

Attacks on a Presumptive Argument Scheme in Multi-Agent Systems: pre-conditions in terms of beliefs and desires

Katie Atkinson, Trevor Bench-Capon and Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 3BX
UK

{k.m.atkinson,tbc,p.j.mcburney}@csc.liv.ac.uk

Abstract

In this paper we give definitions and pre-conditions to allow for the representation of a previously specified theory of persuasion over action by the authors to be used in multi agent systems.

1 Introduction

In [1] and [2] we presented a presumptive argument scheme with attacks to enable participants engaged in a dialogue to argue about the justification for a proposed course of action. Here we set out the pre-conditions and the form which the argument will take, which embodies the attacks. This formalism can then be used to program autonomous agents to reason and argue about practical action, in accordance with our theory.

2 Definitions

We now present the definitions of how an agent can put forward a position regarding the justification of an action, in accordance with our theory.

Definition 1: *The Beliefs of an Agent.* The beliefs of an Agent J is a four tuple $\langle W_J, A_J, D_J, V_J \rangle$ where,

W_J represents beliefs of Agent J about the world;
 A_J represents beliefs of Agent J about actions;
 D_J represents beliefs about the desires of Agent J;
 V_J represents beliefs about the values of Agent J;

Definition 2: Beliefs about the World. The beliefs about the world of Agent J is a set of triples $\langle p, \text{cert}_{pJ}, t \rangle$ where,

p is a proposition; $\text{cert}_{pJ} = -1 \leq \text{cert}_{pJ} \leq 1$; t is a time.

We interpret this as J has cert_{pJ} regarding p at time t . If $\text{cert}_{pJ} = -1$, J believes p to be definitely false, if $\text{cert}_{pJ} = 1$, J believes p to be definitely true, and if $\text{cert}_{pJ} = 0$, J has no opinion as to the truth of p .

Let M denote the set of all agents in the system and T the set of all times.

The set P denotes the set of all p such that $\langle p, \text{cert}_{pJ}, t \rangle \in W_J$ for some agent $J \in M$ and some time $t \in T$.

Definition 3: Beliefs about Actions. The beliefs about action of Agent J is a set of triples $\langle a, \text{Pre}_{aJ}, \text{Post}_{aJ} \rangle$ where,

action a is an action; Pre_{aJ} is a set of pairs $\langle p, \text{threshold}_{pJ} \rangle$ and Post_{aJ} is a set of pairs $\langle p, \text{truth}_{pJ} \rangle$, $-1 \leq \text{threshold}_{pJ} \leq 1$, and $-1 \leq \text{truth}_{pJ} \leq 1$.

Pre_{aJ} is a set of preconditions for *action* a recognised by agent J. The interpretation is that J believes that *action* a can be performed at t if all elements of Pre_{aJ} are satisfied with respect to W_J at t .

$\langle p, \text{threshold}_{pJ} \rangle$ is satisfied with respect to W_J if $\langle p, \text{cert}_{pJ}, t \rangle$ and if $\text{threshold}_{pJ} > 0$, then $\text{cert}_{pJ} \geq \text{threshold}_{pJ}$, else if $\text{threshold}_{pJ} < 0$, $\text{cert}_{pJ} \leq \text{threshold}_{pJ}$. J believes that if *action* a is performed at t , then for all $\langle p, \text{truth}_{pJ} \rangle \in \text{Post}_{aJ}$, $\langle p, \text{truth}_{pJ}, t+1 \rangle$ will be an element of W_J .

W_{Ja} is the state of the world that J believes will result from performing *action* a .

Additionally, J may *assume* that *action* a can be performed at t if all elements of Pre_{aJ} can be *assumed to be satisfied* with respect to W_J at t . $\langle p, \text{threshold}_{pJ} \rangle$ can be assumed satisfied with respect to W_J if $\langle p, \text{cert}_{pJ}, t \rangle$ and if $\text{threshold}_{pJ} > 0$, then $\text{cert}_{pJ} \geq 0$ and if $\text{threshold}_{pJ} < 0$, $\text{cert}_{pJ} \leq 0$.

The set A denotes the set of all actions such that $\langle a, \text{Pre}_{aJ}, \text{Post}_{aJ} \rangle \in A_J$ for some agent $J \in M$.

Definition 4: Desires of an Agent. The desires of an Agent J is a set of pairs $\langle d, \text{Cond}_{dJ} \rangle$ such that,

d is a desire and Cond_{dJ} is a set of pairs $\langle p, \text{threshold}_{pJ} \rangle$. The interpretation is that J believes that d is satisfied at t if Cond_{dJ} is satisfied with respect to W_J at

t. The notions of satisfaction and assumed satisfaction for Cond_{dJ} is the same as that for Pre_{aJ} .

The set D denotes the set of all desires such that $\langle d, \text{Cond}_{dJ} \rangle \in D_J$ for some agent $J \in M$.

Definition 5: *Values of an Agent.* The values of an Agent J is a set of triples $\langle v, d, \text{prom}_{vJ} \rangle$ such that,

v is a value,

d is a desire,

prom_{vJ} a number $-1 \leq \text{prom}_{vJ} \leq 1$, representing the degree to which the satisfaction of d promotes v .

The set V denotes the set of all values such that $\langle v, d, \text{prom}_{vJ} \rangle \in V_J$ for some agent $J \in M$.

Definition 6: Let $\text{satA}(\text{Formula}, W_J)$ be true if Formula can be assumed to be satisfied with respect to W_J .

Let $\text{satS}(\text{Formula}, W_J)$ be true if Formula can be satisfied with respect to W_J .

Now J has a presumptive argument for *action a* at time t if:

there is an $\langle a, \text{Pre}_{aJ}, \text{Post}_{aJ} \rangle \in A_J$ such that:

$\text{satA}(\text{Pre}_{aJ}, J)$ at t ;

$\text{satA}(\text{Cond}_{dJ}, J)$ at $t+1$ and

Cond_{dJ} will be satisfied at $t+1$ with respect to W_J ;

there is a $\langle v, d, \text{prom}_{vJ} \rangle$, such that $\text{prom}_{vJ} > 0$.

The position is expressed as:

In circumstances r , where each $r \in R$ is the first term in each element of Pre_{aJ} ,

Performing *action a*,

Will result in s , where each $s \in S$ is the first term in each element of Post_{aJ} ,

Which will realise d ,

Which promotes v .

3 Attacks

We now present a formalism showing how the initial position for the justification of an action can be attacked, in accordance with our theory.

There is an attacking agent $K \in M$ such that $\langle W_K, A_K, D_K, V_K \rangle$ and agent K may attack the position put forward by agent J using the set of attacks subject to the following conditions:

Source CQ: Are the believed circumstances true? (CQ1).

Attack 1a: *Pre-conditions for A_K to make an attack:*

satA(Pre_{aK}, W_K) and,
not satS(Pre_{aK}, W_K).

Argument: p may not be true.

Attack 1b: *Pre-conditions for A_K to make an attack:*

not satA(Pre_{aK}, W_K).

Argument: p is not true.

Source CQ: Assuming the circumstances are true, does the action have the stated consequences? (CQ2).

Attack 2a: *Pre-conditions for A_K to make an attack:*

satA(Post_{aK}, W_K) and,
not satS(Post_{aK}, W_K).

Argument: action a may not have the desired consequences.

Attack 2b: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K).

Argument: action a will not have the desired consequences.

Attack 2c: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K) and,
for no $\langle d, \text{Cond}_{dK} \rangle$ does satA(Cond_{dK}, W_{Ka}) hold.

Argument: action a will not have the desired consequences.

Attack 2d: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K) and,
satA(Cond_{dK}, W_{Ka}) and,
<v, d, prom_{vK}> and,
prom_{vK} ≤ 0.

Argument: action a will not have the desired consequences.

Attack 2e: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K) and,
satA(Cond_{dK}, W_{Ka}) and,
<v, d, prom_{vK}> and,
prom_{vK} < 0.

Argument: action a will not have the desired consequences.

Attack 2f: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K) and,
satA(Cond_{dK}, W_{Ka}) and,
there is a w, w ≠ v such that <w, d, prom_{wK}> and,
prom_{wK} > 0.

Argument: action a will not have the desired consequences.

Attack 2g: *Pre-conditions for A_K to make an attack:*

not satA(Post_{aK}, W_K) and,
satA(Cond_{dK}, W_{Ka}) and,
there is a w, w ≠ v such that <w, d, prom_{wK}> and,
prom_{wK} < 0.

Argument: action a will not have the desired consequences.

Source CQ: Assuming the circumstances are true and the actions has the stated consequences, will the action bring about the desired goal? (CQ3).

Attack 3a: *Pre-conditions for A_K to make an attack:*

for no < d, Cond_{dK}> does satA(Cond_{dK}, W_{Ka}) hold.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Attack 3b: *Pre-conditions for A_K to make an attack:*

for no $\langle d, \text{Cond}_{dK} \rangle$ does $\text{satA}(\text{Cond}_{dK}, W_{Ka})$ hold and,
for some $e, e \neq d, \text{satA}(\text{Cond}_{eK}, W_{Ka})$.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Attack 3c: *Pre-conditions for A_K to make an attack:*

for no $\langle d, \text{Cond}_{dK} \rangle$ does $\text{satA}(\text{Cond}_{dK}, W_{Ka})$ hold and,
for some $e, e \neq d, \text{satA}(\text{Cond}_{eK}, W_{Ka})$ and,
 $\langle v, e, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} > 0$.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Attack 3d: *Pre-conditions for A_K to make an attack:*

for no $\langle d, \text{Cond}_{dK} \rangle$ does $\text{satA}(\text{Cond}_{dK}, W_{Ka})$ hold and,
for some $e, e \neq d, \text{satA}(\text{Cond}_{eK}, W_{Ka})$ and,
 $\langle v, e, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} < 0$.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Attack 3e: *Pre-conditions for A_K to make an attack:*

for no $\langle d, \text{Cond}_{dK} \rangle$ does $\text{satA}(\text{Cond}_{dK}, W_{Ka})$ hold and,
for some $e, e \neq d, \text{satA}(\text{Cond}_{eK}, W_{Ka})$ and,
there is a $w, w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Attack 3f: *Pre-conditions for A_K to make an attack:*

for no $\langle d, \text{Cond}_{dK} \rangle$ does $\text{satA}(\text{Cond}_{dK}, W_{Ka})$ hold and,
for some $e, e \neq d, \text{satA}(\text{Cond}_{eK}, W_{Ka})$ and,
there is a $w, w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,

$\text{prom}_{wK} < 0$.

Argument: the state of affairs resulting from performing the action will not bring about the goal.

Source CQ: Does the goal realise the value intended? (CQ4).

Attack 4a: *Pre-conditions for A_K to make an attack:*

$\langle v, d, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} \leq 0$.

Argument: the goal may not promote the value.

Attack 4b: *Pre-conditions for A_K to make an attack:*

$\langle v, d, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} < 0$.

Argument: the goal will not promote the value.

Attack 4c: *Pre-conditions for A_K to make an attack:*

$\langle v, d, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} \leq 0$ and,
there is a $w, w \neq v$ such that $\langle w, d, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$.

Argument: the goal will not promote the value.

Attack 4d: *Pre-conditions for A_K to make an attack:*

$\langle v, d, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} \leq 0$ and,
there is a $w, w \neq v$ such that $\langle w, d, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} < 0$.

Argument: the goal will not promote the value.

Source CQ: Are there alternative ways of realising the same consequences? (CQ5).

Attack 5: *Pre-conditions for A_K to make an attack:*

$\text{satA}(\text{Pre}_{bK}, W_K)$ and,

$\text{satA}(\text{Post}_{aK}, W_{Kb})$ and $b \neq a$.

Argument: there is an alternative action which will realise the same consequences.

Source CQ: Are there alternative ways of realising the same goal? (CQ6).

Attack 6: *Pre-conditions for A_K to make an attack:*

$\text{satA}(\text{Pre}_{bK}, W_K)$ and,
 $\text{satA}(\text{Cond}_{dK}, W_{Kb})$ and $b \neq a$.

Argument: there is an alternative action which will realise the same goal.

Source CQ: Are there alternative ways of promoting the same values? (CQ7).

Attack 7a: *Pre-conditions for A_K to make an attack:*

$\text{satA}(\text{Pre}_{bK}, W_K)$ and,
for some e , $e \neq d$, $\text{satA}(\text{Cond}_{eK}, W_{Kb})$ and $b \neq a$ and,
 $\langle v, e, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} > 0$.

Argument: there is an alternative action, leading to an alternative goal, which will promote the value.

Attack 7b: *Pre-conditions for A_K to make an attack:*

$\text{satA}(\text{Cond}_{eK}, W_{Ka})$, $e \neq d$ and,
 $\langle v, e, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} > 0$.

Argument: action a has a side effect which satisfies an alternative goal, which promotes the value.

Source CQ: Does doing A have a side effect which demotes the value V ? (CQ8).

Attack 8: *Pre-conditions for A_K to make an attack:*

$\text{satA}(\text{Cond}_{eK}, W_{Ka})$, $e \neq d$ and,
 $\langle v, e, \text{prom}_{vK} \rangle$ and,
 $\text{prom}_{vK} < 0$.

Argument: action a has a side effect which satisfies an alternative goal, which demotes the value.

Source CQ: Does doing A have a side effect which demotes some other value? (CQ9).

Attack 9: *Pre-conditions for A_K to make an attack:*

satA(Cond $_{eK}$, W_{Ka}), $e \neq d$ and,
there is a w , $w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} < 0$.

Argument: action a has a side effect which satisfies an alternative goal, which demotes some other value.

Source CQ: Would doing A promote some other value? (CQ10).

Attack 10: *Pre-conditions for A_K to make an attack:*

satA(Cond $_{eK}$, W_{Ka}), $e \neq d$ and,
there is a w , $w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$.

Argument: action a has a side effect which satisfies an alternative goal, which promotes some other value.

Source CQ: Does doing A preclude some other action which would promote some other value? (CQ11).

Attack 11a: *Pre-conditions for A_K to make an attack:*

satA(Pre $_{aK}$, W_K) and,
satA(Cond $_{eK}$, W_{Kb}), $e \neq d$ and,
there is a w , $w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$ and,
not satA(Pre $_{aK}$, W_{Kb}) and,
not satA(Pre $_{bK}$, W_{Ka}).

Argument: doing action a precludes some other action which would promote some other value.

Attack 11b: *Pre-conditions for A_K to make an attack:*

there is a w , $w \neq v$ such that $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$ and,

for no $\langle e, \text{Cond}_{eK} \rangle$ does $\text{satA}(\text{Cond}_{eK}, W_{Ka})$ hold.

Argument: there is some other goal, which promotes some other value, but the goal is not derivable from the state of affairs s .

Attack 11c: *Pre-conditions for A_K to make an attack:*

there is no e , $\langle e, \text{Cond}_{eK} \rangle$ such that
for $w, w \neq v$, $\langle w, e, \text{prom}_{wK} \rangle$ and,
 $\text{prom}_{wK} > 0$ and,
 $\text{satA}(\text{Cond}_{eK}, W_{Ka})$.

Argument: if there is some other goal, which promotes some other value, then this goal is not derivable from the state of affairs s .

Source CQ: Are the believed circumstances possible?

Attack 12: *Pre-conditions for A_K to make an attack:*

it is not the case that $\langle a, \text{Pre}_{aK}, \text{Post}_{aK} \rangle \in A_K$.

Argument: r is not a possible state of affairs.

Source CQ: Is it possible to do action A ?

Attack 13: *Pre-conditions for A_K to make an attack:*

for some $\langle p, \text{cert}_{pJ}, t \rangle \in \text{Pre}_{aJ}$, $\langle p, \text{cert}_{pK}, t \rangle \notin W_K$.

Argument: action a is not a possible action.

Source CQ: Is the situation S , believed by agent J to result from doing A , a possible state of affairs?

Attack 14: *Pre-conditions for A_K to make an attack:*

for some $\langle p, \text{cert}_{pJ}, t \rangle \in \text{Post}_{aJ}$, $\langle p, \text{cert}_{pK}, t \rangle \notin W_{Ka}$.

Argument: s is not a possible state of affairs.

Source CQ: Are the particular aspects of S represented by G possible?

Attack 15: *Pre-conditions for A_K to make an attack:*

$\langle d, \text{Cond}_{dK} \rangle \notin D_K$.

Argument: g is not a possible goal.

Source CQ: Is the value proposed indeed a legitimate value?

Attack 16: *Pre-conditions for A_K to make an attack:*

$\langle v, d, \text{prom}_{vK} \rangle \notin V_K$.

Argument: v is not a possible value.

References

- [1] K. Atkinson, T. Bench-Capon, and P. McBurney. Justifying practical reasoning. In F. Grasso, C. Reed, and G. Carenini, editors, *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 87–90, Valencia, Spain, 2004.
- [2] K. Greenwood, T. Bench-Capon, and P. McBurney. Towards a computational account of persuasion in law. In *Proceedings of the Ninth International Conference on AI and Law (ICAIL-2003)*, pages 22–31, New York, NY, USA, 2003. ACM Press.