Coherence in Finite Argument Systems

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

Dept. of Computer Science, University of Liverpool, Liverpool L69 7ZF, United Kingdom

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.

Key words: Argument Systems, Coherence, Credulous and 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 been used to examine the resolution of conflicting norms, e.g. [12], especially where this is studied through the mechanism of

* Corresponding author.
Email address: `ped@csc.liv.ac.uk` (Paul E. Dunne).

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 Defn. 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.

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 [7] does 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*

	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

Table 1
Decision Problems in Finite Argument Systems and their Complexity

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, Thm 5.1, Lemma 5.2, Prop. 5.3] are certainly implied by these. In this context, it is worth drawing attention to some significant points regarding [7, Thm. 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 [8, Defn. 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 argument systems is that proofs of sceptical acceptance are (potentially) rather more readily demonstrated in coherent systems via devices such as those of [13]. The complexity of sceptical acceptance is considered (in the context of membership in preferred extensions) for various non-monotonic Logics by [5], where completeness results at the third-level of the polynomial-time hierarchy are demonstrated. Although [5] argue that their 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 extremely hard, there is an efficiently testable property that can be used 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,

$$w \in L \Leftrightarrow (\forall x \in X : |x| \leq p(|w|))(\exists y \in Y : |y| \leq p(|w|)) \ \langle w, x, y \rangle \in R_L$$

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,

[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)$

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 $\text{COHERENT}(\mathcal{H})$ is in $\Pi_2^{(p)}$.

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

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

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

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, $\neg\text{PREF-EXT}(\mathcal{H}, S)$ is in NP, i.e. $\Sigma_1^{(p)}$ and $\text{STAB-EXT}(\mathcal{H}, S)$ 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 often 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 vari-*

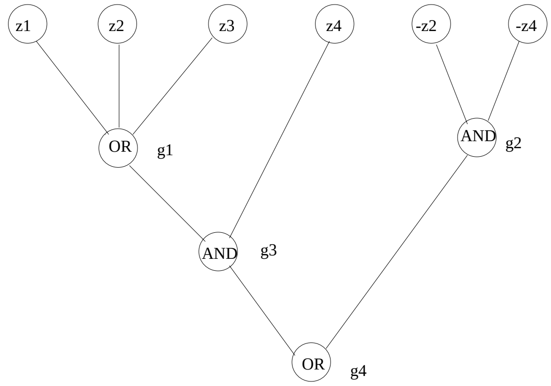

Fig. 1. $T_\Phi(z_1, z_2, z_3, z_4)$ for $(z_1 \vee z_2 \vee z_3) \wedge z_4 \vee (\neg z_2 \wedge \neg z_4)$

ables $Z = \langle z_1, z_2, \ldots, z_n \rangle$ 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 $\Phi(Z) = \wedge_{i=1}^{k} \Psi_i(Z)$, 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 $\Phi(Z) = \vee_{i=1}^{k} \Psi_i(Z)$, 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, $\pi : Z \to \{$**true**, **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 $\{$**true**, **false**$\}$ and* partial *otherwise. We define a partial ordering*

[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.

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

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

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 3 sets $\langle True(\pi), False(\pi), Open(\pi) \rangle$.

T1) The subset $True(\pi)$ consists of literals and gates, h, for which $\nu(h) = \mathbf{true}$.
T2) The subset $False(\pi)$ consists of literals and gates, h, for which $\nu(h) = \mathbf{false}$.
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 $\pi = \langle z_1 = \mathbf{true}, z_4 = \mathbf{false} \rangle$ 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 $Open(\pi) = \{z_2, \neg z_2, z_3, \neg z_3, g_2, g_4\}$.

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 trans-

[3] i.e. $\wedge_{j=1}^{k} x_j$ is $*$ unless all x_j are **true** or at least one x_j is **false**; $\vee_{j=1}^{k} x_j$ is $*$ unless all x_j are **false** or at least one is **true**.

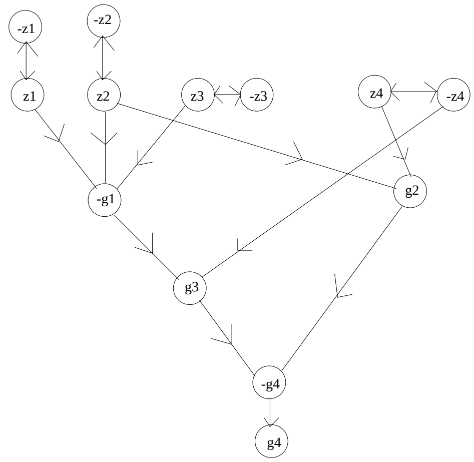

Fig. 2. The Argument System $\mathcal{R}_\Phi$ from the formula of Fig. 1

lation 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* $\{z_i, \neg z_i : 1 \leq i \leq n\}$.
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 $\{\langle z_i, \neg z_i\rangle, \langle \neg z_i, z_i\rangle : 1 \leq i \leq n\}$
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 $\{h_1, h_2, \ldots, h_r\}$: $\{\langle \neg h_i, g_k\rangle : 1 \leq i \leq r\}$.*
A4 *If g_k is an $\vee$-gate with inputs $\{h_1, h_2, \ldots, h_r\}$: $\{\langle h_i, \neg g_k\rangle : 1 \leq i \leq r\}$.*

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, 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:

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

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

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

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 Lemma 15 and Lemma 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 (resp. h_p) denotes the corresponding gate (resp. node) in T_Φ, hence if g_p is an $\vee$-gate, then p is the argument labelled $\neg g_p$; $\mathcal{PE}^{\mathcal{M}}$ (resp. $\mathcal{SE}^{\mathcal{M}}$) denotes the set of *all* preferred (resp. 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. $\square$

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

Proof: From Lemma 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. $\square$

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) $T_U := T_U / \{q \in T_U : \langle p, q\rangle \in \mathcal{A}_\Phi$ for some $p \in S_U\}$.
(5) $S_U := S_U \cup \{q \in T_U :$ 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 $In(\gamma(U)) \cap \{z_i, \neg z_i : 1 \le i \le n\}$, 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 $\Phi(\alpha_X, \beta_Y) =$ **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 $\Phi(\alpha_X, \beta_Y) =$ **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))

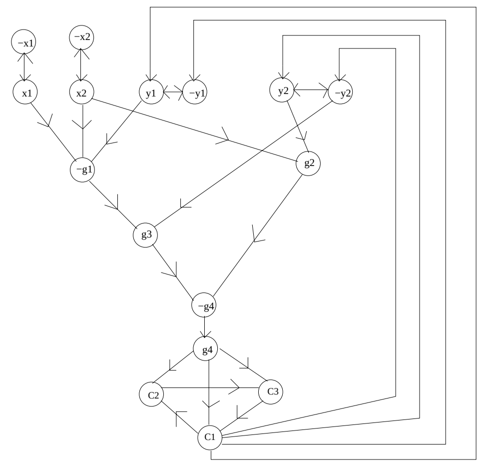

Fig. 3. An Augmented Argument Representation $\mathcal{H}_\Phi$

will yield a stable extension in $\mathcal{H}_\Phi$.

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,*

$$\{\langle C_1, y_i\rangle, \langle C_1, \neg y_i\rangle \ : \ 1 \le i \le n\}$$
$$\{\langle C_1, C_2\rangle, \langle C_2, C_3\rangle, \langle C_3, C_1\rangle\}$$
$$\{\langle g_m, C_1\rangle, \langle g_m, C_2\rangle, \langle g_m, C_3\rangle\}$$

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 Lemma 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 $\Phi(\alpha_X, \beta_Y) =$ **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 $1 \leq \kappa(g) \leq m$ 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 $\Phi(\alpha_X, \beta_Y) =$ **true**, and since the control arguments are not attacked, $In(\alpha_X) \notin \mathcal{SE}^{\mathcal{H}}$. ▫

[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 $In(\alpha_X)$.

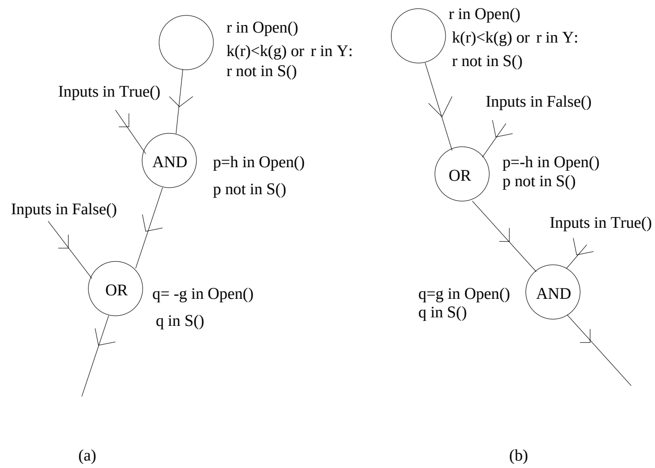

Fig. 4. Final cases in the proof of Lemma 16: $q \in Poss(\alpha_X)$ is not admissible

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

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) $newT_V := newT_V/\{q \in newT_V : \langle p, q\rangle \in \mathcal{B}_\Phi$ for some $p \in S_V\}$.
(5) $S_V := S_V \cup \{q \in newT_V :$ 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**. $\square$

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 Lemma 15 and Lemma 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, $\Phi(\alpha_X, \beta_Y) =$ **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. $\square$

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 $\mathcal{H}_\Phi$. $\square$

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 $0 \leq p \leq 1$. 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

[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.

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.

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 and R. Dechter. Default reasoning using classical logic. *Artificial Intelligence*, 84(1–2):113–150, 1996.

[2] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101, 1997.

[3] C. Cayrol, S. Doutre, and J. Mengin. Dialectical proof theories for the credulous preferred semantics of argumentation frameworks. In *Sixth European Conference on*

[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].

Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001), pages 668–679. Springer LNAI 2143, Springer-Verlag, September 2001.

[4] W. Fernandez de la Vega. Kernels in random graphs. *Discrete Mathematics*, 82:213–217, 1990.

[5] Y. Dimopoulos, B. Nebel, and F. Toni. Preferred arguments are harder to compute than stable extensions. In T. Dean, editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99-Vol1)*, pages 36–43, San Francisco, 1999. Morgan Kaufmann Publishers.

[6] Y. Dimopoulos, B. Nebel, and F. Toni. Finding admissible and preferred arguments can be very hard. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 53–61, San Francisco, 2000. Morgan Kaufmann.

[7] Y. Dimopoulos and A. Torres. Graph theoretical structures in logic programs and default theories. *Theoretical Computer Science*, 170:209–244, 1996.

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

[9] P.E. Dunne. *The Complexity of Boolean Networks*. Academic Press, 1988.

[10] P.E. Dunne and T.J.M. Bench-Capon. Two party immediate response disputes: Properties and efficiency. Technical report, Dept. of Comp. Sci., Univ. of Liverpool, `http://www.csc.liv.ac.uk/˜ped/papers/tpi.ps` (submitted), October 2001.

[11] H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law (ICAIL-99)*, ACM SIGART, pages 53–62, N.Y., June 1999. ACM Press.

[12] H. Prakken. *Logical Tools for Modelling Legal Argument*. Kluwer Academic Publishers, 1997.

[13] G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of JELIA'2000, The 7th European Workshop on Logic for Artificial Intelligence.*, pages 224–238, Berlin, 2000. Springer LNAI 1919, Springer Verlag.

[14] C. Wrathall. Complete sets and the polynomial-time hierarchy. *Theoretical Computer Science*, 3:23–33, 1976.