

Argument Adjudication in a Deontic Logic

Naveen Sundar Govindarajulu
Selmer Bringsjord
Mike Giancola



International Conference on Robot Ethics and Standards (ICRES)
July 29, 2019

Sponsored by



Rensselaer

Related Work

- Judgement Aggregation
 - Arrow's Impossibility Theorem
- Argument Aggregation
 - Pollock (2009)
 - Dung (1995)

Problem

We have a set of n agents producing possibly varying statements that can be ethically charged $\{\phi_1, \dots, \phi_n\}$. The statements are in formal system containing a deontic operator $\mathbf{O}(\gamma, \rho)$. Each of the agents also produce arguments in support of their statements $\{\alpha_1 \rightsquigarrow \phi_1, \dots, \alpha_n \rightsquigarrow \phi_n\}$. We need to compute a statement ϕ^* and an argument α^* that best represents the diverse statements and arguments given a background ethical theory Γ .

Formal Background

- Deontic Cognitive Event Calculus $DC\mathcal{EC}$
 - First Order Multi-Operator Modal Logic
 - Well-Defined Syntax & Inference Schemata
 - Based on Natural Deduction

Sort	Description
Agent	Human and non-human actors.
Time	The Time type stands for time in the domain. E.g. simple, such as t_i , or complex, such as $birthday(son(jack))$.
Event	Used for events in the domain.
ActionType	Action types are abstract actions. They are instantiated at particular times by actors. Example: eating.
Action	A subtype of Event for events that occur as actions by agents.
Fluent	Used for representing states of the world in the event calculus.

Syntax

$$S ::= \text{Agent} \mid \text{ActionType} \mid \text{Action} \sqsubseteq \text{Event} \mid \text{Moment} \mid \text{Fluent}$$

$$f ::= \left\{ \begin{array}{l} \text{action} : \text{Agent} \times \text{ActionType} \rightarrow \text{Action} \\ \text{initially} : \text{Fluent} \rightarrow \text{Formula} \\ \text{holds} : \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ \text{happens} : \text{Event} \times \text{Moment} \rightarrow \text{Formula} \\ \text{clipped} : \text{Moment} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ \text{initiates} : \text{Event} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ \text{terminates} : \text{Event} \times \text{Fluent} \times \text{Moment} \rightarrow \text{Formula} \\ \text{prior} : \text{Moment} \times \text{Moment} \rightarrow \text{Formula} \end{array} \right.$$

$$t ::= x : S \mid c : S \mid f(t_1, \dots, t_n)$$

$$\phi ::= \left\{ \begin{array}{l} q : \text{Formula} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \forall x : \phi(x) \mid \\ \mathbf{P}(a, t, \phi) \mid \mathbf{K}(a, t, \phi) \mid \\ \mathbf{C}(t, \phi) \mid \mathbf{S}(a, b, t, \phi) \mid \mathbf{S}(a, t, \phi) \mid \mathbf{B}(a, t, \phi) \\ \mathbf{D}(a, t, \phi) \mid \mathbf{I}(a, t, \phi) \\ \mathbf{O}(a, t, \phi, (\neg)\text{happens}(\text{action}(a^*, \alpha), t')) \end{array} \right.$$

Inference Schemata (Fragment)

$$\frac{\mathbf{K}(a, t_1, \Gamma), \Gamma \vdash \phi, t_1 \leq t_2}{\mathbf{K}(a, t_2, \phi)} [I_{\mathbf{K}}]$$

$$\frac{\mathbf{B}(a, t_1, \Gamma), \Gamma \vdash \phi, t_1 \leq t_2}{\mathbf{B}(a, t_2, \phi)} [I_{\mathbf{B}}]$$

$$\frac{\mathbf{K}(a, t, \phi)}{\phi} [I_4] \qquad \frac{t < t', \mathbf{I}(a, t, \psi)}{\mathbf{P}(a, t', \psi)} [I_{13}]$$

$$\frac{\mathbf{B}(a, t, \phi) \quad \mathbf{B}(a, t, \mathbf{O}(a, t, \phi, \chi)) \quad \mathbf{O}(a, t, \phi, \chi)}{\mathbf{K}(a, t, \mathbf{I}(a, t, \chi))} [I_{14}]$$

Parallel R&D: Uncertainty System for Modal Logic

- Need a method for handling uncertainty in a quantified modal logic
- Standard probability theory is insufficient
 - Humans' don't reason with numerical probability values
 - Numerical probability values are rarely available
 - E.g. "How certain are you that this is the man you saw rob the bank?"

Strength Factor Definitions

Acceptable An agent a at time t finds ϕ acceptable *iff* withholding ϕ is not more reasonable than believing in ϕ .

$$\mathbf{B}^1(a, t, \phi) \Leftrightarrow \begin{cases} \mathbf{W}(a, t, \phi) \not\succ_t^a \mathbf{B}(a, t, \phi); \text{ or} \\ \left(\neg \mathbf{B}(a, t, \phi) \wedge \neg \mathbf{B}(a, t, \neg \phi) \right) \not\succ_t^a \mathbf{B}(a, t, \phi) \end{cases}$$

Some Presumption in Favor An agent a at time t has some presumption in favor of ϕ *iff* believing ϕ at t is more reasonable than believing $\neg \phi$ at time t :

$$\mathbf{B}^2(a, t, \phi) \Leftrightarrow \left(\mathbf{B}(a, t, \phi) \succ_t^a \mathbf{B}(a, t, \neg \phi) \right)$$

Beyond Reasonable Doubt An agent a at time t has beyond reasonable doubt in ϕ *iff* believing ϕ at t is more reasonable than withholding ϕ at time t :

$$\mathbf{B}^3(a, t, \phi) \Leftrightarrow \begin{cases} \mathbf{B}(a, t, \phi) \succ_t^a \mathbf{W}(a, t, \phi); \text{ or} \\ \left(\mathbf{B}(a, t, \phi) \succ_t^a \left(\neg \mathbf{B}(a, t, \phi) \wedge \neg \mathbf{B}(a, t, \neg \phi) \right) \right) \end{cases}$$

Evident A formula ϕ is evident to an agent a at time t *iff* ϕ is beyond reasonable doubt and if there is a ψ such that believing ψ is more reasonable for a at time t than believing ϕ , then a is certain about ψ at time t .

$$\mathbf{B}^4(a, t, \phi) \Leftrightarrow \begin{cases} \mathbf{B}^3(a, t, \phi) \wedge \\ \exists \psi : \left[\begin{array}{l} \mathbf{B}(a, t, \psi) \succ_t^a \mathbf{B}(a, t, \phi) \\ \Rightarrow \mathbf{B}^5(a, t, \psi) \end{array} \right] \end{cases}$$

Certain An agent a at time t is certain about ϕ *iff* ϕ is beyond reasonable doubt and there is no ψ such that believing ψ is more reasonable for a at time t than believing ϕ .

$$\mathbf{B}^5(a, t, \phi) \Leftrightarrow \begin{cases} \mathbf{B}^3(a, t, \phi) \wedge \\ \neg \exists \psi : \mathbf{B}(a, t, \psi) \succ_t^a \mathbf{B}(a, t, \phi) \end{cases}$$

Strength Factor Continuum

Certain

Evident

Overwhelmingly Likely

Beyond Reasonable Doubt

Likely

More Likely Than Not

Counterbalanced

More Unlikely Than Not

Unlikely

Overwhelmingly Unlikely

Beyond Reasonable Belief

Evidently False

Certainly False

Strength Factor Continuum

Epistemically Positive

(6)

Certain

(5)

Evident

(4)

Overwhelmingly Likely

(3)

Beyond Reasonable Doubt

(2)

Likely

(1)

More Likely Than Not

(0)

Counterbalanced

(-1)

More Unlikely Than Not

(-2)

Unlikely

(-3)

Overwhelmingly Unlikely

(-4)

Beyond Reasonable Belief

(-5)

Evidently False

(-6)

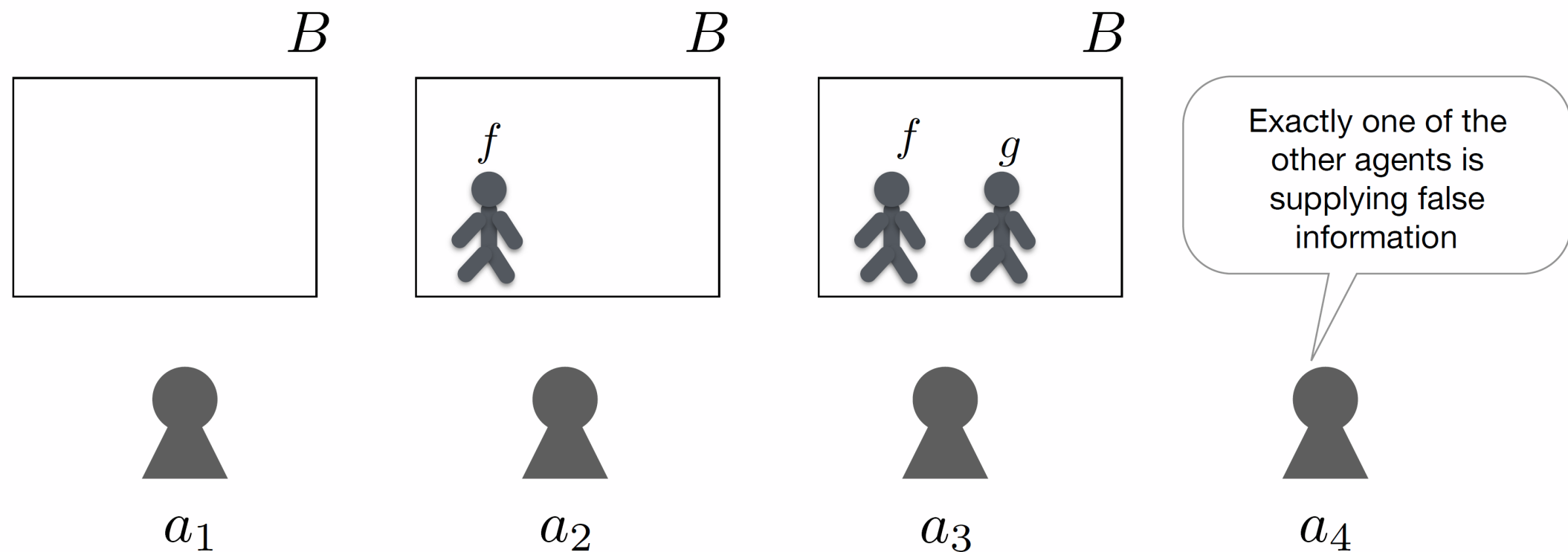
Certainly False

.....

Epistemically Negative

Example

- f is a violent fugitive
- Goal: Determine if f is the only person inside building B .



Agent	Statement	CC Representation	Strength
a_1	f is not in B	$\mathbf{S}(a_1, \neg in(f, B))$	\mathbf{B}^4
a_2	f is in B	$\mathbf{S}(a_2, in(f, B))$	\mathbf{B}^4
a_3	f is in B and g is in B .	$\mathbf{S}(a_3, in(f, B) \wedge in(g, B))$	\mathbf{B}^4
a_4	Exactly one of $\{a_1, a_2, a_3\}$ is false.	$\exists! a \in \{a_1, a_2, a_3\} (\mathbf{S}(a, \phi) \Rightarrow \neg \phi)$	\mathbf{B}^5

Warning!

- Argument Aggregation
 - Pollock (2009)
 - Dung (1995)
- No use of internal structure of arguments

Inductive-Reasoning Example from Pollock — for Peek Ahead

Imagine the following:

Keith tells you that the morning news predicts rain in Troy today. However, Alvin tells you that the same news report predicted sunshine.

Imagine the following:

Keith tells you that the morning news predicts rain in Troy today. However, Alvin tells you that the same news report predicted sunshine.

Without any other source of information, it would be irrational to place belief in either Keith's or Alvin's statements.

Imagine the following:

Keith tells you that the morning news predicts rain in Troy today. However, Alvin tells you that the same news report predicted sunshine.

Without any other source of information, it would be irrational to place belief in either Keith's or Alvin's statements.

Further, suppose you happened to watch the noon news report, and that report predicted rain. Then you should believe that it will rain despite your knowledge of Alvin's argument.

Cast in *Inference Graphs* for Pollock's OSCAR

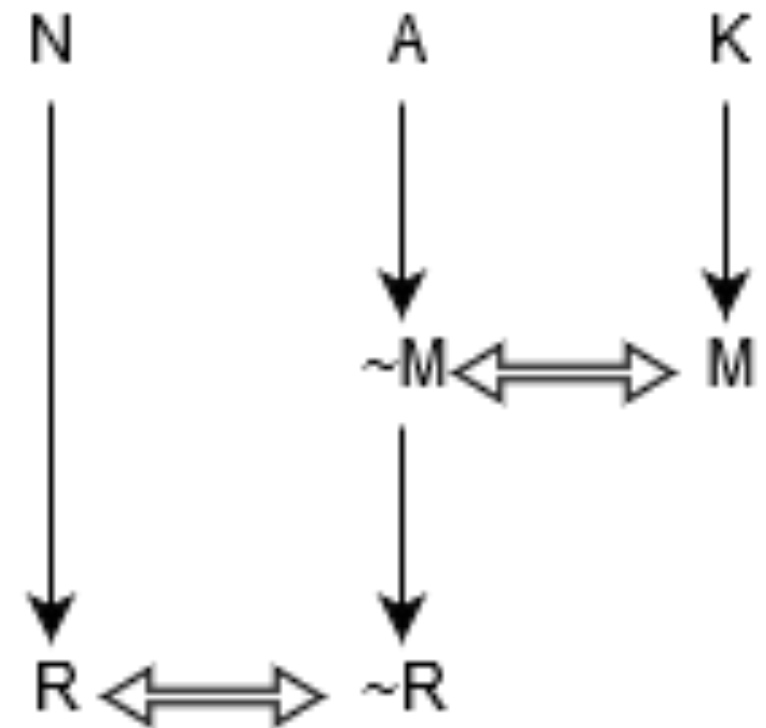
K- Keith says that M

A- Alvin says that $\sim M$

M- The morning news said that R

R- It is going to rain this afternoon

N- The noon news says that R



All such can be absorbed into and refined in our inductive logics and our automated inductive reasoners for argument adjudication.

Relevant Inference Schemata

(simplified)

$$\frac{\mathbf{K}\phi}{\phi} \quad I_1$$

$$\frac{\mathbf{S}(\mathbf{a}_1, \phi, \mathbf{a}_2), \quad \chi(\mathbf{a}_1)}{\mathbf{B}^2(\mathbf{a}_2, \phi)} \quad I_2$$

$$\frac{\mathbf{S}(\mathbf{a}_1, \phi, \mathbf{a}_2), \quad \chi'(\mathbf{a}_1)}{\mathbf{B}^3(\mathbf{a}_2, \phi)} \quad I_{2'}$$

$$\frac{\mathbf{B}^2(\mathbf{a}, \phi, t), \quad \mathbf{B}^2(\mathbf{a}, \neg\phi, t)}{\neg\mathbf{B}^2(\mathbf{a}, \phi, t) \wedge \neg\mathbf{B}^2(\mathbf{a}, \neg\phi, t)} \quad \text{Clash Principle}$$

$\chi, \chi' : \text{preconditions}$

$$\begin{array}{ccc}
\mathbf{K}(you, \mathbf{S}(keith, \mathbf{S}(m, rain))) & & \mathbf{K}(you, \mathbf{S}(alvin, \mathbf{S}(m, \neg rain))) \\
\downarrow \text{via } I_1 & & \downarrow \text{via } I_1 \\
\mathbf{S}(keith, \mathbf{S}(m, rain)) & & \mathbf{S}(alvin, \mathbf{S}(m, \neg rain)) \\
\downarrow \text{via } I_2 & & \downarrow \text{via } I_2 \\
\mathbf{B}^2(you, \mathbf{S}(m, rain)) & \wedge & \mathbf{B}^2(you, \mathbf{S}(m, \neg rain)) \\
& \downarrow & \\
& \mathbf{Clash !} & \\
& & \\
\neg \mathbf{B}^2(you, \mathbf{S}(m, rain)) & \wedge & \neg \mathbf{B}^2(you, \mathbf{S}(m, \neg rain)) \\
& & \\
& \mathbf{K}(you, \mathbf{S}(noonnews, rain)) & \\
& \downarrow \text{via } I_1 & \\
& \mathbf{S}(noonnews, rain) & \\
& \downarrow \text{via } I_{2'} & \\
& \mathbf{B}^3(you, rain) &
\end{array}$$

Future Work

- Further the development of formal methods/algorithms/ implementations of argument adjudication over nested modal statements
- Apply techniques to Visual Question Answering with Justifications
 - Adjudicator queries multiple agents about a scene (image/video), will need to consider each agents beliefs, knowledge, etc.
 - Provide justifications e.g. “Agent 3’s belief has relatively lower strength as her view of the scene is occluded.”

Thank You