The Relevance of Formal Logics for Cognitive Logics, and Vice Versa

Gabriele Kern-Isberner
TU Dortmund, Germany

joint work with Marco Ragni, Christian Eichhorn, Lars-Phillip Spiegel

Invited Talk at FoIKS 2022, Helsinki, June 2022

Cognitive Logics

What are Cognitive Logics?

- Cognitive Logics are formal, logic-based approaches to reasoning that are able to model human reasoning behaviour even if this is in conflict with (classical) logical standards.
- NB: In cognitive logics, not the logic is the norm, but the human is the norm.

Why are cognitive logics relevant?

- Smart devices and AI systems have become ubiquitous, all people have to deal with them in many contexts.
- Cognitive logics can ensure that machine reasoning aligns smoothly with human reasoning, and can prevent situations where AI systems act wrongly or even in a disastrous way because their world model is in conflict with the user's world model due to logical limits.

Contributions and topics of this talk

This talk

- shows famous benchmark examples (paradoxes) that challenge classical logical and probabilistic reasoning;
- aims to sensitize for the general importance and usefulness of cognitive logics in AI, and for crucial details in correct modelling;
- presents approaches to cognitive logics that are able to resolve paradoxes.

Disclaimer: This talk is not against logic and probabilities (just to the contrary), but points out crucial details and techniques regarding

- how to choose the right logical framework,
- how to use logics and probabilities,
- in order to understand and model human reasoning adequately.

(These are just first steps.)

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Observation 1: The Wason Selection Task [Wason 1968]

A	2	D	7
✓	✗	✗	✓

- Given:
  - Four cards with a letter on one and a number on the other side
  - A rule to check: *If there is a vowel on one side then there is an even number on the other side of the card*
- Decide:
  - Exactly which cards to turn in order to check that the rule holds?

Observation 2: Probabilities [Tversky, Kahnemann 1983]

Linda is 31 years old, single, outspoken and very intelligent. As a student she concerned herself thoroughly with subjects of discrimination and social justice and participated in protest against nuclear energy.

Rank the following statements by their probabilities.

- Linda works as a bank teller.
- Linda works as a bank teller and is an active feminist.

- Result: More than 80% judge Linda works as a bank teller and is an active feminist to be more likely than Linda works as a bank teller.
- BUT: $P(a \wedge b) \leqslant P(a)$ or $P(b)$
- Hence, most answer falsely from the perspective of probability!

The Suppression Task [Byrne 1989]

- *If she has an essay to write, she will study late in the library.*
- *If the library is open, she will study late in the library.*
- *She has an essay to write.*

95% of all subjects conclude (modus ponens): Only 38% of all subjects conclude:

- She will study late in the library.

A logic is called non-monotonic if the set of (logical) conclusions from a knowledge base is not necessarily preserved when new information is added to the knowledge base.

$$bird \mid\!\sim fly, bird \wedge penguin \mid\!\sim \neg fly$$

- Commonsense reasoning is usually non-monotonic.

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Basics of propositional logic

$\mathcal{L} = \mathcal{L}(\Sigma)$	propositional language $\mathcal{L}$ over a set of atoms Σ
$\neg, \wedge, \vee$	junctors for negation, conjunction, disjunction
$A \Rightarrow B$	$\equiv \neg A \vee B$ material implication
Ω	set of interpretations/models/possible worlds over Σ
$\omega \models A$	ω is a model of $A (\in \mathcal{L})$
$\mathit{Mod}(A)$	set of models of A
$A \models B$	iff $\mathit{Mod}(A) \subseteq \mathit{Mod}(B)$ classical deduction
$\mathit{Cn}(A)$	$= \{B \in \mathcal{L} \mid A \models B\}$ classical consequence operator

Basic strategies of (nonmonotonic) commonsense reasoning

Like in classical logic, and although Modus Ponens is invalid in general,

RULES

are the main carriers of nonmonotonic inference. However, syntax and/or semantics of rules are different from implications in classical logic.

Basically, two types of rules are used:

- Rules with default assumptions: Reiter's default logic, answer set programming, weak completion semantics;
- Defeasible rules: Conditional reasoning, Poole's default logic

Defeasible rules and conditionals

Defeasible rules establish an uncertain, defeasible connection between antecedent A and consequent B of a rule and can be (logically) implemented by conditionals

$(B|A)$ – "If A then (usually, probably, plausibly ...) B"

- Conditionals encode semantical relationships (plausible inferences) between the antecedent A and the consequent B.
- Conditionals implement nonmonotonic inferences via "$(B|A)$ is accepted iff $A \mid\!\sim B$ holds".
- Conditionals occur in different shapes in many approaches (e.g., as conditional probabilities in Bayesian approaches),
- Conditionals seem to be similar to classical (material) implications "If A then (definitely) B", but are substantially different!

  Indeed, many fallacies observed when applying classical logic to uncertain domains are caused by mixing up implications and conditionals!

Conditionals and implications – example

Christmas on the northern hemisphere

- If Christmas were in summer, there would be no snow at Christmas.
  plausible, approved
- If Christmas were in summer, there would be no Christmas gifts.
  strange, why?
- If Christmas were in summer, there would be no gravitation.
  downright nonsense!

All these statements are logically true, when understood as (material) implications (because Christmas is in winter on the northern hemisphere, hence the antecedent is false!).
However, understood as conditionals, crucial differences appear!

What makes conditionals so special?

A conditional $(B|A)$ focusses on cases where the premise A is fulfilled but does not say anything about cases when A does not hold – conditionals go beyond classical logic, as they are three-valued entities.

A conditional leaves more semantical room for modelling acceptance in case its confirmation $A \wedge B$ is more plausible than its refutation $A \wedge \neg B$.

Conditional acceptance and preferential entailment $\mid\!\sim_{\prec}$ [Makinson 89]

Let $\prec$ be a (well-behaved) relation on models (expressing , e.g., plausibility via a total preorder $\preceq$).

$$(B|A) \text{ is accepted} \quad \text{iff} \quad A \mid\!\sim_{\prec} B$$

iff in the most plausible models of A (wrt $\prec$), B holds also.

$\mid\!\sim_{\prec}$ is a semantic-based nonmonotonic inference relation that is encoded by conditionals on the syntax level.

Logics of conditionals and nonmonotonic reasoning

The basic trick is to leave the narrow frames of 2-valued logics and enter into (at least) 3-valued conditional logics.

- Conditional logics have a long tradition in logics and philosphy, going back to the Old Greeks, with lots of formal systems and axiomatizations.
- There are also lots of formal properties and axiomatic systems for nonmonotonic inference relations $\mid\!\sim$.
- Well-behaved relations on possible worlds expressing (e.g.) plausibility provide semantics to both conditionals and nonmonotonic inference relations.
- Note that plausibility relations are similar to, but significantly weaker than probabilities.

Ranking functions and conditionals

Ordinal conditional functions (OCF, ranking functions[1]) [Spohn 1988]

$\kappa : \Omega \to \mathbb{N}(\cup\{\infty\})$ (Ω set of possible worlds, $\kappa^{-1}(0) \neq \emptyset$)

$\kappa(\omega_1) < \kappa(\omega_2)$	ω_1 is more plausible than ω_2
$\kappa(\omega) = 0$	ω is maximally plausible
$\kappa(A)$	$:= \min\{\kappa(\omega) \mid \omega \models A\}$
Bel (κ)	$:= \{A \mid \kappa(\neg A) > 0\}$

Validating conditionals

$$\kappa \models (B|A) \text{ iff } \kappa(AB) < \kappa(A\overline{B})$$

κ accepts a conditional $(B|A)$ iff its verification AB is more plausible than its falsification $A\overline{B}$.

[1] Rankings can be understood as qualitative abstractions of probabilities

Ranking functions – example

Example (ranked flyers)

$\kappa(\omega) = 4$	$p\overline{b}\,f$
$\kappa(\omega) = 2$	$pbf \quad p\overline{b}\,\overline{f}$
$\kappa(\omega) = 1$	$pb\overline{f} \quad \overline{p}\,b\overline{f}$
$\kappa(\omega) = 0$	$\overline{p}\,bf \quad \overline{p}\,\overline{b}\,f \quad \overline{p}\,\overline{b}\,\overline{f}$

$$\mathit{Bel}\,(\kappa) = Cn(\overline{p}\,(f \vee \overline{b}\,\overline{f}\,)$$

$$\kappa(bf) = 0 < 1 = \kappa(b\overline{f}\,) \Longrightarrow \kappa \models (f|b),$$

$$\text{but } \kappa(p\overline{f}\,) = 1 < 2 = \kappa(pf) \Longrightarrow \kappa \models (\overline{f}\,|p)$$

$$(\text{also } \kappa \models (b|p))$$

Ranking functions make conditional and nonmonotonic reasoning particularly easy!

A general framework for cognitive logics

- Conditionals provide syntactical entities to encode meaningful relationships between propositions – going beyond classical logics, but don't (necessarily) need numbers.
- 3-valued logics, as well as plausibilistic relations and ranking functions on possible worlds explore the semantic field between classical logics and probabilities.

Together, conditionals and plausibilistic relations provide an intuitive and easily accessible, yet formal and powerful modelling tool for commonsense reasoning that may help to explore human rationality.

Rationality is more than logic, but how can it be "defined" adequately?

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Commonsense inference rules

From a conditional statement "If A then B",
Modus ponens and Modus tollens are logically valid inference rules:

(MP) From A, infer B

(MT) From $\neg B$, infer $\neg A$

However, people also use other inference rules in commonsense reasoning:

(AC) Affirmation of the Consequent: From B, infer A

(DA) Denial of the Antecedent: From $\neg A$, infer $\neg B$

Both (AC) and (DA) are logically invalid, but are they irrational?

Logical invalidity in the Suppression Task

In the Suppression Task [Byrne 1989], participants had to draw inferences with respect to the arguments

Suppression Task (plus Additional Argument)

"If Lisa has an essay to write, she will study late in the library."
"If the library stays open, she will study late in the library."
"Lisa has an essay to write."

Here, the majority of the participants (students without tuition in logic)

- did not apply MP (38%) nor MT (33%),
- but did apply AC (63%) and DA (54%).

This inference behaviour (no MP nor MT, but AC and DA) was deemed to be completely irrational, i.e., rationality is usually assessed according to classical logic. However, obviously, the "irrational" inference behaviour was triggered by the additional information
$\rightarrow$ Context of reasoning tasks must be taken into account!

Sensitivity of inference behavior

Different wordings and slightly different information can change human inferences drastically –

- What do people understand from the reasoning task?
  $\rightarrow$ implicit assumptions, background knowledge
- Additional information may suggest implicitly exceptions, alternatives, strengthening etc
  $\rightarrow$ nonmonotonic reasoning
- "If ... then"-statements often are not strict
  $\rightarrow$ conditionals

Rationality needs context!

(My) Crucial hypothesis for cognitive logics

Rationality of statements can be assessed only if context is taken into account!

My most favourite example – rational or irrational???

At BRAON 2017, one of the (famous) *Madeira Workshops on Belief Revision, Argumentation, Ontologies, and Norms* locally and generally organized by *Eduardo Fermé*, Eduardo introduced himself presenting some slides and saying:

> *I have a picture of myself on my first slide because there are no cangaroos on Madeira.*

Everyone understood, and laughed . . .

Context: Dongmo Zhang from Australia introduced himself immediately before, and instead of a picture of himself, he had a picture of a cute cangaroo on his slide.

Dongmo Zhang

WESTERN SYDNEY UNIVERSITY

Affiliation: School of Computing, Engineering and Mathematics, Western Sydney University, Australia

Area of expertise: Belief revision, reasoning about action, multi-agent systems, knowledge representation and reasoning

A picture (optional):

Eduardo Fermé
University of Madeira

Belief Revision
KRR

Conditional theory of rational reasoning

People deviate so systematically from (MP) and (MT) and apply so frequently (AC) and (DA) that cognitive logics have to find a model for this. Obviously, classical logic is not cognitively adequate for cognitive logics.
Instead, we suggest:

[Eichhorn, Kern-Isberner & Ragni AAAI-2018]

- Using a (nonmonotonic) conditional logic as normative theory to evaluate human inferences
- Result: (basically) all irrationality can be eliminated!

The aim of that paper was to devise a novel (descriptive and/or normative) theory of a generic rational reasoner that emerges from a group of people.

Generic rational reasoner

When exploring rationality, we encounter the following

Dilemma of assessing rationality

Thesis: Overall, humans reason and behave rational in the sense that they are successful survivors. However,

- not all individuals reason rationally all the times – even worse, maybe each individual reasons and behaves irrationally at least from time to time ...
- no individual reasoner can be a norm for their own rational reasoning.

Possible solution of this dilemma: Observe groups of people and try to extract a generic reasoning behaviour by

- aggregating reasoning behaviour over the group, and
- finding a formal theory to model this generic rational reasoner

Inference patterns

Basic idea: Consider all four inference rules (MP, MT, AC, DA) together in a 4-tuple to model coherent generic inference behaviour:

Definition

An inference pattern ϱ is a 4-tuple that for each inference rule MP, MT, AC, and DA indicates whether the rule is used (positive rule, e.g., MP) or not used (negated rule, e.g., $\neg\mathrm{MP}$) in an inference scenario.

Inference patterns – examples

- Suppression Task: (MP (38%), MT (33%), AC (63%), DA (54%)) yields the inference pattern $\varrho_{Supp} = (\neg\mathrm{MP}, \neg\mathrm{MT}, \mathrm{AC}, \mathrm{DA})$.
- Counterfactuals [Thompson & Byrne 2002]: "If the car had been out of $\underline{g}$as, then it would have $\underline{s}$talled."
  Overall inferences: (MP (78%), MT (85%), AC (41%), DA (50%)), yielding the inference pattern $\varrho_{Counter} = (\mathrm{MP}, \mathrm{MT}, \neg\mathrm{AC}, \mathrm{DA})$.
  Since DA was observed with exactly half of the participants, one might also argue for the inference pattern $\varrho^{alt}_{Counter} = (\mathrm{MP}, \mathrm{MT}, \neg\mathrm{AC}, \neg\mathrm{DA})$.

$\rightarrow$ Basics of nonmonotonic logics and conditionals

Remember the basics of nonomotonic logics and plausibility:

Total preorders $\preccurlyeq$ on possible worlds Ω expressing plausibility are of crucial importance both for nonmonotonic reasoning and conditionals:

$\omega_1 \preccurlyeq \omega_2$	ω_1 is deemed at least as plausible as ω_2
$A \preccurlyeq B$	iff minimal models of A are at least as plausible as all models of B
$A \mid\!\sim_{\preccurlyeq} B$	iff $AB \prec A\overline{B}$ – in the context of A, B is more plausible than $\overline{B}$; iff the conditional $(B\vert A)$ is accepted
Ψ	epistemic state equipped with a total preorder $\preccurlyeq_\Psi$ (you might think of Ψ as a ranking function)
Bel (Ψ)	$= Th(\min(\preccurlyeq_\Psi))$ most plausible beliefs in Ψ

Inference patterns → conditionals → plaus. constraints

With each inference rule, we associate a nonmonotonic inference relation resp. a conditional which implies a plausibility contraint:

Rule	Inference	Conditional	Plaus. constraint
MP	$A \mid\!\sim B$	$(B\|A)$	$A\,B \prec A\,\overline{B}$
MT	$\overline{B} \mid\!\sim \overline{A}$	$(\overline{A}\|\overline{B})$	$\overline{A}\,\overline{B} \prec A\,\overline{B}$
AC	$B \mid\!\sim A$	$(A\|B)$	$AB \prec \overline{A}\,B$
DA	$\overline{A} \mid\!\sim \overline{B}$	$(\overline{B}\|\overline{A})$	$\overline{A}\,\overline{B} \prec \overline{A}\,B$

Inference patterns → conditionals → plaus. constraints (cont'd)

Negated inference rules (e.g., ¬MP) are implemented simply by negating the constraint (e.g., $A\,\overline{B} \not\preceq A\,B$), being implemented by weak conditionals:

Definition

A weak conditional $(\!|B|A|\!)$ is accepted if $AB \preceq A\overline{B}$.

¬Rule	Weak Conditional	Plaus. constraint
¬MP	$(\!\|\overline{B}\|A\|\!)$	$A\,\overline{B} \preceq A\,B$
¬MT	$(\!\|A\|\overline{B}\|\!)$	$A\,\overline{B} \preceq \overline{A}\,\overline{B}$
¬AC	$(\!\|\overline{A}\|B\|\!)$	$\overline{A}B \preceq AB$
¬DA	$(\!\|B\|\overline{A}\|\!)$	$\overline{A}B \preceq \overline{A}\,\overline{B}$

Rationality in terms of nonmonotonic/conditional logic

reasoning pattern ϱ $\longrightarrow$ set of plausibility constraints $\mathcal{C}(\varrho)$
$\longrightarrow$ set of (weak) conditionals Δ_ϱ

$\mathcal{C}(\varrho)$ is satisfiable

- iff there is a plausibility relation (i.e., a (total) preorder) $\preceq$ on possible worlds that satisfies all constraints in $\mathcal{C}(\varrho)$
- iff the associated set of (weak) conditionals Δ_ϱ is consistent

$\longrightarrow$ novel definition of rationality in terms of conditional consistency:

Definition

- An inference pattern $\varrho \in \mathcal{R}$ is called rational iff there is a plausibility relation $\preceq$ that satisfies $\mathcal{C}(\varrho)$.
- Otherwise, the inference pattern is irrational.

. . . and irrationality disappears

Only 2 out of 16 patterns are irrational:

- $(\mathrm{MP}, \neg\mathrm{MT}, \neg\mathrm{AC}, \mathrm{DA})$: $\overline{A}\,\overline{B} \prec \overline{A}B \preccurlyeq AB \prec A\overline{B} \preccurlyeq \overline{A}\,\overline{B}$ – unsatsifiable
- $(\neg\mathrm{MP}, \mathrm{MT}, \mathrm{AC}, \neg\mathrm{DA})$: $\overline{A}\,\overline{B} \prec A\overline{B} \preccurlyeq AB \prec \overline{A}B \preccurlyeq \overline{A}\,\overline{B}$ – unsatisfiable

How often do they appear in practical reasoning tasks?

In over 60 empirical studies investigated so far, hardly any irrational patterns could be found (less than 2%).

(more on this later)

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Implicit assumptions and background knowledge

With the help of conditionals and nonmonotonic logics/plausibility logics as a normative theory, we are able to model human reasoning much better. Using this framework, we can also deal with the following two issues:

- What implicit assumptions are used?
  How do people understand the task?
  $\rightarrow$ beliefs;
- What (conditional) beliefs are people actually using for the task?
  $\rightarrow$ elaborating on sets of conditionals giving rise to the total preorders compatible with the respective inference pattern
  $\rightarrow$ reverse engineering human reasoning

Example Suppression Task: beliefs

$$\varrho_{Supp} = (\neg\mathrm{MP}, \neg\mathrm{MT}, \mathrm{AC}, \mathrm{DA}) \rightarrow \begin{array}{ccc} A\overline{B} & \preceq & AB \\ A\overline{B} & \preceq & \overline{A}\,\overline{B} \\ AB & \prec & \overline{A}B \\ \overline{A}\,\overline{B} & \prec & \overline{A}B \end{array}$$

$$\rightarrow \qquad A\overline{B} \preceq \left\{ \begin{array}{c} AB \\ \overline{A}\,\overline{B} \end{array} \right\} \prec \overline{A}B$$

Choosing minimal, i.e., most conservative total preorder $\preceq^{min}_{Supp}$:

$$A\overline{B} \approx^{min}_{Supp} AB \approx^{min}_{Supp} \overline{A}\,\overline{B} \prec^{min}_{Supp} \overline{A}B$$

Example Suppression Task: beliefs (cont'd)

From this, we compute the beliefs

$$Bel(\preceq_{Supp}^{min}) = Cn(A\overline{B} \vee AB \vee \overline{A}\,\overline{B}) = Cn(B \Rightarrow A).$$

Here, we have $A = e$ (essay writing), $B = l$ (studying in the library), hence

$$Bel(\preceq_{Supp}^{min}) = Cn(l \Rightarrow e), \text{not } Cn(e \Rightarrow l)!$$

This explains the rationality of the inference pattern:

Participants might have understood the given conditional information in its inverse form, and hence applied AC and DA which, in fact, amount to MP and MT for the inverse conditional.

Example counterfactuals: beliefs

Constraints for the inference pattern $\varrho_{Counter} = (\mathrm{MP}, \mathrm{MT}, \neg\mathrm{AC}, \mathrm{DA})$:

$$\begin{array}{ll} & \{AB \prec A\overline{B}, \overline{A}\,\overline{B} \prec A\overline{B}, \overline{A}B \preccurlyeq AB, \overline{A}\,\overline{B} \prec \overline{A}B\} \\ \equiv & \overline{A}\,\overline{B} \prec \overline{A}B \preccurlyeq AB \prec A\overline{B} \end{array}$$

In this example, *Bel*$(\varrho_{Counter}) = Cn(\overline{A}\,\overline{B})$.

$\rightarrow$ Finding: In the counterfactual case, people believe not only that the antecedent is false[2], but also that the consequent is false!

[2] This is usually assumed in the counterfactual case

C-representations [Kern-Isberner 2001]

For reverse engineering human reasoning, we build on an alternative to system Z [Pearl 1990]: $\Delta = \{(B_1|A_1), \ldots, (B_n|A_n)\}$

c-representation of Δ is defined by

$$\kappa_\Delta(\omega) = \sum_{\omega \models A_i \overline{B_i}} \kappa_i^-$$

with parameters $\kappa_1^-, \ldots, \kappa_n^- \in \mathbb{N}_0$ chosen such that

$$\kappa_\Delta \models (B_j|A_j), 1 \leqslant j \leqslant n,$$

holds, i.e.,

$$\kappa_j^- > \min_{\omega \models A_j B_j} \sum_{\substack{i \neq j \\ \omega \models A_i \overline{B_i}}} \kappa_i^- - \min_{\omega \models A_j \overline{B_j}} \sum_{\substack{i \neq j \\ \omega \models A_i \overline{B_i}}} \kappa_i^-$$

For weak conditionals, one simply has to use $\geqslant$ instead of $>$.

Background beliefs and reasoning

$\kappa_\Delta(\omega) = \sum_{\omega \models A_i \overline{B_i}} \kappa_i^-$ with parameters $\kappa_1^-, \ldots, \kappa_n^- \in \mathbb{N}_0$ chosen such that

$$\kappa_j^- \overset{>}{\geqslant} \min_{\omega \models A_j B_j} \sum_{\substack{i \neq j \\ \omega \models A_i \overline{B_i}}} \kappa_i^- - \min_{\omega \models A_j \overline{B_j}} \sum_{\substack{i \neq j \\ \omega \models A_i \overline{B_i}}} \kappa_i^-$$

Using c-representations of (weak) conditional belief bases Δ and their parameters κ_i^-, we can further elaborate on the background (conditional) beliefs that people (may) have used for reasoning:

- Each κ_i^- symbolizes the impact of (weak) conditional $(B_i|A_i)$ on reasoning with c-representations;
- this impact has to obey a constraint that reveals the impact of $(B_i|A_i)$ in the interaction with the other conditionals from Δ.

$\rightarrow$ Each κ_i^- whose constraint is covered by other constraints can be eliminated.

Explanation generator

With the algorithm Explanation generator [Eichhorn, Kern-Isberner, Ragni, AAAI 2018] we are able to extract most basic conditionals from inference patterns:

Algo Explanation Generator

Input: Inference pattern $\varrho \in \mathcal{R}$

Output: Knowledge base of (weak) conditionals compatible with ϱ

1. Set up Δ_ϱ with a conditional for each rule in pattern ϱ
2. Set up the system of inequalities for Δ_ϱ and simplify:
   - For each inequality that is implied by the other inequalities, remove the line from the system of inequalities and the respective conditional from Δ_ϱ to obtain a (wrt. set inclusion) minimal explaining knowledge base Δ_ϱ^{expl}.
3. Return the knowledge base Δ_ϱ^{expl}.

Generating belief bases: Examples

Inference pattern	Δ example	*Bel*(Δ)
(MP, MT, AC, DA)	$\{(B\|A), (A\|B)\}$	$Cn(A \Leftrightarrow B)$
$(MP, \neg MT, AC, DA)$	$\{(B\|A), (\|A\|\overline{B}\|), (\overline{B}\|\overline{A})\}$	$Cn(AB)$
$(MP, MT, AC, \neg DA)$	$\{(\overline{A}\|\overline{B}), (A\|B), (\|B\|\overline{A}\|)\}$	$Cn(AB)$
$(MP, \neg MT, AC, \neg DA)$	$\{(B\|A), (\|A\|\overline{B}\|), (A\|B)\}$	$Cn(AB)$
$(MP, MT, \neg AC, \neg DA)$	$\{(B\|A)\}$	$Cn(A \Rightarrow B)$

Inference patterns with a generating conditional knowledge base and most plausible beliefs of their appertaining total preorder

Reverse engineering: Suppression Task

Here we have the inference pattern $\varrho_{Supp} = (\neg\mathrm{MP}, \neg\mathrm{MT}, \mathrm{AC}, \mathrm{DA})$ $\rightarrow \Delta_{Supp} = \{\delta_1 : (\!|\bar{l}|e|\!), \delta_2 : (\!|e|\bar{l}|\!), \delta_3 : (e|l), \delta_4 : (\bar{l}|\bar{e})\}$.

Schema of c-representation:

ω	$\kappa_{\Delta_{Supp}}(\omega)$	ω	$\kappa_{\Delta_{Supp}}(\omega)$
el	κ_1^-	$\bar{e}l$	$\kappa_3^- + \kappa_4^-$
$e\bar{l}$	0	$\bar{e}\bar{l}$	κ_2^-

System of constraints:

$$\kappa_1^- \geqslant \min_{e\bar{l}}\{0\} - \min_{el}\{0\} = 0 \qquad \kappa_3^- > \min_{el}\{\kappa_1^-\} - \min_{\bar{e}l}\{\kappa_4^-\}$$
$$\kappa_2^- \geqslant \min_{e\bar{l}}\{0\} - \min_{\bar{e}\bar{l}}\{0\} = 0 \qquad \kappa_4^- > \min_{\bar{e}\bar{l}}\{\kappa_2^-\} - \min_{\bar{e}l}\{\kappa_3^-\}$$

Reverse engineering: Suppression Task (cont'd)

In the end, the only relevant constraint is

$$\kappa_3^- + \kappa_4^- > \max\{\kappa_1^-, \kappa_2^-\}, \text{i.e., minimally } \kappa_3^- > 0 \text{ or } \kappa_4^- > 0$$

$\rightarrow$ two KBs can explain the inference pattern ϱ_{Supp}:

- $\Delta_{Supp}^{expl} = \{(e|l)\}$
  "If Lisa is in the library, then she (usually) has an essay to write"
- $\Delta'^{expl}_{Supp} = \{(\bar{l}|\overline{e})\}$
  "If Lisa does not have an essay to write, then she (usually) is not in the library"

Again: Participants might have understood the given conditional information in its inverse (contraposed) form, and then $\varrho_{Supp} = (\neg\mathrm{MP}, \neg\mathrm{MT}, \mathrm{AC}, \mathrm{DA})$ appears to be rational.

Reverse engineering: counterfactuals

$\varrho_{counter} = (\mathrm{MP}, \mathrm{MT}, \neg\mathrm{AC}, \mathrm{DA})$
$\rightarrow \Delta_{counter} = \{\delta_1 : (s|g), \delta_2 : (\overline{g}|\overline{s}), \delta_3 : (\!(\overline{g}|s)\!), \delta_4 : (\overline{s}|\overline{g})\}$

Constraints:

$$\kappa_1^- + \kappa_2^- > \kappa_3^- \geqslant 0,\ \kappa_1^- + \kappa_2^- > 0,\ \kappa_3^- \geqslant \kappa_4^-,\ \kappa_4^- > 0$$

$\rightarrow \delta_2$ and κ_2^- can be eliminated
$\rightarrow \Delta_{counter}^{expl} = \{\delta_1 : (s|g), \delta_3 : (\!(\overline{g}|s)\!), \delta_4 : (\overline{s}|\overline{g})\}$:

- δ_1 "If the car is out of gas, then (usually) it stalls."
- δ_3 "If the car stalls, then it might not be out of gas." ($\rightarrow$ other possible, more plausible causes)
- δ_4 "If the car is not out of gas, then (usually) it will not stall."($\rightarrow$ possible, but not very plausible cause because drivers usually take care of gas (implicit assumption))

Reverse engineering: counterfactuals (alternative)

Let's look at the alternative inference pattern
$\varrho_{\text{counter-alt}} = (\mathrm{MP}, \mathrm{MT}, \neg\mathrm{AC}, \neg\mathrm{DA})$

$\rightarrow \Delta_{\text{counter-alt}} = \{\delta_1 : (s|g), \delta_2 : (\overline{g}|\overline{s}), \delta_3 : (\!|\overline{g}|s|\!), \delta_4' : (\!|s|\overline{g}|\!)\}$

$\rightarrow \Delta_{counter-alt}^{expl} = \{(s|g)\}$ and $\Delta'^{expl}_{counter-alt} = \{(\overline{g}|\overline{s})\}$, and

$$Bel(\Delta^{expl}_{\text{counter-alt}}) = Cn(g \Rightarrow s)$$

$\rightarrow$ classical-logical reasoner

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Inference patterns in empirical studies

Focus on 22 studies with 35 experiments [Spiegel, BSc Thesis TU Dortmund 2018] –

Only six inference patterns were ever drawn at a frequency of more than 5%. The proportion of irrational patterns is only 1.1%.

Most frequent inference patterns:

(MP, MT, AC, DA)	perc.	meaning
TTTT	33.9	"credulous reasoner"
TTFF	23.6	"the logical reasoner"
TTTF	12.1	"partly logical reasoner"
TFTF	9.2	"reasoner rejecting negations"
TFTT	5.7	"bold reasoner" (all but MT)
TFFF	5.7	"basic reasoner (only MP)

Features of tasks in empirical studies

Wordings, suggestions etc can have a major impact on human reasoning (formalized by inference patterns).

[Spiegel, GKI, Ragni, PRICAI 2019] investigated empirical studies and classified reasoning behavior ($\equiv$ inference pattern) by features that reasoning tasks may have:

Features	
age group	task type
negation	alternatives
abstraction	familiarity
meaning	(counter)factual
strictness	wording

Features and inference patterns: Suppression

Argument Type	MP	MT	AC	DA	Inference Pattern
Simple	96	92	71	46	$(MP, MT, AC, \neg DA)$
Alternative	96	96	13	4	$(MP, MT, \neg AC, \neg DA)$
Additional condition	38	33	54	63	$(\neg MP, \neg MT, AC, DA)$

Evaluation and inference patterns for the Suppression Task

Features and inference patterns: Negation

Argument Type	MP	MT	AC	DA	Inference Pattern
If p, then q	95	60	60	35	$(MP, MT, AC, \neg DA)$
If p, then not q	100	75	40	20	$(MP, MT, \neg AC, \neg DA)$
If not p, then q	100	50	85	50	(MP, MT, AC, DA)
If not p, then not q	100	35	60	30	$(MP, \neg MT, AC, \neg DA)$

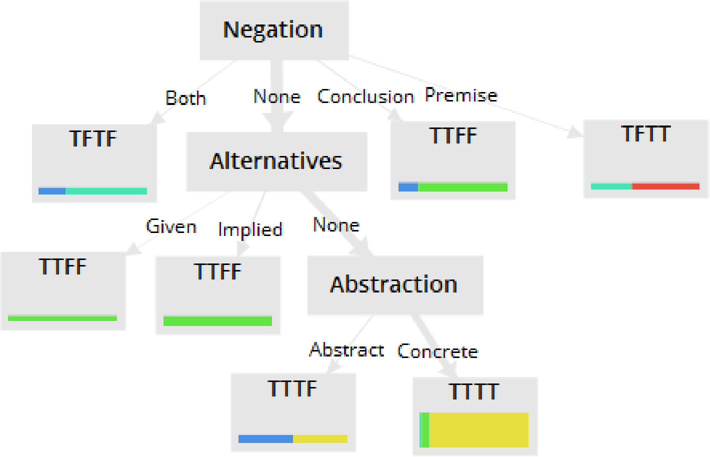
Evaluation and inference patterns for negation

Features and inference patterns: Counterfactuals

Argument Type	MP	MT	AC	DA	Inference Pattern
Normal	80	58	40	20	$(MP, MT, \neg AC, \neg DA)$
Counterfactual	86	81	46	46	$(MP, MT, \neg AC, \neg DA)$
Fict. story	49	45	53	59	$(\neg MP, \neg MT, AC, DA)$

Evaluation and inference patterns for counterfactuals

A small decision tree



Decision tree based on three core features: negation, alternatives, abstraction

Overview of this talk

- Motivation and overview
- Human or logical fallacies?
- A general framework for cognitive logics
- A formal approach to rationality
- Reverse engineering human reasoning
- The effect of features in tasks
- Conclusions

Conclusion and future work

- We presented a novel framework for cognitively adequate logics based on conditionals and qualitative/semi-quantitative relations on possible worlds.
- First approaches to model, explain, and reverse-engineer human reasoning adequately have been presented.
- Well-known paradoxes could be easily resolved, a novel definition of rationality could be established and verified in studies.

Future work:

- Analyze individual reasoning behaviour with inference patterns.
- Take also quantitative endorsements into account [NMR 2021, with Sara Todorovikj].
- Applications to more complex reasoning tasks like syllogisms.

Observation 1: The Wason Selection Task [Wason 1968]

A	2	D	7

A rule: *If a vowel is on one side then an even number is on the other side*

Percentage Humans	Card turned	Response
89%	Vowel (A)	Correct!
62%	Even number (2)	Unnecessary! Really?
25%	Odd number (7)	Correct!
16%	Consonant (D)	Unnecessary!

Explanation: People maybe seeking to verify the conditional instead of the classical implication

Observation 2: Probabilities [Tversky, Kahnemann 1983]

Linda is 31 years old, single, outspoken and very intelligent. As a student she concerned herself thoroughly with subjects of discrimination and social justice and participated in protest against nuclear energy.

Rank the following statements by their probabilities.

- Linda works as a bank teller.
- Linda works as a bank teller and is an active feminist.

- Result: More than 80% judge Linda works as a bank teller and is an active feminist to be more likely than Linda works as a bank teller.
- BUT: $P(A \wedge B) \leqslant P(A)$ or $P(B)$
- My explanation: People are not considering $P(A, B|Linda)$, but the explaining likelihood $P(Linda|A, B)$ (no comparison possible between $P(Linda|A)$ and $P(Linda|A, B)$!).

Observation 2': Probabilities New Linda [Kern-Isberner 2022]

Linda is sitting on a luxurious yacht in the Caribbean Sea, enjoying the sun and sipping on a cocktail that has just brougt to her by the boat steward.

Rank the following statements by their probabilities.

- Linda is receiving social welfare.
- Linda is receiving social welfare and has a rich boy friend.

- BUT: $P(A \wedge B) \leqslant P(A)$ or $P(B)$
- My explanation: People are not considering $P(A, B|Linda)$, but the explaining likelihood $P(Linda|A, B)$ (no comparison possible between $P(Linda|A)$ and $P(Linda|A, B)$!).