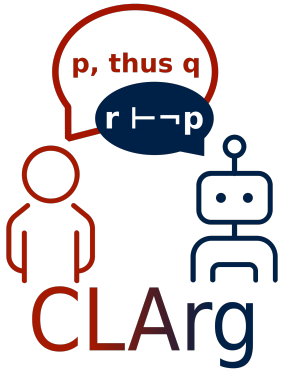
Imperial College London



Monotonicity, Noise-Tolerance, and Explanations in Case-Based Reasoning with Abstract Argumentation

Guilherme Paulino-Passos

Department of Computing
Computational Logic and Argumentation Group

Introduction

Introduction

- Case-based reasoning (CBR)

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example?

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example? k-nearest neighbours!

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example? k-nearest neighbours!
- An abstract argumentation approach to CBR: *AA-CBR*

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example? k-nearest neighbours!
- An abstract argumentation approach to CBR: *AA-CBR*
- Usages:

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example? k-nearest neighbours!
- An abstract argumentation approach to CBR: *AA-CBR*
- Usages:
  - as a (white-box) classification system

Introduction

- Case-based reasoning (CBR)
  - solving new situations based on how similar cases wered solved in the past
  - well known example? k-nearest neighbours!
- An abstract argumentation approach to CBR: *AA-CBR*
- Usages:
  - as a (white-box) classification system
  - as a source of explanations (exploiting the argumentative structure)

Summary of the talk

1. Present *AA-CBR* and how it has been used

Summary of the talk

1. Present *AA-CBR* and how it has been used
2. Approaches to explainability with it

Summary of the talk

1. Present *AA-CBR* and how it has been used
2. Approaches to explainability with it
3. Non-monotonicity and incoherence in *AA-CBR* (our contribution - joint work with Francesca Toni)

Guilherme Paulino-Passos and Francesca Toni. "Monotonicity and Noise-Tolerance in Case-Based Reasoning with Abstract Argumentation". In: *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*. 2021.

Summary of the talk

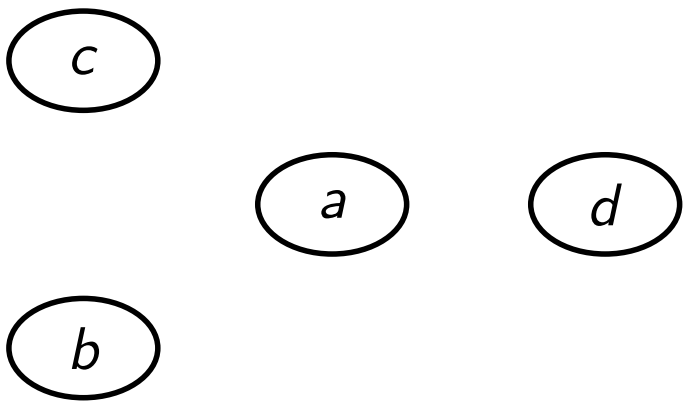
1. Present *AA-CBR* and how it has been used
2. Approaches to explainability with it
3. Non-monotonicity and incoherence in *AA-CBR* (our contribution - joint work with Francesca Toni)
4. Discussions on non-monotonicity and possible impacts on trustworthiness and explainability

---

Guilherme Paulino-Passos and Francesca Toni. "Monotonicity and Noise-Tolerance in Case-Based Reasoning with Abstract Argumentation". In: *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning*. 2021.

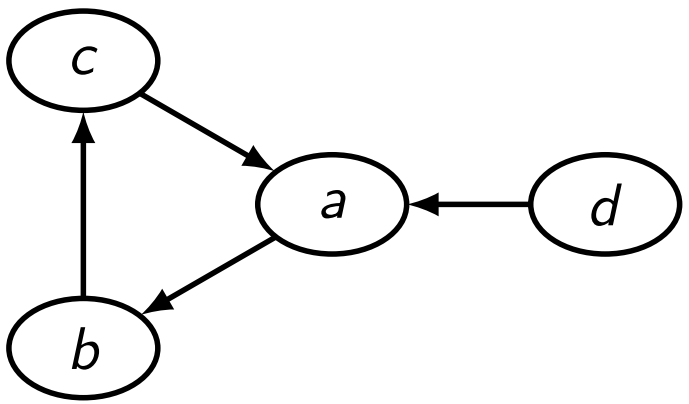
Abstract argumentation in brief

- Arguments



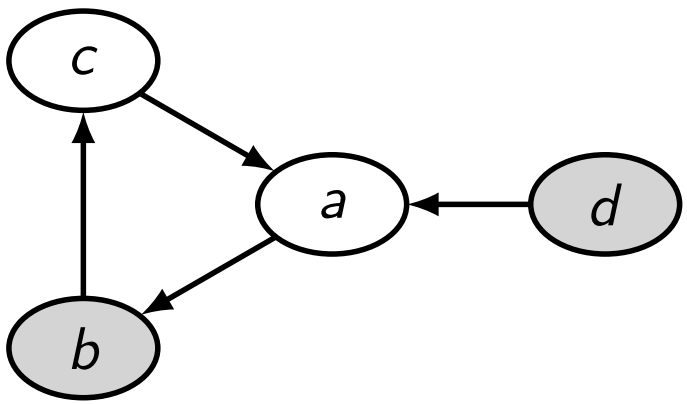
Abstract argumentation in brief

- Arguments
- Attacks



Abstract argumentation in brief

- Arguments
- Attacks
- Grounded extension



AA-CBR

Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation

Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation
- Inspiration from legal domain

Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation
- Inspiration from legal domain
- Some of those approaches have been used as classifiers in different scenarios:

Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation
- Inspiration from legal domain
- Some of those approaches have been used as classifiers in different scenarios:
  - image classification

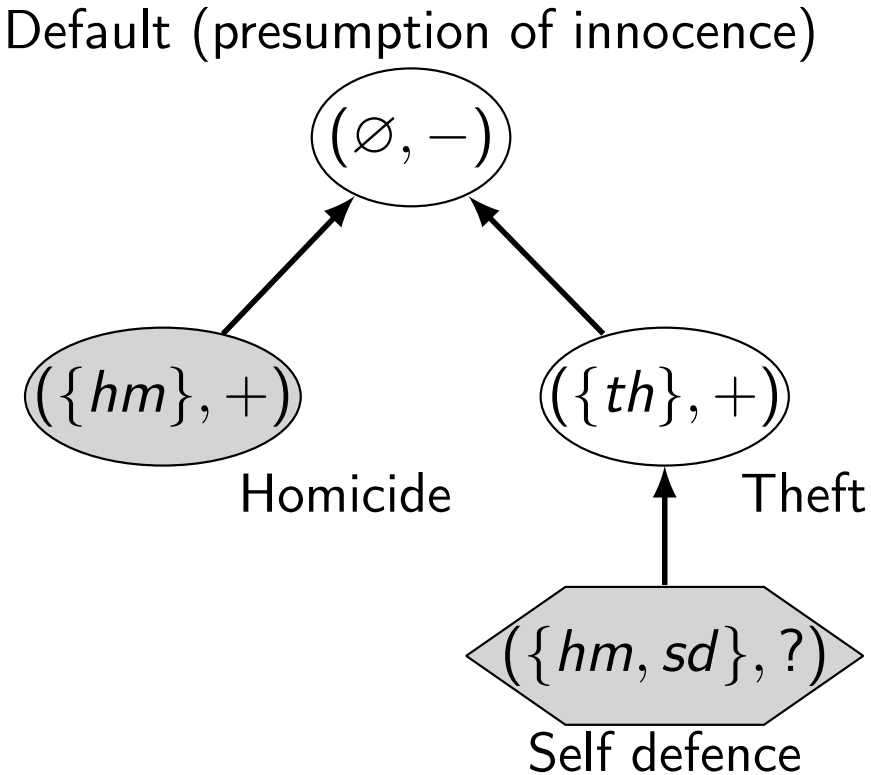
Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation
- Inspiration from legal domain
- Some of those approaches have been used as classifiers in different scenarios:
  - image classification
  - sentiment analysis in text

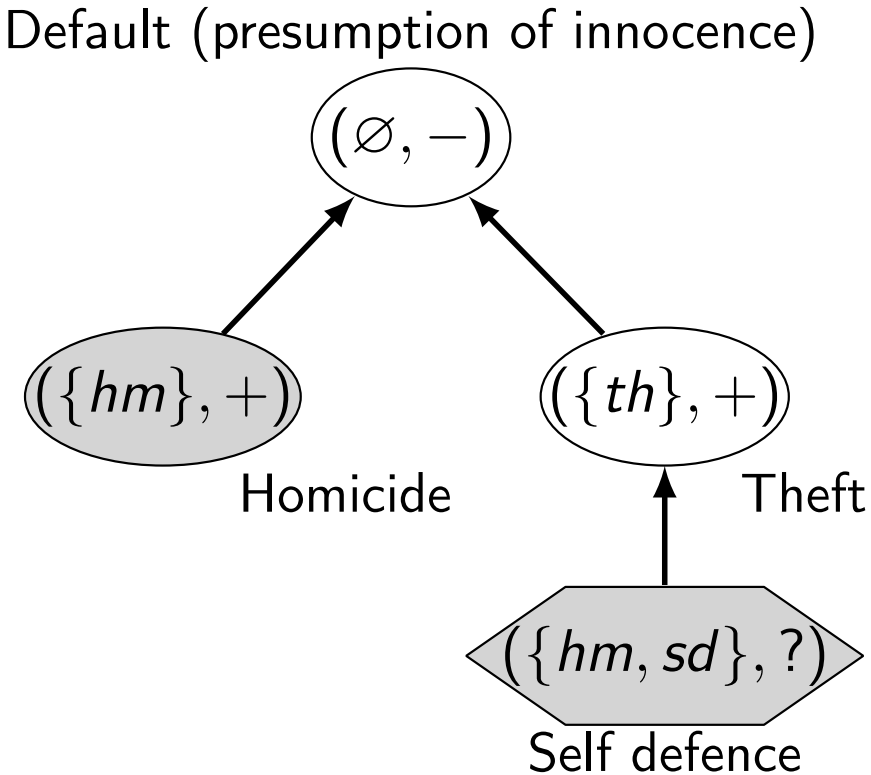
Abstract argumentation for case-based reasoning

- Modelling case-based reasoning with argumentation
- Inspiration from legal domain
- Some of those approaches have been used as classifiers in different scenarios:
  - image classification
  - sentiment analysis in text
  - predicting and explaining passage of bills in the UK parliament

AA-CBR: example

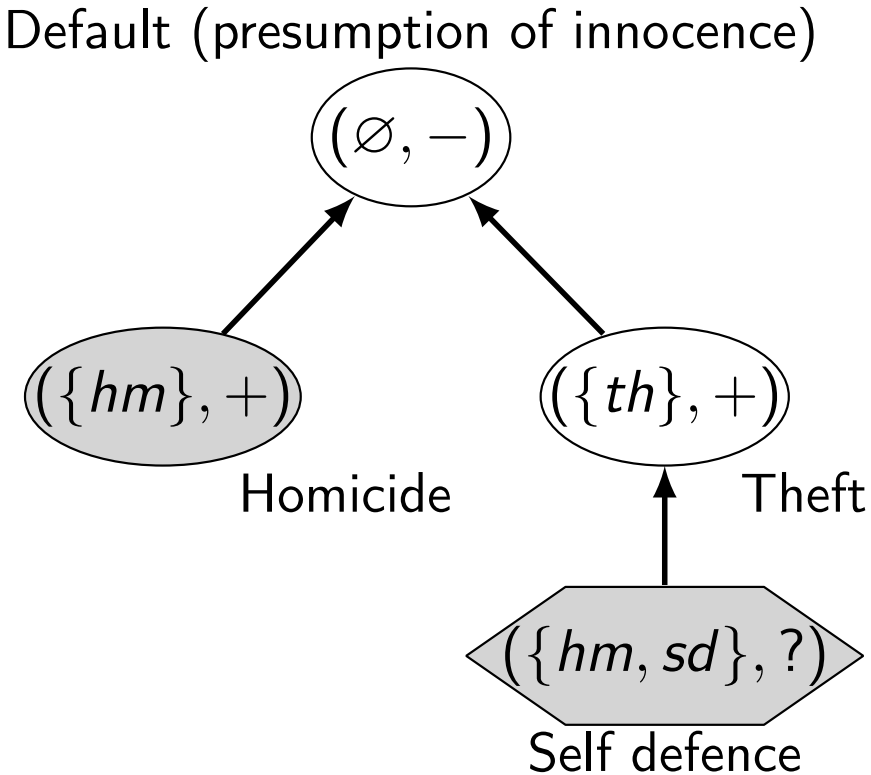


AA-CBR: example



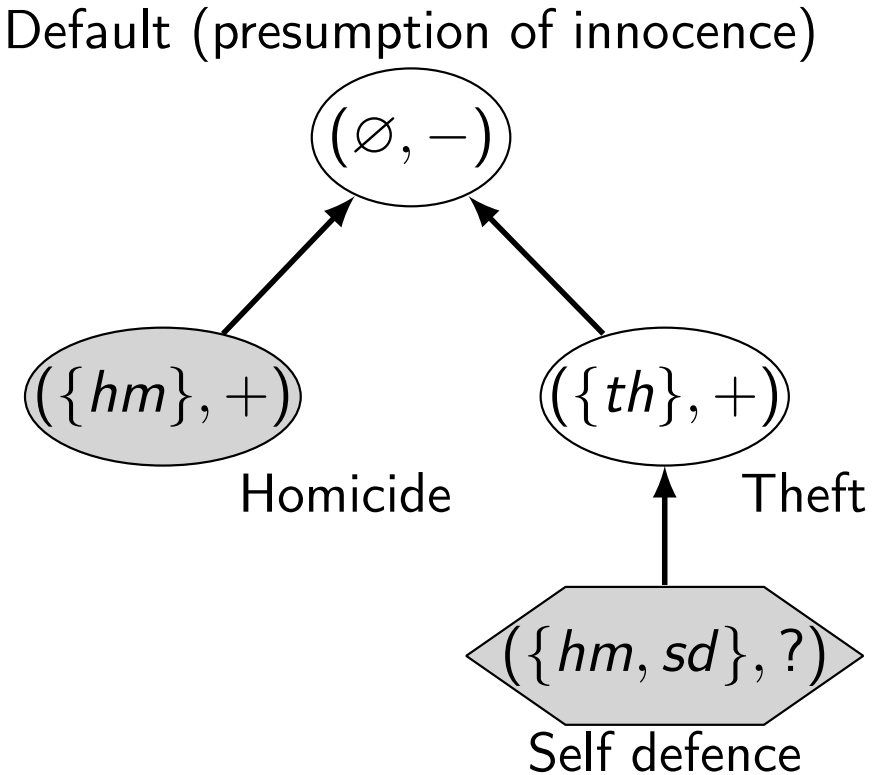
- case descriptions are partially ordered ($\succeq$)

AA-CBR: example



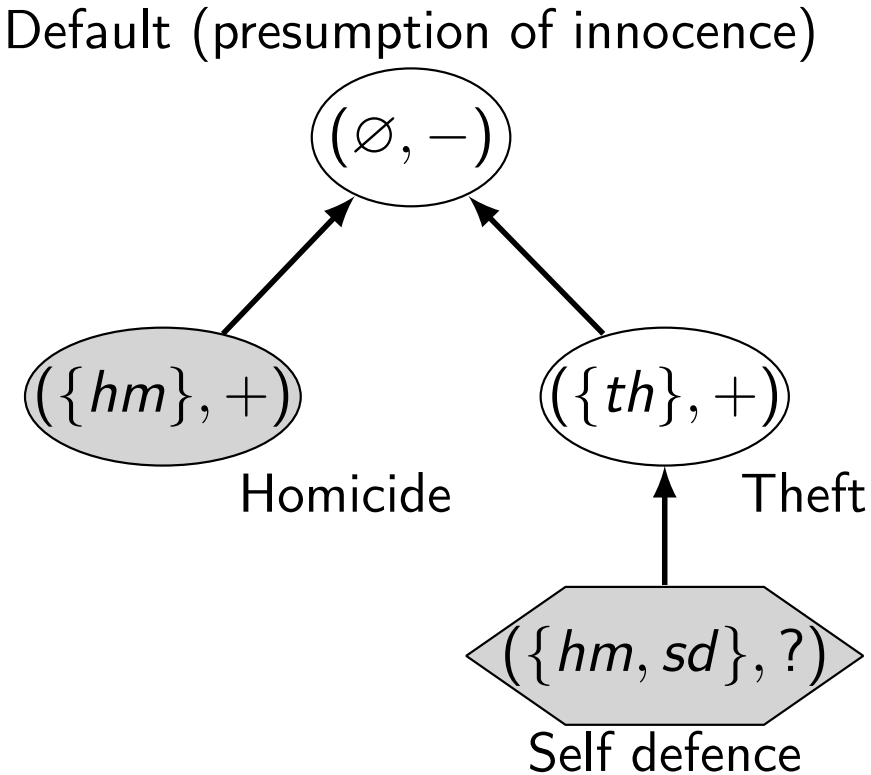
- case descriptions are partially ordered ($\succeq$)
- prediction is default outcome iff default argument is in grounded extension

AA-CBR: example



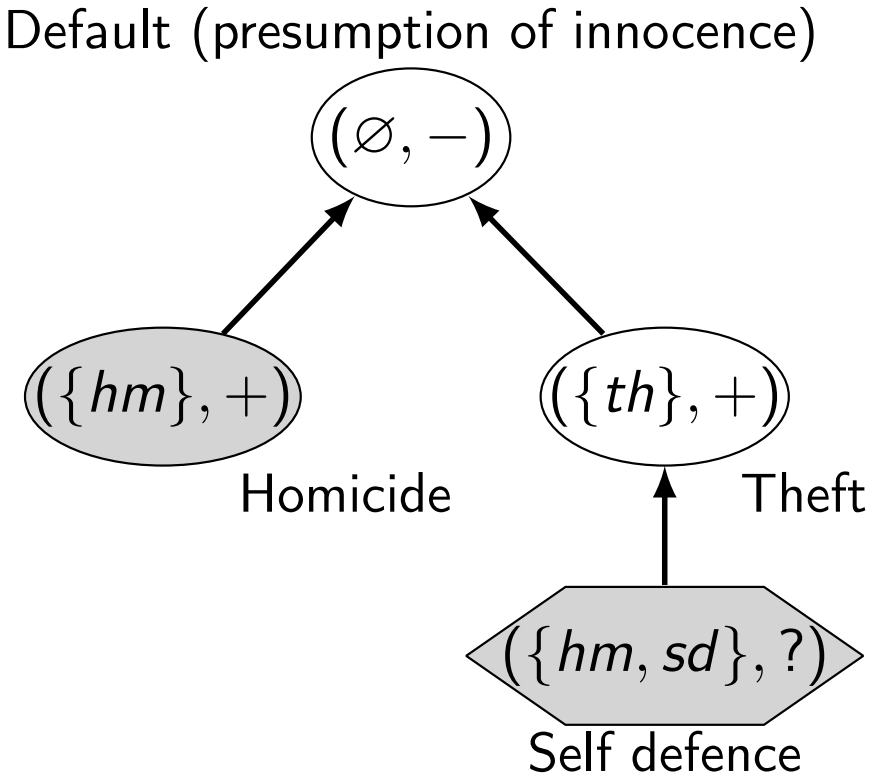
- case descriptions are partially ordered ($\succeq$)
- prediction is default outcome iff default argument is in grounded extension
- attacks from more specific to less specific with opposing outcomes

AA-CBR: example



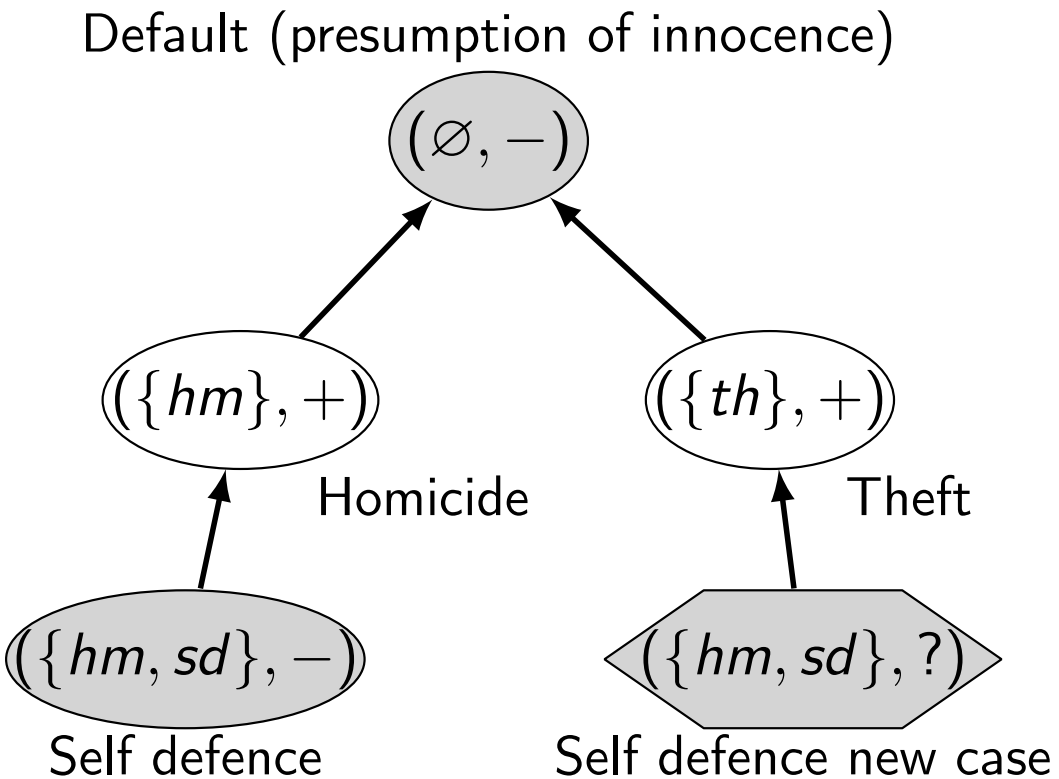
- case descriptions are partially ordered ($\succeq$)
- prediction is default outcome iff default argument is in grounded extension
- attacks from more specific to less specific with opposing outcomes
  - but only if there is no other past case in between with opposing outcome already

AA-CBR: example



- case descriptions are partially ordered ($\succeq$)
- prediction is default outcome iff default argument is in grounded extension
- attacks from more specific to less specific with opposing outcomes
  - but only if there is no other past case in between with opposing outcome already
- new case attacks irrelevant past cases (not less specific)

AA-CBR: example



Partial order

- Partial order?

Partial order

- Partial order?
  - reflexive: $x \preceq x$

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$
- Anti-symmetry is in a sense bureaucratic

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$
- Anti-symmetry is in a sense bureaucratic
  - use equivalence classes!

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$
- Anti-symmetry is in a sense bureaucratic
  - use equivalence classes!
- Intuitively

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$
- Anti-symmetry is in a sense bureaucratic
  - use equivalence classes!
- Intuitively
  - $x \preceq y$ means: x is less specific than y, or y is more informative, or x is more typical

Partial order

- Partial order?
  - reflexive: $x \preceq x$
  - transitive: $x \preceq y$ and $y \preceq z$ imply $x \preceq z$
  - anti-symmetric: $x \preceq y$ and $y \preceq x$ imply $x = y$
- Anti-symmetry is in a sense bureaucratic
  - use equivalence classes!
- Intuitively
  - $x \preceq y$ means: x is less specific than y, or y is more informative, or x is more typical
- In a sense, product of feature engineering

Approaches in past work

- subset relation (sets of binary feature)

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way
  - e.g. feature selection via autoencoder neural networks, ranking the most important

---

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way
  - e.g. feature selection via autoencoder neural networks, ranking the most important
- subset, but change what is the default outcome

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way
  - e.g. feature selection via autoencoder neural networks, ranking the most important
- subset, but change what is the default outcome
- in sentiment analysis, with textual data:

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way
  - e.g. feature selection via autoencoder neural networks, ranking the most important
- subset, but change what is the default outcome
- in sentiment analysis, with textual data:
  - occurring words are clustered and given a weight based on their sentiment

---

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches in past work

- subset relation (sets of binary feature)
- select a subset of binary features in some way
  - e.g. feature selection via autoencoder neural networks, ranking the most important
- subset, but change what is the default outcome
- in sentiment analysis, with textual data:
  - occurring words are clustered and given a weight based on their sentiment
  - then $x_1 \preceq y_1$ iff $\Sigma_{word \in x_1} weight(word) < \Sigma_{word \in x_2} weight(word)$ and $|x_1| < |x_2|$ (Cocarascu et al. 2020)

Oana Cocarascu et al. "Data-Empowered Argumentation for Dialectically Explainable Predictions". In: *ECAI 2020 - 24th European Conference on Artificial Intelligence*. 2020.

Approaches to explainability

Visualise argumentation framework

- problematic when there are many cases

Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Visualise argumentation framework

- problematic when there are many cases
- even then, positively received before, in the context of predicting passage of bills, to allow users to create a mental model

Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Nearest cases

- can be helpful showing the most related

Nearest cases

- can be helpful showing the most related

Theorem

If D is coherent and every nearest case to the new case has same outcome, then the $AA\text{-}CBR_{\succeq}$ outcome for the new case is exactly this outcome.

Nearest cases

- can be helpful showing the most related

Theorem

If D is coherent and every nearest case to the new case has same outcome, then the $AA\text{-}CBR_{\succeq}$ outcome for the new case is exactly this outcome.

- unclear interpretation in case more than one nearest case, with opposing outcomes

Transformation into rules

- Conversion into logic programming

Oana Cocarascu et al. "Explanatory predictions with artificial neural networks and argumentation". In: *2nd Workshop on XAI at the 27th IJCAI and the 23rd ECAI*. 2018.

Transformation into rules

- Conversion into logic programming
- Each argument becomes a predicate which is proven iff the argument is in the grounded extension, what happens iff its features are satisfied and its attackers are `not` proven

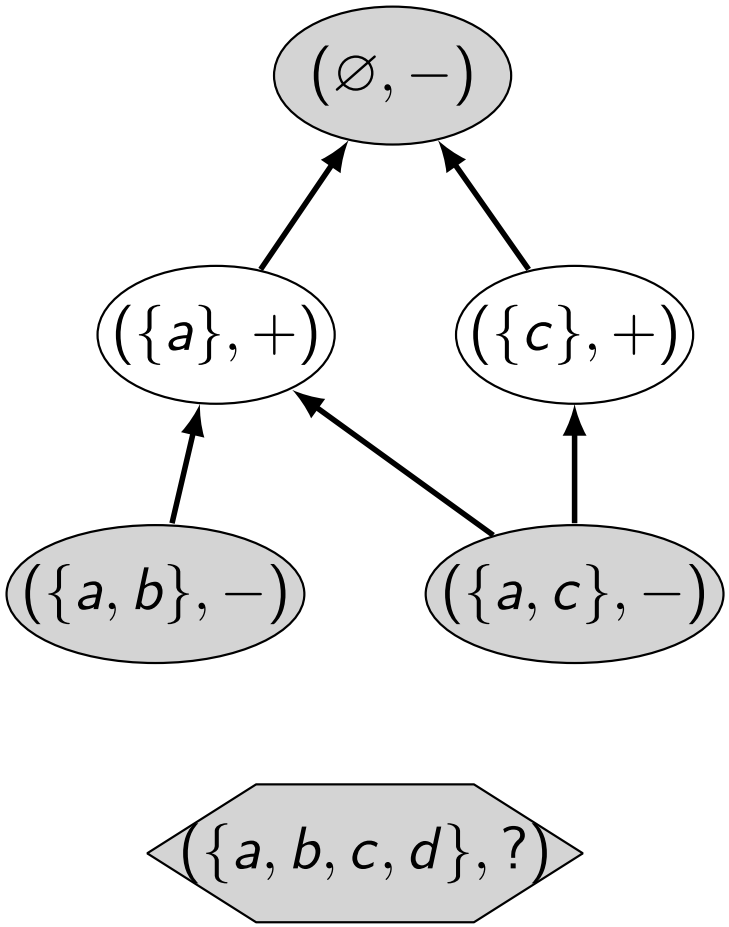
---

Oana Cocarascu et al. "Explanatory predictions with artificial neural networks and argumentation". In: *2nd Workshop on XAI at the 27th IJCAI and the 23rd ECAI*. 2018.

Dispute trees - dialectical explanations

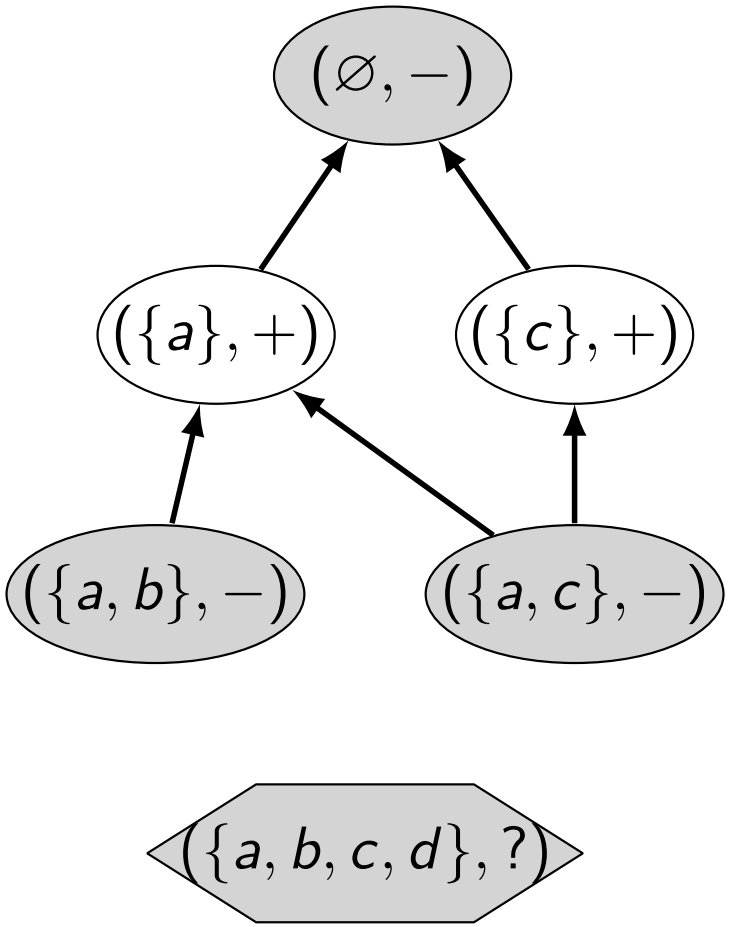
Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

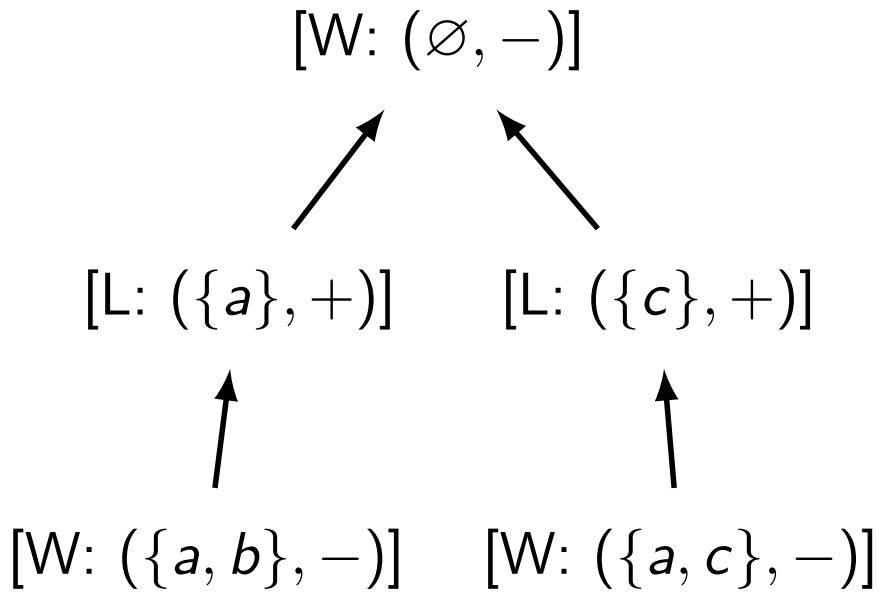
Dispute trees - dialectical explanations



Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Dispute trees - dialectical explanations



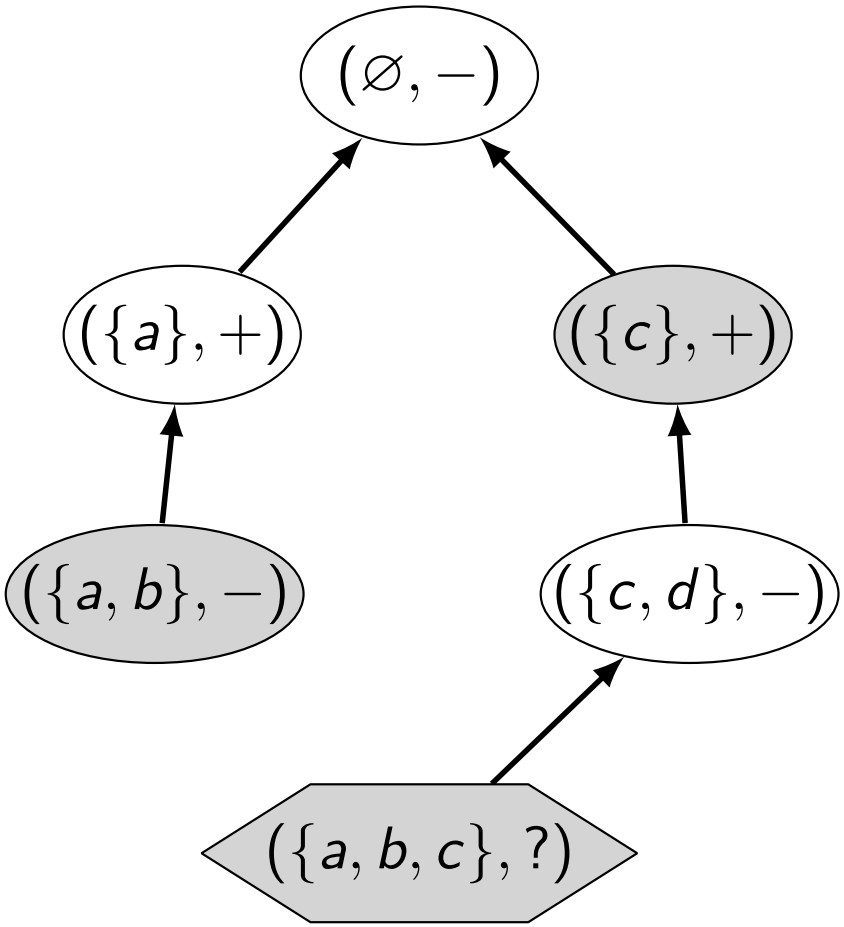


Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Dispute trees - dialectical explanations

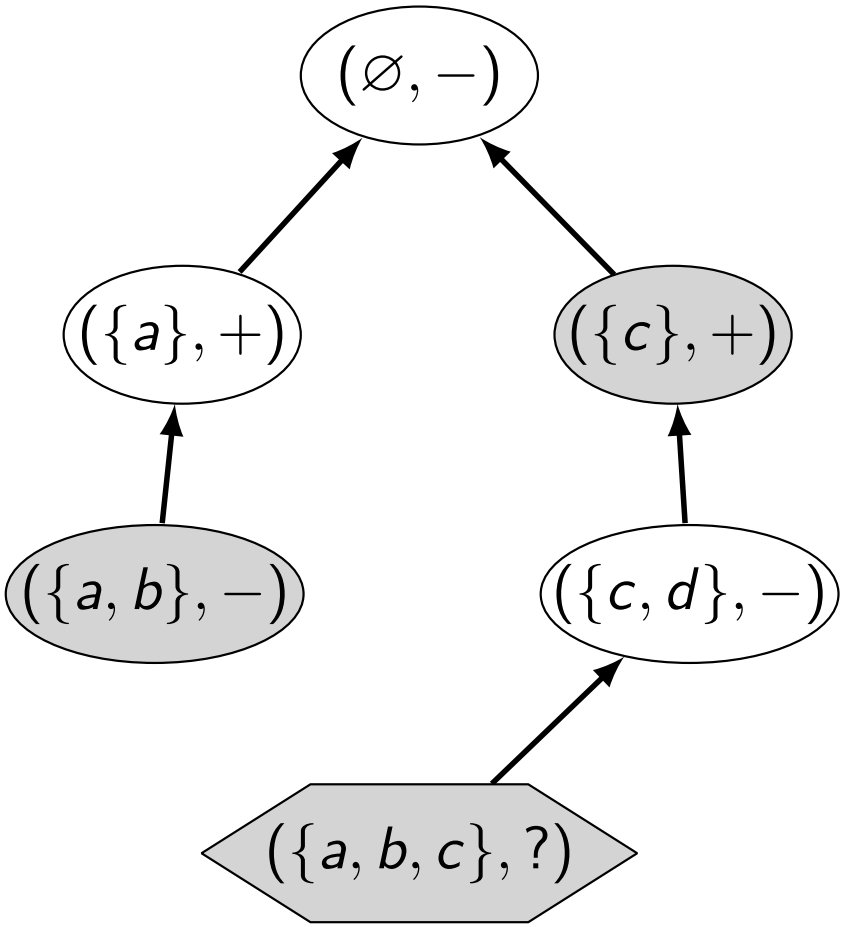
Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

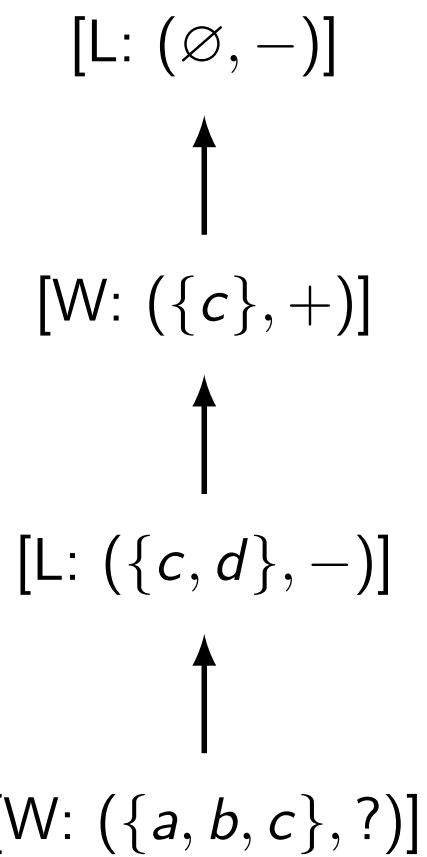
Dispute trees - dialectical explanations



Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Dispute trees - dialectical explanations





Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Excess features

- Features not contained in new case but contained in arguments attacked by the new case

Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Excess features

- Features not contained in new case but contained in arguments attacked by the new case
- In the previous example: d.

Kristijonas Čyras et al. "Explanations by arbitrated argumentative dispute". In: *Expert Syst. Appl.* 127 (2019).

Future work

- Conversational and interactive explanations

Oana Cocarascu et al. "Extracting Dialogical Explanations for Review Aggregations with Argumentative Dialogical Agents". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. 2019.

Antonio Rago et al. "Argumentative explanations for interactive recommendations". In: *Artif. Intell.* 296 (2021).

Non-monotonicity

Non-monotonicity

non-monotonicity $K \vdash a$ and $K \subseteq B$ do not imply that $B \vdash a$

Non-monotonicity

non-monotonicity $K \vdash a$ and $K \subseteq B$ do not imply that $B \vdash a$

- studied in the field of non-monotonic reasoning (NMR)

Classifiers as reasoning

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

Classifiers as reasoning

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,
define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

Classifiers as reasoning

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;

Classifiers as reasoning

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;
- $D \vdash_{\mathbb{C}} \neg(x, y)$, iff there is a y' such that $\mathbb{C}(D, x) = y'$ and $y' \neq y$.

Classifiers as reasoning

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;
- $D \vdash_{\mathbb{C}} \neg(x, y)$, iff there is a y' such that $\mathbb{C}(D, x) = y'$ and $y' \neq y$.

Cautious monotonicity

- $K \vdash a$ and $K \vdash b$ imply that $K \cup \{a\} \vdash b$

$AA\text{-}CBR_{\succeq}$ is not cautiously monotonic

Theorem

$\vdash_{AA\text{-}CBR_{\succeq}}$ *is not cautiously monotonic.*

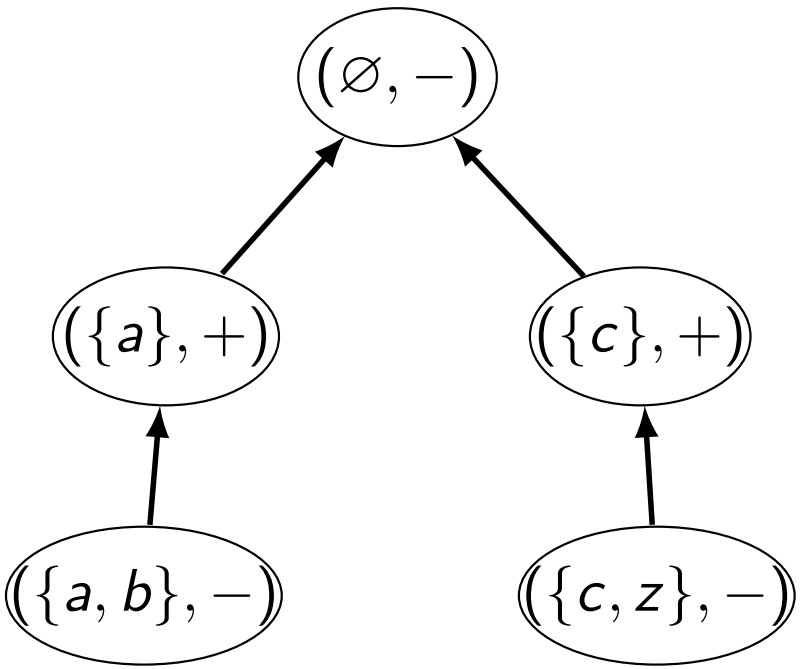
$AA\text{-}CBR_{\succeq}$ is not cautiously monotonic

Theorem

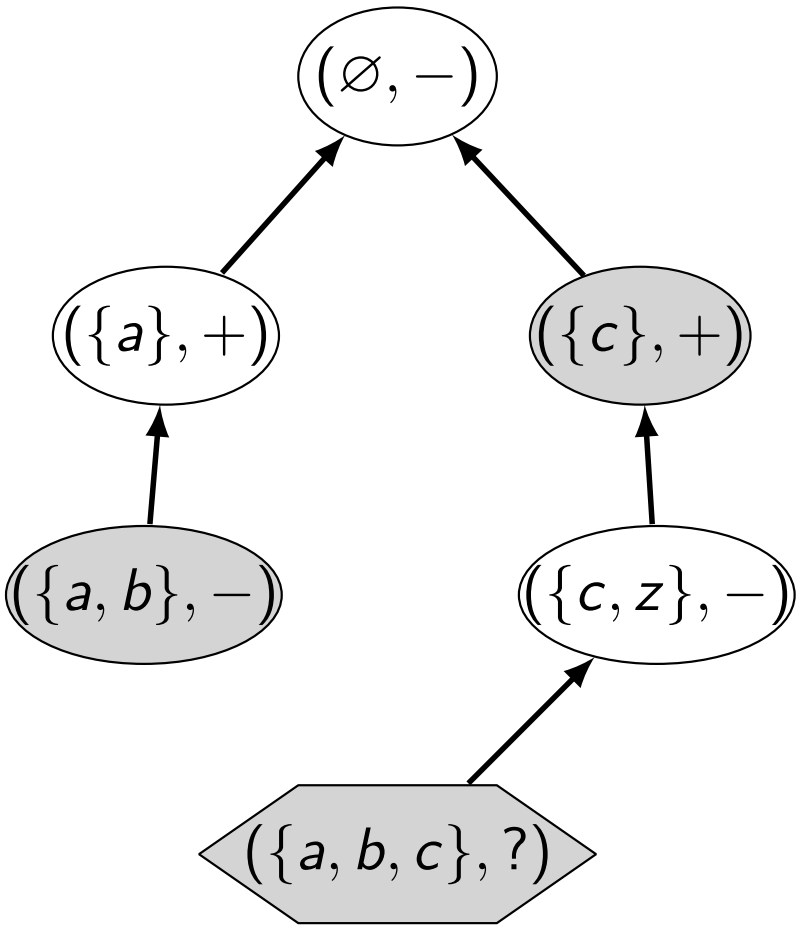
$\vdash_{AA\text{-}CBR_{\succeq}}$ *is not cautiously monotonic.*

- However, we can define a cautiously monotonic variation

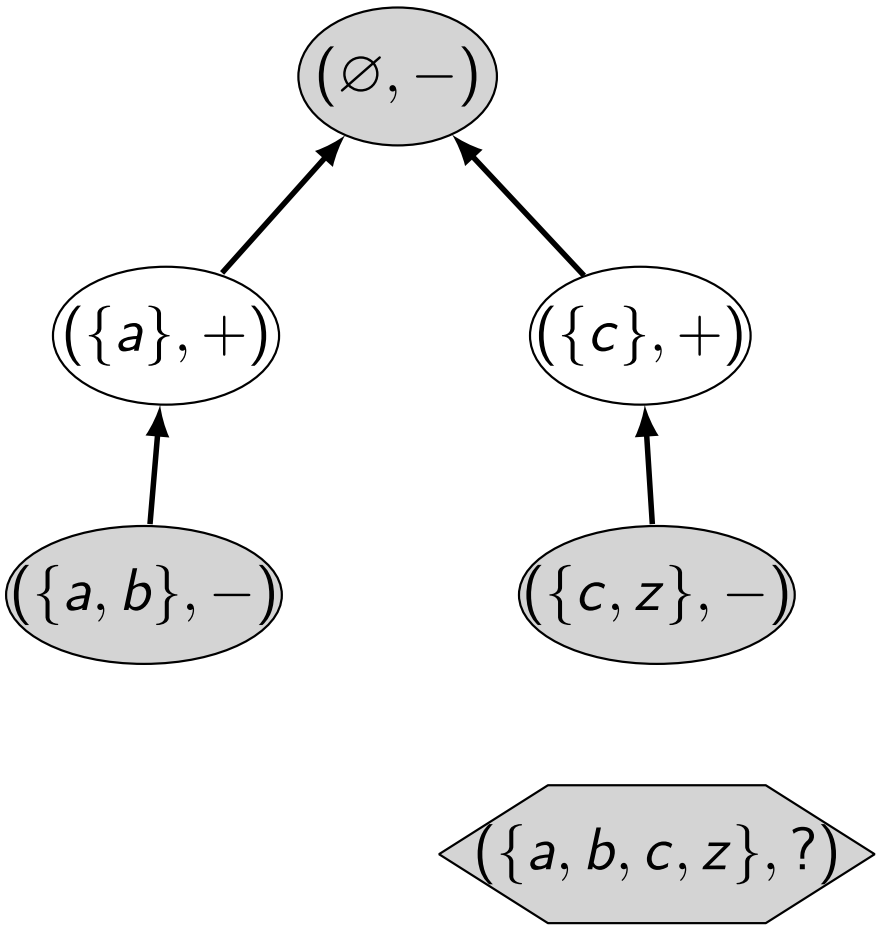
Counterexample



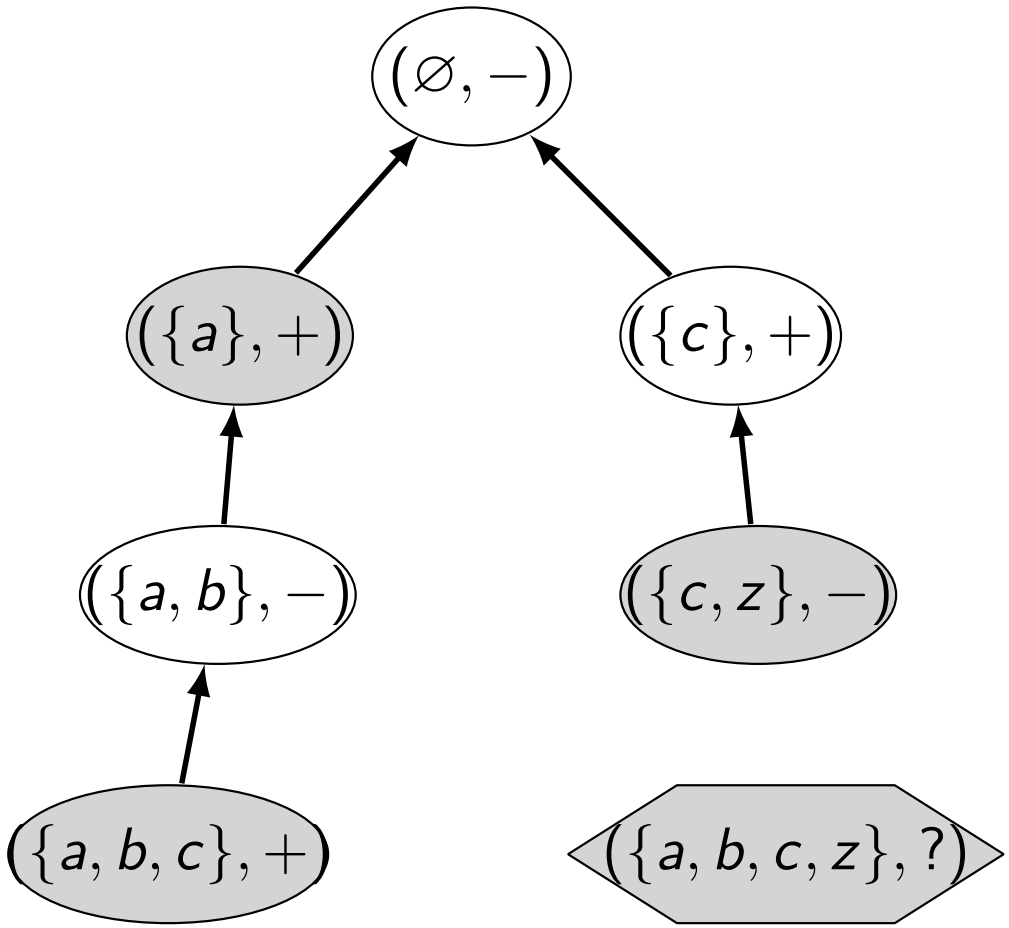
Counterexample



Counterexample



Counterexample



Counterexample

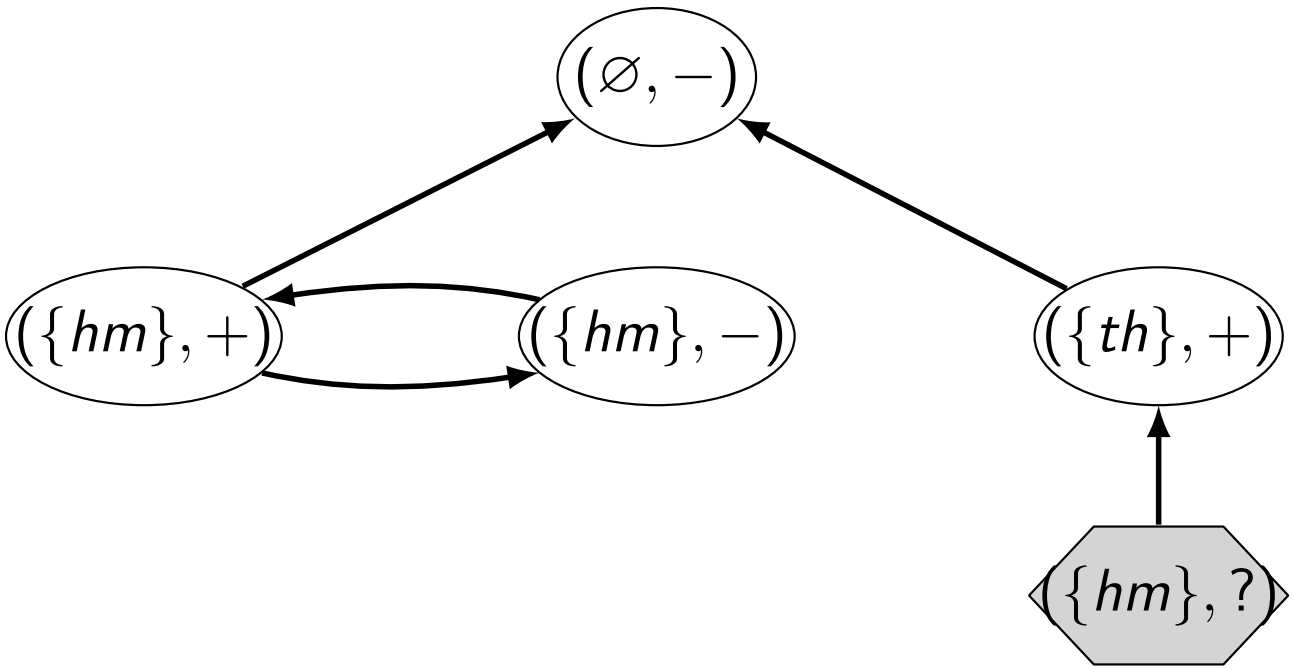
We can then conclude that $D \cup \{(N_1, +)\} \vdash_{AA\text{-}CBR_{\succeq}} (N_2, +)$ even though $D \vdash_{AA\text{-}CBR_{\succeq}} (N_1, +)$ and $D \vdash_{AA\text{-}CBR_{\succeq}} (N_2, -)$.

$AA\text{-}CBR_{\succeq}$: incoherences

- When there are incoherences, *AA-CBR* may trivialise

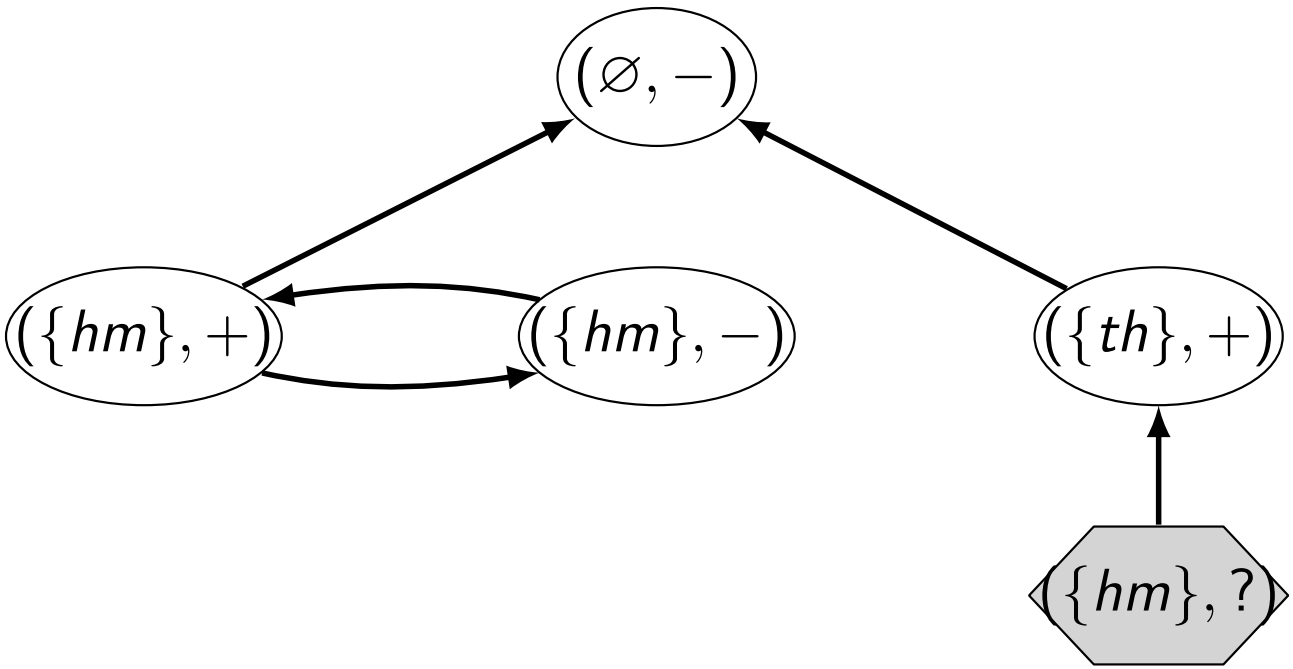
$AA\text{-}CBR_{\succeq}$: incoherences

- When there are incoherences, *AA-CBR* may trivialise
- Example:



$AA\text{-}CBR_{\succeq}$: incoherences

- When there are incoherences, *AA-CBR* may trivialise
- Example:



- Default case is not in the grounded extension, but no attacker of it is as well

Surprising cases

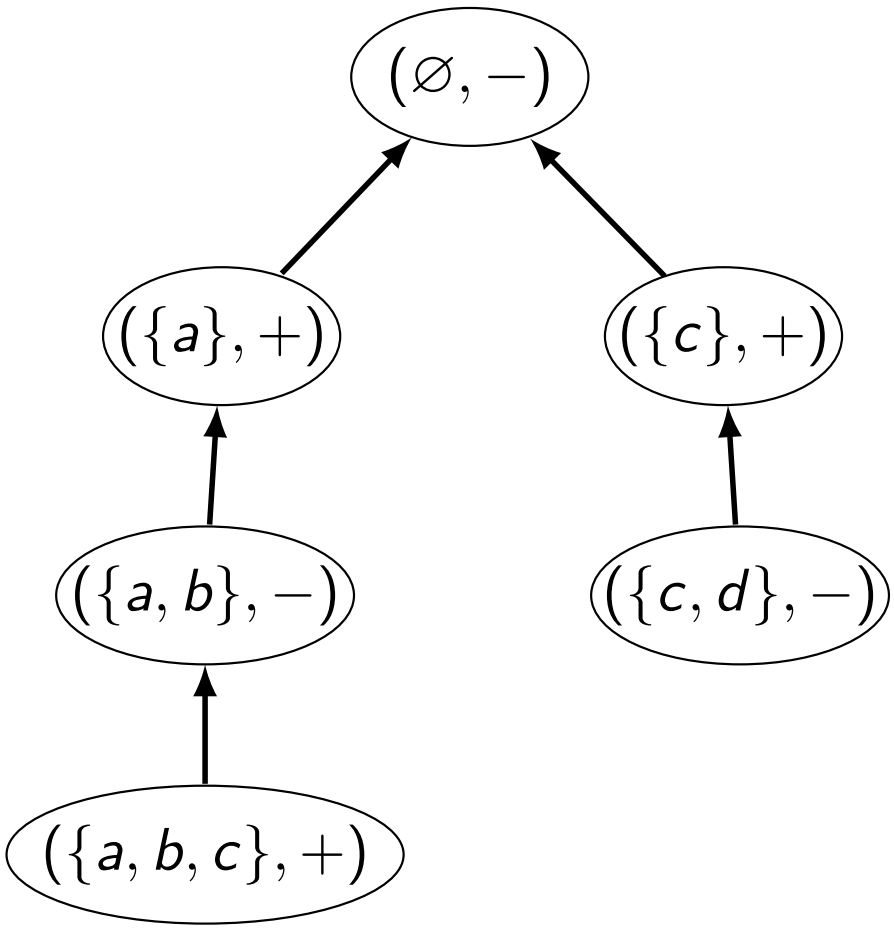
Definition

An example $(x, y) \in X \times Y$ is *surprising* w.r.t. D iff $D \setminus \{(x, y)\} \not\vdash_{\mathbb{C}} (x, y)$.

Surprising cases

Definition

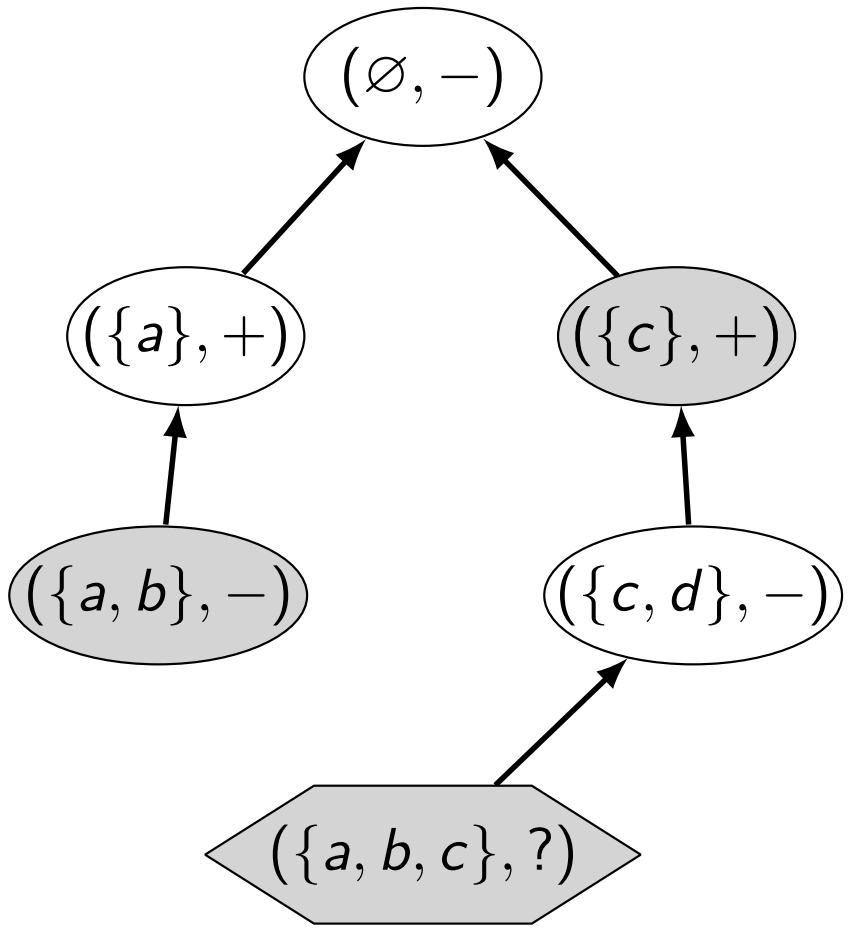
An example $(x, y) \in X \times Y$ is *surprising* w.r.t. D iff $D \setminus \{(x, y)\} \not\vdash_{\mathbb{C}} (x, y)$.



Surprising cases

Definition

An example $(x, y) \in X \times Y$ is *surprising* w.r.t. D iff $D \setminus \{(x, y)\} \not\vdash_{\mathbb{C}} (x, y)$.



$AA\text{-}CBR_{\succeq}$ and order

- A prediction in $AA\text{-}CBR_{\succeq}$ for a new case $x \in X$ depends only on the cases x' in the casebase such that $x' \preceq x$

$AA\text{-}CBR_{\succeq}$ and order

- A prediction in $AA\text{-}CBR_{\succeq}$ for a new case $x \in X$ depends only on the cases x' in the casebase such that $x' \preceq x$
- This suggests a natural way of restricting the casebase for $AA\text{-}CBR_{\succeq}$ in order to guarantee cautious monotonicity

Algorithm (sketch)

1. (topologically) sort the casebase

Algorithm (sketch)

1. (topologically) sort the casebase
2. select the $\prec$-minimal unprocessed cases

Algorithm (sketch)

1. (topologically) sort the casebase
2. select the $\prec$-minimal unprocessed cases
3. for each, consider whether they are surprising w.r.t. the selected cases

Algorithm (sketch)

1. (topologically) sort the casebase
2. select the $\prec$-minimal unprocessed cases
3. for each, consider whether they are surprising w.r.t. the selected cases
4. add the surprising ones to the selected cases, to be used for inference

Algorithm (sketch)

1. (topologically) sort the casebase
2. select the $\prec$-minimal unprocessed cases
3. for each, consider whether they are surprising w.r.t. the selected cases
4. add the surprising ones to the selected cases, to be used for inference
5. go back to 2, unless all cases have been processed

Illustration of the algorithm

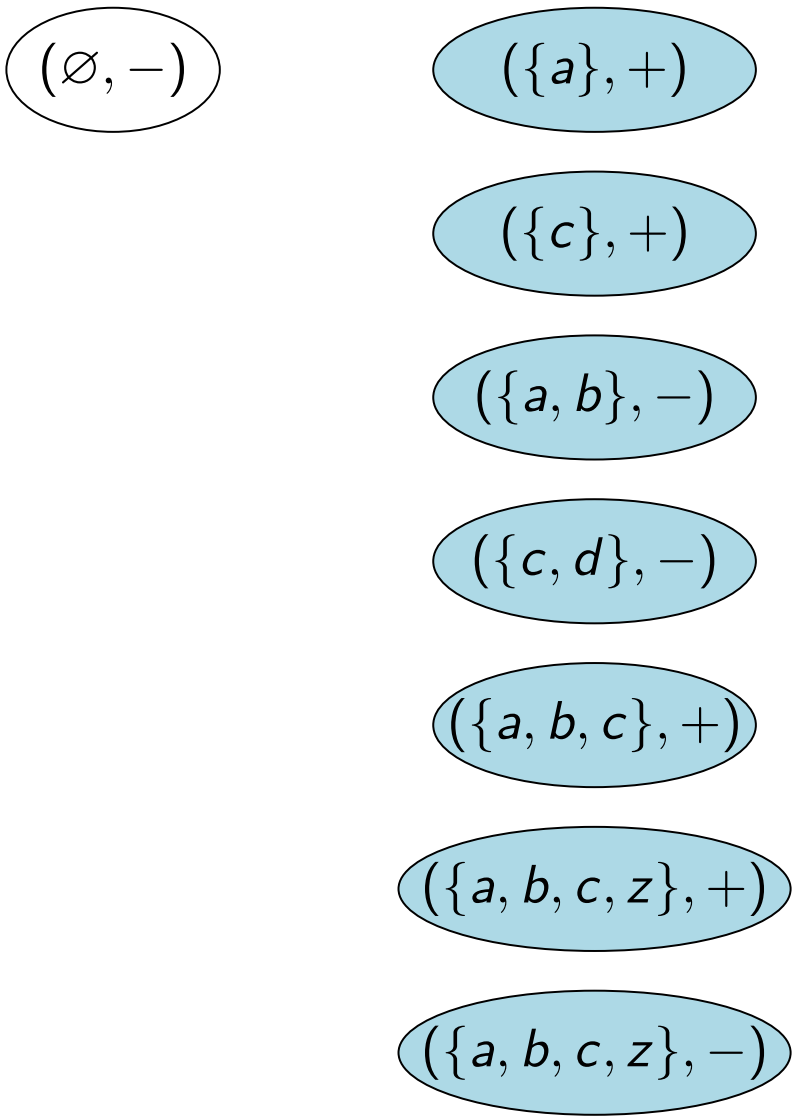


Illustration of the algorithm

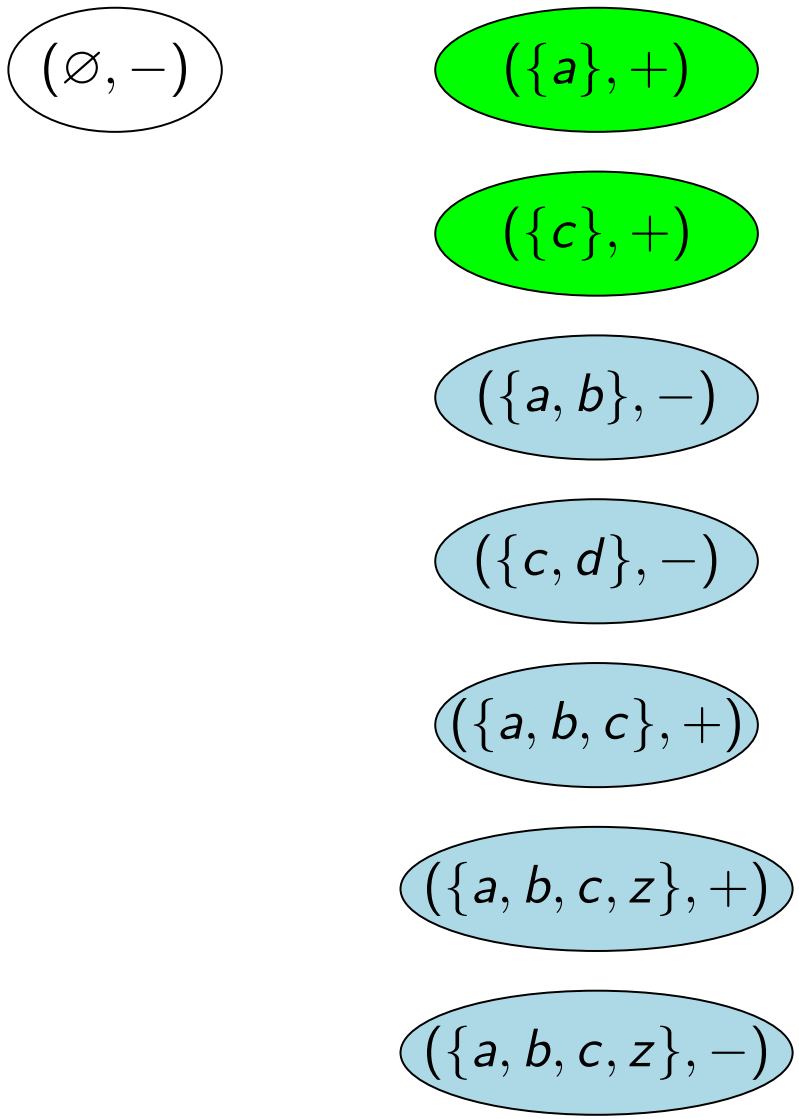


Illustration of the algorithm

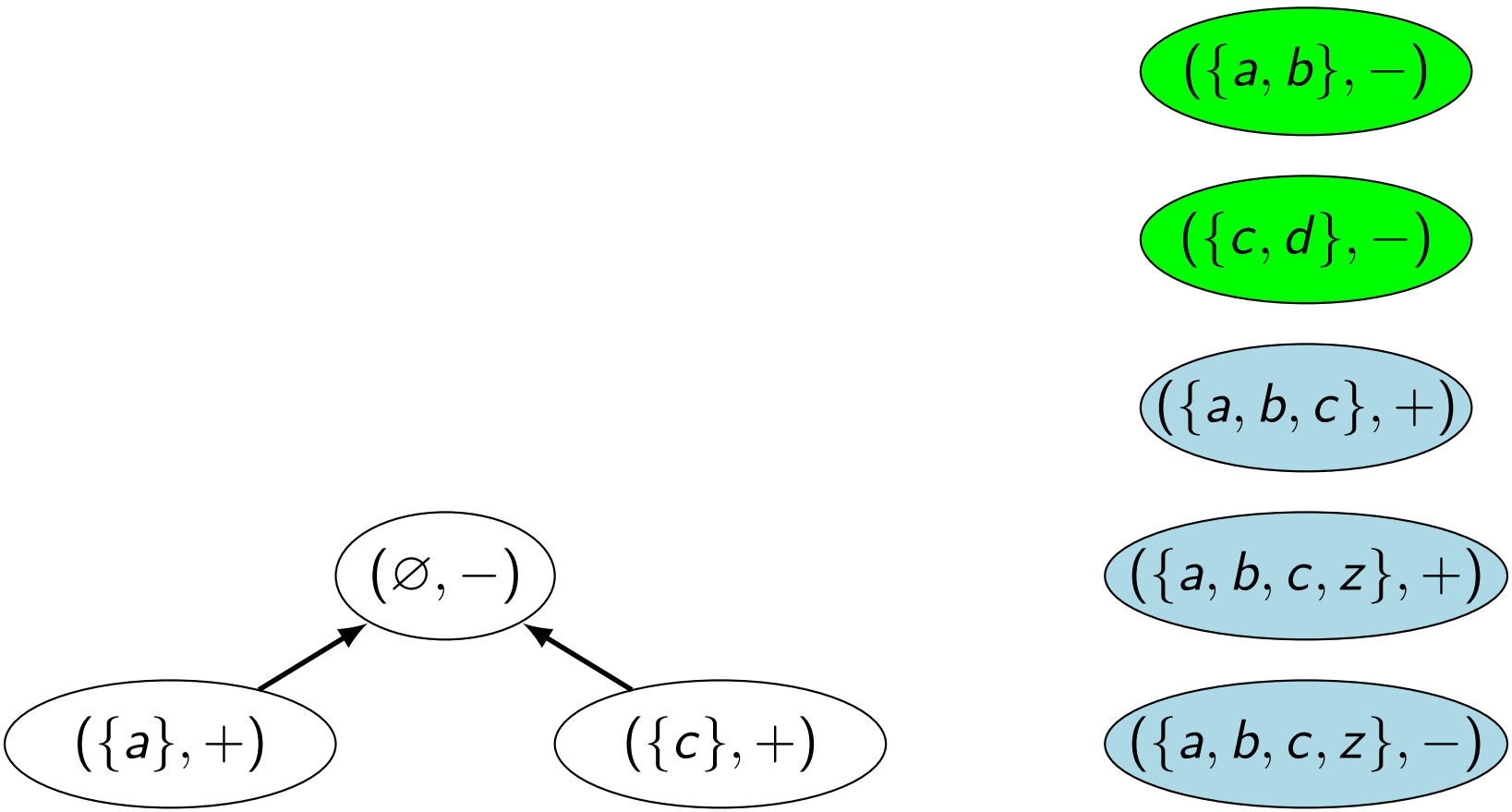


Illustration of the algorithm

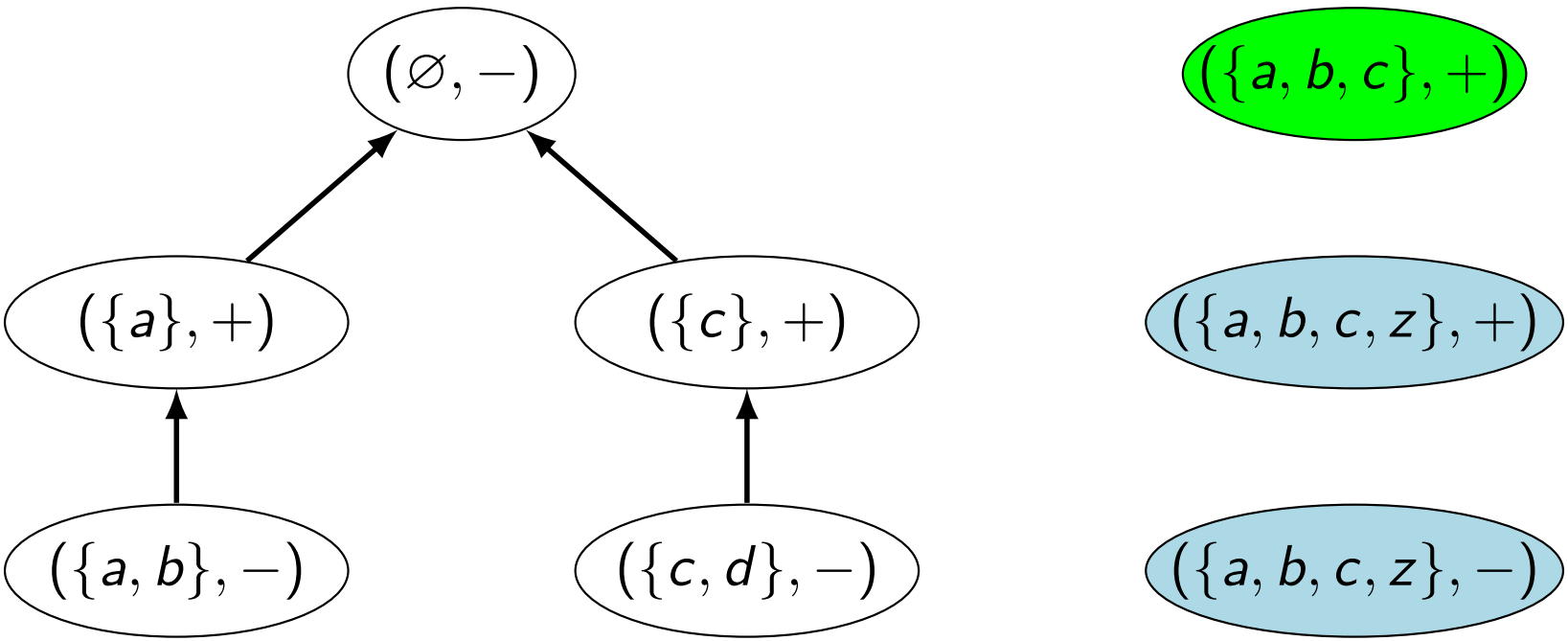


Illustration of the algorithm

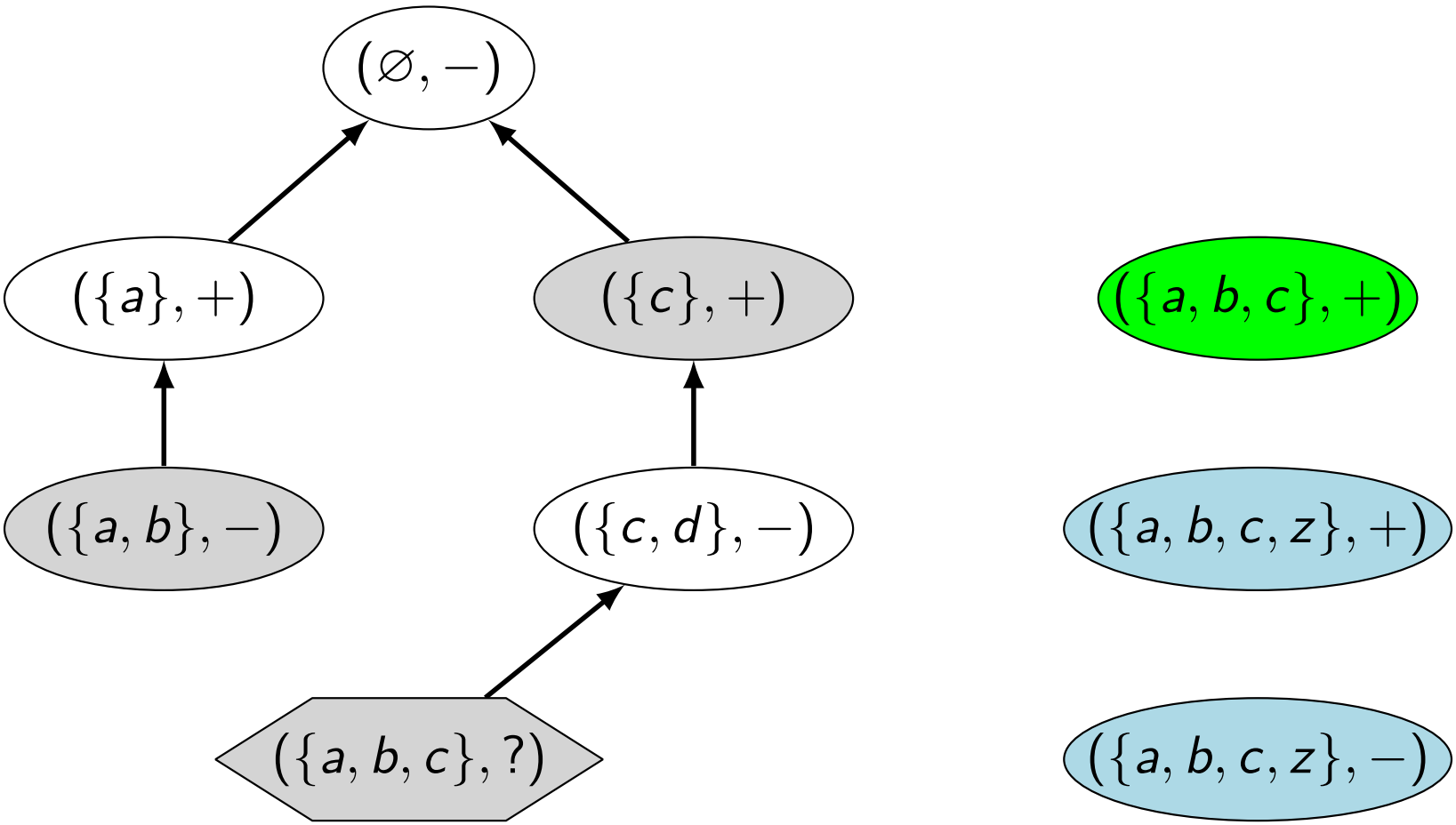


Illustration of the algorithm

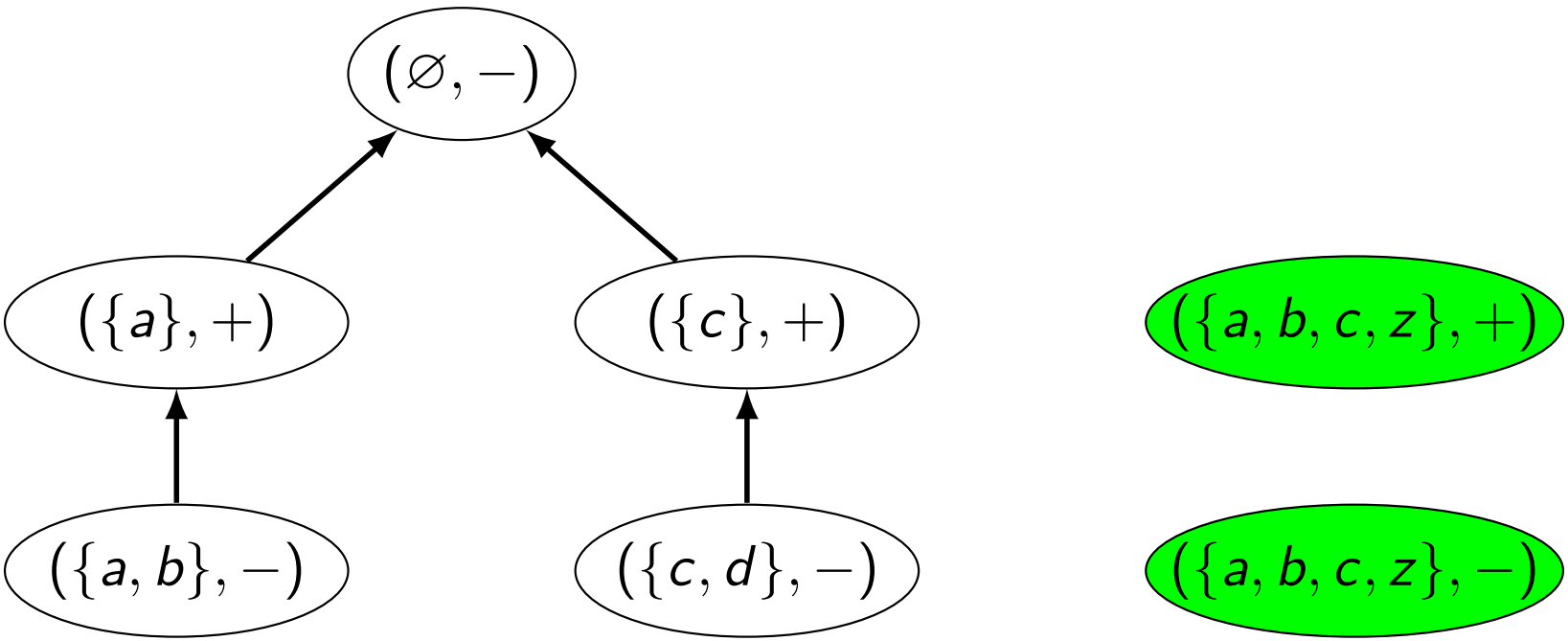


Illustration of the algorithm

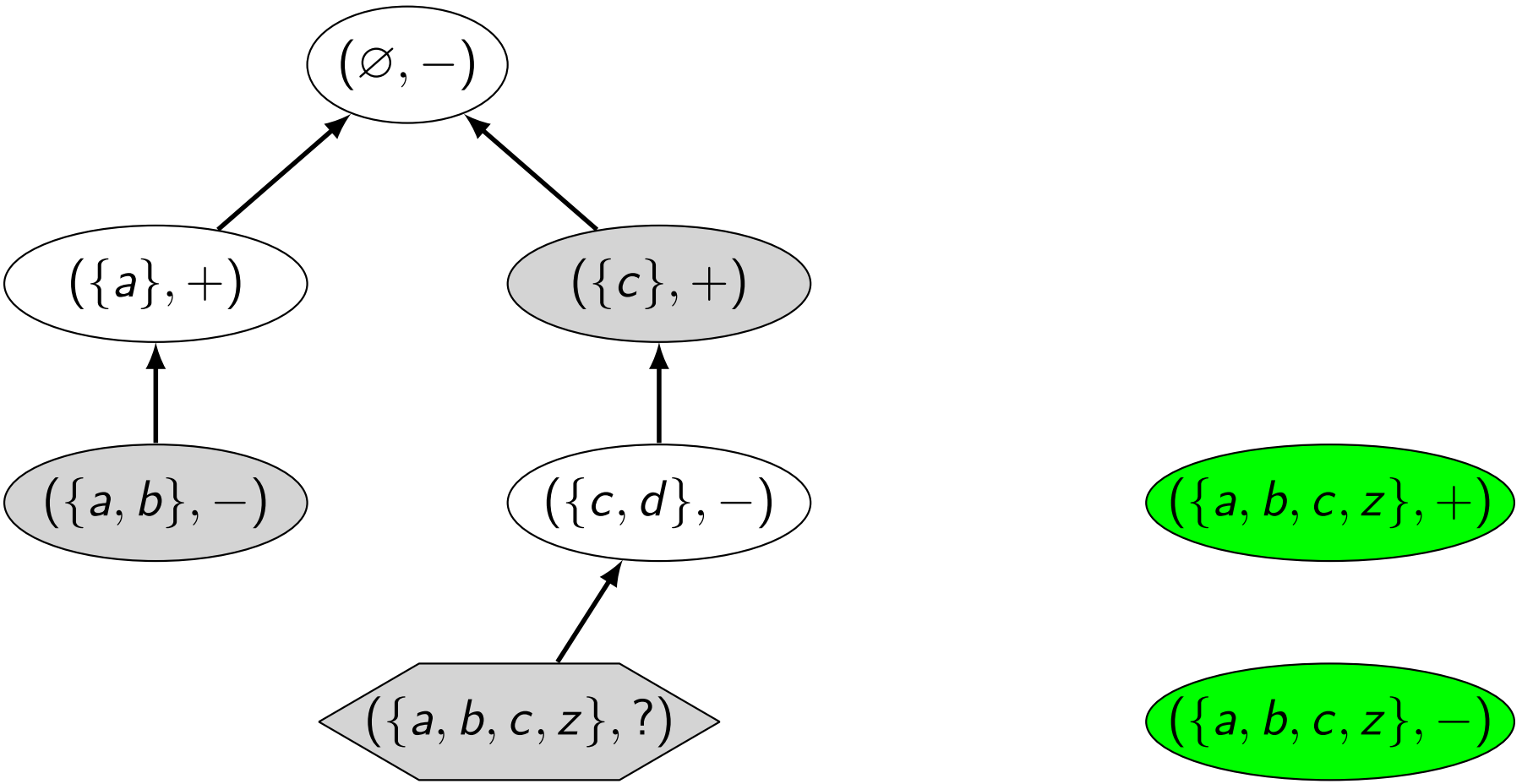
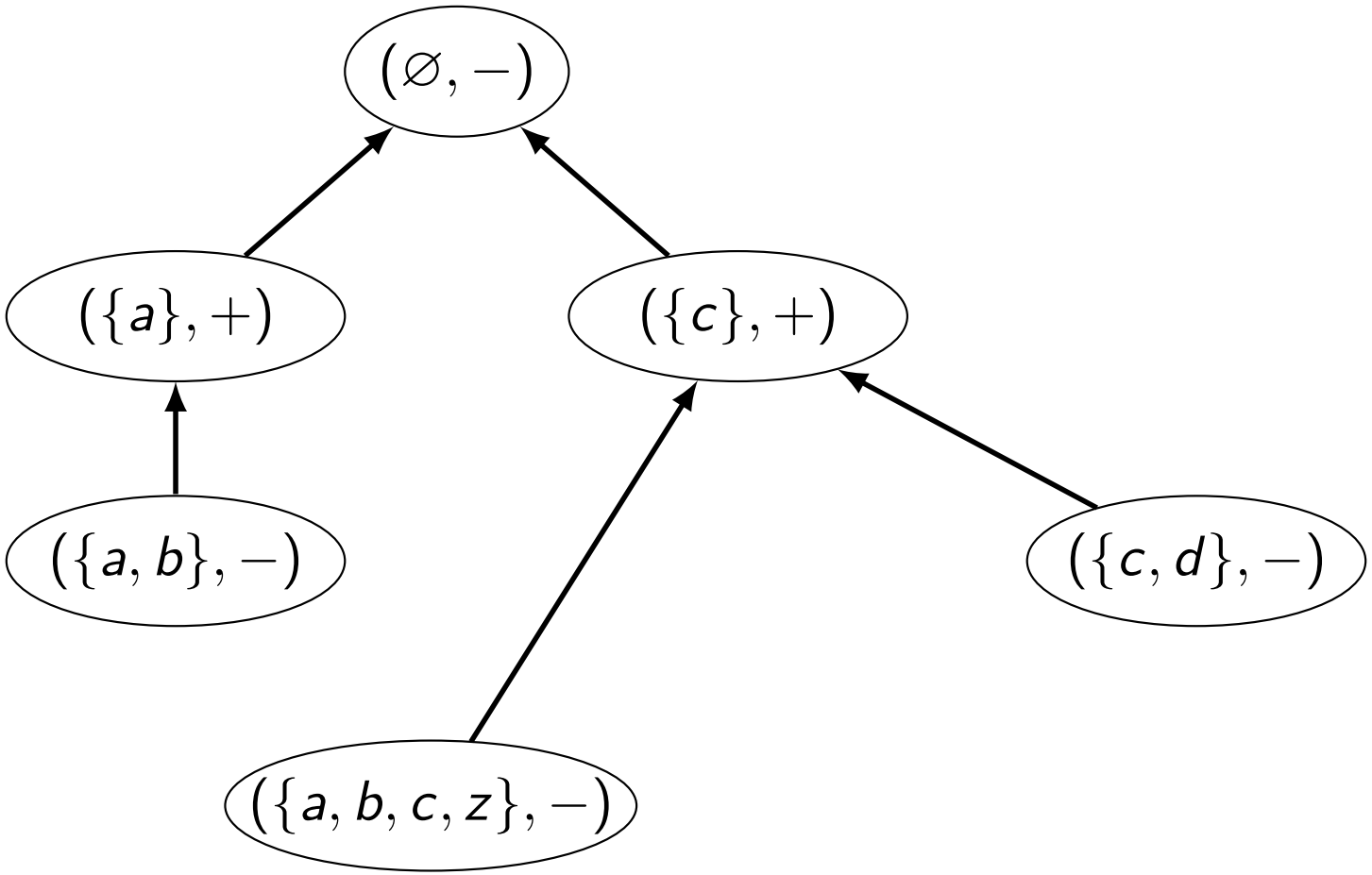


Illustration of the algorithm



Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Precedential constraint

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Precedential constraint

- given a case with $+$ factors and $-$ factors and outcome (say) $+$,

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Precedential constraint

- given a case with $+$ factors and $-$ factors and outcome (say) $+$,
- any case with all of $+$ factors and only some of the $-$ factors...

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Precedential constraint

- given a case with $+$ factors and $-$ factors and outcome (say) $+$,
- any case with all of $+$ factors and only some of the $-$ factors...
- ...should also have outcome $+$

Case study

- Dataset: 32 publicly available cases of the US Trade Secrets domain
- Cases are described as factors (features **for** and **against**)

Precedential constraint

- given a case with $+$ factors and $-$ factors and outcome (say) $+$,
- any case with all of $+$ factors and only some of the $-$ factors...
- ...should also have outcome $+$
- modelled as additional cases by subsets of the "losing" factors

Case study

- We compare $AA\text{-}CBR_{\succeq}$ with $cAA\text{-}CBR_{\succeq}$

Case study

- We compare $AA\text{-}CBR_{\succeq}$ with $cAA\text{-}CBR_{\succeq}$
- Results: No difference! No incoherence or violation of cautious monotonicity in this (small) dataset

Case study

- We compare $AA\text{-}CBR_{\succeq}$ with $cAA\text{-}CBR_{\succeq}$
- Results: No difference! No incoherence or violation of cautious monotonicity in this (small) dataset
- Should we expect case-based reasoning in law to be typically cautiously monotonic?

Discussions

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions
  - Past decisions are taken into account for future cases

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions
  - Past decisions are taken into account for future cases
- e.g. a court

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions
  - Past decisions are taken into account for future cases
- e.g. a court
- If not cautious monotonic, one querying the decision-maker could manipulate it:

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions
  - Past decisions are taken into account for future cases
- e.g. a court
- If not cautious monotonic, one querying the decision-maker could manipulate it:
  - for a pair of cases a, b which violate CM

Manipulation of a legal decision-maker

- Suppose a decision-maker that keeps some coherence to past decisions
  - Past decisions are taken into account for future cases
- e.g. a court
- If not cautious monotonic, one querying the decision-maker could manipulate it:
  - for a pair of cases a, b which violate CM
  - decide whether to present b before a or after it depending on desired outcome

Non-monotonicity properties (more)

1. *completeness*: either $K \vdash a$ or $K \vdash \neg a$.

Non-monotonicity properties (more)

1. *completeness*: either $K \vdash a$ or $K \vdash \neg a$.
2. *cautious monotonicity*: $K \vdash a$ and $K \vdash b$ imply that $K \cup \{a\} \vdash b$

Non-monotonicity properties (more)

1. *completeness*: either $K \vdash a$ or $K \vdash \neg a$.
2. *cautious monotonicity*: $K \vdash a$ and $K \vdash b$ imply that $K \cup \{a\} \vdash b$
3. *cut*: $K \vdash a$ and $K \cup \{a\} \vdash b$ imply that $K \vdash b$

Non-monotonicity properties (more)

1. *completeness*: either $K \vdash a$ or $K \vdash \neg a$.
2. *cautious monotonicity*: $K \vdash a$ and $K \vdash b$ imply that $K \cup \{a\} \vdash b$
3. *cut*: $K \vdash a$ and $K \cup \{a\} \vdash b$ imply that $K \vdash b$
4. *cumulativity*: $\vdash$ is both cautiously monotonic and satisfies cut

Non-monotonicity properties (more)

1. *completeness*: either $K \vdash a$ or $K \vdash \neg a$.
2. *cautious monotonicity*: $K \vdash a$ and $K \vdash b$ imply that $K \cup \{a\} \vdash b$
3. *cut*: $K \vdash a$ and $K \cup \{a\} \vdash b$ imply that $K \vdash b$
4. *cumulativity*: $\vdash$ is both cautiously monotonic and satisfies cut
5. *rational monotonicity*: $K \vdash a$ and $K \nvdash \neg b$ imply that $K \cup \{b\} \vdash a$;

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,
define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;

Revisiting the definition

Definition

Given a set of inputs X, a set of binary outputs Y, and a classifier $\mathbb{C}$: $2^{(X \times Y)} \times X \rightarrow Y$,

define $\vdash_{\mathbb{C}}$ as an *inference relation* such that

- $D \vdash_{\mathbb{C}} (x, y)$, iff $\mathbb{C}(D, x) = y$;
- $D \vdash_{\mathbb{C}} \neg(x, y)$, iff there is a y' such that $\mathbb{C}(D, x) = y'$ and $y' \neq y$.

Classifiers as reasoning (properties)

Theorem

Classifiers as reasoning (properties)

Theorem

1. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it satisfies cut.*

Classifiers as reasoning (properties)

Theorem

1. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it satisfies cut.*
2. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it is cumulative.*

Classifiers as reasoning (properties)

Theorem

1. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it satisfies cut.*
2. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it is cumulative.*
3. $\vdash_{\mathbb{C}}$ *is cautiously monotonic iff it satisfies rational monotonicity.*

Alternative?

- Instead of any output of the classifier, only "confident" ones (however this is defined)

Alternative?

- Instead of any output of the classifier, only "confident" ones (however this is defined)
- Completeness is not necessary!

Counter-intuitive behaviour?

User x Machine interaction

Counter-intuitive behaviour?

User x Machine interaction

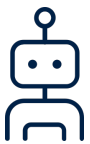
Counter-intuitive behaviour?

User x Machine interaction

-  : Show me some examples in which you are confident.

Counter-intuitive behaviour?

User x Machine interaction

-  : Show me some examples in which you are confident.
-  : I am confident that  is a cat and  is a tiger.

Counter-intuitive behaviour?

User x Machine interaction

- : Show me some examples in which you are confident.
- : I am confident that  is a cat and  is a tiger.
- :  is indeed a tiger.

Counter-intuitive behaviour?

User x Machine interaction

- : Show me some examples in which you are confident.
- : I am confident that  is a cat and  is a tiger.
- :  is indeed a tiger.
- : Thanks! I am not confident that  is a cat anymore...

Open questions

- Limited interplay between machine learning and NMR

Open questions

- Limited interplay between machine learning and NMR
- Can one go further in analysing classifiers as non-monotonic systems?

Open questions

- Limited interplay between machine learning and NMR
- Can one go further in analysing classifiers as non-monotonic systems?
- Does satisfying NMR properties impact the explanation? Could we measure this?

Conclusion

Conclusion

- We can study classifiers as reasoners (under some assumptions)

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons
  - as a classifier

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons
  - as a classifier
  - as a surrogate model for explanation

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons
  - as a classifier
  - as a surrogate model for explanation
  - a larger case study on legal data

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons
  - as a classifier
  - as a surrogate model for explanation
  - a larger case study on legal data
- Applying $cAA\text{-}CBR_{\succeq}$ to more types of data; learning the partial order

Conclusion

- We can study classifiers as reasoners (under some assumptions)
- Some are cautiously monotonic and others are not
- A cautiously monotonic version can be defined for *AA-CBR*

Future work

- Empirical comparisons
  - as a classifier
  - as a surrogate model for explanation
  - a larger case study on legal data
- Applying $cAA\text{-}CBR_{\succeq}$ to more types of data; learning the partial order
- Impacts of cautious monotonicity in interactions with users

Thank you!