Logic, Accountability and Design

David Pearce

Universidad Politécnica de Madrid

June 5, 2024

Motivation

- Artificial Intelligence (AI)
  $\Rightarrow$ solutions, predictions, decisions, actions

  Large, complex systems: big data/knowledge

Motivation

- Artificial Intelligence (AI)
  $\Rightarrow$ solutions, predictions, decisions, actions

  Large, complex systems: big data/knowledge

- Explainable AI:
  most efforts focused on the system's (technical) behavior

Motivation

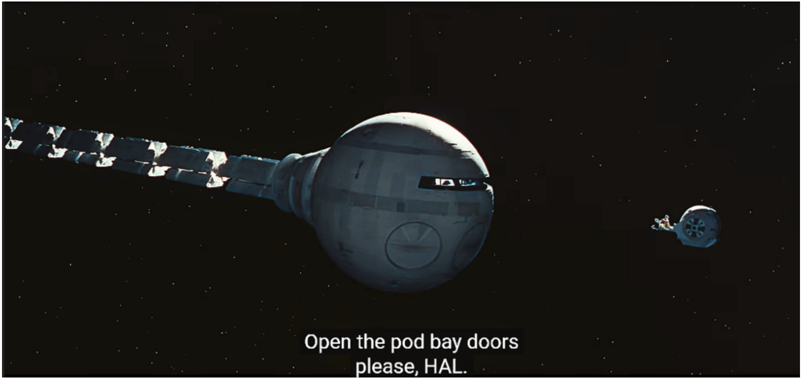
- Artificial Intelligence (AI)
  $\Rightarrow$ solutions, predictions, decisions, actions

  Large, complex systems: big data/knowledge

- Explainable AI:
  most efforts focused on the system's (technical) behavior

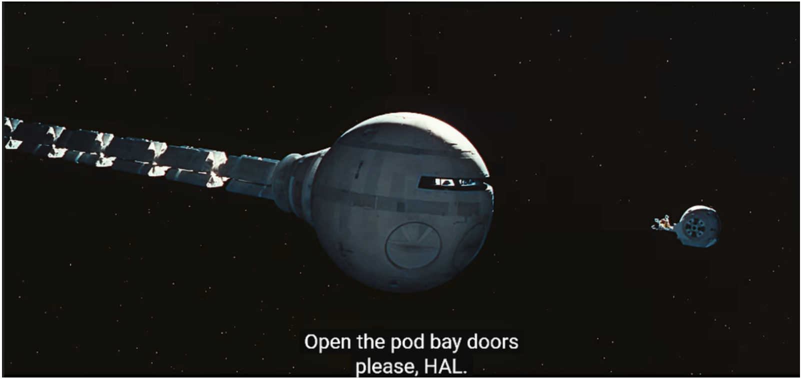
  But technical explanations $\neq$ justifications

Technical explanation



- HAL: *I am afraid I can't do that*

Technical explanation

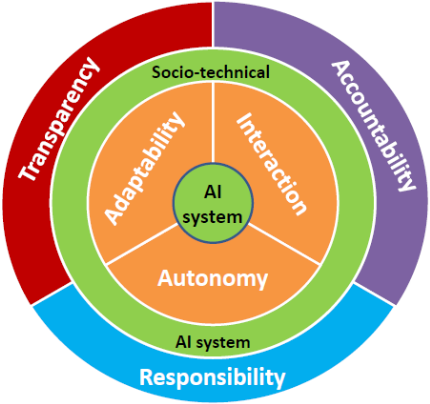


- HAL: *I am afraid I can't do that*
- Dave: *Why?*
- HAL: *Because I read your lips, you plan to disconnect me, and this decreases the probability of success of the mission to 0.05 %*

Responsible AI

ART

Accountability - Responsibility - Transparency



The ART of AI

ART

Accountability - Responsibility - Transparency

- Accountability: explain and justify results in a comprehensible way for the end user.

The ART of AI

ART

Accountability - Responsibility - Transparency

- Accountability: explain and justify results in a comprehensible way for the end user.
  Justification w.r.t. moral values and societal norms

The ART of AI

ART
Accountability - Responsibility - Transparency

- Accountability: explain and justify results in a comprehensible way for the end user.
  Justification w.r.t. moral values and societal norms
- Responsibility: how AI sytems incorporate the role of people
  Link AI system's decisions to a fair use of data and to the actions of stakeholders

The ART of AI

ART

Accountability - Responsibility - Transparency

- Accountability: explain and justify results in a comprehensible way for the end user.
  Justification w.r.t. moral values and societal norms
- Responsibility: how AI sytems incorporate the role of people
  Link AI system's decisions to a fair use of data and to the actions of stakeholders
- Transparency: describe, inspect and reproduce the mechanisms through which AI systems make decisions

Responsible AI

- Logic-based AI seems better suited for these purposes

Responsible AI

- Logic-based AI seems better suited for these purposes
- But systems can be large and complex, with many reasoning steps

Responsible AI

- Logic-based AI seems better suited for these purposes
- But systems can be large and complex, with many reasoning steps
- Workshops on Explainable Logic-Based Knowledge Representation (XLoKR) cover ASP, description logics, argumentation theory, NMR, etc
- Answering the "Why" in Answer Set Programming [Fandinno & Schultz 19]

Responsible AI

- Logic-based AI seems better suited for these purposes
- But systems can be large and complex, with many reasoning steps
- Workshops on Explainable Logic-Based Knowledge Representation (XLoKR) cover ASP, description logics, argumentation theory, NMR, etc
- Answering the “Why” in Answer Set Programming [Fandinno & Schultz 19]
- However, most work takes the adequacy of the primary reasoning formalism for granted. Choosing a formalism or its semantics is “up to the user” (*à la carte*)

Responsible AI

- Logic-based AI seems better suited for these purposes
- But systems can be large and complex, with many reasoning steps
- Workshops on Explainable Logic-Based Knowledge Representation (XLoKR) cover ASP, description logics, argumentation theory, NMR, etc
- Answering the "Why" in Answer Set Programming [Fandinno & Schultz 19]
- However, most work takes the adequacy of the primary reasoning formalism for granted. Choosing a formalism or its semantics is "up to the user" (*à la carte*)
- This is a serious drawback for accountability!

  ☞ We need suitable adequacy criteria.

Responsible AI

- Logic-based AI seems better suited for these purposes
- But systems can be large and complex, with many reasoning steps
- Workshops on Explainable Logic-Based Knowledge Representation (XLoKR) cover ASP, description logics, argumentation theory, NMR, etc
- Answering the "Why" in Answer Set Programming [Fandinno & Schultz 19]
- However, most work takes the adequacy of the primary reasoning formalism for granted. Choosing a formalism or its semantics is "up to the user" (*à la carte*)
- This is a serious drawback for accountability!

  ☞ We need suitable adequacy criteria.

Fundación BBVA *Project LIANDA*

If Logic can provide explanations and justifications: Which Logic(s)?

- Classical logic and extensions: infinitary logics, generalised quantifiers, epistemic, modal and temporal logics

If Logic can provide explanations and justifications: Which Logic(s)?

- Classical logic and extensions: infinitary logics, generalised quantifiers, epistemic, modal and temporal logics
- Deviant logics: constructive logics, multi-valued logics, paraconsistent logics

If Logic can provide explanations and justifications: Which Logic(s)?

- Classical logic and extensions: infinitary logics, generalised quantifiers, epistemic, modal and temporal logics
- Deviant logics: constructive logics, multi-valued logics, paraconsistent logics
- Nonmonotonic logics: default logic, autoepistemic logic, defeasible logics, stable reasoning

What are the *grounds* for choice?

- Internal principles of truth and inference: excluded middle, disjunctive syllogism, explosive axioms

What are the *grounds* for choice?

- Internal principles of truth and inference: excluded middle, disjunctive syllogism, explosive axioms
- General properties of inference and semantics: constructivity, computability, compactness, interpolation, cumulative inference, rationality

What are the *grounds* for choice?

- Internal principles of truth and inference: excluded middle, disjunctive syllogism, explosive axioms
- General properties of inference and semantics: constructivity, computability, compactness, interpolation, cumulative inference, rationality
- Expressive needs for applications: modal operators, special quantifiers, infinitary languages

Back to general principles of inference

Sometimes we may find a Lindström-style theorem, ie a property or properties that narrow down the class of logics to one or a small number

Back to general principles of inference

Sometimes we may find a Lindström-style theorem, ie a property or properties that narrow down the class of logics to one or a small number

Lindström (1969): classical first-order logic is the strongest logic satisfying both:

- (countable) compactness: if a countable set of sentences has no model then some finite subset has no model

- (downward) Löwenheim-Skolem: if a sentence has an infinite model, it has a countable model

Back to general principles of inference

What happens when we extend first-order logic?

Back to general principles of inference

What happens when we extend first-order logic?

$L(Q_1)$ (“there exist at least $\aleph_1$ many") is countably compact

Back to general principles of inference

What happens when we extend first-order logic?

$L(Q_1)$ ("there exist at least $\aleph_1$ many") is countably compact

$L_{\omega_1,\omega}$ satisfies the Löwenheim property

Early days of Logics in AI: Preference for...

“Desirable" properties of inference: cumulative, rational

$$\Pi \mid\!\sim \varphi, \Pi \mid\!\sim \psi \Rightarrow \Pi \cup \varphi \mid\!\sim \psi$$

$$\Pi \mid\!\sim \psi, \Pi \cup \varphi \not\mid\!\sim \psi \Rightarrow \Pi \mid\!\sim \neg\varphi$$

Computability: polynomial is better (but what about Datalog?)

Supraclassicality: add to classical logic rather than revise it (but remember Ptolomaic epicycles!)

Stable reasoning does not fare well

Not cumulative, not rational

Not polynomial

Not supraclassical (fails left and right absorption)

Oh Dear!!

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century
  Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century
  Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum
  2. Exactness: exact rules, connection to a scientific system

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century
  Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum
  2. Exactness: exact rules, connection to a scientific system
  3. Fruitfulness: allows formulating many universal statements

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century
  Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum
  2. Exactness: exact rules, connection to a scientific system
  3. Fruitfulness: allows formulating many universal statements
  4. Simplicity: as simple as 1,2,3 allow

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century
  Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum
  2. Exactness: exact rules, connection to a scientific system
  3. Fruitfulness: allows formulating many universal statements
  4. Simplicity: as simple as 1,2,3 allow
- Most Logic-based AI has shied away from a methodology like Carnap's.

Carnap's explication

- Formal analysis of concepts: logical empiricism, 20th century Carnap's method of explication:
  - explicandum: inexact, informal concept
  - explicatum: formal, definition + rules for its use
- Adequacy conditions for explicatum:
  1. (A degree of) similarity to explicandum
  2. Exactness: exact rules, connection to a scientific system
  3. Fruitfulness: allows formulating many universal statements
  4. Simplicity: as simple as 1,2,3 allow
- Most Logic-based AI has shied away from a methodology like Carnap's. Exceptions: [Herzig 2014, 2017, et al 2018] and notably

. . .

Michael Gelfond's programme

[Gelfond 2011] names two main objectives. To understand:

1. *basic commonsense notions we use to think about the world: beliefs, knowledge, defaults, causality, intentions, probability, etc., and to learn how one ought to reason about them*

Michael Gelfond's programme

[Gelfond 2011] names two main objectives. To understand:

1. *basic commonsense notions we use to think about the world: beliefs, knowledge, defaults, causality, intentions, probability, etc., and to learn how one ought to reason about them*
2. *how to build software components of agents – entities which observe and act upon an environment and direct its activity towards achieving goals*

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.

2. Elegance: the corresponding mathematics should be simple and elegant, usable for KR and programming

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.
2. Elegance: the corresponding mathematics should be simple and elegant, usable for KR and programming $\sim$ Carnap's exactness

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.
2. Elegance: the corresponding mathematics should be simple and elegant, usable for KR and programming $\sim$ Carnap's exactness
3. Expressiveness: L should suggest systematic and elaboration tolerant representations of a broad class of phenomena of natural language, including belief, knowledge, defaults, causality and others

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.
2. Elegance: the corresponding mathematics should be simple and elegant, usable for KR and programming $\sim$ Carnap's exactness
3. Expressiveness: L should suggest systematic and elaboration tolerant representations of a broad class of phenomena of natural language, including belief, knowledge, defaults, causality and others
4. Relevance: a large number of interesting computational problems should be reducible to reasoning about theories formulated in L

Gelfond's 4 adequacy criteria

Formal language L

1. Clarity: logical vocabulary with clear and intuitive meaning.
2. Elegance: the corresponding mathematics should be simple and elegant, usable for KR and programming $\sim$ Carnap's exactness
3. Expressiveness: L should suggest systematic and elaboration tolerant representations of a broad class of phenomena of natural language, including belief, knowledge, defaults, causality and others
4. Relevance: a large number of interesting computational problems should be reducible to reasoning about theories formulated in L

3,4 $\sim$ Carnap's fruitfulness

Gelfond's programme

Other criteria from [Gelfond & Zhang 14]

- Naturalness: constructs of L should be close to (the parts of) natural language that L is designed to formalise

Gelfond's programme

Other criteria from [Gelfond & Zhang 14]

- Naturalness: constructs of L should be close to (the parts of) natural language that L is designed to formalise
  $\sim$ Carnap's similarity

Gelfond's programme

Other criteria from [Gelfond & Zhang 14]

- Naturalness: constructs of L should be close to (the parts of) natural language that L is designed to formalise
  $\sim$ Carnap's similarity
- Stability: Informally equivalent transformations of a text should correspond to formally equivalent ones in L
- (language) Elaboration Tolerance: possibility to expand L by new relevant constructs without substantial changes in its syntax and semantics

Gelfond's programme

Other criteria from [Gelfond & Zhang 14]

- Naturalness: constructs of L should be close to (the parts of) natural language that L is designed to formalise
  $\sim$ Carnap's similarity

- Stability: Informally equivalent transformations of a text should correspond to formally equivalent ones in L

- (language) Elaboration Tolerance: possibility to expand L by new relevant constructs without substantial changes in its syntax and semantics

Note that Gelfond does not include computational efficiency as a primary criterion

Socio-technical systems

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Method for reconstruction of social concepts (trust, role, normative power, ...) in computational systems

- Step 1: natural language description of social phenomena
- Step 2: formal language or calculus

Socio-technical systems

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Method for reconstruction of social concepts (trust, role, normative power, . . .) in computational systems

- Step 1: natural language description of social phenomena
- Step 2: formal language or calculus
  - Phase 1: conceptual framework expressed in formal (logical) terms

Socio-technical systems

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Method for reconstruction of social concepts (trust, role, normative power, . . .) in computational systems

- Step 1: natural language description of social phenomena
- Step 2: formal language or calculus
  - Phase 1: conceptual framework expressed in formal (logical) terms
    It only considers expressive capacity but not . . .

Socio-technical systems

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Method for reconstruction of social concepts (trust, role, normative power, . . .) in computational systems

- Step 1: natural language description of social phenomena
- Step 2: formal language or calculus
  - Phase 1: conceptual framework expressed in formal (logical) terms
    It only considers expressive capacity but not . . .
  - Phase 2: considers computational tractability, fragments, simplifications, approximations, etc

Socio-technical systems

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Method for reconstruction of social concepts (trust, role, normative power, . . .) in computational systems

- Step 1: natural language description of social phenomena
- Step 2: formal language or calculus
  - Phase 1: conceptual framework expressed in formal (logical) terms
    It only considers expressive capacity but not . . .
  - Phase 2: considers computational tractability, fragments, simplifications, approximations, etc
- Step 3: computer model of artificial system

Adequacy criteria

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Adequacy criteria for Step 2-Phase 1 languages. Capacity to:

- identify the principal elements (concept "building-blocks")

Adequacy criteria

The design of intelligent socio-technical systems
[Jones, Artikis & Pitt 2013]

Adequacy criteria for Step 2-Phase 1 languages. Capacity to:

- identify the principal elements (concept “building-blocks”)
- test for consistency (allow inference)
  $\sim$ Carnap’s exactness
- articulate specific, characteristic aspects of the concept
- ’place’ the concept in relation to its near relative
  $\sim$ Carnap’s similarity

Example: Non-Monotonic Reasoning

Examples of criteria we may use in NMR

- Strong equivalence (SE) coverage:
  find a logic L acting as monotonic basis
  Is L-equivalence necessary and sufficient for strong equivalence?

Example: Non-Monotonic Reasoning

Examples of criteria we may use in NMR

- Strong equivalence (SE) coverage:
  find a logic L acting as monotonic basis
  Is L-equivalence necessary and sufficient for strong equivalence?

- If so, L provides a powerful tool!
  We inherit its mathematical machinery to act in the monotonic level keeping SE

Example: Non-Monotonic Reasoning

Examples of criteria we may use in NMR

- Strong equivalence (SE) coverage:
  find a logic L acting as monotonic basis
  Is L-equivalence necessary and sufficient for strong equivalence?
- If so, L provides a powerful tool!
  We inherit its mathematical machinery to act in the monotonic level keeping SE

In our case, we studied other properties in different contexts

- ASP: well-supportedness, atom definability
- Epistemic ASP: splitting, foundedness, constraint monotonicity, supra-S5

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?
  - May seem circular. But some reasoning mechanisms in KR lack some natural properties, or perhaps defined for very restricted language fragments.

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?
  - May seem circular. But some reasoning mechanisms in KR lack some natural properties, or perhaps defined for very restricted language fragments.
- Is the reasoning based on a known underlying logic?

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?
  - May seem circular. But some reasoning mechanisms in KR lack some natural properties, or perhaps defined for very restricted language fragments.
- Is the reasoning based on a known underlying logic?
  - If so, we may profit from known properties and successes for a certain domain

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?
  - May seem circular. But some reasoning mechanisms in KR lack some natural properties, or perhaps defined for very restricted language fragments.
- Is the reasoning based on a known underlying logic?
  - If so, we may profit from known properties and successes for a certain domain
- Is it a combination of known logics?

Methodology: three types of conditions

Type I: Good design and sound methodology

- Is it logic?
  - May seem circular. But some reasoning mechanisms in KR lack some natural properties, or perhaps defined for very restricted language fragments.
- Is the reasoning based on a known underlying logic?
  - If so, we may profit from known properties and successes for a certain domain
- Is it a combination of known logics?
  - Related to requirements of Carnap and Gelfond. Much is known about combining logics and different operators, eg knowledge and belief, tense and modality, space and time. Typical design criterion: Does the combined formalism have a clear connection to its constituent logics.

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?
  - Have a clear semantics and adequate inference relation?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?
  - Have a clear semantics and adequate inference relation?
- Does it accommodate new cases in a clear and natural manner?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?
  - Have a clear semantics and adequate inference relation?
- Does it accommodate new cases in a clear and natural manner?
  - (Relates to Carnap and Gelfond's ideas of simplicity and clarity). Does it provide a general approach beyond a few isolated cases? (Fruitfulness)

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?
  - Have a clear semantics and adequate inference relation?
- Does it accommodate new cases in a clear and natural manner?
  - (Relates to Carnap and Gelfond's ideas of simplicity and clarity). Does it provide a general approach beyond a few isolated cases? (Fruitfulness)
- Does it possess desirable metatheoretic properties?

Methodology: three types of conditions

Type II: Specific adequacy for the logical concepts to be formalised

- Does it adequately reconstruct/formalise the intended concepts?
  - Is it sufficiently expressive? Is it based on a rigorous, informal analysis of the concept?
- Does it offer suitable reasoning mechanisms for those concepts?
  - Have a clear semantics and adequate inference relation?
- Does it accommodate new cases in a clear and natural manner?
  - (Relates to Carnap and Gelfond's ideas of simplicity and clarity). Does it provide a general approach beyond a few isolated cases? (Fruitfulness)
- Does it possess desirable metatheoretic properties?
  - May be properties of tractability or others that are desirable in a given KR context?

Methodology: three types of conditions

Type III: Methods of reasoning that may lead to explainable AI and support the rational acceptability of conclusions

- Can it be combined with methods of explanation?

Methodology: three types of conditions

Type III: Methods of reasoning that may lead to explainable AI and support the rational acceptability of conclusions

- Can it be combined with methods of explanation?
- Can explanations be broken down into simple steps for human comprehension and rational acceptance?

Methodology: three types of conditions

Type III: Methods of reasoning that may lead to explainable AI and support the rational acceptability of conclusions

- Can it be combined with methods of explanation?
- Can explanations be broken down into simple steps for human comprehension and rational acceptance?
  - This is currently an important topic of inquiry. May involve ability to provide the primary reasoning mechanism with a simple, secondary type of logic that can add justification steps, proof trees, explanation graphs, etc that are convincing to a rational agent.

Examples that fail the sound methodology principle

What may happen when we try to keep classical logic against all odds?

Examples that fail the sound methodology principle

What may happen when we try to keep classical logic against all odds?

Consider the simple program rule $p \vee \neg p \rightarrow p$

Examples that fail the sound methodology principle

What may happen when we try to keep classical logic against all odds?

Consider the simple program rule $p \vee \neg p \rightarrow p$

On one approach (so-called FLP semantics) this rule has the single intended model $\{p\}$. Why? Because $p \vee \neg p$ is a tautology

Examples that fail the sound methodology principle

What may happen when we try to keep classical logic against all odds?

Consider the simple program rule $p \vee \neg p \rightarrow p$

On one approach (so-called FLP semantics) this rule has the single intended model $\{p\}$. Why? Because $p \vee \neg p$ is a tautology

But the rule has no stable (equilibrium) model ($\langle \{\}, \{p\} \rangle$ is a (non-stable) equilibrium model).

So, what is a tautology?

Perhaps it is whatever we can add to a program without changing its stable models

So, what is a tautology?

Perhaps it is whatever we can add to a program without changing its stable models

In that case $p \vee \neg p$ is not a tautology; adding $p \vee \neg p$ to the program $p \rightarrow q; \neg p \rightarrow r$ (whose answer set is $\{r\}$) produces an additional answer set $\{p, q\}$.

So, what is a tautology?

Perhaps it is whatever we can add to a program without changing its stable models

In that case $p \vee \neg p$ is not a tautology; adding $p \vee \neg p$ to the program $p \rightarrow q; \neg p \rightarrow r$ (whose answer set is $\{r\}$) produces an additional answer set $\{p, q\}$.

And in this case we have a disjunctive program whose semantics is not in dispute. So, why regard as a tautology something that changes the meaning of a simple program?

another example

Consider the program $\neg\neg p \rightarrow p$

another example

Consider the program $\neg\neg p \rightarrow p$

According to critics of equilibrium logic, this program should not have $\{p\}$ as an answer set (it has a 'circular justification').

another example

Consider the program $\neg\neg p \rightarrow p$

According to critics of equilibrium logic, this program should not have $\{p\}$ as an answer set (it has a 'circular justification').

But in equilibrium logic $\{p\}$ is an answer set. You can see this in two ways:

another example

Consider the program $\neg\neg p \to p$

According to critics of equilibrium logic, this program should not have $\{p\}$ as an answer set (it has a 'circular justification').

But in equilibrium logic $\{p\}$ is an answer set. You can see this in two ways:

To get the equilibrium fixpoint you add negated literals. If you add $\neg p$ you satisfy the formula but the answer set is { }. If you add $\neg\neg p$ you get the answer set $\{p\}$.

another example

Consider the program $\neg\neg p \rightarrow p$

According to critics of equilibrium logic, this program should not have $\{p\}$ as an answer set (it has a 'circular justification').

But in equilibrium logic $\{p\}$ is an answer set. You can see this in two ways:

To get the equilibrium fixpoint you add negated literals. If you add $\neg p$ you satisfy the formula but the answer set is { }. If you add $\neg\neg p$ you get the answer set $\{p\}$.

You can use **monotonic reasoning**. (As we saw last week), in all (and only) those logics that capture the strong equivalence of logic programs (KC to HT), $\neg\neg p \rightarrow p$ **is equivalent to** $p \vee \neg p$. No one in ASP denies that the second formula has an answer set $\{p\}$!

It gets worse ...

Let Π be the propositional program:

$$\neg p \rightarrow p \tag{1}$$

$$\neg\neg p \rightarrow p \tag{2}$$

- This has $\{p\}$ as its equilibrium or general stable model. Yet critics say this suffers from a circular justification. Oh Dear!

It gets worse ...

Let Π be the propositional program:

$$\neg p \rightarrow p \tag{1}$$

$$\neg\neg p \rightarrow p \tag{2}$$

- This has $\{p\}$ as its equilibrium or general stable model. Yet critics say this suffers from a circular justification. Oh Dear!
- But (1) is logically equivalent, even in constructive logic, to the formula $\neg\neg p$. So in Π p follows directly from (2) and re-written (1) by *modus ponens*! The inference to p is entirely monotonic and there is no issue of circular justification.

An alternative analysis

Since Π has the form $A \rightarrow C$ and $B \rightarrow C$, we should be able to infer that also $A \vee B \rightarrow C$. This holds as an axiom:

$$\vdash (A \rightarrow C) \wedge (B \rightarrow C) \rightarrow (A \vee B \rightarrow C) \tag{3}$$

in INT and even in minimal logic and in Anderson and Belnap's basic relevance logic **R**.

An alternative analysis

Since Π has the form $A \to C$ and $B \to C$, we should be able to infer that also $A \vee B \to C$. This holds as an axiom:

$$\vdash (A \to C) \wedge (B \to C) \to (A \vee B \to C) \qquad (3)$$

in INT and even in minimal logic and in Anderson and Belnap's basic relevance logic **R**.
Applying to Π we should obtain

$$\neg p \vee \neg\neg p \to p \qquad (4)$$

Since $\neg p \vee \neg\neg p$ is a tautology in classical logic as well as in **HT**, we should be able to infer p. Yet this is not the case, neither in FLP-semantics nor in modified versions. Since they accept $\neg p \vee \neg\neg p$ as a tautology, the failure to infer p must be due to a failure to accept (3).

The same example in terms of rules

Think of Π as a set of rules

$$\dfrac{A \vee B \qquad \begin{matrix} A \\ \vdots \\ C \end{matrix} \qquad \begin{matrix} B \\ \vdots \\ C \end{matrix}}{C}$$

Figure: Rule of disjunction elimination.

- So $\neg p \vee \neg\neg p \rightarrow p$ is derivable in constructive reasoning and the inference to p will follow in logics admitting the weak law of excluded middle.

The same example in terms of rules

Think of Π as a set of rules

$$\frac{A \vee B \qquad \begin{array}{c} A \\ \vdots \\ C \end{array} \qquad \begin{array}{c} B \\ \vdots \\ C \end{array}}{C}$$

Figure: Rule of disjunction elimination.

- So $\neg p \vee \neg\neg p \rightarrow p$ is derivable in constructive reasoning and the inference to p will follow in logics admitting the weak law of excluded middle.

- The approach of FLP lacks coherence because the type of logical reasoning that is permitted in determining when a rule atom is (non-circularly) inferable is quite different from the type of reasoning which would allow us to move from two different rules to a third one.