal{H}, x$)	Is x in *every* stable extension?	CO-NP-complete
8	SA($\mathcal{H}, x$)	Is x in *every* preferred extension?	$\Pi_2^{(p)}$-complete
9	COHERENT($\mathcal{H}$)	Is $\mathcal{H}$ coherent?	$\Pi_2^{(p)}$-complete

Before proceeding with this, it is useful to discuss important related work of Dimopoulos and Torres [7], in which various semantic properties of the Logic Programming paradigm are interpreted with respect to a (directed) graph translation of *reduced negative* logic programs: graph vertices are associated with rules and the concept of '*attack*' modelled by the presence of edges $\langle r, s\rangle$ whenever there is a non-empty intersection between the set of literals defining the head of r and the negated set of literals in the body of s, i.e., if $z \in body(s)$ then $\neg z$ is in this negated set. Although Dimopoulos and Torres [7] do not employ the terminology—in terms of credulous acceptance, admissible sets, etc.—from [8] used in the present article it is clear that similar forms are being considered: the structures referred to as '*semi-kernel*', '*maximal semi-kernel*' and '*kernel*' in [7] corresponding to 'admissible set', 'preferred extension' and 'stable extension' respectively. The complexity results for problems (3–6) if not immediate from [7, Theorem 5.1, Lemma 5.2, Proposition 5.3] are certainly implied by these. In this context, it is worth drawing attention to some significant points regarding [7, Theorem 5.1] which, translated into the terminology of the present article states:

> The problem of deciding whether an argument system $\mathcal{H}(\mathcal{X}, \mathcal{A})$ has a *non-empty* preferred extension is NP-complete.

First, this implies the complexity classification for PREF-EXT stated, *even* when the subset S forming part of an instance is *the empty set*.

A second point, also relevant to our proof of (9) concerns the transformation used: [7] present a translation of propositional formulae Φ in 3-CNF (this easily generalises for arbitrary CNF formulae) into a finite argument system $\mathcal{H}_\Phi$. It is not difficult, however, given $\mathcal{H}(\mathcal{X}, \mathcal{A})$ to define CNF-formulae $\Phi_\mathcal{H}$ whose satisfiability properties are dependent on the presence of particular structures within $\mathcal{H}$, e.g., stable extensions, admissible subsets containing specific arguments, etc. We thus have a mechanism for transforming a given $\mathcal{H}$ into an 'equivalent' system $\mathcal{F}$ the point being that $\mathcal{F}$ *may* provide a 'better' basis for graph-theoretic analyses of structures within $\mathcal{H}$.

Our final observation, concerns problem (7): although the given complexity classification is neither explicitly stated in nor directly implied by the results of [7], that ALL-STAB is CO-NP-complete can be shown using some minor 're-wiring' of the argument graph G_Φ constructed from an instance Φ of 3-SAT.[1]

The concept of *coherence* was formulated by Dung [8, Definition 31(1), p. 332], to describe those argument systems whose stable and preferred extensions coincide. One significant benefit of coherence as a property has been established in recent work of Vreeswijk and Prakken [13] with respect to proof mechanisms for establishing *sceptical* acceptance: problem (8) of Table 1. In [13] a sound and complete reasoning method for credulous acceptance—using a dialogue game approach—is presented. This approach, as the authors observe, provides a sound and complete mechanism for *sceptical* acceptance in precisely those argument systems that are coherent. Thus a major advantage of coherent

[1] This involves removing all except the edge $\langle Aux, A\rangle$ for edges $\langle A, x\rangle$ or $\langle x, A\rangle$: then ALL-STAB$(G_\Phi, A) \Leftrightarrow \neg$3–SAT$(\Phi)$.

argument systems is that proofs of sceptical acceptance are (potentially) rather more readily demonstrated in coherent systems via devices such as those of [3,13]. The complexity of sceptical acceptance is considered (in the context of membership in preferred extensions) for various non-monotonic logics in [5], where completeness results at the third-level of the polynomial-time hierarchy are demonstrated. Although Dimopoulos et al. [5] conclude that these complexity results 'discredit sceptical reasoning as ... "unnecessarily" complex', it might be argued that within finite systems where coherence is 'promised' this view may be unduly pessimistic. Notwithstanding our main result that testing coherence is, in general, extremely hard, there is an efficiently testable property that suffices to guarantee coherence. Some further discussion of this is presented in Section 3.

In the next section we present the main technical contribution of this article, that COHERENT is $\Pi_2^{(p)}$-complete: the complexity class $\Pi_2^{(p)}$ comprising those problems decidable by CO-NP computations given (unit cost) access to an NP oracle. Alternatively, $\Pi_2^{(p)}$ can be viewed as the class of languages, L, membership in which is certified by a (deterministic) polynomial-time testable ternary relation $R_L \subseteq W \times X \times Y$ such that, for some polynomial bound $p(|w|)$ in the number of bits encoding w,

$$w \in L \Leftrightarrow \big(\forall x \in X\colon\ |x| \leqslant p(|w|)\big)\big(\exists y \in Y\colon\ |y| \leqslant p(|w|)\big) \quad \langle w, x, y\rangle \in R_L.$$

Our result in Theorem 2 provides some further indications that decision questions concerning preferred extensions are (under the usual complexity-theoretic assumptions) likely to be harder than the analogous questions concerning *stable* extensions: line (8) of Table 1 is an easy corollary of our main theorem. Similar conclusions had earlier been drawn in [5,6], where the complexity of reasoning problems in a variety of non-monotonic Logics is considered under both preferred and stable semantics. This earlier work establishes a close link between the complexity of the reasoning problem and that of the *derivability problem* for the associated logic. One feature of our proof is that the result is established purely through a graph-theoretic interpretation of argument, similar in spirit, to the approach adopted in [7]: thus, the differing complexity levels may be interpreted in purely graph-theoretic terms, independently of the logic that the graph structure is defined from.

In Section 3 we discuss some consequences of our main theorem in particular with respect to its implications for designing *dialogue game* style mechanisms for Sceptical Reasoning. Conclusions are presented in Section 4.

2. Complexity of deciding coherence

Theorem 2. COHERENT *is* $\Pi_2^{(p)}$*-complete.*

In order to clarify the proof structure we establish it via a series of technical lemmata. The bulk of these are concerned with establishing $\Pi_2^{(p)}$-hardness, i.e., with reducing a known $\Pi_2^{(p)}$-complete problem to COHERENT.

We begin with the, comparatively easy, proof that COHERENT($\mathcal{H}$) is in $\Pi_2^{(p)}$.

Lemma 3. COHERENT$(\mathcal{H}) \in \Pi_2^{(p)}$.

Proof. Given an instance, $\mathcal{H}(\mathcal{X}, \mathcal{A})$ of COHERENT, it suffices to observe that

$$\text{COHERENT}(\mathcal{H}) \Leftrightarrow \forall S\big(\neg\text{PREF-EXT}(\mathcal{H}, S) \vee \text{STAB-EXT}(\mathcal{H}, S)\big),$$

i.e., $\mathcal{H}$ is coherent if and only if for each subset S of $\mathcal{X}$: either S is *not* a preferred extension or S *is* a stable extension. Since $\neg$PREF-EXT$(\mathcal{H}, S)$ is in NP, i.e., $\Sigma_1^{(p)}$ and STAB-EXT$(\mathcal{H}, S)$ in P, we have COHERENT in $\Pi_2^{(p)}$ as required. □

The decision problem we use as the basis for our reduction is QSAT_2. An instance of QSAT_2 is a well-formed propositional formula, $\Phi(X, Y)$, defined over disjoint sets of propositional variables, $X = \langle x_1, x_2, \ldots, x_n\rangle$ and $Y = \langle y_1, y_2, \ldots, y_t\rangle$. Without loss of generality we may assume that: $n = t$; Φ is formed using only the Boolean operations $\wedge$, $\vee$, and $\neg$; and negation is only applied to variables in $X \cup Y$. An instance, $\Phi(X, Y)$ of QSAT_2 is accepted if and only if $\forall\alpha_X \exists\beta_Y\ \Phi(\alpha_X, \beta_Y)$. That is, no matter how the variables in X are instantiated (α_X) there is *some* instantiation (β_Y) of Y such that $\langle\alpha_X, \beta_Y\rangle$ satisfies Φ. That QSAT_2 is $\Pi_2^{(p)}$-complete was shown in [14].

We start by presenting some technical definitions. The first of these describes a standard presentation of propositional formulae as *directed rooted trees* that has been widely used in applications of Boolean formulae, see, e.g., [9, Chapter 4].

Definition 4. Let $\Phi(Z)$ be a well-formed propositional formula (wff) over the variables $Z = \langle z_1, z_2, \ldots, z_n\rangle$ using the operations $\{\wedge, \vee, \neg\}$ with negation applied only to variables of Φ. The *tree representation of* Φ (denoted T_Φ) is a rooted directed tree with root vertex denoted $\rho(T_\Phi)$ and inductively defined by the following rules.

(a) If $\Phi(Z) = w$ – a single literal z or $\neg z$, then T_Φ consists of a single vertex $\rho(T_\Phi)$ labelled w.
(b) If $\Phi(Z) = \bigwedge_{i=1}^{k} \Psi_i(Z)$, for wff $\langle\Psi_1, \Psi_2, \ldots, \Psi_k\rangle$, T_Φ is formed from the k tree representations $\langle T_{\Psi_i}\rangle$ by directing edges from each $\rho(T_{\Psi_i})$ into a new root vertex $\rho(T_\Phi)$ labelled $\wedge$.
(c) If $\Phi(Z) = \bigvee_{i=1}^{k} \Psi_i(Z)$, for wff $\langle\Psi_1, \Psi_2, \ldots, \Psi_k\rangle$, T_Φ is formed from the k tree representations $\langle T_{\Psi_i}\rangle$ by directing edges from each $\rho(T_{\Psi_i})$ into a new root vertex $\rho(T_\Phi)$ labelled $\vee$.

In what follows we use the term *node* of T_Φ to refer to an arbitrary tree vertex, i.e., a leaf or internal vertex.

In the tree representation of Φ, each leaf vertex is labelled with some literal w, (several leaves may be labelled with the same literal), and each internal vertex with an operation in $\{\wedge, \vee\}$. We shall subsequently refer to the internal vertices of T_Φ as the *gates* of the tree. Without loss of generality we may assume that the successor of any $\wedge$-gate (tree vertex labelled $\wedge$) is an $\vee$-gate (tree vertex labelled $\vee$) and *vice versa*. The *size* of $\Phi(Z)$ is the number of *gates* in its tree representation T_Φ. For formulae of size m we denote by

$\langle g_1, g_2, \ldots, g_m \rangle$ the gates in T_Φ with g_m always taken as the root $\rho(T_\Phi)$ of the tree. Finally for any edge $\langle h, g \rangle$ in T_Φ we refer to the node h as an *input* of the gate g.[2]

Definition 5. For a formula, $\Phi(Z)$, an *instantiation* of its variables is a mapping, $\pi : Z \to \{\textbf{true}, \textbf{false}, *\}$ associating a truth value or unassigned status $(*)$ with each variable z_i. We use π_i to denote $\pi(z_i)$. An instantiation is *total* if every variable is assigned a value in $\{\textbf{true}, \textbf{false}\}$ and *partial* otherwise. We define a partial ordering over instantiations γ and δ to Z by writing $\gamma < \delta$ if: for each i with $\gamma_i \neq *$, $\delta_i = \gamma_i$, and there is at least one i, for which $\gamma_i = *$ and $\delta_i \neq *$.

Given $\Phi(Z)$ any instantiation $\pi : Z \to \{\textbf{true}, \textbf{false}, *\}$ induces a mapping from the nodes defining T_Φ onto values in $\{\textbf{true}, \textbf{false}, *\}$. Assuming the natural generalisations of $\wedge$ and $\vee$ to the domain $\langle \textbf{true}, \textbf{false}, * \rangle$,[3] we define for h a node in T_Φ, its value $\nu(h, \pi)$ under the instantiation π of Z as

$$\nu(h,\pi) = \begin{cases} * & \text{if } h \text{ is a leaf node labelled } z_i \text{ or } \neg z_i \text{ and } \pi_i = *, \\ \pi_i & \text{if } h \text{ is a leaf node labelled } z_i \text{ and } \pi_i \neq *, \\ \neg\pi_i & \text{if } h \text{ is a leaf node labelled } \neg z_i \text{ and } \pi_i \neq *, \\ \bigvee_{j=1}^{k} \nu(h_j, \pi) & \text{if } h \text{ is an } \vee\text{-gate with inputs } \langle h_1, \ldots, h_k \rangle, \\ \bigwedge_{j=1}^{k} \nu(h_j, \pi) & \text{if } h \text{ is an } \wedge\text{-gate with inputs } \langle h_1, \ldots, h_k \rangle, \end{cases}$$

where π is clear from the context, we write $\nu(h)$ for $\nu(h, \pi)$.

With this concept of the value induced at a node of T_Φ via an instantiation π, we can define a partition of the *literals* and *gates* in T_Φ that is used extensively in our later analysis.

The *value partition* $Val(\pi)$ of T_Φ comprises three sets $\langle True(\pi), False(\pi), Open(\pi) \rangle$.

(T1) The subset $True(\pi)$ consists of literals and gates, h, for which $\nu(h) = \textbf{true}$.
(T2) The subset $False(\pi)$ consists of literals and gates, h, for which $\nu(h) = \textbf{false}$.
(T3) The subset $Open(\pi)$ consists of literals and gates, h, for which $\nu(h) = *$.

The following properties of this partition can be easily proved:

Fact 6.

(a) $Open(\pi) = \emptyset \Leftrightarrow \pi$ *is total.*
(b) *If* $\gamma < \delta$, *then* $True(\gamma) \subset True(\delta)$ *and* $False(\gamma) \subset False(\delta)$.

For example in Fig. 1 under the partial instantiation $\pi = \langle z_1 = \textbf{true}, z_4 = \textbf{false} \rangle$ with all other variables unassigned, we have: $True(\pi) = \{z_1, \neg z_4, g_1\}$; $False(\pi) = \{\neg z_1, z_4, g_3\}$; and $Open(\pi) = \{z_2, \neg z_2, z_3, \neg z_3, g_2, g_4\}$.

[2] We note that since any gate may be assumed to have at most n distinct literals among its inputs, our measure of formula size as 'number of gates' is polynomially equivalent to the more usual measure of size as 'number of literal occurrences', i.e., leaf nodes.

[3] I.e., $\bigwedge_{j=1}^{k} x_j$ is $*$ unless all x_j are **true** or at least one x_j is **false**; $\bigvee_{j=1}^{k} x_j$ is $*$ unless all x_j are **false** or at least one is **true**.

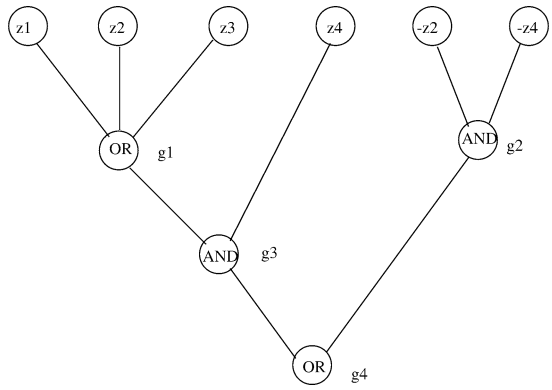

Fig. 1. $T_\Phi(z_1, z_2, z_3, z_4)$ for $(z_1 \vee z_2 \vee z_3) \wedge z_4 \vee (\neg z_2 \wedge \neg z_4)$.

At the heart of our proof that QSAT$_2$ is polynomially reducible to COHERENT is a translation from the tree representation T_Φ of a formula $\Phi(X, Y)$ to an argument system $\mathcal{H}_\Phi(\mathcal{X}_\Phi, \mathcal{A}_\Phi)$. It will be useful to proceed by presenting a preliminary translation that, although not in the final form that will be used in the reduction, will have a number of properties that will be important in deriving our result.

Definition 7. Let $\Phi(Z)$ be a propositional formula with tree representation T_Φ having size m. The *Argument Representation* of Φ is the argument system $\mathcal{R}_\Phi(\mathcal{X}_\Phi, \mathcal{A}_\Phi)$ defined as follows. $\mathcal{R}_\Phi$ contains the following arguments $\mathcal{X}_\Phi$:

(X1) $2n$ *literal* arguments $\{z_i, \neg z_i\colon\ 1 \leqslant i \leqslant n\}$.
(X2) For each gate g_k of T_Φ, an argument $\neg g_k$ (if g_k is an $\vee$-gate) or an argument g_k (if g_k is an $\wedge$-gate). If g_m, i.e., the root of T_Φ, happens to be an $\vee$-gate, then an additional argument g_m is included. We subsequently denote this set of arguments by $\mathcal{G}_\Phi$.

The attack relationship—$\mathcal{A}_\Phi$—over $\mathcal{X}_\Phi$ contains:

(A1) $\{\langle z_i, \neg z_i\rangle, \langle \neg z_i, z_i\rangle\colon\ 1 \leqslant i \leqslant n\}$.
(A2) $\langle \neg g_m, g_m\rangle$ if g_m is an $\vee$-gate in T_Φ.
(A3) If g_k is an $\wedge$-gate with inputs $\{h_1, h_2, \ldots, h_r\}$: $\{\langle \neg h_i, g_k\rangle\colon\ 1 \leqslant i \leqslant r\}$.
(A4) If g_k is an $\vee$-gate with inputs $\{h_1, h_2, \ldots, h_r\}$: $\{\langle h_i, \neg g_k\rangle\colon\ 1 \leqslant i \leqslant r\}$.

Fig. 2 shows the result of this translation when it is applied to the tree representation of the formula in Fig. 1.

The arguments defining $\mathcal{R}_\Phi$ fall into one of two sets: $2n$ arguments corresponding to the $2n$ distinct literals over Z; and m (or $m+1$) '*gate*' arguments. The key idea is the following: any instantiation π of the propositional variables Z of Φ, induces the partition $Val(\pi)$ of literals and gates in T_Φ. In the argument system $\mathcal{R}_\Phi$ the attack relationship for *gate* arguments, reflects the conditions under which the corresponding argument is admissible (with respect to the subset of literal arguments marked out by π). For example,

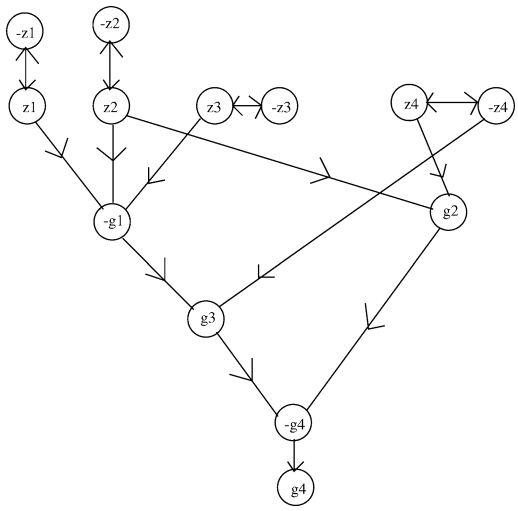

Fig. 2. The argument system $\mathcal{R}_\Phi$ from the formula of Fig. 1.

suppose g_1 is an $\vee$-gate with literals z_1, $\neg z_2$, z_3 as its inputs. In the simulating argument system, g_1 is represented by an argument labelled $\neg g_1$ which is attacked by the (arguments labelled with) literals z_1, $\neg z_2$, and z_3: the interpretation being that "the assertion 'g_1 is **false**' is attacked by instantiations in which z_1 or $\neg z_2$ or z_3 are **true**". Similarly were g_1 an $\wedge$-gate it would appear in $\mathcal{R}_\Phi$ as an argument labelled g_1 which was attacked by literals $\neg z_1$, z_2, and $\neg z_3$: the interpretation now being that "the assertion 'g_1 is **true**' is attacked by instantiations in which z_1 or $\neg z_2$ or z_3 are **false**". With this viewpoint, any instantiation π will induce a selection of the literal arguments and a selection of the *gate* arguments (i.e., those for which no attacking argument has been included).

Suppose π is an instantiation of Z. The key idea is to map the partition of the tree representation T_Φ as $Val(\pi)$ onto an analogous partition of the literal and gate arguments in $\mathcal{R}_\Phi$. Given π this partition comprises 3 sets, $\langle In(\pi),\ Out(\pi),\ Poss(\pi)\rangle$ defined by:

(R1) An argument p is in the subset $In(\pi)$ of $\mathcal{X}_\Phi$ if:

(p is the argument z_i, $\pi_i = \mathbf{true}$) or (p is the argument $\neg z_i$, $\pi_i = \mathbf{false}$)
or ($p = \neg g \in \mathcal{G}_\Phi$ and $g \in T_\phi$ is in $False(\pi)$)
or ($p = g \in \mathcal{G}_\Phi$ and $g \in T_\phi$ is in $True(\pi)$).

(R2) An argument p is in the subset $Out(\pi)$ of $\mathcal{X}_\Phi$ if:

(p is the argument z_i, $\pi_i = \mathbf{false}$) or (p is the argument $\neg z_i$, $\pi_i = \mathbf{true}$)
or ($p = \neg g \in \mathcal{G}_\Phi$ and $g \in T_\phi$ is in $True(\pi)$)
or ($p = g \in \mathcal{G}_\Phi$ and $g \in T_\phi$ is in $False(\pi)$).

(R3) An argument p is in the subset $Poss(\pi)$ of $\mathcal{X}_\Phi$ if:

$p \notin In(\pi) \cup Out(\pi)$.

With the formulation of the argument system $\mathcal{R}_\Phi(\mathcal{X}_\Phi, \mathcal{A}_\Phi)$ from the formula $\Phi(Z)$ and the definition of the partition $\langle In(\pi), Out(\pi), Poss(\pi)\rangle$ via the value partition $Val(\pi)$ of T_Φ we are now ready to embark on the sequence of technical lemmata which will culminate in the proof of Theorem 2.

Our proof strategy is as follows. We proceed by characterising the set of preferred extensions of $\mathcal{R}_\Phi$ showing—in Lemma 8 through Lemma 11—that these consist of exactly the subsets defined by $In(\gamma_Z)$ where γ_Z is a *total* instantiation of Z. In Lemma 12 we deduce that these are all stable extensions and thus that $\mathcal{R}_\Phi$ is itself coherent. In the remaining lemmata, we consider the argument systems arising by transforming instances $\Phi(X, Y)$ of QSAT_2. In these, however, we add to the basic system defined by $\mathcal{R}_\Phi$ (which will have $4n$ literal arguments and m (or $m+1$) gate arguments) an additional set of 3 *control arguments* one of which attacks all of the Y-literal arguments: we denote this augmented system by $\mathcal{H}_\Phi(\mathcal{W}_\Phi, \mathcal{B}_\Phi)$. As will be seen in Lemma 15, it follows easily from Lemma 10 that for any $\langle \alpha_X, \beta_Y\rangle$ satisfying $\Phi(X, Y)$ the subset $In(\alpha_X, \beta_Y)$ is a stable extension of both $\mathcal{R}_\Phi$ *and* H_Φ. The crucial property provided by the additional control arguments in $\mathcal{H}_\Phi$ is proved in Lemma 16: *if* for α_X there is no β_Y for which $\langle \alpha_X, \beta_Y\rangle$ satisfies $\Phi(X, Y)$ then the subset $In(\alpha_X)$ (defined from $\mathcal{R}_\Phi$) is a preferred *but not stable* extension of $\mathcal{H}_\Phi$, where $In(\alpha_X)$ denotes the set $In(\alpha_X, *, *, \ldots, *)$ in which every y_i is unassigned. The reason for introducing the control arguments in moving from $\mathcal{R}_\Phi$ to $\mathcal{H}_\Phi$ is that $In(\alpha_X)$ is *not* a preferred extension of $\mathcal{R}_\Phi$: although it is admissible, it could be extended by adding, for example, Y-literal arguments. The design of $\mathcal{H}_\Phi$ will be such that unless the gate argument g_m can be used in an *admissible* extension of $In(\alpha_X)$ then $In(\alpha_X)$ is already maximal in $\mathcal{H}_\Phi$ and not a stable extension since the control arguments are not attacked. Finally, in Lemma 17, it is demonstrated that the *only* preferred extensions of $\mathcal{H}_\Phi$ are those arising as a result of Lemmas 15 and 16. Theorem 2 will follow easily from Lemma 17, since the argument g_m—corresponding to the root node $\rho(T_\Phi)$ of the instance $\Phi(X, Y)$—must necessarily belong to any stable extension in $\mathcal{H}_\Phi$: hence $\mathcal{H}_\Phi$ is coherent if and only if for each instantiation α_X there is an instantiation β_Y such that $\langle \alpha_X, \beta_Y\rangle$ satisfies $\Phi(X, Y)$, i.e., for which $g_m \in In(\alpha_X, \beta_Y)$ in the system $\mathcal{R}_\Phi$ and thence in the corresponding stable extension of $\mathcal{H}_\Phi$.

We employ the following notational conventions: α_X, β_Y, (and γ_Z) denote *total* instantiations of X, Y, (and Z); for an argument p in $\mathcal{X}_\Phi$, g_p (respectively h_p) denotes the corresponding gate (respectively node) in T_Φ, hence if g_p is an $\vee$-gate, then p is the argument labelled $\neg g_p$; $\mathcal{PE}^{\mathcal{M}}$ (respectively $\mathcal{SE}^{\mathcal{M}}$) denotes the set of *all* preferred (respectively stable) extensions in the argument system $\mathcal{M}_\Phi$, where $\mathcal{M}_\Phi$ is one of $\mathcal{R}_\Phi$ or $\mathcal{H}_\Phi$.

Lemma 8. $\forall \gamma_Z$ $In(\gamma_Z)$ *is conflict-free.*

Proof. Let γ_Z be an instantiation of Z and consider the subset $In(\gamma_Z)$ of $\mathcal{X}_\Phi$ in $\mathcal{R}_\Phi$. Suppose that there are arguments p and q in $In(\gamma_Z)$ for which $\langle p, q\rangle \in \mathcal{A}_\Phi$. It cannot be the case that $h_p = u_i$ and $h_q = \neg u_i$ for u_i some literal over z_i, since exactly one of $\{z_i, \neg z_i\}$ is in $True(\gamma_Z)$ hence exactly one of the corresponding literal arguments is in $In(\gamma_Z)$. Thus q must be a gate argument. Suppose g_q is an $\vee$-gate: $q \in In(\gamma_Z)$ only if $g_q \in False(\gamma_Z)$ and therefore h_p, which (since $\langle p, q\rangle \in \mathcal{A}_\Phi$) must be an input of g_q is also in $False(\gamma_Z)$.

This leads to a contradiction: if h_p is a gate then it is an $\wedge$-gate, so $p \in In(\gamma_Z)$ only if $h_p \in True(\gamma_Z)$; if h_p is a literal u_i, then $h_p \in False(\gamma_Z)$ would mean that $\neg u_i \in True(\gamma_Z)$ and hence $u_i \notin In(\gamma_Z)$. The remaining possibility is that g_q is an $\wedge$-gate: $q \in In(\gamma_Z)$ only if $g_q \in True(\gamma_Z)$ and thus $h_p \in True(\gamma_Z)$. If h_p is a gate it must be an input of g_q and an $\vee$-gate: $h_p \in True(\gamma_Z)$ would force $p \notin In(\gamma_Z)$. Finally if the input h_p is a literal u_i in T_Φ then in $\mathcal{R}_\Phi$ the literal $\neg u_i$ attacks q: $u_i \in True(\gamma_Z)$ implies $\neg u_i \notin In(\gamma_Z)$. We deduce that $In(\gamma_Z)$ must be conflict-free. □

Lemma 9. $\forall \gamma_Z$ $In(\gamma_Z)$ *is admissible.*

Proof. From Lemma 8, $In(\gamma_Z)$ is conflict-free, so it suffices to show for all arguments $p \notin In(\gamma_Z)$ that attack some $q \in In(\gamma_Z)$ there is an argument $r \in In(\gamma_Z)$ that attacks p. Let p, q be such that $p \notin In(\gamma_Z)$, $q \in In(\gamma_Z)$ and $\langle p, q\rangle \in \mathcal{A}_\Phi$. If q is a literal argument, u_i say, then p must be the literal argument $\neg u_i$ and choosing $r = q$ provides a counter-attacker to p. Suppose q is a gate argument. One of the inputs to g_q must be the node h_p. If g_q is an $\vee$-gate then $g_q \in False(\gamma_Z)$ and $h_p \in False(\gamma_Z)$. If h_p is a literal u_i then the literal argument $r = \neg u_i \in In(\gamma_Z)$ attacks p; if h_p is an $\wedge$-gate then $h_p \in False(\gamma_Z)$ implies there is some input h_r to h_p with $h_r \in False(\gamma_Z)$, so that $r = \neg h_r$ is in $In(\gamma_Z)$ (whether h_r is an $\vee$-gate or literal) and r attacks p. Similarly, if g_q is an $\wedge$-gate then $g_q \in True(\gamma_Z)$ and $h_p \in True(\gamma_Z)$. If h_p is a literal u_i then the attacking argument (on q in $\mathcal{R}_\Phi$) is the literal $\neg u_i \in Out(\gamma_Z)$, thus $r = u_i \in In(\gamma_Z)$ provides a counter-attack on p. If h_p is an $\vee$-gate then $h_p \in True(\gamma_Z)$ indicates that some input h_r of h_p is in $True(\gamma_Z)$, so that $r = h_r$ is in $In(\gamma_Z)$ and r attacks p. No more cases remain thus $In(\gamma_Z)$ is admissible. □

Lemma 10. $\forall \gamma_Z$ $In(\gamma_Z) \in \mathcal{PE}^{\mathcal{R}}$.

Proof. From Lemmas 8, 9 and the fact that every argument in $\mathcal{X}_\Phi$ is allocated to either $In(\gamma_Z)$ or $Out(\gamma_Z)$ by γ_Z, cf. Fact 6(a), it suffices to show that for any argument $p \in Out(\gamma_Z)$ there is some $q \in In(\gamma_Z)$ such that p and q conflict. Certainly this is the case for literal arguments, $u \in Out(\gamma_Z)$ since the complementary literal $\neg u$ is in $In(\gamma_Z)$. Suppose $p \in Out(\gamma_Z)$ is a gate argument. If g_p is an $\vee$-gate then $p \in Out(\gamma_Z)$ implies $g_p \in True(\gamma_Z)$ and hence some input h_q of g_p must be in $True(\gamma_Z)$. The argument q corresponding to this input node will therefore be in $In(\gamma_Z)$. If g_p is an $\wedge$-gate then $p \in Out(\gamma_Z)$ implies $g_p \in False(\gamma_Z)$ and some input h_q of g_p must be in $False(\gamma_Z)$. The argument $\neg h_q$ will be in $In(\gamma_Z)$ and conflicts with p. □

Lemma 11. $\forall S \in \mathcal{PE}^{\mathcal{R}}$ $\exists \gamma_Z$: $S = In(\gamma_Z)$.

Proof. First observe that all $S \in \mathcal{PE}^{\mathcal{R}}$ must contain exactly n literal arguments: exactly one representative from $\{z_i, \neg z_i\}$ for each i. Let us call such a subset of the literal arguments a *representative set* and suppose that U is any representative set with S_U any preferred extension containing U. We will show that there is exactly one possible choice for S_U and that this is $S_U = In(\gamma(U))$ where $\gamma(U)$ is the instantiation of Z by: $z_i =$ **true** if $z_i \in U$; $z_i =$ **false** if $\neg z_i \in U$. Consider the following procedure that takes as input a representative set U and returns a subset $S_U \in \mathcal{PE}^{\mathcal{R}}$ with $U \subseteq S_U$.

(1) $S_U := U$; $T_U := \mathcal{X}_\Phi$.
(2) $T_U := T_U / S_U$.
(3) **if** $T_U = \emptyset$ **then return** S_U and stop.
(4) $T_U := T_U / \{q \in T_U\colon \langle p, q\rangle \in \mathcal{A}_\Phi \text{ for some } p \in S_U\}$.
(5) $S_U := S_U \cup \{q \in T_U\colon \text{ for all } p \in T_U,\ \langle p, q\rangle \notin \mathcal{A}_\Phi\}$.
(6) **goto** step (2).

We can note three properties of this procedure. Firstly, it always halts: once the literal arguments in the representative set U and their complements have been removed from T_U (in steps (2) and (4)), the directed graph-structure remaining is acyclic and thus has at least one argument that is attacked by no others. Thus each iteration of the main loop removes at least one argument from T_U which eventually becomes empty. Secondly, the set S_U *is* in $\mathcal{PE}^{\mathcal{R}}$: the initial set ($U$) is admissible and the arguments removed from T_U at each iteration are those that have just been added to S_U (step (2)) as well as those attacked by such arguments (step (4)); in addition the arguments added to S_U at each stage are those that have had counter-attacks to all potential attackers already placed in S_U. Finally for any given U the subset S_U returned by this procedure is uniquely defined. In summary, every $S \in \mathcal{PE}^{\mathcal{R}}$ is defined through exactly one representative set, U_S, and every representative set U develops to a unique $S_U \in \mathcal{PE}^{\mathcal{R}}$. Each representative set, U, however, has the form $In(\gamma(U)) \cap \{z_i, \neg z_i\colon 1 \leqslant i \leqslant n\}$, and hence the unique preferred extension, S_U, consistent with U is $In(\gamma(U))$. □

Lemma 12. *The argument system $\mathcal{R}_\Phi(\mathcal{X}_\Phi, \mathcal{A}_\Phi)$ is coherent.*

Proof. The procedure of Lemma 11 only excludes an argument, q, from the set S_U under construction if q is attacked by some argument $p \in S_U$. Thus, S_U is always a stable extension, and since Lemma 11 accounts for all $S \in \mathcal{PE}^{\mathcal{R}}$, we deduce that $\mathcal{R}_\Phi$ is coherent. □

Although our preceding three results characterise $\mathcal{R}_\Phi$ as coherent, this, in itself, does not allow $\mathcal{R}_\Phi$ be used *directly* as the transformation for instances $\Phi(X, Y)$ of QSAT_2. The overall aim is to construct an argument system from $\Phi(X, Y)$ which is coherent if and only if $\Phi(X, Y)$ is a positive instance of QSAT_2. The problem with $\mathcal{R}_\Phi$ is that, even though $\Phi(X, Y)$ may be a positive instance, there could be instantiations, $\langle \alpha_X, \beta_Y\rangle$ which *fail* to satisfy $\Phi(X, Y)$ but give rise to a stable extension $In(\alpha_X, \beta_Y)$, e.g., for β_Y with which $\Phi(\alpha_X, \beta_Y) = \textbf{false}$. In order to deal with this difficulty, we need to augment $\mathcal{R}_\Phi$ (giving a system $\mathcal{H}_\Phi$) in such a way that the admissible set $In(\alpha_X)$ is a preferred (but not stable) extension (in $\mathcal{H}_\Phi$) *only if* no instantiation β_Y allows $\langle \alpha_X, \beta_Y\rangle$ to satisfy $\Phi(X, Y)$. Thus, in our augmented system, we will have *exactly two* mutually exclusive possibilities for each total instantiation α_X of X: either there is no β_Y for which $\Phi(\alpha_X, \beta_Y) = \textbf{true}$, in which event the set $In(\alpha_X)$ will produce a non-stable preferred extension of $\mathcal{H}_\Phi$; or there is an appropriate β_Y, in which case $In(\alpha_X, \beta_Y)$ (of which $In(\alpha_X)$ is a *proper subset, cf.* Fact 6(b)) will yield a stable extension in $\mathcal{H}_\Phi$.

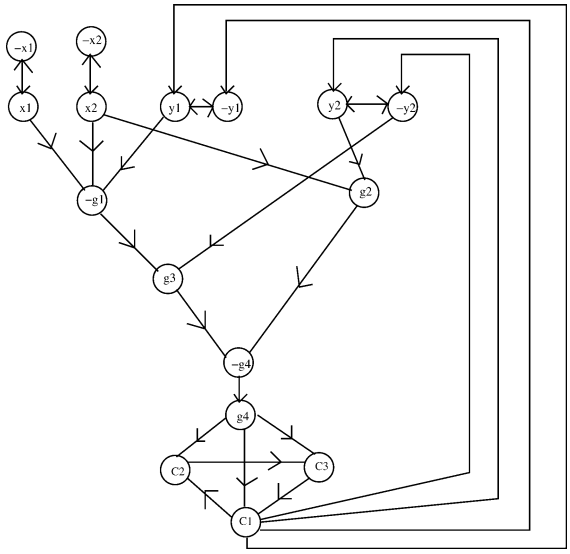

Fig. 3. An Augmented Argument Representation $\mathcal{H}_\Phi$.

Definition 13. For $\Phi(X, Y)$ an instance of QSAT_2, the *Augmented Argument Representation* of Φ—denoted $\mathcal{H}_\Phi(\mathcal{W}_\Phi, \mathcal{B}_\Phi)$—has arguments, $\mathcal{W}_\Phi = \mathcal{X}_\Phi \cup \mathcal{C}_\Phi$, where $\mathcal{X}_\Phi$ are the arguments arising in the Argument Representation of $\Phi(X, Y)$—$\mathcal{R}_\Phi$—as given in Definition 7 and $\mathcal{C}_\Phi = \{C_1, C_2, C_3\}$ are 3 new arguments called the *control arguments*. The attack relationship $\mathcal{B}_\Phi$ contains all of the attacks $\mathcal{A}_\Phi$ in the system $\mathcal{R}_\Phi$ together with new attacks,

$$\big\{\langle C_1, y_i\rangle, \langle C_1, \neg y_i\rangle\colon 1 \leqslant i \leqslant n\big\},$$
$$\big\{\langle C_1, C_2\rangle, \langle C_2, C_3\rangle, \langle C_3, C_1\rangle\big\},$$
$$\big\{\langle g_m, C_1\rangle, \langle g_m, C_2\rangle, \langle g_m, C_3\rangle\big\}.$$

Using the relabelling of variables in our example formula—Figs. 1, 2—as $\langle x_1, x_2\rangle = \langle z_1, z_2\rangle$, $\langle y_1, y_2\rangle = \langle z_3, z_4\rangle$, the Augmented Argument Representation for the system in Fig. 2 is shown in Fig. 3.

Lemma 14. *If $S \in \mathcal{PE}^{\mathcal{H}}$ then $C_i \notin S$ for any of $\{C_1, C_2, C_3\}$. If $S \in \mathcal{SE}^{\mathcal{H}}$ then $g_m \in S$.*

Proof. Suppose $S \in \mathcal{PE}^{\mathcal{H}}$. If $g_m \in S$ then each of the control arguments is attacked by g_m and so cannot be in S. If $g_m \notin S$ then $C_3 \notin S$ since the only counter-attack to C_2 is the argument C_1 which conflicts with C_3. By similar reasoning it follows that $C_2 \notin S$ and $C_1 \notin S$. For the second part of the lemma, given $S \in \mathcal{SE}^{\mathcal{H}}$, since $\{C_1, C_2, C_3\} \not\subseteq S$, there must be some attacker of these in S. The only choice for this attacker is g_m. □

Lemma 15. *$\forall\langle\alpha_X, \beta_Y\rangle$ that satisfy $\Phi(X, Y)$: $In(\alpha_X, \beta_Y) \in \mathcal{SE}^{\mathcal{H}}$.*

Proof. From Lemmas 10 and 12, the subset $In(\alpha_X, \beta_Y)$ is in $\mathcal{SE}^{\mathcal{R}}$. Furthermore, since $g_m \in True(\alpha_X, \beta_Y)$ it follows that the gate argument g_m of $\mathcal{R}_\Phi$ is in $In(\alpha_X, \beta_Y)$. For the augmented system, $\mathcal{H}_\Phi$, the arguments in $In(\alpha_X, \beta_Y)$ remain admissible: attacks on Y-literal arguments by the control argument C_1 are attacked in turn by the gate argument g_m. In addition, using the arguments of Lemma 10 no arguments in $Out(\alpha_X, \beta_Y)$ can be added to the set $In(\alpha_X, \beta_Y)$ within $\mathcal{H}_\Phi$ without conflict. Thus $In(\alpha_X, \beta_Y) \in \mathcal{SE}^{\mathcal{H}}$ whenever $\Phi(\alpha_X, \beta_Y)$ holds. □

Lemma 16. *If α_X is such that* no *instantiation β_Y of Y, leads to $\langle \alpha_X, \beta_Y \rangle$ satisfying $\Phi(X, Y)$ then $In(\alpha_X) \in \mathcal{PE}^{\mathcal{H}}/\mathcal{SE}^{\mathcal{H}}$.*

Proof. The subset $In(\alpha_X)$ of $\mathcal{R}_\Phi$ can be shown to be admissible (in both $\mathcal{R}_\Phi$ and $\mathcal{H}_\Phi$) by an argument similar to that of Lemma 9.[4] Suppose for all β_Y, we have $\Phi(\alpha_X, \beta_Y) = \textbf{false}$, and consider any subset S of $\mathcal{W}_\Phi$ in $\mathcal{H}_\Phi$ for which $In(\alpha_X) \subset S$. We show that $S \notin \mathcal{PE}^{\mathcal{H}}$. Assume the contrary holds. From Lemma 14 no control argument is in S. If $g_m \in S$ then S must contain a *representative set*, V_Y say, of the Y-literal arguments matching some instantiation β_Y. From the argument used to prove Lemma 11, $In(\alpha_X, \beta_Y)$ is the only preferred extension in $\mathcal{R}_\Phi$ consistent with the literal choices indicated by α_X and β_Y, and thus would be the only such possibility for $\mathcal{H}_\Phi$. Now we obtain a contradiction since $g_m \notin In(\alpha_X, \beta_Y)$ (in either system), and so cannot be used in $\mathcal{H}_\Phi$ to counter the attack by C_1 on the representative set V_Y. Thus we can assume that $g_m \notin S$. From this it follows that no Y-literal argument is in S (as g_m is the only attacker of the control argument C_1 which attacks Y-literals). Now consider the gates in T_Φ topologically sorted, i.e., assigned a number $1 \leqslant \kappa(g) \leqslant m$ such that all of the inputs for a gate numbered $\kappa(g)$ are from literals or gates h with $\kappa(h) < \kappa(g)$. Let q be an argument such that g_q is the first gate in this topological ordering for which $q \in S/In(\alpha_X)$. We must have $g_q \in Open(\alpha_X)$ otherwise—i.e., $q \in Out(\alpha_X)$—q would already be excluded from any admissible set having $In(\alpha_X)$ as a subset. Consider the set of arguments in $\mathcal{W}_\Phi$ that attack q. At least one attacker, p, must be a node h_p in T_Φ for which $h_p \in Open(\alpha_X)$. Now our proof is completed: S has no available counter-attack to the attack by p on q since such could only arise from a Y-literal argument (all of which have been excluded) or from another gate argument r with $g_r \in Open(\alpha_X)$, however, $\kappa(g_r) < \kappa(h_p) < \kappa(g_q)$ and $r \in S$ contradicts the choice of q. Fig. 4 illustrates the possibilities. We conclude that the subset $In(\alpha_X)$ of $\mathcal{W}_\Phi$ is in $\mathcal{PE}^{\mathcal{H}}$ whenever there is no β_Y with which $\Phi(\alpha_X, \beta_Y) = \textbf{true}$, and since the control arguments are not attacked, $In(\alpha_X) \notin \mathcal{SE}^{\mathcal{H}}$. □

Lemma 17. *If $S \in \mathcal{SE}^{\mathcal{H}}$ then $S = In(\alpha_X, \beta_Y)$ (with $\Phi(\alpha_X, \beta_Y) = \textbf{true}$). If $S \in \mathcal{PE}^{\mathcal{H}}/\mathcal{SE}^{\mathcal{H}}$ then $S = In(\alpha_X)$ and $\Phi(\alpha_X, \beta_Y) = \textbf{false}$ for all β_Y.*

[4] A minor addition is required in that since α_X is a partial instantiation (of $\langle X, Y \rangle$) it has to be shown that all arguments p that attack arguments $q \in In(\alpha_X)$ belong to the subset $Out(\alpha_X)$, i.e., are not in $Poss(\alpha_X)$. With the generalisation of $\wedge$ and $\vee$ to allow unassigned values, it is not difficult to show that if $p \in Poss(\alpha_X)$ then any argument q attacked by p in $\mathcal{R}_\Phi$ cannot belong to $In(\alpha_X)$.

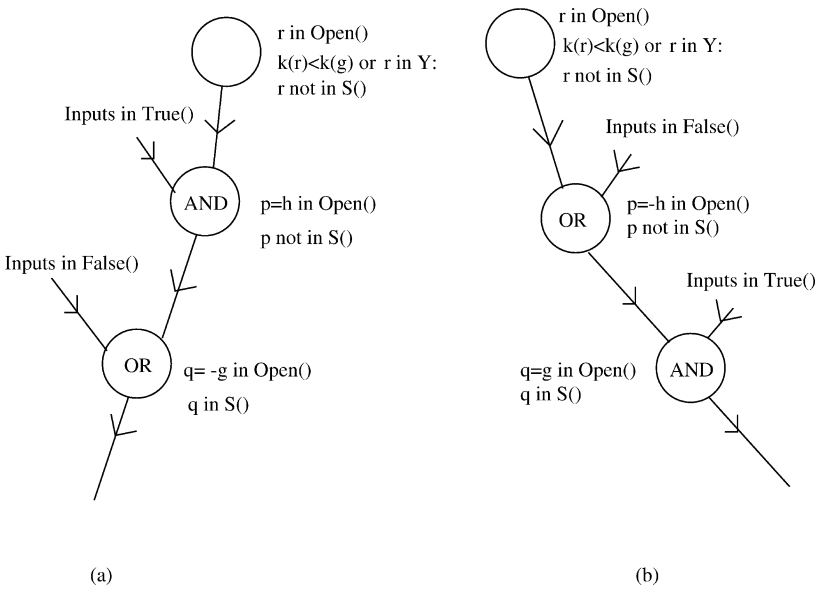

Fig. 4. Final cases in the proof of Lemma 16: $q \in Poss(\alpha_X)$ is not admissible.

Proof. Consider any $S \in \mathcal{PE}^{\mathcal{H}}$. It is certainly the case that S has as a subset some representative set, V_X from the X-literal arguments. Suppose we modify the procedure described in the proof of Lemma 11, to one which takes as input a representative set V of the X-literals and returns a subset S_V of the arguments $\mathcal{W}_\Phi$ of $\mathcal{H}_\Phi$ in the following way:

(1) $S_V := V$; $newT_V := \mathcal{W}_\Phi$;
(2) $oldT_V := newT_V$; $newT_V := oldT_V / S_V$;
(3) **if** $newT_V = oldT_V$ **then return** S_V and stop;
(4) $newT_V := newT_V / \{q \in newT_V\colon \langle p, q\rangle \in \mathcal{B}_\Phi$ for some $p \in S_V\}$;
(5) $S_V := S_V \cup \{q \in newT_V\colon$ for all $p \in newT_V,\ \langle p, q\rangle \notin \mathcal{B}_\Phi\}$;
(6) **goto** step (2).

The set S_V is an admissible subset of $\mathcal{W}_\Phi$ that contains only X-literal arguments and a (possibly empty) subset G of the gate arguments $\mathcal{G}_\Phi$. Furthermore, given V, there is a unique S_V returned by this procedure. It follows that for any $S \in \mathcal{PE}^{\mathcal{H}}$, $V \subseteq S \Rightarrow S_V \subseteq S$ for the representative set V associated with S. This set, V, matches the literal arguments selected by some instantiation $\alpha(V)$ of X, and so as in the proof of Lemma 11, we can deduce that $S_V = In(\alpha(V))$. This suffices to complete the proof: we have established that every set S in $\mathcal{PE}^{\mathcal{H}}$ contains a subset $In(\alpha_X)$ for some instantiation α_X: from Lemma 16, $In(\alpha_X)$ is not maximal if and only if $S = In(\alpha_X, \beta_Y)$ for some β_Y with $\Phi(\alpha_X, \beta_Y) =$ **true**. □

The proof of our main theorem is now easy to construct.

Proof of Theorem 2. It has already been shown that COHERENT$\in \Pi_2^{(p)}$ in Lemma 3. To complete the proof we need only show that $\Phi(X, Y)$ is a positive instance of QSAT_2 if and only if $\mathcal{H}_\Phi$ is coherent.

First suppose that for all instantiations α_X there is some instantiation β_Y for which $\Phi(\alpha_X, \beta_Y)$ holds. From Lemmas 15 and 17 it follows that all preferred extensions in $\mathcal{H}_\Phi$ are of the form $In(\alpha_X, \beta_Y)$, and these are all stable extensions, hence $\mathcal{H}_\Phi$ is coherent. Similarly, suppose that $\mathcal{H}_\Phi$ is coherent. Let α_X be any total instantiation of X. Suppose, by way of contradiction, that for all β_Y, $\Phi(\alpha_X, \beta_Y) = \textbf{false}$. From Lemma 16, $In(\alpha_X)$ is a preferred extension in this case, and hence (since $\mathcal{H}_\Phi$ was assumed to be coherent) a stable extension. From Lemma 14 this implies that $g_m \in In(\alpha_X)$ which could only happen if $g_m \in True(\alpha_X)$ for T_Φ, i.e., the value of Φ is determined in this case, independently of the instantiation of Y, contradicting the assumption that $\Phi(\alpha_X, \beta_Y)$ was **false** for every choice of β_Y. Thus we deduce that $\Phi(X, Y)$ is a positive instance of QSAT_2 if and only if $\mathcal{H}_\Phi$ is coherent so completing the proof that COHERENT is $\Pi_2^{(p)}$-complete. □

An easy corollary of the reduction in Theorem 2 is

Corollary 18. SA *is* $\Pi_2^{(p)}$*-complete.*

Proof. That SA$\in \Pi_2^{(p)}$ follows from the fact that x is sceptically accepted in $\mathcal{H}(\mathcal{X}, \mathcal{A})$ if and only if: for every subset S of $\mathcal{X}$ either S is not a preferred extension or x is in S. To see that SA is $\Pi_2^{(p)}$-hard, we need only observe that in order for $\mathcal{H}_\Phi$ to be coherent, the gate argument g_m must occur in in every preferred extension of $\mathcal{H}_\Phi$ in the reduction of Theorem 2 Thus, $\mathcal{H}_\Phi$ is coherent if and only if g_m is sceptically accepted in $\mathcal{H}_\Phi$. □

3. Consequences of Theorem 2 and open questions

A number of authors have recently considered mechanisms for establishing credulous acceptance of an argument p in a finitely presented system $\mathcal{H}(\mathcal{X}, \mathcal{A})$ through *dialogue games*. The protocol for such games assumes two players—the *Defender* (D) and *Challenger* (C)—and prescribe a move (or *locution*) repertoire together with the criteria governing the application of moves and concepts of 'winning' or 'losing'. The typical scenario is that following D asserting p the players take alternate turns presenting counter-arguments (consistent with the structure of $\mathcal{H}$) to the argument asserted by their opponent in the previous move. A player loses when no legal move (within the game protocol) is available. An important example of such a game is the TPI-dispute formalism of [13] which provides a sound and complete basis for credulous argumentation. An abstract framework for describing such games was presented in [11], and is used in [3] also to define a game-theoretic approach to Credulous Acceptance. Coherent systems are important with respect to the game formalism of [13]: TPI-disputes define a sound and complete proof theory for both Sceptical and Credulous games on coherent argument systems; the Sceptical Game is not, however, complete in the case of incoherent systems. The sequence of moves describing a completed Credulous Game (for both [3,13]) can be interpreted as certificates

of admissibility or inadmissibility for the argument disputed. It may be noted that this view makes apparent a computational difficulty arising in attempting to define similar 'Sceptical Games' applicable to incoherent systems: the shortest certificate that $\mathrm{CA}(\mathcal{H}, x)$ holds, is the size of the smallest admissible set containing x—it is shown in [10] that there is always a strategy for D that can achieve this; it is also shown in [10] that TPI-disputes won by C, i.e., certificates that $\neg\mathrm{CA}(\mathcal{H}, x)$, can require exponentially many (in $|\mathcal{X}|$) moves.[5] If we consider a sound and complete dialogue game for *sceptical* reasoning, then the moves of a dispute won by D constitute a certificate of membership in a $\Pi_2^{(p)}$-complete language: we would expect such certificates 'in general' to have exponential length; similarly, the moves in a dispute won by C constitute a certificate of membership in a $\Sigma_2^{(p)}$-complete language and again these are 'likely' to be exponentially long. Thus a further motivation of coherent systems is that sceptical acceptance is 'at worst' CO-NP-complete: short certificates that an argument is *not* sceptically accepted always exist.

The fact that sceptical acceptance is 'easier' to decide for coherent argument systems, raises the question of whether there are efficiently testable properties that can be exploited in establishing coherence. The following is not difficult to prove:

Fact 19. *If $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is not coherent then it contains a* (*simple*) *directed cycle of* odd *length.*

Thus an absence of odd cycles (a property which can be efficiently decided) ensures that the system is coherent. An open issue concerns coherence in *random* systems. One consequence of [4] is that random argument systems of n arguments in which each attack occurs (independently) with probability p, almost surely have a stable extension when p is a fixed probability in the range $0 \leqslant p \leqslant 1$. Whether a similar result can be proven for coherence is open.

As a final point, we observe that the interaction between graph-theoretic models of argument systems and propositional formulae may well provide a fruitful source of further techniques. We noted earlier that [7] provides a translation from CNF-formulae, Φ into an argument system $\mathcal{H}_\Phi$; our constructions above define similar translations for arbitrary propositional formulae. We can equally, however, consider translations in the reverse direction, e.g., given $\mathcal{H}(\mathcal{X}, \mathcal{A})$ it is not difficult to see that the CNF-formula, $\Phi_\mathcal{H} = \bigwedge_{\langle x,y\rangle\in\mathcal{A}}(\neg x \vee \neg y) \wedge \bigwedge_{x\in\mathcal{X}}(x \vee \bigvee_{\{z:\ \langle z,x\rangle\in\mathcal{A}\}} z)$ is satisfiable if and only $\mathcal{H}$ has a stable extension. Similar encodings can be given for many of the decision problems of Table 1. Translating such forms *back* to argument systems, in effect gives an alternative formulation of the original argument system from which they were generated, and thus these provide mechanisms whereby any system, $\mathcal{H}$ can be translated into another system $\mathcal{H}_{dec}$ with properties of concern holding of $\mathcal{H}$ if and only if related properties hold in $\mathcal{H}_{dec}$. Potentially this may permit both established methodologies from classical propositional logic[6] and graph-theory to be imported as techniques in argumentation.

[5] Since these are certificates of membership in a CO-NP-complete language, this is unsurprising: [10] relates dispute lengths for such instances to the length of validity proofs in the CUT-free Gentzen calculus.

[6] Translations from non-classical logics into propositional forms have also been considered in a more general setting in work of Ben-Eliyahu and Dechter [1].

4. Conclusion

In this article the complexity of deciding whether a finitely presented argument system is coherent has been considered and shown to be $\Pi_2^{(p)}$-complete, employing techniques based entirely around the directed graph representation of an argument system. An important property of coherent systems is that sound and complete methods for establishing credulous acceptance adapt readily to provide similar methods for deciding sceptical acceptance, hence sceptical acceptance in coherent systems is CO-NP-complete. In contrast, as an easy corollary of our main result it can be shown that sceptical acceptance is $\Pi_2^{(p)}$-complete in general. Finally we have outlined some directions by which the relationship between argument systems, propositional formulae, and graph-theoretic concepts offers potential for further research.

References

[1] R. Ben-Eliyahu, R. Dechter, Default reasoning using classical logic, Artificial Intelligence 84 (1–2) (1996) 113–150.
[2] A. Bondarenko, P.M. Dung, R.A. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artificial Intelligence 93 (1–2) (1997) 63–101.
[3] C. Cayrol, S. Doutre, J. Mengin, Dialectical proof theories for the credulous preferred semantics of argumentation frameworks, in: Proc. Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001), in: Lecture Notes in Artificial Intelligence, Vol. 2143, Springer, Berlin, 2001, pp. 668–679.
[4] W. Fernandez de la Vega, Kernels in random graphs, Discrete Math. 82 (1990) 213–217.
[5] Y. Dimopoulos, B. Nebel, F. Toni, Preferred arguments are harder to compute than stable extensions, in: T. Dean (Ed.), Proc. IJCAI-99, Stockholm, Sweden, Vol. 1, Morgan Kaufmann, San Francisco, CA, 1999, pp. 36–43.
[6] Y. Dimopoulos, B. Nebel, F. Toni, Finding admissible and preferred arguments can be very hard, in: A.G. Cohn, F. Giunchiglia, B. Selman (Eds.), Principles of Knowledge Representation and Reasoning (KR-2000), Morgan Kaufmann, San Francisco, CA, 2000, pp. 53–61.
[7] Y. Dimopoulos, A. Torres, Graph theoretical structures in logic programs and default theories, Theoret. Comput. Sci. 170 (1996) 209–244.
[8] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and N-person games, Artificial Intelligence 77 (1995) 321–357.
[9] P.E. Dunne, The Complexity of Boolean Networks, Academic Press, New York, 1988.
[10] P.E. Dunne, T.J.M. Bench-Capon, Two party immediate response disputes: Properties and efficiency, Technical Report, Department of Computer Science, Univ. of Liverpool, http://www.csc.liv.ac.uk/~ped/papers/tpi.ps, October 2001, submitted.
[11] H. Jakobovits, D. Vermeir, Dialectic semantics for argumentation frameworks, in: Proc. Seventh International Conference on Artificial Intelligence and Law (ICAIL-99), ACM SIGART, ACM Press, New York, 1999, pp. 53–62.
[12] H. Prakken, Logical Tools for Modelling Legal Argument, Kluwer Academic, Dordrecht, 1997.
[13] G. Vreeswijk, H. Prakken, Credulous and sceptical argument games for preferred semantics, in: Proc. JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence, in: Lecture Notes in Artificial Intelligence, Vol. 1919, Springer, Berlin, 2000, pp. 224–238.
[14] C. Wrathall, Complete sets and the polynomial-time hierarchy, Theoret. Comput. Sci. 3 (1976) 23–33.

Two Party Immediate Response Disputes: Properties and Efficiency

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Artificial Intelligence 149 (2003) 221–250

Artificial Intelligence

www.elsevier.com/locate/artint

Two party immediate response disputes: Properties and efficiency

Paul E. Dunne *, T.J.M. Bench-Capon

Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK

Received 28 June 2002

Abstract

Two Party Immediate Response Disputes (TPI-disputes) are one class of *dialogue* or *argument game* in which the protagonists take turns producing counter arguments to the 'most recent' argument advanced by their opponent. Argument games have been found useful as a means of modelling dialectical discourse and in providing semantic bases for proof theoretic aspects of reasoning. In this article we consider a formalisation of TPI-disputes in the context of finite *Argument Systems*. Our principal concern may, informally, be phrased as follows: given a specific argument system, $\mathcal{H}$, and argument x within $\mathcal{H}$, what can be stated concerning the number of moves a dispute might take for one of its protagonists to accept that x has *some* defence respectively *cannot* be defended?

Keywords: Argument systems; Dialogue game; Gentzen system; Proof complexity

1. Introduction

In this paper we are concerned with two important formalisms that have been the subject of much interest with respect to their application in modelling dialectical process: *Argument Systems* [17], and *Argument Games* [22,29]. Our principal concern is with the *length* of disputes when they are conducted in accordance with the etiquette prescribed by a particular formal protocol. The protocol of interest—TPI-dispute—was outlined in the work of [38] and in Section 1.2 we present a rigorous formalisation of this with examples of its operation being described in Section 2. The main technical concerns are dealt with in Section 3, wherein two questions are examined. Informally, these may be viewed as

* Corresponding author.
E-mail address: ped@csc.liv.ac.uk (P.E. Dunne).

0004-3702/$ – see front matter
doi:10.1016/S0004-3702(03)00076-6

follows: suppose we are presented with an argument system and an argument within this. If it is required to observe the dispute rules prescribed in some dispute protocol,

(a) when the given argument *can* be defended, how many moves *could* it take to prove to a challenging party that the argument may be defended against any attack?
(b) when the given argument *cannot* be defended against all possible attacks, how many moves *must* it take to convince putative defenders that their position is untenable?

We obtain a precise characterisation answering (a) (Theorem 4, below). In the case of (b), by developing a construction similar to that used in [16], the question is related to the widely studied issue of Proof Complexity. Specifically, we demonstrate that by representing an unsatisfiable CNF-formula, φ, as an argument system the dispute protocol defines a *proof calculus* that may be employed to show $\neg\varphi$ is a *propositional tautology*. Thus, we obtain a partial answer to (b) (in Theorem 5) by establishing that when interpreted as a calculus for Propositional Logic, the TPI-dispute protocol is 'not very powerful': formally we show that it may be efficiently simulated by a Gentzen system in which the CUT inference rule is not available.

In the remainder of this section we review the Argument System formalism from [17] and formally develop the argument game TPI-dispute, originally outlined in [38]. In Section 2 some illustrative examples of how disputes evolve in this protocol are presented. As we have already noted, Section 3 presents the core technical contribution, while Section 4 discusses some issues arising from our results and presents some directions for further work. Conclusions are given in Section 5.

1.1. Argument systems

Argument systems as a mechanism for studying formalisations of reasoning, acceptability, and defeasibility were introduced by Dung [17] and have since received considerable attention with respect to their use in non-classical logics, e.g., [8,13–15]. The basic definition of finite argument system below is derived from that given in [17].

Definition 1. An *argument system* is a pair $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$, in which $\mathcal{X}$ is a set of *arguments* and $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is the *attack relationship* for $\mathcal{H}$. A pair $\langle x, y \rangle \in \mathcal{A}$ is referred to as 'y *is attacked by* x' or 'x *attacks* (or *is an attacker of*) y'. The *range* of an argument x—denoted *range*(x)—is the set of arguments that are attacked by x; the *range of a set* of arguments S, is the union over all x in S of *range*(x).

For R, S subsets of $\mathcal{X}$ in $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$, we say that

(a) $s \in S$ is *attacked* by R if there is some $r \in R$ such that $\langle r, s \rangle \in \mathcal{A}$.
(b) $x \in \mathcal{X}$ is *acceptable with respect to* S if for every $y \in \mathcal{X}$ that attacks x there is some $z \in S$ that attacks y.
(c) S is *conflict-free* if no argument in S is attacked by any other argument in S.
(d) A conflict-free set S is *admissible* if every argument in S is acceptable with respect to S.
(e) S is a *preferred extension* if it is a maximal (with respect to $\subseteq$) admissible set.
(f) S is a *stable extension* if S is conflict free and every argument $y \notin S$ is attacked by S.

While some argument systems may not have any stable extension, it is always the case that *some* preferred extension is present: the reason being that the empty set is always admissible.

Definition 2. The decision problem *Credulous Acceptance* (CA) takes as an instance: an argument system $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$ and an argument $x \in \mathcal{X}$. The result true is returned if and only if *at least one* preferred extension S of $\mathcal{X}$ contains x. If $\text{CA}(\langle \mathcal{H}, x \rangle)$ holds then x is said to be *credulously accepted* in $\mathcal{H}$.

The decision problem *Sceptical Acceptance* (SA) takes as an instance: an argument system $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$ and an argument $x \in \mathcal{X}$. The result true is returned if and only if *every* preferred extension S in $\mathcal{X}$ contains x. If $\text{SA}(\langle \mathcal{H}, x \rangle)$ holds x is said to be *sceptically accepted* in $\mathcal{H}$.

1.2. Argument games and TPI*-disputes*

A widely studied concept that has received some attention in the context of argument systems is that of employing *argument games* both as models of dialectical discourse and as a basis for a formal proof theory. The form of such games involves a sequence of interactions between two protagonists—hereafter referred to as the *Defender* (D) and *Challenger* (C)—wherein the Defender attempts to establish a particular argument in the face of counterarguments advanced by the Challenger, see, e.g., [10,24,26,29,37]. In [38] descriptions of games—*Two Party Immediate Response Disputes* (TPI-disputes)—are presented for Credulous and Sceptical Argument within the framework considered in the present article. We consider a rather more tightly specified definition of TPI-disputes: the form presented in [38] defines notions of move, attack, winning and losing within a dispute. These, however, are illustrated through a series of examples rather than presenting a precise semantics for the game as a whole. Our main point of interest concerns the fact that whilst such games always terminate for finitely specified systems we wish to address how many steps (as a function of $|\mathcal{X}|$) some disputes may take.

We begin by developing the idea of TPI-disputes, using as a basis the informal schema of [38]. In informal terms, a TPI-dispute starts from a named argument, x in a given argument system $\mathcal{H}$. For the *Credulous Game*, a defender attempts to construct an admissible set containing x. For a select class of Argument Systems,[1] *Sceptical Acceptance* can be established by the Defender proving that no attacker of x is credulously accepted. The Challenger's aim is to prevent successful construction. The game proceeds by the players alternately presenting arguments within $\mathcal{H}$ that attack the previous arguments proposed by the other player. The concept of *immediate response* concerns the requirement in the game for both players to identify arguments that attack the most recent argument put forward by the opponent. A number of examples given in [38] indicate that both players must have the capability of 'back-tracking', e.g., if the line of attack followed by the Challenger fails, it must be possible to adopt a different attack on some previous argument.

We can view the progress of such disputes as a sequence of *directed trees* each of which is constructed by a *depth-first* expansion, the root of each tree being the argument x at the

[1] But not *all*, cf. Theorem 3, and Fig. 1 subsequently.

heart of the dispute. In this way the game is characterised by the *moves* through which a tree is expanded and the rules which force back-tracking by either party.

1.2.1. A model of TPI*-disputes*

Definition 3. Let $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$ be an argument system and x an argument in $\mathcal{X}$. A *dispute tree for x in $\mathcal{H}$*, $T_x^{\mathcal{H}}$, is a tree whose vertices are a subset of $\mathcal{X}$ and whose root is x. The edges of a dispute tree are directed *from* vertices *to* their parent vertex. If t is a leaf vertex in $T_x^{\mathcal{H}}$ the path

$$t = v_k \to v_{k-1} \to \cdots \to v_2 \to v_1 \to v_0 = x$$

is called a *dispute line*.

A dispute line (*to* v) is a *failing attack on* x if the number of vertices on the path from v up to (and including) x is *odd*. A dispute line is a *failing defence of* x if this number is *even*.

A vertex, v, is *open* in $T_x^{\mathcal{H}}$ if there is an argument, w in $\mathcal{X}$, which attacks v and is '*available*' (in a sense which is made precise below). If no such argument exists, v is *closed*. A dispute line is closed or open according to whether its leaf vertex is closed or open.

Given a system $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$ and $x \in \mathcal{X}$ a TPI-dispute consists of a sequence of moves

$$M = \langle \mu_1, \mu_2, \ldots, \mu_i, \ldots \rangle.$$

Moves, μ are chosen from a finite *repertoire* of *move types*, some (or all) of which may not be available (depending on the current 'state' of a dispute). This *state* is represented after the kth move ($k \geqslant 0$), by a tuple $\sigma_k = \langle T_k, v_k, \Delta_k, \Gamma_k, P_k, Q_k \rangle$. Here

T_k: the dispute tree after k moves;

v_k: the 'current' argument (vertex of) T_k;

Δ_k: arguments available to D;

Γ_k: arguments available to C;

P_k: arguments proposed as a (subset) of some admissible set by D;

Q_k: the set of *subsets* of arguments that C has *shown* not to be a subset of an admissible set.

The initial state (σ_0) is $\langle \langle x \rangle, x, \Delta_0, \Gamma_0, P_0, Q_0 \rangle$ where

$$\Delta_0 = \mathcal{X} / \big(\{x\} \cup \{y\colon \langle x, y \rangle \in \mathcal{A} \text{ or } \langle y, x \rangle \in \mathcal{A}\} \big),$$

$$\Gamma_0 = \mathcal{X} / \big(\{x\} \cup \mathit{range}(x) \big),$$

$$P_0 = \{x\},$$

$$Q_0 = \emptyset.$$

A dispute, $M = \langle \mu_1, \mu_2, \ldots, \mu_k \rangle$, is *active* if there is a legal move μ_{k+1} available to the current player, i.e., C if $k+1$ is odd, D otherwise. A dispute, M, is *terminated* if M is *not* active. For a terminated dispute, we use $|M|$ to denote the number of *moves* in M.

In informal terms the 'state' describes the progress so far of a dispute over the argument x. The defender is attempting to construct in the subset P_k an admissible set

containing x. In order to achieve this, D, has to respond to attacks put forward by C so that (if k is odd), the argument v_k requires D to employ an 'available' argument in Δ_k to attack v_k: the chosen argument will form the component v_{k+1} of the next state. The Challenger in attempting to show that x is not credulously accepted maintains a set of subsets of $\mathcal{X}$ (the set Q_k) comprising subsets that cannot form part of an admissible set with x.

Before defining the move repertoire we outline the notions of '*availability*' that are used. Suppose D must find an argument z with which to attack v_k proposed by C, i.e., with $\langle z, v_k\rangle \in \mathcal{A}$. Since D aims to construct an admissible set, certainly any z that conflicts with any argument in P_k cannot be used—P_k must be conflict-free. In addition, should z be such that $P_k \cup \{z\}$ has already been shown not to be an admissible set, i.e., for some $S \in Q_k$ it holds that $S \subseteq P_k \cup \{z\}$, then z cannot be used to counter-attack v_k. Thus, in summary, an argument is 'available' to D if it attacks the most recent argument put forward by C, does not conflict with any argument that D is currently defending *and* can be added to this set without forming a 'known' inadmissible set. Similarly, C, in finding a counterattack to v_k needs to identify some z that attacks v_k and *is not attacked* by any argument in P_k. Thus the 'available' arguments for C at any point are simply those that are not attacked by any argument in P_k.

A detailed description of how the sets of available arguments develop between moves is given when we describe the move repertoire.

1.2.2. The move repertoire

It remains to describe the move repertoire, conditions determining applicability, and consequent changes to σ_{i-1} after performing a move μ_i.

The various implementations of argument games allow a variety of different moves. Some, such as [25], provide a small number of basic moves, intended to model disputes in a generic manner, while others allow a larger number in order to attempt to reflect the moves made by the participants in particular kinds of dispute, e.g., [22] or to reflect particular notions of what constitutes an argument. For example Bench-Capon [6] models arguments as described by Toulmin [34]. Since our framework uses Dung's very abstract notion of argument [17], we do not need moves to reflect particular procedures or forms of argument, and so can use a rather small set of moves.

The repertoire of moves[2] we allow comprises just,

$$\{\text{COUNTER}, \text{BACKUP}, \text{RETRACT}\}.$$

The first move can be made by either player, whereas BACKUP is only employed by C, and RETRACT only by D. These two moves arise from the need to allow back-tracking. In the description that follows it should be remembered that *odd* indexed moves are made by the Challenger and *even* indexed moves by the Defender.

$\mu_k = \text{COUNTER}(y)$. Let $\sigma_{k-1} = \langle T_{k-1}, v_{k-1}, \Delta_{k-1}, \Gamma_{k-1}, P_{k-1}, Q_{k-1}\rangle$. If k is odd, μ_k is made by C, and $\text{COUNTER}(y)$ can be applied only if $\langle y, v_{k-1}\rangle \in \mathcal{A}$ and $y \in \Gamma_{k-1}$, i.e., y attacks the current argument (v_{k-1}) and is available. The new state, σ_k, is now

[2] The terminology we use is not employed in [38] which is given simply in terms of attacking moves and back-tracking.

$$T_k := T_{k-1} + \langle y, v_{k-1}\rangle,$$
$$v_k := y,$$
$$\Delta_k := \Delta_{k-1},$$
$$\Gamma_k := \Gamma_{k-1}/\{y\},$$
$$P_k := P_{k-1},$$
$$Q_k := Q_{k-1}.$$

If k is even, so that μ_k is made by D, then COUNTER(y) can be applied only if: $y \in \Delta_{k-1}$; $\langle y, v_{k-1}\rangle \in \mathcal{A}$; and for each set R in Q_{k-1}, R is not contained in $P_{k-1} \cup \{y\}$, i.e., D has available an argument y with which to attack v_{k-1} and, if y is added to the set of arguments that D is (currently) committed to then the resulting set has *not* been ruled inadmissible earlier.

The new state, σ_k is now

$$T_k := T_{k-1} + \langle y, v_{k-1}\rangle,$$
$$v_k := y,$$
$$\Delta_k := \Delta_{k-1}/\big(\{y\} \cup \big\{z \in \Delta_{k-1}\colon \langle y, z\rangle \in \mathcal{A} \text{ or } \langle z, y\rangle \in \mathcal{A}\big\}\big),$$
$$\Gamma_k := \Gamma_{k-1}/\big(\{y\} \cup \mathit{range}(y)\big),$$
$$P_k := P_{k-1} \cup \{y\},$$
$$Q_k := Q_{k-1}.$$

The definition of Δ_k from Δ_{k-1} and y captures the fact that D (in attempting to form an admissible set) may not violate the requirement to be conflict free. The form taken by Γ_k indicates that in adding y to its (currently) accepted arguments, D now has a defence to all arguments in Γ_{k-1} that y attacks. It follows that there is no gain in these being available to C.

$\mu_k =$ BACKUP(j, y) (where j is *even* and $0 \leqslant j \leqslant k-3$). The BACKUP move is only invoked by C and corresponds to the situation where C has no available attack with which to continue the current dispute line. The BACKUP move returns the dispute to the *most recent* point (σ_j) from which C can mount a fresh attack. Thus, if the (currently open) dispute line is,

$$L_{k-1} = \langle v_{k-1} \to v_{k-2} \to \cdots \to v_{j+1} \to v_j \to \cdots \to v_2 \to v_1 \to v_0\rangle$$

then

BC1. L_{k-1} is a closed failing attack, i.e., there are no arguments $z \in \Gamma_{k-1}$ for which $\langle z, v_{k-1}\rangle \in \mathcal{A}$.

BC2. For each r in the set $\{j+2, j+4, j+6, \ldots, k-3\}$ there are no arguments

$$z \in \Gamma_r/\big(\{v_r, v_{r+1}, v_{r+2}, \ldots, v_{k-2}\} \cup \mathit{range}\big(\{v_r, v_{r+2}, \ldots, v_{k-3}\}\big)\big)$$

for which $\langle z, v_r\rangle \in \mathcal{A}$.

BC3. The parameters j and y specified in the move BACKUP(j, y) are such that

$$y \in \Gamma_j / \big(\{v_j, v_{j+1}, v_{j+2}, \ldots, v_{k-2}\} \cup \mathit{range}\big(\{v_j, v_{j+2}, \ldots, v_{k-3}\}\big)\big)$$

and $\langle y, v_j \rangle \in \mathcal{A}$.

In summary, the conditions for the move BACKUP(j, y) to be applicable are: C cannot continue the current dispute line since there is no argument in C's arsenal that can be used to attack the last argument proposed by D (BC1); C cannot mount a *new* line of attack on any argument put forward by D in the set $\{v_{j+2}, v_{j+4}, \ldots, v_{k-3}\}$ (BC2); C, by using y, can launch a different attack on v_j (BC3).

The new state σ_k effected by the move BACKUP(j, y) is given by:

$$\begin{aligned}
T_k &:= T_{k-1} + \langle y, v_j \rangle, \\
v_k &:= y, \\
\Delta_k &:= \Delta_{k-1}, \\
\Gamma_k &:= \Gamma_j / \big(\{y, v_{j+1}, v_{j+2}, \ldots, v_{k-1}\} \cup \mathit{range}\big(\{v_{j+2}, v_{j+4}, \ldots, v_{k-1}\}\big)\big), \\
P_k &:= P_{k-1}, \\
Q_k &:= Q_{k-1}.
\end{aligned}$$

Note that Δ_k does *not* revert to its content at the 'backup' position Δ_j: D has 'committed' to defending these in order to force C to adopt a new line of dispute. Secondly, the set, Γ_k, of available arguments for C, has all of the arguments advanced in progressing from v_{j+1} to v_{k-3} removed (rather than simply the 'old' attack v_{j+1} and the 'new' attack y on v_j): since D has already established a suitable line of defence to each of these, their only utility to the challenger would be in *prolonging* a dispute, rather than winning it.

$\mu_k =$ RETRACT. The RETRACT move is only made by D. Suppose

$$\sigma_{k-1} = \langle T_{k-1}, v_{k-1}, \Delta_{k-1}, \Gamma_{k-1}, P_{k-1}, Q_{k-1} \rangle$$

is the current state (so that $k-1$ is odd). For RETRACT to be applicable D must have no available attack on v_{k-1} *and* $P_{k-1} \neq \{x\}$. In this case, the Challenger has succeeded in showing that the set P_{k-1} cannot be extended to form an admissible set. Thus the only option available to the Defender is to try constructing a new admissible set containing x. Formally, the next state σ_k is given as

$$\begin{aligned}
T_k &:= \langle x \rangle, \\
v_k &:= x, \\
\Delta_k &:= \Delta_0, \\
\Gamma_k &:= \Gamma_0, \\
P_k &:= P_0, \\
Q_k &:= Q_{k-1} \cup \{P_{k-1}\}.
\end{aligned}$$

1.2.3. Discussion

The main point that should be noted is the asymmetry concerning BACKUP and RETRACT. Firstly, BACKUP may be seen as the Challenger invoking a *new line of attack* within the *same* dispute tree. On the other hand, RETRACT represents the dispute over x being *started again*, this time, however, with the knowledge that some lines of defence are not available, i.e., those that would result in a 'known' inadmissible set being constructed. Of course, as will be shown later, if x is *credulously accepted* then D, employing 'best play' will never need to make a retraction. In defining the game rules, however, we cannot assume that D will play 'intelligently' and thus may, inadvertently, call upon arguments that are eventually exposed as collectively indefensible. It may be observed that the position from which the dispute is resumed (following a retraction) is the *opening* dispute tree: while, in principle, one could define the next dispute tree to result from some variant of the current one, such an approach affords no significant gain.

1.2.4. Credulous and sceptical games

Definition 4. Let $M_{\langle\mathcal{H},x\rangle} = \langle\mu_1, \mu_2, \ldots, \mu_k\rangle$ be a *terminated* TPI-dispute over an argument x in the argument system $\mathcal{H}$. $M_{\langle\mathcal{H},x\rangle}$ is a *successful* (*credulous*) *defence of* x if k is *even*, and a *successful rebuttal of* x if k is *odd*.

The following result reformulates Proposition 1 of [38] in terms of the formal framework introduced above.

Theorem 1. $\text{CA}(\mathcal{H}, x) \Leftrightarrow (\exists M_{\langle\mathcal{H},x\rangle}\colon M_{\langle\mathcal{H},x\rangle}$ *is a successful defence of* $x)$.

Proof. First suppose that $\text{CA}(\mathcal{H}(\langle\mathcal{X}, \mathcal{A}\rangle, x)$ holds, i.e., that x is credulously accepted in $\mathcal{H}$. Consider any admissible set, S_x, of $\mathcal{H}$ containing x. It is certainly the case that using *only* the arguments in S_x, D can always COUNTER attacks available to C (recall that in replying to $\text{COUNTER}(y)$ from C the response $\text{COUNTER}(z)$ will remove from C's arsenal of attacks any argument attacked by z). Furthermore, D never has to invoke the RETRACT move. It follows that such a dispute will eventually terminate with C having no further move, i.e., as a successful defence of x.

Conversely, suppose that $M_{\langle\mathcal{H},x\rangle}$ is a successful defence of x. Consider the set P_k pertaining after μ_k the final move of the dispute. It is certainly the case that $x \in P_k$ (since this holds throughout the dispute). In addition, P_k is conflict-free (since Δ_j never makes available to D, arguments that conflict with those in P_j). Finally, since C has no move available, every attack on arguments $y \in P_k$ must have been countered, i.e., is defended by some $p \in P_k$. The three properties just identified establish that P_k is an admissible set containing x, hence x is credulously accepted. □

Theorem 2. *For all* TPI*-dispute instances,* $\langle\mathcal{H}, x\rangle$ *either* all *terminated* $M_{\langle\mathcal{H},x\rangle}$ *are successful defences of* x *or all are successful rebuttals.*

Proof. Suppose the contrary and

$$M^{(1)} = \langle \mu_1^1, \mu_2^1, \ldots, \mu_m^1 \rangle \quad \text{with } \sigma_m^1 = \langle T_m^1, v_m^1, \Delta_m^1, \Gamma_m^1, P_m^1, Q_m^1 \rangle,$$
$$M^{(2)} = \langle \mu_1^2, \mu_2^2, \ldots, \mu_n^2 \rangle \quad \text{with } \sigma_n^2 = \langle T_n^2, v_n^2, \Delta_n^2, \Gamma_n^2, P_n^2, Q_n^2 \rangle$$

are different TPI-disputes with $M^{(1)}$ a successful defence of x and $M^{(2)}$ a successful rebuttal of x within $\mathcal{H}$. Since $M^{(1)}$ is a successful defence, the subset P_m^1 is an admissible set (containing x). If $M^{(2)}$ is a successful rebuttal of x, then D must reach the point where no RETRACT move is applicable. Consider the admissible set, P_m^1 found by $M^{(1)}$ and the first move t at which some $Q \subseteq P_m^1$ is added to Q_{t-1}^2. It must be the case that $\mu_t^2 = \text{RETRACT}$ (or $t = n+1$) and that D has no available defence with which to counter v_{t-1}^2. Now we derive a contradiction: v_{t-1}^2 attacks $y \in P_{t-2}^2 = Q \subseteq P_m^1$ and the progress of $M^{(2)}$ has left no counter attack on v_{t-1}^2 available to D. On the other hand, such a defence (z, say) *is* present in P_m^1 since it is an admissible set and z would only be unavailable if it attacked or was attacked by Q, contradicting the fact that P_m^1 (of which Q is a subset) must be conflict-free. □

Definition 5. For an argument system $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$ and $x \in \mathcal{X}$, the *x-augmented system*, $\mathcal{H}_x$ is the system formed by adding a new argument $\{x_a\}$ to $\mathcal{X}$ together with attack $\{\langle x, x_a \rangle\}$.

The following reformulates Proposition 2 of [38].

Theorem 3. *Let $\mathcal{H}$ be an argument system in which* every *preferred extension is also a stable extension and let x be an argument in $\mathcal{H}$.*[3] *The argument x is* sceptically accepted *in $\mathcal{H}$ if and only if, there is a dispute, M, providing a successful rebuttal of x_a in the x-augmented system $\mathcal{H}_x$.*

Proof. Let $\mathcal{H}$ be an argument system in which every preferred extension is stable. First suppose that x is sceptically accepted in $\mathcal{H}$, the first part of the theorem will follow (via Theorem 1) by showing that x_a is not credulously accepted in the x-augmented system. Suppose the contrary and that $S_a \subset \mathcal{X} \cup \{x_a\}$ is a preferred extension in $\mathcal{H}_x$ that contains x_a. The set S_a cannot contain x, and must contain at least one attacker of x. The set, $S_a/\{x_a\}$, however, is an admissible set in $\mathcal{H}$ and cannot be developed to a preferred extension containing x. This contradicts the premise that x is sceptically accepted in $\mathcal{H}$.

Conversely, suppose that x_a is not credulously accepted in the x-augmented system $\mathcal{H}_x$. Consider any preferred extension S of $\mathcal{H}$. Suppose $x \notin S$. Since S is a stable extension, there is some attacker, y, of x, in S and since y attacks x which is the only attack on x_a in the x-augmented system, we deduce that $S \cup \{x_a\}$ would form a preferred extension in $\mathcal{H}_x$ contradicting the premise that x_a is not credulously accepted. □

The example in Fig. 1 is adapted from [38], and shows that the stability condition is needed. In this example of an x-augmented system, x_a is not in any preferred extension since there is no defence to the attack by x (y is inadmissible since it is effectively self-attacking). Within the original system, however, x is *not* sceptically accepted: there are

[3] Argument systems satisfying this condition are termed *coherent* in [17, Definition 31(1), p. 332].

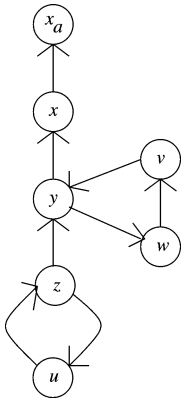

Fig. 1. x-augmented system with x_a not credulously accepted nor x sceptically accepted.

two preferred extensions—$\{x, z\}$ and $\{u\}$—the latter containing neither x nor its attacker y. We note that testing if an argument system is coherent, i.e., every preferred extension is also stable, is likely to be difficult: Dunne and Bench-Capon [19] having demonstrated this to be Π_2^p-complete, although there is an efficiently decidable property that guarantees coherence.

2. Examples

In order to clarify how particular disputes develop we give two examples based on the argument systems, shown in Fig. 2. It may be observed that the system in Fig. 2(b) can be interpreted as a representation of the tautology,

$$\neg F(y, z) = \neg\big((y \vee z) \wedge (y \vee \neg z) \wedge (\neg y \vee z) \wedge (\neg y \vee \neg z)\big) \tag{1}$$

and so serves to illustrate dispute progression for proving *credulous acceptance* of the argument $\neg F$ *and sceptical acceptance* of the same argument, i.e., that the argument F in this system is *not* credulously accepted. A general translation from CNF formulae to argument systems will be given in Definition 7.

For Fig. 2(a) one possible TPI-dispute over x (in which we abbreviate COUNTER, BACKUP, and RETRACT to C,B,R) is,

k	μ_k	v_k	Δ_k	Γ_k	P_k	Q_k
0	–	x	$\{u, v, w\}$	$\{y, z, u, v, w\}$	$\{x\}$	$\emptyset$
1	C(y)	y	$\{u, v, w\}$	$\{z, u, v, w\}$	$\{x\}$	$\emptyset$
2	C(v)	v	$\{u\}$	$\{z, u\}$	$\{x, v\}$	$\emptyset$
3	B$(0, z)$	z	$\{u\}$	$\{u\}$	$\{x, v\}$	$\emptyset$
4	R	x	$\{u, v, w\}$	$\{y, z, u, v, w\}$	$\{x\}$	$\{x, v\}$
5	C(y)	y	$\{u, v, w\}$	$\{z, u, v, w\}$	$\{x\}$	$\{x, v\}$
6	C(u)	u	$\{v, w\}$	$\{z, v, w\}$	$\{x, u\}$	$\{x, v\}$
7	B$(4, z)$	z	$\{v, w\}$	$\{v, w\}$	$\{x, u\}$	$\{x, v\}$
8	C(w)	w	$\emptyset$	$\emptyset$	$\{x, u, w\}$	$\{x, v\}$

(2)

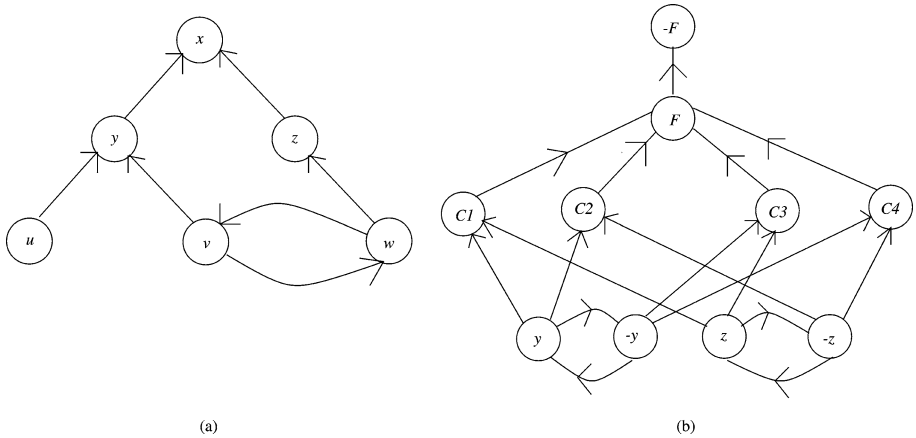

Fig. 2. Two example argument systems.

It may be observed that D, at μ_2, makes an 'incorrect' move in attacking y using v (instead of u) thus removing w from the set of available arguments and allowing C to force a retraction by attacking x with z. Of course, D could have shortened the length of the dispute by playing COUNTER(u) as the second move. As we noted earlier, the intention is to define the protocol for disputes in such a way that even if D advances what turn out to be ill-advised counter-attacks, this does not result in the game being lost since there are opportunities to correct. For Fig. 2(a) there are exactly three possible TPI-disputes over x: two in which C first counter-attacks with y, and one in which the initial counter-attack is using z.

As a final illustration we give an example of a dispute establishing sceptical acceptance of $-F$ in the system of Fig. 2(b). It is not difficult to see that this follows by showing that F is not credulously accepted, so the description is given in terms of a successful rebuttal of F;

k	μ_k	v_k	P_k	Q_k
0	–	F	$\{F\}$	$\emptyset$
1	C($C1$)	$C1$	$\{F\}$	$\emptyset$
2	C(y)	y	$\{F, y\}$	$\emptyset$
3	B($0, C3$)	$C3$	$\{F, y\}$	$\emptyset$
4	C(z)	z	$\{F, y, z\}$	$\emptyset$
5	B($0, C4$)	$C4$	$\{F, y, z\}$	$\emptyset$
6	R	F	$\{F\}$	$\{\{F, y, z\}\}$
7	C($C1$)	$C1$	$\{F\}$	$\{\{F, y, z\}\}$
8	C(z)	z	$\{F, z\}$	$\{\{F, y, z\}\}$
9	B($6, C2$)	$C2$	$\{F, z\}$	$\{\{F, y, z\}\}$
10	R	F	$\{F\}$	$\{\{F, y, z\}, \{F, z\}\}$
11	C($C1$)	$C1$	$\{F\}$	$\{\{F, y, z\}, \{F, z\}\}$

(3)

12	C(y)	y	$\{F, y\}$	$\{\{F, y, z\}, \{F, z\}\}$
13	B$(10, C3)$	$C3$	$\{F, y\}$	$\{\{F, y, z\}, \{F, z\}\}$
14	R	F	$\{F\}$	$\{\{F, y, z\}, \{F, z\}, \{F, y\}\}$
15	C$(C1)$	$C1$	$\{F\}$	$\{\{F, y, z\}, \{F, z\}, \{F, y\}\}$

and now, D cannot counter-attack $C1$ without constructing an already shown to be inadmissible set nor RETRACT since $P_{15} = \{F\}$.

3. Complexity of argument games

The preceding sections have largely been concerned with a rigorous formulation of the concept of TPI-dispute as first outlined in [38]. The principal aim of the present paper, however, is to consider the following questions.

Question 1. Given a TPI-dispute instance—$\langle \mathcal{H}, x \rangle$—such that x *is* credulously accepted in $\mathcal{H}$, how many moves are required (in the worst case) in a dispute M defining a *successful defence* of x?

Question 2. Given a TPI-dispute instance—$\langle \mathcal{H}, x \rangle$—such that x *is not* credulously accepted in $\mathcal{H}$, how many moves are necessary (in the best case) for a dispute M establishing a *successful rebuttal* of x?

In view of Theorem 3, Question 2, is of interest with respect to the number of moves required to establish *sceptical acceptance* of an argument.

In order to make these precise, we introduce the idea of *Dispute Complexity*. Given an instance of a TPI-dispute, $\langle \mathcal{H}, x \rangle$, its *dispute complexity*, denoted $\delta(\mathcal{H}, x)$ is,

$$\delta(\mathcal{H}, x) = \min_{M\text{: } M \text{ is a terminated dispute over } x \text{ in } \mathcal{H}} |M|.$$

Question 1 turns out to have a relatively straightforward characterisation using the following idea.

Definition 6. Let $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$ be an argument system and $x \in \mathcal{X}$ an argument that is credulously accepted in $\mathcal{H}$. The *rank* of x *in* $\mathcal{H}$, denoted $\rho(\mathcal{H}, x)$, is defined by

$$\min_{S \subseteq \mathcal{X}/\{x\}\text{: } S \cup \{x\} \text{ is admissible in } \mathcal{H}} |S|.$$

Theorem 4. *For any* TPI*-dispute instance—*$\langle \mathcal{H}, x \rangle$*—in which* x *is credulously accepted in* $\mathcal{H}$*,*

$$\delta(\mathcal{H}, x) = 2\rho(\mathcal{H}, x).$$

Proof. To see that $\delta(\mathcal{H}, x) \leqslant 2\rho(\mathcal{H}, x)$, consider the subset S of $\mathcal{X}$ that attains the value $\rho(\mathcal{H}, x)$. By an argument similar to that in the proof of Theorem 1, x can be defended

in a TPI-dispute, with D employing only arguments in S. Adopting such a strategy, D never needs to invoke the RETRACT move. The size of the set, P, to which D is committed increases by one with each move made by D as more members of S are added. It follows that, since S is admissible, the Challenger will have no further moves open once D has committed to every argument in S. To complete the proof we show that $\delta(\mathcal{H}, x) \geqslant 2\rho(\mathcal{H}, x)$. Consider a TPI-dispute, M, that attains $\delta(\mathcal{H}, x)$ and the dispute tree, $T_{|M|}$ that is active when the Challenger admits defeat. Certainly, $|M|$ must be at least twice the number of arguments in $T_{|M|}$ (excluding x). The arguments to which D is committed after the $|M|$th move must define an admissible set (otherwise C could continue the dispute by finding an appropriate y attacking some member of $P_{|M|}$). It follows that $|P_{|M|}/\{x\}| \geqslant \rho(\mathcal{H}, x)$ and thence $\delta(\mathcal{H}, x) \geqslant 2\rho(\mathcal{H}, x)$ as required. □

Theorem 4, in its characterisation of the answer to our first question, can be interpreted in the following way: *if* an argument x is credulously accepted in the system $\mathcal{H}$ then there is a 'short proof' of this, i.e., using the TPI-dispute that achieves $\delta(\mathcal{H}, x)$ moves. It is important to note that this does not imply that *deciding* if such a proof *exists* can be accomplished efficiently: given the results of [16][4] (from which it may be deduced that CA is NP-complete) it seems unlikely that such a decision method could be found.

For the remainder of this paper we are concerned with the second question raised. As with the view proposed in the preceding paragraph, we can interpret results concerning this question in terms of properties of the '*size*' of '*proofs*' that an argument is not credulously accepted. The decision problem CA being NP-complete, indicates that such proofs are concerned with a CO-NP-complete problem. While all NP-complete problems are such that positive instances of these have concise proofs that they *are* positive instances (this being one of the defining characteristics of the class NP as a whole) it is suspected that *no* CO-NP-complete problem has this property. In other words, we have the following (long-standing) conjecture: if L is a CO-NP-complete problem, then there are (infinitely many) instances, x of L, for which $L(x)$ is true but the 'shortest proof' of this is of length superpolynomial in the number of bits needed to encode x.[5]

The discussion above suggests that (assuming NP $\neq$ CO-NP) there must be infinitely many instances $\langle \mathcal{H}, x \rangle$ for which x is *not* credulously accepted in $\mathcal{H}$ *and* for which $\delta(\mathcal{H}, x)$—the dispute complexity of the instance—is superpolynomial in $|\mathcal{X}|$, the number of arguments in the system.

Our goal in the remainder of this paper is to establish the existence of a sequence of TPI-dispute instances—$\langle \mathcal{H}_N, x \rangle$—having N arguments, x not credulously accepted in $\mathcal{H}_N$, and with the number of moves in any terminated TPI-dispute being exponential in N. Of course, since these bounds apply *only* to our specific formalisation, this raises the question of defining 'more powerful' dispute protocols.

[4] Dimopoulos and Torres [16] employ rather different terminology from that introduced by Dung [17], however, it is not difficult to relate the two: a brief discussion interpreting the contribution of [16] in terms of Dung's argument systems is presented in [19].

[5] In complexity-theoretic terms, this is the assertion that NP $\neq$ CO-NP. It is worth noting that if true, it implies P $\neq$ NP. The converse, however, is not (necessarily) true: in principle one might have NP = CO-NP and P $\neq$ NP.

3.1. Propositional tautologies and argument systems

The proof that CA is NP-complete obtained in [16] is effected through a reduction from 3-SAT, this construction extending easily to CNF-SAT, i.e., without the restriction of three literals per clause. The class of argument systems that result via this translation of CNF formulae turn out to be central to the analysis of dispute complexity, we therefore review the details of this in,

Definition 7. Given,

$$\varphi(Z_n) = \bigwedge_{i=1}^{m} C_i = \bigwedge_{i=1}^{m} \left(\bigvee_{j=1}^{k_i} y_{i,j} \right)$$

a propositional formula in CNF comprising m clauses—C_i—the ith containing exactly $k_i \geqslant 1$ distinct literals over the propositional variables $Z_n = \langle z_1, z_2, \ldots, z_n \rangle$, the argument system $\mathcal{H}_\varphi(\langle \mathcal{X}_\varphi, \mathcal{A}_\varphi \rangle)$ has $2n + m + 1$ arguments

$$\mathcal{X}_\varphi = \{\varphi\} \cup \{C_1, C_2, \ldots, C_m\} \cup \{z_1, \neg z_1, z_2, \neg z_2, \ldots, z_n, \neg z_n\}$$

and attack relationship $\mathcal{A}_\varphi$ in which,

(1) $\forall C_i \ \langle C_i, \varphi \rangle \in \mathcal{A}_\varphi$.
(2) $\forall z_j \ \langle z_j, \neg z_j \rangle \in \mathcal{A}_\varphi$ and $\langle \neg z_j, z_j \rangle \in \mathcal{A}_\varphi$.
(3) $\langle z_j, C_i \rangle \in \mathcal{A}_\varphi$ if z_j is a literal in the clause C_i.
(4) $\langle \neg z_j, C_i \rangle \in \mathcal{A}_\varphi$ if $\neg z_j$ is a literal in the clause C_i.

For convenience we will subsequently write $y \in C$ rather than 'y is a literal in the clause C'.

This system is similar (although not identical) to the mechanism defined in [16, Theorem 5.1, p. 227]. It is straightforward to show as a consequence,

Fact 1. *The* CNF *formula $\varphi(Z_n)$ is satisfiable if and only if the argument φ is credulously accepted in the system $\mathcal{H}_\varphi(\langle \mathcal{X}_\varphi, \mathcal{A}_\varphi \rangle)$.*

Thus in attempting to derive lower bounds on dispute complexity for cases in which x is not credulously accepted in $\mathcal{H}$, we could focus on bounding $\delta(\mathcal{H}_\varphi, \varphi)$ for appropriate instances in which $\neg\varphi(Z_n)$ is a *tautology*, i.e., $\varphi(Z_n)$ is not satisfiable.

Our approach to establishing such lower bounds will be rather less direct than that of examining $\delta(\mathcal{H}_\varphi, \varphi)$ for a specific propositional tautology $\neg\varphi$. Instead, we shall show that the progression of a TPI-dispute over φ can be '*efficiently simulated*' within a specific Proof Calculus for Propositional Logic: since the calculus we employ is *known* to require exponentially long proofs to validate certain tautologies, it will then follow that $\delta(\mathcal{H}_\varphi, \varphi)$ for such φ must also be exponential (in the number of arguments defining $\mathcal{H}_\varphi$).

It is worth noting, at this point, that there is a rich corpus of research concerning the *length* of proofs in various proof systems. Results on the complexity of General Resolution date back to the seminal paper of Haken [23] in which this approach was shown to require

exponential length proofs for tautologies corresponding to the combinatorial *Pigeon-Hole Principle*, with important subsequent work in, e.g., [1,3,4,30], etc. Excellent introductory surveys discussing progress involving proof complexity may be found in the articles by Pudlák [31] and Beame and Pitassi [5].

3.2. The Gentzen Calculus for Propositional Logic

The Proof Calculus around which our simulation is built is the *Gentzen* (or *Sequent*) *Calculus*, [21], with, however, one of its standard inference rules being unavailable.

In its most general (propositional) form, the Gentzen Calculus, prescribes rules for deriving *sequents*—$\Gamma \Rightarrow \Delta$—where Γ, Δ are *sets* of propositional formulae (over a set of atomic propositional variables $\{x_1, x_2, x_3, \ldots\}$) built using some finite (complete) logical basis. A *proof* of the sequent $\Gamma \Rightarrow \Delta$, consists of a sequence of *derivation steps* each of which is either an axiom or follows by applying one of the rules to (at most) two previously derived sequents. In what follows we observe the convention of employing upper case Roman letters—$\{A, B, C, \ldots, \}$—to denote propositional formulae, and upper case Greek letters—$\{\Gamma, \Delta, \ldots\}$—to denote *sets* of such formulae. We use Γ, A to denote the set $\Gamma \cup \{A\}$.

Definition 8 (*Gentzen Calculus for Propositional Formulae*). Let $\mathcal{L}$ be the language of well-formed formulae using the basis $\{\wedge, \vee, \neg\}$ and propositional variables drawn from $\{z_1, z_2, z_3, \ldots\}$.

A *sequent* is an expression the form $\Gamma \Rightarrow \Delta$ where Γ, Δ are (finite) subsets of $\mathcal{L}$, i.e., sets of well-formed formulae. For a sequent $S = \Gamma \Rightarrow \Delta$ we use $\text{LHS}(S)$ to denote Γ and, similarly, $\text{RHS}(S)$ to denote Δ. A *Gentzen System* is defined by a set $\mathcal{GS}$ of *axioms* and *inference rules*. A sequent $\Gamma \Rightarrow \Delta$ is *provable in the Gentzen System* $\mathcal{GS}$ (written $\vdash_{\mathcal{GS}} \Gamma \Rightarrow \Delta$), if there is a finite sequence of sequents,

$$S_1, S_2, \ldots, S_{k-1}, S_k, S_{k+1}, \ldots, S_t$$

for which S_t is the sequent $\Gamma \Rightarrow \Delta$ and for all k $(1 \leqslant k \leqslant t)$, the sequent S_k is either an axiom of $\mathcal{GS}$ or there are sequents S_i, S_j (with $i, j < k$) and an inference rule r of $\mathcal{GS}$ such that S_k may be inferred from S_i and S_j as a consequence of the rule r. The *Proof Complexity* of a sequent S in the *Gentzen System* $\mathcal{GS}$ (denoted $\pi(S, \mathcal{GS})$ is defined for *provable* sequents, to be the least t such that S is derived by a sequence of t sequents.[6]

We shall use a modification of the Gentzen system, $\mathcal{G}$ shown in Table 1, wherein A and B are members of $\mathcal{L}$, and Γ, Δ, etc. subsets of $\mathcal{L}$.

It may be observed that the Resolution Rule is, in fact, a special case of the CUT rule: if we consider clauses

$$P = x \vee \bigvee_{i=1}^{r} y_i; \qquad Q = \neg x \vee \bigvee_{i=1}^{s} z_i$$

[6] We note that some authors choose to define proof complexity in terms of the total number of *symbol* occurrences over the derivation. For the class of propositional formulae we will be considering, the two measures are polynomially equivalent.

Table 1
The Gentzen system $\mathcal{G}$

Axioms	$\{A\} \Rightarrow \{A\}$		
Rules			
$(\theta \Rightarrow)$	$\dfrac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta}$	$\dfrac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}$	$(\Rightarrow \theta)$
(CUT)	$\dfrac{\Gamma \Rightarrow \Delta, A \ ; \ \Gamma', A \Rightarrow \Delta'}{\Gamma \cup \Gamma' \Rightarrow \Delta \cup \Delta'}$		
$(\neg \Rightarrow)$	$\dfrac{\Gamma \Rightarrow \Delta, A}{\Gamma, \neg A \Rightarrow \Delta}$	$\dfrac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A}$	$(\Rightarrow \neg)$
$(\vee \Rightarrow)$	$\dfrac{\Gamma, A \Rightarrow \Delta \ ; \ \Gamma', B \Rightarrow \Delta'}{\Gamma \cup \Gamma', A \vee B \Rightarrow \Delta \cup \Delta'}$	$\dfrac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B}$	$(\Rightarrow \vee)$

these are resolved (on x) to the clause

$$\bigvee_{i=1}^{r} y_i \vee \bigvee_{j=1}^{s} z_j .$$

The clause P may be expressed as the sequent, $\{y_1, \ldots, y_r\} \Rightarrow \{\neg x\}$ and Q as $\{\neg x\} \Rightarrow \{\neg z_1, \ldots, \neg z_s\}$ whence the sequent $\{y_1, \ldots, y_r\} \Rightarrow \{\neg z_1, \ldots, \neg z_s\}$ follows from the CUT rule. For a more detailed comparison of General Resolution and Gentzen calculi the reader is referred to [2].

The Gentzen system that we will be considering is $\mathcal{G}/\text{CUT}$, i.e., that which allows all of the rules of the system $\mathcal{G}$ *except for* the CUT rule. We recall some standard results concerning the systems $\mathcal{G}$ and $\mathcal{G}/\text{CUT}$.

Fact 2 (Gentzen [21]). *The propositional formula $\mathcal{F} \in \mathcal{L}$ is a tautology if and only if* $\vdash_{\mathcal{G}} \emptyset \Rightarrow \{\mathcal{F}\}$.

Fact 3 (The Gentzen Cut-Elimination Theorem [21]).

$$\vdash_{\mathcal{G}} \emptyset \Rightarrow \{\mathcal{F}\} \quad \textit{if and only if} \quad \vdash_{\mathcal{G}/\text{CUT}} \emptyset \Rightarrow \{\mathcal{F}\}.$$

Fact 3 establishes that the CUT rule is not needed in order to derive any provable sequent. Nonetheless, CUT turns out to be an extremely powerful operation:

Fact 4 (Urquhart [35,36]). *There are* (*infinite*) *sequences of formulae $\langle \mathcal{F}_n \rangle$ in $\mathcal{L}$ for which*:

(a) *$\mathcal{F}_n$ is a propositional tautology of n propositional variables.*
(b) $\pi(\emptyset \Rightarrow \{\mathcal{F}_n\}, \mathcal{G}) = \text{O}(n^k)$ (*for* $k \in \mathbb{N}$).
(c) $\pi(\emptyset \Rightarrow \{\mathcal{F}_n\}, \mathcal{G}/\text{CUT}) = \Omega(2^{n^{\varepsilon}})$ (*where* $\varepsilon > 0$).

These constructions by Urquhart are explicit, i.e., a specific sequence $\langle \mathcal{F}_n \rangle$ is proved to have the properties stated in Fact 4.

We now state and prove the main theorem of this paper.

Theorem 5. *Let*

$$\varphi(Z_n) = \bigwedge_{i=1}^{m} C_i = \bigwedge_{i=1}^{m} \left(\bigvee_{j=1}^{k_i} y_{i,j} \right)$$

be any unsatisfiable CNF*-formula,* $\mathcal{H}_\varphi$ *be the argument system defined from* $\varphi(Z_n)$ *as given in Definition* 7*, and* S_φ *the* (*provable*) *sequent,*

$$\emptyset \Rightarrow \bigcup_{i=1}^{m} \left\{ \neg \left(\bigvee_{j=1}^{k_i} y_{i,j} \right) \right\}.$$

Then,

$$\pi\left(S_\varphi, \mathcal{G}/\text{CUT}\right) \leqslant \delta(\mathcal{H}_\varphi, \varphi) + 2n + m. \tag{4}$$

Less formally, Theorem 5 states that the length of the shortest proof of $\neg\varphi$ (φ being in CNF) being a tautology within the CUT-free Gentzen system cannot be 'much greater than' the number of moves needed to form a successful rebuttal of φ in the argument system $\mathcal{H}_\varphi$.

Proof of Theorem 5. Let, φ, $\mathcal{H}_\varphi$, and S_φ be as described in the Theorem statement. Given *any* terminated TPI-dispute, M over φ in $\mathcal{H}_\varphi$ we describe how its progress may simulated in the Gentzen system $\mathcal{G}/\text{CUT}$. We first observe two important properties of the dispute M.

Firstly, M may be encoded as a sequence of *ordered* sets, R_i, (for which the term *retraction round* will subsequently be employed). Each of these takes the form

$$R_i = \langle D_1, y_1, D_2, y_2, \ldots, D_j, y_j, \ldots, D_q, y_q, F_i \rangle, \tag{5}$$

where

$$\{D_1, D_2, \ldots, D_q, F_i\} \subseteq \{C_1, C_2, \ldots, C_m\}, \quad y_j \in D_j,\ y_j \notin F_i\ \forall 1 \leqslant j \leqslant q.$$

In other words, R_i describes the alternation between clauses (D) used to attack φ and counterattacks (y) used to repel these attacks. The final attack by the clause F_i is the position at which the retraction of $\{\varphi, y_1, y_2, \ldots, y_q\}$ is forced. We observe that $|R_1|$ is the number of moves made in M prior to the *first* RETRACT move; and in general, $|R_i|$ is the number of moves between the retraction arising from R_{i-1} and the next such in M.

In the final move of M, the corresponding set R, contains just a single clause: i.e., that clause of φ upon which the Defender, by reason of the totality of earlier retractions, can mount no attack.

The second property of interest concerns the relationship between the literals defining a *retraction forcing clause,* F, and those used to defend against attacks on φ within the

current dispute tree, i.e., the literals $\{y_1, y_2, \ldots, y_q\}$. The literals in F may be partitioned into two sets,

$$W = \{w_1, w_2, \ldots, w_r\}; \qquad U = \{u_1, u_2, \ldots, u_s\} \tag{6}$$

wherein the literals in W cannot be used to attack F since for each $w \in W$, $\neg w \in \{y_1, y_2, \ldots, y_q\}$ and those in U are unavailable since for each $u \in U$, there is some subset V of $\{y_1, y_2, \ldots, y_q\}$ such that the Defender has retracted $\{\varphi, V, u\}$ in an earlier move.

With the two observations above, the idea underlying the proof may be described, informally, as efficiently deriving sequents that simulate the reasoning through which retractions are forced. More precisely, given

$$\langle R_1, R_2, \ldots, R_i, \ldots, R_t\rangle$$

the sequence of retraction rounds describing the dispute M, we construct a mapping $\beta : \{1, 2, \ldots, t\} \to \mathbb{N}$ and sequents

$$\langle S_1, S_2, S_3, \ldots, S_p\rangle$$

for which

$$\beta(i+1) > \beta(i) \geqslant 1 \quad (1 \leqslant i < t)$$
$$\beta(t) \leqslant p \leqslant \beta(t) + m$$

and

$$S_p = S_\varphi = \emptyset \Rightarrow \bigcup_{i=1}^{m} \left\{ \neg \left(\bigvee_{j=1}^{k_i} y_{i,j} \right) \right\}.$$

In general, the sequent $S_{\beta(i)}$ will express the fact that the Defender must retract the set $\{\varphi, y_1, y_2, \ldots, y_q\}$ in the ith round, since this leaves no defence available to an attack by the clause F_i on φ.

To avoid a surfeit of subscripts, we use Y_i to denote the set $\{y_1, y_2, \ldots, y_q\}$ of literals defining R_i, with W_i and U_i being the partition of the retraction forcing clause, F_i, as described in (6) (obviously the *exact number* of literals in each of these will be dependent on which retraction round R_i is relevant).

When $U_i \neq \emptyset$, for each $u \in U_i$, $ret(u, Y_i)$ is a minimal (with respect to $\subseteq$) subset of Y_i for which the set of arguments $\{\varphi, u, ret(u, Y_i)\}$ has been the subject of an *earlier* retraction.[7] Finally, $index(u, Y_i)$ is,

$$index(u, Y_i) = \max\{k \leqslant \beta(i-1) : \text{LHS}(S_k) = ret(u, Y_i) \cup \{u\}\}. \tag{7}$$

Note. That $index(u, Y_i)$ is well-defined will be clear from the remainder of the proof.

The theorem will follow from the claim below.

[7] An indefinite article is required here, since there may be more than one such subset, e.g., $\{\varphi, y_1, u\}$ and $\{\varphi, y_2, u\}$ could *both* have been retracted: the subsequent argument will show that in such cases, $ret(u, Y)$ can be chosen to be either $\{y_1\}$ or $\{y_2\}$.

Claim 1. *Given* $\langle R_1, \ldots, R_t \rangle$ *the sequence of retraction rounds defined by* M*, there is a mapping* $\beta : \{1, 2, \ldots, t\} \to \mathbb{N}$*, with the following properties*: $\beta(i+1) > \beta(i) > 1$; *and, if the sequent,* $S_{\beta(i)}$ *is defined to be*

$$S_{\beta(i)} = \{Y_i\} \Rightarrow \{\neg F_i\} \cup \bigcup_{u \in U_i} \text{RHS}(S_{index(u,Y_i)}),$$

then

(a) $S_{\beta(i)}$ *is well-defined, i.e., index*(u, Y_i) *is defined for each* $u \in U_i$.
(b) $S_{\beta(i)}$ *is provable in* $\mathcal{G}/\text{CUT}$ *with* $\pi(S_{\beta(i)}, \mathcal{G}/\text{CUT}) \leqslant \beta(i)$.

Proof. First note that we may use the following derivations as the first $2n$ lines, prior to establishing $S_{\beta(1)}$. In consequence, $\beta(1) > 2n$.

Sequent	via	Line
$\{z_j\} \Rightarrow \{z_j\}$	Axiom	$2j-1$
$\{z_j, \neg z_j\} \Rightarrow \emptyset$	$(\neg \Rightarrow)$ and $2j-1$	$2j$

We complete the proof of the claim by induction on $i \geqslant 1$. The inductive base, $i = 1$, deals with the retraction enforced by R_1, i.e., we need to show that the sequent

$$S_{\beta(1)} = \{Y_1\} \Rightarrow \{\neg F_1\}$$

is derivable. Noting that R_1 represents the *first* occurrence of a RETRACT move by the Defender, the set U_1 must be empty, i.e., the retraction is forced because each literal that could be used to attack F_1 is unavailable by reason of Y_1 containing its negation. It follows that,

$$F_1 = W_1 = \{w_1, w_2, \ldots, w_r\},$$
$$Y_1 = \{\neg w_1, \neg w_2, \ldots, \neg w_r, y_{r+1}, y_{r+2}, \ldots, y_q\}.$$

Let T_k (for $1 \leqslant k \leqslant r$) be the sequent,

$$\{\neg w_1, \neg w_2, \ldots, \neg w_k\}, A_k \Rightarrow \emptyset \quad \text{where } A_k = \bigvee_{j=1}^{k} w_j.$$

For $k = 1$, the sequent $T_1 = \{w_1, \neg w_1\} \Rightarrow \emptyset$ has already been derived. For $k > 1$, T_k is derived in one step from the sequent $\{w_k, \neg w_k\} \Rightarrow \emptyset$ and T_{k-1} by a single application of the rule $(\vee \Rightarrow)$. We deduce that,

$$\{\neg w_1, \neg w_2, \ldots, \neg w_k\}, F_1 \Rightarrow \emptyset$$

is derived in $k-1$ steps, and the required sequent—$S_{\beta(1)}$—by a single application of $(\Rightarrow \neg)$ to T_r followed by $q-r$ applications of $(\theta \Rightarrow)$ in order to construct

$$\{\neg w_1, \ldots, \neg w_r, y_{r+1}, \ldots, y_q\} \Rightarrow \{\neg F_1\}.$$

This gives the value of $\beta(1)$ as $2n+q$, where we note that μ_{2q+2} is the *first* RETRACT move occurring in M.

For the Inductive Step, we assume for all retraction rounds R_j with $1 \leqslant j < i$ that the following hold:

(IH1) The value of $\beta(j)$ has been defined.
(IH2) The sequent,

$$S_{\beta(j)} = \{Y_j\} \Rightarrow \{\neg F_j\} \cup \bigcup_{u \in U_j} \text{RHS}(S_{index(u,Y_j)})$$

has been derived in $\mathcal{G}/\text{CUT}$ after $\beta(j)$ steps.

To complete the inductive proof of Claim 1, we 'simulate' the retraction round R_i and to this end it is necessary to,

(C1) define a value of $\beta(i)$ which is greater than $\beta(i-1)$, and
(C2) show that the sequent,

$$\{Y_i\} \Rightarrow \{\neg F_i\} \cup \bigcup_{u \in U_i} \text{RHS}(S_{index(u,Y_i)})$$

is well-defined *and* derivable in a further $\beta(i) - \beta(i-1)$ steps.

Consider the retraction forcing clause, $F_i = \{W_i, U_i\}$, so that

$$Y_i = \{\neg w_1, \neg w_2, \ldots, \neg w_r, y_{r+1}, y_{r+2}, \ldots, y_q\}.$$

If $U_i = \emptyset$, then with $\beta(i) = \beta(i-1) + q$, the sequent,

$$S_{\beta(i)} = \{Y_i\} \Rightarrow \{\neg F_i\}$$

is derivable in a further q steps using exactly the same approach as employed in the Inductive Base. Thus we may assume that U_i is non-empty with

$$U_i = \{u_1, u_2, \ldots, u_s\}.$$

Recalling that $\langle W_i, U_i \rangle$ is a *partition* of F_i it is certainly the case that neither $\neg u \in Y_i$ nor $u \in Y_i$ (the latter holding since F_i was available to the Challenger with which to attack φ). This being so and u being unavailable to the Defender to attack F_i it follows that there has been a retraction round in which some subset of Y_i *together with* u (and φ) have been retracted. Therefore, some such subset of Y_i must satisfy the criteria defining $ret(u, Y_i)$ with respect to u. Suppose R_j is the round at which a committment to $\{\varphi, u, ret(u, Y_i)\}$ was retracted by the Defender. Clearly, $j < i$ and hence from the Inductive Hypothesis, the sequent, $S_{\beta(j)}$, with,

$$S_{\beta(j)} = \{u, ret(u, Y_i)\} \Rightarrow \Delta \quad \text{where } \emptyset \subset \Delta \subseteq \{\neg C_1, \ldots, \neg C_m\}$$

has been derived. As a result we deduce that for each $u \in U_i$, the value $index(u, Y_i)$ is defined and does not exceed $\beta(i-1)$. In summary, we have proven (via the Inductive Hypothesis) the existence of $s = |U_i|$ sequents,

$$\langle S_{i,1}, S_{i,2}, \ldots, S_{i,s} \rangle$$

for which

$$\text{LHS}(S_{i,k}) = ret(u_k, Y_i) \cup \{u_k\} \quad \text{and} \quad \text{RHS}(S_{i,k}) \subseteq \{\neg C_1, \neg C_2, \ldots, \neg C_m\}.$$

We can now complete the derivation of the required sequent $S_{\beta(i)}$.

From $s-1$ applications of $(\vee \Rightarrow)$ using $S_{i,1}, S_{i,2}, \ldots, S_{i,s}$ we obtain

$$S_{\beta(i-1)+s-1} = \left\{ \bigcup_{k=1}^{s} ret(u_k, Y_i) \right\}, \qquad \bigvee_{k=1}^{s} u_k \Rightarrow \left\{ \bigcup_{k=1}^{s} \mathrm{RHS}(S_{i,k}) \right\}.$$

A further r applications of $(\vee \Rightarrow)$ involving $S_{\beta(i-1)+s-1}$ and the sequents

$$\{w_k, \neg w_k\} \Rightarrow \emptyset$$

yields $S_{\beta(i-1)+r+s-1}$ as

$$\left\{ \bigcup_{k=1}^{s} ret(u_k, Y_i) \right\} \cup \left\{ \bigcup_{k=1}^{r} \neg w_k \right\}, \qquad \bigvee_{k=1}^{s} u_k \vee \bigvee_{k=1}^{r} w_k \Rightarrow \left\{ \bigcup_{k=1}^{s} \mathrm{RHS}(S_{i,k}) \right\}.$$

Recalling that,

$$F_i = \left(\bigvee_{k=1}^{s} u_k \vee \bigvee_{k=1}^{r} w_k \right)$$

a single application of $(\Rightarrow \neg)$ to $S_{\beta(i-1)+r+s-1}$ gives $S_{\beta(i-1)+r+s}$ as,

$$\left\{ \bigcup_{k=1}^{s} ret(u_k, Y_i) \right\} \cup \left\{ \bigcup_{k=1}^{r} \neg w_k \right\} \Rightarrow \left\{ \bigcup_{k=1}^{s} \mathrm{RHS}(S_{i,k}) \right\}, \qquad \neg F_i.$$

Finally, since it may be the case that

$$\left\{ \bigcup_{k=1}^{s} ret(u_k, Y_i) \right\} \cup \left\{ \bigcup_{k=1}^{r} \neg w_k \right\} \subset Y_i$$

(i.e., a *strict* subset of Y_i) a total of,

$$\left| Y_i \Big/ \left\{ \bigcup_{k=1}^{s} ret(u_k, Y_i) \cup \bigcup_{k=1}^{r} \neg w_k \right\} \right|$$

applications of $(\theta \Rightarrow)$ will give $S_{\beta(i)}$ as,

$$S_{\beta(i)} = \{Y_i\} \Rightarrow \{\neg F_i\} \cup \bigcup_{u \in U_i} \mathrm{RHS}(S_{index(u,Y_i)}),$$

where

$$\beta(i-1) + r + s \leqslant \beta(i) \leqslant \beta(i-1) + r + s + q \leqslant \beta(i-1) + 2q.$$

Note that $2q = |R_i| - 1$ is the total number of moves occurring in M between the retraction round R_{i-1} and R_i. This completes the inductive proof of the claim. □

To complete the proof of the theorem we need only observe that the total number of steps required to derive S_φ is bounded above by $\beta(t) + m$.

(The additional m arises from the possibility that $S_{\beta(t)}$ may be of the form $\emptyset \Rightarrow \Delta$ with Δ a (non-empty) *strict* subset of

$$\{\neg C_1, \neg C_2, \ldots, \neg C_m\}.$$

This could occur if some subset ψ of φ's clauses defined an unsatisfiable CNF-formula. In such cases $S_{\beta(t)}$ would not be *identical* to the sequent S_φ of the theorem statement, however, at most m applications of $(\Rightarrow \theta)$ (adding the 'missing' $\neg C_i$ clauses) will suffice to derive S_φ from $S_{\beta(t)}$.)

From the analysis in the proof of the claim it is clear that the values $\beta(i)$ satisfy:

$$\beta(i) \leqslant \beta(i-1) + |R_i| \quad \text{when } 1 < i \leqslant t,$$
$$\beta(1) \leqslant 2n + |R_1|,$$

hence $\beta(t) \leqslant 2n + \sum_{i=1}^{t} |R_i| \leqslant 2n + |M|$.

Thus from any terminated TPI-dispute, M, over the unsatisfiable CNF-formula φ in the argument system $\mathcal{H}_\varphi$ we may construct a proof in $\mathcal{G}/\text{CUT}$ that $\neg\varphi$ is a tautology, i.e., of the sequent S_φ. Since this proof involves at most $|M| + 2n + m$ steps we conclude that

$$\pi(S_\varphi, \mathcal{G}/\text{CUT}) \leqslant \delta(\mathcal{H}_\varphi, \varphi) + 2n + m$$

as required. □

From Theorem 5 we get,

Corollary 1. *There are (infinite) sequences of argument systems with arguments $x \in \mathcal{X}$ not credulously accepted but with the number of moves in any* TPI*-dispute establishing such exponential in $|\mathcal{X}|$.*

To conclude this section, we illustrate how the example of Fig. 2(b) that resulted in the dispute given in (3) translates into a derivation of the required sequent following the proof in Theorem 5.

3.3. Example

Recall that Fig. 2(b) could be interpreted as the tautology

$$\neg F(y,z) = \neg\big((y \vee z) \wedge (y \vee \neg z) \wedge (\neg y \vee z) \wedge (\neg y \vee \neg z)\big). \tag{8}$$

From (3) using the encoding of retraction rounds described in the proof of Theorem 5

$$\begin{aligned} R_1 &= \big\langle (y \vee z), y, (\neg y \vee z), z, (\neg y \vee \neg z)\big\rangle, \\ R_2 &= \big\langle (y \vee z), z, (y \vee \neg z)\big\rangle, \\ R_3 &= \big\langle (y \vee z), y, (\neg y \vee z)\big\rangle, \\ R_4 &= \big\langle (y \vee z)\big\rangle. \end{aligned} \tag{9}$$

The sequent we wish to derive is

$$\emptyset \Rightarrow \big\{\neg(y \vee z), \neg(y \vee \neg z), \neg(\neg y \vee z), \neg(\neg y \vee \neg z)\big\}. \tag{10}$$

Following the mechanism in the theorem, for R_1 we wish to derive

$$S_{\beta(1)} = \{y, z\} \Rightarrow \{\neg(\neg y \vee \neg z)\}.$$

This is obtained by

Sequent	via	Line
$\{y\} \Rightarrow \{y\}$	Axiom	1
$\{y, \neg y\} \Rightarrow \emptyset$	1, $(\neg \Rightarrow)$	2
$\{z\} \Rightarrow \{z\}$	Axiom	3
$\{z, \neg z\} \Rightarrow \emptyset$	3, $(\neg \Rightarrow)$	4
$\{y, z\}, (\neg y \vee \neg z) \Rightarrow \emptyset$	2, 4, $(\vee \Rightarrow)$	5
$\{y, z\} \Rightarrow \{\neg(\neg y \vee \neg z)\}$	5, $(\Rightarrow \neg)$	6

Hence $\beta(1) = 6$.

For R_2 the sequent required is

$$S_{\beta(2)} = \{z\} \Rightarrow \{(\neg(y \vee \neg z), \neg(\neg y \vee \neg z)\}$$

where we use the fact that $ret(y, \{z\}) = \{z\}$, so that $index(y, \{z\}) = 6$.

Sequent	via	Line
$\{z\}, (y \vee \neg z) \Rightarrow \{\neg(\neg y \vee \neg z)\}$	4, 6, $(\vee \Rightarrow)$	7
$\{z\} \Rightarrow \{\neg(\neg y \vee \neg z), \neg(y \vee \neg z)\}$	7, $(\Rightarrow \neg)$	8

whence $\beta(2) = 8$. Notice that in deriving S_7, $\text{LHS}(S_4)$ is viewed as $\{z\}, \neg z$ and $\text{LHS}(S_6)$ as $\{z\}, y$, i.e., with $\Gamma = \Gamma' = \{z\}$, $A = \neg z$, and $B = y$ when the inference rule $(\vee \Rightarrow)$ of Table 1 is used.

For R_3 the sequent required is

$$S_{\beta(3)} = \{y\} \Rightarrow \{\neg(\neg y \vee z), \neg(y \vee \neg z), \neg(\neg y \vee \neg z)\},$$

where we use the fact that $ret(z, \{y\}) = \emptyset$, so that $index(z, \{y\}) = 8$.[8]

Sequent	via	Line
$\{y\}, (\neg y \vee z) \Rightarrow \{\neg(\neg y \vee \neg z), \neg(y \vee \neg z)\}$	2, 8, $(\vee \Rightarrow)$	9
$\{y\} \Rightarrow \{\neg(\neg y \vee \neg z), \neg(y \vee \neg z), \neg(\neg y \vee z)\}$	9, $(\Rightarrow \neg)$	10

giving $\beta(3) = 10$.

Finally for R_4 we have

$$ret(y, \emptyset) = \emptyset \quad \text{with } index(y, \emptyset) = 10,$$
$$ret(z, \emptyset) = \emptyset \quad \text{with } index(z, \emptyset) = 8,$$

[8] Were $ret(z, \{y\})$ not subject to a minimality condition, it could also be chosen as $\{y\}$, giving $index(z, \{y\}) = 6$. This choice would, in fact, still lead to a proof of the required final sequent. We also note the need for $index(z, \{y\})$ to be maximal since $\text{LHS}(S_3) = \{z\}$.

so that using S_8, S_{10} and $(\vee \Rightarrow)$ gives

$$S_{11} = \big\{(y \vee z)\big\} \Rightarrow \big\{\neg(\neg y \vee \neg z), \neg(y \vee \neg z), \neg(\neg y \vee z)\big\}$$

and with a single application of $(\Rightarrow \neg)$ to S_{11}, we derive the required sequent

$$S_{12} = \emptyset \Rightarrow \big\{\neg(y \vee z), \neg(y \vee \neg z), \neg(\neg y \vee z), \neg(\neg y \vee \neg z)\big\}.$$

The results above show that TPI–disputes can be interpreted as a proof calculus with which to establish unsatisfiability of propositional formula presented in CNF, and that viewed thus, the number of 'moves' taken to resolve a dispute—i.e., prove that Φ is unsatisfiable—is bounded below by the number of lines in the shortest derivation of $\Rightarrow \neg\Phi$ in a CUT-free *Gentzen System*.

It may be shown that for any unsatisfiable CNF-formula Φ, the number of moves required in a TPI-dispute over $\langle \mathcal{H}_\Phi, \Phi\rangle$ cannot be 'much larger' than the size of the smallest *clausal tableau refutation* of Φ. An immediate consequence of this result being that the propositional proof system afforded by TPI-disputes is polynomially equivalent—in sense of [12]—to CUT-free Gentzen Systems and Clausal Tableaux, i.e., if $\langle \Pi_1, \Pi_2\rangle$ are any two proof systems from

{Gentzen/CUT, Clausal Tableaux, TPI-dispute}

then the length of the shortest validity proofs of $\neg\Phi$ for CNF-formulae Φ in Π_1 is at worst polynomially larger than the length of the shortest proof in the system Π_2. This follows from the equivalence of Clausal Tableaux and CUT-free Gentzen Systems, details of which may be found in [33, Chapter XI].

Definition 9. Let $\Phi(Z_n) = \bigwedge_{i=1}^{m} C_i$ be an unsatisfiable CNF-formula with clause set $\{C_1, C_2, \ldots, C_m\}$. A *clausal tableau* for Φ is a tree $T(V, E)$ in which the non-leaf vertices, v, are associated with a clause $C(v)$ of Φ, in accordance with the following rules.

On any path $\rho = v_0 \to v_1 \to \cdots \to v_r$ from the root (ρ) to a leaf v_r, each clause C_i of Φ labels *at most one* of $\{v_0, v_1, \ldots, v_{r-1}\}$. If $\rho \in V$ is the root of T and $C(\rho) = \bigvee_{i=1}^{k} y_{\rho,i}$ the clause associated with ρ, then ρ has exactly k children—$\langle v_1, v_2, \ldots, v_k\rangle$ with the edge $\langle \rho, v_j\rangle$ labelled $y_{i,j}$. If $v \in V$ is a non-leaf vertex (other than the root) let U be the set of literals labelling edges on the (unique) path from ρ to v and $C(v) = \bigvee_{i=1}^{s} y_{v,i}$ be the clause associated with v. Again, v has exactly s children $\langle w_1, w_2, \ldots, w_s\rangle$ with the edges $\langle v, w_i\rangle$ labelled $y_{v,i}$. In this case, however, the vertex w_i is a leaf labelled $\perp$ if the literal $\neg y_{v,i} \in U$. A vertex is *closed* if every path from it leads to a leaf (labelled $\perp$). A clausal tableau is a *refutation* for Φ if its root is closed.

The *size* of a clausal tableau $T(V, E)$—denoted $\tau(T)$—is the total number of internal vertices contained in it. The *clausal tableau complexity* of an unsatisfiable CNF-formula $\Phi(Z_n)$, is

$$\tau\big(\Phi(Z_n)\big) \underset{\text{def}}{=} \min\big\{\tau(T)\colon\ T \text{ is a clausal tableau refutation of } \Phi(Z_n)\big\}.$$

Theorem 6. *Let $\Phi(Z_n) = \bigwedge_{i=1}^{m} C_i$ be an unsatisfiable* CNF*-formula and $T(V, E)$ any clausal tableau refutation of $\Phi(Z_n)$, then $\delta(\mathcal{H}_\Phi, \Phi) \leqslant (2n+1)\tau(T)$.*

Proof. See [18]. □

4. Discussion and further work

In this paper our primary goal has been to formalise the argument game (TPI-dispute) introduced in [38] and to analyse this in terms of one particular computational measure—dispute complexity. For what is technically the most interesting case—the length of dispute required to convince Defenders of an argument that their position is untenable—we have shown in Theorem 5 that applying this dispute regime to simple argument system representations of propositional tautologies occasions a form of proof calculus. This calculus is in one sense, however, extremely limited: any proof within it being capable of description by a comparable length proof in a CUT-free Gentzen System. Since examples are known of tautologies where allowing CUT admits exponentially shorter proofs[9] the protocol enforced by TPI-disputes when applied to certain propositional argument systems may take significantly longer to reach a conclusion than 'more powerful' deductive systems. We noted earlier, in describing the semantics of the RETRACT move that the position reverted to is the *initial* argument, rather than some 'intermediate' state of the dispute tree being developed. Among the reasons for favouring returning to the initial position, is that the length of disputes (as indicated by our simulation using CUT-free Gentzen Systems) does not, primarily, result from potentially repeating chains of defence which will ultimately fail: if the retraction mechanism were to revert to a 'sub-tree' of the dispute tree, cf. in a similar manner to that of the Challenger's BACKUP move, then this could be simulated from the initial argument just by repeating the relevant COUNTER and BACKUP moves. Since the size of any dispute tree can be at most the number of arguments within the system itself, a more sophisticated RETRACT semantics could only shorten the length of a dispute by a polynomial factor—not reduce it exponentially.

Before dealing with some questions that are raised by the main result of this paper, it may be useful to place our concerns in the general context of argument systems, dialogue games, reasoning systems, etc. While the view of dialogue process as a 2-player game has been long established, e.g., MacKenzie's DC [27], interpretations of Toulmin's Argument Schema [34] as a game-based method [6], etc., the direction towards which such work has tended is in attempting formally to capture different types of dialogue process: e.g., [22] is, primarily, concerned with argument in a legal reasoning context. As a result there is a wealth of differing models of dialogue ranging from taxonomies of dialogue types as in Reed [32] and Walton and Krabbe [39] to frameworks modelling divers concepts of what 'winning' a dialogue game might mean, e.g., [24]. Despite this variety of approaches, one unifying trend is that the central concern is primarily semantic, i.e., in defining the form(s) that games take, the rules and processes by which games evolve, the conditions under which games terminate, and in establishing degrees of soundness and completeness of the game capabilities. The question of how 'efficient' such processes might be, however, seems to have been largely neglected, with the exception of general complexity-theoretic classifications of Argumentation Frameworks within specific non-classical logics, e.g., [13–15] or analyses of termination properties. Thus, little work is evident concerning more general contexts for the two questions which this paper has considered, i.e., with different

[9] In fact, Urquhart [36], shows $\mathcal{G}/\text{CUT}$ can be weaker than simple truth-tables proving worst-case lower bounds of $\Omega(n!)$ for the former as opposed to upper bounds of $n2^n$ for the latter.

protocols for the conduct of dialogues, different attack semantics, concepts of 'winning' other than credulous acceptance. If practical applications of dialectic and reasoning games are to be realised—as has become widely posited with the advent of autonomous agent systems—then measures analogous to our concept of dispute complexity may be of importance in evaluating implemented systems.

A rather different situation to that outlined in the preceding paragraph, pertains with respect to concepts of Proof Complexity, that we have used as the basis of our analysis of dispute complexity: Cook and Reckhow [12] introduced a formal mechanism for comparing the *complexity* of different proof calculi so that two 'different' systems are regarded as equipotent if a formal proof in one can be 'simulated' in the other with only a small increase in size. An important feature of this approach is that it can be developed to address questions concerning proof strategies for acceptance of instances in CO-NP-problems other than UNSAT, e.g., the Graph Stability Number calculus of Chvátal [11], or the Hajós Calculus for proving a graph has chromatic number greater than 3, [7,28]. It is the case, however, that these analyses are effectively only dealing with Classical (Propositional) Logic, and such results as extend to non-classical Logics do so only by virtue of propositional logic being treatable as a sub-case, e.g., Haken [23] trivially applies to the Resolution Calculus for Temporal Logic of [20] simply by expressing the relevant tautology without the use of any temporal operators, i.e., exactly as its propositional form.

We conclude by reviewing some directions for further research, that encompass both argument and dialogue game developments as well as extensions to the concept of dispute complexity.

Within the framework of [12] while it is known that the Gentzen System $\mathcal{G}/\text{CUT}$ is weaker than both the system $\mathcal{G}$ and Propositional Proof systems employing General Resolution only, it is an open problem as to whether $\mathcal{G}$ and Resolution are equivalent, i.e., it has yet to be shown that, e.g., the Pigeon-Hole Principle tautologies require exponential length proofs in $\mathcal{G}$, however no (*efficient*) simulation of $\mathcal{G}$ by Resolution has been constructed. Theorems 5 and 6 establish that using the TPI-dispute protocol as a vehicle for constructing proofs of propositional tautologies, $\neg\varphi$, affords a system which is equivalent to $\mathcal{G}/\text{CUT}$ and Clausal Tableaux, thus we might represent the respective power of various proof calculi for propositional tautologies informally as,

$$\mathcal{G} \geqslant \text{Resolution} > (\mathcal{G}/\text{CUT} \equiv \text{TPI} \equiv \text{Clausal Tableaux}). \tag{11}$$

The situation depicted in (11) raises some interesting questions. Firstly, it may be noted that Theorem 5 operates in only 'one direction', that is we express the problem of proving a propositional formula to be a tautology as a problem of showing an argument is not credulously accepted in an argument system, thence relating a calculus for the latter to a calculus for the former. We have not considered, however, translations of argument systems *into* propositional formulae. For example, given $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), x\rangle$, the CNF-formula $\varphi_{\langle \mathcal{H}, x\rangle}$ over variables $\mathcal{X}$ is,

$$x \wedge \bigwedge_{\langle y,z\rangle \in \mathcal{A}} (\neg y \vee \neg z) \wedge \bigwedge_{y \in \mathcal{X}} \left(y \vee \bigvee_{\{z:\ \langle z,y\rangle \in \mathcal{A}\}} z \right). \tag{12}$$

It is easy to show that there is a *stable* extension of $\mathcal{H}$ containing x if and only if $\varphi_{\langle\mathcal{H},x\rangle}(\mathcal{X})$ is satisfiable.[10]

Translations such as (12) also allow us to give a more precise interpretation of what might be meant by 'more powerful' dispute protocol. Thus, let Π be a (2-player) dispute protocol for argument systems (i.e., prescribing the repertoire of moves, state changes, move applicability, termination conditions, etc.) with the properties that: given an instance $\langle\mathcal{H}, x\rangle$ of CA

(a) Π can produce a successful defence of x if and only if x is credulously accepted in $\mathcal{H}$.
(b) Π either always produces a successful defence *or* always results in a successful rebuttal of x.

We can define analogous notions of *dispute complexity* with respect to arbitrary protocols—say, $\delta(\langle\mathcal{H}, x\rangle, \Pi)$—and hence regard protocol Π_1 as 'at least as powerful' as protocol Π_2 (denoted $\Pi_1 \geqslant \Pi_2$) if there is a constant k with which: for all dispute instances $\langle\mathcal{H}, x\rangle$

$$\delta\big(\langle\mathcal{H}, x\rangle, \Pi_1\big) = \mathrm{O}\big(\delta\big(\langle\mathcal{H}, x\rangle, \Pi_2\big)^k\big).$$

Problem 1. What features must be incorporated in a dispute protocol, Π, in order for it to be *more* powerful than TPI? That is, for the dispute complexity of infinitely many TPI-disputes to be superpolynomial in the dispute complexity of Π on the same instances.

Problem 2. Similarly, what features must be incorporated in Π for it to be at least as powerful as General Resolution, Gentzen Systems, etc.?

It should be noted that there are subtle differences between Problems 1 and 2. The former could be examined directly without recourse to phrasing in terms of propositional proofs, the latter however is specifically concerned with the use of dispute protocols as a propositional proof mechanism.

With respect to Problem 1 it has been observed earlier that something other than 'local' modifications to the state following a RETRACT move is needed.

A rather more general concern is that of what criteria must a 'reasonable' dispute protocol satisfy. From complexity-theoretic considerations, the move repertoire and its implementation cannot be permitted to be 'too powerful', e.g., treating as single operations moves which are predicated on identifying structures in an argument graph whose construction is NP-hard. While the TPI-dispute protocol is 'realistic' in the sense that the applicability of a proposed move can be validated efficiently (this, of course, is *not* the same as identifying a 'best' move), in addressing the issues raised by Problem 1 one may wish

[10] Although it is possible to construct a ('short') CNF encoding 'preferred extension containing x' rather than *stable*, this has a rather more opaque form. In any event since the absence of a preferred extension of x implies the absence of a stable extension of x, for the constructions of interest (i.e., negative instances) the TPI-dispute protocol defined still applies. Furthermore, Dimopoulos and Torres [16] show that deciding if $\mathcal{H}$ has a *stable* extension containing a given argument x is also NP-complete.

to restrict consideration to 'reasonable' protocols.[11] It is, of course, unlikely (given the conjecture NP $\neq$ CO-NP) that there is a 'reasonable' dispute protocol capable of resolving any dispute within a number of moves polynomial in the size of the argument system concerned. Nevertheless, just as the fact that existing lower bounds on Proof Complexity in failing to encompass *all possible* systems—as would be needed to prove NP $\neq$ CO-NP—motivates consideration of more powerful proof systems, so it is reasonable to examine and precisely formulate 'increasingly powerful' dispute protocols.

Finally, even for 'weak' systems such as TPI-disputes in the case of instances which lead to successful rebuttals of an argument, there is the issue of the Challenger *constructing* the 'best' line of attack, i.e., of finding the dispute that minimises dispute complexity. An analogous situation in Proof Complexity was formulated in Bonet et al. [9]: suppose φ is an unsatisfiable CNF with m clauses and n variables. Letting $\pi(\varphi, S)$ denote the size of the shortest proof of $\neg\varphi$ in some Propositional Proof System S, then for a function, $q : \mathbb{N}^3 \to \mathbb{N}$, S is said to be *q-automatizable* if there exists a (deterministic) algorithm that produces a proof (in the system S) of $\neg\varphi$ in time $q(\pi(\varphi, S), n, m)$. The cases of interest are where q is polynomially bounded in $\pi(\varphi, S)$. Informally, if a proof system is *polynomially-bounded* automatizable, then this gives an algorithm that can 'efficiently' construct a proof that is 'not much larger' than the optimal proof. The concept of q-automatizability can be reformulated in the obvious way to refer to dispute complexity (or indeed verification calculi for other CO-NP-complete problems). This motivates,

Problem 3. Let $\langle \mathcal{H}, x \rangle$ be any TPI-dispute instance in which there are n arguments and for which x is not credulously accepted in $\mathcal{H}$. Is there a deterministic algorithm that in $q(\delta(\mathcal{H}, x), n)$ steps returns a terminated TPI-dispute M establishing a successful rebuttal of x *and* with q bounded by a polynomial in $\delta(\mathcal{H}, x)$? In other words, is the TPI-dispute protocol q-automatizable for some polynomial q?

To conclude our discussion of possible directions for further research, we note that our model of dispute assumes both protagonists have complete knowledge of the argument system (i.e., the finite directed graph structure). Thus the Defender may choose counterattacks which are known to eliminate particular (subsequent) attacks by the Challenger; similarly, as may be evinced by the development of the disputes from unsatisfiable CNF-formulae, the Challenger may invoke attacks, potential defences to which have been ruled out, e.g., when the Defender uses a literal y to attack a clause C, the Challenger may continue using an available clause containing $\neg y$, knowing that $\neg y$ cannot be used as a defence. In many situations it may not be the case that such complete knowledge is held *ab initio*. The modelling of disputes where the protagonists' views of the system evolve over several moves would provide a significant development of the preliminary formalism described in this paper. Such an extension would have considerable practical interest, since many of the implementations require such evolution. For example, Gordon's [22] game is intended to induce the participants to present the arguments that

[11] Similar considerations arise in Proof Complexity and an accepted formalism has evolved to distinguish 'reasonable' from 'unreasonable' proof calculi. For the complexity-theoretic aspects affecting dispute protocols such a formalism seems a plausible basis.

they wish to deploy, essentially establishing the argumentation framework which will be subsequently used when the question comes to trial. In [6] it is assumed that each participant has only a partial view on the argumentation framework which is extended by elements recognised by their opponents as the dialogue proceeds. If we consider disputes between autonomous agents, it is perhaps unrealistic to expect them to begin with a shared understanding of the overall argumentation framework.

5. Conclusion

In this paper we have introduced a formal concept of dispute complexity with which to consider questions regarding the number of moves required in a dialogue over a given argument before one player accepts that the argument is/is not defensible. Building on the Argument System formalism of [17] and the argument game—TPI-dispute—discussed in [38], a precise formulation of the latter has been presented. With this formulation at hand, we are able to prove that there are instances representing a win for the Challenger but for which exponentially many moves must be played before the Defender is convinced of this. Our techniques exploit the close relationship between such dispute protocols and the concept of formal proof calculi for propositional tautologies by showing that the TPI-dispute protocol applied to representations of these can be used to build a proof of validity in a CUT-free Gentzen System whose length is comparable to the number of moves needed in a TPI-dispute. The ideas and techniques put forward in this paper represent just a preliminary foundation: an extensive range of open questions and further directions for research arise from this, only a selection of which have been discussed.

References

[1] M. Ajtai, The complexity of the pigeonhole principle, in: Proc. 29th Annual Symposium on Foundations of Computer Science, White Plains, NY, IEEE, 1988, pp. 346–355.
[2] A. Avron, Gentzen-type systems, resolutions, and tableaux, J. Automat. Reason. 10 (2) (1993) 265–281.
[3] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, P. Pudlák, A. Woods, Exponential lower bounds for the pigeonhole principle, in: Proc. Twenty-Fourth Annual ACM Symposium on the Theory of Computing, Victoria, BC, 1992, pp. 200–220.
[4] P. Beame, T. Pitassi, An exponential separation between the parity principle and the pigeonhole principle, Ann. Pure Appl. Logic 80 (3) (1996) 195–228.
[5] P. Beame, T. Pitassi, Propositional proof complexity: Past, present, and future, Bull. EATCS 65 (1998) 66–89.
[6] T.J.M. Bench-Capon, Specification and implementation of Toulmin dialogue game, in: J.C. Hage, et al. (Eds.), Legal Knowledge Based Systems, GNI, 1998, pp. 5–20.
[7] C. Berge, Graphs and Hypergraphs, North-Holland, Amsterdam, 1973.
[8] A. Bondarenko, P.M. Dung, R.A. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artificial Intelligence 93 (1–2) (1997) 63–101.
[9] M. Bonet, T. Pitassi, R. Raz, Lower bounds for cutting planes proofs with small coefficients, J. Symbolic Logic 62 (3) (1997) 708–728.
[10] G. Brewka, A reconstruction of Rescher's theory of formal disputation based on default logic, in: A.G. Cohn (Ed.), Proc. Eleventh European Conference on Artificial Intelligence, Wiley, Chichester, 1994, pp. 366–370.
[11] V. Chvátal, Determining the stability number of a graph, SIAM J. Comput. 6 (4) (1977) 643–662.

[12] S.A. Cook, R.A. Reckhow, The relative complexity of propositional proof systems, J. Symbolic Logic 44 (1) (1979) 36–50.
[13] Y. Dimopoulos, B. Nebel, F. Toni, Preferred arguments are harder to compute than stable extensions, in: D. Thomas (Ed.), Proc. IJCAI-99, Stockholm, Sweden, Vol. 1, Morgan Kaufmann, San Francisco, CA, 1999, pp. 36–43.
[14] Y. Dimopoulos, B. Nebel, F. Toni, Finding admissible and preferred arguments can be very hard, in: A.G. Cohn, F. Giunchiglia, B. Selman (Eds.), Principles of Knowledge Representation and Reasoning, KR2000, Morgan Kaufmann, San Francisco, CA, 2000, pp. 53–61.
[15] Y. Dimopoulos, B. Nebel, F. Toni, On the computational complexity of assumption-based argumentation by default reasoning, Artificial Intelligence 141 (2002) 57–78.
[16] Y. Dimopoulos, A. Torres, Graph theoretical structures in logic programs and default theories, Theoret. Comput. Sci. 170 (1996) 209–244.
[17] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reason, logic programming, and N-person games, Artificial Intelligence 77 (1995) 321–357.
[18] P.E. Dunne, TPI-disputes and proof by clausal tableaux, Technical Report ULCS-02-008, Department of Comp. Sci., University of Liverpool, 2002.
[19] P.E. Dunne, T.J.M. Bench-Capon, Coherence in finite argument systems, Artificial Intelligence 141 (2002) 187–203.
[20] M. Fisher, A resolution method for temporal logic, in: R. Myopoulos, J. Reiter (Eds.), Proc. IJCAI-91, Sydney, Australia, Morgan Kaufmann, San Mateo, CA, 1991, pp. 99–104.
[21] G. Gentzen, Investigations into logical deductions, 1935, in: M.E. Szabo (Ed.), The Collected Papers of Gerhard Gentzen, North-Holland, Amsterdam, 1969, pp. 68–131.
[22] T.F. Gordon, The Pleadings Game: An Artificial Intelligence Model of Procedural Justice, Kluwer Academic, Dordrecht, 1995.
[23] A. Haken, The intractability of resolution, Theoret. Comput. Sci. 39 (2–3) (1985) 297–308.
[24] H. Jakobovits, D. Vermeir, Dialectic semantics for argumentation frameworks, in: Proc. Seventh International Conference on Artificial Intelligence and Law (ICAIL-99), ACM SIGART, ACM Press, New York, 1999, pp. 53–62.
[25] A.R. Lodder, Dialaw: On legal justification and Dialogue Games, PhD Thesis, University of Maastricht, 1998.
[26] R.P. Loui, Process and policy: Resource-bounded nondemonstrative reasoning, COMPINT: Computational Intelligence: An Internat. J. 14 (1998) 1–38.
[27] J.D. MacKenzie, Question-begging in non-cumulative systems, J. Philos. Logic 8 (1) (1978) 117–133.
[28] T. Pitassi, A. Urquhart, The complexity of the Hajós calculus, SIAM J. Discrete Math. 8 (3) (1995) 464–483.
[29] H. Prakken, G. Sartor, Argument-based extended logic programming with defeasible priorities, J. Appl. Non-Classical Logics 7 (1997) 25–75.
[30] P. Pudlák, The lengths of proofs, in: S.R. Buss (Ed.), Handbook of Proof Theory, in: Studies in Logic and the Foundations of Mathematics, Vol. 137, North-Holland, Amsterdam, 1998, Chapter 8, pp. 547–637.
[31] P. Pudlák, On the complexity of propositional calculus, in: Sets and Proofs, Invited papers from Logic Colloquium'97, Cambridge Univ. Press, Cambridge, 1999, pp. 197–218.
[32] C. Reed, Dialogue frames in agent communications, in: Y. Demazeau (Ed.), Proc. 3rd International Conference on Multi-Agent Systems (ICMAS-98), IEEE Press, 1998, pp. 246–253.
[33] R.M. Smullyan, First-order Logic, Springer, New York, 1968 (reprinted Dover Press, New York, 1995).
[34] S. Toulmin, The Uses of Argument, Cambridge University Press, Cambridge, 1959.
[35] A. Urquhart, The complexity of Gentzen systems for propositional logic, Theoret. Comput. Sci. 66 (1) (1989) 87–97.
[36] A. Urquhart, The complexity of propositional proofs, Bull. Symbolic Logic 1 (4) (1995) 425–467.
[37] G. Vreeswijk, Defeasible dialectics: A controversy-oriented approach towards defeasible argumentation, J. Logic Comput. 3 (3) (1993) 317–334.
[38] G. Vreeswijk, H. Prakken, Credulous and sceptical argument games for preferred semantics, in: Proc. JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence, in: Lecture Notes in Artificial Intelligence, Vol. 1919, Springer, Berlin, 2000, pp. 224–238.
[39] D.N. Walton, E.C.W. Krabbe, Committment in Dialogue: Basic Concepts of Interpersonal Reasoning, Univ. of New York Press, New York, 1995.

On the Complexity of Linking Deductive and Abstract Argument Systems

On the Complexity of Linking Deductive and Abstract Argument Systems

Michael Wooldridge and **Paul E. Dunne**
Dept of Computer Science
University of Liverpool
Liverpool L69 3BX, UK
mjw,ped@csc.liv.ac.uk

Simon Parsons
Brooklyn College
CUNY
Brooklyn 11210 NY
parsons@sci.brooklyn.cuny.edu

Abstract

We investigate the computational complexity of a number of questions relating to deductive argument systems, in particular the complexity of linking deductive and more abstract argument systems. We start by presenting a simple model of deductive arguments based on propositional logic, and define logical equivalence and defeat over individual arguments. We then extend logical equivalence to sets of arguments, and show that the problem of checking equivalence of argument sets is co-NP-complete. We also show that the problem of checking that an argument set contains no two logically equivalent arguments is NP-complete, while the problem of checking that a set of arguments is maximal (i.e., that no argument could be added without such an argument being logically equivalent to one that is already present) is co-NP-complete. We then show that checking whether a digraph over an argument set is sound with respect to the defeat relation is co-NP-complete, while the problem of showing that such a digraph is complete is NP-complete, and the problem of showing both soundness and completeness is D^p-complete.

Introduction

Argumentation is the process of attempting to construct rationally justifiable set of beliefs (Prakken & Vreeswijk 2001), and is increasingly used as a mechanism to support interaction in multiagent systems (Parsons, Wooldridge, & Amgoud 2003). The argumentation process typically starts with a knowledge base that contains logical conflicts, and is hence inconsistent: argumentation can be understood as the process of extracting a rationally justifiable position from this inconsistent starting point. Essentially two different approaches to formalizing arguments have been put forward in the literature. The first is the *abstract argument* framework of (Dung 1995). In this framework, the starting point is simply a digraph, with vertices in the graph corresponding to arguments, and edges in the graph representing the notion of attack, or defeat, between arguments. Abstract argument systems are so-called because they abstract away from the internal structure and properties of individual arguments, and focus instead simply on the attack relation between arguments. An alternative approach, which instead gives arguments some internal, logical structure, is that of *deductive* argument systems (Pollock 1992; 1994; Krause *et al.* 1995; Amgoud 1999; Besnard & Hunter 2001; Parsons, Wooldridge, & Amgoud 2003).

Our aim in this paper is to investigate the computational complexity of a number of questions relating to deductive argument systems, but in particular, the complexity of *linking* deductive argument systems and abstract argument systems. Such a linking seems to be necessary if we are to use argumentation systems, for example as the basis of inter-agent communication in multiagent systems. To capture the internal structure and intended meaning of arguments, we need something of deductive arguments, while to capture the interactions between arguments, and to formulate appropriate solution concepts, we clearly need something akin to an abstract argument system. Thus, to be practically useful, it seems an argumentation framework must have linked elements of both deductive and abstract argument systems.

We start by presenting a simple model of deductive arguments, and define notions logical equivalence and defeat over individual arguments. We then extend logical equivalence to sets of arguments, and show that the problem of checking equivalence of argument sets is co-NP-complete. We also show that the problem of checking that an argument set is *distinct* (i.e., contains no two logically equivalent arguments) is NP-complete, while the problem of checking that a set of arguments is *maximal* (i.e., that no argument could be added without such an argument being logically equivalent to one that is already present) is co-NP-complete. We then show that checking whether a graph over an argument set is sound with respect to the defeat relation is co-NP-complete, while the problem of showing that such a graph is complete is NP-complete, and the problem of showing both soundness and completeness is D^p-complete.

Deductive Arguments, Defeat, & Equivalence

We present the model of deductive arguments that we work with throughout the remainder of this paper. This model is closely related to those of (Besnard & Hunter 2001; Parsons, Wooldridge, & Amgoud 2003). Let $\Phi_0 = \{p, q, \ldots\}$ be a finite, fixed, non-empty vocabulary of Boolean variables, and let Φ denote the set of (well-formed) formulae of propositional logic over Φ_0, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" (for truth) and "$\bot$" (for

falsity). We refer to a finite subset of Φ as a *database*, and use $\Delta, \Delta', \Delta_0, \ldots$ as variables ranging over the set of Φ-databases. We assume a conventional semantic consequence relation "$\models$" for propositional logic, writing $\Delta \models \varphi$ to mean that φ is a logical consequence of the database Δ. We write $\models \varphi$ as a shorthand for $\emptyset \models \varphi$; thus $\models \varphi$ means that φ is a tautology. We denote the fact that formulae $\varphi, \psi \in \Phi$ are *logically equivalent* by $\varphi \sim \psi$; thus $\varphi \sim \psi$ means that $\models \varphi \leftrightarrow \psi$. Note that "$\sim$" is a *meta-language* relation symbol, which should not be confused with the object-language bi-conditional operator "$\leftrightarrow$".

If $\Delta \subseteq \Phi$ is a database, then an argument, α, over Δ is a pair $\alpha = (C, S)$ where $C \in \Phi$ is a propositional formula which we refer to as the *conclusion* of the argument, and $S \subseteq \Delta$ ($S \neq \emptyset$) is a subset of Δ which we refer to as the *support* of the argument, such that $S \models C$, i.e., C is a logical consequence of S. Notice that we omit two constraints on arguments that are commonly assumed in the literature, namely that S is consistent, and that S is minimal (Besnard & Hunter 2001; Parsons, Wooldridge, & Amgoud 2003). Of the two, minimality is generally regarded as an aesthetic criterion, rather than technically essential. Consistency is more important, and of course by relaxing this constraint we admit into our analysis some scenarios that do not seem to have any useful interpretation; but of course this does not invalidate the results we present. Let $A(\Delta)$ denote the set of arguments over Δ. If α is an argument, then we denote the support of α by $S(\alpha)$ and the conclusion of α by $C(\alpha)$.

There are two common ways of defining defeat between two deductive arguments (Prakken & Vreeswijk 2001): rebuttal (where the conclusion of each argument is logically equivalent to the negation of the conclusion of the other) and undercut (where the conclusion of the attacker contradicts some part of the support of the other). It is not hard to see that the rebuttal relation between arguments will be symmetric, and this potentially limits its value as an analytical concept (Besnard & Hunter 2001). We therefore focus on undercutting (Besnard & Hunter 2001). There are in fact a number of ways of defining undercuts (Besnard & Hunter 2001), and our choice here is largely motivated by simplicity. We say an argument α_1 *defeats* an argument α_2 (written $def(\alpha_1, \alpha_2)$) if $\exists \varphi \in S(\alpha_2)$ such that $C(\alpha_1) \sim \neg\varphi$. The problem of checking whether $def(\alpha_1, \alpha_2)$ is obviously co-NP-complete.

Now, consider the circumstances under which two arguments may be said to be *equivalent*. First, consider the equivalence of formulae: we have two obvious interpretations of "equivalence" w.r.t. formulae. The first is simply that of syntactic equivalence, which we denote by equality ("="); and the second is that of *logical* equivalence, which you will recall is denoted by $\sim$. Let us now extend these notions to arguments. We write $\alpha_1 = \alpha_2$ to mean that α_1 and α_2 are *syntactically equivalent*, i.e., that $C(\alpha_1) = C(\alpha_2)$ and $S(\alpha_1) = S(\alpha_2)$. What about logical equivalence of arguments? We will say that arguments α_1 and α_2 over a database Δ are logically equivalent (written: $\alpha_1 \approx \alpha_2$) iff $C(\alpha_1) \sim C(\alpha_2)$, i.e., if they are in complete logical agreement w.r.t. the conclusion. The point here is that this ignores syntactic variations in the presentation of the argument's conclusion. In general, for any database Δ, there will be more syntactically distinct arguments over Δ than there will be logically distinct arguments over Δ. To see this, simply consider a database $\Delta_1 = \{p \rightarrow q, p\}$, and the arguments $(q, \{p, p \rightarrow q\})$ and $(\neg\neg q, \{p, p \rightarrow q\})$, both of which are syntactically distinct, but $(q, \{p, p \rightarrow q\}) \approx (\neg\neg q, \{p, p \rightarrow q\})$. It is evident that checking whether two arguments are logically equivalent is co-NP-complete.

Argument Sets, Distinctness, & Maximality

We now change our focus to consider subsets of arguments. We extend our notion of equivalence of arguments to subsets of arguments as follows. We say $X_1 \subseteq A(\Delta)$ and $X_2 \subseteq A(\Delta)$ are logically equivalent (written: $X_1 \lessapprox X_2$) iff there exists a bijection $f : X_1 \rightarrow X_2$ such that $\forall \alpha \in X_1$, we have $\alpha \approx f(\alpha)$. The $\exists\forall$ pattern of quantifiers in the checking of equivalence of argument sets suggests that this is computationally harder than checking equivalence of formulae or arguments – perhaps Σ_2^p-complete (Papadimitriou 1994, pp.424–425). However, this turns out not to be the case:

Theorem 1 *The problem of checking equivalence of argument sets is co-*NP*-complete.*

Proof: It will be convenient to work with the complementary problem – INEQUIVALENT ARGUMENT SETS (IAS) – and to prove that IAS is NP-complete. Showing NP-hardness is straightforward, so we focus on membership of NP. Let $X_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ and $X_2 = \{\beta_1, \ldots, \beta_m\}$. where $\alpha_i = (\varphi_i, S_i)$ and $\beta_i = (\psi_i, T_i)$ are arguments in $A(\Delta)$. We use Φ and Ψ to denote the sets $\{\varphi_1, \ldots, \varphi_m\}$ and $\{\psi_1, \ldots, \psi_m\}$.

Let $B(\Phi, \Psi, E)$ be the bipartite graph on (disjoint) sets of m vertices labelled Φ and Ψ and whose edges are $E = \{\{\varphi_i, \psi_j\} : S(\alpha_i) = S(\beta_j)\}$. For $\underline{a} \in \langle \bot, \top \rangle^n$, $B_{\underline{a}}(\Phi, \Psi, F_{\underline{a}})$ is the subgraph of $B(\Phi, \Psi, E)$ containing only the edges $F_{\underline{a}} = \{\{\varphi_i, \psi_j\} : \varphi_i(\underline{a}) = \psi_j(\underline{a})\}$. For $Y \subseteq \langle \bot, \top \rangle^n$, $B_Y(\Phi, \Psi, F_Y)$ is the subgraph of $B(\Phi, \Psi, E)$ whose edges are $F_Y = \{\{\varphi_i, \psi_j\} : \varphi_i(\underline{a}) = \psi_j(\underline{a}) \text{ for every } \underline{a} \in Y\}$. Letting TOT denote the set $\langle \bot, \top \rangle^n$ it is easy to see the following: $X_1 \lessapprox X_2$ iff $B_{\text{TOT}}(\Phi, \Psi, F_{\text{TOT}})$ contains a perfect matching, i.e., a subset of m edges defining a bijective mapping between Φ and Ψ.

For $V \subset \Phi$, let $\Gamma(V, B_Y)$ denote the subset of Ψ formed by $\Gamma(V, B_Y) = \{\psi_j : \{\varphi_i, \psi_j\} \in F_Y \text{ for some } \varphi_i \in V\}$. From the König-Hall Theorem ((Berge 1976, Ch. 7, Thm. 5, p. 134)), there is a perfect matching in $B_{\text{TOT}}(\Phi, \Psi, F_{\text{TOT}})$ if and only if $\forall\ V \subset \Phi\ |\Gamma(V, B_{\text{TOT}})| \geq |V|$. Suppose it is the case that $X_1 \lessapprox X_2$, i.e., $\langle X_1, X_2 \rangle$ is accepted as an instance of IAS. From the argument above, this happens if and only if $B_{\text{TOT}}(\Phi, \Psi, F_{\text{TOT}})$ does *not* contain a perfect matching, and thus there will be some strict subset of Φ, V say, for which $|\Gamma(V, B_{\text{TOT}})| < |V|$.

These observations lead to the following NP algorithm to decide IAS.

1. Non-deterministically choose some $V \subset \Phi$.
2. Non-deterministically choose some $W \subset \Psi$ of size $|V| - 1$.

3. Non-deterministically select a set F of $|V|.(m - |W|) < m^2$ distinct $\underline{a} \in \text{TOT}$.
4. For each $\varphi \in V$ and each $\psi \in \Psi \setminus W$ check that if $\{\varphi, \psi\} \in E$ then there is some $\underline{a} \in F$ for which $\psi(\underline{a}) \neq \varphi(\underline{a})$.

The last stage involves only polynomially many tests, each of which requires simply evaluating two formulae on a given instantiation.

To see that this algorithm is correct it suffices to observe that the structure $\langle V, W, F\rangle$ witnesses that $\langle X_1, X_2\rangle$ is a positive instance of IAS: there are at most $|V|.(m - |V|)$ pairs $\langle\varphi_i, \psi_j\rangle$ with $\varphi_i \in V$ and $\psi_j \notin W$. If $S(\alpha_i) = S(\beta_j)$ then $\{\varphi_i, \psi_j\} \in E$, however, we only require one instantiation $\underline{a} \in \text{TOT}$ in order to eliminate this edge from F_{TOT}. Thus we need at most $|V|.(m - |V|) < m^2$ instantiations to remove all edges between V and $\Psi \setminus W$. It follows that $\langle V, W, F\rangle$ provides a polynomial length certificate for membership in IAS. □

Next, we define the notion of *distinctness* for sets of arguments. The intuition is that a set of arguments is distinct if it does not contain duplicated arguments, where duplication is measured with respect to logical equivalence of arguments. Formally, we say argument set $X \subseteq A(\Delta)$ is distinct iff $\forall\alpha_1, \alpha_2 \in X$: if $(\alpha_1 \neq \alpha_2)$ then $(\alpha_1 \not\sim \alpha_2)$.

Theorem 2 *The problem of checking whether an argument set is distinct is* NP*-complete.*

Proof: Membership of NP follows from the fact that checking distinctness of an argument set $X \subseteq A(\Delta)$ reduces to the $|X|^2$ independent satisfiability checks, i.e., verifying that for all $\alpha_1, \alpha_2 \in X$, such that $\alpha_1 \neq \alpha_2$, the formula $(C(\alpha_1) \wedge \neg C(\alpha_2)) \vee (\neg C(\alpha_1) \wedge C(\alpha_2))$ is satisfiable. For NP-hardness, we reduce SAT. Given a SAT instance φ, simply check that the argument set $X_1 = \{(\varphi, \{\varphi\}), (\bot, \{\bot\})\}$ is distinct. □

We say a set of arguments $X \subseteq A(\Delta)$ is *maximal* w.r.t. Δ if it is not possible to add an argument from $A(\Delta)$ to X without X becoming indistinct. Intuitively, if a set of arguments X is maximal with respect to some database Δ, then it contains all the arguments that can be made about Δ: it is not possible to pick an argument from $A(\Delta)$ without duplicating a member of X (where duplication is measured with respect to logical equivalence). Notice that distinctness does not of course imply maximality, but *neither does maximality imply distinctness*. That is, an argument set can be maximal but contain duplicates. Indeed, the set $A(\Delta)$ of all arguments that can be made with respect to a database Δ is an obvious example of such a maximal but indistinct set.

In analysing the computational complexity of checking maximality it will be convenient to work with its complementary form, which we dub NON-MAS. In this problem, instances $\langle\Delta, X\rangle$ are accepted iff there is some $\beta \in A(\Delta)$ whose conclusion is *not* logically equivalent to that of any argument in X. We observe that the "natural" formulation of NON-MAS in determining the status of an instance $\langle\Delta, X\rangle$ is as

$$\exists\, S \subseteq \Delta,\ \varphi\ :\ \langle\varphi, S\rangle \in A(\Delta)\ \wedge\ \bigwedge_{\alpha\in X}\ (\varphi\ \not\sim\ C(\alpha))$$

This formulation raises two difficulties: unless restrictions are placed on φ, the structure $\langle\varphi, S\rangle$ which must be validated as an argument in $A(\Delta)$ may have size that is not polynomially bounded in the size of the instance $\langle\Delta, X\rangle$[1]; we have to validate $S \models \varphi$ (in general CO-NP-hard) and $\varphi \not\sim C(\alpha)$ for each $\alpha \in X$ (in general, NP-hard). Overall, even assuming the restriction to formulae whose size, $|\varphi|$ (measured as the number of literals occurring in φ) is bounded by some polynomial in the instance size, it would appear that NON-MAS is "unlikely" to be decidable by an NP computation: with the formulation and the restriction imposed we get only a Σ_2^p algorithm. However, not only is it unnecessary *explicitly* to restrict φ, we may also validate $S \models \varphi$ for all *relevant* φ in *polynomial time* (in the size of the instance $\langle\Delta, X\rangle$). In this way we can show that NON-MAS $\in$ NP a result which, coupled with the easy proof that NON-MAS is NP-hard, allows us to deduce that NON-MAS is NP-complete.

The following result is central to our subsequent proof that NON-MAS $\in$ NP.

Lemma 1 *Let $\langle\Delta, X\rangle$ be an instance of* NON-MAS *and n be the number of Boolean variables in the vocabulary Φ_0 of Δ. The instance $\langle\Delta, X\rangle$ is accepted if and only if there is a propositional formula, φ, over Φ_0 for which* all *of the following properties hold:*

a. $\Delta \models \varphi$.
b. *φ is a* CNF*-formula containing at most $|X|$ clauses, each of which is defined by* exactly *n literals, so that $|\varphi| \leq n|X|$.*
c. $\forall\,\alpha \in X$, $C(\alpha) \not\sim \varphi$.

Proof: From the definition of NON-MAS it is immediate that if φ with the properties (a) through (c) exists, then it is certainly the case the $\langle\Delta, X\rangle$ is accepted as an instance of NON-MAS: the argument (φ, Δ) being distinguished from all arguments in X.

For the converse implication, suppose it is the case that $(\psi, S) \in A(\Delta)$ and that for each $\alpha \in X$, we have $\psi \not\sim C(\alpha)$. We first observe that, since $S \models \psi$ it is certainly the case that $\Delta \models \psi$. For any instantiation $\underline{a} \in \langle\top, \bot\rangle^n$, let $\chi_{\underline{a}}$ be the propositional formula given as the disjunction over all literals over Φ_0 that take the value $\bot$ under $\underline{a}$: thus $\chi_{\underline{a}}(\underline{b}) = \bot \Leftrightarrow \underline{b} = \underline{a}$. Consider the set of (full) instantiations of Φ_0, $\bot(\psi)$, defined by,

$$\bot(\psi)\ =\ \{\underline{a} \in \langle\top, \bot\rangle^n\ :\ \psi(\underline{a}) = \bot\}$$

It is well-known that for any propositional formula, ψ, is logically equivalent to the formula ψ_{CNF} defined via

$$\psi_{\text{CNF}}\ =\ \bigwedge_{\underline{a}\in\bot(\psi)}\ \chi_{\underline{a}}$$

Thus from $\Delta \models \psi$ we have $\Delta \models \psi_{\text{CNF}}$. In addition, however, for any subset R of $\bot(\psi)$ it further holds that

$$\Delta \models\ \left(\bigwedge_{\underline{a}\in R}\ \chi_{\underline{a}}\right)$$

[1]Given a database Δ, it may be the case that there is some φ such that $\Delta \models \varphi$ and the shortest formula ψ such that $\varphi \sim \psi$ is of length exponential in the size of Δ. In other words, there could be arguments that we can construct from a database whose conclusion is necessarily exponential in the size of the database.

Since $\psi_{\text{CNF}} \not\sim C(\alpha)$ for any $\alpha \in X$, it follows that we can identify $k = |X|$ instantiations, $\langle \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k \rangle$ for which $\psi_{\text{CNF}}(\underline{a}_i) \neq C(\alpha_i)(\underline{a}_i)$. We now define the subset R_X of $\bot(\psi)$ to contain $\{ \underline{a}_i : C(\alpha_i)(\underline{a}_i) = \top \}$ and fix φ (the propositional formula whose existence we wish to establish) as $\bigwedge_{\underline{a} \in R_X} \chi_{\underline{a}}$. For φ defined in this way, from our earlier analysis: $\Delta \models \varphi$, as required by (a); φ is in CNF, and contains at most $|X|$ clauses (since $|R_X| \leq |X|$) with each clause defined from exactly n literals – as required by (b); finally $\varphi \not\sim C(\alpha)$ for any $\alpha \in X$ – as required by (c). To see that (c) does hold true of φ it suffices to observe that $\bot(\varphi) = R_X$ so that if $\underline{a}_i \in R_X$ then $\varphi(\underline{a}_i) = \bot$ and (from the definition of R_X) $C(\alpha_i)(\underline{a}_i) = \top$; similarly if $\underline{a}_i \notin R_X$ then $C(\alpha_i)(\underline{a}_i) = \bot$ and (from the fact that $\bot(\varphi) = R_X$) $\varphi(\underline{a}_i) = \top$.

In total if it is the case that $\langle \Delta, X \rangle$ is accepted as an instance of NON-MAS, then we can identify some φ with the properties (a)–(c) described in the Lemma statement. □

Given this, we can now prove that:

Theorem 3 *The problem of checking maximality of argument sets is* CO-NP*-complete.*

Proof: We prove the equivalent result that NON-MAS is NP-complete. We first show NON-MAS is NP-hard using a reduction from SAT. Given an instance φ of SAT, consider the database $\Delta = \{\neg\varphi\}$ with $X \subseteq A(\Delta)$ chosen to be $\{(\top, \{\neg\varphi\})\}$. We claim that φ is satisfiable iff X is not maximal w.r.t. Δ.

We now show that NON-MAS $\in$ NP. Consider the following non-deterministic algorithm.

1. For each $\alpha_i \in X$, non-deterministically choose an instantiation, $\underline{a}_i$ of Φ_0.
2. Construct the formula $\varphi = \bigwedge_{\underline{a}_i : C(\alpha_i)(\underline{a}_i) = \top} \chi_{\underline{a}_i}$
3. Test if $\Delta \models \varphi$, accepting if this is the case.

By Lemma 1 it is certainly the case that $\langle \Delta, X \rangle$ is accepted as an instance of NON-MAS if and only if the algorithm described has an accepting computation. Stages (1) and (2) can clearly be realised in non-determininistic polynomial time. The final stage, however, is easily completed in deterministic polynomial-time: φ is a CNF formula for which $\bot(\varphi) = \{\underline{a}_i : C(\alpha_i)(\underline{a}_i) = \top\}$. Thus to verify $\Delta \models \varphi$ it suffices to check that for each $\underline{a} \in \bot(\varphi)$ some $\psi \in \Delta$ has $\psi(\underline{a}) = \bot$. Since $|\bot(\varphi)| \leq |X|$ this final stage takes time polynomial in the size of the instance $\langle \Delta, X \rangle$. □

Of particular interest to us are argument sets over Δ that are *both* maximal *and* distinct. We say a set of arguments X is *canonical* with respect to Δ if it is both maximal and distinct w.r.t. Δ. A canonical argument set thus represents a limit of what can be argued from a database without repetition. We will let $can(\Delta)$ denote the canonical argument sets of Δ, so $can(\Delta) \subseteq 2^{A(\Delta)}$. First, we prove that every non-empty database Δ has a canonical argument set.

Theorem 4 *For all $\Delta \neq \emptyset \subseteq \Phi$, $can(\Delta) \neq \emptyset$.*

Proof: We use the same proof idea as Lindenbaum's lemma. Let $\sigma : \alpha_0, \alpha_1, \ldots$ be an enumeration of arguments over Δ: such an enumeration clearly exists. Corresponding to σ, define a sequence of argument sets $X_0, X_1, \ldots$ where $X_0 = \{\alpha_0\}$, and for $n > 0$,

$$X_n = \begin{cases} X_{n-1} \cup \{\alpha_n\} & \text{if } X_{n-1} \cup \{\alpha_n\} \text{ is distinct} \\ X_{n-1} & \text{otherwise.} \end{cases}$$

Finally, define an argument set X by: $X = \bigcup_{n=0}^{\infty} X_n$. By construction, X will be a canonical argument set of Δ. □

The following, easily established result gives our motivation for using the term "canonical".

Fact 1 *Canonical argument sets are logically equivalent. That is, $\forall X_1, X_2 \in can(\Delta)$: $X_1 \approxeq X_2$.*

We note, in addition, the following consequence of Lemma 1, the proof of which is omitted.

Corollary 1 *If $X \in can(\Delta)$, then $|X| = 2^{|\bot(\delta)|}$, where $\delta = \bigwedge_{\varphi \in \Delta} \varphi$.*

Argument Graphs

Let us now consider the issue of linking deductive and abstract argument systems. Given a set of arguments $X \subseteq A(\Delta)$, the defeat predicate $def(\cdots)$ induces a graph $(X, \{(\alpha_1, \alpha_2) \mid \alpha_1, \alpha_2 \in X, def(\alpha_1, \alpha_2)\})$ over X, which can obviously be understood as being analogous to the graph structures of Dung's abstract argument systems (Dung 1995). Note that there are some technical difficulties involved in "lifting" a defeat relation to a Dung argumentation graph in this way. In particular, Besnard and Hunter show that unless modified, Dung's notion of an admissible set turns out to collapse under this interpretation; although the notion of an admissible set can be refined to make more sense when interpreted for deductive argument systems, this comes at the cost of eliminating some apparently reasonable cases (Besnard & Hunter 2001). Thus, solution concepts which make sense when studied with respect to arbitrary graphs do not necessarily make sense when the defeat relation is given a concrete interpretation in terms of deductive arguments, suggesting a need for refined versions of these. However, the issue of formulating appropriate Dung-style solution concepts for deductive argument systems is somewhat tangential to the paper at hand, and we shall not investigate this particular issue. Instead, we focus on the problems of establishing links between deductive and more abstract argument systems.

To motivate the discussion, suppose we are given a set of arguments $X \subseteq A(\Delta)$ (for some Δ), and a graph $G_X = (X, E \subseteq X \times X)$, so that the vertices of G_X are the members of X. How might X and G_X be related? Two obvious questions then suggest themselves:

1. *Soundness*: Does G_X "correctly" represents the defeat relation $def(\cdots)$ over X? Formally, G_X will be sound with respect to X iff $\forall \alpha_1, \alpha_2$, if $G_X(\alpha_1, \alpha_2)$ then $def(\alpha_1, \alpha_2)$.
2. *Completeness*: Does G_X "completely" represents the defeat relation $def(\cdots)$ over X? Formally, G_X will be complete with respect to X iff $\forall \alpha_1, \alpha_2$, if $def(\alpha_1, \alpha_2)$ then $G_X(\alpha_1, \alpha_2)$.

Theorem 5 *Given a set of arguments X, the problem of checking whether a graph $G_X = (X, E \subseteq X \times X)$ is sound with respect to the defeat relation def$(\cdots)$ over X is co-*NP*-complete.*

Proof: Consider membership of co-NP. Recall that soundness of G_X with respect to X means that $\forall \alpha_i, \alpha_j \in X$, if $G_X(\alpha_i, \alpha_j)$ then *def*(α_i, α_j). We work with the complement of the problem, i.e., the problem of showing that $\exists \alpha_i, \alpha_j \in X$ such that $G_X(\alpha_i, \alpha_j)$ and not *def*(α_i, α_j). The following NP algorithm decides the problem: (i) Guess $\alpha_i, \alpha_j \in X$ and k propositional valuations $\xi_1, \ldots, \xi_k$, where $k = |S(\alpha_j)|$; (ii) Verify that $G_X(\alpha_i, \alpha_j)$ and that $\xi_1 \models (C(\alpha_i) \wedge \psi_1) \vee \neg(C(\alpha_i) \vee \psi_1), \xi_2 \models (C(\alpha_i) \wedge \psi_2) \vee \neg(C(\alpha_i) \vee \psi_2), \ldots, \xi_k \models (C(\alpha_i) \wedge \psi_k) \vee \neg(C(\alpha_i) \vee \psi_k)$, where $S(\alpha_j) = \{\psi_1, \ldots, \psi_k\}$. The algorithm is clearly in NP. For hardness, we reduce TAUT, the problem of deciding whether a propositional logic formula φ is a tautology. Given a TAUT instance φ, define $G_X = (\{\alpha_1, \alpha_2\}, \{(\alpha_1, \alpha_2)\})$ where $\alpha_1 = (\varphi, \{\varphi\})$ and $\alpha_2 = (\bot, \{\bot\})$. G_X is sound w.r.t. X iff φ is a tautology. □

Theorem 6 *Given a set of arguments X, the problem of checking whether a graph $G_X = (X, E \subseteq X \times X)$ is complete with respect to the defeat relation def$(\cdots)$ over X is* NP*-complete.*

Proof: Membership of NP is by the following algorithm: For each $\alpha_1, \alpha_2 \in X$ such that not $G_X(\alpha_1, \alpha_2)$, and for each $\varphi \in S(\alpha_2)$ guess a valuation ξ and verify that $\xi \models C(\alpha_1) \wedge \varphi$. For NP-hardness we reduce SAT. Given a SAT instance φ, define $G_X = (\{\alpha_1, \alpha_2\}, \{(\alpha_1, \alpha_1), (\alpha_2, \alpha_2), (\alpha_2, \alpha_1)\})$ where $\alpha_1 = (\varphi, \{\varphi\})$ and $\alpha_2 = (\top, \{\top\})$. G_X is complete w.r.t. X iff φ is satisfiable. □

Now consider the problem of determining whether a graph G_X over a set of arguments X is both sound *and* complete.

Theorem 7 *Given a set of arguments X, the problem of checking whether a graph $G_X = (X, E \subseteq X \times X)$ is both sound and complete with respect to the defeat relation def$(\cdots)$ over X is D^p-complete.*

Proof: Membership in D^p follows from Theorems 6 and Theorem 5. For completeness, we reduce SAT-UNSAT (Papadimitriou 1994, p.415), instances of which comprise a pair $\langle \varphi, \psi \rangle$ of propositional formulae. Such an instance is accepted if φ is satisfiable *and* ψ is unsatisfiable. First, from φ we create a new formula $\varphi^* = \varphi \wedge p$, where p is a new Boolean variable, which does not appear in either φ or ψ. The formula φ^* has the following properties, all of which are used in what follows: (i) φ^* will be satisfiable iff φ is satisfiable; (ii) φ^* is *not* a tautology, even if φ is; and (iii) neither $\varphi^* \sim \psi$ nor $\varphi^* \sim \neg\psi$. We then create an argument set $X = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ where: $\alpha_1 = (\varphi^*, \{\varphi^*\})$, $\alpha_2 = (\psi, \{\psi\})$, $\alpha_3 = (\top, \{\top\})$, and $\alpha_4 = (\bot, \{\bot\})$. Our argument graph $G_X = (X, E)$ has X as defined above, and $E = \{(\alpha_2, \alpha_3), (\alpha_2, \alpha_2), (\alpha_3, \alpha_4), (\alpha_4, \alpha_3)\}$. Table 1 relates the graph G_X in this construction to the defeat relation *def*$(\cdots)$ induced over X for possible properties of φ and ψ. Note that the properties in the *def*$(\cdots)$ column of Table 1 are by established by simple propositional logic reasoning.

(α_i, α_j)	*def*(α_i, α_j)?	$G_X(\alpha_i, \alpha_j)$?
(α_1, α_1)	no	no
(α_1, α_2)	no	no
(α_1, α_3)	iff φ^* is unsatisfiable	no
(α_1, α_4)	no (since φ^* is not a tautology)	no
(α_2, α_1)	no	no
(α_2, α_2)	no	no
(α_2, α_3)	iff ψ is unsatisfiable	yes
(α_2, α_4)	iff ψ is a tautology	no
(α_3, α_1)	iff φ^* is unsatisfiable	no
(α_3, α_2)	iff ψ is unsatisfiable	yes
(α_3, α_3)	no	no
(α_3, α_4)	yes	yes
(α_4, α_1)	no (since φ^* is not a tautology)	no
(α_4, α_2)	iff ψ is a tautology	no
(α_4, α_3)	yes	yes
(α_4, α_4)	no	no

Table 1: Defeat relation and argument graph properties for the construction of Theorem 7.

We claim that φ is satisfiable and ψ is unsatisfiable iff G_X is sound and complete w.r.t. X. ($\rightarrow$) Suppose φ is satisfiable and ψ is unsatisfiable. We must show that G_X is sound and complete w.r.t. X. Soundness means that if $G_X(\alpha_i, \alpha_j)$ then *def*(α_i, α_j). With respect to Table 1, this means showing that a "yes" in the $G_X(\alpha_i, \alpha_j)$ column implies a "yes" in the *def*(α_i, α_j) column. That *def*(α_3, α_4) and *def*(α_4, α_3) is obvious, so consider whether *def*(α_2, α_3): since by assumption ψ is unsatisfiable, then it must defeat α_3. Completeness means that if not $G_X(\alpha_i, \alpha_j)$ then not $G_X(\alpha_i, \alpha_j)$. This can be verified by examination of Table 1. ($\leftarrow$) Suppose G_X is sound and complete w.r.t. X, i.e., that $G_X(\alpha_i, \alpha_j)$ iff *def*(α_i, α_j). We must show that this implies φ is satisfiable and ψ is unsatisfiable. This can be done by examination of cases in Table 1. □

Suppose that instead of being given a graph over a set of arguments, we are given an *arbitrary* graph, $G = (V, E)$, where V is simply an abstract set of vertices and $E \subseteq V \times V$, and we are asked whether G "captures" a given deductive argument system. Here, G really is simply a Dung-style argument system: the nodes in the graph are not arguments, and hence we are not given any interpretation of them with respect to the given deductive argument system. How might we establish a link between such a graph and a deductive argument system? It depends on the way in which the deductive argument system itself is presented:

1. as a graph G_X over $A(\Delta)$;
2. as a (sub)set of arguments $X \subseteq A(\Delta)$; or
3. as a database $\Delta \subseteq \Phi$.

In the first case, we are given *both* a graph $G = (V, E)$ and an argument graph $G_X = (X, E_X \subseteq X \times X)$. The problems of soundness and completeness in this case reduce to standard graph theoretic problems: Soundness means checking whether G is isomorphic to some subgraph of G_X, while completeness means checking whether G_X is isomorphic to some subgraph of G, and checking soundness *and*

completeness means checking that G and G_X are isomorphic. From standard results in complexity theory, (in particular the fact that the SUBGRAPH ISOMORPHISM problem is NP-complete) it follows immediately that the problems of checking soundness or completeness for this representation are both NP-complete. The problem of checking both soundness *and* completeness, however, is exactly the well known open problem GRAPH ISOMORPHISM. A classification of the complexity of this problem would in itself represent a major event in the theory of computational complexity.

With respect to the second representation, we are given a set of arguments $X \subseteq A(\Delta)$ and a graph $G = (V, E)$. Here, we have less information: we have no argument graph to compare G with, just a set of arguments X. Thus we do not know a priori what the vertices of G are supposed to correspond to in X – we thus need to "interpret" vertices in G with respect to members of X in our definitions of soundness and completeness. Formally, we will say:

- a graph $G = (V, E)$ is a *sound abstraction* of a set of arguments $X \subseteq A(\Delta)$ if there exists an injective function $f : V \rightarrow X$ such that $\forall v_1, v_2 \in V$, if $G(v_1, v_2)$ then $def(f(v_1), f(v_2))$; and
- a graph $G = (V, e)$ is a *complete abstraction* of $X \subseteq A(\Delta)$ iff there exists an injective function $f : X \rightarrow V$ such that $\forall \alpha_1, \alpha_2 \in X$, if $def(\alpha_1, \alpha_2)$ then $G(f(\alpha_1), f(\alpha_2))$.

The proofs of Theorems 5 and 6 can be readily adapted to show the following:

Theorem 8 *The problem of checking whether a graph G is a sound abstraction of a set of arguments $X \subseteq A(\Delta)$, is co-*NP*-hard, while the problem of checking whether a graph G is a complete abstraction of a set of arguments $X \subseteq A(\Delta)$, is* NP*-complete.*

With respect to the third representation, we are given simply a database Δ and a graph $G = (V, E)$. This case seems the most elaborate computationally but also perhaps the least interesting practically. Once again, we are given even less information to work with: we only have the database of formulae from which arguments may be constructed. So, how are we to interpret the soundness and completeness questions? Recalling that for any database $\Delta \subseteq \Phi$, the set of canonical argument sets over Δ is denoted by $can(\Delta)$, we can give the following interpretation to soundness and completeness for graphs $G = (V, E)$ against databases Δ:

- a graph $G = (V, E)$ is a *sound canonical abstraction* of a database Δ if $\exists X \in can(\Delta)$ such that G is a sound abstraction of X; and
- a graph $G = (V, E)$ is a *complete canonical abstraction* of a database Δ if $\exists X \in can(\Delta)$ such that G is a complete abstraction of X.

It should be clear that these concepts, are much more baroque (and much less amenable to formal analysis) than those we have studied above. They involve quantifying over canonical argument sets, which as we noted in Corollary 1 will be exponentially large in the number of falsifying assignments for Δ, and hence in general doubly exponential in the number of Boolean variables. We will thus not investigate these latter problems further here.

Related Work & Conclusions

The work described in this paper has been concerned with the computational complexity of answering certain questions about sets of arguments. The particular questions we have considered have not previously been considered, but there are several authors whose work is related to ours in one way or another. For example, the work of Besnard and Hunter (Besnard & Hunter 2001) has some elements in common with our work — a definition of equivalence between arguments that is the same as ours, and a notion of canonicity of argument. Their work, however, is focussed exclusively on the properties of a specific deductive argumentation system while ours deals with properties that apply to a range of argumentation systems. Other authors have considered the computational complexity of answering questions related to arguments. Most notable, perhaps, is the work of Dimopoulos *et al.* who have investigated the complexity of computing the acceptability of individual deductive arguments — a surprisingly hard process because of the recursive nature of the relationships between arguments (Dimopoulos, Nebel, & Toni 2002).

References

Amgoud, L. 1999. *Contribution a l'integration des préferences dans le raisonnement argumentatif*. Ph.D. Dissertation, l'Université Paul Sabatier, Toulouse. (In French).

Berge, C. 1976. *Graphs and Hypergraphs*. North-Holland.

Besnard, P., and Hunter, A. 2001. A logic-based theory of deductive arguments. *Artificial Intelligence* 128:203–235.

Dimopoulos, Y.; Nebel, B.; and Toni, F. 2002. On the computational complexity of assumption-based argumentation for default reasoning. *Artificial Intelligence* 141(1):57–78.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *Artificial Intelligence* 77:321–357.

Krause, P.; Ambler, S.; Elvang-Gøransson, M.; and Fox, J. 1995. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence* 11:113–131.

Papadimitriou, C. H. 1994. *Computational Complexity*. Addison-Wesley: Reading, MA.

Parsons, S.; Wooldridge, M.; and Amgoud, L. 2003. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation* 13(3):347–376.

Pollock, J. L. 1992. How to reason defeasibly. *Artificial Intelligence* 57:1–42.

Pollock, J. L. 1994. Justification and defeat. *Artificial Intelligence* 67:377–407.

Prakken, H., and Vreeswijk, G. 2001. Logics for defeasible argumentation. In Gabbay, D., and Guenther, F., eds., *Handbook of Philosophical Logic (second edition)*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Computational Properties of Argument Systems Satisfying Graph-theoretic Constraints

Available online at www.sciencedirect.com
ScienceDirect
Artificial Intelligence 171 (2007) 701–729
Artificial Intelligence
www.elsevier.com/locate/artint

Computational properties of argument systems satisfying graph-theoretic constraints

Paul E. Dunne

Department of Computer Science, The University of Liverpool, Liverpool, United Kingdom

Received 25 October 2006; received in revised form 9 March 2007; accepted 19 March 2007

Available online 30 March 2007

Abstract

One difficulty that arises in abstract argument systems is that many natural questions regarding argument acceptability are, in general, computationally intractable having been classified as complete for classes such as NP, co-NP, and Π_2^p. In consequence, a number of researchers have considered methods for specialising the structure of such systems so as to identify classes for which efficient decision processes exist. In this paper the effect of a number of graph-theoretic restrictions is considered: k-partite systems ($k \geqslant 2$) in which the set of arguments may be partitioned into k sets each of which is conflict-free; systems in which the numbers of attacks originating from and made upon any argument are bounded; planar systems; and, finally, those of k-bounded treewidth. For the class of *bipartite* graphs, it is shown that determining the acceptability status of a *specific* argument can be accomplished in polynomial-time under both credulous and sceptical semantics. In addition we establish the existence of polynomial time methods for systems having bounded treewidth when deciding the following: whether a given (set of) arguments is credulously accepted; if the system has a non-empty preferred extension; has a stable extension; is coherent; has at least one sceptically accepted argument. In contrast to these positive results, however, deciding whether an arbitrary *set* of arguments is "collectively acceptable" remains NP-complete in bipartite systems. Furthermore for both planar and bounded degree systems the principal decision problems are as hard as the unrestricted cases. In deriving these latter results we introduce various concepts of "simulating" a general argument system by a restricted class so allowing any argument system to be translated to one which has both bounded degree and is planar. Finally, for the development of basic argument systems to so-called "value-based frameworks", we present results indicating that decision problems known to be intractable in their most general form remain so even under quite severe graph-theoretic restrictions. In particular the problem of deciding "subjective acceptability" continues to be NP-complete even when the underlying graph is a binary tree.

Keywords: Computational properties of argumentation; Argumentation frameworks; Computational complexity

1. Introduction

Since their introduction in the seminal work of Dung [23] abstract argument systems have proven to be a valuable paradigm with which to formalise divers semantics defining argument "acceptability". In these a key component is the concept of an "*attack*" relationship wherein the incompatibility of two arguments—p and q, say—may be expressed

E-mail address: ped@csc.liv.ac.uk.

0004-3702/$ – see front matter
doi:10.1016/j.artint.2007.03.006

in terms of one of these "attacking" the other: such relationships may be presented independently of any internal structure of the individual arguments concerned so that the properties of the overall argument system, e.g. which of its arguments may be defended against any attack and which are indefensible, depend solely on the attack relationship rather than properties of individual argument schemata. Among other applications, this abstract view of argumentation has proven to be a powerful and flexible approach to modelling reasoning in a variety of non-classical logics, e.g. [15, 20,23].

We present the formal definitions underpinning argument systems in Section 2, including two of the widely-studied admissibility semantics—preferred and stable—introduced in [23]: at this point we simply observe that these describe differing conditions which a maximal set of mutually compatible arguments, S, must satisfy in order to be admissible within some argument system comprising arguments $\mathcal{X}$ with attack relationship $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$.

Despite the descriptive power offered by abstract argument systems one significant problem is the apparent intractability of many natural questions concerning acceptability under all but the most elementary semantics: such intractability classifications encompassing NP-completeness and co-NP-completeness results of Dimopoulos and Torres [21] and the Π_2^p-completeness classifications presented in Dunne and Bench-Capon [27]. Motivated, at least to some degree, by these negative results a number of researchers have considered mechanisms by which argument systems may be specialised or enriched so that the resulting structures admit efficient decision procedures. Two main strategies are evident: the first, and the principal focus of the present paper, has been to identify purely graph-theoretic conditions leading to tractable methods for those cases within which these are satisfied; the second, which itself may be coupled with graph-theoretic restrictions, is to consider additional structural aspects in developing the basic argument and attack relationship form. Under the first category, [23] already identifies *directed acyclic graphs* (DAGs) as a suitable class, while recent work of Coste-Marquis et al. [17] has shown that *symmetric* argument systems—those in which p attacks q if and only if q attacks p—also form a tractable class. Graph-theoretic considerations also feature significantly in work of Baroni et al. [5,6].

Probably the two most important exemplars of the second approach are the *Preference based* argumentation frameworks of Amgoud and Cayrol [1] and *Value based* argumentation frameworks introduced by Bench-Capon [9]. While the supporting motivation for both formalisms is, perhaps, more concerned with providing interpretations and resolution of issues arising from the presence of multiple maximal admissible sets which are mutually inconsistent, both approaches start with an arbitrary argument system, $\langle \mathcal{X}, \mathcal{A} \rangle$, and reduce it to an *acyclic* system, $\langle \mathcal{X}, \mathcal{B} \rangle$ in which $\mathcal{B} \subseteq \mathcal{A}$ this reduction being determined via some additional relationship $\mathcal{R}$: the main distinction between [1] and [9] being the exact manner in which $\mathcal{R}$ is defined.

In this paper some further classes of graph-theoretic restrictions are considered: *k-partite* directed graphs, *bounded degree* systems, planar argument systems, and those with k-bounded treewidth. In the first class, for which the case $k = 2$ is of particular interest, the argument set $\mathcal{X}$ may be partitioned into k pairwise disjoint subsets—$\langle \mathcal{X}_1, \ldots, \mathcal{X}_k \rangle$ such that every attack in $\mathcal{A}$ involves arguments belonging to *different* sets in this partition: the special case, $k = 2$, defines the class of *bipartite* directed graphs. The bounded degree class limits the number of attacks on (the argument's *in-degree*) and by (its *out-degree*) any $x \in \mathcal{X}$, i.e. $|\{y\colon \langle y, x \rangle \in \mathcal{A}\}|$ and $|\{y\colon \langle x, y \rangle \in \mathcal{A}\}|$ are bounded by given values (p, q): again the special case $p = q = 2$ is of particular interest. The concept of *treewidth*, introduced in work of Robertson and Seymour, e.g. [37], has proven to be a useful aid in developing efficient methods for many computationally hard problems, e.g. via very general approaches such as those of Arnborg et al. [3], Courcelle [18,19], even in the case of problems which are not directly graph-theoretic in nature, e.g. Gottlob et al. [32].

In the remainder of this paper formal background and definitions are given in Section 2 together with the decision questions considered. Section 3 describes two important systems from [21,27] that feature in a number of subsequent hardness proofs, while Sections 4 and 5 present results concerning, respectively, k-partite and bounded degree directed graphs. Planarity is discussed in Section 6 and properties of bounded treewidth systems are given in Section 7. The range of results proved indicate that for many of these restrictions it is possible to obtain efficient decision processes: both credulous and sceptical acceptability of individual arguments may be determined in polynomial time within bipartite systems. In the case of systems with bounded treewidth, similar positive results for a number of properties are derivable using a number of deep results originally obtained in [18,19] and developed in [3]. It turns out, however, that for the development of standard argument systems into value-based frameworks we do not obtain more efficient mechanisms simply by limiting the graph structure: in Section 8 we show that two basic decision problems in this model remain hard even when the underlying graph structure is a binary tree. Conclusions and developments are discussed in Section 9.

2. Finite argument systems—basic definitions

The following concepts were introduced in Dung [23].

Definition 1. An *argument system* is a pair $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$, in which $\mathcal{X}$ is a finite set of *arguments* and $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ is the *attack relationship* for $\mathcal{H}$. A pair $\langle x, y \rangle \in \mathcal{A}$ is referred to as '*y is attacked by x*' or '*x attacks y*'. For $S \subseteq \mathcal{X}$, the set of arguments $\mathcal{N}^+(S)$ is given by

$$\mathcal{N}^+(S) = \bigcup_{x \in S} \{y\colon\ \langle x, y \rangle \in \mathcal{A}\}$$

The convention of excluding "self-attacking" arguments, also observed in [17], is assumed, i.e. for all $x \in \mathcal{X}$, $\langle x, x \rangle \notin \mathcal{A}$. For R, S subsets of arguments in the system $\mathcal{H}(\mathcal{X}, \mathcal{A})$, we say that

(a) $s \in S$ is *attacked* by R if there is some $r \in R$ such that $\langle r, s \rangle \in \mathcal{A}$, i.e.

$$\mathit{attacks}(R, s) \equiv_{\text{def}} \exists r \in R\ s.t. \langle r, s \rangle \in \mathcal{A}$$

(b) $x \in \mathcal{X}$ is *acceptable with respect to* S if for every $y \in \mathcal{X}$ that attacks x there is some $z \in S$ that attacks y, i.e.

$$\mathit{acceptable}(x, S) \equiv_{\text{def}} \forall y \in \mathcal{X}\ \langle y, x \rangle \in \mathcal{A} \Rightarrow \mathit{attacks}(S, y)$$

(c) S is *conflict-free* if no argument in S is attacked by any other argument in S,

$$\mathit{cf}(S) \equiv_{\text{def}} \forall y \in S\ \neg \mathit{attacks}(S, y)$$

(d) A conflict-free set S is *admissible* if every $y \in S$ is acceptable w.r.t S. That is,

$$\mathit{adm}(S) \equiv_{\text{def}} \mathit{cf}(S) \wedge \big(\forall y \in S\ \mathit{acceptable}(y, S)\big)$$

(e) S is a *preferred extension* if it is a maximal (with respect to $\subseteq$) admissible set.

$$\mathit{pref}(S) \equiv_{\text{def}} \mathit{adm}(S) \wedge \big(\forall T \subseteq \mathcal{X}\ S \subset T \Rightarrow \neg \mathit{adm}(T)\big)$$

(f) S is a *stable extension* if S is conflict free and every $y \notin S$ is attacked by S.

$$\mathit{stable}(S) \equiv_{\text{def}} \mathit{cf}(S) \wedge \big(\forall x \in \mathcal{X}\ (x \notin S) \Rightarrow \mathit{attacks}(S, x)\big)$$

(g) $\mathcal{H}$ is *coherent* if every preferred extension in $\mathcal{H}$ is also a stable extension.

$$\mathit{coherent}\big(\mathcal{H}(\mathcal{X}, \mathcal{A})\big) \equiv_{\text{def}} \forall S \subseteq \mathcal{X}\ \mathit{pref}(S) \Rightarrow \mathit{stable}(S)$$

Following the terminology of [17], $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is *symmetric* if for every pair of arguments x, y in $\mathcal{X}$ it holds that $\langle x, y \rangle \in \mathcal{A}$ if and only if $\langle y, x \rangle \in \mathcal{A}$.

An argument x is *credulously accepted* if there is *some* preferred extension containing it; x is *sceptically accepted* if it is a member of *every* preferred extension.

Combining the ideas of credulous and sceptical with preferred and stable, provides a number of differing formalisations for the concept of a set of arguments being acceptable: these are sometimes referred to as the credulous preferred/stable semantics and sceptical preferred/stable semantics. Unless we explicitly state otherwise we will usually be considering the preferred variant of these.

We make one further assumption regarding the *graph-theoretic* structure of argument systems: as an *undirected* graph, $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is *connected*. In informal terms, this states that systems do *not* consist of two or more "isolated" graphs.[1]

The concepts of credulous and sceptical acceptance motivate a number of decision problems, summarised in Table 1, that have been considered in [21,27].

[1] With a single exception—that of Corollary 2—all of our results hold without this assumption.

Table 1
Decision problems in finite argument systems

	Problem	Instance	Question	Complexity
(a)	CA	$\mathcal{H}(\mathcal{X},\mathcal{A}), x \in \mathcal{X}$	Is x credulously accepted?	NP-complete
(b)	CA^{S}	$\mathcal{H}(\mathcal{X},\mathcal{A}), x \in \mathcal{X}$	Is x in any *stable* extension?	NP-complete
(c)	PREF-EXT	$\mathcal{H}(\mathcal{X},\mathcal{A})$	Does $\mathcal{H}$ have a *non-empty* preferred extension?	NP-complete
(d)	STAB-EXT	$\mathcal{H}(\mathcal{X},\mathcal{A})$	Does $\mathcal{H}$ have any stable extension?	NP-complete
(e)	SA^{S}	$\mathcal{H}(\mathcal{X},\mathcal{A}), x \in \mathcal{X}$	Is x in *every* stable extension?	CO-NP–complete
(f)	SA	$\mathcal{H}(\mathcal{X},\mathcal{A}), x \in \mathcal{X}$	Is x sceptically accepted?	Π_2^p-complete
(g)	COHERENT	$\mathcal{H}(\mathcal{X},\mathcal{A})$	Is the system $\mathcal{H}$ coherent?	Π_2^p-complete

These problems (a–d) are NP-complete,[2] while (e) is CO-NP-complete follows from results of [21]. Problems (f) and (g) were shown to be Π_2^p-complete in [27].

These questions are formulated in terms of *single* arguments, it will be useful to consider analogous concepts with respect to *sets*. Thus $\text{CA}_{\{\}}$ denotes the decision problem whose instances are an argument system $\langle \mathcal{X}, \mathcal{A} \rangle$ together with a subset S of $\mathcal{X}$: the instance being accepted if there is a preferred extension T for which $S \subseteq T$. Similarly, $\text{SA}_{\{\}}$ accepts instances for which S is a subset of *every* preferred extension.

In contrast, we have the following more positive results.

Fact 2.

(a) *Every argument system $\mathcal{H}$ has at least one preferred extension* (*Dung* [23]).
(b) *If $\mathcal{H}(\mathcal{X},\mathcal{A})$ is a* DAG *then $\mathcal{H}$ has a unique preferred extension. This is also a stable extension and may be found in time linear in $|\mathcal{X}| + |\mathcal{A}|$* (*Dung* [23]).
(c) *If $\mathcal{H}(\mathcal{X},\mathcal{A})$ is symmetric then* CA, SA, $\text{CA}_{\{\}}$, *and* $\text{SA}_{\{\}}$ *are all polynomial-time decidable. Furthermore $\mathcal{H}$ is coherent* (*Coste-Marquis et al.* [17]).
(d) *If $\mathcal{H}(\mathcal{X},\mathcal{A})$ contains no odd-length simple directed cycles, then $\mathcal{H}$ is coherent* (*Dunne and Bench-Capon* [27]).
(e) *If $\mathcal{H}(\mathcal{X},\mathcal{A})$ is coherent then* $\text{SA}(\mathcal{H}, x)$ *can be decided in co-*NP.

Fact 2(e) is an easy consequence of the sceptical acceptance methods described in work of Vreeswijk and Prakken [40].

While Fact 2(a) ensures the existence of a preferred extension—a property that is not guaranteed to be the case for stable extensions—it is possible that the *empty set* of arguments (which is always admissible) is the unique such extension. Noting Table 1(c), whether a given argument system $\mathcal{H}(\mathcal{X},\mathcal{A})$ has a non-empty preferred extension is unlikely to be efficiently decidable in general.

3. The argument systems $\mathcal{H}_\Phi$ and $\mathcal{G}_\Phi$ and their properties

A number of our subsequent hardness proofs regarding various graph-theoretic restrictions are obtained by transforming argument systems used in earlier reductions of [21,27] in classifying the decision problems CA and SA. In order to avoid repetition it will be useful formally to introduce the two systems used in these contexts. Noting that both systems take as their starting point some CNF formula Φ, we denote these subsequently by $\mathcal{H}_\Phi$ and $\mathcal{G}_\Phi$.

3.1. The system $\mathcal{H}_\Phi$

The form we describe is virtually identical to that first presented by Dimopoulos and Torres [21, Theorem 5.1, p. 227] where it is used to establish NP-hardness of CA via a reduction from 3-SAT.

Given a CNF formula $\Phi(Z_n) = \bigwedge_{j=1}^{m} C_j$ with each C_j a disjunction of literals from $\{z_1, \ldots, z_n, \neg z_1, \ldots, \neg z_n\}$, the argument system, $\mathcal{H}_\Phi(\mathcal{X},\mathcal{A})$ has

[2] An earlier, unpublished, NP-completeness proof for (d) is attributed to Chvatal in [31, GT57, p. 204]. We note also the result of Fraenkel [30] mentioned in Section 6.

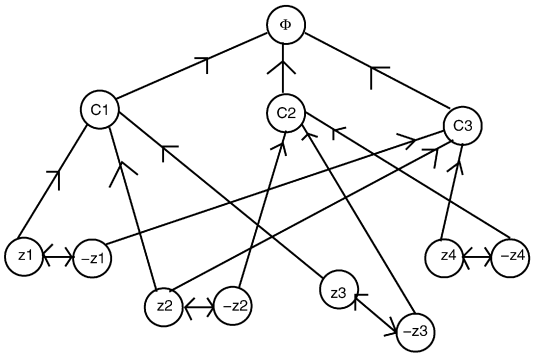

Fig. 1. The argument system $\mathcal{H}_\Phi$.

$$\mathcal{X} = \{\Phi, C_1, \ldots, C_m\} \cup \{z_i, \neg z_i \colon 1 \leqslant i \leqslant n\}$$
$$\mathcal{A} = \big\{\langle C_j, \Phi\rangle \colon 1 \leqslant j \leqslant m\big\} \cup \big\{\langle z_i, \neg z_i\rangle, \langle \neg z_i, z_i\rangle \colon 1 \leqslant i \leqslant n\big\}$$
$$\cup \big\{\langle z_i, C_j\rangle \colon z_i \text{ occurs in } C_j\big\} \cup \big\{\langle \neg z_i, C_j\rangle \colon \neg z_i \text{ occurs in } C_j\big\}$$

Fig. 1 illustrates $\mathcal{H}_\Phi$ for the CNF $\Phi(z_1, z_2, z_3, z_4) = (z_1 \vee z_2 \vee z_3)(\neg z_2 \vee \neg z_3 \vee \neg z_4)(\neg z_1 \vee z_2 \vee z_4)$.

Fact 3. *(See Dimopoulos and Torres [21].) Let $\Phi(Z_n)$ be an instance of 3-*SAT*, i.e. a 3-*CNF *formula. Then $\Phi(Z_n)$ is satisfiable if and only if* $\text{CA}(\mathcal{H}_\Phi(\mathcal{X}, \mathcal{A}), \Phi)$.

3.2. The system $\mathcal{G}_\Phi$

The proof that SA is Π_2^p-complete from [27] uses a reduction from QSAT_2^Π instances of which may, without loss of generality, be restricted to 3-CNF formulae,[3] $\Phi(Y_n, Z_n)$, accepted if $\forall\alpha_Y \exists\beta_Z \Phi(\alpha_Y, \beta_Z)$, i.e. for every instantiation of the propositional variables Y_n (α_Y) there is some instantiation of Z_n (β_Z) for which $\langle\alpha_Y, \beta_Z\rangle$ satisfies Φ.

The system $\mathcal{G}_\Phi(\mathcal{W}, \mathcal{B})$ is formed from the system $\mathcal{H}_\Phi(\mathcal{X}, \mathcal{A})$, i.e. $\mathcal{X} \subset \mathcal{W}$ and $\mathcal{A} \subset \mathcal{B}$, so that

$$\mathcal{W} = \{\Phi, C_1, \ldots, C_m\} \cup \{y_i, \neg y_i, z_i, \neg z_i \colon 1 \leqslant i \leqslant n\} \cup \{b_1, b_2, b_3\}$$
$$\mathcal{B} = \big\{\langle C_j, \Phi\rangle \colon 1 \leqslant j \leqslant m\big\}$$
$$\cup \big\{\langle y_i, \neg y_i\rangle, \langle \neg y_i, y_i\rangle, \langle z_i, \neg z_i\rangle, \langle \neg z_i, z_i\rangle \colon 1 \leqslant i \leqslant n\big\}$$
$$\cup \big\{\langle y_i, C_j\rangle \colon y_i \text{ occurs in } C_j\big\} \cup \big\{\langle \neg y_i, C_j\rangle \colon \neg y_i \text{ occurs in } C_j\big\}$$
$$\cup \big\{\langle z_i, C_j\rangle \colon z_i \text{ occurs in } C_j\big\} \cup \big\{\langle \neg z_i, C_j\rangle \colon \neg z_i \text{ occurs in } C_j\big\}$$
$$\cup \big\{\langle \Phi, b_1\rangle, \langle \Phi, b_2\rangle, \langle \Phi, b_3\rangle, \langle b_1, b_2\rangle, \langle b_2, b_3\rangle, \langle b_3, b_1\rangle\big\}$$
$$\cup \big\{\langle b_1, z_i\rangle, \langle b_1, \neg z_i\rangle \colon 1 \leqslant i \leqslant n\big\}$$

The resulting system is shown in Fig. 2.

Fact 4. *(See Dunne and Bench-Capon [27].)*

(a) *$\Phi(Y_n, Z_n)$ is accepted as an instance of* QSAT_2^Π *if and only if* $\text{SA}(\mathcal{G}_\Phi, \Phi)$.
(b) *$\Phi(Y_n, Z_n)$ is accepted as an instance of* QSAT_2^Π *if and only if $\mathcal{G}_\Phi$ is coherent.*

[3] The proof in [27], in fact presents a more general translation from arbitrary propositional formulae over the logical basis $\{\wedge, \vee, \neg\}$. Exploiting such translations is a significant motivating device underlying Theorem 12 and, in particular, accounts for the original context of Fig. 8.

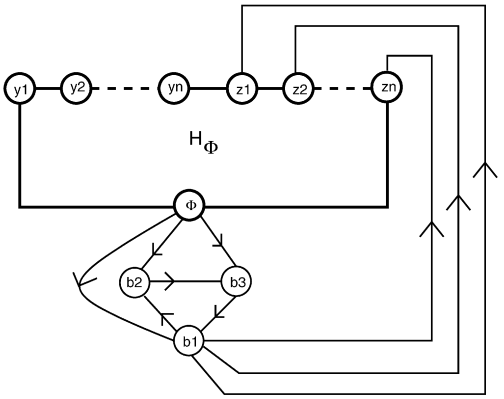

Fig. 2. The argument system $\mathcal{G}_\Phi$.

Table 2
Complexity-theoretic properties of k-partite argument systems

	Decision problem	Complexity
(a)	$\text{CA}^{(2)}$	Polynomial-time
(b)	$\text{CA}^{(3)}$	NP-complete
(c)	$\text{CA}^{(2)}_{\{\}}$	NP-complete
(d)	$\text{SA}^{(2)}$	Polynomial-time
(e)	$\text{SA}^{(3)}$	Π_2^p-complete
(f)	$\text{SA}^{(2)}_{\{\}}$	Polynomial-time
(g)	$\text{SA}^{(3)}_{\{\}}$	Π_2^p-complete
(h)	$\text{COHERENT}^{(2)}$	Trivial
(i)	$\text{COHERENT}^{(3)}$	Π_2^p-complete

4. k-partite argument systems

In this section[4] we consider the effect on problem complexity of restricting systems to be k-partite. Our results are summarised in Table 2.

Definition 5. An argument system $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is *k-partite* if there is a partition of $\mathcal{X}$ into k sets $\langle \mathcal{X}_1, \ldots, \mathcal{X}_k \rangle$ such that

$$\forall \langle y, z \rangle \in \mathcal{A} \quad y \in \mathcal{X}_i \Rightarrow z \notin \mathcal{X}_i$$

The term *bipartite* will be used for the case $k = 2$. It should be noted that, since there is no insistence that each of the partition members be non-empty, any k-partite system is, trivially, also a $(k + t)$-partite system for every $t \geqslant 0$. We use the notation $\Gamma^{(k)}$ for the set of all k-partite argument systems.

The notations $\text{CA}^{(k)}$, $\text{SA}^{(k)}$, $\text{CA}^{(k)}_{\{\}}$, and $\text{SA}^{(k)}_{\{\}}$ will be used to distinguish the various avatars of the decision problems of interest when instances are required to be k-partite argument systems. Similarly we use $\text{COHERENT}^{(k)}$ to denote the problem of deciding whether a k-partite argument system is coherent. In instances of these problems it is assumed that $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is presented using an appropriate partition of $\mathcal{X}$ into k disjoint sets $\langle \mathcal{X}_1, \ldots, \mathcal{X}_k \rangle$.[5]

[4] The results presented in Theorems 6, 7, and 8 first appeared in a preliminary version of this paper in [26].

[5] Without this, problems arise when checking if an *arbitrary* argument system, $\mathcal{H}$, is k-partite: for $k \geqslant 3$ the corresponding decision question is NP-complete.

1: $i := 0$; $\mathcal{Y}_0 := \mathcal{Y}$; $\mathcal{A}_0 := \mathcal{A}$
2: **repeat**
3: $\quad i := i + 1$
4: $\quad \mathcal{Y}_i := \mathcal{Y}_{i-1} \setminus \{y \in \mathcal{Y}_{i-1}\colon \exists z \in \mathcal{Z}\colon \langle z, y\rangle \in \mathcal{A}_{i-1}$ and $|\{y \in \mathcal{Y}_{i-1}\colon \langle y, z\rangle \in \mathcal{A}_{i-1}\}| = 0\}$
5: $\quad \mathcal{A}_i := \mathcal{A}_{i-1} \setminus \{\langle y, z\rangle\colon y \notin \mathcal{Y}_i \setminus \mathcal{Y}_{i-1}\}$
6: **until** $\mathcal{Y}_i = \mathcal{Y}_{i-1}$
7: **return** $\mathcal{Y}_i$

Algorithm 1. Credulous acceptance in bipartite systems.

We first deal with the case of bipartite argument systems ($k = 2$). For other values of k it is noted that the classifications are largely straightforward consequences of the graph-theoretic constructions described in Section 3.[6] Notice that it is straightforward to deal with the claim made in Table 2(h): a bipartite argument system cannot have any odd-length cycles, and thus coherence is ensured via Fact 2(d). In contrast to *undirected* graph structures, the *absence* of odd-length directed cycles, while necessary, is not a *sufficient* condition for an argument system to be bipartite; *symmetric* systems, however, are bipartite systems if and only if the associated undirected graph contains no odd-length cycles. We note that if we consider the class of systems OCF such that $\mathcal{H} \in$ OCF if $\mathcal{H}$ has no odd-length cycles—bipartite systems being a strict subset of OCF—then CA restricted to instances in OCF remains NP-complete; while SA, for such instances, is CO-NP-complete: the former is immediate from the system $\mathcal{H}_\Phi$ of Section 3.1 since $\mathcal{H}_\Phi \in$ OCF; the latter is easily derived by introducing an additional argument Ψ into $\mathcal{H}_\Phi$ whose sole attacker is Φ. The resulting system is in OCF and is such that Ψ is sceptically accepted if and only if $\Phi(Z_n)$ is unsatisfiable. Membership in CO-NP follows since it suffices to check that Ψ belongs to every *stable* extension.

The main idea underlying Algorithm 1 in proving Theorem 6 is as follows: in a bipartite argument system, $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ attackers of an argument $y \in \mathcal{Y}$ can only be arguments $z \in \mathcal{Z}$, and defences to such attacks must, themselves, also be arguments in $\mathcal{Y}$. It follows, therefore, that those arguments of $\mathcal{Y}$ that are attacked by members of $\mathcal{Z}$ upon which no counterattack exists cannot be admissible. Moreover, attacks on $\mathcal{Z}$ furnished by such arguments play no useful function (as counterattacks) and may be eliminated from $\mathcal{A}$, a process that can lead to further arguments in $\mathcal{Z}$ becoming unattacked. By iterating the process of removing indefensible arguments in $\mathcal{Y}$ and their associated attacks on $\mathcal{Z}$, this algorithm identifies an admissible subset of $\mathcal{Y}$.

Theorem 6.

(a) $\text{CA}^{(2)}$ *is polynomial-time decidable.*
(b) $\text{SA}^{(2)}$ *is polynomial-time decidable.*

Proof. For (a), given a bipartite argument system, $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ and $x \in \mathcal{Y} \cup \mathcal{Z}$, without loss of generality assume that $x \in \mathcal{Y}$. Consider the subset, S of $\mathcal{Y}$ that is formed by Algorithm 1.

We claim that $\text{CA}^{(2)}(\mathcal{B}, x)$ holds if and only if $x \in S$.

Suppose first that $x \in S \subseteq \mathcal{Y}$. Since $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ is a bipartite argument system it follows that S is conflict-free. Now consider any argument $z \in \mathcal{Z}$ that attacks S: it must be the case that there is some $y \in S$ that counterattacks z for otherwise at least one argument would have been removed from S at Step 4. In total, S is conflict-free and every argument in S is acceptable with respect to S, i.e. S is an admissible set containing x which is, hence, credulously accepted.

On the other hand, suppose that x is credulously accepted. Let S be the subset of $\mathcal{Y}$ returned and suppose for the sake of contradiction that $x \notin S$: then there must be some iteration of the algorithm during which $x \in \mathcal{Y}_{i-1}$ but $x \notin \mathcal{Y}_i$. In order for this to occur, we must have a sequence of arguments $\langle z_0, z_1, \ldots, z_i\rangle$ in $\mathcal{Z}$ with the property that $|\{y \in \mathcal{Y}_j\colon \langle y, z_j\rangle \in \mathcal{A}_j\}| = 0$ with $\langle z_i, x\rangle \in \mathcal{A}_i$. Now any argument y' of $\mathcal{Y}$ attacked by z_0 cannot be credulously accepted since there is no counterattack on z_0 available. It follows that the attacks $\langle y', z\rangle$ provided by such arguments cannot play an effective role in defending another argument and thus can be removed. Continuing in this way, it follows that no argument y'' that is attacked by z_1 is credulously accepted: the only attackers of z_1 are arguments of

[6] It is noted, however, that some extension of the basic construction in Section 3.2 is needed for the results of Table 2(g) and (i).

$\mathcal{Y}$ that are attacked by z_0 and these, we have seen, are indefensible. In total, $x \notin S$ would imply that x is indefensible, a conclusion which contradicts the assumption that x was credulously accepted.

The preceding analysis establishes the algorithm's correctness. The proof of (a) is completed by noting that it runs in polynomial-time: there are at most $|\mathcal{Y}|$ iterations of the main loop each taking only polynomially many (in $|\mathcal{Y} \cup \mathcal{Z}| + |\mathcal{A}|$) steps.

Part (b) follows from (a), Table 2(h) and the observation of [40] that, in coherent systems, an argument is sceptically accepted if and only if none of its attackers are credulously accepted. □

Examining the structure of Algorithm 1 allows the following characterisation of the set of preferred extensions in bipartite systems.

Corollary 1. *Given a bipartite argument system $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ let $S_\mathcal{Y}$ and $S_\mathcal{Z}$ be the subsets of $\mathcal{Y}$ and $\mathcal{Z}$ returned by Algorithm* 1. *Let $T \subseteq \mathcal{Y} \cup \mathcal{Z}$ and for $T_\mathcal{Y}$ (resp. $T_\mathcal{Z}$) denote $T \cap \mathcal{Y}$ (resp. $T \cap \mathcal{Z}$). Then T is a preferred extension of $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ if and only if*

$$T_\mathcal{Y} \subseteq S_\mathcal{Y} \quad \textit{and} \quad \mathcal{N}^+(T_\mathcal{Y}) = \mathcal{Z} \setminus T_\mathcal{Z}$$
$$T_\mathcal{Z} \subseteq S_\mathcal{Z} \quad \textit{and} \quad \mathcal{N}^+(T_\mathcal{Z}) = \mathcal{Y} \setminus T_\mathcal{Y}$$

Proof. Suppose that T is a preferred extension of $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$. It is certainly the case that $T_\mathcal{Y} \subseteq S_\mathcal{Y}$ and $T_\mathcal{Z} \subseteq S_\mathcal{Z}$: each argument in $T_\mathcal{Y}$ is credulously accepted and, from Theorem 6(a), $y \in \mathcal{Y}$ is so accepted if and only if $y \in S_\mathcal{Y}$. Furthermore, since T must also be a stable extension any argument not belonging to $T_\mathcal{Y}$ (resp. $T_\mathcal{Z}$) must be attacked by an argument in $T_\mathcal{Z}$ (resp. $T_\mathcal{Y}$), i.e. $\mathcal{N}^+(T_\mathcal{Y}) = \mathcal{Z} \setminus T_\mathcal{Z}$ and $\mathcal{N}^+(T_\mathcal{Z}) = \mathcal{Y} \setminus T_\mathcal{Y}$. We deduce that if T is a preferred extension of $\mathcal{B}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ then it has the form required.

Conversely suppose that T satisfies $T_\mathcal{Y} \subseteq S_\mathcal{Y}$, $T_\mathcal{Z} \subseteq S_\mathcal{Z}$, $\mathcal{N}^+(T_\mathcal{Y}) = \mathcal{Z} \setminus T_\mathcal{Z}$, and $\mathcal{N}^+(T_\mathcal{Z}) = \mathcal{Y} \setminus T_\mathcal{Y}$. We claim that T is a preferred extension. Certainly T is conflict-free: $T_\mathcal{Z}$ and $T_\mathcal{Y}$ are conflict-free and it cannot be the case that $\langle y, z\rangle \in \mathcal{A}$ or $\langle z, y\rangle \in \mathcal{A}$ for any $y \in T_\mathcal{Y}$ and $z \in T_\mathcal{Z}$: this would contradict $\mathcal{N}^+(T_\mathcal{Y}) = \mathcal{Z} \setminus T_\mathcal{Z}$ or $\mathcal{N}^+(T_\mathcal{Z}) = \mathcal{Y} \setminus T_\mathcal{Z}$. That T must be a stable (and hence preferred) extension now follows by observing that any $x \notin T$ either is a member of $\mathcal{Y}$ (and thus is attacked by some $z \in T_\mathcal{Z}$) or a member of $\mathcal{Z}$ (and so attacked by some $y \in T_\mathcal{Y}$). □

Turning to the problems $\text{CA}_{\{\}}$ and $\text{SA}_{\{\}}$, [17] note that in many cases decision problems involving *sets* are "no harder" than the related questions formulated for specific arguments, e.g. for unrestricted argument systems, symmetric argument systems and DAGs, the upper bounds for $\text{CA}_{\{\}}$ and $\text{SA}_{\{\}}$ are identical to the corresponding upper bounds for CA and SA. In this light, the next result may appear somewhat surprising: although, as has just been shown, $\text{CA}^{(2)}$ is polynomial-time decidable, $\text{CA}^{(2)}_{\{\}}$ is likely to be noticeably harder.

Theorem 7.

(a) $\text{CA}^{(2)}_{\{\}}$ *is* NP*-complete, even for sets containing exactly two arguments.*

(b) $\text{SA}^{(2)}_{\{\}}$ *is polynomial-time decidable.*

Proof. For (a), that $\text{CA}^{(2)}_{\{\}} \in \text{NP}$ is easily demonstrated via the non-deterministic algorithm that guesses a subset T, checks $S \subseteq T$ and that T is admissible.

To show that $\text{CA}^{(2)}_{\{\}}$ is NP-hard we use a reduction from the problem *Monotone 3-CNF Satisfiability* (MCS) [31, p. 259], instances of which comprise a 3-CNF formula over a set of propositional variables $\{x_1, \ldots, x_n\}$,

$$\Phi(x_1, x_2, \ldots, x_n) = \bigwedge_{i=1}^{m} C_i = \bigwedge_{i=1}^{m} (y_{i,1} \vee y_{i,2} \vee y_{i,3})$$

and each clause, C_i, is defined using exactly three *positive* literals or exactly three *negated* literals, e.g. $(x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_4)$ would define a valid instance of MCS, however $(x_1 \vee \neg x_2 \vee x_3)$ would not. An instance Φ of MCS is accepted if and only if there is an instantiation, $\alpha \in \langle \top, \bot \rangle^n$ under which $\Phi(\alpha) = \top$.

Given $\Phi(x_1, \ldots, x_n)$ an instance of MCS let $\{C_1^+, \ldots, C_r^+\}$ be the subset of its clauses in which only positive literals occur and $\{D_1^\neg, \ldots, D_s^\neg\}$ those in which only negated literals are used. Consider the bipartite argument system $\mathcal{B}_{\text{MCS}}(\mathcal{Y}, \mathcal{Z}, \mathcal{A})$ whose arguments we denote by

$$\mathcal{Y} = \{\Phi^\neg, C_1^+, \ldots, C_r^+, \neg x_1, \ldots, \neg x_n\}$$
$$\mathcal{Z} = \{\Phi^+, D_1^\neg, \ldots, D_s^\neg, x_1, \ldots, x_n\}$$

and whose attack set $\mathcal{A}$ contains

$$\begin{aligned}&\left\{\langle x_j, \neg x_j\rangle, \langle \neg x_j, x_j\rangle\colon\ 1 \leqslant j \leqslant n\right\}\\ &\quad\cup \left\{\langle C_i^+, \Phi^+\rangle\colon\ 1 \leqslant i \leqslant r\right\} \cup \left\{\langle D_i^\neg, \Phi^\neg\rangle\colon\ 1 \leqslant i \leqslant s\right\}\\ &\quad\cup \left\{\langle \neg x_j, D_i^\neg\rangle\colon\ \neg x_j \text{ occurs in } D_i^\neg\right\}\\ &\quad\cup \left\{\langle x_j, C_i^+\rangle\colon\ x_j \text{ occurs in } C_i^+\right\}\end{aligned}$$

The instance of $\text{CA}_{\{\}}^{(2)}$ is completed by setting $S = \{\Phi^+, \Phi^\neg\}$.

Suppose that there is some preferred extension, T, of $\mathcal{B}_{\text{MCS}}$ for which $\{\Phi^+, \Phi^\neg\} \subseteq T$, i.e. that $\langle \mathcal{B}_{\text{MCS}}, S\rangle$ defines a positive instance of $\text{CA}_{\{\}}^{(2)}$. Then, for each C_i^+ some argument x_j with $\langle x_j, C_i^+\rangle \in \mathcal{A}$ must be in T (otherwise the attack $\langle C_i^+, \Phi^+\rangle$ is undefended); similarly for each $D_i^\neg$ some argument $\neg x_k$ with $\langle \neg x_k, D_i^\neg\rangle \in \mathcal{A}$ must be in T. It cannot be the case, however, that *both* x_j and $\neg x_j$ are in T. We can, thus, construct a satisfying instantiation of Φ via $x_j := \top$ if $x_j \in T$, and $x_j := \bot$ if $\neg x_j \in T$.

On the other hand suppose the instance Φ of MCS is satisfiable, using some instantiation α. In this case the set

$$\{\Phi^+, \Phi^\neg\} \cup \{x_j^+\colon\ x_j = \top \text{ under } \alpha\} \cup \{x_j^\neg\colon\ x_j = \bot \text{ under } \alpha\}$$

is easily seen to be admissible, so that $\langle \mathcal{B}_{\text{MCS}}, \{\Phi^+, \Phi^\neg\}\rangle$ defines a positive instance of $\text{CA}_{\{\}}^{(2)}$.

Part (b) follows easily from Theorem 6(b) since a set of arguments S is sceptically accepted if and only if each of its constituent members is sceptically accepted. □

The remaining cases in Table 2 are considered in the following theorem.

Theorem 8.

(a) $\forall k \geqslant 3$, $\text{CA}^{(k)}$ *is* NP*-complete.*
(b) $\forall k \geqslant 3$, $\text{SA}^{(k)}$ *and* $\text{COHERENT}^{(k)}$ *are* Π_2^p*-complete.*

Proof. The membership proofs are identical to those that hold for the unrestricted versions of each problem. For (a), NP-hardness follows by observing that the argument system $\mathcal{H}_\Phi$ given in Section 3.1 is 3-partite: using three colours—$\{R, B, G\}$ say—$\mathcal{H}_\Phi$ may be vertex 3-coloured by assigning R to $\{\Phi, z_1, \ldots, z_n\}$; B to $\{\neg z_1, \ldots, \neg z_n\}$ and G to $\{C_1, \ldots, C_m\}$. The proof of (b) requires techniques introduced in Section 5 applied to the construction $\mathcal{G}_\Phi$ of Section 3.2: details are given in Appendix A. □

5. Bounded degree systems

In contrast to many of the results of Section 4, the restriction considered in this section[7] does not lead to improved algorithmic methods. Our principal interest is in introducing the concept of a given class of argument systems being capable of "representing" another class. This is of interest for the following reason. Suppose that Π and Θ are properties of argument systems (where the formal definition of "property" will be clarified subsequently). Furthermore, suppose that any system with property Θ can be "represented" (in a sense to be made precise) by another system with property Π. Assuming such a representation can be constructed efficiently, we would be able to exploit algorithmic

[7] The presentation here is a revised and expanded treatment of ideas originally outlined in [26].

methods tailored to systems with property Π also to operate on systems with property Θ: given $\mathcal{H}$ (satisfying Θ), form $\mathcal{G}_{\mathcal{H}}$ (with property Π) and use an algorithm operating on this to decide the question posed of $\mathcal{H}$. In a more precise sense, we have the formalism presented below.

Definition 9. A *property*, Π of finite argument systems is a (typically infinite) subset of all possible finite argument systems. We say $\mathcal{H}$ *has property* Π if $\mathcal{H} \in \Pi$.

The argument system $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is *simulated by* the argument system $\mathcal{G}(\mathcal{X} \cup \mathcal{Y}, \mathcal{B})$ with respect to *credulous admissibility* (denoted $\mathcal{G} \sim_{\text{ca}} \mathcal{H}$) if

$$\forall S \subseteq \mathcal{X}\, \text{CA}_{\{\}}\big(\mathcal{G}(\mathcal{X} \cup \mathcal{Y}, \mathcal{B}), S\big) \Leftrightarrow \text{CA}_{\{\}}\big(\mathcal{H}(\mathcal{X}, \mathcal{A}), S\big)$$

Similarly $\mathcal{H}$ is simulated by $\mathcal{G}$ w.r.t. *sceptical admissibility* ($\mathcal{G} \sim_{\text{sa}} \mathcal{H}$) if

$$\forall S \subseteq \mathcal{X}\, \text{SA}_{\{\}}\big(\mathcal{G}(\mathcal{X} \cup \mathcal{Y}, \mathcal{B}), S\big) \Leftrightarrow \text{SA}_{\{\}}\big(\mathcal{H}(\mathcal{X}, \mathcal{A}), S\big)$$

For $\alpha \in \{\text{CA}, \text{SA}\}$, a property, Π *α-represents* a property Θ if for every $\mathcal{H}(\mathcal{X}, \mathcal{A}) \in \Theta$ there is some $\mathcal{G}(\mathcal{X} \cup \mathcal{Y}, \mathcal{B}) \in \Pi$ such that $\mathcal{G} \sim_{\alpha} \mathcal{H}$. We say that Π *polynomially α-represents* Θ if there is some constant k such that, for every $\mathcal{H}(\mathcal{X}, \mathcal{A}) \in \Theta$ there is some $\mathcal{G}(\mathcal{X} \cup \mathcal{Y}, \mathcal{B}) \in \Pi$ such that $|\mathcal{X} \cup \mathcal{Y}| \leqslant |\mathcal{X}|^k$ and $\mathcal{G} \sim_{\alpha} \mathcal{H}$. Finally we say that a property is (*polynomially*) *α-universal* if it (polynomially) α-represents all argument systems.

It will be useful also to view as "polynomially α-universal" those properties that α-represent all but finitely many argument systems.

The class of argument systems considered in this section are those defined by the property, $\Delta^{(p,q)}$ introduced below,

Definition 10. An argument system $\mathcal{H}(\mathcal{X}, \mathcal{A})$ has (p, q)*-bounded degree* if

$$\forall x \in \mathcal{X} \quad \big|\{y \in \mathcal{X}\colon \langle y, x\rangle \in \mathcal{A}\}\big| \leqslant p \quad \text{and} \quad \big|\{y \in \mathcal{X}\colon \langle x, y\rangle \in \mathcal{A}\}\big| \leqslant q$$

The notation $\Delta^{(p,q)}$ will be used for the set of all (p, q)-bounded degree systems.

Our main result in this section is

Theorem 11.

(a) $\Delta^{(2,2)}$ *is polynomially* CA*-universal.*
(b) $\Delta^{(2,2)}$ *is polynomially* SA*-universal.*

Proof. We prove part (a) only. An identical construction serves for part (b) with the analysis needed for the conditions of simulation w.r.t. sceptical admissibility proceeding in a similar style to the case of credulous admissibility.

Let $\mathcal{H}(\mathcal{X}, \mathcal{A})$ be any finite argument system. Suppose $\mathcal{H} \notin \Delta^{(2,2)}$. Consider any $x \in \mathcal{X}$ for which

$$\{y\colon \langle y, x\rangle \in \mathcal{A}\} = \{y_1, y_2, \ldots, y_k\} \quad \text{and} \quad k \geqslant 3$$

Consider the system $\mathcal{G}_x^{(k-1)}(\mathcal{X} \cup \{z_1, z_2\}, \mathcal{B})$ formed by introducing new arguments z_1 and z_2 with

$$\mathcal{B} = \mathcal{A} \setminus \{\langle y_i, x\rangle\colon 2 \leqslant i \leqslant k\} \cup \{\langle z_1, x\rangle, \langle z_2, z_1\rangle\} \cup \{\langle y_i, z_2\rangle\colon 2 \leqslant i \leqslant k\}$$

i.e. formed by replacing the attacks on x in Fig. 3 with the system in Fig. 4.

We claim that $\mathcal{G}_x^{(k-1)}(\mathcal{X} \cup \{z_1, z_2\}, \mathcal{B}) \sim_{\text{ca}} \mathcal{H}(\mathcal{X}, \mathcal{A})$.

Consider any $T \subseteq \mathcal{X} \cup \{z_1, z_2\}$ defining an admissible set in $\mathcal{G}_x^{(k-1)}$ and let $S = T \setminus \{z_1, z_2\}$. To see that S is conflict-free it suffices to observe that the only way in which T can be conflict-free and S fail to be so is if $\{x, y_i\} \subseteq T$ for some $2 \leqslant i \leqslant k$: but in this case, since z_1 attacks x and the only counterattack on z_1 is z_2, $x \in T$ forces $z_2 \in T$ from which $y_i \notin T$, for every $1 \leqslant i \leqslant k$. To see that S also defends itself against any attack if T does so, first suppose that $x \in T$. In this case, not only must some attacker of y_1 be in T (and thus the same attacker is in S) but also since $z_2 \in T$ to defend the attack on x by z_1, we require that for each attack $\langle y_i, z_2\rangle$, T must contain some attacker of y_i: again all

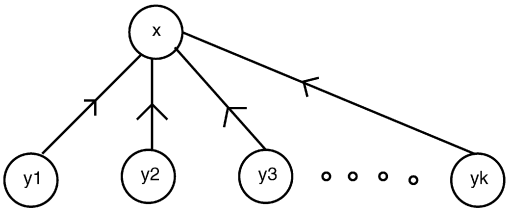

Fig. 3. Argument x attacked by $k \geqslant 3$ arguments.

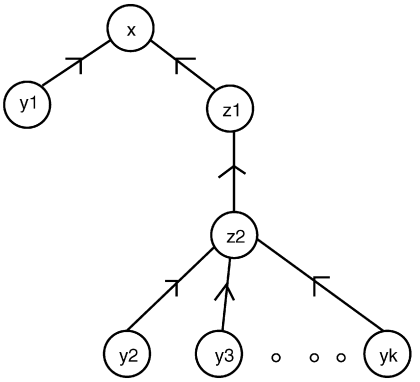

Fig. 4. Reducing k attacks to $k-1$ attacks.

of these attacks will be members of S. If, on the other hand, $x \notin T$, without loss of generality suppose that $\langle x, p\rangle \in \mathcal{A}$ and that $p \in T$. Then either $y_1 \in T$ (and thus also in S) or $z_1 \in T$. The second of these, however, requires that at least one of $\{y_2, \ldots, y_k\}$ is in T to counterattack z_2. It follows that if $x \notin S$ and attacks $p \in S$ then $\{y_1, \ldots, y_k\} \cap S \neq \emptyset$.

In the reverse direction, suppose that $S \subseteq \mathcal{X}$ is admissible in $\mathcal{H}$. If $x \in S$ then $S \cup \{z_2\}$ is an admissible set of $\mathcal{G}_x^{(k-1)}$. If $x \notin S$ either S is also an admissible set of $\mathcal{G}_x^{(k-1)}$ (if $y_1 \in S$ or x does not attack any argument of S) or $S \cup \{z_1\}$ is such a set (whenever $S \cap \{y_2, \ldots, y_k\} \neq \emptyset$). Thus, $\mathcal{G}_x^{(k-1)}(\mathcal{X} \cup \{z_1, z_2\}, \mathcal{B}) \sim_{\text{ca}} \mathcal{H}(\mathcal{X}, \mathcal{A})$.

Noting that the construction does change the number of attacks on arguments other than x, a similar procedure can be applied to any remaining argument attacked by at least three arguments. A near identical construction (in which the direction of attacks is reversed) serves when dealing with those arguments that attack more than two others. □

Recalling from Definition 5 that $\Gamma^{(k)}$ is the set of all k-partite argument systems we obtain

Corollary 2. *The property $\Gamma^{(4)} \cap \Delta^{(2,2)}$ is polynomially* CA*-universal and polynomially* SA*-universal.*

Proof. Viewed as undirected graphs, via Brooks' Theorem ([11, Thm 6, Ch. 15, p. 337]), with a single exception,[8] every argument system in $\Delta^{(2,2)}$ is vertex 4-colourable. It follows that these are 4-partite. □

Corollary 3. *Let* $\text{Q}^{(2,2)}$ *denote either of the decision problems* {CA, SA} *restricted to argument systems with the property $\Delta^{(2,2)}$.*

(a) $\text{CA}^{(2,2)}$ *is* NP*-complete.*
(b) $\text{SA}^{(2,2)}$ *is Π_2^p-complete.*

Proof. Apply the construction of Theorem 11 to the systems $\mathcal{H}_\Phi$ and $\mathcal{G}_\Phi$ presented in Section 3. □

[8] It is because of this case—the complete graph on 5 vertices—that the "connectivity" assumption mentioned following Definition 1 is needed: without it there are infinitely many $(2, 2)$-bounded systems which are not vertex 4-colourable.

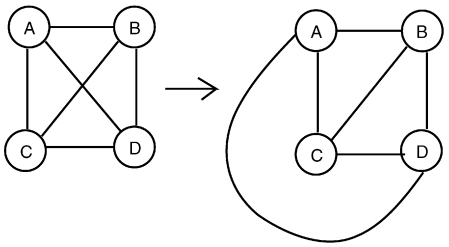

Fig. 5. Planar drawing of K_4 the complete graph on four vertices.

6. Planar argument systems

We recall that a graph $G(V, E)$ is *planar* if it can be drawn (in the plane) in such a way that no two edges of the graph cross each other. Thus, the complete graph on four vertices is planar, e.g. Fig. 5, whilst the complete graph on five vertices is non-planar.

Several graph-theoretic decision problems whose general versions are NP-hard are known to admit polynomial time algorithms when instances must be planar graphs. Examples include not only questions that are immediately resolvable from established properties of planar graphs, e.g. vertex 4-colouring and maximal clique, but also for questions where it is far from obvious that planarity assists in developing efficient algorithms, e.g. the problem of determining whether a graph has a bipartite subgraph containing at least some specified number of edges, [31, GT25, p. 196]. For problems whose complexity status is still open, most notably that of deciding if two given graphs are isomorphic, linear time methods have been found for planar graphs, e.g. [33]. Planarity, however, does not help in the construction of efficient decision procedures for the problems of Table 1. The reductions employed to prove this make use of a device which is of some independent interest: in terms of the formalism introduced in the preceding section this allows us to argue that planarity is a polynomially CA-universal property.

Prior to presenting our main analyses we make two observations: firstly, from work of Fraenkel [30], it is known that STAB-EXT restricted to planar instances is NP-complete. This result, however, does not allow us to deduce anything concerning the complexity of credulous or sceptical acceptance: in particular, the planar systems constructed when addressing CA are guaranteed to have stable extensions so the existence problem STAB-EXT is trivial for such cases. As a second point we observe that using the (NP-complete) decision problem PLANAR-3-SAT, whose instances are 3-CNF formulae having planar *clause incidence graphs*,[9] it is not too difficult to show that $\text{CA}_{\{\}}(\mathcal{H}, S)$ is NP-complete when $\mathcal{H}$ is required to be a planar graph.[10] We do not consider the proof of this result in any further detail, simply noting that it is subsumed by our proof that $\text{CA}(\mathcal{H}, x)$ is NP-complete with $\mathcal{H}$ restricted to planar graphs.

For Q any of the decision problems of Table 1, we let Q^{P} denote the variant in which the argument system forming part of the instance is planar.

Theorem 12. CA^{P} *is* NP*-complete.*

Proof. It suffices to prove that CA^{P} is NP-hard, for which purpose we use a reduction from 3-SAT. Given $\Phi(Z_n)$ we first form the system $\mathcal{H}_\Phi(\mathcal{X}, \mathcal{A})$ of Section 3.1 and recall that $\Phi(Z_n)$ is satisfiable if and only if $\text{CA}(\mathcal{H}_\Phi, \Phi)$ holds.

The argument system $\mathcal{H}_\Phi$, however, will not in general be planar, e.g. in Fig. 1 there are eleven distinct points where edges cross and thus $\mathcal{H}_\Phi$ must be modified to obtain a planar graph, $\mathcal{H}^{\text{P}}_\Phi$, whilst retaining the property that the argument Φ is credulously accepted if and only if $\Phi(Z_n)$ is satisfiable.

[9] The clause incidence graph of a CNF $\Phi(x_1, \ldots, x_n) = \bigwedge_{j=1}^{m} C_j$, is the bipartite graph with vertex sets $\{x_1, \ldots, x_n\}$ and $\{C_1, \ldots, C_m\}$ and edges $\{x_i, C_j\}$ for each case when $\neg x_i$ occurs in C_j or x_i occurs in C_j.

[10] For readers familiar with the relevant graph-theoretic concepts, the instance of $\text{CA}_{\{\}}$ is formed using a planar embedding of the clause incidence graph of Φ—an instance of PLANAR-3-SAT—augmenting it with arguments $\{\phi_1, \phi_2, \ldots, \phi_r\}$ one for each *face* of the embedding in which a clause of Φ occurs. These arguments are then attacked by the individual clauses within the relevant face. Following some minor adjustments to represent the presence of *negated* literals in clauses, we can then show that the *set* $\{\phi_1, \ldots, \phi_r\}$ is credulously accepted if and only if Φ is satisfiable.

The system $\mathcal{H}^{\mathrm{P}}_{\Phi}$ is formed from $\mathcal{H}_{\Phi}$ in two stages. First for each position where two edges[11] cross, e.g. $\langle p, q\rangle$ and $\langle r, s\rangle$, replace the "crossing point" by an argument which attacks q and s and is attacked by p and r. If the chosen realisation of $\mathcal{H}_{\Phi}$ contains r crossings we denote these new arguments $X_c = \{x_1, x_2, \ldots, x_r\}$. We note that $r = \mathrm{O}(|\mathcal{A}|^2)$ so the translation is polynomial time computable. Fig. 6 illustrates the outcome of this translation when applied to the argument system of Fig. 1 after replacing the eleven crossings.

Of course this new system will no longer have the same admissibility properties of the one it replaces: in particular it is not guaranteed to be the case that an admissible set containing Φ can be built if and only if $\Phi(Z_n)$ is satisfiable. For example, for the system shown in Fig. 6, the set $\{\Phi, z_1, z_2, z_3, z_4, x_3, x_8, x_9\}$ is admissible, however, the corresponding instantiation of $\langle z_1, z_2, z_3, z_4\rangle$ by $z_i := \top$ gives $\Phi(\top, \top, \top, \top) \equiv \bot$. In order to restore the desired behaviour we systematically replace each new argument introduced with a planar argument system.

The typical environment in this case is shown in Fig. 7(a). We have arguments (z and y) that (in $\mathcal{H}_{\Phi}$) attacked arguments q and p: the attacks $\langle z, q\rangle$ and $\langle y, p\rangle$ crossing in the drawing of $\mathcal{H}_{\Phi}$ and the crossing point replaced by an argument (x) so that the attacks present are now $\langle z, x\rangle$, $\langle y, x\rangle$, $\langle x, p\rangle$, and $\langle x, q\rangle$. In Fig. 7(b), x in turn is replaced by a *planar system* linking arguments z and y with new arguments y_b and z_d with y_b attacking p and z_d attacking q. In order to ensure this replacing system operates correctly it must have the property that in any preferred extension, S, of the resulting system it holds: $z \in S$ if and only if $z_d \in S$ and $y \in S$ if and only if $y_b \in S$.

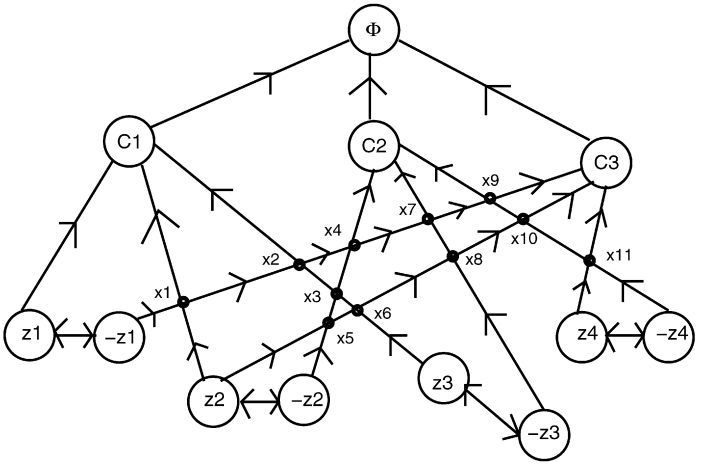

Fig. 6. $\mathcal{H}_{\Phi}$ after crossings replaced by new arguments x_i.

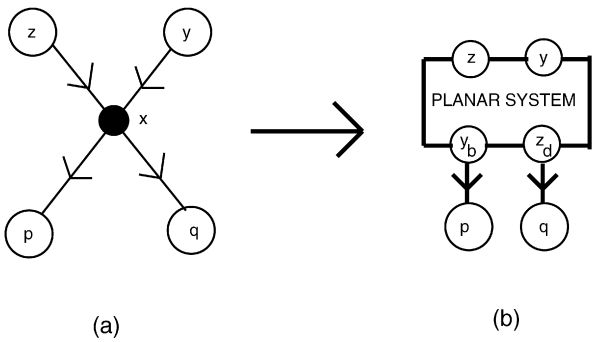

Fig. 7. Crossing edges in $\mathcal{H}_{\Phi}$ and their replacement.

[11] It is not necessary to consider the case of three of more edges having a common crossing point: any graph may be drawn in such a way that this case does not arise.

$\lambda(x) := 0 \forall x \in X_c$
$T := X_c$; $k := 1$
while $k \leqslant r$ **do**
  if $\exists x \in T$: x is attacked by two literals **then**
    $\lambda(x) := k$; $T := T \setminus \{x\}$
  else if $\exists x \in T$: x is attacked by a literal and $x' \in X_c \setminus T$ **then**
    $\lambda(x) := k$; $T := T \setminus \{x\}$
  else
    Choose any $x \in T$ with both attackers of x in $X_c \setminus T$
    $\lambda(x) := k$; $T := T \setminus \{x\}$
  end if
  $k := k + 1$
end while

Algorithm 2. Ordering of arguments in X_c.

Before describing the exact design of the replacing system, however, we specify the order in which the X_c are replaced. We say that the argument y of $\mathcal{H}_\Phi$ is a *literal* if $y \in \{z_i, \neg z_i 1 \leqslant i \leqslant n\}$ and now observe that the set of arguments, X_c may be ordered using the labelling approach presented in Algorithm 2 to assign a unique number $\lambda(x)$ to each $x \in X_c$ with $1 \leqslant \lambda(x) \leqslant r$. For the example of Fig. 6 an ordering produced by this algorithm is $\langle x_1, x_5, x_{11}, x_6, x_8, x_3, x_2, x_4, x_7, x_{10}, x_9\rangle$.

The construction of the planar system $\mathcal{H}^{\mathrm{P}}_\Phi$ is completed by replacing the arguments $x \in X_c$ in order of increasing value of their label $\lambda(x)$ with a copy of the planar crossover gadget given in Fig. 8.[12] We denote by $\mathcal{W}$ the arguments of $\mathcal{H}^{\mathrm{P}}_\Phi$ (noting that $\mathcal{X} \subseteq \mathcal{W}$ and $X_c \cap \mathcal{W} = \emptyset$); and its attacks by $\mathcal{B}$ (observing that each of the attacks $\langle C_j, \Phi\rangle \in \mathcal{A}$ is also in $\mathcal{B}$).

The resulting system, $\mathcal{H}^{\mathrm{P}}_\Phi$, is planar: it remains to show that Φ is credulously accepted in $\mathcal{H}^{\mathrm{P}}_\Phi$ if and only if $\Phi(Z_n)$ is satisfiable. In $\mathcal{H}_\Phi$, S is a preferred extension containing Φ if and only if $S = \{\Phi, y_1, y_2, \ldots, y_n\}$ with $y_i \in \{z_i, \neg z_i\}$ defining an instantiation satisfying $\Phi(Z_n)$, i.e. $z_i = \top$ if $y_i = z_i$ and $z_i = \bot$ if $y_i = \neg z_i$, it therefore is sufficient to prove for the crossover gadget of Fig. 8 that whenever S is a preferred extension of $\mathcal{H}^{\mathrm{P}}_\Phi$, $z \in S \Leftrightarrow z_d \in S$ and $y \in S \Leftrightarrow y_b \in S$. We need only consider the first of these as an identical proof covers the second. To simplify the analysis, it is useful to note that $\mathcal{H}_\Phi$ and $\mathcal{H}^{\mathrm{P}}_\Phi$ are both *coherent*: the only cycles are those of length two formed by the n pairs $\{z_i, \neg z_i\}$, i.e. $\mathcal{H}_\Phi$ and $\mathcal{H}^{\mathrm{P}}_\Phi$ contain no odd length cycles and coherence follows from Fact 2(d). Given this, every preferred extension, S, of $\mathcal{H}^{\mathrm{P}}_\Phi$ is also a *stable* extension so that any $q \notin S$ must be attacked by some $p \in S$.

Consider any preferred extension $S \subseteq \mathcal{W}$ of $\mathcal{H}^{\mathrm{P}}_\Phi$ and an occurrence of the crossover gadget which, without loss of generality, we take as labelled in Fig. 8. Suppose $z \in S$ and consider the two possibilities $y \in S$ and $y \notin S$. The first of these, gives $a_4 \in S$: each of $\{a_1, a_2, a_3\}$ is attacked by $\{y, z\}$, however a_4 is only attacked by $\{a_2, a_3\}$ and so (from stability) must be in S. From $\{y, z, a_4\} \subseteq S$ it follows that $\{b_1, b_2, b_3, d_1, d_2, d_3\} \cap S = \emptyset$ and thence, again via stability, $\{y_b, z_d\} \subset S$ since no attackers of these can belong to S. For the second possibility, $y \notin S$, some attacker of y (y' say) must belong to S and we deduce that $\{z, y', a_3\} \subseteq S$ and $a_4 \notin S$. In this case, however, it must hold that $d_1 \in S$ (this is only attacked by a_4 and y) and thence $z_d \in S$ (since neither of its attackers—d_2 and d_3 can belong to S). In summary if $z \in S$ then $z_d \in S$ regardless of the status of y.

On the other hand suppose that $z \notin S$ so that some attacker of z, z' is in S. Again we have the two possibilities $y \in S$ and $y \notin S$. In the former case, $\{z', y, a_2\} \subseteq S$ and $\{a_4, d_1, d_3\} \cap S = \emptyset$. From this we must have $d_2 \in S$ (since its only attackers are d_1 and a_4) from which it follows that $z_d \notin S$ as required. Finally in the second case with $y \notin S$, some attacker y' of y is in S. From $\{y, z\} \cap S = \emptyset$ we deduce that $\{y', z', a_1, a_4\} \subseteq S$ (y and z are the only attackers of a_1), and thence $\{d_1, d_2\} \cap S = \emptyset$. In this case, however, it must hold that $d_3 \in S$ as its only attackers are y and d_1: in consequence $z_d \notin S$ as required. In total we have that $z \notin S$ implies $z_d \notin S$, completing the proof that the crossover gadget has the desired behaviour.

It is now easy to see that Φ is credulously accepted in the planar system $\mathcal{H}^{\mathrm{P}}_\Phi$ if and only if $\Phi(Z_n)$ is satisfiable. If $\{y_1, \ldots, y_n\}$ is a set of literals defining a satisfying instantiation of $\Phi(Z_n)$ then each clause C_j must contain a literal

[12] Readers familiar with research literature on planar realisations of Boolean networks may recognise that the structure of Fig. 8 derives from that of the planar crossover formed from twelve binary $\neg\wedge$-elements, cf. [35] and [25, Ch. 6, pp. 404–405].

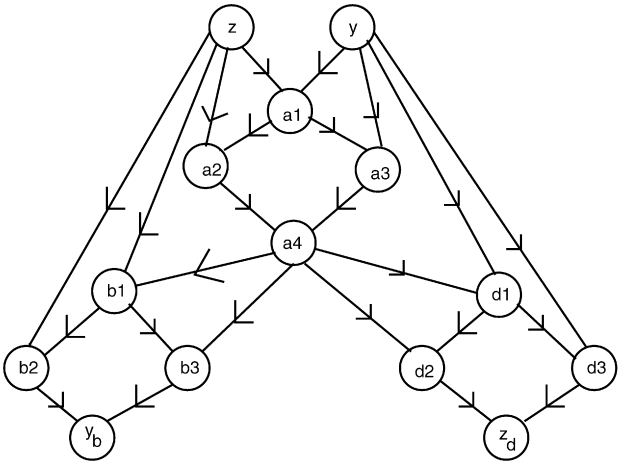

Fig. 8. Planar crossover gadget.

from this set. Choosing the argument z_i in $\mathcal{W}$ if $y_i = z_i$ and the argument $\neg z_i$ otherwise, we can build an admissible subset S of $\mathcal{W}$ which attacks each argument C_j (either the literal itself or that propagated via the crossover gadget that replaced $\langle x, C_j\rangle$), so that Φ can be added to S in forming a preferred extension. On the other hand if Φ is credulously accepted then from a preferred extension containing Φ and the attacks on each C_j in S we identify a set of literals that will satisfy $\Phi(Z_n)$.

We deduce that CA^{P} is NP-complete as claimed. □

In the analysis demonstrating that the crossover gadget of Fig. 8 operated correctly, we relied on the fact that the system in which it was used was coherent and that thus for any given preferred extension, S, arguments $q \notin S$ could be assumed to be attacked by some argument $p \in S$. We cannot, however, rely on this assumption in attempting to translate arbitrary non-planar argument systems to planar schemes, and thus it is unclear whether directly replacing crossing points using the crossover gadget would produce a system with similar admissibility properties. It turns out, however, that it *is* possible to transform *any* argument system, $\mathcal{H}$, into a planar system, $\mathcal{H}^{\text{P}}$ in such a way that questions regarding credulous admissibility of arguments in $\mathcal{H}$ may be posed of corresponding arguments in $\mathcal{H}^{\text{P}}$. In order to do this a rather more indirect construction is needed.

Theorem 13. *Planarity is polynomially* CA*-universal.*

Proof. Let $\mathcal{H}(\mathcal{X}, \mathcal{A})$ be any finite argument system with $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. Consider the propositional formula $\Psi^{\mathcal{H}}(X_n)$ defined from $\mathcal{H}(\mathcal{X}, \mathcal{A})$ as

$$\Psi^{\mathcal{H}}(X_n) = \bigwedge_{\langle x_i, x_j\rangle \in \mathcal{A}} (\neg x_i \vee \neg x_j) \wedge \left(\neg x_j \vee \bigvee_{x_k:\ \langle x_k, x_i\rangle \in \mathcal{A}} x_k\right)$$

With this system $\alpha = \langle a_1, \ldots, a_n\rangle$ is a satisfying instantiation of $\Psi^{\mathcal{H}}$ if and only if the subset $S(\alpha)$ of $\mathcal{X}$ chosen via $x_i \in S \Leftrightarrow a_i = \top$ is an admissible set in $\mathcal{H}$.

The formula $\Psi^{\mathcal{H}}$ is in CNF and so we can define another argument system—$\mathcal{F}_{\Psi}^{\mathcal{H}}$ simply by using the construction of Section 3.1. Furthermore we can now apply the planarization method of Theorem 12 ($\mathcal{F}_{\Psi}^{\mathcal{H}}$ is coherent irrespective of whether $\mathcal{H}$ is so). Let $\mathcal{F}_{\Psi}^{\mathcal{H},\text{P}}$ be the resulting planar argument system. Now although it is not the case that $\mathcal{F}_{\Psi}^{\mathcal{H},\text{P}} \sim_{\text{ca}} \mathcal{H}$-every subset of $\{x_1, x_2, \ldots, x_n\}$ describes an admissible set in $\mathcal{F}_{\Psi}^{\mathcal{H},\text{P}}$—it is easily modified to a system $\mathcal{H}_{\Psi}^{\mathcal{H},\text{P}}$ which is planar and satisfies $\mathcal{H}_{\Psi}^{\mathcal{H},\text{P}} \sim_{\text{ca}} \mathcal{H}$. To achieve this, we add a new argument, u, to the set of arguments forming $\mathcal{F}_{\Psi}^{\mathcal{H},\text{P}}$ together with attacks $\{\langle \Psi, u\rangle\} \cup \{\langle u, x_i\rangle, \langle u, \neg x_i\rangle\colon\ 1 \leqslant i \leqslant n\}$.

Notice that a planar realisation of $\mathcal{H}_{\Psi}^{\mathcal{H},\text{P}}$ is straightforward to construct from the planar realisation of $\mathcal{F}_{\Psi}^{\mathcal{H},\text{P}}$. Now let $\mathcal{Y}$ consist of the arguments $\{\neg x_i\colon\ 1 \leqslant i \leqslant n\}$ together with $\{u, \Psi\}$, the arguments representing clauses of $\Psi^{\mathcal{H}}$ and

those introduced during the transformation of $\mathcal{F}_{\Psi}^{\mathcal{H}}$ to the planar system $\mathcal{F}_{\Psi}^{\mathcal{H},\mathrm{P}}$, i.e. arising by replacing crossing points with copies of the schema in Fig. 8.

We claim that $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}} \sim_{\mathrm{ca}} \mathcal{H}$. Consider any admissible subset T of $\mathcal{X} \cup \mathcal{Y}$, the arguments of $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}}$. To see that $S = T \setminus \mathcal{Y}$ is an admissible set in $\mathcal{H}$, notice that

$$(\Psi \notin T) \text{ and } (T \text{ is admissible}) \quad \Leftrightarrow \quad (T = \emptyset)$$

since the argument u attacks each y in $\{x_i, \neg x_i \colon 1 \leqslant i \leqslant n\}$ and is only attacked by Ψ, so it is no longer the case that every non-empty subset of $\{x_1, x_2, \ldots, x_n\}$ describes an admissible set in $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}}$ (as happened with $\mathcal{F}_{\Psi}^{\mathcal{H},\mathrm{P}}$). So without loss of generality we may assume $\Psi \in T$. Now the definition of $\Psi^{\mathcal{H}}(X_n)$ and the properties of $\mathcal{F}_{\Psi}^{\mathcal{H}}$ ensure that since the instantiation $x_i = \top$ if $x_i \in T$, $x_i = \bot$ if $x_i \notin T$ satisfies $\Psi^{\mathcal{H}}$ the set $\{x_i \colon x_i \in T\}$ is an admissible subset in $\mathcal{H}$: this, set however, is exactly the set of arguments in $S = T \setminus \mathcal{Y}$.

Similarly, if $S \subseteq \mathcal{X}$ is admissible in $\mathcal{H}$, it may be extended to an admissible set in $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}}$ by adding the arguments $\{\Psi\}$, $\{\neg x_i \colon x_i \notin S\}$ and those from the crossover elements whose inclusion is forced by the subset of $\{x_i, \neg x_i \colon 1 \leqslant i \leqslant n\}$ corresponding to the satisfying instantiation of $\Psi^{\mathcal{H}}(X_n)$. □

Corollary 4. PREF-EXT$^{\mathrm{P}}$ *is* NP*-complete.*

Proof. Immediate by noting that $\mathcal{H}_{\Phi}^{\mathrm{P}}$ modified by the addition of the argument u as described in the proof of Theorem 13, has a non-empty preferred extension if and only CA$(H_{\Phi}^{\mathrm{P}}, \Phi)$. □

Corollary 5. *Let $\mathcal{P}^{(p,q),k}$ be the class of planar argument systems in the set $\Delta^{(p,q)} \cap \Gamma^{(k)}$. The property $\mathcal{P}^{(2,2),4}$ is polynomially* CA*-universal.*

Proof. From Theorem 13 planarity is polynomially CA-universal. The transformation described in Theorem 11 preserves planarity, thus the result follows by combining Theorem 13, Theorem 11 and Corollary 3. □

In fact, analysing the structure of $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}}$ from the proof of Theorem 13 we obtain a stronger result,

Corollary 6. *The property $\mathcal{P}^{(3)}$ satisfied by* 3*-partite planar argument systems is polynomially* CA*-universal.*

Proof. Given $\mathcal{H}(\mathcal{X}, \mathcal{A})$ form the planar system $\mathcal{H}_{\Psi}^{\mathcal{H},\mathrm{P}}$ of Theorem 13. It is straightforward to show that this system is vertex 3-colourable and hence 3-partite. □

Finally, parallelling the result of Theorem 12 we have,

Theorem 14.

(a) SA$^{\mathrm{P}}$ *is Π_2^p-complete.*
(b) COHERENT$^{\mathrm{P}}$ *is Π_2^p-complete.*

Proof. Exactly as the reduction from QSAT$_2^{\Pi}$ outlined in Section 3.2, however, with the CNF instance $\Phi(Y_n, Z_n)$ implemented as the argument system $\mathcal{H}_{\Phi}^{\mathrm{P}}$ instead of $\mathcal{H}_{\Phi}$. □

7. Bounded treewidth

Treewidth, which may be informally understood as a measure of the extent to which a graph differs from a tree, is known to provide a significant aid in developing efficient algorithmic approaches, particularly in the case of graphs whose treewidth may be bounded by a constant value k. A useful survey of results concerning graphs with bounded treewidth is presented in [14]. With some minor differences, we follow the treatment given in Arnborg et al. [3] for the definition of treewidth in Definition 15 and for the description of the language of monadic second order logic. The

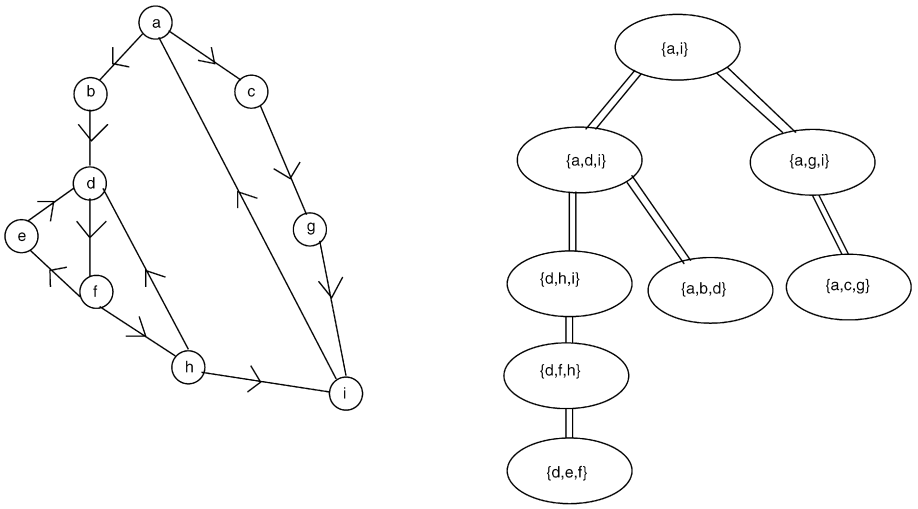

Fig. 9. $\mathcal{H}$ with $tw(\mathcal{H}) = 2$ and its tree decomposition.

second of these admits the use of powerful general tools for synthesising efficient decision algorithms for an extensive range of NP-hard graph problems when the graphs in question have bounded treewidth.

Definition 15. A *tree decomposition* of a directed graph $H(X, A)$ is a pair $\langle T, S\rangle$, where $T(V, F)$ is a tree and $S = \{S_1, S_2, \ldots, S_r\}$ is a family of subsets of X with $r = |V|$ for which

(a) $\bigcup_{i=1}^{r} S_i = X$.
(b) For all $\langle x, y\rangle \in A$ there is at least one[13] $S_i \in S$ for which $\{x, y\} \subseteq S_i$.
(c) For each $x \in X$, the subgraph of $T(V, F)$ induced by the set $V_x = \{V_i\colon\ x \in S_i\}$ is connected, i.e. a subtree of $T(V, F)$.

The *width* of a tree decomposition $\langle T, S\rangle$ of $H(X, A)$ is $\max_{S_i \in S} |S_i| - 1$; the *treewidth* of $H(X, A)$—denote $tw(H)$—is the minimum width over all tree decompositions of H.

Notice that if $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is itself a tree, then $tw(\mathcal{H}) = 1$: simply choose the system of subsets $S = \{S_1, \ldots, S_r\}$ with $r = |\mathcal{A}|$ so that for each $\langle x_i, x_j\rangle \in \mathcal{A}$ there is a set $S_k = \{x_i, x_j\} \in S$. The edges, F, in the tree decomposition $\langle T(V, F), S\rangle$, are those pairs $\{V_i, V_j\}$ for which $S_i \cap S_j \neq \emptyset$. An example of $\mathcal{H}(\mathcal{X}, \mathcal{A})$ for which $tw(\mathcal{H}) = 2$ and an associated tree decomposition is given in Fig. 9.

We denote by $W^{(k)}$ the class of all argument systems $\mathcal{H}(\mathcal{X}, \mathcal{A})$ whose treewidth is at most k. We note that although given $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), k\rangle$ deciding if $\mathcal{H} \in W^{(k)}$ is NP-complete from [4], for constant k there are polynomial-time algorithms: $O(n^{k+2})$ methods were first presented in [2], while linear time methods are given in [13]. Both return a width k decomposition, if one exists, although it should be noted that the $O(n)$ method of [13] involves a significant constant factor (dependent on k) in the $O(n)$ analysis.

Consider structures of the form $\langle \mathcal{X}, \mathcal{A}\rangle$ where $\mathcal{X} = \{x_1, \ldots, x_n\}$ is a finite set of arguments and $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ an attack relation. The language, L, of *monadic second-order logic* (*MSOL*) for this class of structures contains the standard propositional connectives—$\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\neg$—individual variable symbols—x, y, z etc.—predicates, and quantifiers ($\exists$, $\forall$). In addition, and distinguishing it from first-order logic, L contains *set* variable symbols, U, V, W, etc., the set membership symbol ($\in$) and allows quantification over set variables.

We note that the scheme presented in [3] is rather more elaborated. The corresponding structure would be $\langle D, \mathcal{X}, \mathcal{A}, \mathbf{hd}, \mathbf{tl}\rangle$ where $\mathcal{X}$ and $\mathcal{A}$ are unary predicates on elements of the set D, i.e. $\mathcal{X}(d)$ holds if and only if d is an

[13] [3, Definition 3.1, p. 314] requires *exactly* one, however, the distinction is not significant.

argument; $\mathcal{A}(d)$ if and only if d is an attack. To relate attacks to their constituent arguments, **hd** and **tl** are binary predicates defined so that $\mathbf{hd}(b, c)$ if b is an attack whose source is the argument c; similarly $\mathbf{tl}(b, c)$ holds whenever b is an attack directed at the argument c: thus $\langle x, y\rangle \in \mathcal{A}$ would be realised as $\mathcal{X}(x) \wedge \mathcal{X}(y) \wedge \exists d\mathcal{A}(d) \wedge \mathbf{hd}(d, x) \wedge \mathbf{tl}(d, y)$. For reasons of clarity we eschew this level of precision. We note that, where we write, e.g. $\exists U \subseteq \mathcal{X}P(U)$ (for some predicate P), within the formal style of [3], this could be expressed by $\exists U(\forall uu \in U \rightarrow \mathcal{X}(u)) \wedge P(U)$; similarly $\forall U \subseteq \mathcal{X}P(U)$ is equivalent to $\forall U(\forall uu \in U \rightarrow \mathcal{X}(u)) \rightarrow P(U)$.

Now given a well-formed MSOL sentence $\Phi(\mathcal{X}, \mathcal{A})$ typically some argument systems, $\mathcal{H}$, will satisfy[14] Φ and others fail to do so, i.e. such sentences provide a mechanism for specifying *properties* of finite argument systems. Formally we say an argument system property, Π, is *MSOL-definable* if there is a well-formed MSOL sentence, $\Phi(\mathcal{X}, \mathcal{A})$ such that

$$\forall \mathcal{H}(\mathcal{X}, \mathcal{A})\mathcal{H} \in \Pi \Leftrightarrow \Phi \models \mathcal{H}$$

For example, the property of an argument system being bipartite, $\mathcal{H}(\mathcal{X}, \mathcal{A}) \in \Gamma^{(2)}$, is MSOL-definable as shown by the sentence,

$$BI(\mathcal{X}, \mathcal{A}) = \exists U\ \exists V\ \forall x(x \in U \vee x \in V) \wedge \big(\neg(x \in U) \vee \neg(x \in V)\big) \wedge \big(\forall y\big(\langle x, y\rangle \in \mathcal{A}\big) \rightarrow (x \in U \leftrightarrow y \in V)\big)$$

That is the system $\langle \mathcal{X}, \mathcal{A}\rangle$ is bipartite whenever there are two sets (U and V) such that: every x belongs to at least one of these ($x \in U$ or $x \in V$); no x belongs to both; and should $\langle x, y\rangle$ be an attack in $\mathcal{A}$, exactly one of x and y is in U. Thus, the system with $\mathcal{X} = \{x_1, x_2, x_3\}$ and $\mathcal{A} = \{\langle x_1, x_2\rangle, \langle x_2, x_3\rangle, \langle x_3, x_1\rangle\}$ fails to satisfy $BI(\mathcal{X}, \mathcal{A})$ whereas with $\mathcal{A}' = \{\langle x_1, x_2\rangle, \langle x_2, x_3\rangle, \langle x_2, x_1\rangle\}$ $BI(\mathcal{X}, \mathcal{A}')$ is satisfied (choose $U = \{x_1, x_3\}$ and $V = \{x_2\}$).

Although not all graph-theoretic properties are MSOL-definable, for those which are—irrespective of the computational complexity for instances in general—the following result of Courcelle [18,19] is of significance respecting decision methods for MSOL-definable properties restricted to graphs with treewidth k.

Fact 16. *(Courcelle's Theorem, see [18,19] also [3].) Let $\mathcal{K}$ be a class of graphs for which $\forall G \in \mathcal{K}tw(G) \leqslant k$ for some constant $k \in \mathbf{N}$ and Π be any MSOL-definable property. Given $G \in \mathcal{K}$ and a width k tree decomposition of G, $G \in \Pi$ is decidable in linear time.*

Recall that $W^{(k)}$ is the class of finite argument systems $\mathcal{H}(\mathcal{X}, \mathcal{A})$ for which a tree decomposition of width k exists.

Theorem 17. *For all constant $k \in \mathbf{N}$, given $\mathcal{H}(\mathcal{X}, \mathcal{A}) \in W^{(k)}$ together with a width k tree decomposition of $\mathcal{H}(\mathcal{X}, \mathcal{A})$ each of the following decision problems are decidable in linear time.*

(a) PREF-EXT($\mathcal{H}$).
(b) STAB-EXT($\mathcal{H}$).
(c) COHERENT($\mathcal{H}$).
(d) *There is at least one sceptically accepted argument in $\mathcal{H}$.*

Proof. Given Fact 16 it suffices to give MSOL sentences expressing each of these properties.

(a) PREF-EXT($\mathcal{X}, \mathcal{A}$)

$$\exists U \subseteq \mathcal{X}(U \neq \emptyset) \wedge ADM(\mathcal{X}, \mathcal{A}, U)$$

where $ADM(\mathcal{X}, \mathcal{A}, U)$ is the predicate

$$\forall x \in \mathcal{X}\ \forall y \in \mathcal{X}\langle x, y\rangle \in \mathcal{A} \rightarrow \big(\neg(x \in U) \vee \neg(y \in U)\big) \wedge \big(y \in U \rightarrow \big(\exists z(z \in U) \wedge \langle z, x\rangle \in \mathcal{A}\big)\big)$$

Note that we use the abbreviated form $U \neq \emptyset$ rather than the more involved $\exists u \in \mathcal{X}(u \in U)$. Thus, the given expression represents the property of $\langle \mathcal{X}, \mathcal{A}\rangle$ having a non-empty preferred extension via the conditions that there is some non-empty subset (U) of $\mathcal{X}$ which is admissible, i.e. U is conflict-free and for any argument $x \notin U$ attacking an argument $y \in U$, there is some $z \in U$ that counterattacks x.

[14] The satisfaction relation $\Phi \models \mathcal{H}$ is defined in the usual inductive style via the structure of the MSOL sentence Φ.

(b) STAB-EXT$(\mathcal{X}, \mathcal{A})$

$$\exists U \subseteq \mathcal{X} ADM(\mathcal{X}, \mathcal{A}, U) \wedge \forall x \in \mathcal{X} \neg(x \in U) \rightarrow \big(\exists z \in U \langle z, x\rangle \in \mathcal{A}\big)$$

That is, $\langle \mathcal{X}, \mathcal{A}\rangle$ has a stable extension if there is a subset U of $\mathcal{X}$ which is admissible and attacks any argument not contained in it.

(c) COHERENT$(\mathcal{X}, \mathcal{A})$

$$\forall U \subseteq \mathcal{X} PREF(\mathcal{X}, \mathcal{A}, U) \rightarrow STABLE(\mathcal{X}, \mathcal{A}, U)$$

where $STABLE(\mathcal{X}, \mathcal{A}, U)$ is the predicate,

$$ADM(\mathcal{X}, \mathcal{A}, U) \wedge \forall x \in \mathcal{X} \neg(x \in U) \rightarrow (\exists z \in U \langle z, x\rangle \in \mathcal{A})$$

and $PREF(\mathcal{X}, \mathcal{A}, U)$ the predicate

$$ADM(\mathcal{X}, \mathcal{A}, U) \wedge MAXIMAL(\mathcal{X}, \mathcal{A}, U)$$

with $MAXIMAL(\mathcal{X}, \mathcal{A}, U)$ defined as,

$$\forall W \subseteq \mathcal{X}\ \forall Z \subseteq \mathcal{X}\big((Z = U \cup W) \wedge ADM(\mathcal{X}, \mathcal{A}, Z)\big) \rightarrow (W \subseteq U)$$

Again we use abbreviated forms $Z = U \cup W$ and $W \subseteq U$ noting,

$$Z = U \cup W \equiv \forall x \in \mathcal{X}(x \in Z) \leftrightarrow (x \in U \vee x \in W)$$
$$W \subseteq U \equiv \forall x \in \mathcal{X}(x \in W) \rightarrow (x \in U)$$

In total this expression captures the concept of coherence via: any subset of $\mathcal{X}$ which is a preferred extension is also stable. A subset, U, being a preferred extension if it is both admissible and maximal, i.e. for every W for which $U \cup W$ is admissible it holds that W is a subset of U.

(d) There is at least one sceptically accepted argument in $\mathcal{H}(\mathcal{X}, \mathcal{A})$.

$$\exists x \in \mathcal{X} \forall U \subseteq \mathcal{X} PREF(\mathcal{X}, \mathcal{A}, U) \rightarrow (x \in U) \qquad \square$$

Although Theorem 17 establishes the existence of efficient algorithms for decision problems whose complexities in general are NP and Π_2^p-complete, it does not aid with problems concerning the properties of specific arguments within a given system, e.g. CA$(\mathcal{H}, x)$. Suppose, however, we define $D(\mathcal{H})$ as

$$\max_{x \in \mathcal{X}} \big|\{y\colon \langle y, x\rangle \in \mathcal{A} \text{ or } \langle x, y\rangle \in \mathcal{A}\}\big|$$

That is, in standard graph-theoretic terminology, $D(\mathcal{H})$ is the maximum *degree*—number of attacks made on and by—taken over all arguments x of $\mathcal{H}$. We can obtain algorithms whose run-time is $O(f(q)n^c)$ for $\text{CA}_{\{\}}(\mathcal{H}, S)$: here f is some fixed function $f : \mathbf{N} \rightarrow \mathbf{N}$, $n = |\mathcal{X}|$, c is a constant (independent of $\mathcal{H}$) and q is the parameter $tw(\mathcal{H}) \times D(\mathcal{H})$, that is, in terms of the framework of *fixed-parameter complexity* pioneered by Downey and Fellows [22], $\text{CA}_{\{\}}(\mathcal{H}, S)$ is fixed-parameter tractable (FPT) with respect to the parameter q.

In order to prove this we exploit results from Gottlob et al. [32] in which a parameter with to respect which CNF-SAT is FPT was presented.

Definition 18. Let $\Phi(Z_n)$ be a CNF formula with clause set $\{C_1, C_2, \ldots, C_m\}$. The *primal graph* of Φ, denoted $\text{P}_\Phi(Z_n, E)$, is the (undirected) graph with vertices labelled by the propositional variables defining Φ, and whose edge set, E, is,

$$\big\{\{z_i, z_j\}\colon z_i \text{ and } z_j \text{ occur as variables in some clause } C \text{ of } \Phi\big\}$$

Fact 19. *(See Gottlob et al. [32].)* CNF-SAT *is* FPT *w.r.t. the parameter* $tw(\text{P}_\Phi)$.

There have, subsequently, been a number of FPT approaches to CNF-SAT—notably work discussed in Szeider [38]—that consider alternative graph-theoretic representations of CNF formulae. In principle by adopting approaches related to the methods we now describe in Theorem 20 these too would lead to (potentially, improved) FPT methods for CA.

Theorem 20. $\text{CA}_{\{\}}$ *is* FPT *w.r.t. the parameter* $tw(\mathcal{H}) \times D(\mathcal{H})$.

Proof. Let $\mathcal{H}(\mathcal{X}, \mathcal{A})$ have $tw(\mathcal{H}) = r$ and consider the CNF formula, $\Psi^{\mathcal{H}}(X_n)$, as defined in the proof of Theorem 13, i.e.

$$\Psi^{\mathcal{H}}(X_n) = \bigwedge_{\langle x_i, x_j\rangle \in \mathcal{A}} (\neg x_i \vee \neg x_j) \wedge \left(\neg x_j \vee \bigvee_{x_k:\ \langle x_k, x_i\rangle \in \mathcal{A}} x_k\right)$$

Notice that $\text{P}_{\Psi}^{\mathcal{H}}(X_n, E)$ contains the undirected form of $\mathcal{H}(\mathcal{X}, \mathcal{A})$ as a subgraph by virtue of the clause set $\bigwedge_{\langle x_i, x_j\rangle \in \mathcal{A}} (\neg x_i \vee \neg x_j)$. The additional edges of $\text{P}_{\Psi}^{\mathcal{H}}$ are those arising from the clauses $(\neg x_j \vee \bigvee_{x_k:\ \langle x_k, x_i\rangle \in \mathcal{A}} x_k)$. The edges, $E_{\langle x_i, x_j\rangle}$ in E contributed by this clause associated with the attack $\langle x_i, x_j\rangle$ being

$$E_{\langle x_i, x_j\rangle} = \big\{\{x_j, x_k\}\colon\ \langle x_i, x_j\rangle \in \mathcal{A} \text{ and } \langle x_k, x_i\rangle \in \mathcal{A}\big\} \cup \big\{\{x_k, x_l\}\colon\ \langle x_k, x_i\rangle \in \mathcal{A} \text{ and } \langle x_l, x_i\rangle \in \mathcal{A}\big\}$$

For each $x_i \in \mathcal{X}$ define the set of edges X_i by

$$\begin{aligned} X_i = {} & \big\{\{y_j, z_k\}\colon\ \langle y_j, x_i\rangle \in \mathcal{A} \text{ and } \langle x_i, z_k\rangle \in \mathcal{A}\big\} \\ & \cup \big\{\{y_j, y_k\}\colon\ \langle y_j, x_i\rangle \in \mathcal{A} \text{ and } \langle y_k, x_i\rangle \in \mathcal{A}\big\} \\ & \cup \big\{\{z_j, z_k\}\colon\ \langle x_i, z_j\rangle \in \mathcal{A} \text{ and } \langle x_i, z_k\rangle \in \mathcal{A}\big\} \end{aligned}$$

Then if $H(X, A)$ is the undirected form of $\mathcal{H}(\mathcal{X}, \mathcal{A})$ then not only is $H(X, A)$ a subgraph of $\text{P}_{\Psi}^{\mathcal{H}}(X_n, E)$, but $\text{P}_{\Psi}^{\mathcal{H}}(X_n, E)$ is in turn a subgraph of H^{aug} where H^{aug} has vertex set X_n and edge set

$$F^{\text{aug}} = A \cup \bigcup_{x_i \in X} X_i$$

From these it observations it follows that $tw(H) \leqslant tw(\text{P}_{\Psi}^{\mathcal{H}}) \leqslant tw(H^{\text{aug}})$ and thus bounding the width of a tree-decomposition of H^{aug} gives an upper bound on the treewidth of the primal graph $\text{P}_{\Psi}^{\mathcal{H}}(X_n, E)$ of $\Psi^{\mathcal{H}}(X_n)$.

Let $\langle T, S\rangle$ be a width r tree decomposition of $\mathcal{H}(\mathcal{X}, \mathcal{A})$, with $S = \{S_1, S_2, \ldots, S_m\}$, $S_i \subseteq \mathcal{X}$ and $T(V, F)$ the tree structure linking the family of sets indexed by $V = \{V_1, \ldots, V_m\}$. Form the family of sets $Y = \{Y_1, Y_2, \ldots, Y_m\}$ via

$$Y_i = S_i \cup \bigcup_{x_i \in \mathcal{X}} \big\{y, z\colon\ \langle y, x\rangle \in \mathcal{A} \text{ or } \langle x, z\rangle \in \mathcal{A}\big\}$$

With this, $\langle T, Y\rangle$ is a tree decomposition of $H^{\text{aug}}(X, F^{\text{aug}})$. Furthermore, its width is at most $(D(H) + 1)(tw(H) + 1) - 1$: each S_i contains at most $tw(H) + 1$ members, each of which can contribute at most $D(H)$ new elements to S_i in addition to those already present. It follows that

$$tw\big(\text{P}_{\Psi}^{\mathcal{H}}\big) \leqslant tw(H) + D(H)\big(tw(H) + 1\big) = \big(D(H) + 1\big)\big(tw(H) + 1\big) - 1$$

Thus, given an instance $\langle \mathcal{H}, S\rangle$ of $\text{CA}_{\{\}}$ and a width r tree decomposition of $\mathcal{H}$, we may now apply the methods described in [32] to test satisfiability of the CNF formula

$$\Phi(X_n) = \left(\bigwedge_{x \in S} x\right) \wedge \Psi^{\mathcal{H}}(X_n)$$

via a tree decomposition of P_{Φ} having width at most $(D(H) + 1)(r + 1) - 1$. □

8. Value-based argument frameworks

To conclude we consider the effect that restricting the underlying graph structure has with respect to value-based argument systems. We recall the following definitions from Bench-Capon [9].

Definition 21. A *value-based argumentation framework* (VAF), is defined by a triple $\langle \mathcal{H}(\mathcal{X}, \mathcal{A}), \mathcal{V}, \eta\rangle$, where $\mathcal{H}(\mathcal{X}, \mathcal{A})$ is an argument system, $\mathcal{V} = \{v_1, v_2, \ldots, v_k\}$ a set of k *values*, and $\eta : \mathcal{X} \to \mathcal{V}$ a mapping that associates a value $\eta(x) \in \mathcal{V}$ with each argument $x \in \mathcal{X}$.

Table 3
Decision problems in value-based argument frameworks

Problem	Instance	Question
Subjective Acceptance (SBA)	$\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle; x\in\mathcal{X}$	$\exists\alpha$: $x\in P(\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle,\alpha)$?
Objective Acceptance (OBA)	$\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle; x\in\mathcal{X}$	$\forall\alpha$: $x\in P(\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle,\alpha)$?

An *audience* for a VAF $\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle$, is a binary relation $\mathcal{R}\subset\mathcal{V}\times\mathcal{V}$ whose (irreflexive) transitive closure, $\mathcal{R}^*$, is asymmetric, i.e. at most one of $\langle v,v'\rangle$, $\langle v',v\rangle$ are members of $\mathcal{R}^*$ for any distinct $v,v'\in\mathcal{V}$. We say that v_i *is preferred to* v_j *in the audience* $\mathcal{R}$, denoted $v_i\succ_{\mathcal{R}} v_j$, if $\langle v_i,v_j\rangle\in\mathcal{R}^*$. We say that α is a *specific* audience if α yields a *total* ordering of $\mathcal{V}$.

Using VAFs, ideas analogous to those introduced in Definition 1 by relativising the concept of "attack" using that of *successful* attack with respect to an audience. Thus,

Definition 22. Let $\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle$ be a VAF and $\mathcal{R}$ an audience. For arguments x, y in $\mathcal{X}$, x is a *successful attack* on y (or x *defeats* y) *with respect to the audience* $\mathcal{R}$ if: $\langle x,y\rangle\in\mathcal{A}$ *and* it is *not* the case that $\eta(y)\succ_{\mathcal{R}}\eta(x)$.

Replacing "attack" by "successful attack w.r.t. the audience $\mathcal{R}$", in Definition 1(b)–(f) yields definitions of "conflict-free", "admissible set" etc. relating to value-based systems, e.g. S is conflict—free w.r.t. to the audience $\mathcal{R}$ if for each x, y in S it is not the case that x successfully attacks y w.r.t. $\mathcal{R}$. It may be noted that a conflict-free set in this sense is not necessarily a conflict-free set in the sense of Definition 1(c): for x and y in S we may have $\langle x,y\rangle\in\mathcal{A}$, provided that $\eta(y)\succ_{\mathcal{R}}\eta(x)$, i.e. the value promoted by y is preferred to that promoted by x for the audience $\mathcal{R}$.

Bench-Capon [9] proves that every specific audience, α, induces a unique preferred extension within its underlying VAF: we use $P(\langle\langle\mathcal{X},\mathcal{A}\rangle,\mathcal{V},\eta\rangle,\alpha)$ to denote this extension. Analogous to the concepts of credulous and sceptical acceptance, in VAFs the ideas of *subjective* and *objective* acceptance arise,

Regarding these questions, [10,29] show the former to be NP-complete and the latter co-NP-complete. Our main result in this section is that, unlike the case of standard argument systems, even within very limited graph classes, both of these problems remain computationally hard.[15] Formally we have,

Theorem 23. *Let* $\text{SBA}^{(T)}$ *and* $\text{OBA}^{(T)}$ *be the decision problems of Table* 3 *with instances restricted to those for which the graph-structure* $\langle\mathcal{X},\mathcal{A}\rangle$ *is a tree.*

(a) $\text{SBA}^{(T)}$ *is* NP*-complete.*
(b) $\text{OBA}^{(T)}$ *is co-*NP*-complete.*

Proof. Membership in NP (for $\text{SBA}^{(T)}$) and co-NP (for $\text{OBA}^{(T)}$) follows from membership in these classes for the general versions.

For part (a), to show that $\text{SBA}^{(T)}$ is NP-hard we use a reduction from 3-SAT. It will be convenient (although is not essential to the proof) to restrict instances, $\Phi(Z_n)=\bigwedge_{j=1}^{m}C_j$, to those in which no variable z of Z_n occurs in *more than* 3 clauses.[16] Notice that given this restriction, without loss of generality, we may assume that for each variable z the literal $\neg z$ occurs in exactly one clause of Φ; the literal z in at most two (and at least one) clause of Φ.

For each variable z_i of Φ let the values $f(i)$, $s(i)$, and $n(i)$ be

$$f(i)=\min\{j\colon z_i \text{ occurs in } C_j\}$$
$$s(i)=\max\{j\colon z_i \text{ occurs in } C_j\}$$
$$n(i)=j\colon \neg z_i \text{ occurs in } C_j$$

Should z_i occur exactly once in positive form then $f(i)=s(i)$.

[15] Theorem 23 subsumes the result presented in [26, Theorem 4, p. 93] where it was proven that $\text{SBA}^{(2)}$ is NP-complete, i.e. when the underlying system is bipartite.

[16] See, e.g. [36, Proposition 9.3] for one proof that this variant of 3-SAT remains NP-hard.

We can now construct the instance $(\langle T_\Phi(\mathcal{X},\mathcal{A}),\mathcal{V}_\Phi,\eta\rangle, x)$ of $\text{SBA}^{(T)}$.

Its argument set $\mathcal{X}$ comprises (at most) $6n+m+1$ arguments,

$$\mathcal{X} = \{\Phi, C_1, C_2, \ldots, C_m\} \cup \left\{z_i^1, z_i^2, z_i^3 \colon 1 \leqslant i \leqslant n\right\} \cup \left\{\neg z_i^1, \neg z_i^2, \neg z_i^3 \colon 1 \leqslant i \leqslant n\right\}$$

(If z_i occurs exactly once in positive form then neither z_i^2 nor $\neg z_i^2$ occur in $\mathcal{X}$.)

The set of attacks, $\mathcal{A}$, is formed by

$$\begin{aligned}\mathcal{A} = &\left\{\langle C_j, \Phi\rangle \colon 1 \leqslant j \leqslant m\right\}\\ &\cup \left\{\left\langle \neg z_i^1, z_i^1\right\rangle, \left\langle \neg z_i^2, z_i^2\right\rangle \colon 1 \leqslant i \leqslant n\right\}\\ &\cup \left\{\left\langle z_i^3, \neg z_i^3\right\rangle \colon 1 \leqslant i \leqslant n\right\}\\ &\cup \left\{\left\langle z_i^1, C_{f(i)}\right\rangle, \left\langle z_i^2, C_{s(i)}\right\rangle \colon 1 \leqslant i \leqslant n\right\}\\ &\cup \left\{\left\langle \neg z_i^3, C_{n(i)}\right\rangle \colon 1 \leqslant i \leqslant n\right\}\end{aligned}$$

The value set, $\mathcal{V}_\Phi$ of the instance contains $2n+1$ members,

$$\mathcal{V}_\Phi = \{c\} \cup \{pos_i, neg_i \colon 1 \leqslant i \leqslant n\}$$

Finally the mapping, η from $\mathcal{X}$ to $\mathcal{V}_\Phi$ is defined via

$$\eta(x) = \begin{cases} c & \text{if } x \in \{\Phi, C_1, \ldots, C_m\}\\ pos_i & \text{if } x \in \{z_i^1, z_i^2, z_i^3\}\\ neg_i & \text{if } x \in \{\neg z_i^1, \neg z_i^2, \neg z_i^3\}\end{cases}$$

The construction for the CNF formula $\Phi(z_1, z_2, z_3, z_4)$ defined by

$$(z_1 \vee z_2 \vee z_3)(\neg z_2 \vee \neg z_3 \vee \neg z_4)(\neg z_1 \vee z_2 \vee z_4)$$

is illustrated in Fig. 10.

It is easy to see that $T_\Phi(\mathcal{X},\mathcal{A})$ is a tree. To complete the instance of $\text{SBA}^{(T)}$ we set the argument x to be Φ. We now claim that $(\langle T_\Phi(\mathcal{X},\mathcal{A}),\mathcal{V}_\Phi,\eta\rangle, \Phi)$ is accepted as an instance of $\text{SBA}^{(T)}$ if and only if $\Phi(Z_n)$ is satisfiable.

Suppose that $\Phi(Z_n)$ is satisfied by some instantiation $\underline{a} = \langle a_1, a_2, \ldots, a_n\rangle$ of Z_n. Consider any specific audience α for which

$$\begin{aligned} &pos_i \succ_\alpha neg_i \quad \text{if } a_i = \top\\ &neg_i \succ_\alpha pos_i \quad \text{if } a_i = \bot\\ &pos_i \succ_\alpha c \quad \forall 1 \leqslant i \leqslant n\\ &neg_i \succ_\alpha c \quad \forall 1 \leqslant i \leqslant n\end{aligned}$$

Consider the subset $S(\underline{a})$ of $\mathcal{X}$ chosen as

$$\{\Phi\} \cup \left\{z_i^1, z_i^2 \colon a_i = \top\right\} \cup \left\{\neg z_i^3 \colon a_i = \bot\right\}$$

We claim that $S(\underline{a})$ is admissible with respect to the audience α. The only attacks on Φ are from the arguments C_j, however, since $\underline{a}$ satisfies Φ, each clause has at least one true literal with this instantiation: thus C_j is successfully attacked by one of $\{z_i^1, z_i^2\}$ whenever $a_i = \top$ and $j \in \{f(i), s(i)\}$; similarly C_j is successfully attacked by $\neg z_i^3$ whenever $a_i = \bot$ and $j = n(i)$. Furthermore the attacks on $\{z_i^1, z_i^2 \colon a_i = \top\}$ by $\{\neg z_i^1, \neg z_i^2\}$ are not successful on account of the value ordering $pos_i \succ_\alpha neg_i$. In the same way, the attack on $\neg z_i^3$ by z_i^3 fails whenever $a_i = \bot$ since $neg_i \succ_\alpha pos_i$. We deduce that $S(\underline{a})$ is admissible and thus Φ subjectively accepted if $\Phi(Z_n)$ is satisfiable.

On the other hand suppose Φ is subjectively accepted and let α be a specific audience with $S \subseteq \mathcal{X}$ an admissible set w.r.t. α that contains Φ. Noting that $\eta(\Phi) = \eta(C_j) = c$ for each C_j it follows that $S \cap \{C_1, \ldots, C_m\} = \emptyset$ and, thus, each C_j must be successfully attacked by some y_i w.r.t. α, with the (unique) attack on this y_i, i.e. $\neg z_i^k$ if $y_i = z_i^k$, z_i^3

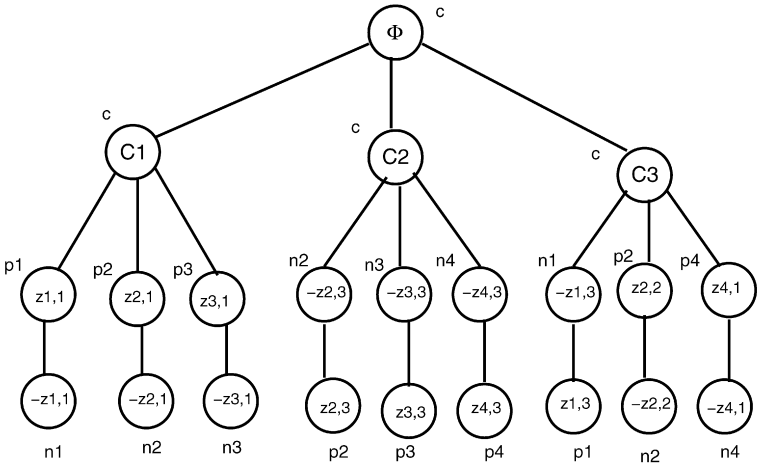

Fig. 10. T_Φ for $\Phi = (z_1 \vee z_2 \vee z_3)(\neg z_2 \vee \neg z_3 \vee \neg z_4)(\neg z_1 \vee z_2 \vee z_4)$.

if $y_i = \neg z_i^3$ failing to succeed. Now let $\{y_1, y_2, \ldots, y_m\}$ be the set of arguments for which y_j successfully attacks C_j w.r.t. α and construct the (partial) instantiation $\langle a_1, \ldots, a_n\rangle$ of Z_n with

$$a_i = \top \quad \text{if } \{z_i^1, z_i^2\} \cap \{y_1, \ldots, y_m\} \neq \emptyset$$
$$a_i = \bot \quad \text{if } \neg z_i^3 \in \{y_1, \ldots, y_m\}$$

It now suffices to observe that this instantiation is well-defined. If both $\neg z_i^3$ and z_i^k occur in $\{y_1, \ldots, y_m\}$, from the fact that α is a specific audience either $pos_i \succ_\alpha neg_i$ or $neg_i \succ_\alpha pos_i$: in the former case, $\neg z_i^3$ is successfully attacked by z_i^3 (and, hence, could not belong to S); in the latter z_i^k is successfully attacked by $\neg z_i^k$ and, again could not belong to S. We deduce that the partial instantiation $\langle a_1, \ldots, a_n\rangle$ is well-defined and satisfies $\Phi(Z_n)$.

In total, Φ is subjectively accepted in the system $\langle T_\Phi(\mathcal{X}, \mathcal{A}), \mathcal{V}_\Phi, \eta\rangle$ if and only if $\Phi(Z_n)$ is satisfiable.

Part (b) uses a similar reduction from UNSAT restricted to 3-CNF instances of the same form as part (a). Given $\langle T_\Phi(\mathcal{X}, \mathcal{A}), \mathcal{V}_\Phi, \eta\rangle$ as described earlier the instance of $\text{OBA}^{(T)}$ is formed by adding one additional argument, Φ', to $\mathcal{X}$ whose sole attacker is the argument Φ and with $\eta(\Phi') = c$. In this construction Φ' is acceptable w.r.t. to every specific audience if and only if Φ is *not* subjectively acceptable. Using an identical argument to (a), the latter holds if and only if $\Phi(Z_n)$ is unsatisfiable. □

Corollary 7. $\text{SBA}^{(T)}$ *is* NP*-complete and* $\text{OBA}^{(T)}$ *is co-*NP*-complete even if instances are restricted to binary trees.*

Proof. Apply the translation of Theorem 11 to the trees constructed in the proof of Theorem 23, assigning the value c to each new argument introduced. This translation and value allocation affects neither the subjective acceptability of Φ nor the objective acceptability of Φ'. With the exception of the root (i.e. the arguments Φ and Φ' respectively), each argument in the trees so formed attacks exactly one other argument. Similarly, with the exception of the leaf arguments which have no attackers and Φ' (which has exactly one attacker), each argument is attacked by exactly two others. □

One feature of the reduction in Theorem 23 (as, indeed, of the reduction for general VAFs given in [10,29]) is that the number of *values* $(2n + 1)$ is of the same order as the number of *arguments* in the system: in the reduction $4n + m + 1 \leqslant |\mathcal{X}| \leqslant 6n + m + 1$, however, given the restrictions on Φ it is easily seen that $2n/3 \leqslant m \leqslant n$ and hence, $|\mathcal{V}| = \Theta(|\mathcal{X}|)$. Our final result indicates that even insisting that $|\mathcal{V}| = o(|\mathcal{X}|)$ does not lead to tractable cases.

Theorem 24. *Let* $\text{SBA}^{(T,\epsilon)}$ *be the decision problem* $\text{SBA}^{(T)}$ *in which instances are restricted to those in which* $|\mathcal{V}| \leqslant |\mathcal{X}|^\epsilon$. $\forall \epsilon > 0$ $\text{SBA}^{(T,\epsilon)}$ *is* NP*-complete.*

Proof. Let $(\langle T_\Phi(\mathcal{X}, \mathcal{A}), \mathcal{V}_\Phi, \eta\rangle, \Phi)$ be the instance of $\text{OBA}^{(T)}$ constructed in the proof of Theorem 23(b). Given $\epsilon > 0$, choose $K_\epsilon \in \mathbf{N}$ as $K_\epsilon = \lceil \epsilon^{-1} \rceil$. An instance of $\text{SBA}^{(T,\epsilon)}$ is formed by taking $r = |\mathcal{X}|^{K_\epsilon - 1}$ copies of $T_\Phi - \{T_1, T_2, \ldots, T_r\}$. Letting ϕ_i denote the argument forming the root of T_i, the instance is completed by adding one further argument, $\Phi^{(\epsilon)}$ with $\eta(\Phi^{(\epsilon)}) = c$ and attacks $\langle \phi_i, \Phi^{(\epsilon)}\rangle$ for each $1 \leqslant i \leqslant r$. Recalling that ϕ_i is objectively accepted if and only if Φ is unsatisfiable it is easily seen that $\Phi^{(\epsilon)}$ is subjectively accepted if and only if Φ is satisfiable. The number of values in the constructed instance is $|\mathcal{V}_\Phi| = \text{O}(|\mathcal{X}|)$, however, the number of arguments is $|\mathcal{X}|^{K_\epsilon}$ and this is now a valid instance of $\text{SBA}^{(T,\epsilon)}$. □

9. Conclusions and development

In this paper we have considered how the complexity of a number of important decision questions in both standard and value-based argument systems is affected under various graph-theoretic restrictions: the system being k-partite; each argument being attacked by and attacking some maximum number of arguments; planar systems; and systems with bounded treewidth.

Overall the picture apparent regarding the efficacy of graph-theoretic restrictions in admitting efficient algorithmic methods is somewhat mixed. For quite general classes—planar and bounded degree systems—the complexity of decision questions remains unchanged from that of the unrestricted case. In contrast, for more limited classes, to the known examples of DAGs and symmetric frameworks can now be added bipartite systems and those with k-bounded treewidth. The nature of what characterises "efficient restrictions" from those which offer no gains may seem rather arbitrary, e.g. bipartite systems are tractable however 3-partite systems are not. A partial explanation of such phenomena is offered by our notions of "polynomial universality". Thus, although, for example, planarity is not a property of every finite argument system, by virtue of Theorem 13 there is no loss of generality (with respect to credulous acceptance issues) in assuming planarity since any system is transformable to a related planar system. Notwithstanding the fact that such translations, in general, do not simplify decision processes, there are potential applications exploiting polynomially universal properties in representing argument systems. For example, consider multiagent environments dedicated to maintaining information about admissible and preferred sets within a dynamically evolving system, knowledge concerning which is distributed over distinct agents. In earlier work, Baroni et al. [6] have shown the graph-theoretic concept of *strongly connected component* (SCC) decompositions provides a useful mechanism with which to approach this environment. One can envisage complementing such techniques by exploiting 4-partiteness and/or planarity as universal properties: the former suggests a natural partition of arguments over four agents with the set maintained by each being conflict-free and questions about a specific argument, p say, requiring local resolution via the (at most two) agents allocated its attackers; similar methods, using properties of planar graphs, e.g. the separator results of Lipton and Tarjan [34], may also offer useful mechanisms. Such treatments are the subject of current work.

We conclude by raising a select number of interesting open issues.

Potentially the most interesting suite of issues arises from the results on bounded treewidth decision problems given in Theorems 17 and 20. Although following the algorithm synthesis template of, for example [3], produces a linear time algorithm via some MSOL sentence and width k tree decomposition, such algorithms are likely to be rather opaque with the linear time method concealing large constant factors that increase rapidly with the treewidth bound.[17] Given such eventualities it is tempting to view the algorithms guaranteed by Courcelle's Theorem as "proof of concept", i.e. that efficient algorithms exist in principle, rather than as viable solutions in themselves. This interpretation then raises the question of forming practical algorithmic methods. Thus suppose one limits attention to systems of treewidth 2 or 3, relying on the nature of argument systems as might arise in real settings to be of this form. Rather than synthesising methods indirectly via Courcelle's Theorem, one could attempt to develop practical *direct* methods. There are several promising indications that this is a realistic objective: the precise characterisation of those graphs having treewidth 2, e.g. [14, Theorem 42, p. 22]; and the dynamic programming templates discussed in [12]. A similar issue arises with respect to the methods discussed for determining credulous acceptability in Theorem 20. Although arguably of a less extreme nature, the algorithm for deciding $\text{CA}(\mathcal{H}, x)$ in the case $tw(\mathcal{H}) = r$ and $D(\mathcal{H}) = d$ is rather

[17] While the comparison is rather unfair the relationship between the property captured by a complex MSOL expression and the width k algorithm synthesised is analogous to that of a high-level programming language description and the binary machine code resulting from its compilation. In addition, we recall that (relative to the full formal description of [3]) the sentences given in the proof of Theorem 17 require further development in order to eliminate constructs such as $U \neq \emptyset$, $V \subseteq W$, etc. prior to applying the algorithm construction process.

indirect involving, as it does, a translation into CNF.[18] Thus there is, again, the issue of finding direct algorithmic solutions, i.e not via CNF-SAT formulations, for systems with small treewidth, e.g. $tw(\mathcal{H}) \leqslant 3$.

A final group of problems regarding bounded treewidth systems concerns combining *dialogue game* methods, e.g. the TPI-disputes studied in [28,40], or the reasoning schema presented in [24], using both the graph-theoretic form of $\mathcal{H}$ and a width k tree decomposition of $\mathcal{H}$. Among the reasons why treewidth decompositions may provide useful representations for both of these approaches are the following. The pathological examples for which exponential length TPI-disputes result constructed in [28], cannot occur in width k systems: the mechanism used to form such cases is via the translation of "provably hard" unsatisfiable CNF instances[19]: such instances, however, necessarily have primal graphs with large treewidth. Regarding the application to the dialogue structure promoted in [24], we observe that one standard design approach for efficient algorithms based on tree decompositions, discussed in [12], is to construct solutions working from the leaves of the tree decomposition building towards its root: such techniques mirror the reasoning methods discussed in [24].

The results presented in Section 8 indicate that efficient methods for the central decision questions—SBA and OBA—are unlikely to come about through simply limiting the underlying directed graph form: binary tree structures being the most basic non-trivial graph class.[20] While Theorem 23 and Corollary 7 seem to offer rather pessimistic prospects for the possibility of developing tractable variants of SBA, these are in some respect unsurprising: a critical distinction between the nature of decision problems in VAFs and in standard argument systems concerns the search space examined.

For SBA this is the set of all specific audiences, i.e. the $k!$ total orderings of $\mathcal{V}$; in decision problems such as CA, this space is the set of all subsets of $\mathcal{X}$, Searching over orderings of structures within combinatorial objects (as opposed to subsets) is known to give rise to decision questions which often remain hard even in restricted instances,[21] a notable example being the *bandwidth minimisation* problem [31, GT40, p. 200]: this, as with SBA, is NP-hard even when restricted to binary trees.

It might, therefore, be argued that in order to identify non-trivial tractable variants of SBA, not only is it needed to restrict the underlying argument graph but also to restrict how the value set $\mathcal{V}$ and mapping $\eta : \mathcal{X} \to \mathcal{V}$ interact with it. While $\mathcal{V}$ defines a parameter w.r.t which SBA is FPT—the procedure described in [9] giving a bound $O(k!|\mathcal{A}|)$ via the brute-force approach of testing each specific audience in turn—an open question is whether alternative approaches can succeed. One aspect of the hardness proofs in Theorem 23 and those of [26,29] is that there is a single value (c) associated with "many" arguments, i.e. $|\eta^{-1}(c)| = \Theta(|\mathcal{X}|)$, and a large number of values (pos_i, neg_i) associated with only a few (at most 3 in the proof of Theorem 23) arguments. This suggests two possible approaches with which to consider alternative restrictions of SBA instances,

(R1) by bounding the minimum and maximum number of occurrences of any given value $v \in \mathcal{V}$
(R2) by bounding the number of occurrences of attacks $\langle x, y \rangle$ in which $\eta(x) = \eta(y)$.

Theorem 24 and the trivial observation that at least one value must be common to $|\mathcal{X}|/|\mathcal{V}|$ arguments limit, however, the possible range of interest in trying to exploit R1: if $|\mathcal{V}| = o(|\mathcal{X}|)$, e.g. the case considered in Theorem 24, then some value is shared by $\omega(1)$ arguments. In trying to limit the number of occurrences of any value to be a constant—thereby forcing $|\mathcal{V}| = \Theta(|\mathcal{X}|)$—another difficulty arises. Thus, suppose $\text{SBA}^{(\mathcal{V},\leqslant k)}$ is the decision problem SBA restricted to instances for which $\forall v \in \mathcal{V} |\eta^{-1}(v)| \leqslant k$, i.e. at most k arguments share a common value, $v \in \mathcal{V}$. Similarly, $\text{SBA}^{(T),(\mathcal{V},\leqslant k)}$ is this problem with instances additionally restricted to trees.

[18] In addition, the methods of [32] require a further translation from CNF to a CSP problem in order to use an algorithm of Yannakakis [41].

[19] The notion of "hardness" is that of proof length within certain weak (but complete) propositional proof systems, see e.g. Cook and Reckhow [16], Beame and Pitassi [7], and Urquhart [39] for technical background. In [28] the TPI formalism is shown equivalent (in the sense of [16]) to the CUT-free Gentzen calculus.

[20] One could limit structures further to, e.g. systems $\mathcal{H}$, with $D(\mathcal{H}) \leqslant 2$. In this case, retaining the connectivity assumption, one has only *paths* and simple cyclic structures: both cases are completely characterised in the original presentations of Bench-Capon [8,9].

[21] The problem of deciding if an n-vertex graph has a hamiltonian cycle may appear to be an exception to this generalisation, however, one can sensibly treat the search space in this instance, not as all possible vertex orderings ($n!$), but rather as n element subsets of the edges: such viewpoints are exploited in efficient algorithms for testing hamiltonicity of graphs with small treewidth by progressively building "partial solutions" defining paths between vertex subsets.

Theorem 25. $\text{SBA}^{(T),(\mathcal{V},\leqslant 3)}$ *is* NP*-complete even if instances are binary trees.*

Proof. The proof uses the binary tree structure of Corollary 7, with a modification of the definition of $\mathcal{V}$ and the associated mapping η. Details are presented in Appendix B. □

The problem, $\text{SBA}^{(\mathcal{V},\leqslant 1)}$ on the other hand is trivial: any argument, x, is subjectively accepted in such instances simply by choosing an audience in which $\eta(x)$ is the most preferred value. Between the extremes of this case and that of Theorem 25, we conjecture that $\text{SBA}^{(\mathcal{V},\leqslant 2)}$ is polynomial time decidable. Regarding the approach suggested by R2, suppose we define the following parameter on VAFs:

$$\sigma\big(\langle\mathcal{H}(\mathcal{X},\mathcal{A}),\mathcal{V},\eta\rangle\big) = \big|\big\{\langle x,y\rangle\in\mathcal{A}\colon\ \eta(x)=\eta(y)\big\}\big|$$

Thus, $\sigma(\langle\mathcal{H},\mathcal{V},\eta\rangle)$ is the number of attacks in $\mathcal{A}$ involving arguments with the same value. We offer as a final conjecture the claim that SBA is FPT with respect to the parameter $\sigma(\langle\mathcal{H},\mathcal{V},\eta\rangle)$. This, again, forms the subject of current work.

Appendix A. Further properties of $\mathcal{G}_{\Phi}$

In this appendix we present the proof of the result stated in Theorem 8(b).

Proof of Theorem 8(b). Recall that this asserts $\text{SA}^{(k)}$ and $\text{COHERENT}^{(k)}$ are Π_2^p-complete for k-partite systems with $k\geqslant 3$.

It suffices to construct a 3-partite argument system $\mathcal{G}_{\Phi}^{(3)}$ from the system $\mathcal{G}_{\Phi}$ of Section 3.2. Noting that Φ is sceptically accepted in the latter system if and only if $\Phi(Y_n,Z_n)$ is accepted as an instance of QSAT_2^{Π}, $\mathcal{G}_{\Phi}^{(3)}$ is designed to preserve this property. In order to form $\mathcal{G}_{\Phi}^{(3)}$ the subsystem of four arguments $\{\Phi,b_1,b_2,b_3\}$ in $\mathcal{G}_{\Phi}$ is replaced by the system of Fig. 11.

From the properties of $\mathcal{G}_{\Phi}$, it is still the case that for every satisfying instantiation of the CNF $\Phi(Y_n,Z_n)$ there is a preferred extension of $\mathcal{G}_{\Phi}^{(3)}$ containing Φ. Such preferred extensions additionally contain the argument p_2. It follows easily from this that $\text{SA}(\mathcal{G}_{\Phi}^{(3)},\Phi)$ holds if and only if $\Phi(Y_n,Z_n)$ is a positive instance of QSAT_2^{Π}. We further observe that the system $\mathcal{G}_{\Phi}^{(3)}$ is coherent if and only if Φ is sceptically accepted. To complete the proof it remains to show that $\mathcal{G}_{\Phi}^{(3)}$ is 3-partite. We can construct a three colouring of $\mathcal{G}_{\Phi}^{(3)}$ by assigning colour R to $\{\Phi,y_1,\ldots,y_n,z_1,\ldots,z_n\}$; colour B to $\{\neg y_1,\ldots,\neg y_n,\neg z_1,\ldots,\neg z_n\}$ and G to $\{C_1,\ldots,C_m\}$. This leaves the arguments $\{b_1,b_2,b_3.p_1.p_2\}$ uncoloured, however, the 3-colouring is completed using G for $\{p_1,b_1\}$; B for $\{b_2,p_2\}$; and R for $\{b_3\}$. □

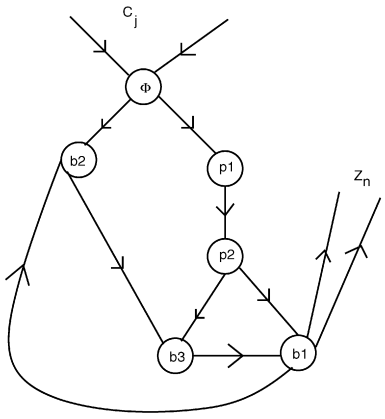

Fig. 11. Local modification of the argument system $\mathcal{G}_{\Phi}$.

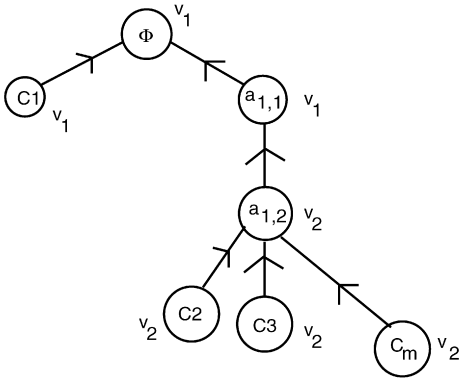

Fig. 12. Reducing number of occurrences of the value c in T_Φ.

Appendix B. Proof of Theorem 25

Recall that Theorem 25 asserts $\text{SBA}^{(T),(\mathcal{V}\leqslant 3)}$ is NP-complete even when instances are restricted to binary trees.

Given an instance, $\Phi(Z_n)$ of 3-SAT as in the proof of Theorem 23, i.e. in which every variable occurs in at most three distinct clauses[22] of Φ, consider the instance of $\text{SBA}^{(T)}$—$\langle T_\Phi(\mathcal{X},\mathcal{A}),\mathcal{V}_\Phi,\eta\rangle$ constructed. In this instance each of the values $v\in\{pos_i,neg_i\colon\ 1\leqslant i\leqslant n\}$ has $|\eta^{-1}(v)|\leqslant 3$. Renaming the value c to v_1, we have $|\eta^{-1}(v_1)|=m+1$—the argument Φ and the m arguments representing clauses. Introduce a new value v_2 together with arguments $a_{1,1}$ and $a_{1,2}$ and replace the sub-tree formed by $\{\Phi,C_1,C_2,\ldots,C_m\}$ with the structure of Fig. 12.

In the resulting tree there are now 3 occurrences of the value v_1 and m occurrences of the new value v_2. Applying the same replacement method to the sub-tree with root $a_{1,2}$ and introducing a further new value v_3, T_Φ will be modified to a tree, $T_\Phi^{(3)}$ with additional arguments

$$\{a_{j,1},a_{j,2}\colon\ 1\leqslant j\leqslant m-2\}$$

New attacks,

$$\begin{aligned}&\big\{\langle a_{1,1},\Phi\rangle,\langle C_m,a_{m-2,2}\rangle\big\}\cup\big\{\langle a_{j,1},a_{j-1,2}\rangle\colon\ 2\leqslant j\leqslant m-2\big\}\\&\quad\cup\big\{\langle a_{j,2},a_{j,1}\rangle\colon\ 1\leqslant j\leqslant m-2\big\}\cup\big\{\langle C_j,a_{j-1,2}\rangle\colon\ 2\leqslant j\leqslant m-1\big\}\end{aligned}$$

and value set

$$\mathcal{V}^{(3)}=\mathcal{V}_\Phi\cup\{v_1,v_2,\ldots,v_{m-1}\}$$

The mapping η as it affects clauses and these new arguments is now,

$$\eta(q)=\begin{cases}v_1 & \text{if } q\in\{\Phi,C_1,a_{1,1}\}\\ v_j & \text{if } q\in\{a_{j,1},a_{j-1,2},C_j\}\text{ and }2\leqslant j\leqslant m-2\\ v_{m-1} & \text{if } q\in\{a_{m-2,2},C_{m-1},C_m\}\end{cases}$$

This now satisfies $|\eta^{-1}(v)|\leqslant 3$ for every value in $\mathcal{V}^{(3)}$.

The final stage is to replace the sub-trees rooted at each clause argument C_j using binary trees. The typical replacement is shown in Fig. 13.

In forming this final (binary) tree $2m$ new arguments are introduced, $\{b_{j,1},b_{j,2}\colon\ 1\leqslant j\leqslant m\}$ and a further m values $\{w_j\colon\ 1\leqslant j\leqslant m\}$. The mapping η being extended for these new arguments via $\eta(b_{j,1})=\eta(b_{j,2})=w_j$.

We now claim that Φ is subjectively accepted in the resulting binary tree if and only if $\Phi(Z_n)$ is satisfiable.

[22] In contrast to Theorem 23 in which this assumption is made for cosmetic purposes of presentational ease, in the current proof this variant of 3-SAT is needed in order to ensure appropriately few occurrences of the values pos_i and neg_i.

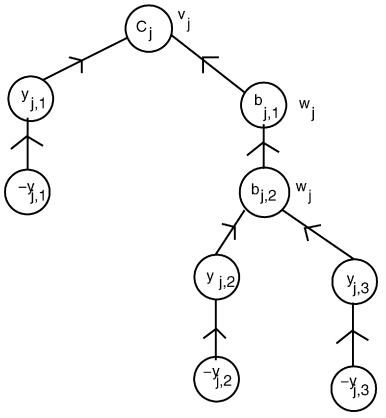

Fig. 13. Reducing clause sub-trees to binary trees in $T_{\Phi}^{(3)}$.

Suppose first that $\Phi(Z_n)$ is satisfied using an instantiation $\underline{a} = \langle a_1, \ldots, a_n \rangle$. Consider any specific audience, α in which

$$
\begin{aligned}
&pos_i \succ_\alpha neg_i \quad \text{if } a_i = \top \\
&neg_i \succ_\alpha pos_i \quad \text{if } a_i = \bot \\
&pos_i \succ_\alpha v_j \quad \forall 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m-1 \\
&neg_i \succ_\alpha v_j \quad \forall 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m-1 \\
&pos_i \succ_\alpha w_j \quad \forall 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m \\
&neg_i \succ_\alpha w_j \quad \forall 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m \\
&w_j \succ_\alpha v_j \quad \forall 1 \leqslant j \leqslant m-1 \text{ and } w_m \succ_\alpha v_{m-1} \\
&v_j \succ_\alpha v_{j-1} \quad \forall 2 \leqslant j \leqslant m-1
\end{aligned}
$$

Since $\underline{a}$ satisfies $\Phi(Z_n)$ each clause C_j has at least one literal, assigned the value $\top$: if the corresponding literal in $T_{\Phi}^{(3)}$ is the (unique) literal attacking the clause C_j then this attack is successful; otherwise the corresponding literal (successfully) attacks $b_{j,2}$ so that $b_{j,1}$ successfully attacks C_j. It follows that in the unique preferred extension, $P(\alpha)$ induced by α, we have $P(\alpha) \cap \{C_1, \ldots, C_m\} = \emptyset$. From this, and the ordering $v_j \succ_\alpha v_{j-1}$ we deduce that the attack by $a_{j,2}$ on $a_{j,1}$ succeeds for each $1 \leqslant j \leqslant m-2$, i.e. $\{a_{1,2}, \ldots, a_{m-2,2}\} \subset P(\alpha)$ and hence $\Phi \in P(\alpha)$ as claimed.

On the other hand suppose the audience α is such that $\Phi \in P(\alpha)$. From the same reasoning as that in the proof of Theorem 23 we can construct an instantiation, $\underline{a} = \langle a_1, \ldots, a_n \rangle$ of Z_n via $a_i = \top$ if and only if $pos_i \succ_\alpha neg_i$. Now since $\Phi \in P(\alpha)$ an easy argument establishes $a_{j,1} \notin P(\alpha)$ and $a_{j,2} \in P(\alpha)$ for every $1 \leqslant j \leqslant m-2$. To complete the proof it suffices to show that this instantiation must satisfy $\Phi(Z_n)$. Suppose, to the contrary, that $\Phi(\underline{a}) = \bot$ and let C_j be any clause that it is falsified by $\underline{a}$. Consider the corresponding argument, C_j within $T_{\Phi}^{(3)}$. It cannot be the case that $C_j = C_1$: for in that case the attack by C_1 on Φ succeeds, contradicting the assumption that $\Phi \in P(\alpha)$. The alternative, however, is that C_j attacks some argument $a_{j-1,2}$ or $a_{m-2,2}$ for $C_j = C_m$. Again C_j falsified by $\underline{a}$ contradicts the property $a_{j,2} \in P(\alpha)$ which holds of any preferred extension with respect to α containing Φ. Thus, every clause of $\Phi(Z_n)$ must be satisfied by $\underline{a}$ and it follows that from a specific audience under which Φ is subjectively accepted we can construct a satisfying instantiation of $\Phi(Z_n)$.

References

[1] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, Ann. Math. Artificial Intelligence 34 (2002) 197–215.
[2] S. Arnborg, D.G. Corneil, A. Proskurowski, Complexity of finding embeddings in a k-tree, SIAM J. Alg. Disc. Methods 8 (1987) 277–284.
[3] S. Arnborg, J. Lagergren, D. Seese, Easy problems for tree-decomposable graphs, J. Algorithms 12 (1991) 308–340.
[4] S. Arnborg, A. Proskurowski, Characterization and recognition of partial 3-trees, SIAM J. Alg. Disc. Methods 7 (1986) 305–314.

[5] P. Baroni, M. Giacomin, Solving semantic problems with odd-length cycles in argumentation, in: Proc. 7th European Conf. on Symbolic and Quantitative Approaches to Reasoning With Uncertainty (ECSQARU), in: Lecture Notes in Artificial Intelligence, vol. 2711, Springer-Verlag, Berlin, 2003, pp. 440–451.
[6] P. Baroni, M. Giacomin, G. Guida, SCC-recursiveness: a general schema for argumentation semantics, Artificial Intelligence 168 (1–2) (2005) 162–210.
[7] P. Beame, T. Pitassi, Propositional proof complexity: Past, present, and future, Bull. EATCS 65 (1998) 66–89.
[8] T.J.M. Bench-Capon, Agreeing to differ: Modelling persuasive dialogue between parties with different values, Informal Logic 22 (3) (2002) 231–245.
[9] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, J. Logic Comput. 13 (3) (2003) 429–448.
[10] T.J.M. Bench-Capon, S. Doutre, P.E. Dunne, Audiences in argumentation frameworks, Artificial Intelligence 171 (2007) 42–71.
[11] C. Berge, Graphs and Hypergraphs, North-Holland, Amsterdam, 1973.
[12] H.L. Bodlaender, Dynamic programming on graphs with bounded treewidth, in: Proc. 15th ICALP, in: Lecture Notes in Computer Science, vol. 317, Springer-Verlag, Berlin, 1988, pp. 105–119.
[13] H.L. Bodlaender, A linear time algorithm for finding tree-decompositions of small treewidth, SIAM J. Comput. 25 (1996) 1305–1317.
[14] H.L. Bodlaender, A partial k-arboretum of graphs with bounded treewidth, Theoret. Comput. Sci. 209 (1998) 1–45.
[15] A. Bondarenko, P.M. Dung, R.A. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artificial Intelligence 93 (1997) 63–101.
[16] S.A. Cook, R.A. Reckhow, The relative complexity of propositional proof systems, J. Symbolic Logic 44 (1) (1979) 36–50.
[17] S. Coste-Marquis, C. Devred, P. Marquis, Symmetric argumentation frameworks, in: L. Godo (Ed.), Proc. 8th European Conf. on Symbolic and Quantitative Approaches to Reasoning With Uncertainty (ECSQARU), in: Lecture Notes in Artificial Intelligence, vol. 3571, Springer-Verlag, Berlin, 2005, pp. 317–328.
[18] B. Courcelle, The monadic second-order logic of graphs. I. Recognizable sets of finite graphs, Inform. Comput. 85 (1) (1990) 12–75.
[19] B. Courcelle, The monadic second-order logic of graphs III: tree-decompositions, minor and complexity issues, Informatique Théorique et Applications 26 (1992) 257–286.
[20] Y. Dimopoulos, B. Nebel, F. Toni, On the computational complexity of assumption-based argumentation for default reasoning, Artificial Intelligence 141 (2002) 55–78.
[21] Y. Dimopoulos, A. Torres, Graph theoretical structures in logic programs and default theories, Theoret. Comput. Sci. 170 (1996) 209–244.
[22] R.G. Downey, M.R. Fellows, Fixed parameter tractability and completeness I: basic results, SIAM J. Comput. 24 (1995) 873–921.
[23] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (1995) 321–357.
[24] P.M. Dung, R.A. Kowalski, F. Toni, Dialectic proof procedures for assumption-based, admissible argumentation, Artificial Intelligence 170 (2) (2006) 114–159.
[25] P.E. Dunne, The Complexity of Boolean Networks, Academic Press, New York, 1988.
[26] P.E. Dunne, Complexity properties of restricted abstract argument systems, in: P.E. Dunne, T.J.M. Bench-Capon (Eds.), Computational Models of Argument (Proc. COMMA 2006), in: Frontiers in Artificial Intelligence and Applications, vol. 144, IOS Press, Amsterdam, 2006, pp. 85–96.
[27] P.E. Dunne, T.J.M. Bench-Capon, Coherence in finite argument systems, Artificial Intelligence 141 (2002) 187–203.
[28] P.E. Dunne, T.J.M. Bench-Capon, Two party immediate response disputes: properties and efficiency, Artificial Intelligence 149 (2003) 221–250.
[29] P.E. Dunne, T.J.M. Bench-Capon, Complexity in value-based argument systems, in: Proc. 9th JELIA, in: Lecture Notes in Artificial Intelligence, vol. 3229, Springer-Verlag, Berlin, 2004, pp. 360–371.
[30] A.S. Fraenkel, Planar kernel and Grundy number with $d \leqslant 3$, $d.out \leqslant 2$, $d.in \leqslant 2$ are NP-complete, Disc. Appl. Math. 3 (4) (1981) 257–262.
[31] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, New York, 1979.
[32] G. Gottlob, F. Scarcello, M. Sideri, Fixed-parameter complexity in AI and nonmonotonic reasoning, Artificial Intelligence 138 (2002) 55–86.
[33] J.E. Hopcroft, J.K. Wong, Linear time algorithm for isomorphism of planar graphs (preliminary report), in: STOC '74: Proc. of the 6th Annual ACM Symposium on Theory of Computing, ACM Press, New York, 1974, pp. 172–184.
[34] R.J. Lipton, R.E. Tarjan, A separator theorem for planar graphs, SIAM J. Appl. Math. 36 (1979) 177–189.
[35] W.F. McColl, Planar crossovers, IEEE Trans. Comput. C-30 (1981) 223–225.
[36] C.H. Papadimitriou, Computational Complexity, Addison-Wesley, Reading, MA, 1994.
[37] N. Robertson, P.D. Seymour, Graph minors II: algorithmic aspects of treewidth, J. Algorithms 7 (1986) 309–322.
[38] S. Szeider, On fixed-parameter tractable parameterizations of SAT, in: SAT'03, in: Lecture Notes in Computer Science, vol. 2919, Springer-Verlag, Berlin, 2003, pp. 188–202.
[39] A. Urquhart, The complexity of propositional proofs, Bull. Symbolic Logic 1 (4) (1995) 425–467.
[40] G. Vreeswijk, H. Prakken, Credulous and sceptical argument games for preferred semantics, in: Proceedings of JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence, in: Lecture Notes in Artificial Intelligence, vol. 1919, Springer-Verlag, Berlin, 2000, pp. 224–238.
[41] M. Yannakakis, Algorithms for acyclic database schemes, in: C. Zaniolo, C. Delobel (Eds.), Proc. Internat. Conference on Very Large Data Bases (VLDB'81), 1981, pp. 82–94.

Extremal Behaviour in Multiagent Contract Negotiation

Journal of Artificial Intelligence Research 23 (2005) 41–78 Submitted 07/04; published 01/05

Research Note

Extremal Behaviour in Multiagent Contract Negotiation

Paul E. Dunne PED@CSC.LIV.AC.UK
Department of Computer Science
The University of Liverpool, Liverpool, UK

Abstract

We examine properties of a model of resource allocation in which several agents exchange resources in order to optimise their individual holdings. The schemes discussed relate to well-known negotiation protocols proposed in earlier work and we consider a number of alternative notions of "rationality" covering both quantitative measures, e.g. cooperative and individual rationality and more qualitative forms, e.g. Pigou-Dalton transfers. While it is known that imposing particular rationality and structural restrictions may result in some reallocations of the resource set becoming unrealisable, in this paper we address the issue of the number of restricted rational deals that may be required to implement a particular reallocation when it *is* possible to do so. We construct examples showing that this number may be exponential (in the number of resources m), even when all of the agent utility functions are monotonic. We further show that k agents may achieve in a single deal a reallocation requiring exponentially many rational deals if at most $k-1$ agents can participate, this same reallocation being unrealisable by any sequences of rational deals in which at most $k-2$ agents are involved.

1. Introduction

Mechanisms for negotiating allocation of resources within a group of agents form an important body of work within the study of multiagent systems. Typical abstract models derive from game-theoretic perspectives in economics and among the issues that have been addressed are strategies that agents use to obtain a particular subset of the resources available, e.g. (Kraus, 2001; Rosenschein & Zlotkin, 1994; Sandholm, 1999), and protocols by which the process of settling upon some allocation of resources among the agents involved is agreed, e.g. (Dignum & Greaves, 2000; Dunne, 2003; Dunne & McBurney, 2003; McBurney et al., 2002).

The setting we are concerned with is encapsulated in the following definition.

Definition 1 *A* resource allocation setting *is defined by a triple* $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ *where*

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\} \quad ; \quad \mathcal{R} = \{r_1, r_2, \ldots, r_m\}$$

are, respectively, a set of (at least two) agents and a collection of (non-shareable) resources. A utility function, *u, is a mapping from subsets of $\mathcal{R}$ to rational values. Each agent $A_i \in \mathcal{A}$ has associated with it a particular utility function u_i, so that $\mathcal{U}$ is $\langle u_1, u_2, \ldots, u_n \rangle$. An* allocation *$P$ of $\mathcal{R}$ to $\mathcal{A}$ is a partition $\langle P_1, P_2, \ldots, P_n \rangle$ of $\mathcal{R}$. The value $u_i(P_i)$ is called the* utility *of the resources assigned to A_i. A utility function, u, is* monotone *if whenever $S \subseteq T$ it holds that $u(S) \leq u(T)$, i.e. the value assigned by u to any set of resources, T, is never less than the value u attaches to any subset, S of T.*

Two major applications in which the abstract view of Definition 1 has been exploited are e-commerce and distributed task realisation. In the first $\mathcal{R}$ represents some collection of commodities offered for sale and individual agents seek to acquire a subset of these, the "value" an agent attaches to a specific set being described by that agent's utility function. In task planning, the "resource" set describes a collection of sub-tasks to be performed in order to realise some complex task, e.g. the "complex task" may be to transport goods from a central warehouse to some set of cities. In this example $\mathcal{R}$ describes the locations to which goods must be dispatched and a given allocation defines those places to which an agent must arrange deliveries. The utility functions in such cases model the cost an agent associates with carrying out its alloted sub-tasks.

Within the very general context of Definition 1, a number of issues arise stemming from the observation that it is unlikely that some *initial allocation* will be seen as satisfactory either with respect to the views of all agents in the system or with respect to divers global considerations. Thus, by proposing changes to the initial assignment individual agents seek to obtain a "better" allocation. This scenario raises two immediate questions: how to evaluate a given partition and thus have a basis for forming improved or optimal allocations; and, the issue underlying the main results of this paper, what restrictions should be imposed on the form that proposed deals may take.

We shall subsequently review some of the more widely studied approaches to defining conditions under which some allocations are seen as "better" than others. For the purposes of this introduction we simply observe that such criteria may be either *quantitative* or *qualitative* in nature. As an example of the former we have the approach wherein the "value" of an allocation P is simply the sum of the values given by the agents' utility functions to the subsets of $\mathcal{R}$ they have been apportioned within P, i.e. $\sum_{i=1}^{n} u_i(P_i)$: this is the so-called *utilitarian social welfare*, which to avoid repetition we will denote by $\sigma_u(P)$. A natural aim for agents within a commodity trading context is to seek an allocation under which σ_u is *maximised*. One example of a *qualitative* criterion is "*envy freeness*": informally, an allocation, P, is envy-free if no agent assigns greater utility to the resource set (P_j) held by another agent than it does with respect to the resource set (P_i) it has actually been allocated, i.e. for each distinct pair $\langle i, j\rangle$, $u_i(P_i) \geq u_i(P_j)$.

In very general terms there are two approaches that have been considered in treating the question of how a finite collection of resources might be distributed among a set of agents in order to optimise some criterion of interest: "contract-net" based methods, e.g. (Dunne et al., 2003; Endriss et al., 2003; Endriss & Maudet, 2004b; Sandholm, 1998, 1999) deriving from the work of Smith (1980); and "combinatorial auctions", e.g. (Parkes & Ungar, 2000a, 2000b; Sandholm et al., 2001; Sandholm, 2002; Sandholm & Suri, 2003; Tennenholz, 2000; Yokoo et al., 2004, amongst others). The significant difference between these is in the extent to which a centralized controlling agent determines the eventual distribution of resources among agents.

One may view the strategy underlying combinatorial auctions as investing the computational effort into a "pre-processing" stage following which a given allocation is determined. Thus a controlling agent (the "auctioneer") is supplied with a set of *bids* – pairs $\langle S_j, p_j\rangle$ wherein S_j is some subset of the available resources and p_j the price agent A_j is prepared to pay in order to acquire S_j. The problem faced by the auctioneer is to decide which bids

to accept in order to maximise the overall profit subject to the constraint that each item can be obtained by at most one agent.

What we shall refer to as "contract-net schemes" typically eschew the precomputation stage and subordination to a controlling arbiter employed in auction mechanisms, seeking instead to realise a suitable allocation by an agreed sequence of deals. The *contract-net* (in its most general instantiation) for scenarios of m resources distributed among n agents is the complete directed graph with n^m vertices (each of which is associated with a distinct allocation). In this way a possible deal $\langle P, Q\rangle$ is represented as an edge directed from the vertex labelled with P to that labelled Q. Viewed thus, identifying a sequence of deals can be interpreted as a search process which, in principle, individual agents may conduct in an autonomous fashion.

Centralized schemes can be effective in contexts where the participants cooperate (in the sense of accepting the auctioneer's arbritration). In environments within which agents are highly self-interested to the extent that their aims conflict with the auction process or in which there is a high degree of "uncertainty" about the outcome, in working towards a final allocation, the agents involved may only be prepared to proceed "cautiously": that is, an agent will only accept a proposed reallocation if satisfied that such would result in an *immediate* improvement from its own perspective. In such cases, the process of moving from the initial allocation, P_{init}, to the eventual reallocation P_{fin} is by a sequence of local *rational* deals, e.g. an agent might refuse to accept deals which reduced σ_u because of the possibility that it suffers an uncompensated loss in utility. A key issue here is the following: if the deal protocol allows only moves in which at each stage some agent A_j offers a single resource to another agent A_j then the rational reallocation $\langle P_{init}, P_{fin}\rangle$ can *always* be implemented; if, however, every single move must be "rational" then $\langle P_{init}, P_{fin}\rangle$ may not be realisable.

We may, informally, regard the view of such agents as "myopic", in the sense that they are unwilling to accept a "short-term loss" (a deal $\langle P, Q\rangle$ under they might incur a loss of utility) despite the prospect of a "long-term gain" (assuming $\sigma_u(P_{fin}) > \sigma_u(P_{init})$ holds).

There are a number of reasons why an agent may adopt such views, e.g. consider the following simple protocol for agreeing a reallocation.

> A reallocation of resources is agreed over a sequence of stages, each of which involves communication between two agents, A_i and A_j. This communication consists of A_i issuing a proposal to A_j of the form (buy, r, p), offering to purchase r from A_j for a payment of p; or $(sell, r, p)$, offering to transfer r to A_j in return for a payment p. The response from A_j is simply *accept* (following which the deal is implemented) or *reject*.

This, of course, is a very simple negotiation structure, however consider its operation within a two agent setting in which one agent, A_1 say, wishes to bring about an allocation P_{fin} (and thus can devise a plan – sequence of deals – to realise this from an initial allocation P_{init}) while the other agent, A_2, does *not* know P_{fin}. In addition, assume that A_1 is the only agent that makes proposals and that a final allocation is fixed either when A_1 is "satisfied" or as soon as A_2 *rejects* any offer.

While A_2 *could* be better off if P_{fin} is realised, it may be the case that the only proposals A_2 will accept are those under which it does not lose, e.g. some agents may be sceptical about the *bona fides* of others and will accept only deals from which they can perceive an

immediate benefit. There are several reasons why an agent may embrace such attitudes within the schema outlined: once a deal has been implemented A_2 may lose utility but no further proposals are made by A_1 so that the loss is "permanent". We note that even if we enrich the basic protocol so that A_1 can describe P_{fin}, A_2 may still reject offers under which it suffers a loss, since it is unwilling to rely on the subsequent deals that would ameliorate its loss actually being proposed. Although the position taken by A_2 in the setting just described may appear unduly cautious, we would claim that it does reflect "real" behaviour in certain contexts. Outside the arena of automated allocation and negotiation in multiagent systems, there are many examples of actions by individuals where promised long-term gains are insufficient to engender the acceptance of short term loss. Consider "chain letter" schemes (or their more subtle manifestation as "pyramid selling" enterprises): such have a natural lifetime bounded by the size of the population in which they circulate, but may break down before this is reached. Faced with a request to "send $10 to the five names at the head of the list and forward the letter to ten others after adding your name" despite the possibility of significant gain after a temporary loss of $50, to ignore such blandishments is not seen as overly sceptical and cautious: there may be reluctance to accept that one will eventually receive sufficient recompense in return and suspicion that the name order has been manipulated.

In summary, we can identify two important influences that lead to contexts in which agents prefer to move towards a reallocation via a sequence of "rational" deals. Firstly, the agents are self-interested but operating in an unstable environment, e.g. in the "chain letter" setting, an agent cannot reliably predict the exact point at which the chain will fail. The second factor is that computational restrictions may limit the decisions an individual agent can make about whether or not to accept a proposed deal. For example in settings where all deals involve one resource at a time, A_2 may reject a proposal to accept some resource, r, since r is only "useful" following a further sequence of deals: if this number of further deals is "small" then A_2 could decide to accept the proposed deal since it has sufficient computational power to determine that there is a context in which r is of value; if this number is "large" however, then A_2 may lack sufficient power to scan the search space of future possibilities that would allow it to accept r. Notice that in the extreme case, A_2 makes its decision solely on whether r is of immediate use, i.e. A_2 is myopic. A more powerful A_2 may be able to consider whether r is useful should up to k further deals take place: in this case, A_2 could still refuse to accept r since, although of use, A_2 cannot determine this with a bounded look ahead.

In total for the scenario we have described, if A_1 wishes to bring about an allocation P_{fin} then faced with the view adopted by A_2 and the limitations imposed by the deal protocol, the only "effective plan" that A_1 could adopt is to find a sequence of *rational* deals to propose to A_2.

Our aim in this article is to show that combining "structural" restrictions (e.g. only one resource at a time is involved in a local reallocation) with rationality restrictions can result in settings in which any sequence to realise a reallocation $\langle P, Q\rangle$ must involve exponentially many (in $|\mathcal{R}|$) separate stages. We refine these ideas in the next sub-section.

1.1 Preliminary Definitions

To begin, we first formalise the concepts of *deal* and *contract path*.

Definition 2 *Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ be a resource allocation setting. A* deal *is a pair $\langle P, Q \rangle$ where $P = \langle P_1, \ldots, P_n \rangle$ and $Q = \langle Q_1, \ldots, Q_n \rangle$ are distinct partitions of $\mathcal{R}$. The effect of implementing the deal $\langle P, Q \rangle$ is that the allocation of resources specified by P is replaced with that specified by Q. Following the notation of (Endriss & Maudet, 2004b) for a deal $\delta = \langle P, Q \rangle$, we use $\mathcal{A}^\delta$ to indicate the subset of $\mathcal{A}$ involved, i.e. $A_k \in \mathcal{A}^\delta$ if and only if $P_k \neq Q_k$.*

Let $\delta = \langle P, Q \rangle$ be a deal. A contract path realising δ *is a sequence of allocations*

$$\Delta = \langle P^{(1)},\ P^{(2)}\ , \ldots,\ P^{(t-1)},\ P^{(t)} \rangle$$

in which $P = P^{(1)}$ and $P^{(t)} = Q$. The length *of Δ, denoted $|\Delta|$ is $t-1$, i.e. the number of* deals *in Δ.*

There are two methods which we can use to reduce the number of deals that a single agent may have to consider in seeking to move from some allocation to another, thereby avoiding the need to choose from exponentially many alternatives: *structural* and *rationality* constraints. Structural constraints limit the permitted deals to those which bound the number of resources and/or the number of agents involved, but take no consideration of the view any agent may have as to whether its allocation has improved. In contrast, rationality constraints restrict deals $\langle P, Q \rangle$ to those in which Q "improves" upon P according to particular criteria. In this article we consider two classes of structural constraint: O-contracts, defined and considered in (Sandholm, 1998), and what we shall refer to as $M(k)$-contracts.

Definition 3 *Let $\delta = \langle P, Q \rangle$ be a deal involving a reallocation of $\mathcal{R}$ among $\mathcal{A}$.*

a. δ is a one contract *(O-contract) if*

O1. $\mathcal{A}^\delta = \{i, j\}$.

O2. There is a unique resource $r \in P_i \cup P_j$ for which $Q_i = P_i \cup \{r\}$ and $Q_j = P_j \setminus \{r\}$ (with $r \in P_j$) or $Q_j = P_j \cup \{r\}$ and $Q_i = P_i \setminus \{r\}$ (with $r \in P_i$)

b. For a value $k \geq 2$, the deal $\delta = \langle P, Q \rangle$ is an $M(k)$-contract *if $2 \leq |\mathcal{A}^\delta| \leq k$ and $\cup_{i \in \mathcal{A}^\delta} Q_i = \cup_{i \in \mathcal{A}^\delta} P_i$.*

Thus, O-contracts involve the transfer of *exactly one* resource from a particular agent to another, resulting in the number of deals compatible with any given allocation being exactly $(n-1)m$: each of the m resources can be reassigned from its current owner to any of the other $n-1$ agents.

Rationality constraints arise in a number of different ways. For example, from the standpoint of an individual agent A_i a given deal $\langle P, Q \rangle$ may have three different outcomes: $u_i(P_i) < u_i(Q_i)$, i.e. A_i values the allocation Q_i as superior to P_i; $u_i(P_i) = u_i(Q_i)$, i.e. A_i is indifferent between P_i and Q_i; and $u_i(P_i) > u_i(Q_i)$, i.e. A_i is worse off after the deal. When global optima such as utilitarian social welfare are to be maximised, there is the question of what incentive there is for any agent to accept a deal $\langle P, Q \rangle$ under which it

is left with a less valuable resource holding. The standard approach to this latter question is to introduce the notion of a *pay-off* function, i.e. in order for A_i to accept a deal under which it suffers a reduction in utility, A_i receives some payment sufficient to compensate for its loss. Of course such compensation must be made by other agents in the system who in providing it do not wish to pay in excess of any gain. In defining notions of pay-off the interpretation is that in any transaction each agent A_i makes a payment, π_i: if $\pi_i < 0$ then A_i is given $-\pi_i$ in return for accepting a deal; if $\pi_i > 0$ then A_i contributes π_i to the amount to be distributed among those agents whose pay-off is negative.

This notion of "sensible transfer" is captured by the concept of *individual rationality*, and is often defined in terms of an appropriate pay-off vector existing. It is not difficult, however, to show that such definitions are equivalent to the following.

Definition 4 *A deal $\langle P, Q\rangle$ is* individually rational *(IR) if and only if $\sigma_u(Q) > \sigma_u(P)$.*

We shall consider alternative bases for rationality constraints later: these are primarily of interest within so-called *money free* settings (so that compensatory payment for a loss in utility is not an option).

The central issue of interest in this paper concerns the properties of the contract-net graph when the allowed deals must satisfy both a structural *and* a rationality constraint. Thus, if we consider arbitrary predicates Φ on deals $\langle P, Q\rangle$ – where the cases of interest are Φ combining a structural and rationality condition – we have,

Definition 5 *For Φ a predicate over distinct pairs of allocations, a contract path*

$$\langle P^{(1)},\ P^{(2)}\ ,\ldots,\ P^{(t-1)},\ P^{(t)}\rangle$$

realising $\langle P, Q\rangle$ is a Φ-path *if for each $1 \leq i < t$, $\langle P^{(i)}, P^{(i+1)}\rangle$ is a* Φ-deal*, that is $\Phi(P^{(i)}, P^{(i+1)})$ holds. We say that Φ is* complete *if any deal δ may be realised by a Φ-path. We, further, say that Φ is* complete with respect to Ψ-deals *(where Ψ is a predicate over distinct pairs of allocations) if any deal δ for which $\Psi(\delta)$ holds may be realised by a Φ-path.*

The main interest in earlier studies of these ideas has been in areas such as identifying necessary and/or sufficient conditions on deals to be complete with respect to particular criteria, e.g. (Sandholm, 1998); and in establishing "convergence" and termination properties, e.g. Endriss et al. (2003), Endriss and Maudet (2004b) consider deal types, Φ, such that every maximal[1] Φ-path ends in a Pareto optimal allocation, i.e. one in which any reallocation under which some agent improves its utility will lead to another agent suffering a loss. Sandholm (1998) examines how restrictions e.g. with $\Phi(P, Q) = \top$ if and only if $\langle P, Q\rangle$ is an O-contract, may affect the existence of contract paths to realise deals. Of particular interest, from the viewpoint of heuristics for exploring the contract-net graph, are cases where $\Phi(P, Q) = \top$ if and only if the deal $\langle P, Q\rangle$ is individually rational. For the case of O-contracts the following are known:

Theorem 1

a. O-contracts are complete.

1. "Maximal" in the sense that if $\langle P^{(1)}, \ldots, P^{(t)}\rangle$ is such a path, then for every allocation, Q, $\Phi(P^{(t)}, Q)$ does not hold.

b. IR O-contracts are not *complete with respect to IR deals.*

In the consideration of algorithmic and complexity issues presented in (Dunne et al., 2003) one difficulty with attempting to formulate reallocation plans by rational O-contracts is already apparent, that is:

Theorem 2 *Even in the case $n = 2$ and with monotone utility functions the problem of deciding if an IR O-contract path exists to realise the IR deal $\langle P, Q\rangle$ is* NP*–hard.*

Thus deciding if any rational plan is possible is already computationally hard. In this article we demonstrate that, even if an appropriate rational plan exists, in extreme cases, there may be significant problems: the number of deals required could be exponential in the number of resources, so affecting both the time it will take for the schema outlined to conclude and the space that an agent will have to dedicate to storing it. Thus in his proof of Theorem 1 (b), Sandholm observes that when an IR O-contract path exists for a given IR deal, it may be the case that its length exceeds m, i.e. some agent passes a resource to another and then accepts the same resource at a later stage.

The typical form of the results that we derive can be summarised as:

> For Φ a structural constraint (O-contract or $M(k)$-contract) and Ψ a rationality constraint, e.g. $\Psi(P, Q)$ holds if $\langle P, Q\rangle$ is individually rational, there are resource allocation settings $\langle \mathcal{A}_n, \mathcal{R}_m, \mathcal{U}\rangle$ in which there is a deal $\langle P, Q\rangle$ satisfying all of the following.
>
> a. $\langle P, Q\rangle$ is a Ψ-deal.
> b. $\langle P, Q\rangle$ can be realised by a contract path on which every deal satisfies the structural constraint Φ *and* the rationality constraint Ψ.
> c. Every such contract path has length at least $g(m)$.

For example, we show that there are instances for which the shortest IR O-contract path has length exponential in m.[2] In the next section we will be interested in lower bounds on the values of the following functions: we introduce these in general terms to avoid unnecessary subsequent repetition.

Definition 6 *Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ be a resource allocation setting. Additionally let Φ and Ψ be two predicates on deals. For a deal $\delta = \langle P, Q\rangle$ the partial function $L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi)$ is the length of the* shortest *Φ-contract path realising $\langle P, Q\rangle$ if such a path exists (and is undefined if no such path is possible). The partial function $L^{\max}(\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi, \Psi)$ is*

$$L^{\max}(\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi, \Psi) \;=\; \max_{\Psi\text{-deals } \delta} L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi)$$

Finally, the partial function $\rho^{\max}(n, m, \Phi, \Psi)$ is

$$\rho^{\max}(n, m, \Phi, \Psi) \;=\; \max_{\mathcal{U}=\langle u_1, u_2, \ldots, u_n\rangle} L^{\max}(\langle \mathcal{A}_n, \mathcal{R}_m, \mathcal{U}\rangle, \Phi, \Psi)$$

where consideration is restricted to those Ψ-deals $\delta = \langle P, Q\rangle$ for which a realising Φ-path exists.

2. Sandholm (1998) gives an upper bound on the length of such paths which is also exponential in m, but does not explicitly state any lower bound other than that already referred to.

The three measures, L^{opt}, $L^{\max}$ and $\rho^{\max}$ distinguish different aspects regarding the length of contract-paths. The function L^{opt} is concerned with Φ-paths realising a single deal $\langle P, Q\rangle$ in a given resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$: the property of interest being the number of deals in the *shortest*, i.e. optimal length, Φ-path. We stress that L^{opt} is a *partial* function whose value is undefined in the event that $\langle P, Q\rangle$ cannot be realised by a Φ-path in the setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$. The function $L^{\max}$ is defined in terms of L^{opt}, again in the context of a specific resource allocation setting. The behaviour of interest for $L^{\max}$, however, is not simply the length of Φ-paths realising a specific $\langle P, Q\rangle$ but the "worst-case" value of L^{opt} for deals which are Ψ-deals. We note the qualification that $L^{\max}$ is defined only for Ψ-deals that *are* capable of being realised by Φ-paths, and thus do not consider cases for which no appropriate contract path exists. Thus, if it should be the case that no Ψ-deal in the setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ can be realised by a Φ-path then the value $L^{\max}(\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi, \Psi)$ is undefined, i.e. $L^{\max}$ is also a partial function. We may interpret any *upper* bound on $L^{\max}$ in the following terms: if $L^{\max}(\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi, \Psi) \leq K$ then any Ψ-deal *for which a Φ-path exists* can be realised by a Φ-path of length at most K.

Our main interest will centre on $\rho^{\max}$ which is concerned with the behaviour of $L^{\max}$ as a function of n and m and ranges over *all* n-tuples of utility functions $\langle u : 2^{\mathcal{R}} \rightarrow \mathbf{Q}\rangle^n$. Our approach to obtaining lower bounds for this function is *constructive*, i.e. for each $\langle \Phi, \Psi\rangle$ that is considered, we show how the utility functions $\mathcal{U}$ may be defined in a setting with m resources so as to yield a lower bound on $\rho^{\max}(n, m, \Phi, \Psi)$. In contrast to the measures L^{opt} and $L^{\max}$, the function $\rho^{\max}$ is not described in terms of a single fixed resource allocation setting. It is, however, still a *partial* function: depending on $\langle n, m, \Phi, \Psi\rangle$ it may be the case that in *every* n agent, m resource allocation setting, regardless of which choice of utility functions is made, there is no Ψ-deal, $\langle P, Q\rangle$ capable of being realised by Φ-path, and for such cases the value of $\rho^{\max}(n, m, \Phi, \Psi)$ will be undefined.[3]

It is noted, at this point, that the definition of $\rho^{\max}$ allows *arbitrary* utility functions to be employed in constructing "worst-case" instances. While this is reasonable in terms of general lower bound results, as will be apparent from the given constructions the utility functions actually employed are highly artificial (and unlikely to feature in "real" application settings). We shall attempt to address this objection by further considering bounds on the following variant of $\rho^{\max}$:

$$\rho^{\max}_{\text{mono}}(n, m, \Phi, \Psi) = \max_{\mathcal{U}=\langle u_1, u_2, \ldots, u_n\rangle \,:\, \text{each } u_i \text{ is monotone}} L^{\max}(\langle \mathcal{A}_n, \mathcal{R}_m, \mathcal{U}\rangle, \Phi, \Psi)$$

Thus, $\rho^{\max}_{\text{mono}}$ deals with resource allocation settings within which all of the utility functions must satisfy a monotonicity constraint.

The main results of this article are presented in the next sections. We consider two general classes of contract path: O-contract paths under various rationality conditions in

3. In recognising the *possibility* that $\rho^{\max}(n, m, \Phi, \Psi)$ could be undefined, we are *not* claiming that such behaviour arises with any of the instantiations of $\langle \Phi, \Psi\rangle$ considered subsequently: in fact it will be clear from the constructions that, denoting by $\rho^{\max}_{\Phi,\Psi}(n, m)$ the function $\rho^{\max}(n, m, \Phi, \Psi)$ for a fixed instantiation of $\langle \Phi, \Psi\rangle$, with the restricted deal types and rationality conditions examined, the function $\rho^{\max}_{\Phi,\Psi}(n, m)$ is a *total* function. Whether it is possible to formulate "sensible" choices of $\langle \Phi, \Psi\rangle$ with which $\rho^{\max}_{\Phi,\Psi}(n, m)$ is undefined for some values of $\langle n, m\rangle$ (and, if so, demonstrating examples of such) is, primarily, only a question of combinatorial interest, whose development is not central to the concerns of the current article.

Section 2; and, similarly, $M(k)$-contract paths for arbitrary values of $k \geq 2$ in Section 3. Our results are concerned with the construction of resource allocation settings $\langle \mathcal{A}, \mathcal{R}_m, \mathcal{U} \rangle$ for which given some rationality requirement, e.g. that deals be individually rational, there is some deal $\langle P, Q \rangle$ that satisfies the rationality condition, can be realised by a rational O-contract path (respectively, $M(k)$-contract path), but with the number of deals required by such paths being exponential in m. We additionally obtain slightly weaker (but still exponential) lower bounds for rational O-contract paths within settings of monotone utility functions, i.e. for the measure $\rho^{\max}_{\text{mono}}$, outlining how similar results may be derived for $M(k)$-contract paths.

In the resource allocation settings constructed for demonstrating these properties with $M(k)$-contract paths, the constructed deal $\langle P, Q \rangle$ is realisable with a *single* $M(k+1)$-contract but unrealisable by any *rational* $M(k-1)$-contract path. We discuss related work, in particular the recent study of (Endriss & Maudet, 2004a) that addresses similar issues to those considered in the present article, in Section 4. Conclusions and some directions for further work are presented in the final section.

2. Lower Bounds on Path Length – O-contracts

In this section we consider the issue of contract path length when the structural restriction requires individual deals to be O-contracts. We first give an overview of the construction method, with the following subsections analysing the cases of unrestricted utility functions and, subsequently, *monotone* utility functions.

2.1 Overview

The strategy employed in proving our results involves two parts: for a given class of restricted contract paths we proceed as follows in obtaining lower bounds on $\rho^{\max}(n, m, \Phi, \Psi)$.

a. For the contract-net graph partitioning m resources among n agents, construct a path, $\Delta_m = \langle P^{(1)}, P^{(2)}, \ldots, P^{(t)} \rangle$ realising a deal $\langle P^{(1)}, P^{(t)} \rangle$. For the *structural* constraint, Φ' influencing Φ it is then proved that:

   a1. The contract path Δ_m *is* a Φ'-path, i.e. for each $1 \leq i < t$, the deal $\langle P^{(i)}, P^{(i+1)} \rangle$ satisfies the structural constraint Φ'.

   a2. For any pair of allocations $P^{(i)}$ and $P^{(i+j)}$ occurring in Δ_m, if $j \geq 2$ then the deal $\langle P^{(i)}, P^{(i+j)} \rangle$ is *not* a Φ'-deal.

   Thus (a1) ensures that Δ_m is a suitable contract path, while (a2) will guarantee that there is exactly one allocation, $P^{(i+1)}$, that can be reached *within* Δ_m from any given allocation $P^{(i)}$ in Δ_m by means of a Φ'-deal.

b. Define utility functions $\mathcal{U}_n = \langle u_1, \ldots, u_n \rangle$ with the following properties

   b1. The deal $\langle P^{(1)}, P^{(t)} \rangle$ is a Ψ-deal.

   b2. For the *rationality* constraint, Φ'' influencing Φ, every deal $\langle P^{(i)}, P^{(i+1)} \rangle$ is a Φ''-deal.

b3. For every allocation $P^{(i)}$ in the contract path Δ and every allocation Q other than $P^{(i+1)}$ the deal $\langle P^{(i)}, Q\rangle$ is *not* a Φ-deal, i.e. it violates either the stuctural constraint Φ' or the rationality constraint Φ''.

Thus, (a1) and (b2) ensure that $\langle P^{(1)}, P^{(t)}\rangle$ has a defined value with respect to the function L^{opt} for the Ψ-deal $\langle P^{(1)}, P^{(t)}\rangle$, i.e. a Φ-path realising the deal is possible. The properties given by (a2) and (b3) indicate that (within the constructed resource allocation setting) the path Δ_m is the *unique* Φ-path realising $\langle P^{(1)}, P^{(t)}\rangle$. It follows that $t-1$, the length of this path, gives a *lower bound* on the value of $L^{\max}$ and hence a lower bound on $\rho^{\max}(n, m, \Phi, \Psi)$.

Before continuing it will be useful to fix some notational details.

We use $\mathcal{H}_m$ to denote the m-dimensional hypercube. Interpreted as a directed graph, $\mathcal{H}_m$ has 2^m vertices each of which is identified with a distinct m-bit label. Using $\alpha = a_1 a_2 \dots a_m$ to denote an arbitrary such label, the edges of $\mathcal{H}_m$ are formed by

$$\{\ \langle \alpha, \beta\rangle : \alpha \text{ and } \beta \text{ differ in } \textit{exactly one} \text{ bit position}\}$$

We identify m-bit labels $\alpha = a_1 a_2 \dots a_m$ with subsets S^α of $\mathcal{R}_m$, via $r_i \in S^\alpha$ if and only if $a_i = 1$. Similarly, any subset S of $\mathcal{R}$ can be described by a binary word, $\beta(S)$, of length m, i.e. $\beta(S) = b_1 b_2 \dots b_m$ with $b_i = 1$ if and only if $r_i \in S$. For a label α we use $|\alpha|$ to denote the number of bits with value 1, so that $|\alpha|$ is the size of the subset S^α. If α and β are m-bit labels, then $\alpha\beta$ is a $2m$-bit label, so that if $\mathcal{R}_m$ and $\mathcal{T}_m$ are disjoint sets, then $\alpha\beta$ describes the union of the subset S^α of $\mathcal{R}_m$ with the subset S^β of $\mathcal{T}_m$. Finally if $\alpha = a_1 a_2 \dots a_m$ is an m-bit label then $\overline{\alpha}$ denotes the label formed by changing all 0 values in α to 1 and *vice versa*. In this way, if S^α is the subset of $\mathcal{R}_m$ described by α then $\overline{\alpha}$ describes the set $\mathcal{R}_m \setminus S^\alpha$. To avoid an excess of superscripts we will, where no ambiguity arises, use α both to denote the m-bit label and the subset of $\mathcal{R}_m$ described by it, e.g. we write $\alpha \subset \beta$ rather than $S^\alpha \subset S^\beta$.

For $n = 2$ the contract-net graph induced by O-contracts can be viewed as the m-dimensional hypercube $\mathcal{H}_m$: the m-bit label, α associated with a vertex of $\mathcal{H}_m$ describing the allocation $\langle \alpha, \overline{\alpha}\rangle$ to $\langle A_1, A_2\rangle$. In this way the set of IR O-contracts define a subgraph, $\mathcal{G}_m$ of $\mathcal{H}_m$ with any directed path from $\beta(P)$ to $\beta(Q)$ in $\mathcal{G}_m$ corresponding to a possible IR O-contract path from the allocation $\langle P, \mathcal{R} \setminus P\rangle$ to the allocation $\langle Q, \mathcal{R} \setminus Q\rangle$.

2.2 O-contract Paths – Unrestricted Utility Functions

Our first result clarifies one issue in the presentation of (Sandholm, 1998, Proposition 2): in this an upper bound that is exponential in m is proved on the length of IR O-contract paths, i.e. in terms of our notation, (Sandholm, 1998, Proposition 2) establishes an upper bound on $\rho^{\max}(n, m, \Phi, \Psi)$. We now prove a similar order *lower* bound.

Theorem 3 *Let $\Phi(P, Q)$ be the predicate which holds whenever $\langle P, Q\rangle$ is an IR O-contract and $\Psi(P, Q)$ that which holds whenever $\langle P, Q\rangle$ is IR. For $m \geq 7$*

$$\rho^{\max}(2, m, \Phi, \Psi)\ \geq \left(\frac{77}{256}\right)\ 2^m\ -\ 2$$

Proof. Consider a path $\mathcal{C} = \langle \alpha_1, \alpha_2, \ldots, \alpha_t \rangle$ in $\mathcal{H}_m$, with the following property[4]

$$\forall\ 1 \leq i < j \leq t\ (j \geq i+2)\ \Rightarrow\ (\alpha_i \text{ and } \alpha_j \text{ differ in at least 2 positions}) \qquad \text{(SC)}$$

e.g. if $m = 4$ then

$$\emptyset, \{r_1\},\ \{r_1, r_3\}, \{r_1, r_2, r_3\}, \{r_2, r_3\}, \{r_2, r_3, r_4\}, \{r_2, r_4\}, \{r_1, r_2, r_4\}$$

is such a path as it corresponds to the sequence $\langle 0000, 1000, 1010, 1110, 0110, 0111, 0101, 1101 \rangle$.

Choose $\mathcal{C}^{(m)}$ to be a *longest* such path with this property that could be formed in $\mathcal{H}_m$, letting $\Delta_m = \langle P^{(1)}, P^{(2)}, \ldots, P^{(t)} \rangle$ be the sequence of allocations with $P^{(i)} = \langle \alpha_i, \overline{\alpha_i} \rangle$. We now define the utility functions u_1 and u_2 so that for $\gamma \subseteq \mathcal{R}_m$,

$$u_1(\gamma) + u_2(\overline{\gamma}) = \begin{cases} k & \text{if} \quad \gamma = \alpha_k \\ 0 & \text{if} \quad \gamma \notin \{\alpha_1, \alpha_2, \ldots, \alpha_t\} \end{cases}$$

With this choice, the contract path Δ_m describes the *unique* IR O-contract path realising the IR deal $\langle P^{(1)}, P^{(t)} \rangle$: that Δ_m is an IR O-contract path is immediate, since

$$\sigma_u(P^{(i+1)}) = i + 1 > i = \sigma_u(P^{(i)})$$

That it is unique follows from the fact that for all $1 \leq i \leq t$ and $i + 2 \leq j \leq t$, the deal $\langle P^{(i)}, P^{(j)} \rangle$ is not an O-contract (hence there are no "short-cuts" possible), and for each $P^{(i)}$ there is exactly one IR O-contract that can follow it, i.e. $P^{(i+1)}$.[5]

From the preceding argument it follows that any lower bound on the length of $\mathcal{C}^{(m)}$, i.e. a sequence satisfying the condition (SC), is a lower bound on $\rho^{\max}(2, m, \Phi, \Psi)$. These paths in $\mathcal{H}_m$ were originally studied by Kautz (1958) in the context of coding theory and the lower bound on their length of $(77/256)2^m - 2$ established in (Abbott & Katchalski, 1991). □

Example 1 *Using the path*

$$\begin{aligned} \mathcal{C}^{(4)} &= \langle 0000, 1000, 1010, 1110, 0110, 0111, 0101, 1101 \rangle \\ &= \langle \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8 \rangle \end{aligned}$$

in the resource allocation setting $\langle \{a_1, a_2\}, \{r_1, r_2, r_3, r_4\}, \langle u_1, u_2 \rangle \rangle$*, if the utility functions are specified as in Table 1 below then* $\sigma_u(\langle \alpha_1, \overline{\alpha}_1 \rangle) = 1$ *and* $\sigma_u(\langle \alpha_8, \overline{\alpha}_8 \rangle) = 8$*. Furthermore,* $\mathcal{C}^{(4)}$ *describes the unique IR O-contract path realising the reallocation* $\langle \langle \alpha_1, \overline{\alpha}_1 \rangle,\ \langle \alpha_8, \overline{\alpha}_8 \rangle \rangle$

There are a number of alternative formulations of "rationality" which can also be considered. For example

Definition 7 *Let* $\delta = \langle P, Q \rangle$ *be a deal.*

4. This defines the so-called "snake-in-the-box" codes introduced in (Kautz, 1958).
5. In our example with $m = 4$, the sequence $\langle 0000, 1000, 1001, 1101 \rangle$, although defining an O-contract path gives rise to a deal which is not IR, namely that corresponding to $\langle 1000, 1001 \rangle$.

S	$\mathcal{R} \setminus S$	$u_1(S)$	$u_2(\mathcal{R} \setminus S)$	σ_u		S	$\mathcal{R} \setminus S$	$u_1(S)$	$u_2(\mathcal{R} \setminus S)$	σ_u	
0000	1111	1	0	1	α_1	1000	0111	1	1	2	α_2
0001	1110	0	0	0		1001	0110	0	0	0	
0010	1101	0	0	0		1010	0101	2	1	3	α_3
0011	1100	0	0	0		1011	0100	0	0	0	
0100	1011	0	0	0		1100	0011	0	0	0	
0101	1010	4	3	7	α_7	1101	0010	4	4	8	α_8
0110	1001	3	2	5	α_5	1110	0001	2	2	4	α_4
0111	1000	3	3	6	α_6	1111	0000	0	0	0	

Table 1: Utility function definitions for $m = 4$ example.

a. δ *is* cooperatively rational *if for every agent,* A_i, $u_i(Q_i) \geq u_i(P_i)$ *and there is at least one agent,* A_j, *for whom* $u_j(Q_j) > u_j(P_j)$.

b. δ *is* equitable *if* $\min_{i \in \mathcal{A}^\delta} u_i(Q_i) > \min_{i \in \mathcal{A}^\delta} u_i(P_i)$.

c. δ *is a* Pigou-Dalton deal *if* $\mathcal{A}^\delta = \{i, j\}$, $u_i(P_i) + u_j(P_j) = u_i(Q_i) + u_j(Q_j)$ *and* $|u_i(Q_i) - u_j(Q_j)| < |u_i(P_i) - u_j(P_j)|$ *(where* $|\ldots|$ *is absolute value).*

There are a number of views we can take concerning the rationality conditions given in Definition 7. One shared feature is that, unlike the concept of individual rationality for which some provision to compensate agents who suffer a loss in utility is needed, i.e. individual rationality presumes a "money-based" system, the forms defined in Definition 7 allow concepts of "rationality" to be given in "money-free" enviroments. Thus, in a cooperatively rational deal, no agent involved suffers a loss in utility and *at least one* is better off. It may be noted that given the characterisation of Definition 4 it is immediate that any cooperatively rational deal is perforce also individually rational; the converse, however, clearly does not hold in general. In some settings, an equitable deal may be neither cooperatively nor individually rational. One may interpret such deals as one method of reducing inequality between the values agents place on their allocations: for those involved in an equitable deal, it is ensured that the agent who places least value on their current allocation will obtain a resource set which is valued more highly. It may, of course, be the case that some agents suffer a *loss* of utility: the condition for a deal to be equitable limits how great such a loss could be. Finally the concept of Pigou-Dalton deal originates from and has been studied in depth within the theory of exchange economies. This is one of many approaches that have been proposed, again in order to describe deals which reduce inequality between members of an agent society, e.g. (Endriss & Maudet, 2004b). In terms of the definition given, such deals encapsulate the so-called Pigou-Dalton principle in economic theory: that any transfer of income from a wealthy individual to a poorer one should reduce the disparity between them. We note that, in principle, we could define related rationality concepts based on several extensions of this principle that have been suggested, e.g. (Atkinson, 1970; Chateauneaf et al., 2002; Kolm, 1976).

Using the same O-contract path constructed in Theorem 3, we need only vary the definitions of the utility functions employed in order to obtain,

Corollary 1 *For each of the cases below,*

a. $\Phi(\delta)$ *holds if and only if* δ *is a cooperatively rational* O*-contract.*
$\Psi(\delta)$ *holds if and only if* δ *is cooperatively rational.*

b. $\Phi(\delta)$ *holds if and only if* δ *is an equitable* O*-contract.*
$\Psi(\delta)$ *holds if and only if* δ *is equitable.*

c. $\Phi(\delta)$ *holds if and only if* δ *is a Pigou-Dalton* O*-contract.*
$\Psi(\delta)$ *holds if and only if* δ *is a Pigou-Dalton deal.*

$$\rho^{\max}(2, m, \Phi, \Psi) \;\geq\; \left(\frac{77}{256}\right) 2^m \;-\; 2$$

Proof. We employ exactly the same sequence of allocations Δ_m described in the proof of Theorem 3 but modify the utility functions $\langle u_1, u_2\rangle$ for each case.

a. Choose $\langle u_1, u_2\rangle$ with $u_2(\gamma) = 0$ for all $\gamma \subseteq \mathcal{R}$ and

$$u_1(\gamma) = \begin{cases} k & \text{if} \quad \gamma = \alpha_k \\ 0 & \text{if} \quad \gamma \notin \{\alpha_1, \ldots, \alpha_t\} \end{cases}$$

The resulting O-contract path is cooperatively rational: the utility enjoyed by A_2 remains constant while that enjoyed by A_1 increases by 1 with each deal. Any deviation from this contract path (employing an alternative O-contract) will result in a loss of utility for A_1.

b. Choose $\langle u_1, u_2\rangle$ with $u_2(\gamma) = u_1(\overline{\gamma})$ and

$$u_1(\gamma) = \begin{cases} k & \text{if} \quad \gamma = \alpha_k \\ 0 & \text{if} \quad \gamma \notin \{\alpha_1, \ldots, \alpha_t\} \end{cases}$$

The O-contract path is equitable: both A_1 and A_2 increase their respective utility values by 1 with each deal. Again, any O-contract deviating from this will result in both agents losing some utility.

c. Choose $\langle u_1, u_2\rangle$ as

$$u_1(\gamma) = \begin{cases} k & \text{if} \quad \gamma = \alpha_k \\ 0 & \text{if} \quad \gamma \notin \{\alpha_1, \ldots, \alpha_t\} \end{cases} \quad ; \quad u_2(\gamma) = \begin{cases} 2^m - k & \text{if} \quad \overline{\gamma} = \alpha_k \\ 2^m & \text{if} \quad \overline{\gamma} \notin \{\alpha_1, \ldots, \alpha_t\} \end{cases}$$

To see that the O-contract path consists of Pigou-Dalton deals, it suffices to note that $u_1(\alpha_i) + u_2(\overline{\alpha_i}) = 2^m$ for each $1 \leq i \leq t$. In addition, $|u_2(\overline{\alpha_{i+1}}) - u_1(\alpha_{i+1})| = 2^m - 2i - 2$ which is strictly less than $|u_2(\overline{\alpha_i}) - u_1(\alpha_i)| = 2^m - 2i$. Finally, any O-contract $\langle P, Q\rangle$ which deviates from this sequence will not be a Pigou-Dalton deal since

$$|u_2(Q_2) - u_1(Q_1)| = 2^m \;>\; |u_2(P_2) - u_1(P_1)|$$

which violates one of the conditions required of Pigou-Dalton deals. □

The construction for two agent settings, easily extends to larger numbers.

Corollary 2 *For each of the choices of $\langle \Phi, \Psi \rangle$ considered in Theorem 3 and Corollary 1, and all $n \geq 2$,*

$$\rho^{\max}(n, m, \Phi, \Psi) \;\geq\; \left(\frac{77}{256}\right) 2^m \;-\; 2$$

Proof. Fix allocations in which A_1 is given α_1, A_2 allocated $\overline{\alpha_1}$, and A_j assigned $\emptyset$ for each $3 \leq j \leq n$. Using identical utility functions $\langle u_1, u_2 \rangle$ as in each of the previous cases, we employ for u_j: $u_j(\emptyset) = 1$, $u_j(S) = 0$ whenever $S \neq \emptyset$ ($\langle \Phi, \Psi \rangle$ as in Theorem 3); $u_j(S) = 0$ for all S (Corollary 1(a)); $u_j(\emptyset) = 2^m$, $u_j(S) = 0$ whenever $S \neq \emptyset$ (Corollary 1(b)); and, finally, $u_j(S) = 2^m$ for all S, (Corollary 1(c)). Considering a realisation of the Ψ-deal $\langle P^{(1)}, P^{(t)} \rangle$ the only Φ-contract path admissible is the path Δ_m defined in the related proofs. This gives the lower bound stated. ▫

We note, at this point, some other consequences of Corollary 1 with respect to (Endriss & Maudet, 2004b, Theorems 1, 3), which state

Fact 1 *We recall that a Φ-path, $\langle P^{(1)}, \ldots, P^{(t)} \rangle$ is* maximal *if for each allocation Q, $\langle P^{(t)}, Q \rangle$ is* not *a Φ-deal.*

a. If $\langle P^{(1)}, \ldots, P^{(t)} \rangle$ is any *maximal path of* cooperatively rational *deals then $P^{(t)}$ is Pareto optimal.*

b. If $\langle P^{(1)}, \ldots, P^{(t)} \rangle$ is any *maximal path of* equitable *deals then $P^{(t)}$ maximises the value $\sigma_e(P) = \min_{1 \leq i \leq n} u_i(P_i)$, i.e. the so-called* egalitarian social welfare*.*

The sequence of cooperatively rational deals in Corollary 1(a) terminates in the Pareto optimal allocation $P^{(t)}$: the allocation for A_2 always has utility 0 and there is no allocation to A_1 whose utility can exceed t. Similarly, the sequence of equitable deals in Corollary 1(b) terminates in the allocation $P^{(t)}$, for which $\sigma_e(P^{(t)}) = t$ the maximum that can be attained for the instance defined. In both cases, however, the optima are reached by sequences of exponentially many (in m) deals: thus, although Fact 1 guarantees convergence of particular deal sequences to optimal states it may be the case, as illustrated in Corollary 1(a–b), that the process of convergence takes considerable time.

2.3 O-contract Paths – Monotone Utility Functions

We conclude our results concerning O-contracts by presenting a lower bound on $\rho^{\max}_{\text{mono}}$, i.e. the length of paths when the utility functions are required to be *monotone*.

In principle one could attempt to construct appropriate monotone utility functions that would have the desired properties with respect to the path used in Theorem 3. It is, however, far from clear whether such a construction is possible. We do not attempt to resolve this question here. Whether an exact translation could be accomplished is, ultimately, a question of purely combinatorial interest: since our aim is to demonstrate that exponential length contract paths are needed with monotone utility functions we are not, primarily, concerned with obtaining an optimal bound.

Theorem 4 *With $\Phi(P,Q)$ and $\Psi(P,Q)$ be defined as in Theorem 3 and $m \geq 14$*

$$\rho^{\max}_{\text{mono}}(2,m,\Phi,\Psi) \;\geq\; \begin{cases} \left(\frac{77}{128}\right)\, 2^{m/2} \;-\; 3 & \text{if} \quad m \text{ is even} \\ \left(\frac{77}{128}\right)\, 2^{(m-1)/2} \;-\; 3 & \text{if} \quad m \text{ is odd} \end{cases}$$

Proof. We describe the details only for the case of m being even: the result when m is odd is obtained by a simple modification which we shall merely provide in outline.
Let $m = 2s$ with $s \geq 7$. For any path

$$\Delta_s \;=\; \langle \alpha_1, \alpha_2, \ldots, \alpha_t \rangle$$

in $\mathcal{H}_s$ (where α_i describes a subset of $\mathcal{R}_s$ by an s-bit label), the path $\mathit{double}(\Delta_s)$ in $\mathcal{H}_{2s}$ is defined by

$$\begin{array}{rcl} \mathit{double}(\Delta_s) & = & \langle\; \alpha_1\overline{\alpha_1},\; \alpha_2\overline{\alpha_2}\;, \ldots,\; \alpha_i\overline{\alpha_i},\; \alpha_{i+1}\overline{\alpha_{i+1}}\;, \ldots,\; \alpha_t\overline{\alpha_t}\;\rangle \\ & = & \langle \beta_1, \beta_3, \ldots, \beta_{2i-1}, \beta_{2i+1}, \ldots, \beta_{2t-1} \rangle \end{array}$$

(The reason for successive indices of β increasing by 2 will become clear subsequently)

Of course, $\mathit{double}(\Delta_s)$ does not describe an O-contract path[6]: it is, however, not difficult to interpolate appropriate allocations, β_{2i}, in order to convert it to such a path. Consider the subsets β_{2i} (with $1 \leq i < t$) defined as follows:

$$\beta_{2i} \;=\; \begin{cases} \alpha_{i+1}\overline{\alpha_i} & \text{if} \quad \alpha_i \subset \alpha_{i+1} \\ \alpha_i\overline{\alpha_{i+1}} & \text{if} \quad \alpha_i \supset \alpha_{i+1} \end{cases}$$

If we now consider the path, $\mathit{ext}(\Delta_s)$, within $\mathcal{H}_{2s}$ given by

$$\mathit{ext}(\Delta_s) \;=\; \langle \beta_1, \beta_2, \beta_3\;, \ldots,\; \beta_{2(t-1)}, \beta_{2t-1} \rangle$$

then this satisfies,

a. If Δ_s has property (SC) of Theorem 3 in $\mathcal{H}_s$ then $\mathit{ext}(\Delta_s)$ has property (SC) in $\mathcal{H}_{2s}$.

b. If j is odd then $|\beta_j| = s$.

c. If j is even then $|\beta_j| = s+1$.

From (a) and the bounds proved in (Abbott & Katchalski, 1991) we deduce that $\mathit{ext}(\Delta_s)$ can be chosen so that with $P^{(i)}$ denoting the allocation $\langle \beta_i, \overline{\beta_i} \rangle$

d. $\mathit{ext}(\Delta_s)$ describes an O-contract path from $P^{(1)}$ to $P^{(2t-1)}$.

e. For each pair $\langle i,j \rangle$ with $j \geq i+2$, the deal $\langle P^{(i)}, P^{(j)} \rangle$ is *not* an O-contract.

f. If Δ_s is chosen as in the proof of Theorem 3 then the number of deals in $\mathit{ext}(\Delta_s)$ is as given in the statement of the present theorem.

6. In terms of the classification described by Sandholm (1998), it contains only *swap* deals (S-contracts): each deal swaps exactly one item in β_{2i-1} with an item in $\overline{\beta_{2i-1}}$ in order to give β_{2i+1}.

We therefore fix Δ_s as the path from Theorem 3 so that in order to complete the proof we need to construct utility functions $\langle u_1, u_2\rangle$ that are monotone and with which $ext(\Delta_s)$ defines the unique IR O-contract path realising the reallocation $\langle P^{(1)}, P^{(2t-1)}\rangle$.

The choice for u_2 is relatively simple. Given $S \subseteq \mathcal{R}_{2s}$,

$$u_2(S) \;=\; \begin{cases} 0 & \text{if} \quad |S| \leq s-2 \\ 2t+1 & \text{if} \quad |S| = s-1 \\ 2t+2 & \text{if} \quad |S| \geq s \end{cases}$$

In this t is the number of allocations in Δ_s. The behaviour of u_2 is clearly monotone.

The construction for u_1 is rather more complicated. Its main idea is to make use of the fact that the size of each set β_i occurring in $ext(\Delta_s)$ is very tightly constrained: $|\beta_i|$ is either s or $s+1$ according to whether i is odd or even. We first demonstrate that each set of size $s+1$ can have at most two strict subsets (of size s) occurring within $ext(\Delta_s)$: thus, every S of size $s+1$ has exactly 2 or 1 or 0 subsets of size s on $ext(\Delta_s)$. To see this suppose the contrary. Let γ, β_{2i-1}, β_{2j-1}, and β_{2k-1} be such that $|\gamma| = s+1$ with

$$\beta_{2i-1} \subset \gamma \;\;;\;\; \beta_{2j-1} \subset \gamma \;\;;\;\; \beta_{2k-1} \subset \gamma$$

Noting that $\beta_{2i-1} = \alpha_i\overline{\alpha_i}$ and that Δ_s has the property (SC) it must be the case that (at least) two of the s-bit labels from $\{\alpha_i, \alpha_j, \alpha_k\}$ differ in at least two positions. Without loss of generality suppose this is true of α_i and α_k. As a result we deduce that the sets β_{2i-1} and β_{2k-1} have at most $s-2$ elements in common, i.e. $|\beta_{2i-1} \cap \beta_{2k-1}| \leq s-2$: $\beta_{2i-1} = \alpha_i\overline{\alpha_i}$ and $\beta_{2k-1} = \alpha_k\overline{\alpha_k}$ so in any position at which α_i differs from α_k, $\overline{\alpha_i}$ differs from $\overline{\alpha_k}$ at exactly the same position. In total $|\beta_{2i-1} \setminus \beta_{2k-1}| \geq 2$, i.e. there are (at least) two elements of β_{2i-1} that do not occur in β_{2k-1}; and in the same way $|\beta_{2k-1} \setminus \beta_{2i-1}| \geq 2$, i.e. there are (at least) two elements of β_{2k-1} that do not occur in β_{2i-1}. The set γ, however, has only $s+1$ members and so cannot have *both* β_{2i-1} and β_{2k-1} as subsets: this would require

$$\beta_{2i-1} \cap \beta_{2k-1} \;\cup\; \beta_{2i-1} \setminus \beta_{2k-1} \;\cup\; \beta_{2k-1} \setminus \beta_{2i-1} \subseteq \gamma$$

but, as we have just seen,

$$|\; \beta_{2i-1} \cap \beta_{2k-1} \;\cup\; \beta_{2i-1} \setminus \beta_{2k-1} \;\cup\; \beta_{2k-1} \setminus \beta_{2i-1} \;| \;\;\geq\;\; s+2$$

One immediate consequence of the argument just given is that for any set γ of size $s+1$ there are exactly two strict subsets of γ occurring on $ext(\Delta_s)$ if and only if $\gamma = \beta_{2i-1} \cup \beta_{2i+1} = \beta_{2i}$ for some value of i with $1 \leq i < t$. We can now characterise each subset of $\mathcal{R}_{2s}$ of size $s+1$ as falling into one of three categories.

C1. *Good* sets, given by $\{\gamma \;:\; \gamma = \beta_{2i}\}$.

C2. *Digressions*, consisting of

$$\{\; \gamma \;:\; \beta_{2i-1} \subset \gamma,\, \gamma \neq \beta_{2i} \text{ and } i < t\}$$

C3. *Inaccessible* sets, consisting of

$$\{\; \gamma \;:\; \gamma \text{ is neither } \textit{Good} \text{ nor a } \textit{Digression}\}$$

Good sets are those describing allocations to A_1 within the path defined by $ext(\Delta_s)$; *Digressions* are the allocations that could be reached using an O-contract from a set of size s on $ext(\Delta_s)$, i.e. β_{2i-1}, but differ from the set that actually occurs in $ext(\Delta_s)$, i.e. β_{2i}. Finally, *Inaccessible* sets are those that do not occur on $ext(\Delta_s)$ *and* cannot be reached via an O-contract from any set on $ext(\Delta_s)$. We note that we view any set of size $s+1$ that *could* be reached by an O-contract from β_{2t-1} as being inaccessible: in principle it is possible to extend the O-contract path beyond β_{2t-1}, however, we choose not complicate the construction in this way.

We now define u_1 as

$$u_1(\gamma) \;=\; \begin{cases} 2i-1 & \text{if} \quad \gamma = \beta_{2i-1} \\ 2i+1 & \text{if} \quad \gamma = \beta_{2i} \\ 2i & \text{if} \quad |\gamma| = s+1 \text{ and } \gamma \text{ is a } \textit{Digression} \text{ from } \beta_{2i-1} \\ 0 & \text{if} \quad |\gamma| \leq s-1 \\ 0 & \text{if} \quad |\gamma| = s \text{ and } \gamma \notin ext(\Delta_s) \\ 2t-1 & \text{if} \quad \gamma \text{ is } \textit{Inaccessible} \text{ or } |\gamma| \geq s+2 \end{cases}$$

It remains only to prove for these choices of $\langle u_1, u_2\rangle$ that the O-contract path $\langle P^{(1)}, \ldots, P^{(2t-1)}\rangle$ defined from $ext(\Delta_s)$ is the unique IR O-contract path realising the IR deal $\langle P^{(1)}, P^{(2t-1)}\rangle$ and that u_1 is monotone.

To show that $\langle P^{(1)}, \ldots, P^{(2t-1)}\rangle$ is IR we need to demonstrate

$$\forall\ 1 \leq j < 2t-1 \quad u_1(\beta_j) + u_2(\overline{\beta_j}) \;<\; u_1(\beta_{j+1}) + u_2(\overline{\beta_{j+1}})$$

We have via the definition of $\langle u_1, u_2\rangle$

$$\begin{array}{rcl} u_1(\beta_{2i-1}) + u_2(\overline{\beta_{2i-1}}) & = & 2(t+i)+1 \\ & < & u_1(\beta_{2i}) + u_2(\overline{\beta_{2i}}) \\ & = & 2(t+i)+2 \\ & < & u_1(\beta_{2i+1}) + u_2(\overline{\beta_{2i+1}}) \\ & = & 2(t+i)+3 \end{array}$$

Thus, via Definition 4, it follows that $ext(\Delta_s)$ gives rise to an IR O-contract path.

To see that this path is the unique IR O-contract path implementing $\langle P^{(1)}, P^{(2t-1)}\rangle$, consider any position $P^{(j)} = \langle \beta_j, \overline{\beta_j}\rangle$ and allocation Q other than $P^{(j+1)}$ or $P^{(j-1)}$. It may be assumed that the deal $\langle P^{(j)}, Q\rangle$ is an O-contract. If $j = 2i-1$ then $\sigma_u(P^{(2i-1)}) = 2(t+i)+1$ and $|\beta_j| = s$. Hence $|Q_1| \in \{s-1, s+1\}$. In the former case, $u_1(Q_1) = 0$ and $u_2(Q_2) = 2t+2$ from which $\sigma_u(Q) = 2t+2$ and thus $\langle P^{(j)}, Q\rangle$ is not IR. In the latter case $u_1(Q_1) = 2i$ since Q_1 is a *Digression* from β_{2i-1} and $u_2(Q_2) = 2t+1$ giving $\sigma_u(Q) = 2(t+i)+1$. Again $\langle P^{(j)}, Q\rangle$ fails to be IR since Q fails to give any increase in the value of σ_u. We are left with the case $j = 2i$ so that $\sigma_u(P^{(2i)}) = 2(t+i)+2$ and $|\beta_j| = s+1$. Since $\langle P^{(j)}, Q\rangle$ is assumed to be an O-contract this gives $|Q_1| \in \{s, s+2\}$. For the first possibility Q_1 could not be a set on $ext(\Delta_s)$: β_{2i-1} and β_{2i+1} are both subsets of β_{2i} and there can be at most two such subsets occurring on $ext(\Delta_s)$. It follows, therefore, that $u_1(Q_1) = 0$ giving $\sigma_u(Q) = 2t+2$ so that $\langle P^{(j)}, Q\rangle$ is not IR. In the second possibility, $u_1(Q_1) = 2t-1$ but $u_2(Q_2) = 0$ as $|Q_2| = s-2$ so the deal would result in an overall loss. We deduce that for each $P^{(j)}$ the only IR O-contract consistent with it is the deal $\langle P^{(j)}, P^{(j+1)}\rangle$.

The final stage is to prove that the utility function u_1 is indeed a *monotone* function. Suppose S and T are subsets of $\mathcal{R}_{2s}$ with $S \subset T$. We need to show that $u_1(S) \leq u_1(T)$. We may assume that $|S| = s$, that S occurs as some set within $ext(\Delta_s)$, and that $|T| = s+1$. If $|S| < s$ or $|S| = s$ but does not occur on $ext(\Delta_s)$ we have $u_1(S) = 0$ and the required inequality holds; if $|S| \geq s+1$ then in order for $S \subset T$ to be possible we would need $|T| \geq s+2$, which would give $u_1(T) = 2t-1$ and this is the maximum value that any subset is assigned by u_1. We are left with only $|S| = s$, $|T| = s+1$ and S on $ext(\Delta_s)$ to consider. It has already been shown that there are at most two subsets of T that can occur on $ext(\Delta_s)$. Consider the different possibilities:

a. $T = \beta_{2i}$ so that exactly two subsets of T occur in $ext(\Delta_s)$: β_{2i-1} and β_{2i+1}. Since $u_1(\beta_{2i}) = 2i+1$ and this is at least $\max\{u_1(\beta_{2i-1}), u_1(\beta_{2i+1})\}$, should S be either of β_{2i-1} or β_{2i+1} then $u_1(S) \leq u_1(T)$ as required.

b. T is a *Digression* from $S = \beta_{2i-1}$, so that $u_1(T) = 2i$ and $u_1(S) = 2i-1$ and, again, $u_1(S) \leq u_1(T)$.

We deduce that u_1 is monotone completing our lower bound proof for $\rho^{\max}_{\text{mono}}$ for even values of m.

We conclude by observing that a similar construction can be used if $m = 2s+1$ is odd: use the path $ext(\Delta_s)$ described above but modifying it so that one resource (r_m) is always held by A_2. Only minor modifications to the utility function definitions are needed. ▫

Example 2 *For $s = 3$, we can choose $\Delta_3 = \langle 000, 001, 101, 111, 110\rangle$ so that $t = 5$. This gives double(Δ_3) as*

$$\langle 000111, 001110, 101010, 111000, 110001\rangle$$

with the O-contract path being defined from $ext(\Delta_3)$ which is

$$\begin{array}{ll} & \langle 000111, 001111, 001110, 101110, 101010, 111010, 111000, 111001, 110001\rangle \\ = & \langle \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9\rangle \end{array}$$

Considering the 15 subsets of size $s+1 = 4$, gives

$$\begin{array}{lcl} \textit{Good} & = & \{001111,\ 101110,\ 111010,\ 111001\} \\ \textit{Digression} & = & \{010111, 100111, 101011, 011110, 111100\} \\ \textit{Inaccessible} & = & \{011011, 011101, 101101, 110110, 110011, 110101\} \end{array}$$

Notice that both of the sets in $\{110011, 110101\}$ are Inaccessible: in principle we could continue from $\beta_9 = 110001$ using either, however, in order to simplify the construction the path is halted at β_9.

Following the construction presented in Theorem 4, gives the following utility function definitions with $S \subseteq \mathcal{R} = \{r_1, r_2, r_3, r_4, r_5, r_6\}$.

$$u_2(S) = \begin{cases} 0 & \text{if} \quad |S| \leq 1 \\ 11 & \text{if} \quad |S| = 2 \\ 12 & \text{if} \quad |S| \geq 3 \end{cases}$$

For u_1 we obtain

$$u_1(S) = \begin{cases} 0 & \text{if} \quad |S| \leq 2 \\ 0 & \text{if} \quad |S| = 3 \text{ and } S \notin \{000111, 001110, 101010, 111000, 110001\} \\ 1 & \text{if} \quad S = 000111 \ \ (\beta_1) \\ 2 & \text{if} \quad S = 010111 \ (\text{digression from } \beta_1) \\ 2 & \text{if} \quad S = 100111 \ (\text{digression from } \beta_1) \\ 3 & \text{if} \quad S = 001111 \ \ (\beta_2) \\ 3 & \text{if} \quad S = 001110 \ \ (\beta_3) \\ 4 & \text{if} \quad S = 011110 \ (\text{digression from } \beta_3) \\ 5 & \text{if} \quad S = 101110 \ \ (\beta_4) \\ 5 & \text{if} \quad S = 101010 \ \ (\beta_5) \\ 6 & \text{if} \quad S = 101011 \ (\text{digression from } \beta_5) \\ 7 & \text{if} \quad S = 111010 \ \ (\beta_6) \\ 7 & \text{if} \quad S = 111000 \ \ (\beta_7) \\ 8 & \text{if} \quad S = 111100 \ (\text{digression from } \beta_7) \\ 9 & \text{if} \quad S = 111001 \ \ (\beta_8) \\ 9 & \text{if} \quad S = 110001 \ \ (\beta_9) \\ 9 & \text{if} \quad |S| \geq 5 \text{ or } S \in \{011011, 011101, 101101, 110110, 110011, 110101\} \end{cases}$$

The monotone utility functions, $\langle u_1, u_2 \rangle$, employed in proving Theorem 4 are defined so that the path arising from $ext(\Delta_s)$ is IR: in the event of either agent suffering a loss of utility the gain made by the other is sufficient to provide a compensatory payment. A natural question that now arises is whether the bound obtained in Theorem 4 can be shown to apply when the rationality conditions preclude any monetary payment, e.g. for cases where the concept of rationality is one of those given in Definition 7. Our next result shows that if we set the rationality condition to enforce cooperatively rational or equitable deals then the bound of Theorem 4 still holds.

Theorem 5 *For each of the cases below and $m \geq 14$*

a. *$\Phi(\delta)$ holds if and only if δ is a cooperatively rational O-contract.*
$\Psi(\delta)$ holds if and only if δ is cooperatively rational.

b. *$\Phi(\delta)$ holds if and only if δ is an equitable O-contract.*
$\Psi(\delta)$ holds if and only if δ is equitable.

$$\rho^{\max}_{\text{mono}}(2, m, \Phi, \Psi) \geq \begin{cases} \left(\frac{77}{128}\right) 2^{m/2} - 3 & \text{if} \quad m \text{ is even} \\ \left(\frac{77}{128}\right) 2^{(m-1)/2} - 3 & \text{if} \quad m \text{ is odd} \end{cases}$$

Proof. We again illustrate the constructions only for the case of m being even, noting the modification to deal with odd values of m outlined at the end of the proof of Theorem 4. The path $ext(\Delta_s)$ is used for both cases.

For (a), we require $\langle u_1, u_2\rangle$ to be defined as monotone functions with which $ext(\Delta_s)$ will be the unique cooperatively rational O-contract path to realise the cooperatively rational deal $\langle P^{(1)}, P^{(2t-1)}\rangle$ where $P^{(j)} = \langle \beta_j, \overline{\beta_j}\rangle$. In this case we set $\langle u_1, u_2\rangle$ to be,

$$\langle u_1(\gamma), u_2(\overline{\gamma})\rangle \;=\; \begin{cases} \langle i, i\rangle & \text{if} \quad \gamma = \beta_{2i-1} \\ \langle i+1, i\rangle & \text{if} \quad \gamma = \beta_{2i} \\ \langle i, i-1\rangle & \text{if} \quad |\gamma| = s+1 \text{ and } \gamma \text{ is a } \textit{Digression} \text{ from } \beta_{2i-1} \\ \langle 0, 2t-1\rangle & \text{if} \quad |\gamma| \leq s-1 \\ \langle 0, 2t-1\rangle & \text{if} \quad |\gamma| = s \text{ and } \gamma \notin ext(\Delta_s) \\ \langle 2t-1, 0\rangle & \text{if} \quad \gamma \text{ is } \textit{Inaccessible} \text{ or } |\gamma| \geq s+2 \end{cases}$$

Since,

$$\begin{array}{lcl} \langle u_1(\beta_{2i-1}), u_2(\overline{\beta_{2i-1}})\rangle & = & \langle i, i\rangle \\ \langle u_1(\beta_{2i}), u_2(\overline{\beta_{2i}})\rangle & = & \langle i+1, i\rangle \\ \langle u_1(\beta_{2i+1}), u_2(\overline{\beta_{2i+1}})\rangle & = & \langle i+1, i+1\rangle \end{array}$$

it is certainly the case that $\langle P^{(1)}, P^{(2t-1)}\rangle$ and all deals on the O-contract path defined by $ext(\Delta_s)$ are cooperatively rational. Furthermore if $Q = \langle \gamma, \overline{\gamma}\rangle$ is any allocation other than $P^{(j+1)}$ then the deal $\langle P^{(j)}, Q\rangle$ will fail to be a cooperatively rational O-contract. For suppose the contrary letting $\langle P^{(j)}, Q\rangle$ without loss of generality be an O-contract, with $Q \notin \{P^{(j-1)}, P^{(j+1)}\}$ – we can rule out the former case since we have already shown such an deal is not cooperatively rational. If $j = 2i-1$ so that $\langle u_1(\beta_j), u_2(\overline{\beta_j})\rangle = \langle i, i\rangle$ then $|\gamma| \in \{s-1, s+1\}$: the former case leads to a loss in utility for A_1; the latter, (since γ is a *Digression* from β_{2i-1}) a loss in utility for A_2. Similarly, if $j = 2i$ so that $\langle u_1(\beta_j), u_2(\overline{\beta_j})\rangle = \langle i+1, i\rangle$ then $|\gamma| \in \{s, s+2\}$: for the first $\gamma \notin ext(\Delta_s)$ leading to a loss of utility for A_1; the second results in a loss of utility for A_2. It follows that the path defined by $ext(\Delta_s)$ is the unique cooperatively rational O-contract path that realises $\langle P^{(1)}, P^{(2t-1)}\rangle$.

It remains only to show that these choices for $\langle u_1, u_2\rangle$ define monotone utility functions.

Consider u_1 and suppose S and T are subsets of $\mathcal{R}_{2s}$ with $S \subset T$. If $|S| \leq s-1$, or S does not occur on $ext(\Delta_s)$ then $u_1(S) = 0$. If $|T| \geq s+2$ or is *Inaccessible* then $u_1(T) = 2t-1$ which is the maximum value attainable by u_1. So we may assume that $|S| = s$, occurs on $ext(\Delta_s)$, i.e. $S = \beta_{2i-1}$, for some i, and that $|T| = s+1$ and is either a *Good* set or a *Digression*. From the definition of u_1, $u_1(S) = i$: if $T \in \{\beta_{2i}, \beta_{2i-2}\}$ then $u_1(T) \geq i = u_1(S)$; if T is a *Digression* from β_{2i-1} then $u_1(T) = i = u_1(S)$. We deduce that if $S \subseteq T$ then $u_1(S) \leq u_1(T)$, i.e. the utility function is monotone.

Now consider u_2 with S and T subsets of $\mathcal{R}_{2s}$ having $S \subset T$. If $|T| \geq s+1$ or $\mathcal{R}_{2s} \setminus T$ does not occur in $ext(\Delta_s)$ then $u_2(T) = 2t-1$ its maximal value. If $|S| \leq s-2$ or $\mathcal{R}_{2s} \setminus S$ is *Inaccessible* then $u_2(S) = 0$. Thus we may assume that $T = \overline{\beta_{2i-1}}$ giving $u_2(T) = i$ and $|S| = s-1$, so that $\mathcal{R}_{2s} \setminus S$ is either a *Digression* or one of the *Good* sets $\{\beta_{2i}, \beta_{2i-2}\}$. If $\mathcal{R}_{2s} \setminus S$ is a *Digression* then $u_2(S) = i-1$; if it is the *Good* set β_{2i-2} then $u_2(S) = i-1 \;<\; u_2(T)$; if it is the *Good* set β_{2i} then $u_2(S) = i = u_2(T)$. It follows that u_2 is monotone completing the proof of part (a).

For (b) we use,

$$\langle u_1(\gamma), u_2(\overline{\gamma})\rangle \;=\; \begin{cases} \langle 2i-1, 2i\rangle & \text{if} \quad \gamma = \beta_{2i-1} \\ \langle 2i+1, 2i\rangle & \text{if} \quad \gamma = \beta_{2i} \\ \langle 2i, 2i-1\rangle & \text{if} \quad |\gamma| = s+1 \text{ and } \gamma \text{ is a } \textit{Digression} \text{ from } \beta_{2i-1} \\ \langle 0, 2t-1\rangle & \text{if} \quad |\gamma| \leq s-1 \\ \langle 0, 2t-1\rangle & \text{if} \quad |\gamma| = s \text{ and } \gamma \notin \textit{ext}(\Delta_s) \\ \langle 2t-1, 0\rangle & \text{if} \quad \gamma \text{ is } \textit{Inaccessible} \text{ or } |\gamma| \geq s+2 \end{cases}$$

These choices give $\textit{ext}(\Delta_s)$ as the unique equitable O-contract path to realise the equitable deal $\langle P^{(1)}, P^{(2t-1)}\rangle$, since

$$\begin{array}{lcl} \min\{u_1(\beta_{2i-1}), u_2(\overline{\beta_{2i-1}})\} & = & 2i-1 \\ \min\{u_1(\beta_{2i}), u_2(\overline{\beta_{2i}})\} & = & 2i \\ \min\{u_1(\beta_{2i+1}), u_2(\overline{\beta_{2i+1}})\} & = & 2i+1 \end{array}$$

each deal $\langle P^{(j)}, P^{(j+1)}\rangle$ is equitable. If $Q = \langle \gamma, \overline{\gamma}\rangle$ is any allocation other than $P^{(j+1)}$ then the deal $\langle P^{(j)}, Q\rangle$ is not an equitable O-contract. Assume that $\langle P^{(j)}, Q\rangle$ is an O-contract, and that $Q \notin \{P^{(j-1)}, P^{(j+1)}\}$. If $j = 2i-1$, so that $P^{(j)} = \langle \beta_{2i-1}, \overline{\beta_{2i-1}}\rangle$ and $\min\{u_1(\beta_{2i-1}), u_2(\overline{\beta_{2i-1}})\} = 2i-1$ then $|\gamma| \in \{s-1, s+1\}$. In the first of these $\min\{u_1(\gamma), u_2(\overline{\gamma})\} = 0$; in the second $\min\{u_1(\gamma), u_2(\overline{\gamma})\} = 2i-1$ since γ must be a *Digression*. This leaves only $j = 2i$ with $P^{(j)} = \langle \beta_{2i}, \overline{\beta_{2i}}\rangle$ and $\min\{u_1(\beta_{2i}), u_2(\overline{\beta_{2i}})\} = 2i$. For this, $|\gamma| \in \{s, s+2\}$: if $|\gamma| = s$ then $\min\{u_1(\gamma), u_2(\overline{\gamma})\} \leq 2i-1$ (with equality when $\gamma = \beta_{2i-1}$); if $|\gamma| = s+2$ then $\min\{u_1(\gamma), u_2(\overline{\gamma})\} = 0$. In total these establish that $\textit{ext}(\Delta_s)$ is the unique equitable O-contract path realising the equitable deal $\langle P^{(1)}, P^{(2t-1)}\rangle$.

That the choices for $\langle u_1, u_2\rangle$ describe monotone utility functions can be shown by a similar argument to that of part (a). □

Example 3 *For $s = 3$ using the same O-contract path $\textit{ext}(\Delta_3)$ as the previous example, i.e.*

$$\begin{array}{rl} & \langle 000111, 001111, 001110, 101110, 101010, 111010, 111000, 111001, 110001\rangle \\ = & \langle \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9\rangle \end{array}$$

For $\langle u_1, u_2\rangle$ in (a) we obtain

$$\langle u_1(S), u_2(\mathcal{R}\setminus S)\rangle \;=\; \begin{cases} \langle 0,9\rangle & \text{if} \quad |S| \leq 2 \\ \langle 0,9\rangle & \text{if} \quad |S| = 3 \text{ and } S \notin \{000111, 001110, 101010, 111000, 110001\} \\ \langle 1,1\rangle & \text{if} \quad S = 000111 \;\; (\beta_1) \\ \langle 1,0\rangle & \text{if} \quad S = 010111 \text{ digression from } \beta_1 \\ \langle 1,0\rangle & \text{if} \quad S = 100111 \text{ digression from } \beta_1 \\ \langle 2,1\rangle & \text{if} \quad S = 001111 \;\; (\beta_2) \\ \langle 2,2\rangle & \text{if} \quad S = 001110 \;\; (\beta_3) \\ \langle 2,1\rangle & \text{if} \quad S = 011110 \text{ digression from } \beta_3 \\ \langle 3,2\rangle & \text{if} \quad S = 101110 \;\; (\beta_4) \\ \langle 3,3\rangle & \text{if} \quad S = 101010 \;\; (\beta_5) \\ \langle 3,2\rangle & \text{if} \quad S = 101011 \text{ digression from } \beta_5 \\ \langle 4,3\rangle & \text{if} \quad S = 111010 \;\; (\beta_6) \\ \langle 4,4\rangle & \text{if} \quad S = 111000 \;\; (\beta_7) \\ \langle 4,3\rangle & \text{if} \quad S = 111100 \text{ digression from } \beta_7 \\ \langle 5,4\rangle & \text{if} \quad S = 111001 \;\; (\beta_8) \\ \langle 5,5\rangle & \text{if} \quad S = 110001 \;\; (\beta_9) \\ \langle 9,0\rangle & \text{if} \quad |S| \geq 5 \text{ or } S \in \{011011, 011101, 101101, 110110, 110011, 110101\} \end{cases}$$

Similarly, in (b)

$$\langle u_1(S), u_2(\mathcal{R} \setminus S)\rangle \;=\; \begin{cases} \langle 0,9\rangle & \text{if} \quad |S| \leq 2 \\ \langle 0,9\rangle & \text{if} \quad |S| = 3 \text{ and } S \notin \{000111, 001110, 101010, 111000, 110001\} \\ \langle 1,2\rangle & \text{if} \quad S = 000111 \;\; (\beta_1) \\ \langle 2,1\rangle & \text{if} \quad S = 010111 \text{ digression from } \beta_1 \\ \langle 2,1\rangle & \text{if} \quad S = 100111 \text{ digression from } \beta_1 \\ \langle 3,2\rangle & \text{if} \quad S = 001111 \;\; (\beta_2) \\ \langle 3,4\rangle & \text{if} \quad S = 001110 \;\; (\beta_3) \\ \langle 4,3\rangle & \text{if} \quad S = 011110 \text{ digression from } \beta_3 \\ \langle 5,4\rangle & \text{if} \quad S = 101110 \;\; (\beta_4) \\ \langle 5,6\rangle & \text{if} \quad S = 101010 \;\; (\beta_5) \\ \langle 6,5\rangle & \text{if} \quad S = 101011 \text{ digression from } \beta_5 \\ \langle 7,6\rangle & \text{if} \quad S = 111010 \;\; (\beta_6) \\ \langle 7,8\rangle & \text{if} \quad S = 111000 \;\; (\beta_7) \\ \langle 8,7\rangle & \text{if} \quad S = 111100 \text{ digression from } \beta_7 \\ \langle 9,8\rangle & \text{if} \quad S = 111001 \;\; (\beta_8) \\ \langle 9,10\rangle & \text{if} \quad S = 110001 \;\; (\beta_9) \\ \langle 9,0\rangle & \text{if} \quad |S| \geq 5 \text{ or } S \in \{011011, 011101, 101101, 110110, 110011, 110101\} \end{cases}$$

That we can demonstrate similar extremal behaviours for contract path length with rationality constraints in both money-based (individual rationality) and money-free (cooperative rationality, equitable) settings irrespective of whether monotonicity properties are assumed, has some interesting parallels with other contexts in which monotonicity is relevant. In particular we can observe that in common with the complexity results already noted from (Dunne et al., 2003) – deciding if an allocation is Pareto optimal, if an allocation maximises σ_u, or if an IR O-contract path exists – requiring utility functions to be monotone does not result in a setting which is computationally more tractable.

3. $M(k)$-contract paths

We now turn to similar issues with respect to $M(k)$-contracts, recalling that in one respect these offer a form of deal that does not fit into the classification of Sandholm (1998). This classification defines four forms of contract type: O-contracts, as considered in the previous section; S-contracts, that involve exactly 2 agents swapping single resources; C-contracts, in which one agent tranfers *at least* two of its resources to another; and M-contracts in which *three or more* agents reallocate their resource holding amongst themselves. Our definition of $M(k)$-contracts permits *two* agents to exchange resources (thus are not M-contracts in Sandholm's (1998) scheme) and the deals permitted are not restricted to O, S, and C-contracts. In one regard, however, $M(k)$-contracts are not as general as M-contracts since a preset bound (k) is specified for the number of agents involved.

Our main result on $M(k)$-contract paths is the following development of Theorem 3.

Theorem 6 *Let $\Phi_k(P, Q)$ be the predicate which holds whenever $\langle P, Q\rangle$ is an IR $M(k)$-contract. For all $k \geq 3$, $n \geq k$ and $m \geq \binom{k}{2}$, there is a resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ and an IR deal $\delta = \langle P, Q\rangle$ for which,*

$$\begin{array}{lcll} L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_k) & = & 1 & (a) \\ L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_{k-1}) & \geq & 2^{\lfloor 2m/k(k-1)\rfloor} - 1 & (b) \\ L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_{k-2}) & & \text{is undefined} & (c) \end{array}$$

Before presenting the proof, we comment about the formulation of the theorem statement and give an overview of the proof structure.

We first note that the lower bounds (where defined) have been phrased in terms of the function L^{opt} as opposed to $\rho^{\max}$ used in the various results on O-contract paths in Section 2.2. It is, of course, the case that the bound claimed for $L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, \Phi_{k-1})$ will also be a lower bound on $\rho^{\max}(n, m, \Phi_{k-1}, \Psi)$ when $n \geq k$ and $\Psi(P, Q)$ holds whenever the deal $\langle P, Q \rangle$ is IR. The statement of Theorem 6, however, claims rather more than this, namely that a *specific* resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ can be defined for each $n \geq k$ and each m, together with an IR deal $\langle P, Q \rangle$ in such a way that: $\langle P, Q \rangle$ can be achieved by a *single* $M(k)$-contract and *cannot* be realised by an IR $M(k-2)$-contract path. Recalling that L^{opt} is a *partial* function, the latter property is equivalent to the claim made in part (c) for the deal $\langle P, Q \rangle$ of the theorem statement. Furthermore, this same deal although achievable by an IR $M(k-1)$-contract path can be so realised only by one whose length is as given in part (b) of the theorem statement.

Regarding the proof itself, there are a number of notational complexities which we have attempted to ameliorate by making some simplifying assumptions concerning the relationship between m – the size of the resource set $\mathcal{R}$ – and k – the number of agents which are needed to realise $\langle P, Q \rangle$ in a *single* IR deal. In particular, we shall assume that m is an exact multiple of $\binom{k}{2}$. We observe that by employing a similar device to that used in the proof of Theorem 4 we can deal with cases for which m does not have this property: if $m = s\binom{k}{2} + q$ for integer values $s \geq 1$ and $1 \leq q < \binom{k}{2}$, we simply employ exactly the same construction using $m - q$ resources with the "missing" q resources from $\mathcal{R}_m$ being allocated to A_1 and never being reallocated within the $M(k-1)$-contract path. This approach accounts for the rounding operation ($\lfloor \ldots \rfloor$) in the exponent term of the lower bound. We shall also assume that the number of agents in $\mathcal{A}$ is exactly k. Within the proof we use a running example for which $k = 4$ and $m = 18 = 3 \times 6$ to illustrate specific features.

We first give an outline of its structure.

Given $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ a resource allocation setting involving k agents and m resources, our aim is to define an IR $M(k-1)$-contract path

$$\Delta \;=\; \langle P^{(1)}, P^{(2)}, \ldots, P^{(t)} \rangle$$

that realises the IR $M(k)$ deal $\langle P^{(1)}, P^{(t)} \rangle$. We will use d to index particular allocations within Δ, so that $1 \leq d \leq t$.

In order to simplify the presentation we employ a setting in which the k agents are $\mathcal{A} = \{A_0, A_1, \ldots, A_{k-1}\}$. Recalling that $m = s\binom{k}{2}$, the resource set $\mathcal{R}_m$ is formed by the union of $\binom{k}{2}$ pairwise disjoint sets of size s. Given distinct values i and j with $0 \leq i < j \leq k-1$, we use $\mathcal{R}^{i,j}$ to denote one of these subsets with $\{r_1^{\{i,j\}}, r_2^{\{i,j\}}, \ldots, r_s^{\{i,j\}}\}$ the s resources that form $\mathcal{R}^{\{i,j\}}$.

There are two main ideas underpinning the structure of each $M(k-1)$-contract in Δ.

Firstly, in the initial and subsequent allocations, the resource set $\mathcal{R}^{\{i,j\}}$ is partitioned between A_i and A_j and any reallocation of resources between A_i and A_j that takes place within the deal $\langle P^{(d)}, P^{(d+1)} \rangle$ will involve *only* resources in this set. Thus, for every allocation $P^{(d)}$ and each pair $\{i, j\}$, if $h \notin \{i, j\}$ then $P_h^{(d)} \cap \mathcal{R}^{\{i,j\}} = \emptyset$. Furthermore, for

$\delta = \langle P^{(d)}, P^{(d+1)} \rangle$ should *both* A_i and A_j be involved, i.e. $\{A_i, A_j\} \subseteq \mathcal{A}^\delta$, then this reallocation of $\mathcal{R}^{\{i,j\}}$ between A_i and A_j will be an O-contract. That is, either exactly one element of $\mathcal{R}^{\{i,j\}}$ will be moved from $P_i^{(d)}$ to become a member of the allocation $P_j^{(d+1)}$ or exactly one element of $\mathcal{R}^{\{i,j\}}$ will be moved from $P_j^{(d)}$ to become a member of the allocation $P_i^{(d+1)}$. In total, every $M(k-1)$-contract δ in Δ consists of a *simultaneous* implementation of $\binom{k-1}{2}$ O-contracts: a single O-contract for each of the distinct pairs $\{A_i, A_j\}$ of agents from the $k-1$ agents in $\mathcal{A}^\delta$.

The second key idea is to exploit one well-known property of the s-dimensional hypercube network: for every $s \geq 2$, $\mathcal{H}_s$ contains a *Hamiltonian cycle*, i.e. a simple directed cycle formed using only the edges of $\mathcal{H}_s$ and containing all 2^s vertices.[7] Now, suppose

$$\mathcal{S}^{(v)} = \underline{v}^{(0)}, \underline{v}^{(1)}, \ldots, \underline{v}^{(i)}, \ldots, \underline{v}^{(2^s-1)}, \underline{v}^{(0)}$$

is a Hamiltonian cycle in the hypercube $\mathcal{H}_s$ and

$$\mathcal{S}^{(w)} = \underline{w}^{(0)}, \underline{w}^{(1)}, \ldots, \underline{w}^{(i)}, \ldots, \underline{w}^{(2^s-1)}, \underline{w}^{(0)}$$

the Hamiltonian cycle in which $\underline{w}^{(i)}$ is obtained by complementing each bit in $\underline{v}^{(i)}$. As we have described in the overview of Section 2.1 we can interpret the s-bit label $\underline{v} = v_1 v_2 \ldots v_s$ as describing a particular subset of $\mathcal{R}^{\{i,j\}}$, i.e. that subset in which $r_k^{\{i,j\}}$ occurs if and only if $v_k = 1$. Similarly from any subset of $\mathcal{R}^{\{i,j\}}$ we may define a unique s-bit word. Now suppose that $P_i^{(d)}$ is the allocation held by A_i in the allocation $P^{(d)}$ of Δ. The deal $\delta = \langle P^{(d)}, P^{(d+1)} \rangle$ will affect $P_i^{(d)} \cap \mathcal{R}^{\{i,j\}}$ in the following way: if $i \notin \mathcal{A}^\delta$ or $j \notin \mathcal{A}^\delta$ then $P_i^{(d+1)} \cap \mathcal{R}^{\{i,j\}} = P_i^{(d)} \cap \mathcal{R}^{\{i,j\}}$ and $P_j^{(d+1)} \cap \mathcal{R}^{\{i,j\}} = P_j^{(d)} \cap \mathcal{R}^{\{i,j\}}$. Otherwise we have $\{i,j\} \subseteq \mathcal{A}^\delta$ and the (complementary) holdings $P_i^{(d)} \cap \mathcal{R}^{\{i,j\}}$ and $P_j^{(d)} \cap \mathcal{R}^{\{i,j\}}$ define (complementary) s-bit labels of vertices in $\mathcal{H}_s$: if these correspond to places $\langle \underline{v}^{(h)}, \underline{w}^{(h)} \rangle$ in the Hamiltonian cycles, then in $P_i^{(d+1)}$ and $P_j^{(d+1)}$ the s-bit labels defined from $P_i^{(d+1)} \cap \mathcal{R}^{\{i,j\}}$ and $P_j^{(d+1)} \cap \mathcal{R}^{\{i,j\}}$ produce the s-bit labels $\underline{v}^{(h+1)}$ and $\underline{w}^{(h+1)}$, i.e. the vertices that succeed $\underline{v}^{(h)}$ and $\underline{w}^{(h)}$ in the Hamiltonian cycles. In total, for each j, A_i initially holds either the subset of $\mathcal{R}^{\{i,j\}}$ that maps to $\underline{v}^{(0)}$ or that maps to $\underline{w}^{(0)}$ and, at the conclusion of the $M(k-1)$-path, holds the subset that maps to $\underline{v}^{(2^s-1)}$ (or $\underline{w}^{(2^s-1)}$). The final detail is that the progression through the Hamiltonian cycles is conducted over a series of *rounds* each round comprising k $M(k-1)$-deals.

We have noted that each $M(k-1)$-contract, $\langle P^{(d)}, P^{(d+1)} \rangle$ that occurs in this path Δ can be interpreted as a set of $\binom{k-1}{2}$ distinct O-contracts. An important property of the utility functions employed is that unless $p \geq k-1$ there will be *no individually rational* $M(p)$-contract path that realises the deal $\langle P^{(d)}, P^{(d+1)} \rangle$, i.e. the $\binom{k-1}{2}$ O-contract deals must occur *simultaneously* in order for the progression from $P^{(d)}$ to $P^{(d+1)}$ to be IR. Although the required deal could be realised by a sequence of O-contracts (or, more generally, any suitable $M(k-2)$-contract path), such realisations will *not* describe an IR contract path.

7. This can be shown by an easy inductive argument. For $s = 2$, the sequence $\langle 00, 01, 11, 10, 00 \rangle$ defines a Hamiltonian cycle in $\mathcal{H}_2$. Inductively assume that $\langle \alpha_1, \alpha_2, \ldots, \alpha_p, \alpha_1 \rangle$ (with $p = 2^s$) is such a cycle in $\mathcal{H}_s$ then $\langle 0\alpha_1, 1\alpha_1, 1\alpha_p, 1\alpha_{p-1}, \ldots, 1\alpha_2, 0\alpha_2 \ldots, 0\alpha_p, 0\alpha_1 \rangle$ defines a Hamiltonian cycle in $\mathcal{H}_{s+1}$.

The construction of utility functions to guarantee such behaviour provides the principal component in showing that the IR deal $\langle P^{(1)}, P^{(t)}\rangle$ cannot be realised with an IR $M(k-2)$-contract path: if Q is *any* allocation for which $\langle P^{(1)}, Q\rangle$ is an $M(k-2)$-contract then $\langle P^{(1)}, Q\rangle$ is not IR.

We now proceed with the proof of Theorem 6.

Proof. (of Theorem 6) Fix $\mathcal{A} = \{A_0, A_1, \ldots, A_{k-1}\}$. $\mathcal{R}$ consists of $\binom{k}{2}$ pairwise disjoint sets of s resources

$$\mathcal{R}^{\{i,j\}} = \{r_1^{\{i,j\}}, r_2^{\{i,j\}}, \ldots, r_s^{\{i,j\}}\}$$

For $k = 4$ and $s = 3$ these yield $\mathcal{A} = \{A_0, A_1, A_2, A_3\}$ and

$$\begin{aligned}
\mathcal{R}^{\{0,1\}} &= \{r_1^{\{0,1\}}, r_2^{\{0,1\}}, r_3^{\{0,1\}}\}\\
\mathcal{R}^{\{0,2\}} &= \{r_1^{\{0,2\}}, r_2^{\{0,2\}}, r_3^{\{0,2\}}\}\\
\mathcal{R}^{\{0,3\}} &= \{r_1^{\{0,3\}}, r_2^{\{0,3\}}, r_3^{\{0,3\}}\}\\
\mathcal{R}^{\{1,2\}} &= \{r_1^{\{1,2\}}, r_2^{\{1,2\}}, r_3^{\{1,2\}}\}\\
\mathcal{R}^{\{1,3\}} &= \{r_1^{\{1,3\}}, r_2^{\{1,3\}}, r_3^{\{1,3\}}\}\\
\mathcal{R}^{\{2,3\}} &= \{r_1^{\{2,3\}}, r_2^{\{2,3\}}, r_3^{\{2,3\}}\}
\end{aligned}$$

We use two ordering structures in defining the $M(k-1)$-contract path.

a.

$$\mathcal{S}^{(v)} = \underline{v}^{(0)}, \underline{v}^{(1)}, \ldots, \underline{v}^{(i)}, \ldots, \underline{v}^{(2^s-1)}, \underline{v}^{(0)}$$

a Hamiltonian cycle in $\mathcal{H}_s$, where without loss of generality, $\underline{v}^{(0)} = 111\ldots11$.

b.

$$\mathcal{S}^{(w)} = \underline{w}^{(0)}, \underline{w}^{(1)}, \ldots, \underline{w}^{(i)}, \ldots, \underline{w}^{(2^s-1)}, \underline{w}^{(0)}$$

the complementary Hamiltonian cycle to this, so that $\underline{w}^{(0)} = 000\ldots00$.

Thus for $k = 4$ and $s = 3$ we obtain

$$\begin{aligned}
\text{a.}\quad \mathcal{S}^{(v)} &= \langle 111, 110, 010, 011, 001, 000, 100, 101\rangle\\
\text{b.}\quad \mathcal{S}^{(w)} &= \langle 000, 001, 101, 100, 110, 111, 011, 010\rangle
\end{aligned}$$

We can now describe the $M(k-1)$-contract path.

$$\Delta \;=\; \langle P^{(1)}, P^{(2)}, \ldots, P^{(t)}\rangle$$

Initial Allocation: $P^{(1)}$.

Define the $k \times k$ Boolean matrix, $B = [b_{i,j}]$ (with $0 \le i, j \le k-1$) by

$$b_{i,j} \;=\; \begin{cases} \bot & \text{if} \quad i = j\\ \neg b_{j,i} & \text{if} \quad i > j\\ \neg b_{i,j-1} & \text{if} \quad i < j \end{cases}$$

We then have for each $1 \leq i \leq k$,

$$P_i^{(1)} \;=\; \bigcup_{j=0}^{i-1}\{\; R^{\{j,i\}} \,:\, b_{i,j} = \top\} \;\cup\; \bigcup_{j=i+1}^{k-1} \{\; R^{\{i,j\}} \,:\, b_{i,j} = \top\}$$

Thus, in our example,

$$B \;=\; \begin{bmatrix} \bot & \top & \bot & \top \\ \bot & \bot & \top & \bot \\ \top & \bot & \bot & \top \\ \bot & \top & \bot & \bot \end{bmatrix}$$

Yielding the starting allocation

$$\begin{array}{lllllll} P_0^{(1)} & = & \mathcal{R}^{\{0,1\}} \cup \mathcal{R}^{\{0,3\}} & = & \langle 111,000,111\rangle & \subseteq & \mathcal{R}^{\{0,1\}} \cup \mathcal{R}^{\{0,2\}} \cup \mathcal{R}^{\{0,3\}} \\ P_1^{(1)} & = & \mathcal{R}^{\{1,2\}} & = & \langle 000,111,000\rangle & \subseteq & \mathcal{R}^{\{0,1\}} \cup \mathcal{R}^{\{1,2\}} \cup \mathcal{R}^{\{1,3\}} \\ P_2^{(1)} & = & \mathcal{R}^{\{0,2\}} \cup \mathcal{R}^{\{2,3\}} & = & \langle 111,000,111\rangle & \subseteq & \mathcal{R}^{\{0,2\}} \cup \mathcal{R}^{\{1,2\}} \cup \mathcal{R}^{\{2,3\}} \\ P_3^{(1)} & = & \mathcal{R}^{\{1,3\}} & = & \langle 000,111,000\rangle & \subseteq & \mathcal{R}^{\{0,3\}} \cup \mathcal{R}^{\{1,3\}} \cup \mathcal{R}^{\{2,3\}} \end{array}$$

The third column in $P_i^{(1)}$ indicating the 3-bit labels characterising each of the subsets of $\mathcal{R}^{\{i,j\}}$ for the three values that j can assume.

Rounds: The initial allocation is changed over a series of *rounds*

$$Q^1, Q^2, \ldots, Q^z$$

each of which involves exactly k distinct $M(k-1)$-contracts. We use $Q^{x,p}$ to indicate the allocation resulting after stage p in round x where $0 \leq p \leq k-1$. We note the following:

a. The initial allocation, $P^{(1)}$ will be denoted by $Q^{0,k-1}$.

b. $Q^{x,0}$ is obtained using a single $M(k-1)$-contract from $Q^{x-1,k-1}$ (when $x \geq 1$).

c. $Q^{x,p}$ is obtained using a single $M(k-1)$-contract from $Q^{x,p-1}$ (when $0 < p \leq k-1$).

Our final item of notation is that of the *cube position of i with respect to j in an allocation* P, denoted $\chi(i,j,P)$. Letting $\underline{u}$ be the s-bit string describing $P_i \cap \mathcal{R}^{\{i,j\}}$ in some allocation P, $\chi(i,j,P)$ is the *index* of $\underline{u}$ in the Hamiltonian cycle $S^{(v)}$ (when $\mathcal{R}^{\{i,j\}} \subseteq P_i^{(1)}$) or the Hamiltonian cycle $S^{(w)}$ (when $\mathcal{R}^{\{i,j\}} \subseteq P_j^{(1)}$). When $P = Q^{x,p}$ for some allocation in the sequence under construction we employ the notation $\chi(i,j,x,p)$, noting that one invariant of our path will be $\chi(i,j,x,p) = \chi(j,i,x,p)$, a property that certainly holds true of $P^{(1)} = Q^{0,k-1}$ since $\chi(i,j,0,k-1) = \chi(j,i,0,k-1) = 0$.

The sequence of allocations in Δ is built as follows. Since $Q^{1,0}$ is the immediate successor of the initial allocation $Q^{0,k-1}$, it suffices to describe how $Q^{x,p}$ is formed from $Q^{x,p-1}$ (when $p > 0$) and $Q^{x+1,0}$ from $Q^{x,k-1}$. Let $Q^{y,q}$ be the allocation to be formed from $Q^{x,p}$. The deal $\delta = \langle Q^{x,p}, Q^{y,q}\rangle$ will be an $M(k-1)$ contract in which $\mathcal{A}^\delta = \mathcal{A} \setminus \{A_q\}$. For each pair $\{i,j\} \subseteq \mathcal{A}^\delta$ we have $\chi(i,j,x,p) = \chi(j,i,x,p)$ in the allocation $Q^{x,p}$. In moving to $Q^{y,q}$ exactly one element of $\mathcal{R}^{\{i,j\}}$ is reallocated between A_i and A_j in such a way that in $Q^{y,q}$,

$\chi(i,j,y,q) = \chi(i,j,x,p)+1$, since A_i and A_j are tracing complementary Hamiltonian cycles with respect to $\mathcal{R}^{\{i,j\}}$ this ensures that $\chi(j,i,y,q) = \chi(j,i,x,p)+1$, thereby maintaining the invariant property.

Noting that for each distinct pair $\langle i,j\rangle$, we either have $\mathcal{R}^{\{i,j\}}$ allocated to A_i in $P^{(1)}$ or $\mathcal{R}^{\{i,j\}}$ allocated to A_j in $P^{(1)}$, the description just outlined indicates that the allocation $P^{(d)} = Q^{x,p}$ is completely specified as follows.

The cube position, $\chi(i,j,x,p)$, satisfies,

$$\chi(i,j,x,p) \;=\; \begin{cases} 0 & \text{if} \quad x = 0 \text{ and } p = k-1 \\ 1 \;+\; \chi(i,j,x-1,k-1) & \text{if} \quad x \geq 1,\ p = 0, \text{ and } p \notin \{i,j\} \\ \chi(i,j,x-1,k-1) & \text{if} \quad x \geq 1,\ p = 0, \text{ and } p \in \{i,j\} \\ 1 \;+\; \chi(i,j,x,p-1) & \text{if} \quad 1 \leq p \leq k-1, \text{ and } p \notin \{i,j\} \\ \chi(i,j,x,p-1) & \text{if} \quad 1 \leq p \leq k-1, \text{ and } p \in \{i,j\} \end{cases}$$

For each i, the subset of $\mathcal{R}^{\{i,j\}}$ that is held by A_i in the allocation $Q^{x,p}$ is,

$$\begin{array}{ll} \underline{v}^{(\chi(i,j,x,p))} & \text{if} \quad \mathcal{R}^{\{i,j\}} \subseteq P_i^{(1)} \\ \underline{w}^{(\chi(i,j,x,p))} & \text{if} \quad \mathcal{R}^{\{i,j\}} \subseteq P_j^{(1)} \end{array}$$

(where we recall that s-bit labels in the hypercube $\mathcal{H}_s$ are identified with subsets of $\mathcal{R}^{\{i,j\}}$.)

The tables below illustrates this process for our example.

			A_0			A_1			A_2			A_3			
			$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$\mathcal{A}^{\langle P^{(d-1)},P^{(d)}\rangle}$
d	x	p	0 1	0 2	0 3	1 0	1 2	1 3	2 0	2 1	2 3	3 0	3 1	3 2	
1	0	3	111	000	111	000	111	000	111	000	111	000	111	000	–
2	1	0	111	000	111	000	110	001	111	001	110	000	110	001	$\{A_1, A_2, A_3\}$
3	1	1	111	001	110	000	110	001	110	001	010	001	110	101	$\{A_0, A_2, A_3\}$
4	1	2	110	001	010	001	110	101	110	001	010	101	010	101	$\{A_0, A_1, A_3\}$
5	1	3	010	101	010	101	010	101	010	101	010	101	010	101	$\{A_0, A_1, A_2\}$
6	2	0	010	101	011	101	011	100	010	100	011	101	011	100	$\{A_1, A_2, A_3\}$
7	2	1	010	100	001	101	011	100	011	100	001	100	011	110	$\{A_0, A_2, A_3\}$
8	2	2	011	100	001	100	011	110	011	100	001	110	001	110	$\{A_0, A_1, A_3\}$
9	2	3	001	110	001	110	001	110	001	110	001	110	001	110	$\{A_0, A_1, A_2\}$
⋮	⋮	⋮		⋮			⋮			⋮			⋮		⋮

Subsets of $\mathcal{R}^{\{i,j\}}$ held by A_i in $Q^{x,p}$ ($k = 4$, $s = 3$)

			A_0			A_1			A_2			A_3			
			$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$i\ j$	$\mathcal{A}^{\langle P^{(d-1)},P^{(d)}\rangle}$
d	x	p	0 1	0 2	0 3	1 0	1 2	1 3	2 0	2 1	2 3	3 0	3 1	3 2	
1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	–
2	1	0	0	0	0	0	1	1	0	1	1	0	1	1	$\{A_1, A_2, A_3\}$
3	1	1	0	1	1	0	1	1	1	1	2	1	1	2	$\{A_0, A_2, A_3\}$
4	1	2	1	1	2	1	1	2	1	1	2	2	2	2	$\{A_0, A_1, A_3\}$
5	1	3	2	2	2	2	2	2	2	2	2	2	2	2	$\{A_0, A_1, A_2\}$
6	2	0	2	2	2	2	3	3	2	3	3	2	3	3	$\{A_1, A_2, A_3\}$
7	2	1	2	3	3	2	3	3	3	3	4	3	3	4	$\{A_0, A_2, A_3\}$
8	2	2	3	3	4	3	3	4	3	3	4	4	4	4	$\{A_0, A_1, A_3\}$
9	2	3	4	4	4	4	4	4	4	4	4	4	4	4	$\{A_0, A_1, A_2\}$
⋮	⋮	⋮		⋮			⋮			⋮			⋮		⋮

Cube Positions $\chi(i,j,x,p)$ ($k=4$, $s=3$)

It is certainly the case that this process of applying successive rounds of k deals could be continued, however, we wish to do this only so long as it is not possible to go from some allocation $P^{(d)}$ in the sequence to another $P^{(d+r)}$ for some $r \geq 2$ via an $M(k-1)$-contract.

Now if $Q^{x,p}$ and $Q^{y,q}$ are distinct allocations generated by the process above then the deal $\delta = \langle Q^{x,p}, Q^{y,q}\rangle$ is an $M(k-1)$-contract if and only if for some A_i, $Q_i^{x,p} = Q_i^{y,q}$. It follows that if $\langle P^{(d)}, P^{(d+r)}\rangle$ is an $M(k-1)$-contract for some $r > 1$, then for some i and all $j \neq i$, $P_i^{(d+r)} \cap \mathcal{R}^{\{i,j\}} = P_i^{(d)} \cap \mathcal{R}^{\{i,j\}}$.

To determine the minimum value of $r > 1$ with which $P_i^{(d+r)} = P_i^{(d)}$, we observe that without loss of generality we need consider only the case $d = i = 0$, i.e. we determine the minimum number of deals before $P_0^{(1)}$ reappears. First note that in each round, Q^x, if $\chi(0,j,x-1,k-1) = p$ then $\chi(0,j,x,k-1) = p+k-2$, i.e. each round advances the cube position $k-2$ places: $\chi(0,j,x-1,k-1) = \chi(0,j,x,0)$ and $\chi(0,j,x,j) = \chi(0,j,x,j-1)$. We can also observe that $P_0^{(1)} = Q_0^{0,k-1} \neq Q_0^{x,p}$ for any p with $0 < p < k-1$, since

$$\chi(0,1,x,p) = \chi(0,2,x,p) = \ldots = \chi(0,k-1,x,p)$$

only in the cases $p = 0$ and $p = k-1$. It follows that our value $r > 1$ must be of the form qk where q must be such that $q(k-2)$ is an *exact multiple* of 2^s. From this observation we see that,

$$\min\{\ r > 1 :\ P_0^{(1)} = P_0^{(1+r)}\ \} \ =\ \min\{\ qk :\ q(k-2) \text{ is a multiple of } 2^s\}$$

Now, if k is *odd* then $q = 2^s$ is the minimal such value, so that $r = k2^s$. If k is even then it may be uniquely written in the form $z2^l + 2$ where z is *odd* so giving q as 1 (if $l \geq s$) or 2^{s-l} (if $l \leq s$), so that these give $r = k$ and $r = z2^s + 2^{s-l+1}$, e.g. for $k = 4$ and $s = 3$, we get $k = 1 \times 2^1 + 2$ so that $r = 2^3 + 2^{3-1+1} = 16$ and in our example $P_0^{(1)} = P_0^{(17)}$ may be easily verified. In total,

$$r \geq \begin{cases} k2^s & \text{if} \quad k \text{ is odd} \\ k & \text{if} \quad k = z2^l + 2,\ z \text{ is odd, and } l \geq s \\ 2^s & \text{if} \quad k = z2^l + 2,\ z \text{ is odd and } l \leq s \end{cases}$$

All of which immediately give $r \geq 2^s$ (in the second case $k \geq 2^s$, so the inequality holds trivially), and thus we can continue the chain of $M(k-1)$ contracts for at least 2^s moves. Recalling that $m = s\binom{k}{2}$, this gives the length of the $M(k-1)$-contract path

$$\Delta \;=\; \langle P^{(1)}, P^{(2)}, \ldots, P^{(t)} \rangle$$

written in terms of m and k as at least[8]

$$2^{m/\binom{k}{2}} - 1 \;=\; 2^{\frac{2m}{k(k-1)}} \;-\; 1$$

It remains to define appropriate utility functions $\mathcal{U} = \langle u_0, \ldots, u_{k-1} \rangle$ in order to ensure that Δ is the unique IR $M(k-1)$-contract path realising the IR $M(k)$-deal $\langle P^{(1)}, P^{(t)} \rangle$. In defining $\mathcal{U}$ it will be convenient to denote Δ as the path

$$\Delta = \langle Q^{0,k-1}, Q^{1,0}, Q^{1,1}, \ldots, Q^{1,k-1}, \ldots, Q^{x,p}, \ldots, Q^{r,k-1} \rangle$$

and, since $rk \geq 2^s$, we may without loss of generality, focus on the first 2^s allocations in this contract path.

Recalling that $\chi(i,j,x,p)$ is the index of the s-bit label $\underline{u}$ corresponding to $Q_i^{x,p} \cap \mathcal{R}^{\{i,j\}}$ in the relevant Hamiltonian cycle – i.e. $\mathcal{S}^{(v)}$ if $\mathcal{R}^{\{i,j\}} \subseteq Q_i^{0,k}$, $\mathcal{S}^{(w)}$ if $\mathcal{R}^{\{i,j\}} \subseteq Q_j^{0,k-1}$ – we note the following properties of the sequence of allocations defined by Δ that hold for each distinct i and j.

P1. $\forall\ x, p\ \chi(i,j,x,p) = \chi(j,i,x,p)$

P2. If $Q^{y,q}$ is the immediate successor of $Q^{x,p}$ in Δ then $\chi(i,j,y,q) \leq \chi(i,j,x,p) + 1$ with equality if and only if $q \notin \{i,j\}$.

P3. $\forall\ i', j'$ with $0 \leq i', j' \leq k-1$, $\chi(i,j,x,k-1) = \chi(i',j',x,k-1)$.

The first two properties have already been established in our description of Δ. The third follows from the observation that within each round Q^x, each cube position is advanced by exactly $k-2$ in progressing from $Q^{x-1,0}$ to $Q^{x,k-1}$.

The utility function u_i is now given, for $S \subseteq \mathcal{R}_m$, by

$$u_i(S) = \begin{cases} \sum_{j \neq i} \chi(i,j,x,p) & \text{if } S = Q_i^{x,p} \text{ for some } 0 \leq x \leq r,\ 0 \leq p \leq k-1 \\ -2^{km} & \text{otherwise} \end{cases}$$

We claim that, with these choices,

$$\Delta = \langle Q^{0,k-1}, Q^{1,0}, Q^{1,1}, \ldots, Q^{1,k-1}, \ldots, Q^{x,p}, \ldots, Q^{r,k-1} \rangle$$

is the unique IR $M(k-1)$-contract path realising the IR $M(k)$-deal $\langle Q^{0,k-1}, Q^{r,k-1} \rangle$. Certainly, Δ *is* an IR $M(k-1)$-contract path: each deal $\delta = \langle Q^{x,p}, Q^{y,q} \rangle$ on this path has $|\mathcal{A}^\delta| = k-1$ and since for each agent A_i in $\mathcal{A}^\delta = \mathcal{A} \setminus \{A_q\}$ the utility of $Q_i^{y,q}$ has increased

8. We omit the rounding operation $\lfloor \ldots \rfloor$ in the exponent, which is significant only if m is not an exact multiple of $\binom{k}{2}$, in which event the device described in our overview of the proof is applied.

by exactly $k-2$, i.e. each cube position of i with respect to j whenever $q \notin \{i,j\}$ has increased, it follows that $\sigma_u(Q^{y,q}) > \sigma_u(Q^{x,p})$ and hence $\langle Q^{x,p}, Q^{y,q}\rangle$ is IR.

We now show that Δ is the *unique* IR $M(k-1)$-contract path continuation of $Q^{0,k-1}$ Suppose $\delta = \langle Q^{x,p}, P\rangle$ is a deal that deviates from the contract path Δ (having followed it through to the allocation $Q^{x,p}$). Certainly both of the following must hold of P: for each i, $P_i \subseteq \cup_{j\neq i}\mathcal{R}^{\{i,j\}}$; and there is a k-tuple of pairs $\langle (x_0,p_0),\ldots,(x_{k-1},p_{k-1})\rangle$ with which $P_i = Q_i^{x_i,p_i}$, for if either fail to be the case for some i, then $u_i(P_i) = -2^{km}$ with the consequent effect that $\sigma_u(P) < 0$ and thence not IR. Now, if $Q^{y,q}$ is the allocation that would succeed $Q^{x,p}$ in Δ then $P \neq Q^{y,q}$, and thus for at least one agent, $Q_i^{x_i,p_i} \neq Q_i^{y,q}$. It cannot be the case that $Q_i^{x_i,p_i}$ corresponds to an allocation occurring *strictly later* than $Q_i^{y,q}$ in Δ since such allocations could not be realised by an $M(k-1)$-contract. In addition, since $P_i = Q_i^{x_i,p_i}$ it must be the case that $|\mathcal{A}^\delta| = k-1$ since exactly $k-1$ cube positions in the holding of A_i must change. It follows that there are only two possibilities for (y_i,p_i): P_i reverts to the allocation immediately preceding $Q_i^{x,p}$ or advances to the holding $Q_i^{y,q}$. It now suffices to observe that a deal in which some agents satisfy the first of these while the remainder proceed in accordance with the second either does not give rise to a valid allocation or cannot be realised by an $M(k-1)$-contract. On the other hand if P corresponds to the allocation preceding $Q^{x,p}$ then δ is not IR. We deduce, therefore, that the *only* IR $M(k-1)$ deal that is consistent with $Q^{x,p}$ is that prescribed by $Q^{y,q}$.

This completes the analysis needed for the proof of part (b) of the theorem. It is clear that since the system contains only k agents, any deal $\langle P, Q\rangle$ can be effected with a single $M(k)$-contract, thereby establishing part (a). For part (c) – that the IR deal $\langle P^{(1)}, P^{(t)}\rangle$ cannot be realised using an *individually rational* $M(k-2)$-contract path, it suffices to observe that since the class of IR $M(k-2)$-contracts are a subset of the class of IR $M(k-1)$-contracts, were it the case that an IR $M(k-2)$-contract path existed to implement $\langle P^{(1)}, P^{(t)}\rangle$, this would imply that Δ was not the *unique* IR $M(k-1)$-contract path. We have, however, proved that Δ *is* unique, and part (c) of the theorem follows. ▫

We obtain a similar development of Corollary 1 in

Corollary 3 *For all $k \geq 3$, $n \geq k$, $m \geq \binom{k}{2}$ and each of the cases below,*

a. $\Phi_k(\delta)$ holds if and only if δ is a cooperatively rational $M(k)$-contract.
$\Psi(\delta)$ holds if and only if δ is cooperatively rational.

b. $\Phi_k(\delta)$ holds if and only if δ is δ is an equitable $M(k)$-contract.
$\Psi(\delta)$ holds if and only if δ is is equitable.

there is a resource allocation setting $\langle \mathcal{A},\mathcal{R},\mathcal{U}\rangle$ and a Ψ-deal $\delta = \langle P, Q\rangle$ for which

$$\begin{array}{lcll} L^{\text{opt}}(\delta,\langle \mathcal{A},\mathcal{R},\mathcal{U}\rangle,\Phi_k) & = & 1 & (a)\\ L^{\text{opt}}(\delta,\langle \mathcal{A},\mathcal{R},\mathcal{U}\rangle,\Phi_{k-1}) & \geq & 2^{\lfloor 2m/k(k-1)\rfloor}-1 & (b)\\ L^{\text{opt}}(\delta,\langle \mathcal{A},\mathcal{R},\mathcal{U}\rangle,\Phi_{k-2}) & & \text{is undefined} & (c)\end{array}$$

Proof. As with the proof of Corollary 1 in relation to Theorem 3, in each case we employ the contract path from the proof of Theorem 6, varying the definition of $\mathcal{U} = \langle u_1, u_2, \ldots, u_k\rangle$ in order to establish each result. Thus let

$$\begin{array}{rcl}\Delta_m & = & \langle P^{(1)}, P^{(2)}, \ldots, P^{(r)}, \ldots, P^{(t)}\rangle\\ & = & \langle Q^{0,k-1}, Q^{1,0}, \ldots, Q^{x,p}, \ldots, Q^{z,r}\rangle\end{array}$$

be the $M(k-1)$-contract path realising the $M(k)$-deal $\langle P^{(1)}, P^{(t)}\rangle$ described in the proof of Theorem 6, this path having length $t \geq 2^{\lfloor 2m/k(k-1)} - 1$.

a. The utility functions $\mathcal{U} = \langle u_0, \ldots, u_{k-1}\rangle$ of Theorem 6 ensure that $\langle P^{(1)}, P^{(t)}\rangle$ is cooperatively rational and that Δ_m is a cooperatively rational $M(k-1)$-contract path realising $\langle P^{(1)}, P^{(t)}\rangle$: the utility held by A_i never decreases in value and there is at least one agent (in fact exactly $k-1$) whose utility increases in value. Furthermore Δ_m is the unique cooperatively rational $M(k-1)$-contract path realising $\langle P^{(1)}, P^{(t)}\rangle$ since, by the same argument used in Theorem 6, any deviation will result in some agent suffering a loss of utility.

b. Set the utility functions $\mathcal{U} = \langle u_0, \ldots, u_{k-1}\rangle$ as,

$$u_i(S) = \begin{cases} -1 & \text{if} \quad S \neq Q_i^{x,p} \text{ for any } Q^{x,p} \in \Delta_m \\ xk^2 + k - i & \text{if} \quad S = Q_i^{x,k-1} \\ (x-1)k^2 + k + p & \text{if} \quad S = Q_0^{x,p},\ p < k-1 \text{ and } i = 0 \\ (x-1)k^2 + k - i + p + 1 & \text{if} \quad S = Q_i^{x,p},\ p < i-1 \text{ and } i \neq 0. \\ xk^2 + 1 & \text{if} \quad S = Q_i^{x,i-1} = Q_i^{x,i} \text{ and } i \neq 0. \\ xk^2 + 1 + p - i & \text{if} \quad S = Q_i^{x,p},\ p > i \text{ and } i \neq 0 \end{cases}$$

To see that these choices admit Δ_m as an equitable $M(k-1)$-contract path realising the equitable deal $\langle Q^{0,k-1}, Q^{z,r}\rangle$, we first note that

$$\min_{0\leq i\leq k-1} \{u_i(Q_i^{z,r})\} > 1 = \min_{0\leq i\leq k-1} \{u_i(Q_i^{0,k-1})\}$$

thus, $\langle Q^{0,k-1}, Q^{z,r}\rangle$ is indeed equitable. Consider any deal $\delta = \langle Q^{x,p}, Q^{y,q}\rangle$ occurring within Δ_m. It suffices to show that

$$\min_{0\leq i\leq k-1} \{u_i(Q_i^{x,p})\} \neq u_q(Q_q^{x,p})$$

since $A_q \notin \mathcal{A}^\delta$, and for all other agents $u_i(Q_i^{y,q}) > u_i(Q_i^{x,p})$. We have two possibilities: $q = 0$ (in which case $p = k-1$ and $y = x+1$); $q > 0$ (in which case $p = q-1$). Consider the first of these: $u_0(Q_0^{x,k-1}) = xk^2 + k$, however,

$$\min\{u_i(Q_i^{x,k-1})\} = xk^2 + 1 = u_{k-1}(Q_{k-1}^{x,k-1})$$

and hence every deal $\langle Q^{x,k-1}, Q^{x+1,0}\rangle$ forming part of Δ_m is equitable.

In the remaining case, $u_q(Q_q^{x,q-1}) = xk^2 + 1$ and

$$\begin{aligned} \min\{u_i(Q_i^{x,q-1})\} &\leq u_0(Q_0^{x,q-1}) \\ &= (x-1)k^2 + k + q - 1 \\ &< xk^2 - (k^2 - 2k + 1) \\ &= xk^2 - (k-1)^2 \\ &< xk^2 + 1 \\ &= u_q(Q_q^{x,q-1}) \end{aligned}$$

and thus the remaining deals $\langle Q^{x,q-1}, Q^{x,q}\rangle$ within Δ_m are equitable. By a similar argument to that employed in Theorem 6 it follows that Δ_m is the unique equitable $M(k-1)$-contract path realising $\langle Q^{0,k-1}, Q^{z,r}\rangle$.

□

Monotone Utility Functions and $M(k)$-contract paths

The device used to develop Theorem 3 to obtain the path of Theorem 4 can be applied to the rather more intricate construction of Theorem 6, thereby allowing exponential lower bounds on $\rho^{\max}_{\text{mono}}(n, m, \Phi_k, \Psi)$ to be derived. We will merely outline the approach rather than present a detailed technical exposition. We recall that it became relatively straightforward to define suitable monotone utility functions once it was ensured that the subset sizes of interest – i.e. those for allocations arising in the O-contract path – were forced to fall into a quite restricted range. The main difficulty that arises in applying similar methods to the path Δ of Theorem 6 is the following: in the proof of Theorem 4 we consider two agents so that converting Δ_s from a setting with s resources in Theorem 3 to $ext(\Delta_s)$ with $2s$ resources in Theorem 4 is achieved by combining "complementary" allocations, i.e. $\alpha \subseteq \mathcal{R}_s$ with $\overline{\alpha} \subseteq \mathcal{T}_s$. We can exploit two facts, however, to develop a path $multi(\Delta)$ for which monotone utility functions could be defined: the resource set $\mathcal{R}_m$ in Theorem 6 consists of $\left(\begin{smallmatrix} k \\ 2 \end{smallmatrix} \right)$ disjoint sets of size s; and any deal δ on the path Δ involves a reallocation of $\mathcal{R}^{\{i,j\}}$ between A_i and A_j when $\{i,j\} \subseteq \mathcal{A}^{\delta}$. Thus letting $\mathcal{T}_m$ be formed by $\left(\begin{smallmatrix} k \\ 2 \end{smallmatrix} \right)$ disjoint sets, $\mathcal{T}^{\{i,j\}}$ each of size s, suppose that $P_i^{(d)}$ is described by

$$\alpha_{i,0}^{(d)}\ \alpha_{i,1}^{(d)}\ \cdots\ \alpha_{i,i-1}^{(d)}\ \alpha_{i,i+1}^{(d)}\ \cdots\ \alpha_{i,k-1}^{(d)}$$

with $\alpha_{i,j}^{(d)}$ the s-bit label corresponding to the subset of $R^{\{i,j\}}$ that is held by A_i in $P^{(d)}$. Consider the sequence of allocations,

$$multi(\Delta) \ = \ \langle C^{(1)}, C^{(2)}, \ldots, C^{(t)} \rangle$$

in a resource allocation setting have k agents and $2m$ resources – $\mathcal{R}_m \cup \mathcal{T}_m$ for which $C_i^{(d)}$ is characterised by

$$\beta_{i,0}^{(d)}\ \beta_{i,1}^{(d)}\ \cdots\ \beta_{i,i-1}^{(d)}\ \beta_{i,i+1}^{(d)}\ \cdots\ \beta_{i,k-1}^{(d)}$$

In this, $\beta_{i,j}^{(d)}$, indicates the subset of $\mathcal{R}^{\{i,j\}} \cup \mathcal{T}^{\{i,j\}}$ described by the $2s$-bit label,

$$\beta_{i,j}^{(d)} \ = \ \alpha_{i,j}^{(d)}\ \overline{\alpha_{i,j}^{(d)}}$$

i.e. $\alpha_{i,j}^{(d)}$ selects a subset of $\mathcal{R}^{\{i,j\}}$ while $\overline{\alpha_{i,j}^{(d)}}$ a subset of $\mathcal{T}^{\{i,j\}}$.

It is immediate from this construction that for each allocation $C^{(d)}$ in $multi(\Delta)$ and each A_i, it is always the case that $|C_i^{(d)}| = (k-1)s$. It follows, therefore, that the only subsets that are relevant to the definition of *monotone* utility functions with which an analogous result to Theorem 6 for the path $multi(\Delta)$ could be derived, are those of size $(k-1)s$: if $S \subseteq \mathcal{R}_m \cup \mathcal{T}_m$ has $|S| < (k-1)s$, we can fix $u_i(S)$ as a small enough negative value; similarly if $|S| > (k-1)s$ then $u_i(S)$ can be set to a large enough positive value.[9]

Our description in the preceding paragraphs, can be summarised in the following result, whose proof is omitted: extending the outline given above to a formal lower bound

9. It is worth noting that the "interpolation" stage used in Theorem 4 is not needed in forming $multi(\Delta)$: the deal $\langle C^{(d)}, C^{(d+1)} \rangle$ is an $M(k-1)$-contract. We recall that in going from Δ_s of Theorem 3 to $ext(\Delta_s)$ the intermediate stage – $double(\Delta_s)$ – was not an O-contract path.

proof, is largely a technical exercise employing much of the analysis already introduced, and since nothing signifcantly new is required for such an analysis we shall not give a detailed presentation of it.

Theorem 7 *Let $\Phi_k(P, Q)$ be the predicate which holds whenever $\langle P, Q\rangle$ is an IR $M(k)$-contract. For all $k \geq 3$, $n \geq k$ and $m \geq 2\binom{k}{2}$, there is a resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ in which every $u \in \mathcal{U}$ is* monotone*, and an IR deal $\delta = \langle P, Q\rangle$ for which,*

$$\begin{array}{lcll} L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_k) & = & 1 & (a) \\ L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_{k-1}) & \geq & 2^{\lfloor m/k(k-1)\rfloor} - 1 & (b) \\ L^{\text{opt}}(\delta, \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle, \Phi_{k-2}) & & \text{is undefined} & (c) \end{array}$$

4. Related Work

The principal focus of this article has considered a property of contract paths realising rational reallocations $\langle P, Q\rangle$ when the constituent deals are required to conform to a structural restriction and satisfy a rationality constraint. In Section 2 the structural restriction limited deals to those involving a single resource, i.e. O-contracts. For the rationality constraint forcing deals strictly to improve utilitarian social welfare, i.e. to be individually rational (IR) we have the following properties.

a. There are resource allocation settings $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ within which there are IR reallocations $\langle P, Q\rangle$ that *cannot* be realised by a sequence of IR O-contracts. (Sandholm, 1998, Proposition 2)

b. Every IR reallocation, $\langle P, Q\rangle$, that *can* be realised by an IR O-contract path, can be realised by an IR O-contract path of length at most $n^m - (n-1)m$. (Sandholm, 1998, Proposition 2)

c. Given $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ together with an IR reallocation $\langle P, Q\rangle$ the problem of deciding if $\langle P, Q\rangle$ can be implemented by an IR O-contract path is NP–hard, even if $|\mathcal{A}| = 2$ and both utility functions are monotone. (Dunne et al., 2003, Theorem 11).

d. There are resource allocation settings $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ within which there are IR reallocations $\langle P, Q\rangle$ that *can* be realised by an IR O-contract path, but with any such path having length exponential in m. This holds even in the case $|\mathcal{A}| = 2$ and both utility functions are monotone. (Theorem 3 and Theorem 4 of Section 2)

In a recent article Endriss and Maudet (2004a) analyse contract path length also considering O-contracts with various rationality constraints. Although the approach is from a rather different perspective, the central question addressed – "How many rational deals are required to reach an optimal allocation?", (Endriss & Maudet, 2004a, Table 1, p. 629) – is closely related to the issues discussed above. One significant difference in the analysis of rational O-contracts from Sandholm's (1998) treatment and the results in Section 2 is that in (Endriss & Maudet, 2004a) the utility functions are restricted so that *every* rational reallocation $\langle P, Q\rangle$ *can* be realised by a rational O-contract path. The two main restrictions examined are requiring utility functions to be *additive*, i.e. for every $S \subseteq \mathcal{R}$, $u(S) = \sum_{r \in S} u(r)$;

and, requiring the value returned to be either 0 or 1, so-called $0-1$ utility functions. Additive utility functions are considered in the case of IR O-contracts (Endriss & Maudet, 2004a, Theorems 3, 9), whereas $0-1$ utility functions for cooperatively rational O-contracts (Endriss & Maudet, 2004a, Theorems 4, 11). Using $\rho^{\max}_{\text{add}}(n, m, \Phi, \Psi)$ and $\rho^{\max}_{0-1}(n, m, \Phi, \Psi)$ to denote the functions introduced in Definition 6 where all utility functions are additive (respectively $0-1$), cf. the definition of $\rho^{\max}_{\text{mono}}$, then with $\Phi_1(P, Q)$ holding if $\langle P, Q\rangle$ is an IR O-contract; $\Phi_2(P, Q)$ holding if $\langle P, Q\rangle$ is a cooperatively rational O-contract and $\Psi(P, Q)$ true when $\langle P, Q\rangle$ is IR, we may formulate Theorems 9 and 11 of (Endriss & Maudet, 2004a) in terms of the framework used in Definition 6, as

$$\begin{array}{lcll} \rho^{\max}_{\text{add}}(n, m, \Phi_1, \Psi) & = & m & (\textit{Endriss \& Maudet}, 2004a, \textit{Theorem } 9) \\ \rho^{\max}_{0-1}(n, m, \Phi_2, \Psi) & = & m & (\textit{Endriss \& Maudet}, 2004a, \textit{Theorem } 11) \end{array}$$

We can, of course, equally couch Theorems 3 and 4 of Section 2 in terms of the "shortest-path" convention adopted in (Endriss & Maudet, 2004a), provided that the domains of utility and reallocation instances are restricted to those for which an appropriate O-contract path exists. Thus, we can obtain the following development of (Endriss & Maudet, 2004a, Table 1) in the case of O-contracts.

Utility Functions	Additive	0-1	Unrestricted	Monotone	Unrestricted	Monotone
Rationality	IR	CR	IR	IR	CR	CR
Shortest Path	m	m	$\Omega(2^m)$	$\Omega(2^{m/2})$	$\Omega(2^m)$	$\Omega(2^{m/2})$
Complete	Yes	Yes	No	No	No	No

Table 2: How many O-contract rational deals are required to reach an allocation? Extension of Table 1 from (Endriss & Maudet, 2004a, p. 629)

5. Conclusions and Further Work

Our aim in this article has been to develop the earlier studies of Sandholm (1998) concerning the scope and limits of particular "practical" contract forms. While Sandholm (1998) has established that insisting on individual rationality in addition to the structural restriction prescribed by O-contracts leads to scenarios which are incomplete (in the sense that there are individually rational deals that cannot be realised by individually rational O-contracts) our focus has been with respect to deals which can be realised by restricted contract paths, with the intention of determining to what extent the combination of structural and rationality conditions increases the number of deals required. We have shown that, using a number of natural definitions of rationality, for settings involving m resources, *rational* O-contract paths of length $\Omega(2^m)$ are needed, whereas without the rationality restriction on individual deals, at most m O-contracts suffice to realise any deal. We have also considered a class of deals – $M(k)$-contracts – that were not examined in (Sandholm, 1998), establishing for these cases that, when particular rationality conditions are imposed, $M(k-1)$-contract paths of length $\Omega(2^{2m/k^2})$ are needed to realise a deal that can be achieved by a single $M(k)$-contract.

We note that our analyses have primarily been focused on worst-case lower bounds on path length when appropriate paths exist, and as such there are several questions of

practical interest that merit further discussion. It may be noted that the path structures and associated utility functions are rather artificial, being directed to attaining a path of a specific length meeting a given rationality criterion. We have seen, however, in Theorems 4 and 5 as outlined in our discussion concluding Section 3 that the issue of exponential length contract paths continues to arise even when we require the utility functions to satisfy a monotonicity condition. We can identify two classes of open question that arise from these results.

Firstly, focusing on IR O-contract paths, it would be of interest to identify "natural" restrictions on utility functions which would ensure that, *if* a deal $\langle P, Q\rangle$ can be implemented by an IR O-contract path, then it can be realised by one whose length is polynomially bounded in m, e.g. such as additivity mentioned in the preceding section. We can interpret Theorem 4, as indicating that monotonicity does *not* guarantee "short" IR contract paths. We note, however, that there *are* some restrictions that suffice. To use a rather trivial example, if the number of *distinct* values that σ_u can assume is at most m^p for some *constant* p then no IR O-contract path can have length exceeding m^p: successive deals must strictly increase σ_u and if this can take at most K different values then no IR contract path can have length exceeding K. As well as being of practical interest, classes of utility function with the property being considered would also be of some interest regarding one complexity issue. The result proved in (Dunne et al., 2003) establishing that deciding if an IR O-contract path exists is NP-hard, gives a *lower bound* on the computational complexity of this problem. At present, no (non-trivial) *upper bound* on this problem's complexity has been demonstrated. Our results in Theorems 3 and 4 indicate that *if* this decision problem is in NP (thus its complexity would be NP–complete rather than NP–hard) then the required polynomial length existence certificate *may* have to be something other than the path itself.[10] We note that the proof of NP–hardness in (Dunne et al., 2003) constructs an instance in which σ_u can take at most $O(m)$ distinct values: thus, from our example of a restriction ensuring that if such are present then IR O-contract paths are "short", this result of (Dunne et al., 2003) indicates that the question of *deciding* their existence might remain computationally hard.

Considering restrictions on the form of utility functions is one approach that could be taken regarding finding "tractable" cases. An alternative would be to gain some insight into what the "average" path length is likely to be. In attempting to address this question, however, a number of challenging issues arise. The most immediate of these concerns, of course, the notion of modeling a distribution on utility function given our definitions of rationality in terms of the value agents attach to their resource holdings. In principle an average-case analysis of scenarios involving exactly *two* agents could be carried out in purely graph-theoretic terms, i.e. without the complication of considering utility functions directly. It is unclear, however, whether such a graph-theoretic analysis obviating the need for consideration of literal utility functions, can be extended beyond settings involving exactly two agents. One difficulty arising with three or more agents is that our utility

10. The use of "may" rather than "must" is needed because of the convention for representing utility functions employed in (Dunne et al., 2003).

functions have no allocative externalities, i.e. given an allocation $\langle X, Y, Z\rangle$ to three agents, $u_1(X)$ is unchanged should $Y \cup Z$ be redistributed among A_2 and A_3.[11]

As one final set of issues that may merit further study we raise the following. In our constructions, the individual deals on a contract path must satisfy both a *structural* condition (be an O-contract or involve at most k agents), and a *rationality* constraint. Focusing on O-contracts we have the following extremes: from (Sandholm, 1998), at most m O-contracts suffice to realise any rational deal; from our results above, $\Omega(2^m)$ *rational* O-contracts are needed to realise some rational deals. There are a number of mechanisms we can employ to relax the condition that every single deal be an O-contract and be rational. For example, allow a path to contain some number of deals which are not O-contracts (but must still be IR) or insist that all deals are O-contracts but allow some to be irrational. Thus, in the latter case, if we go to the extent of allowing up to m irrational O-contracts, then any rational deal can be realised efficiently. It would be of some interest to examine issues such as the effect of allowing a *constant* number, t, of irrational deals and questions such as whether there are situations in which t irrational contracts yield a 'short' contract path but $t-1$ force one of exponential length. Of particular interest, from an application viewpoint, is the following: define a $(\gamma(m), O)$-path as an O-contract path containing at most $\gamma(m)$ O-contracts which are not individually rational. We know that if $\gamma(m) = 0$ then individually rational $(0, O)$-paths are not complete with respect to individually rational deals; similarly if $\gamma(m) = m$ then (m, O)-paths *are* complete with respect to individually rational deals. A question of some interest would be to establish if there is some $\gamma(m) = o(m)$ for which $(\gamma(m), O)$-paths are complete with respect to individually rational deals *and* with the maximum length of such a contract path bounded by a polynomial function of m.

Acknowledgements

The author thanks the reviewers of an earlier version of this article for their valuable comments and suggestions which have contributed significantly to its content and organisation. The work reported in this article was carried out under the support of EPSRC Grant GR/R60836/01.

References

Abbott, H. L., & Katchalski, M. (1991). On the construction of snake in the box codes. *Utilitas Mathematica*, *40*, 97–116.

Atkinson, A. (1970). On the measurement of inequality. *Jnl. Econ. Theory*, *2*, 244–263.

Chateauneaf, A., Gajdos, T., & Wilthien, P.-H. (2002). The principle of strong diminishing transfer. *Jnl. Econ. Theory*, *103*, 311–333.

11. A very preliminary investigation of complexity-theoretic questions arising in settings with allocative externalities is presented in (Dunne, 2004) where these are referred to as "context–dependent": such utility functions appear to have been neglected in the computational and algorithmic analysis of resource allocation problems, although the idea is well-known to game-theoretic models in economics from which the term "allocative externality" originates.

Dignum, F., & Greaves, M. (2000). *Issues in Agent Communication*, Vol. 1916 of *LNCS*. Springer-Verlag.

Dunne, P. (2003). Prevarication in dispute protocols. In *Proc. Ninth International Conf. on A.I. and Law (ICAIL'03)*, pp. 12–21, Edinburgh. ACM Press.

Dunne, P. (2004). Context dependence in multiagent resource allocation. In *Proc. ECAI'04*, pp. 1000–01, Valencia.

Dunne, P., & McBurney, P. (2003). Optimal utterances in dialogue protocols. In *Proc. Second International Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS'03)*, pp. 608–615. ACM Press.

Dunne, P., Wooldridge, M., & Laurence, M. (2003). The complexity of contract negotiation. Tech. rep. ULCS-03-002, Dept. of Computer Science, Univ. of Liverpool. (to appear *Artificial Intelligence*).

Endriss, U., & Maudet, N. (2004a). On the communication complexity of multilateral trading. In *Proc. Third International Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS'04)*, pp. 622–629.

Endriss, U., & Maudet, N. (2004b). Welfare engineering in multiagent systems. In Omicini, A., Petta, P., & Pitt, J. (Eds.), *Proc. Fourth International Workshop on Engineering Societies in the Agents World (ESAW-2003)*, Vol. 3071 of *LNAI*, pp. 93–106. Springer-Verlag.

Endriss, U., Maudet, N., Sadri, F., & Toni, F. (2003). On optimal outcomes of negotiations over resources. In *Proc. Second International Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS'03)*, pp. 177–184. ACM Press.

Kautz, W. H. (1958). Unit distance error checking codes. *IRE Trans. on Electronic Computers*, *7*, 179–180.

Kolm, S.-C. (1976). Unequal inequalities. *Jnl. Econ. Theory*, *13*, 82–111.

Kraus, S. (2001). *Strategic negotiation in multiagent environments*. MIT Press.

McBurney, P., Parsons, S., & Wooldridge, M. (2002). Desiderata for argumentation protocols. In *Proc. First International. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pp. 402–409. ACM Press.

Parkes, D. C., & Ungar, L. H. (2000a). Iterative combinatorial auctions: theory and practice. In *Proc. 17th National Conf. on Artificial Intelligence (AAAI-00)*, pp. 74–81.

Parkes, D. C., & Ungar, L. H. (2000b). Preventing strategic manipulation in iterative auctions: proxy agents and price adjustment. In *Proc. 17th National Conf. on Artificial Intelligence (AAAI-00)*, pp. 82–89.

Rosenschein, J. S., & Zlotkin, G. (1994). *Rules of Encounter*. MIT Press.

Sandholm, T. W. (1998). Contract types for satisficing task allocation: I theoretical results. In *AAAI Spring Symposium: Satisficing Models*.

Sandholm, T. W. (1999). Distributed rational decision making. In Weiß, G. (Ed.), *Multiagent Systems*, pp. 201–258. MIT Press.

Sandholm, T. W. (2002). Algorithm for optimal winner determination in combinatorial auctions. *Artificial Intelligence*, *135*, 1–54.

Sandholm, T. W., & Suri, S. (2003). Bob: Improved winner determination in combinatorial auctions and generalizations. *Artificial Intelligence*, *145*, 33–58.

Sandholm, T. W., Suri, S., Gilpin, A., & Levine, D. (2001). Cabob: A fast optimal algorithm for combinatorial auctions.. In *Proc. IJCAI-01*, pp. 1102–1108.

Smith, R. G. (1980). The contract net protocol: high-level communication and control in a distributed problem solver. *IEEE Trans. on Computers*, *C-29*(12), 1104–1113.

Tennenholz, M. (2000). Some tractable combinatorial auctions. In *Proc. 17th National Conf. on Artificial Intelligence (AAAI-00)*.

Yokoo, M., Sakurai, Y., & Matsubara, S. (2004). The effect of false-name bids in combinatorial auctions: new fraud in internet auctions. *Games and Economic Behavior*, *46*(1), 174–188.

Optimal Utterances in Dialogue Protocols

Optimal Utterances in Dialogue Protocols

Paul E. Dunne
Dept. of Computer Science
University of Liverpool
Liverpool L69 7ZF, UK
ped@csc.liv.ac.uk

Peter McBurney
Dept. of Computer Science
University of Liverpool
Liverpool L69 7ZF, UK
p.j.mcburney@csc.liv.ac.uk

ABSTRACT

Dialogue protocols have been the subject of considerable attention with respect to their potential applications in multiagent system environments. Formalisations of such protocols define classes of dialogue *locutions*, concepts of a dialogue *state*, and *rules* under which a dialogue proceeds. One important consideration in implementing a protocol concerns the criteria an agent should apply in choosing which utterance will constitute its next contribution to a discussion in progress: ideally, an agent should select a locution that (by some measure) "*optimises*" the outcome. The precise interpretation of '*optimise*' is, however, something that may vary greatly depending on the nature and intent of a dialogue area. If we consider '*persuasion*' protocols, where one agent's intention is to convince others of the validity or invalidity of a specific proposition, then optimality might be regarded in the sense of "choice of locution that results in a '*minimal length*' debate": thus the agent defending a hypothesis tries to select utterances that will convince other participants of the validity of this hypothesis after 'as few locutions as possible'. We present a formal setting for considering the problem of deciding if a particular utterance in the context of a *persuasion dialogue* is optimal in this sense. We show that, in general, this decision problem is *both* NP–hard *and* CO-NP–hard.

ACM CATEGORIES AND SUBJECT DESCRIPTORS:

F.2.2 **[Analysis of Algorithms and Problem Complexity]:** Nonnumerical algorithms and problems: *Complexity in proof procedures*. I.2.11 **[Artificial Intelligence]** Distributed Artificial Intelligence: *Coherence and co-ordination; Languages and Structures; Multiagent systems*.

GENERAL TERMS: Design, Languages, Theory.

KEYWORDS: Agent Communication Languages, Argumentation and Persuasion, Computational Complexity, Dialogue Protocols, Locution Selection.

AAMAS'03, July 14-18, 2003, Melbourne, Australia.

1. INTRODUCTION

Methods for modeling discussion and dialogue processes have proved to be of great importance in describing multiagent interactions. The study of *dialogue protocols* ranges from perspectives such as argumentation theory, e.g., [23, 26], taxonomies of types of dialogue such as [26, 28], and formalisms for describing and reasoning about protocols, e.g. [19, 21, 22]. Among the many applications that have been considered are bargaining and negotiation processes, e.g. [14, 23, 17]; legal reasoning, e.g. [13, 16, 1, 2, 25], persuasion in argumentation and other systems, e.g. [27, 11, 3, 8], and inquiry and information-discovery, e.g. [18, 20]. The collection of articles presented in [7] gives an an overview of various perspectives relating to multiagent discourse.

While we present a general formal model for dialogue protocols below, informally we may view the core elements of such as comprising a description of the 'locution types' for the protocol ("what participants can *say*"); the topics of discussion ("what participants talk *about*"); and how discussions may start, evolve, and finish.

Despite the divers demands of protocols imposing special considerations of interest with particular applications, there are some properties that might be considered desirable irrespective of the protocol's specific domain, cf. [22]. Among such properties are *termination*; the capability to *validate* that a discussion is being conducted according to the protocol; and the ability for participants to determine "sensible" contributions. In [21] frameworks for uniform comparison of protocols are proposed that are defined independently of the application domain. In principle, if two distinct protocols can be shown 'equivalent' in the senses defined by [21], then termination and other properties need only be proved for one of them.

In this paper our concern is with the following problem: in realising a particular discussion protocol within a multiagent environment, one problem that must be addressed by each participant can, informally, be phrased as "*what do/should I say next*?" In other words, each agent must "be aware of" its permitted (under the protocol rules) *utterances* given the progress of the discussion so far, and following specific criteria, either choose to say nothing or contribute one of these. While the extent to which a protocol admits a 'reasonable' decision-making process is, of course, a property that is of domain-independent interest, one crucial feature distinguishing different types of discussion protocol is the criteria that apply when an agent makes its choice. More precisely, in making a contribution an agent may be seen as "optimising" the outcome. A clear distinction between protocol applications is that the sense of "optimality" in one protocol may be quite different from "optimality" in another. For example, in multiagent bidding and bargaining protocols, a widely-used concept of "optimal utterance" is based on the view that any utterance has the force of affording a par-

ticular "utility value" to the agent invoking it that may affect the utility enjoyed by other agents. In such settings, the *policy* (often modeled as a probability distribution) is "optimal" if no agent can improve its (expected) utility by unilaterally deviating. This – Nash equilibrium – has been the subject of intensive research and there is strong evidence of its computational intractability [4]. While valid as a criterion for utterances in multiagent bargaining protocols, such a model of "optimality" is less well-suited to fields such as persuasion, information-gathering, etc. We may treat a "persuasion protocol" as one in which an agent seeks to convince others of the validity of a given proposition, and interpreting such persuasion protocols as proof mechanisms – a view used in, among others, [27, 11] – we contend that a more appropriate sense of an utterance being "optimal", is that it *allows* the discussion to be *concluded* "as quickly as possible".[1] There are several reasons why such a measure is appropriate with respect to persuasion protocols. In practice, discussions in which one agent attempts to persuade another to carry out some action cannot (reasonably) be allowed to continue indefinitely; an agent may be unable to continue with other tasks which are time-constrained in some sense until other agents in the system have been persuaded through some reasoned discussion to accept particular propositions. It is, of course, the case that describing optimality in terms of length of discussion provides only one measure. We discuss alternative notions of optimality in the concluding sections.

Concentrating on persuasion protocols we formulate the "optimal utterance problem" and establish lower bounds on its complexity. In the next section we outline an abstract computational framework for dialogue protocols and introduce two variants of the optimal utterance decision problem. In Section 3 we present a setting in which this problem is proved to be both NP–hard *and* CO-NP–hard. Conclusions and further work are presented in the final section.

2. DEFINITIONS

DEFINITION 1. *Let $\mathcal{F}$ be the (infinite) set of all well-formed formulae (wff) in some propositional language (where we assume an enumerable set of propositional variables $x_1, x_2, \ldots$).*

A dialogue arena*, denoted $\mathcal{A}$, is a (typically infinite) set of* finite *subsets of $\mathcal{F}$. For a dialogue arena,*

$$\mathcal{A} = \{\Phi_1, \Phi_2, \ldots, \Phi_k, \ldots, \} \quad \Phi_i \subset \mathcal{F}$$

the set of wff in $\Phi_i = \{\psi_1, \psi_2, \ldots, \psi_q\}$ is called a dialogue context *from the dialogue arena $\mathcal{A}$.*

DEFINITION 2. *A* dialogue schema *is a triple $\langle \mathcal{L}, \mathcal{D}, \Phi \rangle$, where $\mathcal{L} = \{L_j | 1 \leq j \leq l\}$ is a finite set of* locution types*, $\mathcal{D}$ is a* dialogue protocol *as defined below, and Φ is a dialogue context.*

We are interested in reasoning about properties of protocols operating in given dialogue arenas. In the following, $\mathcal{A} = \{\Phi_1, \Phi_2, \ldots, \}$ is a dialogue arena, with $\Phi = \{\psi_1, \ldots, \psi_q\}$ a (recall, finite) set of wff constituting a single dialogue context of this arena.

DEFINITION 3. *Let $\mathcal{L} = \{L_j | 1 \leq j \leq l\}$ be a set of* locution types*. A* dialogue fragment *over the dialogue context Φ is a (finite) sequence,*

$$\mu_1 \cdot \mu_2 \cdots \mu_k$$

where $\mu_t = L_{j,t}(\theta_t)$ is the instantiated locution *or* utterance *(with $\theta_t \in \Phi$) at time t. The* commitment *represented by a dialogue fragment δ – denoted $\Sigma(\delta)$ – is a subset of the context Φ.*

The notation $M^*_{\mathcal{L},\Phi}$ is used to denote the set of all dialogue fragments involving instantiated locutions from $\mathcal{L}$; δ to denote an arbitrary member of this set, and $|\delta|$ to indicate the length (number of utterances) in δ.

In order to represent dialogues of interest we need to describe mechanisms by which dialogue fragments and their associated commitments evolve.

DEFINITION 4. *A* dialogue protocol *for the discussion of the context Φ using locution set $\mathcal{L}$ – is a pair $\mathcal{D} = \langle \Pi, \Sigma \rangle$ defined by:*

a. A possible dialogue continuation *function –*

$$\Pi : M^*_{\mathcal{L},\Phi} \rightarrow \wp(\mathcal{L} \times \Phi) \cup \{\bot\}$$

The subset of dialogue fragments δ in $M^*_{\mathcal{L},\Phi}$ having $\Pi(\delta) \neq \bot$ is called the set of legal dialogues over $\langle \mathcal{L}, \Phi \rangle$ in the protocol $\mathcal{D}$*, this subset being denoted $T_{\mathcal{D}}$. It is required that the* empty dialogue fragment*, ϵ containing no locutions is a legal dialogue, i.e. $\Pi(\epsilon) \neq \bot$, and we call the set $\Pi(\epsilon)$ the* legal commencement locutions.[2] *We further require that Π satisfies the following condition:*

$$\forall \delta \in M^*_{\mathcal{L},\Phi} \ (\Pi(\delta) = \bot) \Rightarrow (\forall \mu = \mathcal{L}_j(\theta) \ \Pi(\delta \cdot \mu) = \bot)$$

i.e. if δ is not a legal dialogue then no dialogue fragment starting with δ is a legal dialogue.

b. A commitment function *– $\Sigma : T_{\mathcal{D}} \rightarrow \wp(\Phi)$ associating each legal dialogue with a subset of the dialogue context Φ.*

This definition abstracts away ideas concerning commencement, combination and termination rules into the pair $\langle \Pi, \Sigma \rangle$ through which the possible dialogues of a protocol and the associated states (subsets of Φ) are defined. Informally, given a legal dialogue, δ, $\Pi(\delta)$ delineates all of the utterances that may be used to continue the discussion.

A dialogue, δ, is *terminated* if $\Pi(\delta) = \emptyset$ and *partial* if $\Pi(\delta) \neq \emptyset$.

We now describe mechanisms for assessing dialogue protocols in terms of the *length* of a dialogue. The following notation is used.

$$\Delta = \{\Delta_k\} = \{\langle \mathcal{L}, \mathcal{D} = \langle \Pi, \Sigma \rangle, \Phi_k \rangle\}$$

is a (sequence of) dialogue schemata for an arena

$$\mathcal{A} = \{\Phi_1, \ldots, \Phi_k, \ldots\}$$

Although one can introduce concepts of dialogue length predicated on the number of utterances needed to attain a particular state Θ, the decision problem we consider will focus on the concept of "minimal length terminated *continuation* of a dialogue fragment δ". Formally

DEFINITION 5. *Let $\langle \mathcal{L}, \mathcal{D} = \langle \Pi, \Sigma \rangle, \Phi_k \rangle$ be a dialogue schema Δ_k instantiated with the context Φ_k of $\mathcal{A}$. Let $\delta \in M^*_{\langle \mathcal{L},\Phi_k \rangle}$ be a dialogue fragment. The* completion length *of δ under $\mathcal{D}$ for the context Φ_k, denoted $\chi(\delta, \mathcal{D}, \Phi_k)$, is,*

$$\min\{|\eta| : \eta \in T_{\mathcal{D}}, \eta = \delta \cdot \zeta, \Pi(\eta) = \emptyset\}$$

if *such a dialogue fragment exists, and undefined otherwise.*

[1] An alternative view is proposed in [10], where it is argued that utterances which *prolong* discussions can, in certain settings, be seen as "optimal".

[2] Note that we allow $\Pi(\epsilon) = \emptyset$, although the dialogues that result from this case are unlikely to be of significant interest.

Thus the completion length of a (legal) dialogue, δ, is the least number of utterances in a terminated dialogue that *starts* with δ. We note that if δ is *not* a legal dialogue then $\chi(\delta, \mathcal{D}, \Phi_k)$ is always undefined.

The decision problem whose properties we are concerned with is called the *Generic Optimal Utterance Problem*.

DEFINITION 6. *An instance of the* Generic Optimal Utterance Problem *(*GOUP*) comprises,*

$$\mathcal{U} = \langle \Delta, \delta, \mu \rangle$$

where $\Delta = \langle \mathcal{L}, \mathcal{D}, \Phi \rangle$ is a dialogue schema with locution set $\mathcal{L}$, protocol $\mathcal{D} = \langle \Pi, \Sigma \rangle$, and dialogue context Φ; $\delta \in M^*_{\langle \mathcal{L}, \Phi \rangle}$ is a dialogue fragment, and $\mu \in \mathcal{L} \times \Phi$ is an utterance.

An instance $\mathcal{U}$ is accepted if there exists a dialogue fragment $\eta \in M^*_{\langle \mathcal{L}, \Phi \rangle}$ for which all *of the following hold*

1. $\eta = \delta \cdot \mu \cdot \zeta \in T_{\mathcal{D}}$.
2. $\Pi(\eta) = \emptyset$.
3. $|\eta| = \chi(\delta, \mathcal{D}, \Phi)$.

If any of these fail to hold, the instance is rejected.

Thus, given representations of a dialogue schema together with a *partial* dialogue, δ and utterance μ, an instance is accepted if there is a terminated dialogue (η) which commences with the dialogue fragment $\delta \cdot \mu$ and whose length is the completion length of δ under $\mathcal{D}$ for the context Φ. In other words, the utterance μ is such that it *is* a legal continuation of δ leading to a shortest length terminated dialogue.

Our formulation of GOUP, as given in Definition 6, raises a number of questions. The most immediate of these concerns how the schema Δ is to be represented, specifically the protocol $\langle \Pi, \Sigma \rangle$. Noting that we have (so far) viewed $\langle \Pi, \Sigma \rangle$ in abstract terms as mappings from dialogue fragments to sets of utterances (subsets of the context), one potential difficulty is that in "most" cases these will not be *computable*.[3] We can go some way to addressing this problem by representing $\langle \Pi, \Sigma \rangle$ through (encodings of) Turing machine programs $\langle M_\Pi, M_\Sigma \rangle$ with the following characteristics: M_Π takes as its input a *pair* $\langle \delta, \mu \rangle$, where $\delta \in M^*_{\langle \mathcal{L}, \Phi \rangle}$ and $\mu \in \mathcal{L} \times \Phi$, accepting if $\delta \cdot \mu$ is a legal dialogue and rejecting otherwise; similarly M_Σ takes as its input a pair $\langle \delta, \Psi \rangle$ with $\Psi \in \Phi$ accepting if δ is a legal dialogue having $\Psi \in \Sigma(\delta)$, rejecting otherwise. There remain, however, problems with this approach: it is *not* possible, in general, to validate that a given input *is* an instance of GOUP, cf. Rice's Theorem for Recursive Index Sets in e.g., [9, Chapter 5, pp. 58–61]; secondly, even in those cases where one can interpret the encoding of $\langle \Pi, \Sigma \rangle$ "appropriately" the definition places no time-bound on how long the computation of these programs need take. There are two methods we can use to overcome these difficulties: one is to employ 'clocked' Turing machine programs, so that, for example, if no decision has been reached for an instance $\langle \delta, \mu \rangle$ on M_Π after, say $|\delta \cdot \mu|$ steps, then the instance is rejected. The second is to consider *specific* instantiations of GOUP with protocols that can be established "independently" to have desirable efficient decision procedures. More formally,

DEFINITION 7. *Instances of the* Optimal Utterance Problem in $\Delta - \text{OUP}^{(\Delta)}$ *– where $\{\Delta\} = \{\langle \mathcal{L}, \mathcal{D} = \langle \Pi, \Sigma \rangle, \Phi_k \rangle\}$ is a sequence of dialogue schema over the arena $\mathcal{A} = \{\Phi_k : k \geq 1\}$, comprise*

$$\mathcal{U} = \langle \Phi_k, \delta, \mu \rangle$$

where $\delta \in M^*_{\langle \mathcal{L}, \Phi_k \rangle}$ is a dialogue fragment, and $\mu \in \mathcal{L} \times \Phi_k$ is an utterance.

An instance $\mathcal{U}$ is accepted if there exists a dialogue fragment $\eta \in M^*_{\langle \mathcal{L}, \Phi_k \rangle}$ for which all *of the following hold*

1. $\eta = \delta \cdot \mu \cdot \zeta \in T_{\mathcal{D}}$.
2. $\Pi(\eta) = \emptyset$.
3. $|\eta| = \chi(\delta, \mathcal{D}, \Phi_k)$.

If any of these fail to hold, the instance is rejected.

The crucial difference between the problems GOUP and $\text{OUP}^{(\Delta)}$ is that we can consider the latter in the context of *specific* protocols without being concerned about *how* these are represented – the protocol description does *not* form part of an instance of $\text{OUP}^{(\Delta)}$ (only the specific context Φ_k). In particular, should we wish to consider some 'sense of complexity' for a given schema, we could use the device of employing an 'oracle' Turing machine, M_Δ, to report (at unit-cost) whether properties (1–2) hold of any given η. With such an approach, should Δ be such that the set of legal dialogues for a specific context is *finite*, then the decision problem $\text{OUP}^{(\Delta)}$ is *decidable* (relative to the oracle machine M_Δ). A further advantage is that any *lower* bound that can be demonstrated for a specific incarnation of $\text{OUP}^{(\Delta)}$ gives a lower bound on the "computable fragment" of GOUP. In the next section, we describe a (sequence of) dialogue schemata, $\{\Delta_k^{DPLL}\}$ for which the following computational properties are provable.

1. The set of legal dialogues for Δ_k^{DPLL} is finite: thus every continuation of any legal partial dialogue will result in a legal terminated dialogue.
2. Given $\langle \delta, \mu, \Phi_k \rangle$ with δ a dialogue fragment, μ an utterance, and Φ_k the dialogue context for Δ_k^{DPLL}, there is a deterministic algorithm that decides if $\delta \cdot \mu$ is a legal dialogue using time linear in the number of bits needed to encode the instance.
3. Given $\langle \delta, \Psi, \Phi_k \rangle$ with δ a legal dialogue and Ψ an element of the context Φ_k, there is a deterministic algorithm deciding if $\Psi \in \Sigma(\delta)$ using time linear in the number of bits needed to encode the instance.

We will show that the Optimal Utterance Problem for Δ_k^{DPLL} is both NP–hard and CO-NP–hard.

3. THE OPTIMAL UTTERANCE PROBLEM

Prior to defining the schema used as the basis of our results, we introduce the dialogue arena, $\mathcal{A}_{CNF}$ upon which it operates.

Let $\Theta(n)$ ($n \geq 1$) denote the set of all CNF formulae formed from propositional variables $\{x_1, \ldots, x_n\}$ (so that $|\Theta(n)| = 2^{3^n}$). For $\Psi \in \Theta(n)$ with

$$\Psi = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{i_r} y_{i,j} \quad y_{i,j} \in \{x_k, \neg x_k \ : \ 1 \leq k \leq n\}$$

we use C_i to denote the clause $\bigvee_{j=1}^{i_r} y_{i,j}$. Let Ψ_{rep} be the set of wff given by,

$$\Psi_{rep} = \{\Psi, C_1, \ldots, C_m, x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}$$

The *dialogue arena of formulae in* CNF is

$$\mathcal{A}_{\text{CNF}} = \bigcup_{n=1}^{\infty} \bigcup_{\Psi \in \Theta(n)} \{\{\Psi_{rep}\}\}$$

[3]For example, it is easy to show that the set of distinct protocols that could be defined using only *two* locutions and a *single element* context is not enumerable.

Thus, each different CNF, Ψ gives rise to the dialogue context whose elements are defined by Ψ_{rep}.

We note that $\Phi \in \mathcal{A}_{CNF}$ may be encoded as a word, $\beta(\Phi)$, over alphabet $\{-1, 0, 1\}$

$$\beta(\Phi) = 1^n 0\alpha \quad \text{with } \alpha \in \{-1, 0, 1\}^{nm}$$

where the i'th clause is described by the sub-word

$$\alpha_{(i-1)*n+1} \ldots \alpha_{i*n}$$

so that

$$\begin{array}{lll} \alpha_{(i-1)n+k} = -1 & \text{if} & \neg x_k \in C_i \\ \alpha_{(i-1)n+k} = 1 & \text{if} & x_k \in C_i \\ \alpha_{(i-1)n+k} = 0 & \text{if} & \neg x_k \notin C_i \text{ and } x_k \notin C_i \end{array}$$

It is thus immediate that given any word $w \in \{-1, 0, 1\}^*$ there is an algorithm that accepts w if and only if $w = \beta(\Phi)$ for some CNF Φ and this algorithm runs in $O(|w|)$ steps.

The basis for the dialogue schema we now define is found in the classic DPLL procedure for determining whether a well-formed propositional formula is satisfiable or not [5, 6]. Our protocol – the DPLL-dialogue protocol – is derived from the realisation of the DPLL-procedure on CNF formulae.

In describing this we assume some ordering

$$\langle \Phi_1, \Phi_2, \ldots, \Phi_k, \ldots \rangle$$

of the contexts in the arena $\mathcal{A}_{CNF}$.

DPLL-*Dialogue Schema*

The sequence of DPLL-*Dialogue Schema* – $\Delta_{DPLL} = \{\Delta_k^{DPLL}\}$ – is defined with contexts from the arena $\mathcal{A}_{CNF}$ as

$$\Delta_k^{DPLL} = \langle \mathcal{L}_{DPLL}, \mathcal{D}_{DPLL} = \langle \Pi_{DPLL}, \Sigma_{DPLL} \rangle, \Phi_k \rangle$$

where

$$\mathcal{L}_{DPLL} = \{\text{ASSERT,REBUT,PROPOSE,DENY,MONO,UNIT}\}$$

and the set Φ_k from $\mathcal{A}_{CNF}$ is,

$$\{\bigwedge_{i=1}^{m} C_i, C_1, \ldots, C_m, x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}$$

Recall that Ψ_k denotes the formula $\bigwedge_{i=1}^{m} C_i$, and C_i is the clause $\bigvee_{j=1}^{i_r} y_{i,j}$ from the context Φ_k. It will be convenient to regard a clause C both as a disjunction of literals and as a *set* of literals, so that we write $y \in C$ when C has the form $y \vee B$.

The protocol $\langle \Pi_{DPLL}, \Sigma_{DPLL} \rangle$ is defined through the following cases.

At any stage the commitment state – $\Sigma_{DPLL}(\delta)$ consists of a (possibly empty) subset of the clauses of Ψ_k and a (possibly empty) subset of the literals, subject to the condition that y and $\neg y$ are never simultaneously members of $\Sigma_{DPLL}(\delta)$. With the exception of {ASSERT,REBUT} the instantiated form of any locution involves a literal y.

Case 1: $\delta = \epsilon$ the empty dialogue fragment.

$$\begin{array}{lcl} \Pi_{DPLL}(\epsilon) & = & \{\text{ASSERT}(\Psi_k)\} \\ \Sigma_{DPLL}(\epsilon) & = & \emptyset \\ \Sigma_{DPLL}(\text{ASSERT}(\Psi_k)) & = & \{C_i : 1 \leq i \leq m\} \end{array}$$

In the subsequent development, y is a literal and

$$\begin{array}{lcl} Open(\delta) & = & \{C_i : C_i \in \Sigma_{DPLL}(\delta)\} \\ Lits(\delta) & = & \{y : y \in \Sigma_{DPLL}(\delta)\} \\ Single(\delta) & = & \{y : \neg y \notin Lits(\delta) \textit{ and } \exists C \in Open(\delta) \textit{ s.t.} \\ & & y \in C \textit{ and } \forall z \in C/\{y\}\ \neg z \in Lits(\delta)\} \\ Unary(\delta) & = & \{y : \neg y \notin Lits(\delta) \textit{ and} \\ & & \forall C \in Open(\delta)\ \neg y \notin C \textit{ and} \\ & & \exists C \in Open(\delta) \textit{ with } y \in C\} \\ Bad(\delta) & = & \{C_i : C_i \in Open(\delta) \textit{ and} \\ & & \forall y \in C\ \neg y \in Lits(\delta)\} \end{array}$$

Informally, $Open(\delta)$ indicates clauses of Ψ_k that have yet to be satisfied and $Lits(\delta)$ the set of literals that have been committed to in trying to construct a satisfying assignment to Ψ_k. Over the progress of a dialogue the literals in $Lits(\delta)$ may, if instantiated to **true**, result in some clauses being reduced to a *single* literal – $Single(\delta)$ is the set of such literals. Similarly, either initially or following an instantantiation of the literals in $Lits(\delta)$ to **true**, the set of clauses in $Open(\delta)$ may be such that some variables occurs only positively among these clauses or only negated. The corresponding literals form the set $Unary(\delta)$. Finally, the course of committing to various literals may result in a set that contradicts all of the literals in some clause: thus this set cannot constitute a satisfying instantiation: the set of clauses in $Bad(\delta)$ if non-empty indicate that this has occurred. Notice that the definition of $Single(\delta)$ admits the possibility of a literal y and its negation being in this set: a case which cannot lead to the set of literals in $Lits(\delta)$ being extended to a satisfying set. Thus we say that the literal set $Lits(\delta)$ is a *failing set* if either $Bad(\delta) \neq \emptyset$ or for some y, $\{y, \neg y\} \subseteq Single(\delta)$.

Recognising that $\Sigma_{DPLL}(\delta) = Open(\delta) \cup Lits(\delta)$ it suffices to describe changes to $\Sigma_{DPLL}(\delta)$ in terms of changes to $Open(\delta)$ and $Lits(\delta)$.

Case 2: $\delta \neq \epsilon$, $Open(\delta) = \emptyset$

$$\Pi(\delta) = \emptyset$$

Case 3: $\delta \neq \epsilon$, $Open(\delta) \neq \emptyset$, $Lits(\delta)$ is *not a failing set*.
There are a number of sub-cases depending on $\Sigma_{DPLL}(\delta)$
Case 3.1: $Single(\delta) \neq \emptyset$.

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{UNIT}(y) : y \in Single(\delta)\} \\ Open(\delta \cdot \text{UNIT}(y)) & = & Open(\delta)/\{C : y \in C\} \\ Lits(\delta \cdot \text{UNIT}(y)) & = & Lits(\delta) \cup \{y\} \end{array}$$

Case 3.2: $Single(\delta) = \emptyset, Unary(\delta) \neq \emptyset$

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{MONO}(y) : y \in Unary(\delta)\} \\ Open(\delta \cdot \text{MONO}(y)) & = & Open(\delta)/\{C : y \in C\} \\ Lits(\delta \cdot \text{MONO}(y)) & = & Lits(\delta) \cup \{y\} \end{array}$$

Case 3.3: $Single(\delta) = Unary(\delta) = \emptyset$
Since $Bad(\delta) = \emptyset$ and $Open(\delta) \neq \emptyset$, instantiating the literals in $Lits(\delta)$ will neither falsify nor satisfy Ψ_k. It follows that the set

$$\begin{array}{lcl} Poss(\delta) & = & \{y : y \notin Lits(\delta),\ \neg y \notin Lits(\delta) \textit{ and} \\ & & \exists C \in Open(\delta) \textit{ with } y \in C\} \end{array}$$

is non-empty. We note that since $Unary(\delta) = \emptyset$, $y \in Poss(\delta)$ if and only if $\neg y \in Poss(\delta)$. This gives,

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{PROPOSE}(y) : y \in Poss(\delta)\} \\ Open(\delta \cdot \text{PROPOSE}(y)) & = & Open(\delta)/\{C : y \in C\} \\ Lits(\delta \cdot \text{PROPOSE}(y)) & = & Lits(\delta) \cup \{y\} \end{array}$$

This completes the possibilities for Case 3. We are left with,
Case 4: $\delta \neq \epsilon$, $Lits(\delta)$ *is a failing set.*
Let $\delta = \text{ASSERT}(\Psi_k) \cdots \mu_t$

Given the cases above, there are only three utterances that μ_t could be:

$$\mu_t \in \{\text{ASSERT}(\Psi_k),\ \text{PROPOSE}(y),\ \text{DENY}(y)\}$$

Case 4.1: $\mu_t = \mu_1 = \text{ASSERT}(\Psi_k)$
Sinces $Lits(\text{ASSERT}(\Psi_k)) = \emptyset$, Ψ_k either contains an *empty clause* (one containing no literals), or for some x both (x) and $(\neg x)$ are clauses in Ψ_k.[4] In either case Ψ_k is "trivially" unsatisfiable, giving

$$\begin{array}{lcl} \Pi_{DPLL}(\text{ASSERT}(\Psi_k)) & = & \{\text{REBUT}(\Psi_k)\} \\ \Sigma_{DPLL}(\text{ASSERT}(\Psi_k) \cdot \text{REBUT}(\Psi_k)) & = & \emptyset \\ \Pi_{DPLL}(\text{ASSERT}(\Psi_k) \cdot \text{REBUT}(\Psi_k)) & = & \emptyset \end{array}$$

Case 4.2: $\mu_t = \text{PROPOSE}(y)$

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{DENY}(y)\} \\ Open(\delta \cdot \text{DENY}(y)) & = & Open(\mu_1 \cdots \mu_{t-1})/\{C : \neg y \in C\} \\ Lits(\delta \cdot \text{DENY}(y)) & = & Lits(\mu_1 \cdots \mu_{t-1}) \cup \{\neg y\} \end{array}$$

Notice this corresponds to a 'back-tracking' move under which having failed to complete a satisfying set by employing the literal y, its negation $\neg y$ is tried instead.
Case 4.3: $\mu_t = \text{DENY}(y)$
Consider the sequence of utterances given by

$$\eta = \mu_2 \cdot \mu_3 \cdots \mu_{t-1} \cdot \mu_t = \text{DENY}(y)$$

We say that η is *unbalanced* if there is a position p such that $\mu_p = \text{PROPOSE}(z)$ with $\text{DENY}(z) \notin \mu_{p+1} \cdots \mu_t$ and *balanced* otherwise. If η is unbalanced let $index(\eta)$ be the *highest* such position for which this holds (so that $p < t$).

We now obtain the final cases in our description.
Case 4.3(a): η is unbalanced with $index(\eta)$ equal to p.

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{DENY}(y) : \mu_p = \text{PROPOSE}(y)\} \\ Open(\delta \cdot \text{DENY}(y)) & = & Open(\mu_1 \cdots \mu_{p-1})/\{C : \neg y \in C\} \\ Lits(\delta \cdot \text{DENY}(y)) & = & Lits(\mu_1 \cdots \mu_{p-1}) \cup \{\neg y\} \end{array}$$

Thus this case corresponds to a 'back-tracking' move continuing from the "most recent" position at which a literal $\neg y$ instead of y can be tested.

Finally,
Case 4.3(b): η is balanced.

$$\begin{array}{lcl} \Pi_{DPLL}(\delta) & = & \{\text{REBUT}(\Psi_k)\} \\ \Sigma_{DPLL}(\delta \cdot \text{REBUT}(\Psi_k)) & = & \emptyset \\ \Pi_{DPLL}(\delta \cdot \text{REBUT}(\Psi_k)) & = & \emptyset \end{array}$$

We state the following without proof.

THEOREM 1. *In the following, δ is a dialogue fragment from $M^*_{\langle \mathcal{L}_{DPLL}, \Phi_k \rangle}$; Φ_k is a context from $\mathcal{A}_{CNF}$, and $N(\delta, \Phi_k)$ is the number of bits used to encode δ and Φ_k under some reasonable encoding scheme.*

1. *The problem of determining whether δ is a legal dialogue for the protocol $\mathcal{D}_{DPLL}$ in context Φ_k can be decided in $O(N(\delta, \Phi_k))$ steps.*
2. *The problem of determining whether δ is a* terminated *legal dialogue for the protocol $\mathcal{D}_{DPLL}$ in context Φ_k is decidable in $O(N(\delta, \Phi_k))$ steps.*
3. *For any $\psi \in \Phi_k$, the problem of determining whether $\psi \in \Sigma_{DPLL}(\delta)$ is decidable in $O(N(\delta, \Phi_k))$ steps.*
4. *For all contexts $\Phi_k \in \mathcal{A}_{CNF}$, the set of legal dialogues over Φ_k in the protocol $\mathcal{D}_{DPLL}$ is* finite.
5. *If δ is a terminated dialogue of $\mathcal{D}_{DPLL}$ in context Φ_k then $\Sigma_{DPLL}(\delta) \neq \emptyset$ if and only if Ψ_k is satisfiable. Furthermore, instantiating the set of literals in $Lits(\delta)$ to* **true***, yields a satisfying assignment for Ψ_k.*

Before analysing this protocol we review how it derives from the basic DPLL-procedure. Consider the description of this below.

DPLL-*Procedure*

Input: Set of clauses C
Set of Literals L

if $C = \emptyset$ **return true**. (SAT)
if any clause of C is empty
or C contains clauses (y) and $(\neg y)$ (for some literal y)
return false. (UNSAT)
if C contains a clause containing a single literal y
return DPLL$(C^{|y}, L \cup \{y\})$ (U)
if there is a literal y such that $\neg y$ does not occur in any clause (and y occurs in some clause)
return DPLL$(C^{|y}, L \cup \{y\})$ (M)
choose a literal y. (B)
if DPLL$(C^{|y}, L \cup \{y\})$
then **return true**
else **return** DPLL$(C^{|\neg y}, L \cup \{\neg y\})$ (FAIL).

For a set of clauses and literal, y, the set of clauses $C^{|y}$ is formed by removing all clauses, C_i for which $y \in C_i$ and deleting the literal $\neg y$ from all clauses C_j having $\neg y \in C_j$.

To test if $\Psi = \wedge_{i=1}^m C_i$ is satisfiable, the procedure is called with input $C = \{C_1, \ldots, C_m\}$ and $L = \emptyset$.

Lines (U) and (M) are the "unit-clause" and "monotone literal" rules which improve the run-time of the procedure: these are simulated by the UNIT and MONO locutions. Otherwise a literal is selected – at line (B) – to "branch" on: the PROPOSE locution; should the choice of branching literal FAIL to lead to a satisfying assignment, its negation is tested – the DENY locution. Each time a literal is set to **true**, clauses containing it can be deleted from the current set – the $Open(\delta)$ of the protocol; clauses containing its negation contain one fewer literal. Either all clauses will be eliminated (C is satisfiable) or an empty clause will result (the current set of literals chosen is not a satisfying assignment). When all choices have been exhausted the method will conclude that C is unsatisfiable.

The motivation for the form of the dialogue protocol Δ_k^{DPLL} is the connection between terminated dialogues in T_{DPLL} and *search trees* in the DPLL-procedure above.

DEFINITION 8. *Given a set of clauses C, a* DPLL–search tree *for C is a binary tree, S, recursively defined as follows: if $C = \emptyset$ or C conforms to the condition specified by* UNSAT *in the* DPLL*-procedure, then S is the empty tree, i.e. S contains no nodes. If y is a monotone literal or defines a unit-clause in C, then S comprises a root labelled y whose sole child is a* DPLL*-search tree for the set $C^{|y}$. If none of these four cases apply, S consists of a root labelled with the branching literal y chosen in line (*B*) with at most two children – one comprising a* DPLL*-search tree for the set $C^{|y}$, the other child – if the case (*FAIL*) arises – a* DPLL*-search tree for the set $C^{|\neg y}$.*

[4]Note that we distinguish the wff y (a *literal* used in Ψ_k) and (y) (a *clause* containing the *single* literal y) within the context Φ_k.

A DPLL*-search tree is* full *if no further expansion of it can take place (under the procedure above).*

The size *of a* DPLL*-search tree, S – $\nu(S)$ – is the total number of* edges[5] *in S. A full* DPLL*-search tree, S, is* minimum *for the set of clauses C, if given any full* DPLL*-search tree, R for C, $\nu(S) \leq \nu(R)$. Finally, a literal y is an* optimal branching literal *for a clause set C, if there is a minimum* DPLL*-search tree for C whose root is labelled y.*

We say a set of clauses, C, is *non-trivial* if $C \neq \emptyset$. Without loss of generality we consider only CNF-formulae, Ψ, whose clause set is non-trivial. Of course, during the evolution of the DPLL-procedure and the dialogue protocol $\mathcal{D}_{DPLL}$ sets of clauses which are trivial may result (this will certainly be the case is Ψ is satisfiable): our assumption refers *only* to the initial instance set.

THEOREM 2. *Let $\Psi = \wedge_{i=1}^{m} C_i$ be a* CNF*-formula over propositional variables $\langle x_1, \ldots, x_n \rangle$. Let $C(\Psi)$ and Φ_k be respectively the set of clauses in Ψ and the dialogue context from the arena $\mathcal{A}_{CNF}$ corresponding to Ψ, i.e. the set Ψ_{rep} above.*

1. *Given any full* DPLL*-search tree, S, for $C(\Psi)$ there is a legal terminated dialogue, $\delta_S \in T_{DPLL}$ for which,*

$$\delta_S \;=\; \text{ASSERT}(\Psi_k) \cdot \eta_S \cdot \mu$$

and $|\eta_S| = \nu(S)$, with μ being one of the locution types in {PROPOSE,UNIT,MONO,REBUT}.

2. *Given any terminated legal dialogue $\delta \;=\; \text{ASSERT}(\Psi_k) \cdot \eta \cdot \mu$, with*

$$\mu \in \{\text{REBUT}(\Psi_k), \text{PROPOSE}(y), \text{MONO}(y), \text{UNIT}(y)\}$$

there is a full DPLL*-search tree, S_δ having $\nu(S_\delta) = |\eta|$.*

PROOF. (Outline) We present the proof of Part 1 only. Let Ψ, $C(\Psi)$, and Φ_k be as in the Theorem statement. For Part 1, let S be any full DPLL-search tree for the clause set $C(\Psi)$. We obtain the result by induction on $\nu(S) \geq 0$.

For the inductive base, $\nu(S) = 0$, either S is the empty tree or S contains a single node labelled y. In the former instance, since Ψ is non-trivial it must be the case that Ψ is unsatisfiable (by reason of containing an empty clause or opposite polarity unit clauses). Choosing

$$\delta_S \;=\; \text{ASSERT}(\Psi_k) \cdot \eta_S \cdot \text{REBUT}(\Psi_k)$$

with $\eta_S = \epsilon$ is a legal terminated dialogue (Case 4.1) and $|\eta_S| = 0 = \nu(S)$.

When S contains a single node, so that $\nu(S) = 0$, let y be the literal labelling this. It must be the case that $C(\Psi)$ is *satisfiable* – it cannot hold that $C(\Psi)^{|y}$ *and* $C(\Psi)^{|\neg y}$ *both* yield empty search trees, since this would imply the presence of unit-clauses (y) and $(\neg y)$ in $C(\Psi)$.[6] Thus the literal y occurs in every clause of $C(\Psi)$. If y is a unit-clause, the dialogue fragment,

$$\delta_S \;=\; \text{ASSERT}(\Psi_k) \cdot \text{UNIT}(y)$$

is legal (Case 3.1) and terminated (Case 2). Fixing $\eta_S = \epsilon$ and $\mu = \text{UNIT}(y)$ gives $|\eta_S| = 0 = \nu(S)$ and $\delta = \text{ASSERT}(\Psi_k) \cdot \eta_S \cdot \mu$ a legal terminated dialogue. If y is not a unit clause, we obtain an identical conclusion using $\eta_S = \epsilon$ and $\mu = \text{MONO}(y)$ via Case 3.2 and Case 2.

Now, inductively assume, for some M, that if S_M is a DPLL-search tree for a set of clauses $C(\Psi)$, with $\nu(S_M) < M$ then there is a terminated legal dialogue, δ_{S_M}, over the corresponding context, Φ, with $\delta_{S_M} = \text{ASSERT}(\Psi) \cdot \eta_{S_M} \cdot \mu$ and $|\eta_{S_M}| = \nu(S_M)$.

Let S be a DPLL-search tree for $C(\Psi)$ with $\nu(S) = M \geq 1$. Consider the literal, y, labelling the root of S. Since $\nu(S) \geq 1$, the set $C(\Psi)^{|y}$ is non-empty. If $C(\Psi_k)$ contains a unit-clause, then (y) must be one such, thus S comprises the root labelled y and a single child, $S^{|y}$ forming a full DPLL-search tree for the (non-empty) clause set $C(\Psi)^{|y}$. It is obvious that $\nu(S^{|y}) < \nu(S) \leq M$, so from the Inductive Hypothesis, there is a legal terminated dialogue, $\delta^{|y}$ in the context formed by the CNF $\Psi_k^{|y}$. Hence,

$$\delta^{|y} = \text{ASSERT}(\Psi_k^{|y}) \cdot \eta^{|y} \cdot \mu$$

and $|\eta^{|y}| = \nu(S^{|y})$. From Case(3.1), the dialogue fragment

$$\delta_S = \text{ASSERT}(\Psi_k) \cdot \text{UNIT}(y) \cdot \eta^{|y} \cdot \mu$$

is legal and is terminated. Setting $\eta_S = \text{UNIT}(y) \cdot \eta^{|y}$, we obtain

$$|\eta_S| \;=\; 1 + |\eta^{|y}| \;=\; 1 + \nu(S^{|y}) \;=\; \nu(S)$$

A similar construction applies in those cases where y is a monotone literal – substituting the utterance $\text{MONO}(y)$ for UNIT(y) – and when y is a branching literal with exactly one child $S^{|y}$ – in this case, substituting the utterance $\text{PROPOSE}(y)$ for $\text{UNIT}(y)$.

The remaining case is when S comprises a root node labelled y with *two* children – $S^{|y}$ and $S^{|\neg y}$ – the former a full DPLL-search tree for the clause set $C(\Psi)^{|y}$, the latter a full DPLL-search tree for the set $C(\Psi)^{|\neg y}$. We use $\Phi^{|y}$ and $\Phi^{|\neg y}$ to denote the contexts in $\mathcal{A}_{CNF}$ corresponding to these CNF-formulae. As in the previous case, $\nu(S^{|y}) < \nu(S) = M$ and $\nu(S^{|\neg y}) < \nu(S) = M$. Invoking the Inductive Hypothesis, we identify legal terminated dialogues, over the respective contexts $\Phi^{|y}$ and $\Phi^{|\neg y}$

$$\begin{array}{lcl} \delta^{|y} & = & \text{ASSERT}(\Psi^{|y}) \cdot \eta^{|y} \cdot \mu^{|y} \\ \delta^{|\neg y} & = & \text{ASSERT}(\Psi^{|\neg y}) \cdot \eta^{|\neg y} \cdot \mu^{|\neg y} \end{array}$$

with $|\eta^{|y}| = \nu(S^{|y})$ and $|\eta^{|\neg y}| = \nu(S^{|\neg y})$.

We first note that the set $C(\Psi)^{|y}$ cannot be satisfiable – if it were the search-tree $S^{|\neg y}$ would not occur. We can thus deduce that $\mu^{|y} = \text{REBUT}(\Psi^{|y})$. Now consider the dialogue fragment, δ_S, from the context Φ_k

$$\text{ASSERT}(\Psi_k) \cdot \text{PROPOSE}(y) \cdot \eta^{|y} \cdot \text{DENY}(y) \cdot \eta^{|\neg y} \cdot \mu^{|\neg y}$$

Certainly this is a legal terminated dialogue via the Inductive hypothesis and Cases 4.2, 4.3(a–b). In addition, with

$$\eta_S = \text{PROPOSE}(y) \cdot \eta^{|y} \cdot \text{DENY}(y) \cdot \eta^{|\neg y}$$

we have

$$|\eta_S| \;=\; 2 + |\eta^{|y}| + |\eta^{|\neg y}| \;=\; 2 + \nu(S^{|y}) + \nu(S^{|\neg y}) \;=\; \nu(S)$$

so completing the Inductive proof of Part 1.

Part 2 can be demonstrated by a straightforward inductive argument on $|\eta| \geq 0$. □

COROLLARY 1. *An instance,*

$$U = \langle \Phi_k, \text{ASSERT}(\Psi_k), \text{PROPOSE}(y) \rangle$$

of the Optimal Utterance Problem for Δ_{DPLL} is accepted if and only if y is neither a unit-clause nor a monotone literal and y is an optimal branching literal for the clause set $C(\Psi_k)$.

[5] The usual definition of *size* is as the number of *nodes* in S, however, since S is a tree this value is exactly $\nu(S) + 1$.

[6] It should be remembered that *at most one* of $\{y, \neg y\}$ occurs in any clause.

PROOF. If y defines a unit-clause or monotone literal in Ψ_k then PROPOSE(y) is not a legal continuation of ASSERT(Ψ_k). The corollary is now an easy consequence of Theorem 2: suppose that

$$\delta \ = \ \text{ASSERT}(\Psi_k) \cdot \text{PROPOSE}(y) \cdot \eta \cdot \mu_y$$

is a minimum length completion of ASSERT(Ψ_k), then Part 2 of Theorem 2 yields a full DPLL-search tree, R, for $C(\Psi_k)$ of size $1 + |\eta|$ whose root is labelled y. If R is not minimum then there is smaller full DPLL-search tree, S. From Part 1 of Theorem 2 this yields a legal terminated dialogue

$$\text{ASSERT}(\Psi_k) \cdot \mu_S \cdot \eta_S \cdot \mu$$

with

$$\nu(S) \ = \ |\mu_S \cdot \eta_S \cdot \mu| - 1 \ < \ |\text{PROPOSE}(y) \cdot \eta \cdot \mu_y| - 1 \ = \ \nu(R)$$

which contradicts the assumption that δ is a minimum length completion. □

We now obtain a lower bound on the complexity of OUP$^{(\Delta)}$ via the following result of Liberatore [15].

FACT 1. Liberatore ([15]) *Given an instance $\langle C, y\rangle$ where C is a set of clauses and y a literal in these, the problem of deciding whether y is an optimal branching literal for the set C is* NP*–hard* and CO-NP*–hard.*

THEOREM 3. *The Optimal Utterance in Δ Problem is* NP*–hard and* CO-NP*–hard.*

PROOF. Choose Δ as the sequence of schema $\{\Delta_k^{DPLL}\}$. From Corollary 1 an instance $\langle \Phi_k, \text{ASSERT}(\Psi_k), \text{PROPOSE}(y)\rangle$ is accepted in OUP$^{(\Delta)}$ if and only if y does not form a unit-clause of Ψ_k, is not a monotone literal, and is an optimal branching literal for the clause set $C(\Psi_k)$. We may assume, (since these are easily tested) that the first two conditions do not hold, whence it follows that decision methods for such instances of OUP$^{(\Delta)}$ yield decision methods for determining if y is an optimal branching literal for $C(\Psi_k)$. The complexity lower bounds now follow directly from Liberatore's results stated in Fact 1. □

4. CONCLUSION

The principal contentions of this paper are three-fold: firstly, in order for a dialogue protocol to be realised effectively in a multiagent setting, each agent must have the capability to determine what contribution(s) it must or should or can make to the discussion as it develops; secondly, in deciding which (if any) utterance to make, an agent should (ideally) take cognisance of the extent to which its utterance is 'optimal'; and, finally, the criteria by which an utterance is judged to be 'optimal' are *application dependent*. In effect, the factors that contributors take into consideration when participating in one style of dialogue, e.g. bargaining protocols, are *not* necessarily those that would be relevant in another style, e.g. persuasion protocols.

We have proposed one possible interpretation of "optimal utterance in persuasion protocols": that which leads to the debate terminating 'as quickly as possible'. There are, however, a number of "length-related" alternatives that may merit further study. We have already mentioned in passing the view explored in [10]. One drawback to the concept of "optimal utterance" as we have considered it, is that it presumes the protocol is "well-behaved" in a rather special sense: taking the aim of an agent in a persuasion process as "to convince others that a particular proposition is valid", the extent to which an agent is successful may depend on the 'final' commitment state attained. In the DPLL-protocol this final state is either *always* empty (if Ψ_k is not satisfiable) or *always* non-empty: the protocol is "sound" in the sense that conflicting interpretations of the final state are not possible. Suppose we consider persuasion protocols where there is an 'external' interpretation of final state, e.g. using a method of defining some (sequence) of mappings $\tau : \wp(\Phi_k) \rightarrow \{\textbf{true}, \textbf{false}, \bot\}$, so that a terminated dialogue, δ, with $\tau(\Sigma(\delta)) = \textbf{true}$ indicates that the persuading agent has successful demonstrated its desired hypothesis; $\tau(\Sigma(\delta)) = \textbf{false}$ indicates that its hypothesis is *not* valid; $\tau(\Sigma(\delta)) = \bot$ indicates that no conclusion can be drawn.[7] There are good reasons why we may wish to implement 'seemingly contradictory' protocols, i.e. in which the persuasion process for a given context Φ can terminate in any (or all) of **true**, **false** or $\bot$ states, e.g. to model concepts of cautious, credulous, and sceptical agent belief, cf. [24]. In such cases defining "optimal utterance" as that which can lead to a shortest terminated dialogue may not be ideal: the persuading agent's view of "optimal" is not simply to terminate discussion but to terminate in a **true** state; in contrast, "sceptical" agents may seek utterances that (at worst) terminate in the inconclusive $\bot$ state. We note that, in such settings, there is potentially an "asymmetry" in the objectives of individual agents – we conjecture that in suitably defined protocols and contexts with appropriately defined concepts of "optimal utterance" the decision problems arising are likely to prove at least as intractable as those for the basic variant we consider in Theorem 3.

A natural objection to the use of length-related measures to assess persuasion processes is that these do not provide any sense of how convincing a given discourse might be, i.e. that an argument can be presented concisely does not necessarily render it effective in persuading those to whom it is addressed. One problem with trying formally to capture concepts of persuasiveness is that, unlike measures based on length, this is a subjective measure: a reasoning process felt to be extremely convincing by one party may fail to move another. One interesting problem in this respect concerns modeling the following scenario. Suppose we have a collection of agents with differing knowledge and 'prejudices' each of whom an external agent wishes to persuade to accept some proposition, e.g. election candidates seeking to persuade a cross-section of voters to vote in their favour. In such settings one might typically expect contributions by the persuading party to affect the degree of conviction felt by members of the audience in different ways. As such the concept of an 'optimal' utterance might be better assessed in terms of proportionate increase in acceptance that the individual audience members hold after the utterance is made.

We conclude by mentioning two open questions of interest within the context of persuasion protocols and the optimal utterance problem in these. In practical terms, one problem of interest is, informally, phrased as follows: can one define "non-trivial" persuasion protocols for a "broad" collection of dialogue contexts within which the optimal utterance problem is tractable? We note that, it is unlikely that dialogue arenas encompassing the totality of all propositional formulae will admit such protocols, however, for those subsets which have efficient decision procedures e.g. Horn clauses, 2-CNF formulae, appropriate methods may be available. A second issue is to consider complexity-bounds for other persuasion protocols: e.g. one may develop schema for the arena $\mathcal{A}_{CNF}$ defined via the TPI–dispute mechanism of [27], the complexity (lower and upper bounds) of the optimal utterance problem in this setting is open,

[7]For example, game theorists in economics have considered the situation where two advocates try to convince an impartial judge of the truth or otherwise of some claim, e.g. [12].

although in view of our results concerning $\mathcal{D}_{DPLL}$ it is plausible to conjecture that the optimal utterance problem for $\mathcal{D}_{TPI}$ will also prove intractable.

5. REFERENCES

[1] T. J. M. Bench-Capon. Specification and implementation of Toulmin dialogue game. In J. C. Hage *et al.*, editor, *Legal Knowledge Based Systems*, pages 5–20. GNI, 1998.

[2] T. J. M. Bench-Capon, P. E. Dunne, and P. H. Leng. A dialogue game for dialectical interaction with expert systems. In *Proc. 12th Annual Conf. Expert Systems and their Applications*, pages 105–113, 1992.

[3] C. Cayrol, S. Doutre, and J. Mengin. Dialectical proof theories for the credulous preferred semantics of argumentation frameworks. In *Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001)*, pages 668–679. Lecture Notes in A.I., 2143, Springer, 2001.

[4] V. Conitzer and T. Sandholm. Complexity results about Nash equilibria. Technical Report CMU-CS-02-135, School of Computer Science, Carnegie-Mellon University, May 2002.

[5] M. Davis, G. Logemann, and D. Loveland. A machine program for theorem proving. *Communications of the ACM*, 5:394–397, 1962.

[6] M. Davis and H. Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7:201–215, 1960.

[7] F Dignum and M. Greaves (editors). *Issues in Agent Communication*. Springer-Verlag, 2000.

[8] S. Doutre and J. Mengin. Preferred extensions of argumentation frameworks: Query answering and computation. In *First Intern. Joint Conf. Automated Reasoning (IJCAR 2001)*, pages 272–288. Lecture Notes in A.I., 2083, Springer, June 2001.

[9] P.E. Dunne. *Computability Theory - Concepts and Applications*. Ellis-Horwood, 1991.

[10] P.E. Dunne. Prevarication in dispute protocols. Technical Report ULCS-02-025, Dept. of Comp. Sci., Univ. of Liverpool, 2002.

[11] P.E. Dunne and T.J.M. Bench-Capon. Two party immediate response disputes: Properties and efficiency. Technical Report ULCS-01-005, Dept. of Comp. Sci., Univ. of Liverpool, (to appear *Artificial Intelligence*)

[12] J. Glazer and A. Rubinstein. Debates and decisions: on a rationale of argumentation rules. *Games and Economic Behavior*, 36(2):158–173, 2001.

[13] T. F. Gordon. *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice*. Kluwer Academic, Dordrecht, 1995.

[14] S. Kraus. *Strategic negotiation in multiagent environments*. MIT Press, 2001.

[15] P. Liberatore. On the complexity of choosing the branching literal in DPLL. *Artificial Intelligence*, 116:315–326, 2000.

[16] A. R. Lodder. *Dialaw: On legal justification and Dialogue Games*. PhD thesis, Univ.of Maastricht, 1998.

[17] P. McBurney, R. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *J. Autonomous Agents and Multiagent Systems*, in press.

[18] P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and AI*, 32(1–4):125–169, 2001.

[19] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *J. Logic, Language and Information*, 11:315–334, 2002.

[20] P. McBurney and S. Parsons. Chance Discovery using dialectical argumentation. In T. Terano *et al.*, editors, *New Frontiers in Artificial Intelligence*, pages 414–424. Lecture Notes in A.I., 2253, Springer, 2001.

[21] P. McBurney, S. Parsons, and M. W. Johnson. When are two protocols the same? In M. P. Huget, F. Dignum and J. L. Koning, editors, *Agent Communications Languages and Conversation Policies*, Proc. AAMAS-02 Workshop, Bologna, Italy, 2002.

[22] P. McBurney, S. Parsons, and M. J. Wooldridge. Desiderata for agent argumentation protocols. In *Proc. First Intern. Joint Conf. on Autonomous Agents and Multiagent Systems*, pages 402–409. ACM Press, 2002.

[23] S. Parsons, C. A. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *J. Logic and Computation*, 8(3):261–292, 1998.

[24] S. Parsons, M. J. Wooldridge, and L. Amgoud. An analysis of formal inter-agent dialgoues. In *Proc. First Intern. Joint Conf. Autonomous Agents and Multiagent Systems*, pages 394–401. ACM Press, 2002.

[25] H. Prakken. *Logical Tools for Modelling Legal Argument*. Kluwer Academic, Dordrecht, 1997.

[26] C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proc. 3rd Intern. Conf. Multiagent Systems (ICMAS-98)*, pages 246–253. IEEE Press, 1998.

[27] G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proc. JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence.*, pages 224–238, Berlin, 2000. Lecture Notes in A.I., 1919, Springer.

[28] D. N. Walton and E. C. W. Krabbe. *Committment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, 1995.

The Complexity of Contract Negotiation

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Artificial Intelligence 164 (2005) 23–46
Artificial Intelligence
www.elsevier.com/locate/artint

The complexity of contract negotiation

Paul E. Dunne *, Michael Wooldridge, Michael Laurence

Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, United Kingdom

Received 7 February 2003

Available online 14 March 2005

Abstract

The use of software agents for automatic contract negotiation in e-commerce and e-trading environments has been the subject of considerable recent interest. A widely studied abstract model considers the setting in which a set of agents have some collection of resources shared out between them and attempt to construct a mutually beneficial optimal reallocation of these by trading resources. The simplest such trades are those in which a single agent transfers exactly one resource to another—so-called 'one-resource-at-a-time' or '*O-contracts*'. In this research note we consider the computational complexity of a number of natural decision problems in this setting.

Keywords: Negotiation; Multiagent systems; Computational complexity

1. Introduction

Mechanisms for automatically negotiating the allocation of resources in a group of agents form an important body of work within the multiagent systems field. Typical abstract models derive from game-theoretic perspectives in economics and among the issues that have been addressed are strategies that agents may use to negotiate, e.g., [9,12,14], and protocols for negotiation in agent societies, e.g., [2,10].

In this paper, we investigate the computational complexity of one of the most fundamental questions that may be asked of such a negotiation setting: that of whether a particular

* Corresponding author.
E-mail address: ped@csc.liv.ac.uk (P.E. Dunne).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2005.01.006

outcome is *feasible* under the assumption that negotiation participants will act rationally. The particular negotiation setting we consider—introduced by Sandholm [13]—relates to the reallocation of resources amongst agents. The idea is that, starting from some initial allocation, agents can negotiate to transfer resources between themselves to their mutual benefit. At each stage of negotiation, agents make deals by transferring resources to other agents, and receiving resources in return. The feasibility question in this setting may be informally understood as follows.

> Given some initial allocation P^s of resources to agents, and some potential final allocation P^t, is there a sequence of deals that will be individual rational to all involved, such that at the end of this sequence of deals, the allocation P^t will be realised?

It could be argued that a *positive* answer to this question does not imply that negotiation *will* be successful, as it merely implies the existence of an individual rational sequence of deals to get from P^s to P^t. The agents in question may have their own (perhaps irrational) reasons for rejecting some deals in this sequence. Moreover, unless the feasibility checking process is constructive, the agents may not be able to find the desired sequence of deals. A *negative* answer, however, surely rules out any chance of getting from P^s to P^t: for every possible sequence of deals realising this reallocation, some agent would suffer in the course of its implementation, and would therefore reject it.

Our main result is to show that this problem—and a number of natural variations of it—is NP-hard. We also investigate the complexity of a number of related problems: for example, we show that the problem of determining whether a particular allocation is Pareto Optimal is co-NP-complete.

2. Preliminary definitions

The scenario that we are concerned with is encapsulated in the following definition.

Definition 1. A *resource allocation setting* is defined by a triple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ where

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\}; \qquad \mathcal{R} = \{r_1, r_2, \ldots, r_m\}$$

are, respectively, a set of (at least two) agents and a collection of (non-shareable) resources. A *utility function*, u, is a mapping from subsets of $\mathcal{R}$ to rational values. Each agent $A_i \in \mathcal{A}$ has associated with it a particular utility function u_i, so that $\mathcal{U}$ is $\langle u_1, u_2, \ldots, u_n \rangle$. An *allocation* P of $\mathcal{R}$ to $\mathcal{A}$ is a partition $\langle P_1, P_2, \ldots, P_n \rangle$ of $\mathcal{R}$. The utility function, u_i, is *monotone* if $u_i(S) \leqslant u_i(T)$ whenever $S \subseteq T$. The value $u_i(P_i)$ is called the *utility* of the resources assigned to A_i.

Starting from some initial allocation—P_0—individual agents negotiate in an attempt to improve the utility of their holding. A number of interpretations have been proposed in order to define what constitutes a 'sensible' transfer of resource from both an individual agent's viewpoint and from the perspective of the overall allocation. Thus in negotiating a

change from an allocation P_i to Q_i (with $P_i, Q_i \subseteq \mathcal{R}$ and $P_i \neq Q_i$) there are three possible outcomes for the agent A_i:

- $u_i(P_i) < u_i(Q_i)$ A_i values the allocation Q_i as superior to P_i;
- $u_i(P_i) = u_i(Q_i)$ A_i is indifferent between P_i and Q_i; and
- $u_i(P_i) > u_i(Q_i)$ A_i is worse off after the exchange.

In a setting in which agents are self-interested, in order for an agent to accept an exchange with the last outcome, the notion of a *pay-off* function is used: in order to accept the new allocation, A_i receives some payment sufficient to compensate for the resulting loss in utility. Of course, such compensation must be made by other agents in the system who in providing it do not wish to pay in excess of any gain in resource. In defining notions of 'pay-off', the interpretation is that in any transaction each agent A_i makes a payment, π_i: if $\pi_i < 0$ then A_i is given $-\pi_i$ in return for accepting a contract; if $\pi_i > 0$ then A_i contributes π_i to the amount to be distributed among those agents whose pay-off is negative. Formally, such a notion of 'sensible transfer' is captured by the concept of *individual rationality*.

Definition 2. Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ be a resource allocation setting. A *deal* is a pair $\langle P, Q \rangle$ where $P = \langle P_1, \ldots, P_n \rangle$ and $Q = \langle Q_1, \ldots, Q_n \rangle$ are distinct partitions of $\mathcal{R}$. We use δ to denote an arbitrary deal. The effect of implementing the deal $\langle P, Q \rangle$ is that the allocation of resources specified by P is replaced with that specified by Q.

A deal $\langle P, Q \rangle$ is said to be *individually rational* (IR) if there is a *pay-off vector* $\pi = \langle \pi_1, \pi_2, \ldots, \pi_n \rangle$ satisfying,

(a) $\sum_{i=1}^{n} \pi_i = 0$.
(b) $u_i(Q_i) - u_i(P_i) > \pi_i$, for each agent A_i, *except that* π_i is allowed to be 0 if $P_i = Q_i$, i.e., should the deal (P, Q) leave the agent A_i with no change in its resource then it is not *required* that A_i be rewarded (have $\pi_i < 0$).

Definition 2 captures one view of a deal being 'sensible' with respect to the perspective of single agents. We require also concepts of '*global*' optimality. We consider two commonly used versions of this: Pareto Optimality and (Utilitarian) Social Welfare.

Definition 3. Let P be an allocation of $\mathcal{R}$ among $\mathcal{A}$. The *utilitarian social welfare* resulting from P, denoted $\sigma_u(P)$, is given by $\sum_{i=1}^{n} u_i(P_i)$.

The allocation P is *Pareto optimal* if for all allocations Q differing from P, it holds

$$\left(\bigvee_{i=1}^{n} \left[u_i(Q_i) > u_i(P_i) \right] \right) \Rightarrow \left(\bigvee_{i=1}^{n} \left[u_i(Q_i) < u_i(P_i) \right] \right). \tag{1}$$

Thus a Pareto optimal allocation is one in which no agent can attain better than its current utility except at the cost of leaving some agent worse off.

We make frequent use of the following result throughout the remainder of the paper.

Fact 4 [7]. *A deal $\langle P, Q \rangle$ is IR if and only if $\sigma_u(Q) > \sigma_u(P)$.*

In a typical application it is unlikely that an initial allocation P_0 to $\mathcal{A}$ will either maximise social welfare or be Pareto optimal, thus the agents involved seek to find a sequence of deals that will terminate in an optimal allocation. Given the setting it is clearly the case that there are allocations P_{opt} and Q_{opt} with the properties that $\sigma_u(P_{opt})$ maximises social welfare and for which Q_{opt} is Pareto optimal—of course, P_{opt} and Q_{opt} may not be unique. If the object is to maximise social welfare then clearly the deal $\langle P_0, P_{opt}\rangle$ will achieve this in a single round. It is unreasonable, however, to view such a deal as a viable solution: although always IR (if it represents a strict increase of social welfare) it is questionable whether it could be identified as the *first and only* deal required. The total number of possible allocations is n^m, and so for moderately large numbers of resources (m) there are too many feasibly to enumerate (even when $n = 2$). In addition, it may not be possible to implement the optimising contract in a *single* transaction even if only two agents are involved: the environment in which the trading process is implemented may not be suited to handling transactions in which large numbers of resources are involved; similarly, the protocol used for negotiation and contract description may not allow arbitrarily large numbers of resources to be dealt with.

In order to develop a realistic framework for negotiation, Sandholm [13] (using Smith's Contract-Net model [16]), presents a number of classes of *contract type*. In this article we are concerned with the following of these.

Definition 5 [13]. Let $\delta = \langle P, Q\rangle$ be a deal involving an allocation of $\mathcal{R}$ among $\mathcal{A}$. We say that δ is a *cluster contract* (C-contract) if there are distinct agents A_i and A_j for which,

(C1) $P_k = Q_k$ if and only if $k \notin \{i, j\}$.
(C2) There is a unique (non-empty) set S for which $Q_i = P_i \cup S$ and $Q_j = P_j \setminus S$ (with $S \subseteq P_j$) or $Q_j = P_j \cup S$ and $Q_i = P_i \setminus S$ (with $S \subseteq P_i$).

Thus a C-contract involves one agent transferring a subset of its allocation to another agent (without receiving any subset of resources in return).

The definition of C-contract permits an arbitrarily large number of resources to be transferred from one agent to another in a single deal. For the class of contracts of interest in our subsequent results, we wish to impose a bound on the maximum number of resources that can be moved in one deal. We thus introduce the notion of $C(k)$-*contracts*.

Definition 6. For a resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ and value $k \leqslant m = |\mathcal{R}|$, we say that δ is a *k-bounded cluster contract*, ($C(k)$-contract) if δ is a C-contract in which S—the set of resources transferred—contains *at most* k elements. When $k = 1$, we use the term *one contract* (O-contract): the name given to such deals in [13].

We recall that a $C(k)$-contract $\langle P, Q\rangle$ will be IR if and only if $\sigma_u(Q) > \sigma_u(P)$.

A sequence of deals $\Delta = \langle \delta_1, \delta_2, \ldots, \delta_t\rangle$ for which $\delta_i = \langle Q_{i-1}, Q_i\rangle$ is called a *contract path* realising the deal $\langle Q_0, Q_t\rangle$. The *length* of a contract path is the total number of deals comprising it. Given a predicate Φ over deals, we say that a contract path Δ is a *Φ-path* if $\Phi(\delta_i)$ is true of every deal δ_i within Δ.

Our main results concern Φ-paths where $\Phi(\delta)$ is the predicate which is true if and only if δ is an individually rational $C(k)$-contract. In the case of $k = 1$, i.e., IR O-contracts, such paths are attractive from an implementation viewpoint since these only involve agent-to-agent negotiation concerning a single resource at a time. In addition, starting from a given allocation, the number of O-contracts that are consistent with it is exactly $m(n-1)$, as opposed to n^m possible allocations. Thus heuristic methods may be able to find improved allocations by exploring the search space through O-contracts alone.

Appealing as the latter approach is, there are, nevertheless, problems associated with it. The following results were established by Sandholm [13].

Fact 7. *Let P_0 be any initial allocation of $\mathcal{R}$ to $\mathcal{A}$ and P_t be any other allocation.*

(a) *The deal $\langle P_0, P_t\rangle$ can always be realised by a contract path in which every deal is an O-contract.*
(b) *There are resource allocation settings, $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ within which there are IR deals $\langle P_0, P_t\rangle$ that cannot be realised by any IR C-contract path.*

We note that Fact 7(b) holds even if we are concerned with settings involving only two agents and the allocation P_t concerned is one that maximises social welfare.

In total, IR C-contracts (and thereby also the more restricted IR $C(k)$ and IR O-contracts) in themselves may not suffice to form an IR contract-path realising a specific deal.

In this paper we are concerned with the following decision problem:

Definition 8. The decision problem *IR-k-path* (IR^k) is given by
Instance: A 5-tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^s, P^t\rangle$ in which $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ is a resource allocation setting, $P^{(s)}$ and $P^{(t)}$ are allocations of $\mathcal{R}$ to $\mathcal{A}$ in which $\sigma_u(P^{(t)}) > \sigma_u(P^{(s)})$.
Question: Is there an IR $C(k)$-contract path that realises the deal (P^s, P^t)?

It is important to note that the value k (which restricts the number of resources in a cluster contract), does *not* form part of an *instance* of IR^k.

In keeping with the use of the term O-contract for $C(1)$-contract, we denote the decision problem IR^1 by IRO.

The main results of this article concern IR^k when k is *constant* and IR^k when the cluster size (k) is a predefined function of the number of resources. Specifically we prove the following:

(a) IR^k is NP-hard for all constant values of k. This holds even when $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ is a setting comprising two agents. The special case IRO remains NP-hard when both utility functions are monotone.
(b) For $k : \mathbb{N} \to \mathbb{N}$, satisfying $k(m) \leqslant m/3$, $\text{IR}^{k(m)}$ is NP-hard, again even in the case of resource allocation settings involving exactly two agents.
(c) $\text{IR}^{m/2}$ is NP-hard, again even in the case of resource allocation settings involving exactly two agents.

Our proofs of these results are given in Theorems 12–15.

We first note that the result of Theorem 15 does *not* imply (from the proof presented) either of the preceding theorems. It may seem to be the case that, when $h < k$, a lower bound on the complexity of IR^k implies a similar lower bound on the complexity of IR^h by virtue of the fact that within any resource allocation setting, all IR $C(h)$-contracts are also IR $C(k)$-contracts. As we shall, however, illustrate in proving (c), it is *not* necessarily the case that we can deduce IR^h to be NP-hard from a proof that IR^{h+k} is so: in order for this to hold, the construction used in demonstrating the latter must be such that any positive instances formed by the reduction to IR^{h+k} admit IR $C(h)$-contract paths. In the case of Theorem 14, while it is the case that our proof subsumes the result of Theorem 12, the construction for the latter case is rather less involved and has the additional advantage that the extension to monotone utility functions with IRO follows easily. For this reason, we have presented separate proofs of these results.

Before proceeding, we address one issue that is raised by Fact 7. Consider the following argument deriving from this fact.

(a) *Every* deal $\langle P_0, P_t\rangle$ can be realised by a sequence of O-contracts.
(b) There are IR deals which *cannot* be realised by a sequence of IR C-contracts.
(c) Therefore, to implement any IR deal $\langle P_0, P_t\rangle$ why not use an O-contract path some of whose constituent deals may fail to be IR?

In other words, why might it be necessary for *every* deal to be IR?

One answer to this question is offered by the scenario, outlined in [4], that we now describe. We observe that the issue underlying this argument is relevant with respect to any class of restricted contract types, i.e., the fact that O-contracts are referred to is purely for illustrative purposes. For simplicity, let us assume that we have a resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ involving exactly two agents $\{A_1, A_2\}$. These negotiate an allocation of $\mathcal{R}$ working with the following protocol.

> A reallocation of resources is agreed over a sequence of stages. Each stage consists of A_1 issuing a proposal to A_2 of the form (buy, r, p), offering to purchase r from A_2 for a payment of p; or $(sell, r, p)$, offering to transfer r to A_2 in return for a payment p. The response from A_2 is simply *accept* (following which the exchange is implemented) or *reject*. A final allocation is fixed either when A_1 is 'satisfied' or as soon as A_2 *rejects* any offer.

This is, of course, a very simple negotiation setting; however, consider its operation when A_1 wishes to bring about an allocation P_t and can thus devise a plan—a sequence of O-contracts—to realise this from an initial allocation P_0.

While A_2 *could* be better off if P_t is realised, it may be the case that the only proposals A_2 will accept are those under which it does not lose, i.e., A_2 is not prepared to suffer a short-term loss even if it is suggested that a long-term gain will result. Thus if some agents are sceptical about the *bona fides* of others then they will be inclined to accept *only* deals from which they can perceive an *immediate* benefit, i.e., those which are individually rational.

There are several reasons why an agent may embrace such attitudes within the schema outlined: once a deal has been implemented A_2 may lose utility but no further proposals are made by A_1 so that its loss is 'permanent'. We note that even if we enrich the basic protocol so that A_1 can describe P_t to A_2 *before* any formal exchange of resources takes place, if $\langle P_0, P_t\rangle$ is implemented by an O-contract path (via the sequence of stages outlined), A_2 may still reject offers under which it suffers a loss, since it is unwilling to rely on the subsequent O-contracts that would ameliorate its loss actually being proposed.[1]

Although the position taken by A_2 in the setting just described may appear unduly cautious, we would claim that it clearly reflects actual behaviour in certain arenas. In contexts other than automated allocation and negotiation models in multiagent systems, there are many examples of actions by individuals where promised long-term gains are insufficient to engender the acceptance of short term loss, e.g., 'chain letter' schemes although having a natural lifetime bounded by the size of the population in which they circulate, typically break down before this is reached. Despite the possibility of significant gain after a temporary loss, recipients may be disinclined to invest the expense requested to propagate the chain: such behaviour is not seen as overly sceptical and cautious. In the same way, the 'rational' response to the widespread e-mail fraud by which one is asked to furnish bank account details and working capital in order to facilitate the release of significant funds in return for a percentage of these, is to ignore the request. As a final example, it is considered standard practice to delete without reading, unexpected e-mail attachments regardless of what incentives to open such may be promised by the accompanying message text.

In summary, the critical question underpinning such views is this: in a reallocation of resources conducted over a sequence of stages, should either agent suffer a loss in utility why should they have any 'confidence' that this loss will eventually be reversed? It is inevitable, in view of Fact 7(b) that there will sometimes be IR deals which, if implemented by a sequence of unrestricted O-contracts, will lead to such a loss for one agent.

In the scenario we have described, an agent A_1 wishing to realise an IR deal $\langle P_0, P_t\rangle$ with an extremely cautious agent A_2 faces the following dilemma: whether to formulate a plan to realise $\langle P_0, P_t\rangle$, e.g., an O-contract path, regardless of whether this path is IR; or whether to try and realise $\langle P_0, P_t\rangle$ by an IR O-contract path. In favour of the first option is the fact that such a plan can *always* be formulated; a problem will be, however, that the plan may never be implemented in full: A_2 may reject deals under which it suffers a loss or A_1 may suffer a loss which is never put right. The second alternative—construct an IR O-contract path—has in its favour the fact that neither agent has a *rational* motive to refrain from making or accepting offers until the allocation P_t has been effected. The drawback, however, is that it may not be possible to construct such a plan.

Nevertheless, it would seem reasonable for A_1, before resorting to adopting an arbitrary O-contract path, at least to determine if some IR O-contract path (or, more generally, some IR $C(k)$-contract path) does exist. One consequence of our results is that such an approach is unlikely to be computationally feasible.

[1] We note that even if A_1 attempts to construct an ordering under which any 'irrational' deal reduces the value of its own holding, there is one problem: A_2 may reject subsequent offers after the 'irrational' deals so that A_1 is worse off.

The next section of this article presents these results with conclusions and open questions raised in the final section.

3. Complexity results

Before proceeding with our results we describe our representation for typical instances in which resource allocation settings $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ feature. The key issue here concerns the collection of utility functions $\mathcal{U}$ and how these should be encoded. A form in which the value attached to each subset of $\mathcal{R}$ is explicitly provided will result in an instance occupying space exponential in $|\mathcal{R}|$ and would not be considered reasonable in practice. On the other hand, using some encoding of $\mathcal{U}$ as a set of *Turing machine programs*, $\mathcal{M}$ say, it becomes necessary to assume certain properties in interpreting their computational behaviour, e.g., that the value of $u_i(S)$ as returned by the program M_i is defined from the content of M_i's tape after exactly some specified number of moves such as $|\mathcal{R}|$ since without such it would not be possible to establish membership in NP (or, indeed, any other complexity class).

Ideally, we wish a representation, $\rho(u)$, of the utility function $u : 2^{\mathcal{R}} \to \mathbb{Q}$ to satisfy the following informally phrased criteria:

(a) $\rho(u)$ is 'concise' in the sense that the length, e.g., number of bits, used by $\rho(u)$ to describe the utility function u within an instance is 'comparable' with the time taken by an optimal program that computes the value of $u(S)$.
(b) $\rho(u)$ is 'verifiable', i.e., given some binary word, w, there is an efficient algorithm that can check whether w corresponds to $\rho(u)$ for *some* u.
(c) $\rho(u)$ is 'effective', i.e., given $S \subseteq \mathcal{R}$, the value $u(S)$ can be efficiently computed from the description $\rho(u)$.

It is, in fact, possible to identify a representation form that satisfies all three of these criteria: we represent each member of $\mathcal{U}$ in a manner that does not *require* explicit enumeration of each subset of $\mathcal{R}$ and allows (a) to be met; uses a 'program' form whose syntactic correctness can be efficiently verified, hence satisfying (b); and for which termination in time linear in the program length is guaranteed, so meeting the condition set by (c). The class of programs employed are the so-called *straight-line programs*, which have a natural correspondence with combinational logic networks [3].

Definition 9. An (m, s)-*combinational network* C is a directed acyclic graph in which there are m *input nodes*, Z_m, labelled $\langle z_1, z_2, \ldots, z_m \rangle$ all of which have in-degree 0. In addition, C has s *output nodes*, called the *result vector*. These are labelled $\langle t_{s-1}, t_{s-2}, \ldots, t_0 \rangle$, and have out-degree 0. Every other node of C has in-degree at most 2 and out-degree at least 1. Each non-input node (*gate*) is associated with a Boolean operation of at most two arguments.[2] We use $|C|$ to denote the number of *gate* nodes in C. Any Boolean instantiation

[2] In practice, we can restrict the Boolean operations employed to those of binary conjunction ($\wedge$), binary disjunction ($\vee$) and unary negation ($\neg$).

of the input nodes to $\alpha \in \langle 0,1\rangle^m$ naturally induces a Boolean value at each gate of C: if h is a gate associated with the operation θ, and $\langle g_1, h\rangle$, $\langle g_2, h\rangle$ are edges of C then the value $h(\alpha)$ is $g_1(\alpha)\theta g_2(\alpha)$. Hence α induces some s-tuple $\langle t_{s-1}(\alpha), \ldots, t_0(\alpha)\rangle \in \langle 0,1\rangle^s$ at the result vector. For the (m,s)-combinational network C and $\alpha \in \langle 0,1\rangle^m$, this s-tuple is denoted by $C(\alpha)$.

Although often considered as a model of parallel computation, (m,s)-combinational networks yield a simple form of sequential program—straight-line programs—as follows. Let C be an (m,s)-combinational network to be transformed to a straight-line program, $\text{SLP}(C)$, that will contain exactly $m + |C|$ lines. Since C is directed and acyclic it may be topologically sorted, i.e., each gate, g, given a unique integer label $\tau(g)$ with $1 \leqslant \tau(g) \leqslant |C|$ so that if $\langle g, h\rangle$ is an edge of C then $\tau(g) < \tau(h)$. The line l_i of $\text{SLP}(C)$ evaluates the input z_i if $1 \leqslant i \leqslant m$ and the gate for which $\tau(g) = i - m$ if $i > m$. The gate labelling means that when g with inputs g_1 and g_2 is evaluated at $l_{m+\tau(g)}$ since g_i is either an input node or another gate its value will have been determined at l_j with $j < m + \tau(g)$.

Definition 10. Let $\mathcal{R}$ be as previously with $|\mathcal{R}| = m$, and u a mapping from subsets of $\mathcal{R}$ to rational values, i.e., a utility function. The (m,s)-network C^u is said to *realise* the utility function u if: for every $S \subseteq R$ with α_S the instantiation of Z_m by $z_i = 1$ if and only if $r_i \in S$, it holds

$$u(S) = \frac{val(C(\alpha_S))}{m}$$

where for $\beta = \langle \beta_{s-1}, \beta_{s-2}, \ldots, \beta_0\rangle \in \langle 0,1\rangle^s$, $val(\beta)$ is the whole number[3] whose s-bit binary expansion is β, i.e.,

$$val(\beta) = \sum_{i=0}^{s-1} \beta_i * 2^i,$$

where β_i is treated as the appropriate integer value from $\{0,1\}$.

These ideas allow any utility function u_i in $\mathcal{U}$ to be encoded using an appropriate (m,s_i)-combinational network, $C^{(i)}$ in such a way that $u_i(S)$ can be evaluated in time linear in the number of nodes in $C^{(i)}$ by determining the value of each gate under the related instantiation α_S and then dividing this value by m.

We give some concrete examples of this approach in the proof of Theorem 11. These are primarily intended to illustrate its feasibility and, having presented these, we will not complicate subsequent proofs with similarly detailed constructions. Regarding such constructions with respect to (a) of the representation criteria given, we note as a consequence of the simulations presented in [8,15] (see, e.g., Dunne [3, pp. 28–36]), that any deterministic algorithm with worst-case run-tine, $T(n)$ can be translated into a combinational

[3] Although this definition assumes utility functions to have non-negative values, were it the case that some function with $u(S) < 0$ was to be represented we can achieve this by using an additional output bit, $t_\pm$ to flag whether $val(C(\alpha))$ should be treated as positive ($t_\pm = 0$) or negative ($t_\pm = 1$).

network of size $T(n)\log T(n)$. It follows that from a high-level *algorithmic* description of how u_i is computed, an appropriate combinational network can be built.

The decision problem IR^k concerns the existence of a suitable contract path from one allocation to another having greater social welfare. For completeness, it is useful to present three results concerning the existence of resource allocations meeting particular criteria. These problems are respectively,

Welfare Improvement (WI)
Instance: A tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P\rangle$ where $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{U}$ are as before, and P is an allocation.
Question: Is there an allocation Q for which $\sigma_u(Q) > \sigma_u(P)$?

Welfare Optimisation (WO)
Instance: A tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, K\rangle$ where $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{U}$ are as before, and K is a rational number.
Question: Is there an allocation P for which $\sigma_u(P) \geqslant K$?

Pareto Optimal (PO)
Instance: A tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P\rangle$ as for WI.
Question: Is the allocation P Pareto optimal?

Kraus [9, p. 43] proves NP-hardness of a weaker form of the problem WO, whereby in addition to the total social welfare having to attain some specified value the allocation must be such that each agent accrues some designated guaranteed utility.

Theorem 11. *Even if* $|\mathcal{A}| = 2$ *and the utility functions are monotone*

(a) WI *is* NP*-complete.*
(b) WO *is* NP*-complete.*
(c) PO *is* CO-NP*-complete.*

Proof. We first demonstrate that the three problems are in the classes stated, recalling that the utility functions $\mathcal{U}$ are encoded by (m, s_i)-combinational networks $C^{(i)}$ as described in Definition 10. For (a), given an instance $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P\rangle$ of WI simply non-deterministically guess an allocation $Q = \langle Q_1, \ldots, Q_n\rangle$ and compute

$$\sigma_u(Q) = \sum_{i=1}^{n} \frac{val(C^{(i)}(\alpha_{Q_i}))}{|\mathcal{R}|}$$

accepting if this exceeds $\sigma_u(P)$. For (b) a similar approach is used with an instance accepted if the guessed allocation Q has $\sigma_u(Q) \geqslant K$. Finally, for (c) we may use a CO-NP algorithm to check that for all allocations Q the Pareto Optimality condition given in Definition 3(1) holds.

We now prove NP-hardness for WI, WO and CO-NP-hardness for PO.

For part (a) we use a reduction from 3-SAT, instances of which are propositional formulae $\Phi(X_n)$ in conjunctive normal form with each clause of Φ defined by exactly three literals. Let

$$\Phi(X_n) = \bigwedge_{i=1}^{m} C_i = \bigwedge_{i=1}^{n} (y_{i,1} \vee y_{i,2} \vee y_{i,3})$$

be an instance of this problem, where $y_{i,j}$ is some literal x_k or $\neg x_k$.

Given $\Phi(X_n)$ we construct an instance $\langle \{A_1, A_2\}, \mathcal{R}, \langle u_1, u_2\rangle, P\rangle$ in which

(a) $\mathcal{R} = \{x_1, x_2, \ldots, x_n, \neg x_1, \ldots, \neg x_n, C_1, \ldots, C_m\}$,
(b) $P = \langle \emptyset; \mathcal{R}\rangle$.

For W a set of literals, i.e.,

$$W \subseteq \{x_1, x_2, \ldots, x_n, \neg x_1, \neg x_2, \ldots, \neg x_n\}$$

we say that W is *useful for* $\Phi(X_n)$ if it satisfies *both* of the conditions below

(1) For each $1 \leqslant k \leqslant n$, W contains *at most one* of the literals x_k, $\neg x_k$.
(2) The partial instantiation of X_n under which each $y \in W$ is assigned true, i.e.,

$$x_i := \begin{cases} 1 & \text{if and only if } x_i \in W, \\ 0 & \text{if and only if } \neg x_i \in W, \end{cases}$$

satisfies $\Phi(X_n)$. Note that if neither $x_i \in W$ nor $\neg x_i \in W$ then this partial instantiation does not assign any value to x_i.

Now with $S \subseteq \mathcal{R}$, let *Lits*(S) be the set

$$\mathit{Lits}(S) = S \cap \{x_1, x_2, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}.$$

The utility functions $\langle u_1, u_2\rangle$ are now given by,

$$u_1(S) = \begin{cases} 0 & \text{if } S = \emptyset, \\ \frac{|S|+1}{2n+m} & \text{if } \mathit{Lits}(S) \text{ is useful}, \\ \frac{|S|}{2n+m} & \text{if } \mathit{Lits}(S) \text{ is } \textit{not} \text{ useful}, \end{cases}$$

$$u_2(S) = \begin{cases} 2 & \text{if } S = \mathcal{R}, \\ 1 + \frac{|S|}{2n+m} & \text{if } \mathit{Lits}(\mathcal{R} \setminus S) \text{ is useful}, \\ 1 + \frac{|S|-1}{2n+m} & \text{if } \mathit{Lits}(\mathcal{R} \setminus S) \text{ is not useful}. \end{cases}$$

Both of these are monotone. Furthermore given $\Phi(X_n)$ we may construct the combinational networks $C^{(1)}$ and $C^{(2)}$ as follows. Let the inputs for each network be $\langle z_1, \ldots, z_{2n+m}\rangle$ with z_i set to represent the presence of x_i (if $i \leqslant n$), the presence of $\neg x_{i-n}$ (if $n < i \leqslant 2n$) and the presence of C_{i-2n} if $(2n < i \leqslant 2n+m)$.

For $C^{(1)}$ we simply use a combinational network that computes the binary representation of $\mathit{Useful}(Z_{2n}) + \sum_{i=1}^{2n+m} z_i$ where

$$\mathit{Useful}(Z_{2n}) = \bigwedge_{i=1}^{n} (\neg z_i \vee \neg z_{n+i}) \ \wedge \ \bigwedge_{i=1}^{m} (z_{i,1} \vee z_{i,2} \vee z_{i,3}).$$

Here, $z_{i,j}$ is the variable from $\{z_1, \ldots, z_{2n}\}$ matching the literal $y_{i,j}$ of clause C_i. Thus, given S a subset of the literals over X_n, the term $(\neg z_i \vee \neg z_{n+i})$ in the corresponding instantiation induced over Z_{2n} will evaluate to $\top$ if and only if at most one of the literals $\{x_i, \neg x_i\}$ occurs in S. Similarly, for each clause $C_i = (y_{i,1} \vee y_{i,2} \vee y_{i,3})$ defining $\Phi(X_n)$ S contains at least one literal from C_i if and only if the term $(z_{i,1} \vee z_{i,2} \vee z_{i,3})$ evaluates to $\top$ for the instantiation of Z_{2n} defined from S.

The summation to compute the binary representation of the number of bits set to 1 within Z_{2n+m} can be carried out using the using the schema of Muller and Preparata [11], see, e.g., [3, pp. 112–114]. The whole number $val(C^1(\alpha_S))$ computed will be $|S|$, i.e., the number of variables set to 1 in α_S, if S is empty or not useful; and $|S|+1$ if S is useful.

For $C^{(2)}$, a combinational network computes the binary representation of

$$\sum_{i=1}^{2n+m-1} 1 + \bigwedge_{i=1}^{2n+m} z_i + \sum_{i=1}^{2n+m} z_i + \mathit{Useful}(\neg z_1, \ldots, \neg z_n, \neg z_{n+1}, \ldots, \neg z_{2n}).$$

For $S \subseteq \mathcal{R}$, this will return $val(C^{(2)}(\alpha_S))$ as

$$\begin{aligned}
&4n + 2m = 2n + m - 1 + 1 + 2n + m + 0 \quad \text{when } S = \mathcal{R},\\
&2n + m + |S| = 2n + m - 1 + 0 + |S| + 1 \quad \text{when } \mathit{Lits}(\mathcal{R} \setminus S) \text{ is useful},\\
&2n + m + |S| - 1 = 2n + m - 1 + 0 + |S| + 0 \quad \text{when } \mathit{Lits}(\mathcal{R} \setminus S) \text{ is not useful}.
\end{aligned}$$

It is clearly the case that these descriptions can be constructed in polynomial-time from the formula $\Phi(X_n)$.

Now, noting that $\sigma_u(\langle \emptyset; \mathcal{R}\rangle) = 2$, we claim that there is an allocation, Q, having $\sigma_u(Q) > 2$ if and only if $\Phi(X_n)$ is satisfiable. To see this consider any non-empty $S \subseteq \mathcal{R}$ and the allocation $\langle S, \mathcal{R} \setminus S\rangle$ to $\langle A_1, A_2\rangle$. We have,

$$\sigma_u\big(\langle S, \mathcal{R} \setminus S\rangle\big) = \begin{cases} \frac{|S|+1}{2n+m} + 1 + \frac{|\mathcal{R}\setminus S|}{2n+m} & \text{if } \mathit{Lits}(S) \text{ is useful},\\ \frac{|S|}{2n+m} + 1 + \frac{|\mathcal{R}\setminus S|-1}{2n+m} & \text{otherwise}. \end{cases}$$

In the former case we get, $\sigma_u(\langle S, \mathcal{R} \setminus S\rangle) = 2 + 1/(2n+m)$ and, in the latter, $\sigma_u(\langle S, \mathcal{R} \setminus S\rangle) = 2 - 1/(2n+m)$. Thus the allocation $\langle \emptyset, \mathcal{R}\rangle$ is welfare improvable if and only if there is an allocation S to A_1 for which $\mathit{Lits}(S)$ is useful: a condition that requires $\mathit{Lits}(S)$ to induce a satisfying instantiation of $\Phi(X_n)$, completing the proof that WI is NP-hard.

For part (b) we simply form the instance, $\langle \{A_1, A_2\}, \mathcal{R}, \langle u_1, u_2\rangle, K\rangle$ with $\mathcal{R}$, $\langle u_1, u_2\rangle$ as in part (a) and $K = 2 + 1/(2n+m)$.

For part (c), although continuing to employ a reduction from 3-SAT, we restrict instances of this to formulae that contain *exactly n* clauses, a variant shown to be NP-complete in [5, Theorem 2(b)]. We use $\mathcal{R}$ and $\langle u_1, u_2\rangle$ as previously, but set $P = \langle P_1, P_2\rangle = \langle \{C_1, \ldots, C_n\}, \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}\rangle$. In this case we have $u_1(P_1) = 1/3$ and $u_2(P_2) = 1 + (2n-1)/(3n)$, so that $\sigma_u(P) = 2 - 1/(3n)$. We claim that this allocation is Pareto optimal if and only if $\Phi(X_n)$ is *unsatisfiable*. First suppose $\Phi(X_n)$ is unsatisfiable. Certainly for any allocation $Q = \langle S, \mathcal{R} \setminus S\rangle$ differing from $\langle P_1, P_2\rangle$, it must be the case that $S = \emptyset$ or $\mathit{Lits}(S)$ is not useful. In the former case,

$$u_1(\emptyset) = 0 < u_1(P_1) = \frac{1}{3}$$

so that the Pareto Optimality condition of Definition 3(1) holds for $\langle P_1, P_2\rangle$ with respect to $\langle \emptyset, \mathcal{R}\rangle$.

If S is non-empty then

$$\sigma_u\big(\langle S, \mathcal{R}\setminus S\rangle\big) = u_1(S) + u_2(\mathcal{R}\setminus S) = 2 - \frac{1}{3n}$$

and so does not increase social welfare. It follows that, in this case,

$$\big(\big[u_1(S) > u_1(P_1)\big] \vee \big[u_2(\mathcal{R}\setminus S) > u_2(P_2)\big]\big)$$
$$\Rightarrow$$
$$\big(\big[u_1(S) < u_1(P_1)\big] \vee \big[u_2(\mathcal{R}\setminus S) < u_2(P_2)\big]\big).$$

Hence if $\Phi(X_n)$ is unsatisfiable then P is Pareto optimal. On the other hand suppose $\Phi(X_n)$ *is* satisfiable. We can then demonstrate that P is *not* Pareto optimal by considering any set of literals $\{y_1, \ldots, y_n\}$ whose instantiation to true satisfies Φ. With such a set consider the allocation

$$Q = \langle Q_1, Q_2\rangle = \big\langle \{y_1, \ldots, y_n\}, \{\neg y_1, \ldots, \neg y_n, C_1, \ldots, C_n\}\big\rangle.$$

Certainly $\mathit{Lits}(Q_1)$ is useful, therefore

$$u_1(Q_1) = \frac{n+1}{3n} > u_1(P_1),$$
$$u_2(Q_2) = 1 + \frac{2}{3} > u_2(P_2).$$

We deduce that the allocation P is Pareto optimal if and only if $\Phi(X_n)$ is unsatisfiable. □

We now proceed with the main results of this paper, showing that deciding if an individually rational $C(k)$-contract path exists between two allocations, is NP-hard for all *constant* values of k and when k can be a predefined function of the size of the resource set. In all cases the results hold in setting involving exactly two agents.

Theorem 12. *For all constant, k,* IR^k *is* NP*-hard.*

Corollary 13. IRO *is* NP*-hard in resource allocation settings for which all utility functions are monotone.*

Theorem 14. *For $k : \mathbb{N} \to \mathbb{N}$ satisfying $k(m) \leqslant m/3$,* $\text{IR}^{k(m)}$ *is* NP*-hard.*

Theorem 15. $\text{IR}^{m/2}$ *is* NP*-hard.*

We have commented earlier on the relationship between these results and our reasons for presenting the proofs separately.

Before continuing it is noted that, in contrast to the complexity classifications for the three problems reviewed in Theorem 11, we do not present *upper bounds* for any of the cases considered: we prove NP-hardness but not NP-*completeness*, i.e., do not present algorithms establishing membership in NP.

Some comments on this point are in order, particularly since there may appear to be an 'obvious' NP algorithm available, namely: guess a sequence of $C(k)$-contracts to realise $\langle P^s, P^t\rangle$ and check whether this defines an IR $C(k)$-contract path. This algorithm, however, may not be implementable[4] with an NP computation. For example, in the case of O-contracts, there may be a *unique* IR O-contract path realising the deal $\langle P^s, P^t\rangle$ but containing *exponentially many* (in m) O-contracts: such paths fail to provide the polynomial length certificate required for membership in NP. Constructions, in instances where only two agents are involved, are given in [4, Theorems 3, 4], for both unrestricted and monotone utility functions. Although not presented explicitly in [4], it is easy to extend these to IR $C(k)$-contracts for any constant k. Of course the 'obvious' algorithm we have outlined will be realisable in NP for resource allocation settings that satisfy certain criteria. One such criterion is that the number of *distinct* values which $\sigma_u(P)$ can take is polynomially-bounded in m: i.e., if $|\{w\colon \exists \text{ an allocation } P \text{ for which } \sigma_u(P) = w\}| \leqslant m^p$. In such settings, no IR contract-path can contain more than m^p deals. Thus, if instances of IR^k are restricted to those for which σ_u has this property, then the corresponding decision problem is in NP. While this may seem to be a rather trivial example, we mention it since, as will be clear from the constructions presented in the proofs, the resource allocation settings formed have precisely this property: the number of distinct values that $\sigma_u(P)$ may take is $O(m)$. We can therefore deduce that, with such a restriction applying, the resulting decision problem *is* NP-complete. The question of upper bounds on the complexity of IR^k when arbitrary resource allocation settings may form part of an instance, remains, however, an open issue.

We now proceed with the proofs of Theorems 12 and 14.

Proof of Theorem 12. Given an instance $\Phi(X_n)$ of 3-SAT, we form an instance $T_\Phi = \langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^s, P^t\rangle$ of IR^k for which there is an IR $C(k)$-contract path realising $\langle P^s, P^t\rangle$ if and only if $\Phi(X_n)$ is satisfiable. Without loss of generality, it may be assumed that $n \geqslant 2k$ (recalling that k is constant). We use

$$\begin{aligned}
&\mathcal{A} = \{A_1, A_2\},\\
&\mathcal{R} = \{x_1, x_2, \ldots, x_n, \neg x_1, \ldots, \neg x_n\},\\
&P^s = \big\langle \emptyset; \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}\big\rangle,\\
&P^t = \big\langle \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}; \emptyset\big\rangle,\\
&u_2(S) = |S|.
\end{aligned}$$

In order to define the utility function, u_1 we need to extend our definition of a set of literals S being *useful*. We say that S is an *effective* set of literals for $\Phi(X_n)$ if both of the following hold.

(a) For each $1 \leqslant i \leqslant n$, S contains *at most one* of the literals $x_i, \neg x_i$.

[4] Our use of '*may not*', as opposed to the more emphatic '*cannot*', is intended: there is a rather subtle (and, at present, unresolved) technical complication that precludes the latter form. We discuss this issue further in Section 4.1 below.

(b) If Ψ_S is the sub-formula (defined on at most $n - |S|$ variables) that results from $\Phi(X_n)$ by applying the partial instantiation of X_n under which each $y \in S$ is assigned true[5] then Ψ_S is satisfiable.

We note that every *useful* set S for $\Phi(X_n)$ is also an *effective* set, however, the converse does not hold in general.

Given the definition of an effective set of literals, we now define

$$u_1(S) = \begin{cases} 2|S| & \text{if } |S| \leqslant n-k \text{ or } |S| > n, \\ 2|S| & \text{if } n-k < |S| \leqslant n \text{ and } S \text{ is effective for } \Phi(X_n), \\ |S| & \text{if } n-k < |S| \leqslant n \text{ and } S \text{ is } \textit{not} \text{ effective for } \Phi(X_n). \end{cases}$$

The key feature of this definition concerns how efficiently $u_1(S)$ can be represented: certainly whenever $|S| \leqslant n-k$ or $|S| > n$ this is easy. Similarly, for $|S|$ outside this range, it is straightforward to determine whether S contains a literal y and its negation $\neg y$. This leaves the case: $n - k < |S| \leqslant n$ where for each y, S contains *at most* one of the literals $\{y, \neg y\}$. For this, whether $u_1(S)$ is $2|S|$ or $|S|$ depends on the induced subformula Ψ_S from Φ and whether this is satisfiable. From our definition, Ψ_S is defined over *at most* $k-1$ variables, and was induced from an instance of 3-SAT. It follows therefore that Ψ_S is a CNF formula on $k-1$ variables each of whose distinct clauses contains between 0 and 3 literals. Since k is *constant*, we can construct a suitable combinational network to recognise satisfiable CNF of this form and with the size of this network being constant (albeit a constant value which may be exponential in k). For example with $k = 2$, the *unsatisfiable* CNF formulae on a single variable z are those containing an empty clause or containing *both* (z) and $(\neg z)$ as clauses.

This technical detail dealt with, we can proceed with the argument that $\Phi(X_n)$ is satisfiable if and only if T_Φ is a positive instance of IR^k.

First suppose that $\Phi(X_n)$ is satisfiable and let $\{y_1, \ldots, y_n\}$ be a set of n literals the instantiation of each to true will satisfy $\Phi(X_n)$. Consider the sequence of $2n$ O-contracts, $\Delta = \langle \delta_1, \delta_2, \ldots, \delta_{2n} \rangle$, in which $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$, $P^{(0)} = P^s$ and $P^{(r)}$ is

$$\begin{cases} \langle \{y_1, \ldots, y_r\}; \mathcal{R} \setminus \{y_1, \ldots, y_r\} \rangle & \text{if } r \leqslant n, \\ \langle \{y_1, \ldots, y_n, \neg y_1, \ldots, \neg y_{r-n}\}; \ \mathcal{R} \setminus \{y_1, \ldots, y_n, \neg y_1, \ldots, \neg y_{r-n}\} \rangle & \text{if } r > n. \end{cases}$$

The O-contract path described by Δ realises $\langle P^s, P^t \rangle$. Furthermore each δ_i is IR:

$$\sigma_u(P^{(i-1)}) = 2(i-1) + (2n - i + 1) = 2n + i - 1,$$
$$\sigma_u(P^{(i)}) = 2i + (2n - i) = 2n + i,$$

and for each $n - k + 1 \leqslant i \leqslant n$, the set of literals $P_1^{(i)}$ held by A_1 is effective from the fact that $\{y_1, \ldots, y_n\}$ induces a satisfying instantiation for $\Phi(X_n)$.

On the other hand, suppose that $\Delta = \langle \delta_1, \delta_2, \ldots, \delta_r \rangle$ with $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$, $P^{(0)} = P^s$ and $P^{(r)} = P^t$ is an IR $C(k)$-contract path. Since at most k literals feature in any deal, in order to progress from $P^{(s)}$, in which A_1 holds no literals, to $P^{(t)}$ in which A_1 holds $2n$

[5] I.e., Ψ_S is formed from the set of clauses in Φ by removing any clause $C = y \vee D$ and replacing $C = \neg y \vee D$ with D when $y \in S$.

literals, it must be the case that at some point, $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$ we have $|P_1^{(i-1)}| \leqslant n-k$ and $n-k < |P_1^{(i)}| \leqslant n$. Letting $d_{(less)}$ denote the value $|P_1^{(i-1)}| - (n-k)$ and $d_{(more)}$ the value $n-k-|P_1^{(i)}|$ so that $0 \leqslant d_{(less)} < d_{(more)} \leqslant k$ for this deal δ_i,

$$\sigma_u(P^{(i-1)}) = 3n-k-d_{(less)},$$
$$\sigma_u(P^{(i)}) = \begin{cases} 3n-k+d_{(more)} & \text{if } P_1^{(i)} \text{ is effective,} \\ 2n & \text{if } P_1^{(i)} \text{ is not effective.} \end{cases}$$

Thus if $P_1^{(i)}$ is *not* an effective set then the deal δ_i is not IR: $\delta_{(less)} \leqslant k-1$, and so, $\sigma_u(P^{(i-1)}) \geqslant 3n-2k+1 > 2n$. We deduce that the existence of an IR $C(k)$-contract path implies that $\Phi(X_n)$ is satisfiable. □

In the special case when $k=1$, i.e., the decision problem IRO, we have the result of Corollary 13.

Proof of Corollary 13. Using the reduction from 3-SAT to IR^k from the proof of Theorem 12 the utility function u_2 is clearly monotone but the function u_1 is not. If, however, we modify the definition of u_1 to become

$$u_1(S) = \begin{cases} 2|S| & \text{if } |S| \neq n, \\ 2n & \text{if } |S| = n \text{ and } S \text{ is useful,} \\ 2n-1 & \text{if } |S| = n \text{ and } S \text{ is not useful,} \end{cases}$$

then not only does the argument of Theorem 12 continue to hold but the utility function u_1 is now monotone. □

Our final result deals with the case of IR $C(k(m))$-contract paths. Thus the number of resources that could be transferred in a single deal is not bounded by some constant value, as in the case of O-contracts or $C(k)$-contracts in general, but is now limited by some function of the total number of resources within the setting. For example, suppose $k(m) = \lfloor \sqrt{m} \rfloor$: given $\mathcal{A} = \{A_1, A_2\}, \mathcal{U} = \langle u_1, u_2 \rangle$, in the resource allocation setting $\langle \mathcal{A}, \{r_1, r_2, r_3, r_4\}, \mathcal{U} \rangle$, a $C(k(m))$-contract can move up to two resources between agents in a single deal. In the same setting, but with $|\mathcal{R}| = 16$, $C(k(m))$-contracts can now transfer up to 4 resources in a single deal.

The fact that the bound on the number of resources allowed to feature in a single deal is no longer constant, means that the reduction employed in proving Theorem 12 cannot be applied in general: we need to be able to specify the utility function u_1 in such a way that from a given instance of 3-SAT an appropriate polynomial-size representation of u_1 can be built. In these proofs, we used the fact that k is constant to demonstrate that testing if a set of literals is effective for $\Phi(X_n)$ can be carried out by testing satisfiability of CNF formulae defined on at most $k-1$ variables, and thus a 'compact' description of u_1 was possible. Although this construction can be effected by a polynomial-time reduction provided that $k(m) = \mathrm{O}(\log m)$—since u_1 need recognise only polynomially many (in m) cases—the same device, however, cannot be used for functions such as $k(m) = \lfloor \sqrt{m} \rfloor$ since testing if S is effective requires testing satisfiability of CNF formulae defined on $\sqrt{n}$ variables.

In order to deal with this complication we need to modify our construction.

Proof of Theorem 14. We employ a reduction from 3-SAT restricted to instances in which the number of clauses is exactly n as in the proof of Theorem 11(c). Let

$$\Phi(X_n) = \bigwedge_{i=1}^{n} C_i = \bigwedge_{i=1}^{n} (y_{i,1} \vee y_{i,2} \vee y_{i,3}).$$

We construct $T_\Phi = \langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^s, P^t \rangle$ an instance of IR^n as follows.

$$\begin{aligned}
&\mathcal{A} = \{A_{lits}, A_{clse}\},\\
&\mathcal{R} = \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n, C_1, \ldots, C_n\},\\
&P^s = \big\langle \{C_1, \ldots, C_n\}; \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\} \big\rangle,\\
&P^t = \big\langle \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}; \{C_1, \ldots, C_n\} \big\rangle.
\end{aligned}$$

It remains to define the utility functions u_{lits} and u_{clse} for each agent. If we consider any subset S of $\mathcal{R}$, then this consists of a subset of $\{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}$ (literals) together with a subset of $\{C_1, \ldots, C_n\}$ (clauses). For a given allocation we use Y_{lits} to denote the subset of literals held by A_{lits}. Similarly Y_{clse}, C_{lits}, C_{clse} will describe respectively: the set of literals held by A_{clse}, of clauses held by A_{lits} and clauses held by A_{clse}. The idea underlying the construction of these is that moving literals from A_{clse} to A_{lits} by $C(n)$-contracts, will *only* be IR if at some stage those literals held by A_{lits} define a satisfying instantiation of $\Phi(X_n)$ (by choosing values for the variables which make the corresponding literals true).

$$u_{lits}(Y_{lits} \cup C_{lits}) = \begin{cases} 0 & \text{if } |Y_{lits}| < n \text{ and } Y_{lits} \text{ is } \textit{not} \text{ useful for } \bigwedge_{C_j \in C_{clse}} C_j,\\ 0 & \text{if } |Y_{lits}| = n \text{ and } Y_{lits} \text{ is } \textit{not} \text{ useful for } \Phi(X_n),\\ 0 & \text{if } |Y_{lits}| > n \text{ and } C_{lits} \neq \emptyset,\\ |Y_{lits}| & \text{otherwise,} \end{cases}$$

$$u_{clse}(Y_{clse} \cup C_{clse}) = \begin{cases} 0 & \text{if } |Y_{lits}| < n \text{ and } Y_{lits} \text{ is } \textit{not} \text{ useful}\\ & \quad \text{for } \bigwedge_{C_j \in C_{clse}} C_j,\\ 0 & \text{if } |Y_{lits}| = n \text{ and } Y_{lits} \text{ is } \textit{not} \text{ useful for } \Phi(X_n),\\ 0 & \text{if } |Y_{lits}| > n \text{ and } C_{lits} \neq \emptyset,\\ |C_{clse}| & \text{otherwise.} \end{cases}$$

We note that $|\mathcal{R}| = 3n$ so our bound on cluster size allows at most n elements from $\mathcal{R}$ to feature in a single deal.

We claim that $\Phi(X_n)$ is satisfiable if and only if there is an IR $C(n)$-contract path realising the deal $\langle P^s, P^t \rangle$.

First suppose that $\Phi(X_n)$ is satisfiable and let $\langle y_1, \ldots, y_n \rangle$ be a set of literals the instantiation of each to true satisfies $\Phi(X_n)$. Consider the sequence of O-contracts, $\langle \delta_1, \ldots, \delta_r \rangle$ in which $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$ and $P^{(0)} = P^s$, $P^{(r)} = P^t$, resulting from the algorithm below.

(1) $i := 1$; $j := 1$.
(2) $P^{(j)}$ is formed by moving the literal y_i from Y_{clse} (in $P^{(j-1)}$) to Y_{lits}.

(3) $j := j + 1$;
  (3.1) Let $\{D_1, \ldots, D_p\}$ be the clauses currently in C_{lits} in which y_i occurs.
  (3.2) The next p O-contracts move each $D \in \{D_1, \ldots, D_p\}$ from C_{lits} to C_{clse}.
  (3.3) $j := j + p$; $i := i + 1$;
(4) If $i \leqslant n$ repeat from step (2).
(5) The final n O-contracts transfer each literal $\neg y_i$ from Y_{clse} to Y_{lits}.

To see that this procedure constructs an IR O-contract path realising $\langle P^s, P^t \rangle$ it suffices to note that in the allocation $P^{(j)}$,

$$u_{lits}(Y_{lits}^{(j)} \cup C_{lits}^{(j)}) = |Y_{lits}^{(j)}|,$$
$$u_{clse}(Y_{clse}^{(j)} \cup C_{clse}^{(j)}) = |C_{clse}^{(j)}|.$$

Furthermore with each deal either the number of literals in Y_{lits} increases by exactly one or the number of clauses in C_{clse} increases by exactly one.

Thus, if $\Phi(X_n)$ is satisfiable then this instance T_Φ of IR^n is accepted.

For the converse implication, suppose Δ is a IR $C(n)$-contract path realising the deal $\langle P^s, P^t \rangle$: $\Delta = \langle \delta_1, \delta_2, \ldots, \delta_i, \ldots, \delta_r \rangle$ with $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$, $P^{(0)} = P^s$, $P^{(r)} = P^t$, and $P^{(i)} = \langle Y_{lits}^{(i)} \cup C_{lits}^{(i)}, Y_{clse}^{(i)} \cup C_{clse}^{(i)} \rangle$.

Noting that $\sigma_u(P^s) = 0$, consider the first deal $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$ in Δ for which the following are true: $C_{lits}^{(i-1)} \neq \emptyset$ and $C_{lits}^{(i)} = \emptyset$. Certainly there must be such a deal since the first condition is true of P^s while the second holds for P^t. Consider the various possibilities:

(a) $|Y_{lits}^{(i-1)}| > n$.
  If such a case were to occur then $u_{lits}(Y_{lits}^{(i-1)} \cup C_{lits}^{(i-1)}) = 0$ and $u_{clse}(Y_{clse}^{(i-1)} \cup C_{clse}^{(i-1)}) = 0$: in $P^{(i-1)}$, A_{lits} holds a non-empty set to clauses together with more than n literals. This contradicts the assumption that Δ is IR since it leads to $\sigma_u(P^{(0)}) = \sigma_u(P^{(i-1)} = 0$. We note that we cannot have $i = 1$ because of the premise $|Y_{lits}^{(i-1)}| > n$.
(b) $|Y_{lits}^{(i-1)}| \leqslant n$.
  Since δ_i is a transfer of resources from A_{lits} to A_{clse}, we have $Y_{lits}^{(i)} \subseteq Y_{lits}^{(i-1)}$: if the set $Y_{lits}^{(i-1)}$ is *not* useful for $\Phi(X_n)$ then this would give $\sigma_u(P^{(i)}) = \sigma_u(P^s)$ (since both contributing utilities would be 0). This contradicts the assumption that Δ is IR, hence in this case $Y_{lits}^{(i-1)}$ must be useful and thus Φ is satisfiable. □

In our final result we show that the bound on cluster size may be increased to $m/2$. The argument used in the proof differs in one significant aspect from those presented in Theorem 12 and Theorem 14: it does not allow a lower bound on the complexity of $\text{IR}^{m/2-d}$ ($d > 0$) to be deduced.

Proof of Theorem 15. We again use a reduction from 3-SAT, but without the restrictions on the number of clauses in instances employed in Theorem 14. Given $\Phi(X_n)$ an instance of 3-SAT, the instance T_Φ of $\text{IR}^{m/2}$ has,

$$\mathcal{A} = \{A_1, A_2\},$$
$$\mathcal{R} = \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\},$$
$$P^s = \langle \emptyset; \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\} \rangle,$$
$$P^t = \langle \{x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}; \emptyset \rangle.$$

The utility functions, $\langle u_1, u_2 \rangle$ being

$$u_1(S) = \begin{cases} 0 & \text{if } |S| < n, \\ 0 & \text{if } |S| = n \text{ and } S \text{ is not useful for } \Phi(X_n), \\ n & \text{if } |S| = n \text{ and } S \text{ is useful for } \Phi(X_n), \\ |S| & \text{if } |S| > n, \end{cases}$$
$$u_2(S) = 0.$$

Noting that $|\mathcal{R}| = 2n$, we claim that $\Phi(X_n)$ is satisfiable if and only if there is an IR $C(n)$-contract path realising $\langle P^s, P^t \rangle$, i.e., T_Φ is a positive instance of IR^n.

Suppose that $\Phi(X_n)$ is satisfiable. Let $\{y_1, y_2, \ldots, y_n\}$ be a set of n literals the instantiation of each to true will satisfy $\Phi(X_n)$. Consider the sequence of $C(n)$-contracts, $\langle \delta_1, \delta_2 \rangle$ below in which Y_j^i is the subset of $\mathcal{R}$ held by A_j after δ_i.

i	Y_1^i	Y_2^i	$u_1(Y_1^i)$	$u_2(Y_2^i)$
0	$\emptyset$	$\{y_1, \ldots, y_n, \neg y_1, \ldots, \neg y_n\}$	0	0
1	$\{y_1, \ldots, y_n\}$	$\{\neg y_1, \ldots, \neg y_n\}$	n	0
2	$\{y_1, \ldots, y_n, \neg y_1, \ldots, \neg y_n\}$	$\emptyset$	$2n$	0

This sequence is IR and realises the deal $\langle P^s, P^t \rangle$ as required.

Conversely, suppose that Δ is a IR $C(n)$-contract path realising the deal $\langle P^s, P^t \rangle$: $\Delta = \langle \delta_1, \delta_2, \ldots, \delta_i, \ldots, \delta_r \rangle$ with $\delta_i = \langle P^{(i-1)}, P^{(i)} \rangle$, $P^{(0)} = P^s$, $P^{(r)} = P^t$. Noting that $\sigma_u(P^{(0)}) = 0$, in order for δ_1 to be IR, we must have $\sigma_u(P^{(1)}) > 0$. This, however, can only happen if $|Y_1^1| \geqslant n$, and since δ_1 is a $C(n)$-contract, it therefore follows that $|Y_1^1| = n$. Such an allocation to A_1, however, will only yield $u_1(Y_1^1) > 0$ if the set Y_1^1 is useful for $\Phi(X_n)$, i.e., if $\Phi(X_n)$ is satisfiable. □

4. Further work and development

Our results presented over Theorems 12–15 above, have been concentrated on *lower bounds* on computational complexity. In total for a range of values of cluster size, the problem of deciding whether a particular resource allocation setting admits a rational $C(k)$-contract path between two specified allocations appears unlikely to admits a feasible algorithmic solution, even if the settings of interest comprise only two agents.

In this section we briefly consider approaches and open problems directed towards more positive results. Our review comprises two subsections, the first of which deals with a somewhat abstruse technical point alluded to earlier; the second outlining algorithmic approaches that might be used in tackling formulations of IRO as an 'optimisation' problem. Readers who are more interested in the algorithmic aspects may wish to proceed directly to the second subsection.

4.1. Upper bounds on IRO

We first consider the issue raised earlier, namely whether $\text{IR}^k \in \text{NP}$. The results of [4, Theorems 3, 4], whereby positive instances of IRO in two agent settings are constructed in which the unique witnessing IR O-contract path has length exponential in m, may appear to disqualify the obvious 'guess and verify' algorithm from being realisable in NP. This reasoning, however, does not take into account the fact that an instance of IRO contains not only the elements $\langle \mathcal{A}, \mathcal{R}, P^s, P^t \rangle$ but also an encoding of the collection of utility functions $\mathcal{U}$. While the constructions from [4] are exponentially long in terms of the former, it is far from clear whether these paths are also exponential in the length of an optimal straight-line programs for $\mathcal{U}$. It is this issue that raises the principal difficulty in inferring that the obvious algorithm cannot be realised in NP as a consequence of [4]. The concerns of [4] are in establishing 'extremal' properties, thus the utility functions constructed to these ends are highly artificial in nature: in particular, the question of optimal straight-line programs is not addressed (since this is not relevant in the context). In total, the following question is unresolved:

Question 1. Is there a polynomial-bound, $q()$ with which: if $T = \langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^s, P^t \rangle$ is a *positive* instance of IRO encoded, using the approach described above, in $|T|$ bits, then there is *always* some IR O-contract path realising $\langle P^s, P^t \rangle$ whose length is at most $q(|T|)$?

A *negative* answer would indicate that the obvious algorithm could not be implemented in NP: a result that would *not* rule out the possibility of $\text{IRO} \in \text{NP}$, but it would indicate that such an upper bound requires a structure *other than a witnessing contract-path* to serve as the polynomial-length certificate.

A *positive* answer to Question 1 is likely to be *extremely hard* to obtain: although we have remarked on the 'artificial' nature of the utility functions in [4] these are, nonetheless, well-defined. In consequence, a positive answer would imply that any straight-line program realising these functions has exponential length: to date the largest lower bound proved for a n-argument function within this model is $3n$ given in [1], [3, pp. 91–99].

4.2. Formulating IRO as an optimisation problem

We have considered properties of $C(k)$-contract paths from the perspective of deciding if paths meeting particular criteria *exist*: in these terms our results indicate that feasible algorithms are unlikely to be found. One possibility is to identify 'special cases' which admit tractable decision processes, e.g., recent work reported in [6] considers a class of resource allocation settings motivated from a 'task allocation' context: the resource set is viewed as a set of m locations, $\mathcal{C}$ with $d_{i,j}$ describing the 'cost' of moving between c_i and c_j; the utility that each agent assigns to any subset S of C is the total cost of a minimal spanning tree of S. There are also a number of related problems for which possible approximation techniques may be constructed. We consider one such problems in this section and outline a 'greedy' approach for it.

We begin by observing that if P^s and P^t are distinct allocations with $\sigma_u(P^t) > \sigma_u(P^s)$ then the length of any O-contract (whether or not such is individually rational) is at least

$$\mathit{Diff}(P^s, P^t) = \sum_{A_i \in \mathcal{A}} \left|\{r \in \mathcal{R}\colon\ r \in P_i^s \text{ and } r \notin P_i^t\}\right|.$$

That is, the total number of resources in $\mathcal{R}$ which have to reallocated from their original owner in P^s to a new owner in P^t. Recognising that it may not be possible to identify an IR O-contract path of length $\mathit{Diff}(P^s, P^t)$ to realise $\langle P^s, P^t\rangle$ motivates the problem of finding an O-contract path that achieves this minimal length *and* has the fewest number of *irrational* deals among such paths. More formally,

Definition 16. The problem *Minimal Irrationality* (MI) takes as an instance a resource allocation setting $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ and allocations P^s, P^t of $\mathcal{R}$ to $\mathcal{A}$. The value returned by $\text{MI}(\mathcal{A}, \mathcal{R}, \mathcal{U}, P^s, P^t)$ is

$$\min\big\{k\colon \exists \text{ an } O\text{-contract path, } \Delta = \langle \delta_1, \ldots, \delta_r\rangle, \text{ of length } \mathit{Diff}(P^s, P^t) \text{ realising } \langle P^s, P^t\rangle \text{ and on which there are at most } k \text{ deals, } \delta_i, \text{ that are } \textit{not} \text{ individually rational}\big\}.$$

It is, of course, an immediate consequence of Theorem 12 and Corollary 13 that the *decision problem* form of MI (in which the upper bound on the number of permitted irrational deals, k, occurs as part of an instance) is NP-complete: use the bound $k = 0$ and the reduction of Corollary 13 noting that if the deal $\langle P^s, P^t\rangle$ can be realised by an IR O-contract path of length $\mathit{Diff}(P^s, P^t)$ if and only if the CNF from which the instance is formed is satisfiable.

Suppose we regard MI as a (partial) function[6] whose domain comprises resource allocation settings $T = \langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ and pairs of allocations $\langle P^s, P^t\rangle$ as given in Definition 16, and whose range is $\mathbb{N}$. We may re-interpret the result of [13] given in Fact 7 as indicating: $\text{MI}(T, \langle P^s, P^t\rangle) \leqslant \mathit{Diff}(P^s, P^t)$, i.e., there is always some O-contract path of length $\mathit{Diff}(P^s, P^t)$ available; and, there are instances for which $\text{MI}(T, \langle P^s, P^t\rangle) > 0$, i.e., there deals which cannot be realised by any IR O-contract path. In total, [13] gives

$$\forall \langle T, P^s, P^t\rangle\colon \text{MI}\big(T, \langle P^s, P^t\rangle\big) \leqslant \mathit{Diff}(P^s, P^t),$$
$$\exists \langle T, P^s, P^t\rangle\colon \text{MI}\big(T, \langle P^s, P^t\rangle\big) \geqslant 1.$$

It is a trivial matter to obtain exact bounds improving these to

$$\forall \langle T, P^s, P^t\rangle\colon \text{MI}\big(T, \langle P^s, P^t\rangle\big) \leqslant \mathit{Diff}(P^s, P^t) - 1,$$
$$\exists \langle T, P^s, P^t\rangle\colon \text{MI}\big(T, \langle P^s, P^t\rangle\big) \geqslant \mathit{Diff}(P^s, P^t) - 1.$$

For the upper bound simply note that since $\sigma_u(P^t) > \sigma(P^s)$ there must be *at least one* IR O-contract on any O-contract path of minimal length realising $\langle P^s, P^t\rangle$. For the lower

[6] 'Partial' since it is convenient to regard its value as undefined when $\sigma_u(P^t) \leqslant \sigma_u(P^s)$.

bound, use any $\langle T, P^s, P^t\rangle$ under which $\sigma_u(P^t) = 1$ and $\sigma_u(P) = 0$ for all allocations P differing from P^t.

While the behaviour of $\text{MI}(T, \langle P^s, P^t\rangle)$ from a general perspective is of some interest, e.g., studies of its value 'on average', such investigations are outside the scope of this note. Our main interest here will be to outline a heuristic aimed at constructing O-contract paths which attain the optimal value.

To simplify the presentation we shall assume that exactly two agents are involved, noting that the development to more than two is straightforward. We present the algorithm and then discuss the thinking underpinning it

Input: $\langle\{A_1, A_2\}, \mathcal{R}, \{u_1, u_2\}, P^s, P^t\rangle$
returns O-contract path of length $\mathit{Diff}(P^s, P^t)$ realising $\langle P^s, P^t\rangle$
$Q := P^s$; $i := 1$;
while $Q \neq P^t$ **loop**
  Choose $p \in Q_1 \setminus P_1^t \cup Q_2 \setminus P_2^t$ such that the allocation V formed by moving p from A_1 to A_2 (if $p \in Q_1$) or from A_2 to A_1 (if $p \in Q_2$) has the following properties:
    P1 $\sigma_u(V) > \sigma_u(Q)$.
    P2 $\sigma_u(V) - \sigma_u(Q)$ is *minimal* among possible choices that satisfy P1.
    P3 If no choice of $p \in Q_1 \setminus P_1^t \cup Q_2 \setminus P_2^t$ that satisfies P1 is possible, i.e., $\forall V \sigma_u(V) \leqslant \sigma_u(Q)$ then choose any V for which the value $\sigma_u(Q) - \sigma_u(V)$ is *maximised*.
  $\delta_i := \langle Q, V\rangle$;
  output δ_i;
  $Q := V$; $i := i + 1$;
end loop

It is not difficult to see that the sequence, $\langle\delta_1, \ldots, \delta_r\rangle$, that is output by this algorithm describes an O-contract path of length $r = \mathit{Diff}(P^s, P^t)$: some deal is chosen via (P1–P3); this deal *is* an O-contract; and, since the choice made is in terms of the current allocation (Q) with respect to the final allocation (P^t), it follows that $r = \mathit{Diff}(P^s, P^t)$.

The motivation for the algorithm is the following: given that $\sigma_u(P^t) > \sigma_u(P^s)$ and that the O-contract path to be formed must have minimal length, i.e., $\mathit{Diff}(P^s, P^t)$, the aim is to implement as many 'small increases' in σ_u within a minimal length path. Of course it may happen that a point, Q, is reached where *every* successor O-contract will result in σ_u not being increased. Rather than attempt to minimise any loss, the algorithm does the opposite: P3 implements the deal which *maximises* the loss of welfare. The idea being that the remaining O-contracts (particularly as the subsequent increments in σ_u are kept minimal) will be 'more likely' to be IR as a result.

We outline this approach merely to indicate that there may be reasonable approximation techniques for the class of problems which have been our principal interest. We will not present a detailed analysis of this algorithm's performance: such studies—both experimental and analytic—of this method and several variations are the topic of continuing work.

5. Conclusion

We have considered a number of decision problems that naturally arise from the multiagent contract negotiation models promoted by (among others) [7,13]. In summary, if contracts are restricted to those in which a limited number of resources can be transferred from one agent to another and are required to be rational (in the sense of strictly improving overall worth of an allocation), then not only is it the case that a suitable contract-path to an optimal allocation may fail to exist (as already shown in [13]), but even *deciding* if a path from a given allocation to a specified more beneficial allocation is possible, is intractable. There are a number of directions in which the results above could be developed. The requirement for individuals deals in a contract-path to be IR could be relaxed so that a limited number of 'irrational' deals are permitted, provided that the allocation eventually reached improves upon the initial allocation. Alternatively, we could consider contracts in which deals permitting an *exchange* of resources between two agents are allowed—the so-called *swap* or S-contracts of [13]. We conjecture, however, that even these degrees of freedom will continue to yield decision questions that are intractable.

Acknowledgement

The work reported in this article was carried out under the support of EPSRC Grant GR/R60836/01.

References

[1] N. Blum, A Boolean function requiring $3n$ network size, Theoret. Comput. Sci. 28 (1984) 337–345.
[2] F. Dignum, M. Greaves (Eds.), Issues in Agent Communication, Lecture Notes in Comput. Sci., vol. 1916, Springer, Berlin, 2000.
[3] P.E. Dunne, The Complexity of Boolean Networks, Academic Press, New York, 1988.
[4] P.E. Dunne, Extremal behaviour in multiagent contract negotiation, J. AI Res. 23 (2005) 41–78.
[5] P.E. Dunne, A.M. Gibbons, M. Zito, Complexity-theoretic models of phase transitions in search problems, Theoret. Comput. Sci. 294 (2000) 243–263.
[6] P.E. Dunne, M.R. Laurence, M.J. Wooldridge, Tractability results for automatic contracting, in: Proc. ECAI'04, Valencia, 2004, pp. 1002–1003.
[7] U. Endriss, N. Maudet, F. Sadri, F. Toni, On optimal outcomes of negotiations over resources, in: Proc. Second Intnl. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS-2003), 2003, pp. 177–184.
[8] M. Fischer, N.J. Pippenger, Relations among complexity measures, J. ACM 26 (1979) 361–381.
[9] S. Kraus, Strategic Negotiation in Multiagent Environments, MIT Press, Cambridge, MA, 2001.
[10] P. McBurney, S. Parsons, M. Wooldridge, Desiderata for argumentation protocols, in: Proc. First Intnl. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS-2002), ACM Press, New York, 2002, pp. 402–409.
[11] D.E. Muller, F.P. Preparata, Bounds to complexities of networks for sorting and for switching, J. ACM 22 (1975) 195–201.
[12] J.S. Rosenschein, G. Zlotkin, Rules of Encounter, MIT Press, Cambridge, MA, 1994.
[13] T.W. Sandholm, Contract types for satisficing task allocation: I theoretical results, in: AAAI Spring Symposium: Satisficing Models, 1998.

[14] T.W. Sandholm, Distributed rational decision making, in: G. Weiß (Ed.), Multiagent Systems, MIT Press, Cambridge, MA, 1999, pp. 201–258.
[15] C.P. Schnorr, The network complexity and Turing machine complexity of finite functions, Acta Inform. 7 (1976) 95–107.
[16] R.G. Smith, The contract net protocol: high-level communication and control in a distributed problem solver, IEEE Trans. Comput. C-29 (12) (1980) 1104–1113.

The Complexity of Deciding Reachability Properties of Distributed Negotiation Schemes

Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 396 (2008) 113–144
Theoretical Computer Science
www.elsevier.com/locate/tcs

The complexity of deciding reachability properties of distributed negotiation schemes

Paul E. Dunne[a,*], Yann Chevaleyre[b]

[a] *Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK*
[b] *LAMSADE, University of Paris-Dauphine, 75775 Paris Cedex 16, France*

Received 25 April 2007; received in revised form 13 September 2007; accepted 25 January 2008

Communicated by X. Deng

Abstract

Distributed negotiation schemes offer one approach to agreeing an allocation of resources among a set of individual agents. Such schemes attempt to agree a distribution via a sequence of locally agreed 'deals' – reallocations of resources among the agents – ending when the result satisfies some accepted criteria. Our aim in this article is to demonstrate that some natural decision questions arising in such settings can be computationally significantly harder than questions related to optimal clearing strategies in combinatorial auctions. In particular we prove that the problem of deciding whether it is possible to progress from a given initial allocation to some desired final allocation via a sequence of "rational" steps is PSPACE-complete.

Keywords: Distributed negotiation; Multiagent resource allocation; Computational complexity; PSPACE-completeness; Straight-line program

1. Introduction

The abstraction wherein a triple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ represents sets of agents, resources, and "utility" functions by which individual agents associate values with resource subsets, has proven to be a useful mechanism in which to consider problems concerning how best to distribute a finite collection of items among a group of agents. In very informal terms, two general approaches have been the basis of algorithmic studies concerning how to organise the allocation of resources to agents. Centralised mechanisms of which combinatorial auction techniques are possibly the best-known exemplar. In addition, distributed methods deriving from the contract-net model formulated by Smith [19] whose properties are the subject of the present article. In Combinatorial Auction schemes, e.g. [15,16,20,21,11,12], a centralised controlling agent (the "*auctioneer*") assumes responsibility for determining which agents receive which resources, basing its decisions on the bids submitted by individual agents. Bidding protocols vary in expressive complexity from those that simply allow an agent to submit a single bid of the form $\langle S, p \rangle$ expressing the fact that the agent is prepared to pay some price p in return for the subset S of $\mathcal{R}$. More complex methods allow a number

* Corresponding author. Tel.: +44 1517954247.
E-mail addresses: ped@csc.liv.ac.uk (P.E. Dunne), chevaley@lamsade.dauphine.fr (Y. Chevaleyre).

0304-3975/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.01.031

of different subsets to be described in separate bids, e.g. the so-called XOR language discussed in [15]. A typical aim of the auctioneer is to decide which bids to accept so as to maximise the overall price paid subject to *at most* one agent being granted any resource. This scheme gives rise to the *Winner Determination Problem* of deciding which bids among those submitted are successful. In its most general form Winner Determination is NP-hard, but there are a number of powerful heuristic approaches and winner determination can be efficiently carried out albeit if the bidding language is of very limited expressiveness. Despite the practical effectiveness of these approaches, there has, however, been a recent revival of interest in autonomous distributed negotiation schemes building on the pioneering study of these by Sandholm [13]. It is not difficult to identify motivations underpinning this renewed interest. For example, the implementation overheads in schema where significant numbers of bids (possibly having complex structures) are communicated to a single controlling agent; the potential difficulties that might arise in persuading an individual agent to assume the rôle and responsibilities of auctioneer and the need to ensure that bidding agents comply with the decisions made by the auctioneer. There are, in addition, the issues raised in deciding on a bidding protocol given the extremes from languages that are highly expressive to those which have very rigid and simple structures. The former are typically computationally hard for winner determination. The latter, while tractable, face the problem of no allocation at all being compatible with the bids received. Finally, aside from the computational problems with which the auctioneer is faced, there is the highly non-trivial issue for the agents bidding as regards selecting and pricing resource sets so as to optimise the likelihood of their "most preferred" bid being accepted.

Faced with such computational issues, notwithstanding the advances in combinatorial auction technology, environments whereby allocations are settled following a process of local improvements negotiated by agents agreeing changes, appear attractive. This is particularly so when the protocols for proposing and implementing resource transfers between agents limit the number of possibilities that individual agents may have to review.

The principal results of this paper establish that, far from resulting in a computationally more tractable regime or, indeed, even one that exhibits complexity "no worse" than the NP-hard status of winner determination, a number of natural decision questions concerning simple distributed negotiation protocols, have significantly *greater* complexity. In particular, we show that given a description of a resource allocation setting – $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ – together with some initial and desired allocations $\langle P^{(s)}, P^{(t)} \rangle$ deciding if the desired allocation can be realised by a sequence of rational "local" reallocations is PSPACE-complete. Thus, deciding if a particular type of negotiation will be effective in bringing about a reallocation is at a similar level of complexity to classical AI planning problems, e.g. as considered in the work of Bylander [1]. We, further, note one of our results resolves a question left open from Dunne et al. [6]: specifically we show the problem, of deciding if there is a rational sequence of "one-resource-at-a-time" reallocations to progress between given starting and final allocations, to be PSPACE-complete, improving upon the earlier NP-hardness classification.

In the next section we introduce the formal structures of contract-net derived distributed negotiation reviewing the components of this presented by [13] together with the terminology and notation that will be used subsequently. Section 3 describes the decision questions that are considered, summarises related work concerning these, and presents a formal statement of the results subsequently proved in Section 5. Separating these two sections, we give a high-level, informal overview of the proof mechanisms in Section 4.

The problems analysed in Section 5 are concerned with what might be called "local" properties of a given allocation setting, specifically whether it is possible to progress from a given starting point to a desired allocation via a restricted class of negotiation primitives. In Section 6 we address "global" properties of such schemes which we term *Convergence* and *Accessibility*. The convergence problem, also studied in work of [7,2], considers a property of resource allocation settings using only a restricted class of deals. Namely, is the setting such that no matter what starting allocation is used and whichever sequence of allowed deals is followed, an optimal allocation will *always* be reached? Perhaps surprisingly, for the restricted deal classes under which the questions considered in Section 5 turn out to be PSPACE-complete, deciding convergence properties is "only" coNP-complete. Accessibility, considers whether from a given starting point there is *at least one* sequence of permitted deals that reaches an optimal outcome. This, too, turns out to be PSPACE-complete. We present concluding comments and discuss further developments in Section 7.

2. Resource allocation settings and local negotiation

The principal structure we consider in this paper is presented in the following definition.

Definition 1. A *resource allocation setting* is defined by a triple $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ where

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\}; \qquad \mathcal{R} = \{r_1, r_2, \ldots, r_m\}$$

are, respectively, a set of (at least two) agents and a collection of (non-shareable) resources. A *utility function*, u, is a mapping from subsets of $\mathcal{R}$ to rational values. Each agent $A_i \in \mathcal{A}$ has associated with it a particular utility function u_i, so that $\mathcal{U}$ is $\langle u_1, u_2, \ldots, u_n \rangle$. An *allocation* P of $\mathcal{R}$ to $\mathcal{A}$ is a partition $\langle P_1, P_2, \ldots, P_n \rangle$ of $\mathcal{R}$. The value $u_i(P_i)$ is called the *utility* of the resources assigned to A_i. We use $\Pi_{n,m}$ to denote the set of all partitions of m resources among n agents: it is easy to see that $|\Pi_{n,m}| = n^m$, there being n different choices for the owner of each of the m resources.

Given some starting allocation, $P \in \Pi_{n,m}$, individual agents may wish to "improve" this: for the purposes of this paper, the concept of an allocation Q improving upon an allocation P will be defined in purely quantitative terms. Even within these limits there are, of course, many different methods by which an allocation P may be quantitatively rated. For the settings considered in this paper we concentrate on the measure *utilitarian social welfare*, denoted $\sigma_u(P)$, which is simply the sum of the agents' utility functions for their allocated resources under P, i.e. $\sigma_u(P) = \sum_{i=1}^{n} u_i(P_i)$.

We next formalise the concepts of *deal* and *contract path*.

Definition 2. Let $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ be a resource allocation setting. A *deal* is a pair $\langle P, Q \rangle$ where $P = \langle P_1, \ldots, P_n \rangle$ and $Q = \langle Q_1, \ldots, Q_n \rangle$ are distinct partitions of $\mathcal{R}$. The effect of implementing the deal $\langle P, Q \rangle$ is that the allocation of resources specified by P is replaced with that specified by Q. For a deal $\delta = \langle P, Q \rangle$, we use $\mathcal{A}^\delta$ to indicate the subset of $\mathcal{A}$ involved, i.e. $A_k \in \mathcal{A}^\delta$ if and only if $P_k \neq Q_k$.

Let $\delta = \langle P, Q \rangle$ be a deal. A *contract path for* δ is a sequence of allocations

$$\Delta = \langle P^{(0)}; P^{(1)}; \ldots; P^{(d-1)}; P^{(d)} \rangle$$

in which $P = P^{(0)}$ and $P^{(d)} = Q$. The *length* of Δ, denoted $|\Delta|$ is d, i.e. the number of *deals* in Δ.

Sandholm [13] presents a number of restrictions on the form that deals may take, one motivation for such being to limit the number of deals that a single agent may have to consider. The class of restricted deals presented in the following definition includes those analysed in [13,14].

Definition 3. Let $\delta = \langle P, Q \rangle$ be a deal involving a reallocation of $\mathcal{R}$ among $\mathcal{A}$.

a. δ is *bilateral* if $|\mathcal{A}^\delta| = 2$.
b. δ is *t-bounded* if δ is bilateral *and* the number of resources whose ownership changes after implementing δ is at most t.
c. δ is a *t-swap* if δ is bilateral *and* for some $s \leq t$, Q is formed by exactly s resources in P_i being assigned to A_j and replaced, in turn, by exactly s resources of P_j.

The class of t-bounded and t-swap deals are simple extensions of the classes of O-contracts and S-contracts in [13]: O-contracts being 1-bounded deals and, similarly, S-contracts are 1-swap deals. We note that t-swap deals are a special case of $(2t)$-bounded deals.

We introduce the concept of a deal being *rational* in the following definition. It will be useful to consider two forms: one linked to the particular quantitative measure of utilitarian social welfare; and, more generally, one which is expressed in terms of arbitrary quantitative measures.

Definition 4. A deal $\langle P, Q \rangle$ is *individually rational* (IR) if and only if $\sigma_u(Q) > \sigma_u(P)$.

The deal $\langle P, Q \rangle$ is said to *cooperatively rational* if for every i, $u_i(Q_i) \geq u_i(P_i)$ and there is at least one j for which $u_j(Q_j) > u_j(P_j)$.

For $\langle \mathcal{A}, \mathcal{R} \rangle$ as before, an *evaluation measure* is a (total) mapping $\sigma : \Pi_{n,m} \to \mathbf{Q}$. A deal $\langle P, Q \rangle$ is *σ-rational* if and only if $\sigma(Q) > \sigma(P)$.

We note that δ is individually rational if and only if δ is σ_u-rational. Where there is no ambiguity we will simply refer to a deal being rational without specifying σ.

Almost all of our development is in terms of *individually* rational deals. As we argue in Section 7, our results are easily modified to hold in the case of *cooperatively* rational deals.

It should be noted that the terms "individually rational" and "cooperatively rational" predate the current article, e.g. Sandholm [13], Endriss et al. [8,9]. The definition of "*individually rational*" is sometimes presented in terms of the existence of so-called *payment functions* with particular properties, e.g. [9, Defn. 13, p. 323]. Informally, such a function specifies an amount, $p(i)$ that $a_i \in \mathcal{A}^\delta$ pays ($p(i) > 0$) or receives ($-p(i)$ when $p(i) < 0$) if the deal is implemented. Thus, $\delta = \langle P, Q\rangle$ is defined to be individually rational (in terms of *payments*), if there is a payment function for which $\sum_{a_i \in \mathcal{A}^\delta} p(i) = 0$ *and* for each $a_i \in \mathcal{A}^\delta$, $u_i(Q_i) - u_i(P_i) > p(i)$. With this sense of individually rational should an agent fail to increase its utility – $u_i(Q_i) \leq u_i(P_i)$ – there would still be an incentive for it to participate: the payment, $-p(i)$, received would compensate. Similarly when a_i increases its utility, although a positive payment will have to be contributed, this will be sufficiently small to leave a_i with some profit. The utility gained, i.e. $u_i(Q_i) - u_i(P_i)$ exceeds $p(i)$ the payment made. Although this definition of individually rational (in terms of payments) superficially appears rather different from that in Definition 4, it is well-known and easily shown that the two are *equivalent*. A deal $\langle P, Q\rangle$ is individually rational (in terms of payments) if and only if $\sigma_u(Q) > \sigma_u(P)$ (the form used in Definition 4), see e.g. [9, Lemma 1, p. 324].

The notions of rationality introduced above are now extended in order to introduce the structures that form the main object of study in this paper: σ-rational *paths*.

Definition 5. For $\langle \mathcal{A}, \mathcal{R}\rangle$ and an evaluation measure, σ, a sequence of allocations

$$\Delta = \langle P^{(0)}; P^{(1)}; \ldots; P^{(d)}\rangle$$

is a *σ-rational contract path* for the (σ-rational) deal $\langle P^{(0)}, P^{(d)}\rangle$ if for all $1 \leq i \leq d$, $\langle P^{(i-1)}, P^{(i)}\rangle$ is σ-rational.

More generally, if $\Phi : \Pi_{n,m} \times \Pi_{n,m} \to \{\top, \bot\}$, is some predicate on deals, we say that Δ is a *Φ-path* if $\Phi(P^{(i-1)}, P^{(i)})$ holds for each $1 \leq i \leq d$. We say that Φ-deals are *complete for σ-rationality* if

$$\forall \langle P, Q\rangle \in \Pi_{n,m} \times \Pi_{n,m} : (\langle P, Q\rangle \text{ is } \sigma\text{-rational}) \Rightarrow (\exists\, \Delta : \Delta \text{ is a } \Phi\text{-path for } \langle P, Q\rangle).$$

3. Decision problems in localised negotiation

The ideas introduced in Definitions 3 and 4 combine to focus on deals that not only restrict their *structure* (in the sense of limiting the number of agents and the number of resources involved) but also add the further condition that a deal must result in a better allocation. It is as a result of such *rationality* conditions that significant difficulties arise within local negotiation approaches.

Notice that Definition 5 imposes a monotonicity condition on the sequence of allocations within, for example, 1-bounded σ-rational paths $\langle P^{(0)}; \ldots; P^{(d)}\rangle$: not only must $\langle P^{(i)}, P^{(i+1)}\rangle$ be realisable by moving exactly one resource, but also $\sigma(P^{(i+1)}) > \sigma(P^{(i)})$. Given that an agent is seeking to maximise its utility, why might such a monotonicity constraint be important? In other words, why would an agent be unprepared to accept a "short-term" loss when this will, eventually, be ameliorated? The detailed discussion of this question from [6, pp. 28–30] may be summarised as follows: the insistence on monotonically increasing $\sigma(P^{(i)})$ allows *cautious* agents to participate in negotiation. That is to say, without this constraint, an agent may reject a proposed deal under which it suffers a loss[1] (even though subsequent deals will change this) because it is uncertain whether any future proposal will make good its loss, or even be made at all.[2]

The effects of combining structural and rationality conditions are already apparent in the following result from [13].

Fact 6.

a. 1-*bounded deals are complete for σ-rationality.*
b. IR 1-*bounded deals are* not *complete for individual rationality.*
c. *If* $|\mathcal{A}| \geq 3$, *then* IR *bilateral deals are not complete for individual rationality.*

[1] That is, the agent's utility decreases and (in terms of payment functions) it is offered inadequate compensation.

[2] In [6] so-called "chain-letters" are given as an example where such uncertainty is reasonable. Significant gains *would* be obtained if the chain continued long enough: owing to doubts that such a profitable outcome will result, individuals may decline to incur the "short-term" loss needed to propagate the chain.

Fact 6 motivates a number of natural questions:

Q1. Are there "reasonable" conditions that can be imposed on collections of utility functions, $\mathcal{U}$, so that in settings $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ where these hold, IR 1-bounded deals *are* complete for individual rationality?
Q2. Given $\langle \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, P^{(s)}, P^{(t)} \rangle$ with $\langle P^{(s)}, P^{(t)} \rangle$ being IR, how efficiently can one determine whether there is a rational 1-bounded contract path for $\langle P^{(s)}, P^{(t)} \rangle$?
Q3. When such a path does exist what can be proven regarding its properties, e.g. number of deals involved, etc.?

The first has been considered in [7,2] and while these offer some positive results, the initial analyses regarding the other two questions presented in [4,6] are rather less encouraging.

Fact 7.

a. *Given* $\langle \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, P^{(s)}, P^{(t)} \rangle$ *with* $\langle P^{(s)}, P^{(t)} \rangle$ *being* IR*, the problem of deciding if there is a rational* 1*-bounded contract path for* $\langle P^{(s)}, P^{(t)} \rangle$ *is* NP*-hard. (Dunne et al. [6, Thm. 12])*
b. *For every* $m = |\mathcal{R}| \geq 7$ *there are choices of* $\langle \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, P^{(s)}, P^{(t)} \rangle$ *for which: there is a* unique IR 1*-bounded contract path,* Δ*, for the* IR *deal* $\langle P^{(s)}, P^{(t)} \rangle$ *and* $|\Delta| = \Omega(2^m)$*. (Dunne [4, Thm. 3].)*
c. *For every* $m = |\mathcal{R}| \geq 6$ *there are choices of* $\langle \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, P^{(s)}, P^{(t)} \rangle$ *with* $|\mathcal{A}| = 3$ *and for which: there is a* unique IR bilateral *contract path,* Δ*, for the* IR *deal* $\langle P^{(s)}, P^{(t)} \rangle$ *and* $|\Delta| = \Omega(2^{m/3})$*. (Dunne [4, Thm. 6].)*

Slightly weaker exponential lower bounds for contract path length than those given in Fact 7(b) and (c), continue to hold even if each agent's utility function must be *monotone*, i.e. satisfy for all subsets S and T of $\mathcal{R}$ the condition $S \subseteq T \Rightarrow u(S) \leq u(T)$.

Although the analysis leading to the proof of Fact 7(a) is couched in terms of IR 1-bounded deals, it is straightforward to adapt it to establish NP-hardness for IR 1-*swap* deals. The principal contribution of the present article is in obtaining tight complexity classifications for these decision problems: Theorem 14 proving both to be PSPACE-complete.

We consider two general forms of decision problems in Section 5 where Φ in the description below is a predicate on deals.

Φ-**PATH**$^{\mathrm{E}}$
Instance: $\langle \langle \mathcal{A}, \mathcal{R} \rangle, \sigma, P^{(s)}, P^{(t)} \rangle$ with $\sigma(P^{(t)}) > \sigma(P^{(s)})$.
Question: Is there a Φ-path Δ for the deal $\langle P^{(s)}, P^{(t)} \rangle$?
Φ-**PATH**$^{\mathcal{U}}$
Instance: $\langle \langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle, P^{(s)}, P^{(t)} \rangle$ with $\sigma_u(P^{(t)}) > \sigma_u(P^{(s)})$.
Question: Is there a Φ-path Δ for the deal $\langle P^{(s)}, P^{(t)} \rangle$?

Although, superficially, these are similar problems, the significant distinction that should be noted is that Φ-**PATH**$^{\mathcal{U}}$ is a *restricted special case* of Φ-**PATH**$^{\mathrm{E}}$. We elaborate further on the differences in our overview of Section 4.

The particular instantiations of Φ that we consider are the following.

(1) $\Phi^{\mathrm{E}}_{\text{1-bd},\sigma\text{-R}}(P, Q)$: the predicate which holds if and only if $\langle P, Q \rangle$ is a σ-rational 1-bounded deal. We subsequently denote the resulting specialisation of Φ-**PATH**$^{\mathrm{E}}$ by **1**-PATH.
(2) $\Phi^{\mathcal{U}}_{\text{1-sw,IR}}(P, Q)$: the predicate which holds if and only if $\langle P, Q \rangle$ is an IR 1-swap deal. We use **1**-SWAP to denote the corresponding special case of Φ-**PATH**$^{\mathcal{U}}$.
(3) $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}(P, Q)$: the predicate which holds if and only if $\langle P, Q \rangle$ is an IR 1-bounded deal. Following the form used in [6] (in which an NP-hardness lower bound was obtained), we denote this special case of Φ-**PATH**$^{\mathcal{U}}$ by IRO-PATH.

Of the three decision problems considered, those defined by IRO-PATH and 1-SWAP have been considered in a number of practical settings. Thus, Sandholm [14] considers so-called "hill-climbing" heuristics built on 1-bounded contracts to identify optimal task allocations.[3] In Dunne et al. [5] t-bounded deals are used in modelling task allocation where the context is that of assigning sets of locations to be covered by individual agents in solving transportation problems.

We will show that each of the resulting decision problems is PSPACE-complete.

[3] The problem of allocating a collection of tasks between a group of cooperating agents with the intention of minimising the overall workload can, clearly, be treated in a similar manner to that of distributing resources.

4. Overview of proof methods

This section has three aims: firstly, to address the technical question of how instances of the decision problems introduced at the conclusion of Section 3 are encoded; secondly, to elaborate on the differences between the forms Φ-**PATH**$^{\mathrm{E}}$ and Φ-**PATH**$^{\mathcal{U}}$; and, finally, to outline the organisation and structure of the proofs presented in Section 5.

4.1. Representing instances

In order to describe instances of Φ-**PATH**$^{\mathrm{E}}$ or Φ-**PATH**$^{\mathcal{U}}$ the problem of encoding functions whose domain is exponentially large in $|\mathcal{R}|$, i.e. $\sigma : \Pi_{n,m} \to \mathbf{Q}$; $u_i : 2^{\mathcal{R}} \to \mathbf{Q}$ must be addressed. Of course, one approach would be simply to enumerate values using some ordering of the relevant domain. There are, however, at least two objections that can be made to such solutions: since the domains are exponentially large – n^m and 2^m – exhaustive enumeration would in practical terms be infeasible even in the case of very simple functions, e.g. $u(S) = 1$ if $|S|$ is even; $u(S) = 2$ otherwise. The second objection is that exhaustive enumeration schemes are liable to give misleading assessments of run-time complexity: an algorithm that is polynomial-time in the length of such an encoding, is actually of exponential complexity in terms of the numbers of agents and resources.

In [6] the following *desiderata* are proposed for encoding a utility function, u, as a sequence of bits $\rho(u)$:

a. $\rho(u)$ is 'concise' in the sense that the length, e.g. number of bits, used by $\rho(u)$ to describe the utility function u within an instance is "comparable" with the time taken by an optimal program that computes the value of $u(S)$.
b. $\rho(u)$ is 'verifiable', i.e. given some binary word, w, there is an efficient algorithm that can check whether w corresponds to $\rho(u)$ for *some* u.
c. $\rho(u)$ is 'effective', i.e. given $S \subseteq \mathcal{R}$, the value $u(S)$ can be efficiently computed from the description $\rho(u)$.

It is, in fact, possible to identify a representation form that satisfies all three of these criteria: we represent each member of $\mathcal{U}$ in a manner that does not *require* explicit enumeration of each subset of $\mathcal{R}$ and allows (a) to be met; uses a 'program' form whose syntactic correctness can be efficiently verified, hence satisfying (b); and for which termination in time linear in the program length is guaranteed, so meeting the condition set by (c). The class of programs employed are the so-called *straight-line programs* (SLP) which have a natural correspondence with combinational logic networks [3].

Definition 8. An (m, s)-*combinational network* C is a directed acylic graph in which there are m *input nodes*, Z_m, labelled $\langle z_1, z_2, \ldots, z_m\rangle$ all of which have in-degree 0. In addition, C has s *output nodes*, called the *result vector*. These are labelled $\langle t_{s-1}, t_{s-2}, \ldots, t_0\rangle$, and have out-degree 0. Every other node of C has in-degree at most 2 and out-degree at least 1. Each non-input node (*gate*) is associated with a Boolean operation of at most two arguments.[4] We use $|C|$ to denote the number of *gate* nodes in C. Any Boolean instantiation of the input nodes to $\underline{a} = \langle a_1, a_2, \ldots, a_m\rangle \in \langle 0, 1\rangle^m$ naturally induces a Boolean value, $h(\underline{a})$ at each node h of C. If h is an input node associated with z_i then $h(\underline{a}) = a_i$; if h is associated with the operation $\neg$ and has as its single input a node g, i.e. $\langle g, h\rangle$ is an edge of C, then $h(\underline{a}) = \neg g(\underline{a})$. Finally if h is a gate associated with the operation θ whose inputs are nodes g_1 and g_2 (that is, $\langle g_1, h\rangle$, $\langle g_2, h\rangle$ are edges of C) then the value $h(\underline{a})$ is $g_1(\underline{a})\theta g_2(\underline{a})$. Hence $\underline{a}$ induces some s-tuple $\langle t_{s-1}(\underline{a}), \ldots, t_0(\underline{a})\rangle \in \langle 0, 1\rangle^s$ at the result vector. For the (m, s)-combinational network C and $\underline{a} \in \langle 0, 1\rangle^m$, this s-tuple is denoted by $C(\underline{a})$.

Although often considered as a model of parallel computation, (m, s)-combinational networks yield a simple form of sequential program – straight-line programs – as follows. Let C be an (m, s)-combinational network to be transformed to a straight-line program, SLP(C), that will contain exactly $m + |C|$ lines. Since C is directed and acyclic it may be topologically sorted, i.e. each gate, g, given a unique integer label $\tau(g)$ with $1 \leq \tau(g) \leq |C|$ so that if $\langle g, h\rangle$ is an edge of C then $\tau(g) < \tau(h)$. The line l_i of SLP(C) evaluates the input z_i if $1 \leq i \leq m$ and the gate for which $\tau(g) = i - m$ if $i > m$. The gate labelling means that when g with inputs g_1 and g_2 is evaluated at $l_{m+\tau(g)}$ since g_i is either an input node or another gate its value will have been determined at l_j with $j < m + \tau(g)$.

[4] In practice, we can restrict the Boolean operations employed to those of binary conjunction ($\wedge$), binary disjunction ($\vee$) and unary negation ($\neg$).

Definition 9. Let $\mathcal{R}$ be as previously with $|\mathcal{R}| = m$, and u a mapping from subsets of $\mathcal{R}$ to whole numbers, i.e. a utility function. The (m, s)-network C^u is said to *realise* the utility function u if: for every $S \subseteq R$ with $\underline{s}$ the instantiation of Z_m given by $z_i = 1$ if and only if $r_i \in S$, it holds

$$u(S) = val(C(\underline{s}))$$

where for $\underline{b} = \langle b_{s-1}, b_{s-2}, \ldots, b_0 \rangle \in \langle 0, 1 \rangle^s$, $val(\underline{b})$ is the whole number whose s-bit binary expansion is $\underline{b}$, i.e. $val(\underline{b}) = \sum_{i=0}^{s-1} b_i * 2^i$, where b_i is treated as the appropriate integer value from $\{0, 1\}$.

Definition 9 provides a method of encoding utility functions $u : 2^{\mathcal{R}} \rightarrow \mathbf{N} \cup \{0\}$ in instances of Φ-**PATH**$^{\mathcal{U}}$: each $u_i \in \mathcal{U}$ is represented by a straight-line program, SLP(C^{u_i}) derived from a suitable combinational network. For instances of Φ-**PATH**$^{\mathrm{E}}$, the function $\sigma : \Pi_{n,m} \rightarrow \mathbf{N} \cup \{0\}$ can be encoded in a similar fashion. For example, via a $(mn, s + 1)$-combinational network, C, whose input $z_{i,j}$ indicates if $r_j \in P_i$; $val(C(\alpha))$ is again an s-bit value: the additional output bit being used to flag if the instantiation α is *not* a valid partition, e.g. if $z_{i,j} = 1$ and $z_{k,j} = 1$ for some r_j and $i \neq k$.[5]

A key property of encodings via SLPs is the following result of [10,18].

Fact 10. *If $f : \{0, 1\}^m \rightarrow \{0, 1\}^s$ is computable by a deterministic Turing Machine program in time T, then f may be realised by an* SLP *containing $O(T \log T)$ lines.*

It should be noted that the proof of Fact 10 is *constructive*, i.e. the translation is not merely an existence argument and, in addition, a suitable SLP can be built in time polynomial in T. Thus a further consequence is our subsequent reductions do not need to give explicit detailed constructions of SLPs.[6] It will suffice to specify σ or $\mathcal{U}$ for it to be apparent that these may be computed efficiently: Fact 10 then ensures suitable representations can be formed.

4.2. Distinctions between Φ-**PATH**$^{\mathrm{E}}$ and Φ-**PATH**$^{\mathcal{U}}$

We recall that Φ-**PATH**$^{\mathrm{E}}$ concerns the existence of σ-rational Φ-paths with the evaluation measure, σ, forming part of the instance whereas Φ-**PATH**$^{\mathcal{U}}$ focuses on the particular choice $\sigma = \sigma_u$ with the collection of utility functions forming part of the instance. Given that our primary interest is in the measure σ_u, it may seem that there is some redundancy in considering Φ-**PATH**$^{\mathrm{E}}$, e.g. if we introduce utility functions for which $u_2 = u_3 = \cdots u_n = 0$, defining $u_1(S)$ as $\sigma(\langle S, P_2, P_3, \ldots, P_n \rangle)$, where P_i is the particular subset of $\mathcal{R}$ held by A_i in a specific case of A_1 holding S, then one has $\sigma_u(P)$ (in the "new" setting) equal to $\sigma(P)$ (in the original form). The main objection to such an approach is that the utility function, u_1, is likely to have *allocative externalities*, i.e. its value is dependent not only on the actual resources held by A_1 but also upon how the other resources are distributed. It has tended to be the normal assumption, often not even mentioned directly,[7] that utility functions do not have such externalities, e.g. [4,6,8,13]. While the complexity classification of Φ-**PATH**$^{\mathrm{E}}$ has some interest in itself, our main concern is with the decision problem relating to Φ-**PATH**$^{\mathcal{U}}$, which focuses on a *single* measure of how good an allocation is – σ_u – and, in keeping with standard approaches, assumes utility functions to be free from externalities.

One point of some importance in our proofs concerning the variant of Φ-**PATH**$^{\mathrm{E}}$ given in Section 5, is that the evaluation measure, σ, constructed in the instance $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)} \rangle$ does *not* admit a *direct* translation to $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^{(s)}, P^{(t)} \rangle$ in which $\mathcal{U}$ is externality-free and is such that $\sigma_u(P) = \sigma(P)$. We introduce a "coding trick" by means of which a general translation from *any* $\langle \mathcal{A}, \mathcal{R}, \sigma \rangle$ to a setting $\langle \{A_1, A_2\}, \mathcal{R}', \{u_1, u_2\} \rangle$ results. In particular this translation provides the means by which two special cases of Φ-**PATH**$^{\mathcal{U}}$ can be proven PSPACE-hard, i.e. the problems **1**-SWAP and IRO-path.

Of course, in principle, our proofs that the special cases of Φ-**PATH**$^{\mathcal{U}}$ are PSPACE-hard could be presented directly, i.e. without reference to Φ-**PATH**$^{\mathrm{E}}$ and the coding device used. There are, however, a number of reasons why we avoid such an approach. The first of these is the technical complexity of the proofs themselves: although the translation

[5] Although we describe the range of σ and u to be whole numbers using SLP encodings, it is a trivial matter to extend to integers, e.g. use an additional output bit to indicate whether a value is positive or negative; and to rationals, e.g. treat one section of the output bits as defining a numerator, the remaining section as a denominator.

[6] Some illustrative constructions of SLPs in specific polynomial-time reductions are presented in [6, pp. 33–4].

[7] One of the few exceptions is [21] which explicitly states that the valuation functions considered are assumed to be free of allocative externalities.

from Φ-**PATH**$^{\mathrm{E}}$ to Φ-**PATH**$^{\mathcal{U}}$ turns out to be relatively straightforward; the central result that **1**-PATH is PSPACE-hard on which our subsequent classifications build, is rather more involved. We note that notwithstanding the use of arbitrary evaluation measures, the problem **1**-PATH is a "natural" decision question whose properties, we contend, merit consideration in their own right.

4.3. Proof structure

We begin by recalling the decision problems considered.

1-PATH (special case of Φ-**PATH**$^{\mathrm{E}}$)
Instance: $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)} \rangle$ with $\sigma(P^{(t)}) > \sigma(P^{(s)})$.
Question: Is there a σ-rational 1-bounded path for $\langle P^{(s)}, P^{(t)} \rangle$?

1-SWAP (special case of Φ-**PATH**$^{\mathcal{U}}$)
Instance: $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^{(s)}, P^{(t)} \rangle$ with $\sigma_u(P^{(t)}) > \sigma_u(P^{(s)})$.
Question: Is there an IR 1-swap path for $\langle P^{(s)}, P^{(t)} \rangle$?

IRO-PATH (special case of Φ-**PATH**$^{\mathcal{U}}$)
Instance: $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}, P^{(s)}, P^{(t)} \rangle$ with $\sigma_u(P^{(t)}) > \sigma_u(P^{(s)})$.
Question: Is there an IR 1-bounded path for $\langle P^{(s)}, P^{(t)} \rangle$?

Subject to $\Phi(P, Q)$ being decidable in PSPACE it is straightforward to show that Φ-**PATH**$^{\mathrm{E}} \in$ PSPACE. For each of the problems listed, the corresponding $\Phi(P, Q)$ is decidable in (deterministic) polynomial-time.

On first inspection the approach taken to proving PSPACE-hardness may seen rather indirect: an "auxiliary problem" – *Achievable Circuit Sequence* (ACS) – is defined independently of the arena of multiagent negotiation contexts. The assertion "**1**-PATH is PSPACE-complete", is justified by showing "ACS is PSPACE-complete" (Theorem 12) and then ACS $\leq_p$ **1**-PATH (Theorem 13). This auxiliary problem has, however, two important properties. Firstly, it is "easy" to prove that ACS is PSPACE-complete using standard generic reduction techniques.[8] The second property of ACS is that its formal structure is very similar to that of **1**-PATH.

Thus, ACS is concerned with deciding a property of a given (N, N)-combinational logic network, C, with respect to two distinct binary N-tuples. The N inputs of C are interpreted as a sequence of n *data* bits $\langle x_1, x_2, \ldots, x_n \rangle$ coupled with a sequence of m *value* bits $\langle y_0, y_1, \ldots, y_{m-1} \rangle$; the N outputs are viewed in a similar fashion. Now, suppose that $\underline{a} = \langle data(\underline{a}), value(\underline{a}) \rangle$ and $\underline{b} = \langle data(\underline{b}), value(\underline{b}) \rangle$ are the binary N-tuples given with C to form an instance of ACS.

Recall that $val(\underline{y})$ is the whole number represented by the m value bits of C, i.e. $val(\underline{y}) = \sum_{i=0}^{m-1} (2^i) * y_i$, and define

$$\langle data_k(\underline{a}), value_k(\underline{a}) \rangle = \begin{cases} \langle data(\underline{a}), value(\underline{a}) \rangle & \text{if } k = 0 \\ C(\langle data_{k-1}(\underline{a}), value_{k-1}(\underline{a}) \rangle) & \text{if } k > 0. \end{cases}$$

Since the output of any (N, N)-combinational logic network on a given instantiation of its inputs is uniquely determined, the sequence $[\langle data_k(\underline{a}), value_k(\underline{a}) \rangle]_{k \geq 0}$ is well-defined and unique.

Informally, ACS asks of its instance $\langle C, \underline{a}, \underline{b} \rangle$ if there is some value $t \geq 1$ with which:

a. $\langle data_t(\underline{a}), value_t(\underline{a}) \rangle = \langle data(\underline{b}), value(\underline{b}) \rangle$
b. For each $1 \leq i \leq t$, $val(value_i(\underline{a})) > val(value_{i-1}(\underline{a}))$.

Although the formal technical argument that ACS $\leq_p$ **1**-PATH given in Section 5.2 involves a number of notational complexities, its basic strategy is not difficult to describe. Recalling that an instance of ACS consists of an $(n+m, n+m)$-combinational logic network, C, together with instantiations $\langle \underline{x}, \underline{y} \rangle$, $\langle \underline{z}, \underline{w} \rangle$ from $\langle 0, 1 \rangle^{n+m}$, the instance $\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)} \rangle$ of **1**-PATH that is formed uses 5 agents. The resource set $\mathcal{R}_C$ contains disjoint sets each of

[8] That is to say, "easy" *pace* the notational overheads inherent in most generic simulations of resource-bounded Turing machine classes: the elegant casting of Turing machine behaviour in terms of planning operators presented in [1] being a notable exception.

size $2(n+m)$ – $\mathcal{R}^V$ and $\mathcal{R}^W$ – with "appropriate" subsets of $\mathcal{R}^X$ (for $X \in \{V, W\}$) mapping to elements of $\langle 0, 1\rangle^{n+m}$. In the initial allocation, $P^{(s)}$, A_1 holds the subset of $\mathcal{R}^V$ *and* the subset of $\mathcal{R}^W$ that maps to $\langle \underline{x}, \underline{y}\rangle \in \langle 0, 1\rangle^{n+m}$. In the final allocation, $P^{(t)}$, A_1 should hold the subsets of $\mathcal{R}^V$ and $\mathcal{R}^W$ that map to $\langle \underline{z}, \underline{w}\rangle$. For the agents A_2 and A_3: the former should hold subsets of $\mathcal{R}^V$ while the latter holds subsets of $\mathcal{R}^W$. The evaluation measure, σ, is constructed so that any allocation, Q, for which $Q_2 \not\subseteq \mathcal{R}^V$ or $Q_3 \not\subseteq \mathcal{R}^W$ has $\sigma(Q) < 0$.

The main idea is to simulate the witnessing sequence $\{\langle \underline{x}_i, \underline{y}_i\rangle\}_{0\le i\le t}$ for a positive instance of ACS by a sequence of allocations to A_1, i.e. from the initial allocation to A_1 which we recall mapped to $\langle \underline{x}_0, \underline{y}_0\rangle \in \langle 0, 1\rangle^{n+m}$ subsequent allocations to A_1 will be those subsets of $\mathcal{R}^V$ and $\mathcal{R}^W$ which map to $\langle \underline{x}_i, \underline{y}_i\rangle \in \langle 0, 1\rangle^{n+m}$. The problem that arises in this simulation is that if $Q^{(i)}$ is the allocation in which A_1's holding reflects $\langle \underline{x}_i, \underline{y}_i\rangle$ then the deal $\langle Q^{(i)}, Q^{(i+1)}\rangle$ although σ-rational for the evaluation measure constructed, will not be 1-bounded. In order to effect this deal, a sequence of 1-bounded σ-rational deals is used which involve the following stages:

1. a subset of $\mathcal{R}^V$ is transferred one resource at a time from A_2 to A_4;
2. a subset of $\mathcal{R}^V$ is transferred one resource at a time from A_1 to A_2;
3. the resources moved into A_4 in stage (1) are transferred to A_1.

The subset of $\mathcal{R}^V$ held by A_1 on completion will map to $\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle$, while the subset of $\mathcal{R}^W$ continues to map to $\langle \underline{x}_i, \underline{y}_i\rangle$. These three stages are then repeated, but now with resources from $\mathcal{R}^W$ and the agent A_3 involved, so that the subset of $\mathcal{R}^W$ held by A_1 will, on completion, map to $\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle$.

In order to track whether resources should be moved *out of* A_4 into A_1, a "marker" resource, μ, initially held by A_5 is used: μ is reallocated to A_4 at the end of the second phase and returned to A_5 once the third stage is complete.

The notational overhead in the proof stems from specifying the evaluation measure, σ, in such a way that a σ-rational 1-bounded sequence of deals to go from $P^{(s)}$ to $P^{(t)}$ is possible if and only if the source instance of ACS should be accepted.

5. PSPACE-complete negotiation questions

We begin with the relatively straightforward proof that the decision problems we consider are all decidable by PSPACE algorithms. Since all of these are specialisations of Φ-**PATH**$^{\mathrm{E}}$ and the predicate $\Phi(P, Q)$ is polynomial-time decidable for each, it suffices to prove,

Theorem 11. *For predicates* $\Phi : \Pi_{n,m} \times \Pi_{n,m} \to \{\top, \bot\}$ *such that* $\Phi(P, Q)$ *is polynomial-time decidable,* Φ-**PATH**$^{\mathrm{E}} \in$ PSPACE.

Proof. Given an instance $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ of Φ-**PATH**$^{\mathrm{E}}$ in which $\sigma : \Pi_{n,m} \to \mathbf{Q}$ is described in the form of a straight-line program, use a non-deterministic algorithm which proceeds as follows:

$P := P^{(s)}$
loop
  Non-deterministically choose an allocation $Q \in \Pi_{n,m}$
  if $\neg\Phi(P, Q)$ then **reject**
  else if $Q = P^{(t)}$ then **accept**
  else $P := Q$
end loop

If a Φ-path realising $\langle P^{(s)}, P^{(t)}\rangle$ exists then this non-deterministic algorithm has a computation that will successfully identify it. The algorithm need only record the allocations P and Q occurring in the loop body and thus can be implemented in NPSPACE. The theorem now follows from Savitch's Theorem: NPSPACE = PSPACE, [17]. □

5.1. The achievable circuit sequence problem (ACS)

The following decision problem is central to our subsequent argument.

Achievable Circuit Sequence (ACS)
Instance: (N, N)-combinational logic network, C, with $N = n + m$ inputs $\langle X_n, Y_m\rangle$ and $n + m$ outputs, $\langle Z_n, W_m\rangle$; $\langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle \in \langle 0, 1\rangle^{n+m}$.

Question: Is there a sequence

$$\Gamma = \langle\langle \underline{x}_0, \underline{y}_0\rangle, \langle \underline{x}_1, \underline{y}_1\rangle, \ldots, \langle \underline{x}_k, \underline{y}_k\rangle\rangle$$

such that

a. $\langle \underline{x}_0, \underline{y}_0\rangle = \langle \underline{x}, \underline{y}\rangle$,
b. $\langle \underline{x}_k, \underline{y}_k\rangle = \langle \underline{z}, \underline{w}\rangle$,
c. $\forall 1 \leq i \leq k, C(\underline{x}_{i-1}, \underline{y}_{i-1}) = \langle \underline{x}_i, \underline{y}_i\rangle$ and $val(\underline{y}_i) > val(\underline{y}_{i-1})$?

Theorem 12. ACS *is* PSPACE*-complete.*

Proof. An instance $\langle C, \underline{x}, \underline{y}, \underline{z}, \underline{w}\rangle$ of ACS can be decided by a (deterministic) polynomial-space computation that iterates evaluating

$$\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle = C(\underline{x}_i, \underline{y}_i)$$

(starting with $\langle \underline{x}, \underline{y}\rangle$).

This computation terminates either when $val(\underline{y}_{i+1}) \leq val(\underline{y}_i)$ (the instance is rejected) or when $\langle \underline{z}, \underline{w}\rangle$ occurs with the former condition taking precedence when $\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle = \langle \underline{z}, \underline{w}\rangle$. Since there are only 2^{n+m} possible cases, eventually one of these two termination conditions must arise. The whole computation can be accomplished in polynomial-space since only the current $\langle \underline{x}_i, \underline{y}_i\rangle$ need be remembered.

For PSPACE-hardness we use a generic reduction, i.e. given a Turing machine program, M, input s, and space-bound $S = |s|^c$ we form an instance of ACS that is accepted if and only if s is accepted by M within an S-space bounded computation. We may assume that M has a unique accepting configuration u. It suffices to note that from the description of M we can build a (t, t)-combinational logic network C_M whose input bits encode configurations of M on exactly S tape-cells. For such a configuration, $\underline{\chi}$, $C_M(\underline{\chi}) = \underline{\pi}$ if and only if the configuration $\underline{\pi}$ follows in exactly one move of M from the configuration $\underline{\chi}$. Note we may use the convention that $C_M(\underline{u}) = \underline{u}$ for the unique accepting configuration. Combine C_M with a p-bit counter, D, i.e. $val(D(\underline{v})) = val(\underline{v}) + 1$ with p chosen large enough so that the total number of configurations of S-tape bounded configurations of M can be represented in p bits.[9] Now let $\underline{s}$ be the instantiation of the inputs of C_M corresponding to the initial configuration of M on input s: s is accepted by M if and only if $\langle (C_M, D), \langle \underline{s}, 0^p\rangle, \langle \underline{u}, 1^p\rangle\rangle$ is accepted as an instance of ACS. □

5.2. ACS *is polynomially-reducible to 1*-PATH

It will be convenient to introduce the following notation and definitions.

For $V = \{v_1, v_2, \ldots, v_{n+m}\}$ and $W = \{w_1, w_2, \ldots, w_{n+m}\}$ disjoint sets of $n + m$ propositional variables, we define sets

$$\begin{aligned}\mathcal{R}^V &= \{v_1,\ v_2\ , \ldots,\ v_{n+m},\ \neg v_1\ , \ldots,\ \neg v_{n+m}\}\\ \mathcal{R}^W &= \{w_1,\ w_2\ , \ldots,\ w_{n+m},\ \neg w_1\ , \ldots,\ \neg w_{n+m}\}\\ \mathcal{R} &= \mathcal{R}^V \cup \mathcal{R}^W.\end{aligned}$$

In our subsequent notation, in order to avoid repetition, X refers to either of V or W.

Given $S \subseteq \mathcal{R}$, the subset S^X is defined via $S^X = S \cap \mathcal{R}^X$. We define a partial mapping, $\beta : 2^{\mathcal{R}} \to \langle 0, 1\rangle^{n+m}$ as follows.

For all of the cases below, $\beta(S)$ is undefined, i.e. $\beta(S) = \perp$ whenever

$$\begin{cases} S^V \neq \emptyset \text{ and } S^W \neq \emptyset \\ \text{or} \\ S \subseteq \mathcal{R}^X \text{ and } |S| \neq n + m \\ \text{or} \\ S \subseteq \mathcal{R}^X \text{ and there is some } i \text{ for which } \{x_i, \neg x_i\} \subset S. \end{cases}$$

[9] It is easy to show that $p = O(S)$.

For the remaining cases,

$$\beta(S) = \langle a_1 a_2 \dots a_{n+m} \rangle \text{ where } a_i = \begin{cases} 0 & \text{if } \neg x_i \in S \\ 1 & \text{if } x_i \in S. \end{cases}$$

Given $\underline{a} = \langle a_1 a_2 \dots a_{n+m} \rangle \in \langle 0, 1 \rangle^{n+m}$, there is a uniquely defined set $S \subseteq \mathcal{R}^X$ for which $\beta(S) = \underline{a}$. Thus we can introduce β_X^{-1} as a total mapping from $\langle 0, 1 \rangle^{n+m}$ to subsets from $\mathcal{R}^X$, as

$$\beta_X^{-1}(\underline{a}) = S \subseteq \mathcal{R}^X \text{ such that } \beta(S) = \underline{a}.$$

For $\underline{a} \in \langle 0, 1 \rangle^{n+m}$, we denote by $val_m(\underline{a})$ the whole number whose m bit binary representation is $a_{n+1} a_{n+2} \cdots a_{n+m}$, i.e the value $\sum_{i=n+1}^{n+m} (a_i) * 2^{n+m-i}$.

Let S and T be subsets of $\mathcal{R}^X$ that satisfy all the conditions (CS1)–(CS4).

CS1. $S \cap T = \emptyset$
CS2. For each i $(1 \leq i \leq n+m)$ either $x_i \notin S$ or $\neg x_i \notin S$
CS3. For each i $(1 \leq i \leq n+m)$ either $x_i \notin T$ or $\neg x_i \notin T$
CS4. For each i $(1 \leq i \leq n+m)$ if $(x_i \notin S)$ and $(\neg x_i \notin S)$ then $(x_i \in T)$ or $(\neg x_i \in T)$.

For such sets S, T the *composite set*, $S \otimes T$, is the subset of $\mathcal{R}^X$ given by

$$S \otimes T = S \setminus (\{x \ : \ \neg x \in T\} \cup \{\neg x \ : \ x \in T\}) \bigcup T.$$

Now suppose that C is an (N, N)-combinational logic network with $N = n+m$, $\underline{a} \in \langle 0, 1 \rangle^{n+m}$, and $S \subseteq \mathcal{R}^X$, is such that for each i, either $x_i \notin S$ or $\neg x_i \notin S$. The *difference set* for S with respect to $\underline{a}$ is the subset of $\mathcal{R}^X$,

$$\text{DIFF}_X(S, \underline{a}) = \beta_X^{-1}(\underline{a}) \setminus S.$$

The following lemma establishes some useful relationships between the composite set operation, $\otimes$, and difference sets.

Lemma 1. *Let C be an $(n+m, n+m)$-combinational logic network, $\underline{a} \in \langle 0, 1 \rangle^{n+m}$, and, as in the notation introduced above, let $\mathcal{R}^X$ denote $\{x_1, \dots, x_{n+m}, \neg x_1, \dots, \neg x_{n+m}\}$.*

For every $D \subseteq \beta_X^{-1}(\underline{a}) \setminus \beta_X^{-1}(C(\underline{a}))$, the sets S and T defined by

$$S = \beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})) \cup D$$
$$T = \text{DIFF}_X(S, C(\underline{a}))$$

have the following properties,

a. $T = \beta_X^{-1}(C(\underline{a})) \setminus \beta_X^{-1}(\underline{a})$
b. $S \otimes T = \beta_X^{-1}(C(\underline{a}))$.

Proof. For (a), from the definition of DIFF_X,

$$\begin{aligned} T &= \text{DIFF}_X(S, C(\underline{a})) \\ &= \beta_X^{-1}(C(\underline{a})) \setminus (\beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})) \ \cup \ D) \\ &= \beta_X^{-1}(C(\underline{a})) \setminus \beta_X^{-1}(\underline{a}). \end{aligned}$$

The final line following as $D \subseteq \beta_X^{-1}(\underline{a}) \setminus \beta_X^{-1}(C(\underline{a}))$ and thus $D \cap \beta_X^{-1}(C(\underline{a})) = \emptyset$.

For (b), consider the set $S \otimes T$. This is formed by first removing from S all elements in

$$F = \{x \in S \ : \ \neg x \in T\} \bigcup \{\neg x \in S \ : \ x \in T\}.$$

We claim that this set comprises exactly those elements of the set D. To see this, first observe that F cannot contain any member of the set $\beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a}))$: if $x \in \beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a}))$ then $x \in \beta_X^{-1}(C(\underline{a}))$ and from the fact that $T \subseteq \beta_X^{-1}(C(\underline{a}))$ this precludes $\neg x \in T$. Without loss of generality, suppose for the sake of contradiction, that $x \in D \setminus F$ – a similar argument applies if we assume instead $\neg x \in D \setminus F$. From the fact that $D \subseteq \beta_X^{-1}(\underline{a}) \setminus \beta_X^{-1}(C(\underline{a}))$

we have $x \in \beta_X^{-1}(\underline{a})$ and $x \notin \beta_X^{-1}(C(\underline{a}))$. Since exactly one of x and $\neg x$ must appear in $\beta_X^{-1}(C(\underline{a}))$ we deduce that $\neg x \in \beta_X^{-1}(C(\underline{a}))$. We now have

$$x \in D \subseteq \beta_X^{-1}(\underline{a}) \setminus \beta_X^{-1}(C(\underline{a})) \subseteq S$$
and
$$\neg x \in \beta_X^{-1}(C(\underline{a})) \setminus \beta_X^{-1}(\underline{a}) = T$$

and thus $x \in F$ contradicting our assumption that $x \in D \setminus F$. It follows, therefore, that $D \subseteq F$ and thus, recalling that $F \cap \beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})) = \emptyset$,

$$\begin{aligned} S \setminus F &= (\beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})) \cup D) \setminus F \\ &= \beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})). \end{aligned}$$

Having formed $S \setminus F$, the construction of $S \otimes T$ is completed by adding all elements in T, so that

$$\begin{aligned} S \otimes T &= (S \setminus F) \cup T \\ &= \beta_X^{-1}(\underline{a}) \cap \beta_X^{-1}(C(\underline{a})) \cup \beta_X^{-1}(C(\underline{a})) \setminus \beta_X^{-1}(\underline{a}) \\ &= \beta_X^{-1}(C(\underline{a})) \end{aligned}$$

as was claimed. □

We now prove,

Theorem 13. **1**-PATH *is* PSPACE-*complete.*

Proof. Noting that **1**-PATH $\in$ PSPACE the result will follow via Theorem 12 by showing ACS$\leq_p$**1**-PATH.

Thus given, $\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle\rangle$ an instance of ACS we form $\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)} \rangle$ for which

$$\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle\rangle \in \mathcal{L}_{\text{ACS}} \quad \Leftrightarrow \quad \langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)} \rangle \in \mathcal{L}_{\mathbf{1}\text{-PATH}}.$$

$\mathcal{A}_C$ contains five agents,

$$\mathcal{A}_C = \{A_1, A_2, A_3, A_4, A_5\}.$$

Fix sets $V = \{v_1, v_2, \ldots, v_{n+m}\}$ and $W = \{w_1, w_2, \ldots, w_{n+m}\}$ so that the resource set in the instance of **1**-PATH is,

$$\mathcal{R}_C = \mathcal{R}^V \bigcup \mathcal{R}^W \bigcup \{\mu\}.$$

Here μ is a "new" resource distinct from those in $\mathcal{R}^V \cup \mathcal{R}^W$.

For the source and destination allocations – $P^{(s)}$ and $P^{(t)}$ – we use,

$$\begin{aligned} P_1^{(s)} &= \beta_V^{-1}(\langle \underline{x}, \underline{y} \rangle) \cup \beta_W^{-1}(\langle \underline{x}, \underline{y} \rangle) \\ P_2^{(s)} &= \mathcal{R}^V \setminus P_1^{(s)} \\ P_3^{(s)} &= \mathcal{R}^W \setminus P_1^{(s)} \\ P_4^{(s)} &= \emptyset \\ P_5^{(s)} &= \{\mu\} \end{aligned} \qquad \begin{aligned} P_1^{(t)} &= \beta_V^{-1}(\langle \underline{z}, \underline{w} \rangle) \cup \beta_W^{-1}(\langle \underline{z}, \underline{w} \rangle) \\ P_2^{(t)} &= \mathcal{R}^V \setminus P_1^{(t)} \\ P_3^{(t)} &= \mathcal{R}^W \setminus P_1^{(t)} \\ P_4^{(t)} &= \emptyset \\ P_5^{(t)} &= \{\mu\}. \end{aligned}$$

To complete the construction, we need to specify σ.

Given $Q \in \Pi_{5,4(n+m)+1}$, we will have $\sigma(Q) \geq 0$ only if Q satisfies *all* of the following requirements:

B1. $Q_1 \subseteq \mathcal{R}^V \cup \mathcal{R}^W$.
B2. $Q_2 \subseteq \mathcal{R}^V$.
B3. $Q_3 \subseteq \mathcal{R}^W$.
B4. $Q_4^V = \emptyset$ *or* $Q_4^W = \emptyset$.
B5. $Q_5 \subseteq \{\mu\}$, i.e. either $Q_5 = \emptyset$ or $Q_5 = \{\mu\}$.
B6. For $X \in \{V, W\}$, if $Q_i^X \neq \emptyset$ then for all j, $\{x_j, \neg x_j\} \not\subseteq Q_i^X$.

Assuming that (B1) through (B6) hold, then $\sigma(Q) \geq 0$ if and only if (at least) one of the following six conditions holds true[10] of Q.

C1. $\beta(Q_1^V) = \beta(Q_1^W)$ and $Q_4 \subseteq \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$.
C2. $\beta(Q_1^V \otimes Q_4^V) = C(\beta(Q_1^W))$ and $Q_4 = \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$.
C3. $\beta(Q_1^V \cup Q_4^V) = C(\beta(Q_1^W))$ and $\mu \in Q_4$.
C4. $\beta(Q_1^V) = C(\beta(Q_1^W))$ and $Q_4 \subseteq \text{DIFF}_W(Q_1^W, \beta(Q_1^V))$.
C5. $\beta(Q_1^V) = \beta(Q_1^W \otimes Q_4^W)$ and $Q_4 = \text{DIFF}_W(Q_1^W, \beta(Q_1^V))$.
C6. $\beta(Q_1^V) = \beta(Q_1^W \cup Q_4^W)$ and $\mu \in Q_4$.

One further requirement relating to (C3) is the following. Let $\underline{f}$ and $\underline{g}$ be the instantiations in $\langle 0, 1\rangle^{n+m}$ defined as

$$\begin{aligned} \underline{f} &= \beta(P_1^W) \\ \underline{g} &= C(\beta(P_1^W)) \end{aligned}$$

then, in addition $val_m(\underline{g}) > val_m(\underline{f})$.[11]

We write, C1(Q), C2(Q), etc. if Q satisfies C1, C2, and so on.

In the specification of σ given below, $K_{mn} \in \mathbf{N}$ is a suitably large integer value depending on $n + m$.[12]

For an allocation Q satisfying at least one[13] of these conditions, $\sigma(Q)$ is

C1	$2\,K_{mn} val_m(\beta(Q_1^W))$	$+\lvert Q_4\rvert$	
C2	$2\,K_{mn} val_m(\beta(Q_1^W))$	$+\lvert Q_4\rvert$	$+n + m - \lvert Q_1^V\rvert$
C3	$K_{mn} val_m(\beta(Q_1^W)) + K_{mn} val_m(C(\beta(Q_1^W)))$	$-\lvert Q_4\rvert$	
C4	$2\,K_{mn} val_m(\beta(Q_1^V))$	$+\lvert Q_4\rvert - 2$	$-3\lvert\text{DIFF}_W(Q_1^W, \beta(Q_1^V))\rvert$
C5	$2\,K_{mn} val_m(\beta(Q_1^V))$	$-2\lvert Q_4\rvert - 2$	$+n + m - \lvert Q_1^W\rvert$
C6	$2\,K_{mn} val_m(\beta(Q_1^V))$	$-\lvert Q_4\rvert$.	

For any allocation, Q, in which none of these conditions holds, we set $\sigma(Q) = -1$.

We note, at this juncture, that $\sigma(Q)$ can be evaluated in time polynomial in the number of bits required to encode the instance of ACS: firstly, given C, the relationship between Q_1^V, Q_1^W and Q_4 characterising each of the six conditions is easily checked, and the evaluation of $\sigma(Q)$, given that one of these is satisfied, involves basic arithmetic operations, e.g. multiplication and addition, on values represented in $O(m)$ bits. It follows, via Fact 10, that an appropriate SLP defining σ can be efficiently constructed.

We claim that $\langle C, \underline{x}, \underline{y}, \underline{z}, \underline{w}\rangle$ is accepted as an instance of ACS if and only if $\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)}\rangle$ is accepted as an instance of **1**-PATH.

Suppose that $\langle C, \underline{x}, \underline{y}, \underline{z}, \underline{w}\rangle \in \mathcal{L}_{\text{ACS}}$ and let

$$\Gamma = \langle \underline{x}_0, \underline{y}_0\rangle, \ldots, \langle \underline{x}_i, \underline{y}_i\rangle, \ldots, \langle \underline{x}_p, \underline{y}_p\rangle$$

be the sequence of instantiations in $\langle 0, 1\rangle^{n+m}$ witnessing this. Consider the sequence of allocations

$$\langle Q^{(0)}, Q^{(1)}, \ldots, Q^{(p)}\rangle$$

[10] To avoid excessive repetition, when, for $S \subseteq \mathcal{R}^V \cup \mathcal{R}^W$, we refer to $\beta(S)$ in specifying any of these six conditions, it should be taken that $\beta(S) \neq \perp$: should this fail to be the case then the condition in question is not satisfied.

[11] By imposing this condition, which is not strictly necessary for the subsequent argument, we can simplify the analysis of one particular case in proving the correctness of the reduction.

[12] Choosing $K_{mn} = 3(m + n) + 2$ suffices for σ to have the properties needed in the subsequent proof and since this value is represented in $O(\log mn)$-bits the polynomial-time computability of the reduction from ACS is unaffected.

[13] Although, it is possible for Q to satisfy both of C1 and C2 or both of C4 and C5 in the cases where this arises the value that results for $\sigma(Q)$ applying C1 (resp. C4) is the same as the value that results using C2 (resp. C5).

in which

$$\begin{aligned}
Q_1^{(i)} &= \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle) \cup \beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle)\\
Q_2^{(i)} &= \mathcal{R}^V \setminus Q_1^{(i)}\\
Q_3^{(i)} &= \mathcal{R}^W \setminus Q_1^{(i)}\\
Q_4^{(i)} &= \emptyset\\
Q_5^{(i)} &= \{\mu\}.
\end{aligned}$$

For each of these, C1($Q^{(i)}$) holds: when $Q = Q^{(i)}$ we have

$$\begin{aligned}
\beta(Q_1^V) &= \beta(Q_1^W) = \langle \underline{x}_i, \underline{y}_i\rangle\\
Q_4 &= \emptyset \subseteq \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W))).
\end{aligned}$$

In addition,

$$\begin{aligned}
\sigma(Q^{(i)}) = 2K_{mn}val_m(\langle \underline{x}_i, \underline{y}_i\rangle) &= 2K_{mn}val(\underline{y}_i)\\
&< 2K_{mn}val(\underline{y}_{i+1}) = 2K_{mn}val_m(\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)\\
&= \sigma(Q^{(i+1)}).
\end{aligned}$$

So that the sequence of allocations $\langle Q^{(0)}, Q^{(1)}, \ldots, Q^{(p)}\rangle$ is σ-rational. This sequence, however, is *not* 1-bounded, and so to complete the argument that positive instances of ACS yield positive instances of **1**-PATH with the reduction described, we need to construct a 1-bounded, σ-rational sequence for each of the deals $\langle Q^{(i)}, Q^{(i+1)}\rangle$.

Consider any $Q^{(i)}$ for some $0 \le i < p$ and the following sequences of 1-bounded deals starting with $Q^{(i)}$.

S1. Using 1-bounded deals, transfer the set $\text{DIFF}_V(Q_1^{(i),V}, C(\beta(Q_1^{(i),W})))$ from A_2 to A_4, giving the allocation $S^{(i),1}$.

Let $T^{(j)}$ be the allocation resulting after exactly j resources have been moved from A_2 to A_4, so that $T^{(0)} = Q^{(i)}$ and $T^{(d)} = S^{(i),1}$, (with $d = |\text{DIFF}_V(Q_1^{(i),V}, C(\beta(Q_1^{(i),W})))|$).

Since the resources held by A_1 are unchanged by the deal $\langle T^{(j-1)}, T^{(j)}\rangle$ it follows that each of the allocations $T^{(j)}$ satisfies C1. In addition, $T^{(d)}$ also satisfies C2. Each of these deals is σ-rational, since for $0 \le j \le d$: $\sigma(T^{(j)}) = \sigma(T^{(0)}) + j$. We observe that using C2 to evaluate $T^{(d)}$ returns,

$$\sigma(T^{(d)}) = \sigma(T^{(0)}) + d + (n+m) - |Q_1^{(i),V}| = \sigma(T^{(0)}) + d$$

since, from the fact that C1 holds, $\beta(Q_1^{(i),V}) \neq \perp$ and this requires $|Q_1^{(i),V}| = n + m$.

S2. Using 1-bounded deals, transfer the set

$$D = \{v \in S_1^{(i),1,V} : \neg v \in S_4^{(i),1}\} \bigcup \{\neg v \in S_1^{(i),1,V} : v \in S_4^{(i),1}\}$$

from A_1 to A_2, to give the allocation $S^{(i),2}$.

Again denote by $T^{(j)}$ the allocation resulting after exactly j resources have been moved from A_1 to A_2, with $T^{(0)} = S^{(i),1}$, $T^{(d)} = S^{(i),2}$ and $d = |D|$. Notice that

$$d = |S_4^{(i),1}| = |\text{DIFF}_V(\beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle), C(\langle \underline{x}_i, \underline{y}_i\rangle))|.$$

Each of these allocations satisfies C2. To see this, first observe that the resources held by A_4 are unchanged by any of the deals $\langle T^{(j-1)}, T^{(j)}\rangle$: throughout this stage A_4 holds

$$\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i\rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle).$$

The subset of $\mathcal{R}^V$ held by A_1, initially $\beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle)$, is altered by transferring D to A_2. This set of resources, however, is exactly $\beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle) \setminus \beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i\rangle))$, so that from Lemma 1(a), in the allocation $T^{(j)}$, the subsets of $\mathcal{R}^V$ held by A_1 and A_4 have the respective forms,

$$\begin{aligned}
G &= \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle) \cap \beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i\rangle)) \cup D_j\\
H &= \text{DIFF}_V(G, C(\langle \underline{x}_i, \underline{y}_i\rangle)).
\end{aligned}$$

Applying Lemma 1(b), $\beta(G \otimes H) = C(\langle \underline{x}_i, \underline{y}_i \rangle)$, i.e. each of the allocations $T^{(j)}$ satisfies the conditions specified in C2. Finally we have

$$\sigma(T^{(j)}) = \sigma(T^{(0)}) + n + m - (n + m - j) = \sigma(T^{(0)}) + j$$

so that each of the deals $\langle T^{(j-1)}, T^{(j)} \rangle$ is σ-rational.

It should be noted that, in $S^{(i),2}$ we have

$$|S_1^{(i),2,V}| = n + m - |S_4^{(i),2}| = n + m - |\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i \rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)|$$

so that,

$$\sigma(S^{(i),2}) = 2K_{mn} val(\underline{y}_i) + 2|\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i \rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)|.$$

S3. Transfer the resource μ from A_5 to A_4 to give the allocation $S^{(i),3}$.

The allocation satisfies $S^{(i),3}$ satisfies C3, and has

$$S_1^{(i),3,W} = \beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle).$$

Furthermore,

$$|S_4^{(i),3}| = 1 + |\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i \rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)|.$$

With the evaluation measure σ

$$\begin{aligned} \sigma(S^{(i),2}) &= 2K_{mn} val(\underline{y}_i) + 2|\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i \rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)| \\ &< K_{mn} val(\underline{y}_i) + K_{mn} val(\underline{y}_{i+1}) - |\beta_V^{-1}(C(\langle \underline{x}_i, \underline{y}_i \rangle)) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)| - 1 \\ &= \sigma(S^{(i),3}). \end{aligned}$$

The deal $\langle S^{(i),2}, S^{(i),3} \rangle$ is σ-rational since with $val(\underline{y}_{i+1}) \geq val(\underline{y}_i) + 1$ and K_{mn} large enough,

$$\begin{aligned} \sigma(S^{(i),3}) - \sigma(S^{(i),2}) &\geq K_{mn} - 3|\beta_V^{-1}(\langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)| - 1 \\ &\geq K_{mn} - 3(n + m) - 1 \\ &> 0. \end{aligned}$$

S4. Using 1-bounded deals, transfer the set $S_4^{(i),3,V}$ from A_4 to A_1, giving $S^{(i),4}$.

Let $T^{(j)}$ be the allocation resulting after exactly j resources have been moved from A_4 to A_1, with $T^{(0)} = S^{(i),3}$ and $T^{(d)} = S^{(i),4}$ with

$$d = |S_4^{(i),3,V}| - 1 = |\beta_V^{-1}(\langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle)|.$$

Noting that

$$\begin{aligned} S_1^{(i),3,V} &= \beta_V^{-1}(\langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle) \cap \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle) \\ S_4^{(i),3,V} &= \beta_V^{-1}(\langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle) \setminus \beta_V^{-1}(\langle \underline{x}_i, \underline{y}_i \rangle) \end{aligned}$$

we see that each of the allocations, $T^{(j)}$ satisfies C3: $\beta(T_1^{(j),V} \cup T_4^{(j),V}) = C(\beta(T_1^{(j),W}))$. In addition

$$\sigma(T^{(j)}) = \sigma(T^{(0)}) + j$$

so each deal $\langle T^{(j-1)}, T^{(j)} \rangle$ is σ-rational. For the allocation, $S^{(i),4}$ we have

$$\sigma(S^{(i),4}) = K_{mn} val(\underline{y}_i) + K_{mn} val(\underline{y}_{i+1}) - 1.$$

S5. Transfer the resource μ from A_4 to A_5 giving $S^{(i),5}$.

The allocation $S^{(i),5}$ satisfies C4:

$$\begin{aligned} S_4^{(i),5} &= \emptyset &&\subseteq \text{DIFF}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle) \\ \beta(S_1^{(i),5,V}) &= \langle \underline{x}_{i+1}, \underline{y}_{i+1} \rangle &&= C(\beta(S_1^{(i),5,W})). \end{aligned}$$

The deal $\langle S^{(i),4}, S^{(i),5}\rangle$ is σ-rational since

$$\begin{aligned}\sigma(S^{(i),4}) &= K_{mn}val(\underline{y}_i) + K_{mn}val(\underline{y}_{i+1}) - 1\\ \sigma(S^{(i),5}) &= 2K_{mn}val(\underline{y}_{i+1}) - 2 - 3|\textsc{diff}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|\end{aligned}$$

so that, since $val(\underline{y}_{i+1}) \geq val(\underline{y}_i) + 1$,

$$\sigma(S^{(i),5}) - \sigma(S^{(i),4}) \geq K_{mn} - 1 - 3(n+m) > 0.$$

S6. Using 1-bounded deals, transfer the set $\textsc{diff}_W(S_1^{(i),5,W}, \beta(S_1^{(i),5,V}))$ from A_3 to A_4, to give the allocation $S^{(i),6}$.

Let $T^{(j)}$ be the allocation in place after exactly j resources have been transferred from A_3 to A_4, so that $T^{(0)} = S^{(i),5}$ and $T^{(d)} = S^{(i),6}$ with

$$d = |\textsc{diff}_W(S_1^{(i),5,W}, \beta(S_1^{(i),5,V}))|.$$

By similar arguments to those used when considering S1 above, we see that each of the allocations $T^{(j)}$ satisfies C4. The allocation $T^{(d)}$ in addition satisfies C5. The deal $\langle T^{(j-1)}, T^{(j)}\rangle$ is σ-rational since,

$$\sigma(T^{(j)}) = \sigma(T^{(0)}) + |T_4^{(j)}| = \sigma(T^{(0)}) + j.$$

We, further note, that $\sigma(T^{(d)})$ when evaluated by using C4 is,

$$2K_{mn}val(\underline{y}_{i+1}) - 2 - 2|\textsc{diff}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|$$

(since $|T_4^{(d)}| = |\textsc{diff}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|$), and if evaluated using C5,

$$\begin{aligned}\sigma(T^{(d)}) &= 2K_{mn}val(\underline{y}_{i+1}) - 2 - 2|\textsc{diff}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)| + n + m - |T_1^{(d),W})\\ &= 2K_{mn}val(\underline{y}_{i+1}) - 2 - 2|\textsc{diff}_W(S_1^{(i),5,W}, \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|.\end{aligned}$$

S7. Using 1-bounded deals, transfer the set

$$D = \{w \in S_1^{(i),6,W} \ : \ \neg w \in S_4^{(i),6}\} \bigcup \{\neg w \in S_1^{(i),6,W} \ : \ w \in S_4^{(i),6}\}$$

from A_1 to A_3 to give $S^{(i),7}$.

Let $T^{(j)}$ denote the allocation after exactly j resources have been transferred from A_1 to A_3, so that $T^{(0)} = S^{(i),6}$ and $T^{(d)} = S^{(i),7}$ with $d = |D|$. By a similar argument to that in S2,

$$d = |S_4^{(i),6}| = |\textsc{diff}_W(\beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle), \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|.$$

Again via Lemma 1 and the analysis of S2 it follows that each allocation $T^{(j)}$ satisfies C5. The deal $\langle T^{(j-1)}, T^{(j)}\rangle$ is σ-rational by virtue of the fact that $\sigma(T^{(j)}) = \sigma(T^{(0)}) + j$, so that

$$\begin{aligned}\sigma(S^{(i),7}) &= 2K_{mn}val(y_{i+1}) - 2 - 2|\textsc{diff}_W(\beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle), \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)| + n + m - |S_1^{(i),7,W}|\\ &= 2K_{mn}val(y_{i+1}) - 2 - |\textsc{diff}_W(\beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle), \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle)|.\end{aligned}$$

The last line following from the fact that

$$S_1^{(i),7,W} = \beta_W^{-1}(\langle \underline{x}_i, \underline{y}_i\rangle) \ \cap \ \beta_W^{-1}(\langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle).$$

S8. Transfer μ from A_5 to A_4 to give $S^{(i),8}$.

The allocation $S^{(i),8}$ satisfies C6 with the deal $\langle S^{(i),7}, S^{(i),8}\rangle$ being σ-rational:

$$\begin{aligned}\sigma(S^{(i),8}) &= 2K_{mn}val(y_{i+1}) - 1 - |S_4^{(i),7}|\\ &> 2K_{mn}val(y_{i+1}) - 2 - |S_4^{(i),7}|\\ &= \sigma(S^{(i),7}).\end{aligned}$$

S9. Using 1-bounded deals, transfer the set $S_4^{(i),8,W}$ from A_4 to A_1, giving $S^{(i),9}$.

Letting $T^{(j)}$ be the allocation after exactly j resources have been moved so that $T^{(0)} = S^{(i),8}$ and $T^{(d)} = S^{(i),9}$ with $d = |S_4^{(i),8,W}|$, each $T^{(j)}$ satisfies C6 and the deal $\langle T^{(j-1)}, T^{(j)}\rangle$ is σ-rational since $\sigma(T^{(j)}) = \sigma(T^{(0)}) + j$. The allocation $S^{(i),9}$ has $S_4^{(i),9} = \{\mu\}$ so that, $\sigma(S^{(i),9}) = 2K_{mn}val(\underline{y}_{i+1}) - 1$.

Furthermore, $S^{(i),9}$ has

$$\beta(S_1^{(i),9,V}) = \beta(S_1^{(i),9,W}) = \langle \underline{x}_{i+1}, \underline{y}_{i+1}\rangle.$$

S10. Transfer the resource μ from A_4 to A_5 giving $S^{(i),10}$. This allocation satisfies C1 and, since $\sigma(S^{(i),10}) = 2K_{mn}val(\underline{y}_{i+1})$ the deal $\langle S^{(i),9}, S^{(i),10}\rangle$ is σ-rational.

To complete the argument that positive instances of ACS induce positive instances of **1**-PATH in the reduction described, it suffices to note that the allocation $S^{(i),10}$ is exactly that described by $Q^{(i+1)}$.

It remains only to prove that should $\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)}\rangle$ describe a positive instance of **1**-PATH then the instance $\langle C, \langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle$ from which it arose described a positive instance of ACS.

Thus, let

$$\Gamma = \langle Q^{(0)}; Q^{(1)}; \cdots; Q^{(i)}; \cdots; Q^{(p)}\rangle$$

be a sequence of allocations for which

a. $Q^{(0)} = P^{(s)}$
b. $Q^{(p)} = P^{(t)}$
c. $\forall 1 \leq i \leq p$ $\langle Q^{(i-1)}, Q^{(i)}\rangle$ is 1-bounded *and* σ-rational.

Given an allocation $Q \in \Pi_{5,4(n+m)+1}$ we say that Q has the *assignment property* if

$$(\text{C1}(Q) \text{ holds and } Q_4 = \emptyset) \text{ OR } (\text{C3}(Q) \text{ holds and } Q_4 = \{\mu\}).$$

Consider the sub-sequence of Γ,

$$\Delta = \langle S^{(0)}; S^{(1)}; \cdots; S^{(d)}\rangle$$

such that every $S^{(j)}$ in Δ has the assignment property and if $\langle S^{(j)}, S^{(j+1)}\rangle$ correspond to allocations $\langle Q^{(i)}, Q^{(i+k)}\rangle$ in Γ then for every $1 \leq t < k$, the allocation $Q^{(i+t)}$ does *not* have the assignment property. Noting that $P^{(s)}$ and $P^{(t)}$ both have the assignment property, it is certainly the case that Δ can be formed and will have $S^{(0)} = P^{(s)}$ and $S^{(d)} = P^{(t)}$. Our aim is to use Δ to extract the witnessing sequence of instantiations from $\langle 0, 1\rangle^{n+m}$ certifying $\langle C, \langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle$ as a positive instance of ACS.

From Δ we may define a sequence of pairs – $\langle \underline{a}_i, \underline{b}_i\rangle \in \langle 0, 1\rangle^{n+m} \times \langle 0, 1\rangle^{n+m}$ – via $\underline{a}_i = \beta(S_1^{(i),V})$ and $\underline{b}_i = \beta(S_1^{(i),W})$. Since any allocation, Q, with the assignment property must satisfy either C1 or C3 it follows that $\beta(Q_1^V)$ and $\beta(Q_1^W)$ are both well-defined: if C1(Q) this is immediate from the specification of C1; if C3(Q) then since Q_4 must contain only the element μ it follows that $Q_4^V = \emptyset$ and, again, that $\beta(Q_1^V)$ is well-defined follows from the defining conditions for C3.

In order to extract the appropriate witnessing sequence for $\langle C, \langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle \in \mathcal{L}_{\text{ACS}}$ it suffices to show that $\langle \underline{a}_i, \underline{b}_i\rangle$ behaves as follows:

$$\langle \underline{a}_i, \underline{b}_i\rangle = \begin{cases} \langle \langle \underline{x}, \underline{y}\rangle, \langle \underline{x}, \underline{y}\rangle\rangle & \text{if } i = 0 \\ \langle C(\underline{a}_{i-1}), \underline{b}_{i-1}\rangle & \text{if } i > 0 \text{ and } i \text{ is odd} \\ \langle \underline{a}_{i-1}, C(\underline{b}_{i-1})\rangle & \text{if } i > 0 \text{ and } i \text{ is even.} \end{cases}$$

For the sequence $\{\langle \underline{a}_i, \underline{b}_i\rangle : 0 \leq i \leq d\}$ defined from $\Gamma = \langle S^{(0)}; \cdots S^{(d)}\rangle$ consider the sequence of 1-bounded, σ-rational deals that realise the (σ-rational) deal $\langle S^{(0)}, S^{(1)}\rangle$.

First observe that this must comprise three sequences — $\langle S^{(0)}, T^{(1)}\rangle$, $\langle T^{(1)}, T^{(2)}\rangle$, and $\langle T^{(2)}, T^{(3)}\rangle$ of 1-bounded, σ-rational deals implementing

$\langle S^{(0)}, T^{(1)}\rangle$ with C1($T^{(1)}$), C2($T^{(1)}$), and $T_4^{(1)} = \text{DIFF}_V(S_1^{(0),V}, C(\underline{b}_0))$
$\langle T^{(1)}, T^{(2)}\rangle$ with C3($T^{(2)}$) and $|T_1^{(2),V}| = n + m - |T_4^{(2)}|$
$\langle T^{(2)}, S^{(1)}\rangle$ with C3($S^{(1)}$) and $S_4^{(1),V} = \emptyset$.

To see this[14] consider the allocations, P, such that $\langle S^{(0)}, P\rangle$ is 1-bounded and σ-rational. Given that P must satisfy at least one of the conditions (C1) through (C6), and that C1($S^{(0)}$) holds, we must have $P_1 = S_1^{(0)}$, $P_3 = S_3^{(0)}$ and $P_5 = S_5^{(0)}$, i.e. $\langle S^{(0)}, P\rangle$ involves transferring some resource held by A_2 to A_4. Any such resource, however, must belong to the set $\text{DIFF}_V(S_1^{(0),V}, C(\underline{b}_0))$ or C1(P) will fail to hold. By similar arguments any 1-bounded, σ-rational continuation of P will eventually reach the allocation $T^{(1)}$. In the same way, considering any allocation P for which $\langle T^{(1)}, P\rangle$ is 1-bounded and σ-rational, it follows that $T_3^{(1)} = P_3$, $T_4^{(1)} = P_4$ and $T_5^{(1)} = P_5$ so that $\langle T^{(1)}, P\rangle$ transfers some resource between A_1 and A_2: the only choices for such transfers which preserve condition C2 are those $v \in T_1^{(1),V}$ for which $\neg v \in T_4^{(1)}$ or $\neg v \in T_1^{(1),V}$ for which $v \in T_4^{(1)}$. Eventually such transfers lead to the allocation $T^{(2)}$ described and, in the same way from $T^{(2)}$ to the allocation $S^{(1)}$.

From C1($T^{(1)}$) and C2($T^{(1)}$) we have

$$\beta(T_1^{(1),V}) = \underline{a}_0 = \underline{b}_0 = \beta(T_1^{(1),W}).$$

From C3($T^{(2)}$) we have

$$\beta(T_1^{(2),V} \cup T_4^{(2),V}) = C(\underline{b}_0) = C(\underline{a}_0).$$

So that, in total, from C3($S^{(1)}$) and $S_4^{(1),V} = \emptyset$ we obtain

$$\underline{a}_1 = C(\underline{a}_0) \ ; \ \underline{b}_1 = \underline{b}_0$$

as required.

In the same way, noting that $\langle C(\underline{a}_0), \underline{b}_0\rangle \neq \langle\langle \underline{z}, \underline{w}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle$, it cannot be the case that $S^{(1)} = S^{(d)}$. Thus, by similar arguments to those given above, we may identify further sequences – $\langle S^{(1)}, T^{(3)}\rangle$, $\langle T^{(3)}, T^{(4)}\rangle$ and $\langle T^{(4)}, S^{(2)}\rangle$ – of σ-rational, 1-bounded deals that realise $\langle S^{(1)}, S^{(2)}\rangle$. These have the form

$\langle S^{(1)}, T^{(3)}\rangle$ with C4($T^{(3)}$), C5($T^{(3)}$), and $T_4^{(3)} = \text{DIFF}_W(S_1^{(1),W}, \underline{a}_1)$
$\langle T^{(3)}, T^{(4)}\rangle$ with C6($T^{(4)}$) and $|T_1^{(4),W}| = n + m - |T_4^{(4)}|$
$\langle T^{(4)}, S^{(2)}\rangle$ with C1($S^{(2)}$) and $S_4^{(1)} = \emptyset$.

From C4($T^{(3)}$) and C5($T^{(3)}$) we have

$$\beta(T_1^{(3),V}) = \underline{a}_1 = C(\underline{a}_0)$$
$$\beta(T_1^{(3),W}) = \underline{b}_1 = \underline{b}_0.$$

From C6($T^{(4)}$) we obtain,

$$\beta(T_1^{(4),W} \cup T_4^{(4),W}) = \beta(T_1^{(4),V}) = \underline{a}_1.$$

Finally, C1($S^{(2)}$) and $S_4^{(2)} = \emptyset$ give

$$\underline{a}_2 = \beta(S_1^{(2),V}) = \underline{a}_1$$
$$\underline{b}_2 = \beta(S_1^{(2),W}) = \underline{a}_1 = C(\underline{b}_1) = C(\underline{b}_0).$$

Thus, $\underline{a}_2 = \underline{a}_1$ and $\underline{b}_2 = C(\underline{b}_1)$.

[14] For ease of presentation we give only a brief outline of the argument here. The (somewhat tedious) fuller expansion of individual cases is provided in Appendix.

Thus the assertion regarding $\{\langle \underline{a}_i, \underline{b}_i \rangle\}_{0 \leq i \leq d}$ follows by an identical analysis of the cases

$$\langle \underline{a}_2, \underline{b}_2 \rangle \, , \ldots , \, \langle \underline{a}_{2j}, \underline{b}_{2j} \rangle \, , \ldots .$$

We now easily obtain the witnessing sequence that $\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle \rangle$ is a positive instance of ACS simply by using,

$$\langle \underline{a}_0, \, \underline{a}_2 \, , \ldots , \, \underline{a}_{2j} \, , \ldots , \, \underline{a}_{2k} \rangle$$

where $d = 2k$. We have already seen that this satisfies

$$\begin{aligned} \underline{a}_0 &= \langle \underline{x}, \underline{y} \rangle \\ \underline{a}_{2k} &= \langle \underline{z}, \underline{w} \rangle \\ \forall 1 \leq i \leq k \;\; \underline{a}_{2i} &= C(\underline{a}_{2(i-1)}). \end{aligned}$$

This sequence, however, must also satisfy $val_m(\underline{a}_{2i}) > val_m(\underline{a}_{2(i-1)})$: the deal $\langle S^{(2(i-1))}, S^{(2i)} \rangle$ is σ-rational as it is realised during the 1-bounded, σ-rational implementation of $\langle P^{(s)}, P^{(t)} \rangle$. From the definition of σ, recalling that $\mathrm{C1}(S^{(2i)})$ and $S_4^{(2i)} = \emptyset$ we have

$$\begin{aligned} \sigma(S^{(2(i-1))}) &= 2K_{mn} val_m(\beta(S_1^{(2(i-1)),W})) \\ &= 2K_{mn} val_m(\beta(S_1^{(2(i-1)),V})) \\ &= 2K_{mn} val_m(\underline{a}_{2(i-1)}) \\ \sigma(S^{(2i)}) &= 2K_{mn} val_m(\beta(S_1^{(2i),W})) \\ &= 2K_{mn} val_m(\beta(S_1^{(2i),V})) \\ &= 2K_{mn} val_m(\underline{a}_{2i}) \end{aligned}$$

and hence $\sigma(S^{(2i)}) > \sigma(S^{(2(i-1))})$ gives $val_m(\underline{a}_{2i}) > val_m(\underline{a}_{2(i-1)})$ as required.

In summary we deduce that $\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle \rangle$ is a positive instance of ACS if and only if $\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)} \rangle$ is a positive instance of **1**-PATH, thereby completing the argument that **1**-PATH is PSPACE-complete. □

5.3. Translating from evaluation measures to utilities

In this section we show how settings $\langle \mathcal{A}, \mathcal{R}, \sigma \rangle$ involving *arbitrary* evaluation measures, σ, may be translated in a general way to settings $\langle \mathcal{A}, \mathcal{R}', \mathcal{U} \rangle$ with utility functions so that utilitarian social welfare (σ_u) in the translated context mirrors the evaluation measure (σ) in the original setting.

Consider any $\langle \mathcal{A}, \mathcal{R}, \sigma \rangle$ with $|\mathcal{A}| = n$, $|\mathcal{R}| = m$ and $\sigma \; : \; \Pi_{n,m} \to \mathbf{Q}$, where it is assumed that for all $P \in \Pi_{n,m}$, $\sigma(P) \geq -1$. The *resource translation*

$$\tau(\mathcal{A}, \mathcal{R}) = \mathcal{R}_\tau$$

has $\mathcal{R}_\tau = \mathcal{R} \times \mathcal{A}$. We define a partial mapping $\pi \; : \; 2^{\mathcal{R}_\tau} \; \to \; \Pi_{n,m}$ as follows

If either $\cup_{\langle r, A_i \rangle \in S} \{r\} \neq \mathcal{R}$ or there exists r, A_i, A_j $(i \neq j)$ with $\{\langle r, A_i \rangle, \langle r, A_j \rangle\} \subseteq S$, then $\pi(S) = \perp$, i.e. undefined. Otherwise

$$\pi(S) = \left\langle \bigcup_{\langle r, A_1 \rangle \in S} \{r\}; \; \bigcup_{\langle r, A_2 \rangle \in S} \{r\}; \cdots; \; \bigcup_{\langle r, A_n \rangle \in S} \{r\} \right\rangle.$$

We note that for any $P \in \Pi_{n,m}$ there is a uniquely defined $S \subseteq \mathcal{R}_\tau$ for which $\pi(S) = P$: we employ the notation $\pi^{-1}(P)$ to refer to this S.

The concept of resource translation now allows us to prove.

Theorem 14.

a. **1**-SWAP *is* PSPACE-*complete.*

b. IRO-PATH *is* PSPACE-*complete.*

Proof. In both results we use a reduction from **1**-PATH.

For (a), given $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ an instance of **1**-PATH, consider the instance of **1**-SWAP, $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ in which $\mathcal{B} = \{B_1, B_2\}$, $u_2(S) = 0$ for all $S \subseteq \mathcal{R}_\tau$ and

$$u_1(S) = \begin{cases} -2 & \text{if } \pi(S) = \perp \\ \sigma(\pi(S)) & \text{if } \pi(S) \neq \perp. \end{cases}$$

Since the instance of 1-SWAP has exactly two agents, any allocation $\langle Q_1, Q_2\rangle$ is completely determined by the subset of $\mathcal{R}_\tau$ allocated to B_1. Thus, to complete the reduction we set $Q_1^{(s)} = \pi^{-1}(P^{(s)})$ and, similarly, $Q_1^{(t)} = \pi^{-1}(P^{(t)})$.

We claim that $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ is accepted as an instance of **1**-PATH if and only if $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ is accepted as an instance of **1**-SWAP.

Suppose the former is the case and let

$$\Delta = \langle P^{(0)},\ P^{(1)}, \ldots,\ P^{(d)}\rangle$$

be a witnessing rational 1-bounded path. First notice that, as $u_2(S) = 0$ for all $S \subseteq \mathcal{R}_\tau$, so $\sigma_u(Q) = u_1(Q_1)$. It follows, therefore that

$$\forall 1 \leq k \leq d \quad u_1(\pi^{-1}(P^{(i-1)})) \ < \ u_1(\pi^{-1}(P^{(i)})).$$

That is to say, the sequence of successive allocations, $\langle Q_1^{(0)}, \ldots, Q_1^{(d)}\rangle$ to B_1 given by

$$\langle \pi^{-1}(P^{(0)}),\ \pi^{-1}(P^{(1)}),\ \ldots, \pi^{-1}(P^{(d)})\rangle$$

yields an IR path.

It is also the case, however, that the deal defined from $\langle \pi^{-1}(P^{(i-1)}), \pi^{-1}(P^{(i)})\rangle$ is a 1-SWAP. To see this, recall that $\langle P^{(i-1)}, P^{(i)}\rangle$ is 1-bounded. Let $\{A_j, A_k\}$ be the agents involved and $r \in \mathcal{R}$ be the resource transferred, without loss of generality, from A_j to A_k. Then,

$$\begin{array}{ll} \langle r, A_j\rangle \in \pi^{-1}(P^{(i-1)}); & \langle r, A_k\rangle \in \mathcal{R}_\tau \setminus \pi^{-1}(P^{(i-1)}) \\ \langle r, A_k\rangle \in \pi^{-1}(P^{(i)}); & \langle r, A_j\rangle \in \mathcal{R}_\tau \setminus \pi^{-1}(P^{(i)}) \end{array}$$

so that the deal corresponding to $\langle \pi^{-1}(P^{(i-1)}), \pi^{-1}(P^{(i)})\rangle$ is realised by exchanging $\langle r, A_j\rangle \in Q_1^{(i-1)}$ for $\langle r, A_k\rangle \in Q_2^{(i-1)}$. We deduce that if $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ is accepted as an instance of 1-PATH then $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ is accepted as an instance of **1**-SWAP.

Now suppose that $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ is accepted as an instance of **1**-SWAP, letting

$$\langle Q_1^{(0)},\ Q_1^{(1)}, \ldots,\ Q_1^{(d)}\rangle$$

be the sequence of successive allocations to B_1 witnessing this. Consider the sequence of allocations,

$$\langle \pi(Q_1^{(0)}),\ \pi(Q_1^{(1)}), \ldots,\ \pi(Q_1^{(d)})\rangle$$

of $\mathcal{R}$ among $\mathcal{A}$. It is certainly the case that for each $Q^{(i)}$, $\pi(Q_1^{(i)}) \neq \perp$ and $\sigma(\pi(Q_1^{(i-1)})) < \sigma(\pi(Q_1^{(i)}))$, so it remains to show that each of the deals $\langle \pi(Q_1^{(i-1)}), \pi(Q_1^{(i)})\rangle$ is 1–bounded. Let $\langle r, A_j\rangle \in Q_1^{(i-1)}$ and $\langle r', A_k\rangle \in Q_2^{(i-1)}$ be the resources featuring in the IR 1-SWAP deal $\langle Q^{(i-1)}, Q^{(i)}\rangle$ so that

$$\begin{aligned} Q_1^{(i)} &= Q_1^{(i-1)} \setminus \{\langle r, A_j\rangle\} \cup \{\langle r', A_k\rangle\} \\ Q_2^{(i)} &= Q_2^{(i-1)} \setminus \{\langle r', A_k\rangle\} \cup \{\langle r, A_j\rangle\}. \end{aligned}$$

Since $\pi(Q_1^{(i)}) \neq \perp$, we must have $\cup_{\langle r, A\rangle \in Q_1^{(i)}} r = \mathcal{R}$, and thus $r = r'$. It follows that the deal $\langle \pi(Q^{(i-1)}), \pi(Q^{(i)})\rangle$ corresponds to a single resource, r, being transferred from A_j to A_k, i.e. this deal is 1-bounded. In consequence, if $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ is accepted as an instance of **1**-SWAP then $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ is accepted as an instance of 1-PATH, completing the proof that **1**-SWAP is PSPACE-complete.

For (b), we employ a similar approach: given an instance $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)} \rangle$ of **1**-PATH we form an instance $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)} \rangle$ of IRO–PATH in which $\mathcal{B} = \{B_1, B_2\}$, $u_2(S) = 0$ for all $S \subseteq \mathcal{R}_\tau$ and $u_1(S)$ is now,

$$u_1(S) = \begin{cases} -2 & \text{if } |S| < |\mathcal{R}| \\ -2 & \text{if } |S| = |\mathcal{R}| \text{ and } \pi(S) = \perp \\ 2\sigma(\pi(S)) & \text{if } |S| = |\mathcal{R}| \text{ and } \pi(S) \neq \perp \\ -2 & \text{if } |S| > |\mathcal{R}| + 1 \\ -2 & \text{if } |S| = |\mathcal{R}| + 1 \text{ and for all } \langle r, A_j \rangle \in S, \pi(S \setminus \{r, A_j\}) = \perp. \end{cases}$$

The only unspecified case is that of, $|S| = |\mathcal{R}|$ and with $\pi(S \setminus \{\langle r, A_j \rangle\}) \neq \perp$ for some $\langle r, A_j \rangle \in S$. In this case, $u_1(S)$ is

$$2 \min_{\langle r, A_j \rangle \in S \,:\, \pi(S \setminus \{\langle r, A_j \rangle\}) \neq \perp} \sigma(\pi(S \setminus \{r, A_j\})) + 1.$$

To complete the construction we fix $Q_1^{(s)} = \pi^{-1}(P^{(s)})$ and $Q_1^{(t)} = \pi^{-1}(P^{(t)})$. As before suppose that

$$\Delta = \langle P^{(0)}, P^{(1)}, \ldots, P^{(d)} \rangle$$

witnesses to $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)} \rangle$ as a positive instance of **1**-PATH. The sequence of allocations to B_1, $\langle Q_1^{(0)}, \ldots, Q_1^{(d)} \rangle$ with $Q_1^{(i)} = \pi^{-1}(P^{(i)})$ is IR by the argument used in part (a). Although this sequence is not 1-bounded we can, however, modify it as follows. From the proof of part (a), we know that the deal $\langle Q^{(i-1)}), Q^{(i)} \rangle$ is a **1**-SWAP: let $\langle r, A_j \rangle \in Q_1^{(i-1)}$ and $\langle r, A_k \rangle \in Q_2^{(i-1)}$ be the resources swapped in order to form $Q^{(i)}$. The deal $\langle Q^{(i-1)}), Q^{(i)} \rangle$ may be implemented by,

$$\begin{aligned} Q_1^{(i-1),0} &= Q_1^{(i-1)} \\ Q_1^{(i-1),1} &= Q_1^{(i-1),0} \cup \{\langle r, A_k \rangle\} \\ Q_1^{(i-1),2} &= Q_1^{(i-1),1} \setminus \{\langle r, A_j \rangle\} \\ Q_1^{(i)} &= Q_1^{(i-1),2}. \end{aligned}$$

This defines a sequence of 1-bounded deals implementing $\langle Q^{(i-1)}, Q^{(i)} \rangle$. In addition

$$\begin{aligned} u_1(Q_1^{(i-1),0}) &= 2\sigma(\pi(Q_1^{(i-1)})) \\ &< 2\sigma(\pi(Q_1^{(i-1)})) + 1 \\ &= u_1(Q_1^{(i-1),1}) \\ &< 2\sigma(\pi(Q_1^{(i)})) \\ &= u_1(Q^{(i-1),2}) = u_1(Q_1^{(i)}). \end{aligned}$$

Notice that $u_1(Q_1^{(i-1),1}) = 2\sigma(\pi(Q_1^{(i-1)})) + 1$, follows from the fact that there are exactly two choices of $\langle r, A \rangle \in Q_1^{(i-1),1}$ for which $\pi(Q_1^{(i-1),1} \setminus \{\langle r, A \rangle\}) \neq \perp$: one of these is $\langle r, A_k \rangle$; the other being $\langle r, A_j \rangle$. From the premise that we have a positive instance of **1**–PATH, it follows $\sigma(P^{(i-1)}) < \sigma(P^{(i)})$ so that

$$\begin{aligned} &\sigma(P^{(i-1)}) = \sigma(\pi(Q_1^{(i-1)})) = \sigma(\pi(Q^{(i-1),1} \setminus \{\langle r, A_k \rangle\})) \\ &\sigma(P^{(i)}) = \sigma(\pi(Q_1^{(i)})) = \sigma(\pi(Q^{(i-1),1} \setminus \{\langle r, A_j \rangle\})). \end{aligned}$$

Thus, if $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)} \rangle$ is a positive instance of **1**–PATH then we can construct an IR 1-bounded path in the instance $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)} \rangle$ of IRO–PATH.

For the converse, given

$$\langle Q^{(0)}, Q^{(1)}, \ldots, Q^{(d)} \rangle$$

establishing that $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)} \rangle$ is accepted as an instance of IRO–PATH, it is easy to see that $|Q_1^{(i)}| = |\mathcal{R}|$ if and only if i is even, with $|Q_1^{(i)}| = |\mathcal{R}| + 1$ whenever i is odd. Furthermore, $\pi(Q_1^{(2j)}) \neq \perp$, and

$$\sigma(\pi(Q_1^{(2(j-1))})) = u_1(Q_1^{(2(j-1))})/2 < u_1(Q_1^{(2j)})/2 = \sigma(\pi(Q_1^{(2j)})).$$

By similar arguments used to those in part (a), from the fact that the deal $\langle Q^{2(j-1)}, Q^{(2j)}\rangle$ must be an IR 1-SWAP we deduce that $\langle \pi(Q_1^{(2(j-1))}), \pi(Q_1^{(2j)})\rangle$ is a σ-rational 1-bounded deal. Hence if $\langle \mathcal{B}, \mathcal{R}_\tau, \mathcal{U}, Q^{(s)}, Q^{(t)}\rangle$ is accepted as an instance of IRO–PATH then $\langle \mathcal{A}, \mathcal{R}, \sigma, P^{(s)}, P^{(t)}\rangle$ is a positive instance of **1**–PATH, thus establishing that IRO-path is PSPACE-complete. □

6. Convergence and accessibility

Our analyses of the preceding sections consider one effect of restricting agent negotiation methods. The main focus being on the complexity of deciding whether a particular reallocation may be achieved. As we noted in the introduction, such issues can be seen as addressing a rather localised property. In this section our aim is to consider two different questions, one – Convergence – of a rather more "global" nature, the other – Accessibility – falling in between the extremes represented by Convergence and the variants of Φ-**PATH** examined in Section 5. To clarify this point we now give formal definitions of the problems Φ-Convergence and Φ-Accessibility. In the same style used in defining Φ-**PATH** we give a version (for Φ-Accessibility) both in terms of evaluation measures and social welfare via specific utility functions. For the decision problem Φ-Convergence, however, only the utility form is used, it being possible to determine complexity bounds for this in a straightforward manner, i.e. without recourse to devices such as those used in the proof of Theorem 14.

Recall that $\Phi(P, Q)$ is a predicate on deals and that a sequence of allocations

$$\Delta = \langle P^{(0)}; P^{(1)}; \ldots; P^{(d-1)}; P^{(d)}\rangle$$

is said to be a Φ-path for the deal $\langle P^{(0)}, P^{(d)}\rangle$ if $\Phi(P^{(i-1)}, P^{(i)})$ holds for each $1 \leq i \leq d$. We say that Δ is a *maximal* Φ-path if

$$\Delta = \langle P^{(0)}; P^{(1)}; \ldots; P^{(d-1)}; P^{(d)}\rangle \text{ and } \forall Q \in \Pi_{n,m}\ \ \neg\Phi(P^{(d)}, Q).$$

It is, of course, possible to choose $\Phi(P, Q)$ in such a way that maximal Φ-paths are not well-defined, e.g. consider $\Phi_{1\text{-bd}}(P, Q)$ the predicate which is true if and only if $\langle P, Q\rangle$ is 1-bounded. In this case, if $\Phi_{1\text{-bd}}(P, Q) = \top$ then $\Phi_{1\text{-bd}}(Q, P) = \top$ so that $\langle P; Q; P; Q; P; \ldots\rangle$ is a (non-terminating) $\Phi_{1\text{-bd}}$-path. For the instantiations of $\Phi(P, Q)$ we consider – specifically $\Phi^{\mathrm{E}}_{1\text{-bd},\sigma\text{-R}}$ and $\Phi^{\mathcal{U}}_{1\text{-bd,IR}}$ – infinite length paths cannot occur. More generally, if Φ satisfies $\forall P, Q\ \Phi(P, Q) \Rightarrow \sigma(Q) > \sigma(P)$ then there are no infinite Φ-paths.

For a maximal Φ-path Δ we use $last(\Delta)$ to denote the final allocation of $\mathcal{R}$ that results, i.e. $P^{(d)}$ in the notation above.

Finally, for $P \in \Pi_{n,m}$ we denote by $\max_\Phi(P)$ the set

$$\max_\Phi(P) = \{\Delta\ :\ \Delta \text{ is a maximal } \Phi\text{-path starting from } P\}.$$

For $\Phi \in \{\Phi^{\mathcal{U}}_{1\text{-bd,IR}}, \Phi^{\mathrm{E}}_{1\text{-bd},\sigma\text{-R}}\}$ we note that $\max_\Phi(P)$ is never empty. If there is no allocation Q for which $\Phi(P, Q)$ holds then $\max_\Phi(P) = \{\langle P\rangle\}$, the path containing exactly one allocation. It is also the case, as shown in [4, Thm. 3, p. 50], that there are example in which $\max_{\Phi^{\mathcal{U}}_{1\text{-bd,IR}}}(P)$ contains exactly one path Δ with $|\Delta| = \Omega(2^m)$.

Φ-**Convergence** (denoted by Φ-CONV)
Instance: $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$.
Question: Is it the case that

$$\forall P \in \Pi_{n,m}\ \ \forall \Delta \in \max_\Phi(P)\ \ \forall Q \in \Pi_{n,m}\ \ \sigma_u(last(\Delta)) \geq \sigma_u(Q)\ ?$$

Less formally, Φ-CONV asks whether an instance $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ has the following property. Given *any* initial allocation ($P \in \Pi_{n,m}$), is it the case that regardless of which Φ-path Δ is followed, one will always reach an allocation $last(\Delta)$ that maximises σ_u?

Φ-**Accessible**$^{\mathrm{E}}$ (denoted by Φ-ACC$^{\mathrm{E}}$)
Instance: $\langle \mathcal{A}, \mathcal{R}, \sigma\rangle$ and $P \in \Pi_{n,m}$
Question: Is it the case that

$$\exists\ \Delta \in \max_\Phi(P) \text{ such that } \forall Q \in \Pi_{n,m}\ \ \sigma(last(\Delta)) \geq \sigma(Q)\ ?$$

Φ-**Accessible**$^{\mathcal{U}}$ (denoted by Φ-ACC$^{\mathcal{U}}$)
Instance: $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ and $P \in \Pi_{n,m}$
Question: Is it the case that

$$\exists\, \Delta \in \max_{\Phi}(P) \text{ such that } \forall Q \in \Pi_{n,m}\ \ \sigma_u(last(\Delta)) \geq \sigma_u(Q)\ ?$$

We consider the special cases defined from the predicates $\Phi^{\mathrm{E}}_{\text{1-bd},\sigma\text{-R}}$ and $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$ introduced at the end of Section 3.

In the specific cases of 1-bounded IR deals, both of these problems are of some practical interest: in settings yielding positive instances of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV, it is guaranteed that starting from any allocation and following *any* sequence of 1-bounded IR deals from this will eventually end with an optimal allocation. Similarly, in the case of positive instances of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-ACC$^{\mathcal{U}}$, it will be known that *some* sequence of rational 1-bounded deals will lead to an optimal allocation.

Theorem 15. $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV *is co*NP*-complete.*

Proof. To show $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV is in coNP, given $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ it suffices to test whether the following predicate is true of all pairs of allocations P, Q in $\Pi_{n,m}$:

$$\chi(P, Q) = (\sigma_u(P) < \sigma_u(Q)) \ \Rightarrow\ (\exists\, R \text{ such that } \Phi^{\mathcal{U}}_{\text{1-bd,IR}}(P, R)).$$

Certainly $\chi(P, Q)$ can be evaluated in deterministic polynomial-time since there are exactly $m(n-1)$ 1-bounded deals consistent with P. To see this algorithm correctly decides instances of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV, suppose $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ should be accepted: then any allocation $P \in \Pi_{n,m}$ is either optimal (so $\chi(P, Q)$ always holds since the premise $\sigma_u(P) < \sigma_u(Q)$ is always false) or (if sub-optimal) cannot be $last(\Delta)$ on any maximal $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-path, i.e. there is at least one IR 1-bounded deal $\langle P, R \rangle$ available.

On the other hand, suppose the instance $\langle \mathcal{A}, \mathcal{R}, \mathcal{U} \rangle$ should *not* be accepted. Then there is some maximal $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-path Δ, whose final allocation, $last(\Delta)$ is sub-optimal. Since $last(\Delta)$ is sub-optimal there is an allocation Q with $\sigma_u(last(\Delta)) < \sigma_u(Q)$: as a result $\chi(last(\Delta), Q) = \bot$ and such instances would fail to be accepted.

To prove coNP-hardness we use a reduction from UNSAT, an instance of which is a 3-CNF formula

$$\psi(x_1, x_2, \ldots, x_n) = \bigwedge_{i=1}^{t} (y_{i,1} \vee y_{i,2} \vee y_{i,3})$$

where

$$y_{i,j} \in \{x_1, x_2, \ldots, x_n, \neg x_1, \neg x_2, \ldots, \neg x_n\}.$$

We say that a subset

$$S \subseteq \{x_1, x_2, \ldots, x_n, \neg x_1, \neg x_2, \ldots, \neg x_n\}$$

is *useful for* ψ if $|S| = n$, S contains exactly one of each of the literals $\{x_i, \neg x_i\}$, and the instantiation formed by setting each literal in S to $\top$ satisfies ψ.

Given $\psi(x_1, x_2, \ldots, x_n)$, the instance $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV has

$$\begin{aligned} \mathcal{A}_\psi &= \{a_1, a_2\} \\ \mathcal{R}_\psi &= \{x_1, x_2, \ldots, x_n, \neg x_1, \neg x_2, \ldots, \neg x_n\} \\ \mathcal{U}_\psi &= \langle u_1, u_2 \rangle \end{aligned}$$

with

$$u_1(S) = \begin{cases} 2n+1 & \text{if } S \text{ is useful for } \psi \\ 2|S| & \text{otherwise} \end{cases}$$

$$u_2(S) = |S|.$$

We claim that $\psi(x_1, \ldots, x_n)$ is unsatisfiable if and only if $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ is accepted as an instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV.

First observe that the allocation $P^{opt} = \langle \mathcal{R}_\psi; \emptyset \rangle$ has $\sigma_u(P^{opt}) = 4n$, and every other allocation, Q, has $\sigma_u(Q) < 4n$. Thus to complete the proof, it suffices to show that ψ is unsatisfiable if and only if every maximal $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-path, Δ within $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ has $last(\Delta) = P^{opt}$.

Suppose ψ is unsatisfiable and consider any allocation $\langle S, \mathcal{R}_\psi \setminus S \rangle$. Since ψ is unsatisfiable, it follows that $u_1(S) = 2|S|$ for every $S \subseteq \mathcal{R}_\psi$ (since there are no subsets that are useful for ψ). Thus, the *only* IR 1-bounded deals possible must involve a transfer of a single literal held by a_2 to a_1: any transfer from a_1 to a_2 reduces u_1 by *exactly* 2 while increasing u_2 by exactly one. It follows that any maximal $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-path, Δ, from $\langle S, \mathcal{R}_\psi \setminus S \rangle$ has $last(\Delta) = \langle \mathcal{R}_\psi, \emptyset \rangle$, i.e. if ψ is unsatisfiable then $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ is accepted as an instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV.

On the other hand, suppose that $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ is accepted as an instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV. We show that ψ must be unsatisfiable. Assume the contrary, letting $\{y_1, \ldots, y_{n-1}, y_n\}$ be a set of n literals whose instantiation to $\top$ satisfies ψ. Now consider the allocation

$$P = \langle \{y_1, \ldots, y_{n-1}\}; \mathcal{R}_\psi \setminus \{y_1, \ldots, y_{n-1}\} \rangle.$$

We have $\sigma_u(P) = 2n-2+n+1 = 3n-1$. Consider the 1-bounded deal $\langle P, Q \rangle$ under which y_n is transferred from a_2 to a_1. For this, since the set $\{y_1, \ldots, y_{n-1}, y_n\}$ is *useful* we get $\sigma_u(Q) = 2n+1+n = 3n+1$, so that $\langle P, Q \rangle$ is IR. Any subsequent 1-bounded deal $\langle Q, Q' \rangle$, will not, however, be IR: we have seen that this must involve a single resource transfer from a_2 to a_1, but then $\sigma_u(Q') = 2n + 2 + n - 1 = 3n + 1$ with no increase in welfare, contradicting the premise that $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ is accepted as an instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV. We deduce that the assumption that ψ is satisfiable cannot hold, i.e. if $\langle \mathcal{A}_\psi, \mathcal{R}_\psi, \mathcal{U}_\psi \rangle$ is accepted as an instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV. then ψ is unsatisfiable. □

Thus, in contrast to IRO-PATH considered in Theorem 14(b), whose complexity is PSPACE-complete, the (superficially) more difficult question represented by $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV. is coNP-complete, i.e. under the usual assumptions significantly easier. This reduced complexity is easily accounted for by the properties of the predicate $\chi(P, Q)$ used in the proof. We note that $\chi(P, Q)$ is polynomial-time decidable by virtue of there being only a "small" (polynomially many) number of cases to consider, i.e. 1-bounded deals compatible with the allocation P. If, however, we consider Φ-CONV when $\Phi(P, Q)$ is such that there may be superpolynomially many Φ-deals compatible with any given P, then although we cannot guarantee coNP as an upper bound, provided that $\Phi(P, Q)$ itself is polynomial-time decidable, Φ-CONV is ("at worst") in Π^p_2, i.e. still somewhat easier than Φ-PATH. To see this, it suffices to note that the following predicate, $\chi'(P, Q)$ is decidable by an NP algorithm:

$$\chi'(P, Q) \equiv (\sigma_u(P) < \sigma_u(Q)) \Rightarrow \exists R \in \Pi_{n,m} : \Phi(P, R) \wedge (\sigma_u(R) > \sigma_u(P)).$$

Turning to the problem, Φ-ACC$^{\mathcal{U}}$, notice that we have the following progression

Problem	Number of allocations in Instance	Complexity
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-PATH	2	PSPACE-complete
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-ACC$^{\mathcal{U}}$	1	See below
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV	0	coNP-complete

Thus, in principle, we could hope that the classification of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}} - \text{ACC}^{\mathcal{U}}$ is "closer" to that of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV In practice, as demonstrated in the following results, such hopes turn out to be ill-founded.

Theorem 16. $\Phi^{\text{E}}_{\text{1-bd},\sigma\text{-R}}$-ACC$^{\text{E}}$ *is* PSPACE*-complete.*

Proof. For membership in PSPACE, given $\langle \langle \mathcal{A}, \mathcal{R}, \sigma \rangle, P \rangle$ we may use an NPSPACE algorithm, similar to that of Theorem 11, to choose $last(\Delta)$, for some $\Delta \in \max_\Phi(P)$. We may then test, in PSPACE, whether $\sigma(last(\Delta)) \geq \sigma(Q)$ for every $Q \in \Pi_{n,m}$ accepting if and only if this is the case. Noting that NPSPACE = PSPACE completes the argument.

To establish $\Phi^{\text{E}}_{\text{1-bd},\sigma\text{-R}}$-ACC$^{\text{E}}$ is PSPACE-hard, we show that ACS $\leq_p$ $\Phi^{\text{E}}_{\text{1-bd},\sigma\text{-R}}$-ACC$^{\text{E}}$. Given an instance $\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle \rangle$ of ACS we form an instance $\langle \langle \mathcal{A}_C, \mathcal{R}_C, \sigma' \rangle, P^{(C)} \rangle$ of $\Phi^{\text{E}}_{\text{1-bd},\sigma\text{-R}}$-ACC$^{\text{E}}$. This instance is *identical*

to that described in the proof of Theorem 13 except for the following details: $P^{(C)} = P^{(s)}$ the source allocation in the construction of Theorem 13; σ' is defined via

$$\sigma'(Q) = \begin{cases} -2 & \text{if } \sigma(Q) > \sigma(P^{(t)}) \text{ or } \sigma(Q) = \sigma(P^{(t)}) \text{ and } Q \neq P^{(t)}. \\ \sigma(Q) & \text{otherwise.} \end{cases}$$

This modification ensures that the allocation, $P^{(t)}$, in the proof of Theorem 13 is the unique allocation which *maximises* σ'. We now have, by exactly the same argument, that an optimal allocation is accessible from $P^{(C)}$ if and only if $\langle C, \langle \underline{x}, \underline{y} \rangle, \langle \underline{z}, \underline{w} \rangle \rangle$ is a positive instance of ACS. □

Corollary 17. $\Phi^{\mathcal{U}}_{1\text{-bd,IR}} - \text{ACC}^{\mathcal{U}}$ *is* PSPACE*-complete.*

Proof. Immediate by applying the translation of Theorem 14(b) to instances of $\Phi^{\text{E}}_{1\text{-bd},\sigma\text{-R}}\text{-ACC}^{\text{E}}$. □

7. Conclusions

The negotiation questions analysed in Theorem 14, consider environments in which agents may independently assess their resource holdings and attempt to obtain a "better" resource set by agreeing reallocations with other agents. In the most basic case, where only two agents are involved, extremely simple protocols[15] are sufficiently expressive to agree a partition of the resource set. Such schemes, even when limited to one resource at a time deals, are capable of achieving optimal (in the sense of maximising social welfare) allocations, provided that neither agent insists that given deals be IR. As we observed in the discussion opening Section 3, it is in the extreme case where rationality constraints are introduced, that significant problems arise with simple negotiation regimes. Some reallocations may be unrealisable, as demonstrated by [13]. Even if a particular reallocation *can* be realised by a sequence of 1-bounded rational deals, the number of deals involved may be exponentially larger than the number of 1-bounded deals required without the rationality condition imposed. Finally, deciding if such a sequence exists *at all*, a problem already known to be NP-hard from [6], is, in fact, (under the standard assumptions) unlikely even to belong to NP: Theorem 14 (b) proving this decision problem to be PSPACE-complete. Although we do not develop the proofs in detail here, it is straightforward to demonstrate that this level of complexity is not a property limited to negotiations attempting to improve social welfare. For example, when the notion of $\langle P, Q \rangle$ being "rational" is that of "cooperative rationality", then deciding if $\langle P^{(s)}, P^{(t)} \rangle$ is realisable by a sequence of 1-bounded, cooperatively rational deals is also PSPACE-complete.[16]

To conclude we raise some open questions relating to the computational complexity of the decision problems addressed when alternative formalisms are used for representing utility functions. We have noted that the SLP representation is general enough to describe any set of utility functions and can do so via a program of length comparable to the run-time of an optimal algorithm to compute the function's value. A number of alternative representation approaches have been proposed. While these are not being completely general they are of interest as compact representations. In particular, [7,2] introduced the class of *k-additive* functions as such a mechanism (Table 1).

A function $f : 2^{\mathcal{R}} \to \mathbf{Q}$ is said to k-additive if there are constants

$$\{\alpha_T : T \subseteq \mathcal{R}, |T| \leq k\}$$

for which

$$\forall S \subseteq \mathcal{R} \quad f(S) = \sum_{T \subseteq \mathcal{R} : |T| \leq k} \alpha_T \cdot I_T(S)$$

where $I_T(S)$ is the indicator function whose value is 1 if $T \subseteq S$ and 0 otherwise.

When $k = O(1)$, i.e. a constant, k-additive functions may be represented by the $O(m^k)$ values defining the characterising set of constants $\{\alpha_T\}$. It is, of course, the case that for any constant value of k, there will be functions

[15] For example, allowing an agent to make offers to buy/sell a single resource for a given price; to accept offers; and to decline these.

[16] This is a trivial consequence of the fact that $u_2(S) = 0$ in the reduction presented in Theorem 14 (b).

Table 1
Complexity of negotiation properties for 2-additive utility functions

Problem	Proven complexity
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV	coNP-complete
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-ACC$^{\mathcal{U}}$	NP-hard
$\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-PATH	Open

that cannot be expressed in k-additive form. In the special case of $k = 1$, it is shown in [2], that $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV is trivial: every system $\langle \mathcal{A}, \mathcal{R}, \mathcal{U}\rangle$ in which each u_i is 1-additive, is *a priori* a positive instance of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-CONV. For $k \geq 2$, however, the status of other decision problems is less clear. Thus, for $k = 2$, determining exact bounds for the accessibility and reachability problems when utility functions are 2-additive is likely to present significant problems. In particular, we have one unresolved issue which affects whether $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-PATH belongs to NP for this case. Thus, Dunne [4], introduces the following measures related to Φ-paths.

- $L^{opt}(P, Q)$: the length of the *shortest* Φ-path realising $\langle P, Q\rangle$.
- $L^{\max}(\mathcal{A}, \mathcal{R}, \mathcal{U})$: the maximum value of $L^{opt}(P, Q)$ over those deals for which a Φ-path exists.
- $\rho^{\max}(n, m)$: The maximum value (taken over all choices of utility function) of $L^{\max}(\mathcal{A}, \mathcal{R}, \mathcal{U})$.
- $\rho^{\max}_{C}(n, m)$: As $\rho^{\max}$, but with the maximisation taken over utility functions belonging to some class C.

In the case of $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}(P, Q)$, the function $\rho^{\max}(2, m)$ is shown to be exponential in m, a result which provides indications – justified by Theorem 14(b) – that $\Phi^{\mathcal{U}}_{\text{1-bd,IR}}$-PATH$\notin$NP. It is open, however, as to whether $\rho^{\max}_{\text{2-add}}(2, m)$ is superpolynomial in m. A proof to the contrary, i.e that $\rho^{\max}_{\text{2-add}}(2, m) = O(m^p)$ with $p = O(1)$ would in the light of Theorem 14(b) have some consequences of interest: both the accessibility and reachability problems for such utility functions would belong to NP. This contrasts with the PSPACE-hardness lower bounds for the general case that have been the basis of the main results of this paper.

Acknowledgements

The work reported in this article developed following the AgentLink Technical Forum Group on Multiagent Resource Allocation (TFG-MARA), Ljubljana, 28th February–2nd March 2005. In particular, the authors are grateful for the contributions of Ulle Endriss, Jérôme Lang, and Nicolas Maudet to discussions during this meeting.

Appendix. σ-rational, 1-bounded deals in the proof of Theorem 13

For completeness we present in this appendix the case analysis concerning one aspect of the proof of Theorem 13. This arises in the argument that

$$\langle \mathcal{A}_C, \mathcal{R}_C, \sigma, P^{(s)}, P^{(t)}\rangle \in \mathcal{L}_{\mathbf{1}\text{-PATH}} \;\Rightarrow\; \langle C, \langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle \in \mathcal{L}_{\text{ACS}}.$$

In particular, given $P \in \Pi_{5,4(n+m)+1}$ satisfying at least one of the conditions (C1) through (C6) listed above, we precisely characterise those allocations, Q, for which $\langle P, Q\rangle$ is σ-rational and 1-bounded.

We first note that P satisfies *exactly* one of the following:

$$\begin{array}{lll||lll} \text{a.} & \text{C1}(P) \wedge \neg\text{C2}(P) & & \text{d.} & \text{C4}(P) \wedge \neg\text{C5}(P) & \\ \text{b.} & \text{C2}(P) & & \text{e.} & \text{C5}(P) & (1) \\ \text{c.} & \text{C3}(P) & & \text{f.} & \text{C6}(P). & \end{array}$$

As a second point, although $\mathcal{A}_C$ has five agents and thus there are 20 possible choices for the combination of agent from whom a resource is transferred and to whom this resource is reallocated, in practice the 8 choices arising from

$$\left\{\begin{array}{lll} \langle A_2, A_3\rangle, & \langle A_3, A_2\rangle, & \\ \langle A_1, A_5\rangle, & \langle A_2, A_5\rangle, & \langle A_3, A_5\rangle \\ \langle A_5, A_1\rangle, & \langle A_5, A_2\rangle, & \langle A_5, A_3\rangle \end{array}\right\}, \qquad (2)$$

Table 2
1-bounded, rational successors of P

Line	P satisfies	From	To	Q satisfies	Conditions
1	$\mathrm{C1}(P) \wedge \neg\mathrm{C2}(P)$	A_2	A_4	$\mathrm{C1}(Q) \wedge \neg\mathrm{C2}(Q)$	$Q_4 \subset \mathrm{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$
2	$\mathrm{C1}(P) \wedge \neg\mathrm{C2}(P)$	A_2	A_4	$\mathrm{C2}(Q)$	$Q_4 = \mathrm{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$
3	$\mathrm{C2}(P)$	A_1	A_2	$\mathrm{C2}(Q)$	$\lvert P_1^V\rvert > n+m-\lvert P_4\rvert$
4	$\mathrm{C2}(P)$	A_5	A_4	$\mathrm{C3}(Q)$	$\lvert P_1^V\rvert = n+m-\lvert P_4\rvert$
5	$\mathrm{C3}(P)$	A_4	A_1	$\mathrm{C3}(Q)$	$\lvert P_1^V\rvert < n+m$
6	$\mathrm{C3}(P)$	A_4	A_5	$\mathrm{C4}(Q)$	$\lvert P_1^V\rvert = n+m$
7	$\mathrm{C4}(P) \wedge \neg\mathrm{C5}(P)$	A_3	A_4	$\mathrm{C4}(Q) \wedge \neg\mathrm{C5}(Q)$	$Q_4 \subset \mathrm{DIFF}_W(Q_1^W, \beta(Q_1^V))$
8	$\mathrm{C4}(P) \wedge \neg\mathrm{C5}(P)$	A_3	A_4	$\mathrm{C5}(Q)$	$Q_4 = \mathrm{DIFF}_W(Q_1^W, \beta(Q_1^V))$
9	$\mathrm{C5}(P)$	A_1	A_3	$\mathrm{C5}(Q)$	$\lvert P_1^W\rvert > n+m-\lvert P_4\rvert$
10	$\mathrm{C5}(P)$	A_5	A_4	$\mathrm{C6}(Q)$	$\lvert P_1^W\rvert = n+m-\lvert P_4\rvert$
11	$\mathrm{C6}(P)$	A_4	A_1	$\mathrm{C6}(Q)$	$\lvert P_1^W\rvert < n+m$
12	$\mathrm{C6}(P)$	A_4	A_5	$\mathrm{C1}(Q)$	$\lvert P_1^W\rvert = n+m$

need not be considered. If P satisfies the conditions described in (1) then a 1-bounded deal transferring a resource from A_i to A_j with $\langle A_i, A_j\rangle$ defined by (2), results in an allocation that fails at least one of the conditions (B1)–(B6) presented in the proof[17] of Theorem 13.

Given P satisfying (1), Table 2 characterises possible choices for Q such that $\langle P, Q\rangle$ is σ-rational and 1-bounded.

We wish to show that if the instance of **1**-PATH constructed from $\langle C, \langle \underline{x}, \underline{y}\rangle, \langle \underline{z}, \underline{w}\rangle\rangle$ is accepted then *every* 1-bounded, σ-rational path witnessing this must progress (from $P = P^{(s)}$) according to the sequence specified in Table 2, where we note that $P^{(s)}$ satisfies $\mathrm{C1}(P^{(s)}) \wedge \neg\mathrm{C2}(P^{(s)})$.

For ease of reference we recall the conditions (B1)–(B6) and (C1)–(C6) which must be satisfied in order for P to have $\sigma(P) \geq 0$

B1. $Q_1 \subseteq \mathcal{R}^V \cup \mathcal{R}^W$.
B2. $Q_2 \subseteq \mathcal{R}^V$.
B3. $Q_3 \subseteq \mathcal{R}^W$.
B4. $Q_4^V = \emptyset$ *or* $Q_4^W = \emptyset$.
B5. $Q_5 \subseteq \{\mu\}$, i.e. either $Q_5 = \emptyset$ or $Q_5 = \{\mu\}$.
B6. For $X \in \{V, W\}$, if $Q_i^X \neq \emptyset$ then for all j, $\{x_j, \neg x_j\} \not\subseteq Q_i^X$.

C1. $\beta(Q_1^V) = \beta(Q_1^W)$ and $Q_4 \subseteq \mathrm{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$.
C2. $\beta(Q_1^V \otimes Q_4^V) = C(\beta(Q_1^W))$ and $Q_4 = \mathrm{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$.
C3. $\beta(Q_1^V \cup Q_4^V) = C(\beta(Q_1^W))$ and $\mu \in Q_4$.
C4. $\beta(Q_1^V) = C(\beta(Q_1^W))$ and $Q_4 \subseteq \mathrm{DIFF}_W(Q_1^W, \beta(Q_1^V))$.
C5. $\beta(Q_1^V) = \beta(Q_1^W \otimes Q_4^W)$ and $Q_4 = \mathrm{DIFF}_W(Q_1^W, \beta(Q_1^V))$.
C6. $\beta(Q_1^V) = \beta(Q_1^W \cup Q_4^W)$ and $\mu \in Q_4$.

Similarly, we recall that $\sigma(Q)$ is given as,

C1	$2\,K_{mn}val_m(\beta(Q_1^W))$	$+\lvert Q_4\rvert$	
C2	$2\,K_{mn}val_m(\beta(Q_1^W))$	$+\lvert Q_4\rvert$	$+n+m-\lvert Q_1^V\rvert$
C3	$K_{mn}val_m(\beta(Q_1^W)) + K_{mn}val_m(C(\beta(Q_1^W)))$	$-\lvert Q_4\rvert$	
C4	$2\,K_{mn}val_m(\beta(Q_1^V))$	$+\lvert Q_4\rvert - 2$	$-3\lvert\mathrm{DIFF}_W(Q_1^W, \beta(Q_1^V))\rvert$
C5	$2\,K_{mn}val_m(\beta(Q_1^V))$	$-2\lvert Q_4\rvert - 2$	$+n+m-\lvert Q_1^W\rvert$
C6	$2\,K_{mn}val_m(\beta(Q_1^V))$	$-\lvert Q_4\rvert$	

with all other allocations having $\sigma(Q) = -1$.

[17] We recall that $\sigma(Q) \geq 0$ only if Q satisfies these six conditions.

We proceed by a case analysis of the different possibilities, where we use $from(P)$ to denote the agent from which a resource is transferred, $to(Q)$ for the agent receiving this resource in the 1-bounded deal $\langle P, Q\rangle$, and $r_P \in \mathcal{R}_C$ to denote the featured resource. We note that it suffices to present the analysis with respect to lines (1)–(6) of Table 2: lines (7) through (12) follow through a near identical argument.

Let $\langle P, Q\rangle$ be 1-bounded. Given the cases identified already in (2) we have the following.

Case 1*: C1*$(P) \wedge \neg$*C2*(P)

1(a) $from(P) = A_1; to(Q) = A_2$
If $r_P \in \mathcal{R}^W$ then Q fails to satisfy (B2), so we may assume $r_P = v \in P_1^V$. Since C1$(P) \wedge \neg$C2(P) holds, such a transfer will result in $\beta(Q_1^V)$ being ill-defined, a situation which is only allowed in (C2) and (C3): C3(Q) is ruled out since $\mu \notin Q_4$; C2(Q) requires $\beta(Q_1^V \otimes Q_4)$ to be well-defined and equal to $C(\beta(Q_1^W))$, but where this is the case then $\neg v \in Q_4 = P_4$ and hence $P_4 = \text{DIFF}_V(P_1^V, C(\beta(P_1^W)))$, contradicting the assumption $\neg$C2(P).

1(b) $from(P) = A_2; to(Q) = A_1$
In this case Q fails to satisfy $(B6)$ with respect to the subset Q_1^V.

1(c) $from(P) = A_1; to(Q) = A_3$
If $r_P \in P_1^V$ then Q fails (B3). If $r_P \in P_1^W$ then $\beta(Q_1^W)$ is ill-defined, a state only allowed with C6(Q) or C5(Q). The first cannot hold since $\mu \notin P_4$. The second is impossible also: $Q_4 = P_4$ and therefore $Q_4^W = \emptyset$ ensuring that $Q_1^W \otimes Q_4^W$ is ill-defined.

1(d) $from(P) = A_3; to(Q) = A_1$
For this case, Q fails to satisfy $(B6)$ with respect to the subset Q_1^W.

1(e) $from(P) = A_1; to(Q) = A_4$
If $r_P \in P_1^V$ then $\beta(Q_1^V)$ will be ill-defined and since $\mu \notin Q_4$ by virtue of the fact that C1$(P) \wedge \neg$C2(P), the only possible condition that Q could satisfy is (C2), i.e. $Q_4 = \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$ and $\beta(Q_1^V \otimes Q_4) = C(\beta(Q_1^W))$. Let $v = r_P$. If $\neg v \in Q_4$ then Q fails to meet condition $(B6)$. It now follows, from C2(Q) that $Q_1^V \otimes Q_4 = P_1^V \otimes P_4$, i.e. $P_4 = \text{DIFF}_V(P_1^V, C(\beta(P_1^W)))$ contradicting the assumption $\neg$C2(P).

If $r_P \in P_1^W$ then from the fact that C1$(P) \wedge \neg$C2(P), $\beta(Q_1^W)$ will be ill-defined, and since $\mu \notin P_4$ the only possibility is that C5(Q) holds, and thus $Q_4 = \{r_P\} = \text{DIFF}_W(Q_1^W, \beta(Q_1^V))$: notice that P_4 must be empty (as is implied by $Q_4 = \{r_P\}$), for otherwise Q would breach condition (B4) on account of $Q_4^V \neq \emptyset$ and $Q_4^W \neq \emptyset$. Comparing $\sigma(P)$ with $\sigma(Q)$ in this case, however, it is easily seen that $\langle P, Q\rangle$ cannot be σ-rational. Noting that $P_1^V = Q_1^V$ and $\beta(P_1^V) = \beta(P_1^W)$ we have,

$$\begin{aligned}\sigma(P) &= 2K_{mn}val_m(\beta(P_1^V))\\ \sigma(Q) &= 2K_{mn}val_m(\beta(P_1^V)) - 2|Q_4| - 2 + (n + m - |Q_1^W|)\\ &= 2K_{mn}val_m(\beta(P_1^V)) - 3.\end{aligned}$$

1(f) $from(P) = A_4; to(Q) = A_1$
In this case noting that $P_4 \subset \text{DIFF}_V(P_1^V, C(\beta(P_1^W)))$, via Lemma 1(a) and C1$(P)$ the resulting allocation would fail to satisfy $(B6)$ with respect to the set Q_1^V.

1(g) $from(P) = A_2; to(Q) = A_4$
Discussed at the end of Case 1.

1(h) $from(P) = A_4; to(Q) = A_2$
Given C1$(P) \wedge \neg$C2(P), C1(Q) can hold, however, $\langle P, Q\rangle$ cannot be σ-rational:

$$\begin{aligned}\sigma(P) &= 2K_{mn}val_m(\beta(P_1^V)) + |P_4|\\ \sigma(Q) &= 2K_{mn}val_m(\beta(P_1^V)) + |Q_4|\\ &= 2K_{mn}val_m(\beta(P_1^V)) + |P_4| - 1.\end{aligned}$$

1(i) $from(P) = A_3; to(Q) = A_4$
If $P_4 \neq \emptyset$ then from C1(P), Q will fail condition (B4). Again, from C1(P) both $\beta(P_1^V)$ and $\beta(P_1^W)$ are well defined and, thus, the only option open for Q is that C4(Q). In this case, however, $\langle P, Q\rangle$ cannot be σ-rational:

$$\begin{aligned}\sigma(P) &= 2K_{mn}val_m(\beta(P_1^V))\\ \sigma(Q) &\leq 2K_{mn}val_m(\beta(P_1^V)) + |Q_4| - 2\\ &\leq 2K_{mn}val_m(\beta(P_1^V)) - 1.\end{aligned}$$

1(j) $from(P) = A_4; to(Q) = A_3$
In this case, Q fails to satisfy (B3).

1(k) $from(P) = A_4; to(Q) = A_5$
From the fact that $\mu \notin P_4$, Q would breach (B5).

1(l) $from(P) = A_5; to(Q) = A_4$
The only options allowing $\mu \in Q_4$ are C3(Q) and C6(Q). In the first of these it must be the case that $Q_4^V = \emptyset$ for otherwise $\beta(Q_1^V \cup Q_4^V)$ is ill-defined. In this case, however, since $Q_1 = P_1$, we get from C1(P) that $\beta(P_1^V) = \beta(P_1^W) = C(\beta(P_1^W))$. It now follows that $\langle P, Q\rangle$ is not σ-rational

$$\begin{aligned}\sigma(P) &= 2K_{mn}val_m(\beta(P_1^W))\\ \sigma(Q) &= K_{mn}val_m(\beta(P_1^W)) + K_{mn}val_m(C(\beta(P_1^W))) - |Q_4|\\ &= 2K_{mn}val_m(\beta(P_1^W)) - 1.\end{aligned}$$

We are left only with Case 1(g) – $from(P) = A_2$ and $to(Q) = A_4$ – corresponding to the first two lines of Table 2 – and in order to preserve $\sigma(Q) \geq 0$ the only choice available for r_P is to as a member of the set $\text{DIFF}_V(P_1^V, C(\beta(P_1^W)))\backslash P_4$. Notice that, from $\neg$C2(P) this set is non-empty. We now have two possibilities for Q: C1(Q) $\wedge \neg$C2(Q), arising when

$$r_P \cup P_4 = Q_4 \quad \subset \quad \text{DIFF}_V(P_1^V, C(\beta(P_1^W))) = \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$$

and

$$r_P \cup P_4 = Q_4 = \text{DIFF}_V(P_1^V, C(\beta(P_1^W))) = \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W))).$$

The first is line (1) of Table 2; the second corresponds to line (2).

Case 2: C2(P)

2(a) $from(P) = A_1; to(Q) = A_2$
This is discussed at the end of Case 2.

2(b) $from(P) = A_2; to(Q) = A_1$
Although Q could satisfy (C2), the resulting deal would not be σ-rational: $|Q_1^V| > |P_1^V$ and $|Q_4| = |P_4|$.

2(c) $from(P) = A_1; to(Q) = A_3$
If $r_P \in P_1^V$ then Q fails condition (B3). If $r_P \in P_1^W$, then $\beta(Q_1^W)$ is ill-defined. In this case, however, C6(Q) cannot hold (since $\mu \notin P_4$), and C5(Q) cannot hold: from C2(P), we have $Q_4^W = \emptyset$ and thus $Q_1^W \otimes Q_4^W$ is ill-defined also.

2(d) $from(P) = A_3; to(Q) = A_1$
From C2(P) is follows that $\beta(P_1^W)$ is well-defined, but this would fail to be the case for Q_1^W which would have size $n + m + 1$.

2(e) $from(P) = A_1; to(Q) = A_4$
From C2(P) we have $P_4 = \text{DIFF}_V(P_1^V, C(\beta(P_1^W)))$, thus to retain $B6(Q)$ (with respect to Q_4) and $B4(Q)$, would require

$$r_P \in \beta_V^{-1}(\beta(P_1^W)) \cap \beta_V^{-1}(C(\beta(P_1^W))).$$

The resulting allocation, however, satisfies neither (C5) ($\mu \notin Q_4$) nor (C2) as $Q_4 \neq \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$: Q must satisfy one of these as $\beta(Q_1^V)$, is ill-defined.

2(f) $from(P) = A_4; to(Q) = A_1$
Similarly to 2(b), although Q could satisfy (C2), the resulting deal would not be σ-rational: $|Q_1^V| > |P_1^V$ and $|Q_4| < |P_4|$.

2(g) $from(P) = A_2; to(Q) = A_4$
From C2(P), $P_4 = \text{DIFF}_V(P_1^V, C(\beta(P_1^W)))$: since $Q_4 \neq \text{DIFF}_V(Q_1^V, C(\beta(Q_1^W)))$, Q cannot satisfy any of (C1) through (C6).

2(h) $from(P) = A_4; to(Q) = A_2$
The resulting allocation could satisfy $\text{C1}(Q) \wedge \neg\text{C2}(Q)$ (if $|P_1^V| = n + m$), however, $\langle P, Q\rangle$ would not be σ-rational: $\sigma(Q) = \sigma(P) - 1$.

2(i) $from(P) = A_3; to(Q) = A_4$
If $P_4 \neq \emptyset$ then Q fails to satisfy (B4). Otherwise, from C2(P) we have $\text{DIFF}_V(P_1^V, C(\beta(P_1^W))) = \emptyset$, i.e.

$$\beta(P_1^V) = C(\beta(P_1^W)) = \beta(P_1^W).$$

In this case, however, $P_1^V = Q_1^V$, $P_1^W = Q_1^W$ and both $\beta(P_1^V)$ and $\beta(P_1^W)$ are well-defined and from

$$\beta(P_1^V) = C(\beta(P_1^W)) = \beta(P_1^W)$$

it follows that $\text{DIFF}_W(Q_1^W, \beta(Q_1^V)) = \emptyset$ so that C4(Q) cannot hold.

2(j) $from(P) = A_4; to(Q) = A_3$
If $P_4 \neq \emptyset$ then C2(P) would lead to Q failing to satisfy (B3). If $P_4 = \emptyset$ then no transfer from A_4 to A_3 is possible.

2(k) $from(P) = A_4; to(Q) = A_5$ Since $\mu \notin P_4$ as a consequence of C2(P), any such transfer would result in Q failing to satisfy (B5).

2(l) $from(P) = A_5; to(Q) = A_4$
Dealt with below.

With the exception of Cases 2(a) and 2(l) each of the possible 1-bounded deals from P results in an allocation Q such that the deal $\langle P, Q\rangle$ fails to be σ-rational. For 2(a) – in which $from(P) = A_1$ and $to(Q) = A_2$ – we need only note that $r_P \in P_1^V$ (in order that (B2) is satisfied) and, for the conditions governing (C2) to continue to be true of Q, it must be the case that

$$r_P \in P_1^V \setminus \beta_V^{-1}(C(\beta(P_1^W))).$$

Such a choice of r_P is possible if and only if C2(P) with $|P_1^V| > n + m - |P_1^4|$, i.e. exactly the preconditions relevant for line (3) of Table 2. Case 2(l), with $from(P) = A_5$ and $to(Q) = A_4$, has only $r_P = \mu$ as an option. The resulting allocation, Q, given that C2(P) holds, will satisfy C3(Q) if and only if $\beta(Q_1^V \cup Q_4^V)$ is well-defined and equal to $C(\beta(Q_1^W))$: this is possible only in the conditions prescribed by line (4) or Table 2.

Case 3*: C3*(P)

We first recall the additional condition imposed in order that C3(P) holds. For

$$\underline{f} = \beta(P_1^W)$$
$$\underline{g} = C(\beta(P_1^W))$$

$val_m(\underline{g}) > val_m(\underline{f})$. This is useful for dealing with Case 3(k).

3(a) $from(P) = A_1; to(Q) = A_2$
As with previous cases, we must have $r_P \in P_1^V$ or $B2(Q)$ fails. From C3(P), however, we still have $\mu \in Q_4$ leaving only the option C3(Q): this, however, cannot hold since $\beta(P_1^V \cup P_4^V)$ is well-defined but $\beta(Q_1^V \cup Q_4^V) = \beta(P_1^V \setminus \{r_P\} \cup P_4^V)$ is not.

3(b) $from(P) = A_2; to(Q) = A_1$ In the same way as the previous case, from $\mu \in Q_4$, $\beta(Q_1^V \cup Q_4^V)$ will be ill-defined.

3(c) $from(P) = A_1$; $to(Q) = A_3$ We may assume $r_P \in P_1^W$ (otherwise (B3) fails to hold). As a result we have $\mu \in Q_4$ and $\beta(Q_1^W)$ ill-defined. From C3(P), $Q_4^W = \emptyset$, and so the resulting allocation is unable to satisfy (C6) the only option open.

3(d) $from(P) = A_3$; $to(Q) = A_1$ Again from C3(P), the instantiation $\beta(P_1^W)$ is well-defined: this will not be the case, however, for $\beta(P_1^W \cup \{r_P\})$, i.e. $\beta(Q_1^W)$.

3(e) $from(P) = A_1$; $to(Q) = A_4$
Although C3(Q) will hold, provided that $r_P \in P_1^V$, the deal $\langle P, Q\rangle$ will not be σ-rational: $|P_4| < |Q_4|$ thus $\sigma(P) = \sigma(Q) + 1$ using the evaluation condition for (C3).

3(f) $from(P) = A_4$; $to(Q) = A_1$
Considered at the end of Case 3.

3(g) $from(P) = A_2$; $to(Q) = A_4$ Such a transfer will result in $\beta(Q_1^V \cup Q_4^V)$ being ill-defined.

3(h) $from(P) = A_4$; $to(Q) = A_2$ Similarly, such a transfer results in $\beta(Q_1^V \cup Q_4^V)$ being ill-defined.

3(i) $from(P) = A_3$; $to(Q) = A_4$
From C3(P) it holds that $\mu \in Q_4$: if $Q_4 \neq \{\mu\}$ then ($B6$) fails to hold with respect to Q_4; on the other hand, if $Q_4^V = \emptyset$, then $\beta(Q_1^W \cup Q_4^W)$ is ill-defined thereby preventing the option C6(Q) from the fact that $\beta(P_1^W)$ is well-defined.

3(j) $from(P) = A_4$; $to(Q) = A_3$
Any choice of $r_P \in P_4$ results in Q_3 not satisfying (B3).

3(k) $from(P) = A_4$; $to(Q) = A_5$
Considered below.

3(l) $from(P) = A_5$; $to(Q) = A_4$
Given C3(P) we have $P_5 = \emptyset$ and thus no such transfer is possible.

The remaining two cases are 3(f) ($from(P) = A_4$, $to(Q) = A_1$) and 3(k) ($from(P) = A_4$; $to(Q) = A_5$). In the first of these, given that $r_P \neq \mu$ (condition (B1) must hold for Q), we have the case described by line (5) of Table 2. In the second, from (B5) the only choice is $r_P = \mu$. If it is the case that $Q_4 \neq \emptyset$, then the resulting allocation, Q, would satisfy (C2): now recalling that C3(P) enforces,

$$val_m(C(\beta(P_1^W))) > val_m(\beta(P_1^W))$$

were it the case that $Q_4 \neq \emptyset$ and C2(Q) the deal $\langle P, Q\rangle$ would not be σ-rational,

$$\begin{aligned}\sigma(Q) &\leq 2K_{mn}val_m(\beta(Q_1^W)) + |Q_4| + n + m\\ &= 2K_{mn}val_m(\beta(P_1^W)) + |P_4| - 1 + n + m\\ &< K_{mn}val_m(\beta(P_1^W)) + K_{mn}val_m(C(\beta(P_1^W))) - |P_4|\\ &= \sigma(P).\end{aligned}$$

References

[1] T. Bylander, The computational complexity of propositional STRIPS planning, Artificial Intelligence 69 (1994) 165–204.
[2] Y. Chevaleyre, U. Endriss, N. Maudet, On maximal classes of utility functions for efficient one-to-one negotiation, in: Proc. IJCAI-05, 2005.
[3] P.E. Dunne, The Complexity of Boolean Networks, Academic Press, 1988.
[4] P.E. Dunne, Extremal behaviour in multiagent contract negotiation, J. Artificial Intelligence Res. 23 (2005) 41–78.
[5] P.E. Dunne, M. Laurence, M. Wooldridge, Tractability results for automatic contracting negotiation, in: Proc. ECAI'04, 2004.
[6] P.E. Dunne, M. Wooldridge, M. Laurence, The complexity of contract negotiation, Artificial Intelligence 164 (2005) 23–46.
[7] U. Endriss, N. Maudet, On the communication complexity of multilateral trading, in: Proc. Third International Joint Conf. on Autonomous Agents and Multiagent Systems, AAMAS'04, 2004.
[8] U. Endriss, N. Maudet, F. Sadri, F. Toni, On optimal outcomes of negotiations over resources, in: Proc. Second International Joint Conf. on Autonomous Agents and Multiagent Systems, AAMAS'03, ACM Press, 2003.
[9] U. Endriss, N. Maudet, F. Sadri, F. Toni, Negotiating socially optimal allocations of resources, J. Artificial Intelligence Res. 25 (2006) 315–348.

[10] M. Fischer, N.J. Pippenger, Relations among complexity measures, J. ACM 26 (1979) 361–381.
[11] D.C. Parkes, L.H. Ungar, Iterative combinatorial auctions: Theory and practice, in: Proc. 17th National Conf. on Artificial Intelligence, AAAI-00, 2000.
[12] D.C. Parkes, L.H. Ungar, Preventing strategic manipulation in iterative auctions: Proxy agents and price adjustment, in: Proc. 17th National Conf. on Artificial Intelligence, AAAI-00, 2000.
[13] T.W. Sandholm, Contract types for satisficing task allocation: I theoretical results, in: AAAI Spring Symposium: Satisficing Models, 1998.
[14] T.W. Sandholm, Contract types for satisficing task allocation: II experimental results, in: AAAI Spring Symposium: Satisficing Models, 1998.
[15] T.W. Sandholm, Algorithm for optimal winner determination in combinatorial auctions, Artificial Intelligence 135 (2002) 1–54.
[16] T.W. Sandholm, S. Suri, Bob: Improved winner determination in combinatorial auctions and generalizations, Artificial Intelligence 145 (2003) 33–58.
[17] W.J. Savitch, Relationship between non-deterministic and deterministic tape classes, J. Comput. System Sci. 4 (1970) 177–192.
[18] C.P. Schnorr, The network complexity and turing machine complexity of finite functions, Acta Inform. 7 (1976) 95–107.
[19] R.G. Smith, The contract net protocol: High-level communication and control in a distributed problem solver, IEEE Trans. Comput. C-29 (12) (1980) 1104–1113.
[20] M. Tennenholz, Some tractable combinatorial auctions, in: Proc. 17th National Conf. on Artificial Intelligence, AAAI-00, 2000.
[21] M. Yokoo, Y. Sakurai, S. Matsubara, The effect of false-name bids in combinatorial auctions: New fraud in internet auctions, Games Econom. Behav. 46 (1) (2004) 174–188.