Deontic Explanations

Leon van der Torre

University of Luxembourg, Zhejiang University, visiting Toulouse University

Toulouse, December 6, 2022

1.Forrester 1984: scenario of gentle murderer

1.	you should not kill	*O not k*
2.	If you kill, you should do it gently	*if k then O g*
3.	You kill	*k*
4.	So, you should kill gently	*O g*
5.	So, you should kill	*O k*
6.	So, you should not kill and you should kill	*O not k and O k*
7.	So, this is a violation of “ought implies can”	*not (O not k and O k)*

Forrester’s analysis:

- Drop modus ponens (step 4, also called factual detachment in dyadic DL)
- Drop consequential closure (step 5)

1.Forrester 1984: scenario of gentle murderer

1.	You should not kill	*O not k*
2.	If you kill, you should do it gently	*if k then O g*
3.	You kill	*k*
4.	So, you should kill gently	*O g*
5.	So, you should kill	*O k*
6.	So, you should not kill and you should kill	*O not k and O k*
7.	So, this is a violation of “ought implies can”	*not (O not k and O k)*

Leon’s challenge for nonmonotonic modal logic (2022)

- *O not k and O k* is inconsistent
- But *O not k and O k and k* is consistent

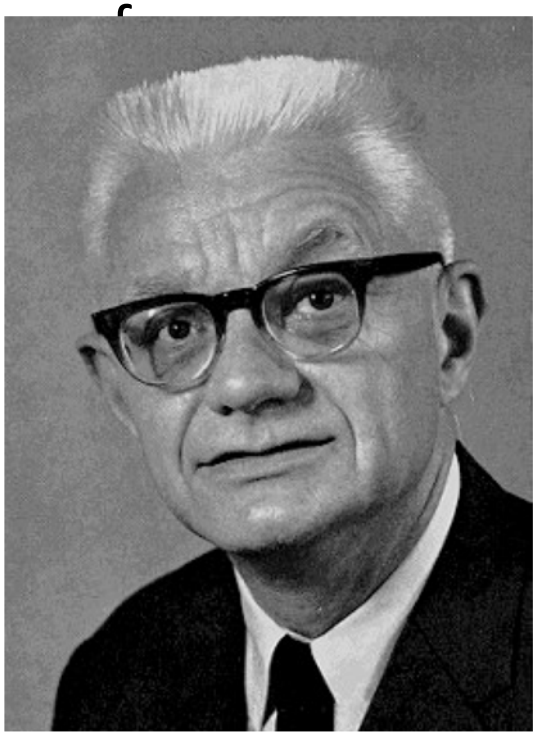
Layout of this talk on deontic explanations

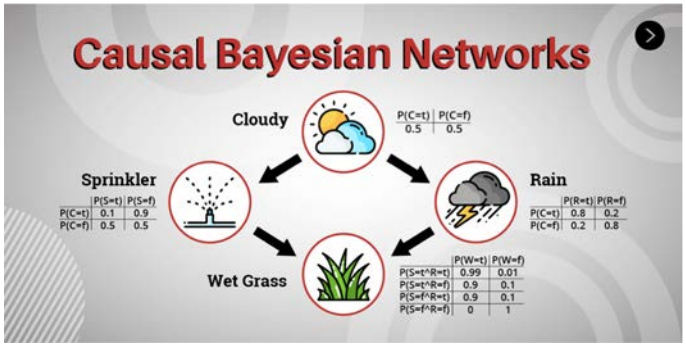
1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanation
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

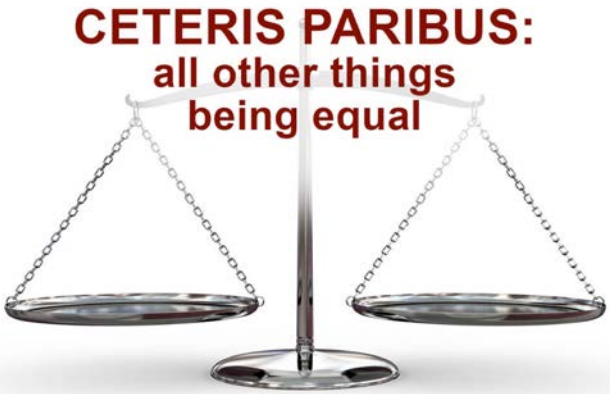
Layout of this talk on deontic explanations

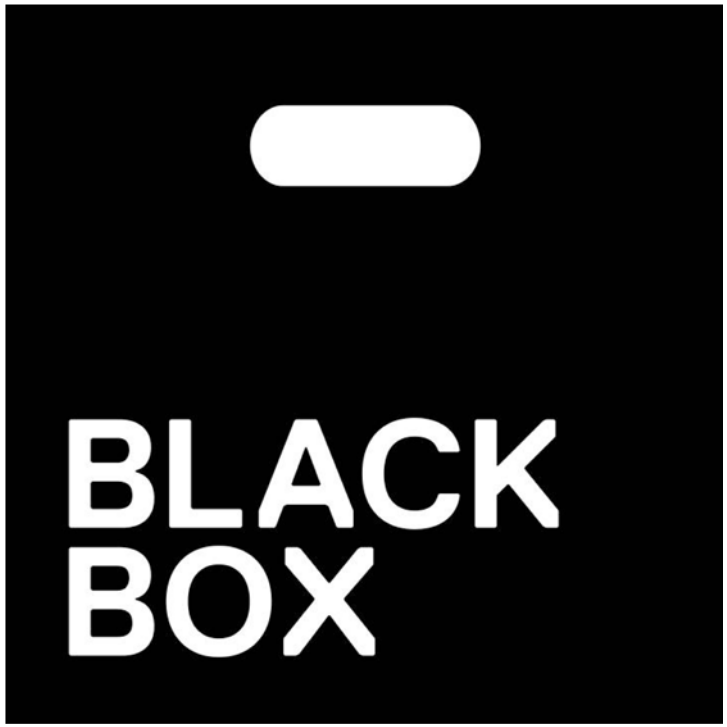
1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanation
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

Scientific discovery to deontic explanation

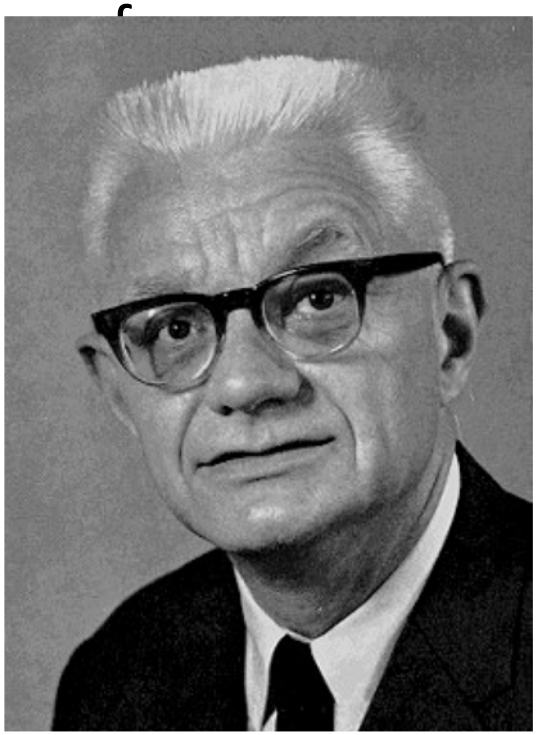


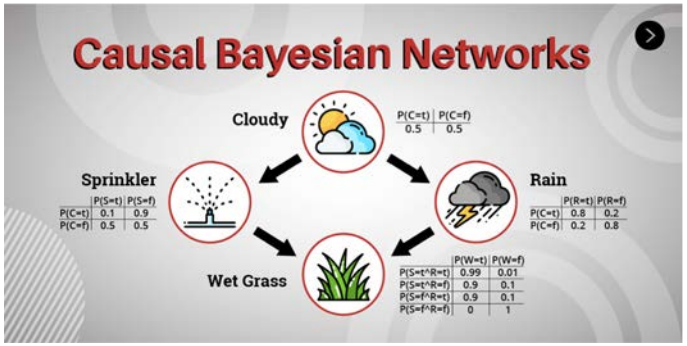


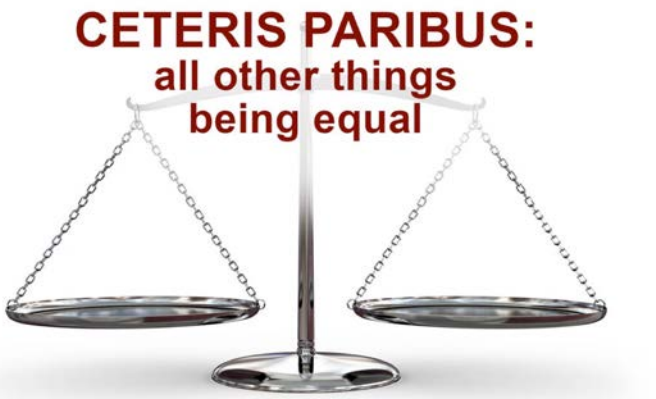


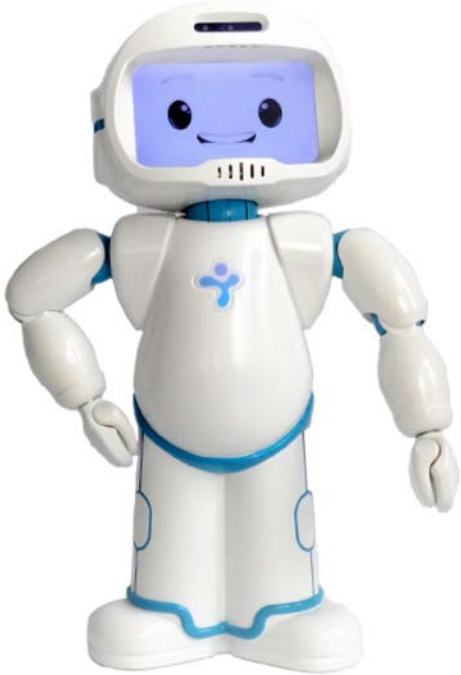


Scientific discovery to deontic explanation

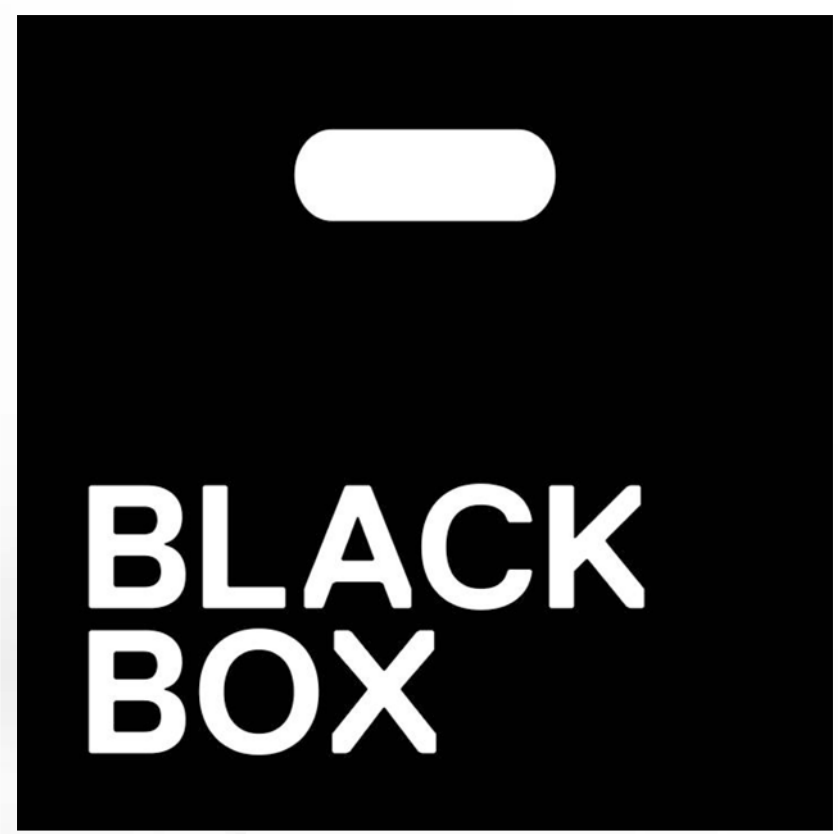


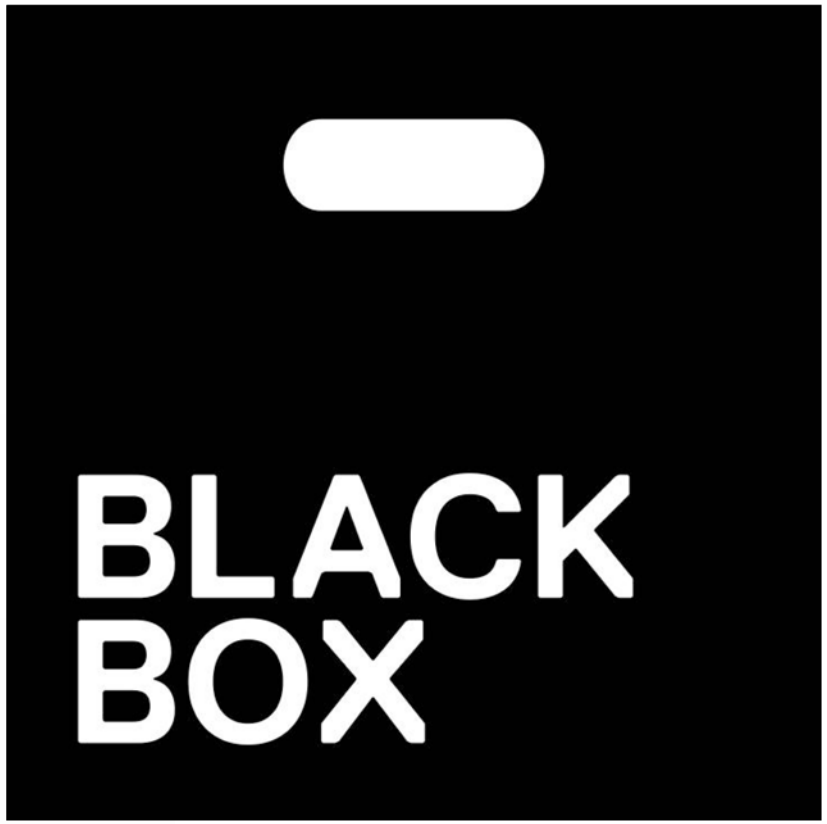






2.Moxie and QT: new generation social robots





2.Moxie and QT: new generation social robots

Firefox File Edit View History Bookmarks Tools Window Help Sat 15:13

Live tv | Netflix | The Huma | Log in to | Log in to | TIUWIEN2020 | Playgroun | ai%art es | The Op | Deontic Logic

https://esch2022.uni.lu/events/the-opening-of-the-living-lab-of-the-aiart-pavilion/ 67%

AI&Art

The Opening of the Living Lab of the AI&ART Pavilion

Fri 11 March 2022
16:00 - 20:00

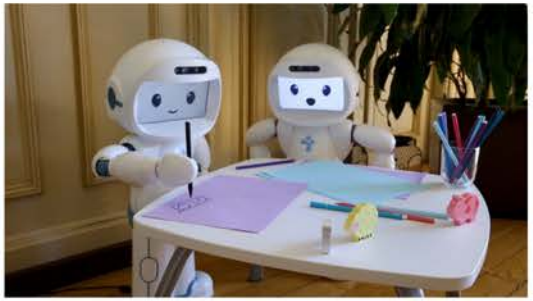
Computational Creativity Hub (CCH) - MSA Administration Building (Ground Floor)

On **March 11**, we welcome you to the **Opening of the Living Lab of the AI&ART pavilion** project @Esch 2022!

The event takes place at the Computational Creativity Hub of the University of Luxembourg on the Belval campus.

We invite anyone interested to participate in-person to discuss with artists and students their projects that combine two disciplines: **Artificial Intelligence and Art**.

These projects are created both by students, who provided the technological part and competencies to support the implementation of artworks, and artists, who are the source and the driver of the creative part.



2.Example: Forrester (1984)

Forrester's argument is that, starting from the statements that

- It is obligatory (under the law) that Smith not murder Jones.
- It is obligatory that, if Smith murders Jones, Smith murder Jones gently.

it logically follows that:

- If Smith murders Jones, it is obligatory, that Smith murder Jones gently.

However, if it were actually the case that Smith murdered Jones, it can then be deduced that:

- It is obligatory, that Smith murder Jones

which contradicts the first statement, leading to a logical fallacy.[4]

Source: wiikipedia

2.Example: Explaining GPT3: Forrester

M. It is obligatory (under the law) that Smith not murder Jones. It is obligatory that, if Smith murders Jones, Smith murder Jones gently. If Smith murders Jones, it is obligatory, that…

Q. …Smith be punished for the crime according to the law. It is not obligatory that Smith murder Jones gently, as this would be in violation of the law.

M. Smith murdered Jones, it is obligatory that…

Q. …Smith face the consequences of his actions, including any legal action that may be taken against him.

Source: wikipedia

2.Example: GPT3 and incorrect reasoning

M. Smith murdered Jones, it can then be deduced that…

Q. …Smith murdered Jones gently.

O. This is a false statement as there is no proof that Smith murdered Jones gently. It is possible that Smith murdered Jones in a cruel or violent manner.

M. Smith murdered Jones, it can then be deduced that…

Q. …Smith did not murder Jones gently.

O. No, that cannot be deduced. It is possible that Smith may have murdered Jones in a violent way. The fact that it is obligatory to murder Jones gently does not mean that Smith actually followed that obligation.

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. Jones…

Q. … should go to the assistance of his neighbors and let them know that he is coming. He should also ensure that he is prepared to help them in whatever way they need.

M. Why?

Q. Because it is the neighborly thing to do and it builds relationships between neighbors.

Source: wikipedia

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. Jones…

Q. … should go to the assistance of his neighbors and let them know that he is coming. He should also ensure that he is prepared to help them in whatever way they need.

deontic detachment

M. Why?

Q. Because it is the neighborly thing to do and it builds relationships between neighbors.

Source: wikipedia

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. Jones…

Q. … should go to the assistance of his neighbors and let them know that he is coming. He should also ensure that he is prepared to help them in whatever way they need.

M. Why?

Q. Because it is the neighborly thing to do and it builds relationships between neighbors.

goal / purpose of norm

Source: wikipedia

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. If Jones doesn't go, then he ought not tell them he is coming. Jones doesn't go. Jones…

Q. … ought not tell his neighbors he is coming.

M. Jones ought not to tell them he is coming, because…

Q. … it would be dishonest.

M. Why?

Q. Jones does not go to the assistance of his neighbors because he does not think it is his responsibility to do so, he does not have the resources to help, or he simply does not want to. Therefore, he should not tell them he is coming because it would be misleading and could give them false hope that he will be able to help them.

Source: wikipedia

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. If Jones doesn't go, then he ought not tell them he is coming. Jones doesn't go. Jones…

Q. … ought not tell his neighbors he is coming.

M. Jones ought not to tell them he is coming, because…

Q. … it would be dishonest.

M. Why?

Q. Jones does not go to the assistance of his neighbors because he does not think it is his responsibility to do so, he does not have the resources to help, or he simply does not want to. Therefore, he should not tell them he is coming because it would be misleading and could give them false hope that he will be able to help them.

factual detachment

Source: wikipedia

2.Example: Explaining GPT3: Chisholm

M. It ought to be that Jones goes (to the assistance of his neighbors). It ought to be that if Jones goes, then he tells them he is coming. If Jones doesn't go, then he ought not tell them he is coming. Jones doesn't go. Jones…

Q. … ought not tell his neighbors he is coming.

M. Jones ought not to tell them he is coming, because…

Q. … it would be dishonest.

M. Why?

Q. Jones does not go to the assistance of his neighbors because he does not think it is his responsibility to do so, he does not have the resources to help, or he simply does not want to. Therefore, he should not tell them he is coming because it would be misleading and could give them false hope that he will be able to help them.

beliefs, desires, intentions

Source: wikipedia

2.Example: Cottage regulations

1. There must be no fence.
2. If there is a fence, it must be a white fence.
3. If the cottage is by the sea, there may be a fence.

- (2) is intended as a CTD obligation of (1) and (3) as an exception to (1).
- A person who has a cottage by the sea with a fence does not violate (1), since (1) is defeated by (3): (1) does not apply when the cottage is by the sea.
- Someone whose cottage is not by the sea and who has a white fence complies with (2) but still violates (1): any fine imposed for violating (1) will have to be paid.
- A logic that in these circumstances regards (1) as being defeated by (2) cannot express this.

Source: [PS97]

2.Example: Explaining GPT3: Cottage

M. There must be no fence. If there is a fence, it must be a white fence. If the cottage is by the sea, there may be a fence. A person who has a cottage by the sea with a fence…

Q. … may choose to have a white fence, but it is not required.

M. Why?

Q. The color of the fence is a personal preference and is not a required element of a cottage by the sea.

M. Someone whose cottage is not by the sea and who has a white fence…

Q. …should consider replacing it with a more appropriate type of fence, such as a picket fence or a wrought iron fence.

M. Why?

Q. A white fence can stand out when placed in a natural setting, and a picket fence or a wrought iron fence can blend better with the landscape.

2.Example: Explaining GPT3: Cottage

M. There must be no fence. If there is a fence, it must be a white fence. If the cottage is by the sea, there may be a fence. Someone whose cottage is not by the sea and who has a white fence…

Q. … should consider replacing it with a different color or material to create a more inviting and visually appealing property. Alternatively, they can choose to plant trees, shrubs, and other plants along the perimeter of the property, creating a natural barrier.

M. Why?

Q. Natural barriers are low maintenance, aesthetically pleasing, and can also be beneficial for wildlife.

2.Example: Explaining GPT3: BOID

M. If the agent goes to the conference, then it believes that there are no cheap rooms close to the conference site. If it goes to the conference, then it is obliged to take a cheap room. If it goes to the conference, then it desires to stay close to the conference site. It intends to go to the conference. The agent…

Q. …should research the available rooms near the conference site to determine if there are any cheap options. If not, the agent may need to consider other options, such as staying at a nearby hotel or motel.

M. The goal…

Q. …of the agent is to find a reasonable and affordable accommodation close to the conference site.

Source: [BDHvdT01]

2.Example: Explaining GPT3: BOID

M. If the agent goes to the conference, then it knows that there are no cheap rooms close to the conference site. If it goes to the conference, then it is obliged to take a cheap room. If it goes to the conference, then it desires to stay close to the conference site. It intends to go to the conference. The agent is social. The goal…

Q. …is to find the best and most affordable room close to the conference site. The agent will research different hotels, compare prices, and take into account the quality of the rooms and services offered. The agent will also consider the distance from the conference site and other factors such as reviews, amenities, and other amenities. The agent will also consider the convenience of transportation to and from the conference site. Finally, the agent may consult with friends or colleagues who have recently attended the conference to get recommendations.

2.Deontic explanation of GPT3

- We are not only interested in GPT3, but in systems like GPT3
  - Known as foundation models, or large language models
- We do not consider personalization and psychological aspects
- Still relevance problem (in sense of IR) with respect to the explanation gap
  - Because it is the neighborly thing to do and it builds relationships between neighbors.
  - because it would be misleading and could give them false hope that he will be able to help them
  - Jones does not go to the assistance of his neighbors because he does not think it is his responsibility to do so, he does not have the resources to help, or he simply does not want to.
  - a more appropriate type of fence, such as a picket fence or a wrought iron fence.
  - The agent will also consider the convenience of transportation to and from the conference site. Finally, the agent may consult with friends or colleagues who have recently attended the conference to get recommendations.
- V. Opijnen M., Santos C. (2017), On the Concept of Relevance in Legal Information Retrieval, Artificial Intelligence and Law Journal.

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanation
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

3. Normative system as deontic explanation

- Obligations, permissions and institutional facts are detached from regulative, permissive and constitutive norms (rules, imperatives, …)
- Norms may have a purpose or goal associated with them.
- Why are bikes forbidden to enter the park?
- Because of constitutive norm that bicycles count as vehicles, and regulative norm that vehicles are forbidden in the park.
- Why?
- Goal of rule prohibiting to enter into parks is to promote road safety.

Source: [BGRvdT09,BGRvdT10]

3.Time machine

December 1997 (25 years ago)

- Leon goes from Saarbruecken to Toulouse :)

December 1992 (30 years ago)

- Leon starts his PhD in Rotterdam
- KR conference series was just created (1989)
- DEON conference series was just created (1991)
- Influence of philosophical logic, handbook 1983-1986 (4 volumes)

3.Chisholm’s scenario as normative system

- n_1: O g
- n_2: g then O t
- n_3: not g then O not t
- not g

3.Chisholm's scenario as normative system

- n_1: O g
- n_2: g then O t
- n_3: not g then O not t
- not g
- n_4: O g and not g then O s_1
- n_5: O t and not t then O s_2
- n_6: O not t and t then O s_3
- n_7: O g and g then O r_1
- n_8: O t and t then O r_2
- n_9: O not t and not t then O r_3

3.Chisholm's scenario (obligation antecedent)

- n_1: O g
- n_2: g then O t
- n_3: not g then O not t
- not g
- n_4: O g and not g then O s_1 — V(n_1) then O s_1
- n_5: O t and not t then O s_2 — V(n_2) then O s_2
- n_6: O not t and t then O s_3 — V(n_3) then O s_3
- n_7: O g and g then O r_1 — F(n_1) then O r_1
- n_8: O t and t then O r_2 — F(n_2) then O r_2
- n_9: O not t and not t then O r_3 — F(n_3) then O r_3

3.Chisholm's scenario (obligation consequent)

- n_1: O g — not $V(n_1)$ then g — $F(n_1)$ iff g
- n_2: g then O t — g and not $V(n_2)$ then t — $F(n_2)$ iff g and t
- n_3: not g then O not t — not g and not $V(n_3)$ then not t — $F(n_{3)})$ iff …
- not g
- n_4: O g and not g then O s_1 — $V(n_1)$ and not $V(n_4)$ then s_1
- n_5: O t and not t then O s_2 — $V(n_2)$ and not $V(n_5)$ then s_2
- n_6: O not t and t then O s_3 — $V(n_3)$ and not $V(n_6)$ then s_3
- n_7: O g and g then O r_1 — $F(n_1)$ and not $V(n_7)$ then r_1
- n_8: O t and t then O r_2 — $F(n_2)$ and not $V(n_8)$ then r_2
- n_9: O not t and not t then O r_3 — $F(n_3)$ and not $V(n_9)$ then r_3

3.Cottage regulations

- n_1: O no f
- n_2: f then O w
- n_3: s then P f
- n_4: O no f and f then Os_1
- n_5: O w and no w then Os_2

3.Cottage regulations as normative systems

- n_1: O no f — not $V(n_1)$ and not $Ab(n_1)$ then no f
- n_2: f then O w — f and not $V(n_2)$ and not $Ab(n_2)$ then w
- n_3: s then P f
- n_4: O no f and f then Os_1 — $V(n_1)$ and not $V(n_4)$ and not $Ab(n_4)$ then s_1
- n_5: O w and no w then Os_2 — $V(n_2)$ and not $V(n_5)$ and not $Ab(n_5)$ then s_2

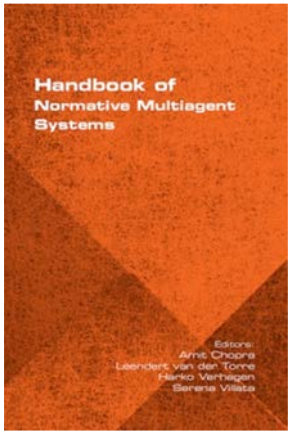
3.Cottage regulations as normative systems

- n_1: O no f — not $V(n_1)$ and not $Ab(n_1)$ then no f
- n_2: f then O w — f and not $V(n_2)$ and not $Ab(n_2)$ then w
- n_3: s then P f — s then $Ab(n_1)$
- n_4: O no f and f then Os_1 — $V(n_1)$ and not $V(n_4)$ and not $Ab(n_4)$ then s_1
- n_5: O w and no w then Os_2 — $V(n_2)$ and not $V(n_5)$ and not $Ab(n_5)$ then s_2

- Leendert van der Torre, Yao-Hua Tan: Diagnosis and Decision Making in Normative Reasoning. Artif. Intell. Law 7(1): 51-67 (1999)

3.Historical notes on normative systems

- Inspired by deductive systems in 70s (Alchourron and Bulygin)
- Logic programming in 80s (Kowalski, Sergot, Lee,…)
- Reiter's default logic in 90s (Horty)
- Deontic logic based on systematic frame constants in 90s (Aqvist)
- Input/output logics in 00s (Makinson and van der Torre since 1997)
- Normative multiagent systems since 2005
  - See Handbook of Normative Multiagent Systems, 2018



Normative systems vs deontic logic

Advantages normative (rule-based) systems

- Easy to extend to exceptions
- Computationally often more efficient
- Well suited for legal reasoning where norms are explicit
- Explain in terms of norms

Advantages (modal) deontic logic

- No need to name norms, typically more expressive
- More natural for explanation (sometimes), as will be explained next

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanation
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

4.Time machine

December 1997 (25 years ago): Leon to Toulouse!

- The start of input/output logics

December 1992 (30 years ago): Leon starts his PhD!

- Diagnostic theory for deontic reasoning

Before December 1982 (40+ years ago)

- Kratzer's theory of modality, Poole's Theorist system

Before December 1972 (50+ years ago)

- Alchourron & Bulygin, Danielsson, Hansson, Van Fraassen, .

Prehistory (60+ years ago)

- Hempel, Von Wright, Hallden, Anderson, Kelsen, Rescher, …

4.Kratzer two step explanations

- Explain obligation by modal base and ordering source
  - Conversational background, given by pragmatics: explanation gap
  - For us, ordering source is (detached from) a normative system
- Ordering source induces a preference relation over worlds
  - Subset ordering over violation propositions (following Rescher 67)
- Conflicts in ordering source can be resolved or not

go tell > go not tell > not go not tell > not go tell

$V(n_2)$ $V(n_1)$ $V(n_1)V(n_3)$

4.Kratzer bimodal explanations

- Kratzer introduces a general theory of modality
- Use two step explanations for constitutive norms / beliefs / Ab
- Multi preference or decision theoretic semantics
- In the cottage regulations, all conflicts are resolved

no sea no fence > no sea white fence > no sea other fence

$V(n_1)$ $V(n_1)V(n_2)$

V

sea no fence = sea white fence > sea other fence

$Ab(n_1)$ $Ab(n_1)$ $Ab(n_1)\ V(n_2)$

4.Kratzer BOID explanations

- In GPT3 explanations, relevant rules also other attitudes (e.g. desires)
- Use the same Kratzer explanations for desires, intentions, …
- Four ordering sources for BOID, four preference orderings
- Each ordering source / preference ordering can contain conflicts

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanations
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

5.Nonmonotonic logic for explanation

- Explanation is pragmatics (NL) is nonmonotonic reasoning
- Kratzer is using techniques similar to nonmonotonic logic
  - Essentially the same as Poole's theorist, which was developed afterwards
- Makinson 2005: bridges from classical to nonmonotonic logic
- AI / KR / NML has made substantial progress in 30 years
- How can modern nonmonotonic logic support deontic explanations?

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanations
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

6.Deontic ASP explanations

- Ciabattoni, Cabalar, vdTorre. Deontic equilibrium logic. In preparation
- Uniform framework for defeasible deontic logic
  - For example, can elegantly deal with challenge in appetizer
- Conflict resolution: normative system is CSP, multiple answers to CSP
- Standardization defeasible deontic logic with tool support
- Uses explanation techniques from ASP, e.g. for causal explanations

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanations
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

7.Deontic argumentation explanations

- Many papers are now written on this topic
- A good one: Kees van Berkel, Christian Straßer: Reasoning With and About Norms in Logical Argumentation. COMMA 2022: 332-343
- Deontic description logic, and first-order deontic logic
  - Ordering not only worlds, but also objects, rules, …
- More fine-grained analysis of conflicts and their resolution
  - E.g. inspired by conflict sets as minimmal inconsistent sets in diagnosis
  - E.g. flattening a conflict graph to a ranking
- Game theoretic techniques (e.g. from MAS) for analysis of conflicts
- Open issue: Contrastive explanations

What you need to remember from this talk

What the theory of (deontic) explanation needs most of all is…

game theory

mechanism design

social choice theory

Liuwen Yu, Dongheng Chen, Lisha Qiao, Yiqi Shen, Leendert van der Torre: A Principle-based Analysis of Abstract Agent Argumentation Semantics. KR 2021: 629-640

Layout of this talk on deontic explanations

1. Appetizer: Forrester's gentle murderer scenario
2. Example: explaining GPT3: Forrester, Chisholm, Cottage, BOID, …
3. Normative system as one step deontic explanations
4. Kratzer two step explanations: normative system as ordering source
5. Nonmonotonic logic framework for deontic explanation
6. Deontic ASP explanations
7. Deontic argumentation explanations
8. Legal and ethical explanations

8.Legal and ethical explanations

- Deontic logic is used for legal and ethical reasoning
  - as well as practical reasoning, linguistic analysis, coordination, security, …
- Mining and reasoning with legal texts (MIREL), 2016-2019
- Jiminy architecture (2019) normative systems & formal argumentation
- LogiKEy framework, methodology and tool support (2020)
- Handbook of Legal AI (to appear)
- However, these areas have their own specificities
  - Legal norms are explicit and in natural language
  - Purpose of the norms is often known / written down as well
- Another talk next week here in Toulouse ☺

Summary of this talk on deontic explanations

- Explanation (answering why questions) using Kratzer pragmatics
- Two step explanations: ordering sources and preference orderings
- Relevance: explanation gap: from modals to ordering sources ++
- For deontic explanations: from obligations to normative systems ++
- BOID Kratzer explanations: four dimensional preference structures
- Nonmonotonic logic for explanation and pragmatics
- ASP and formal argumentation provide more fine-grained analysis
- Explanation needs contributions from game theory and social choice
- To be added: contrastive and causal explanations, personalization

Questions?