Seminar in Logic: Non-Monotonic Reasoning

Konstantin Kueffner

1 Introduction

Non-Monotonic reasoning is a form of reasoning which rejects the notion of monotonicity, in order to capture, among others *defeasible reasoning*. Therefore, it is strongly connected with the notions of assumption and normality. Since its formal origin in 1980 a variety of formalisms and approaches were developed. This text shall serve as a quick introduction into non-monotonic reasoning. Hence, it starts with introducing monotonicity and its limitations. From there, the two major approaches of non-monotonic reasoning are introduced. This is followed by the presentation of two specific formalisms. And concluded by a brief discussion of possible applications. (Bochman 2007). Test citation Bochman 2007

1.1 Monotonicity

When it comes to reasoning, *monotonicity* ensures that additional information can not invalidate previous conclusions. That is, if a conclusion can be inferred with less information, it should still be possible to obtain the same conclusion with more information. Formally, monotonicity can be understood in two different ways.

Firstly, global monotonicity ensures that adding formulas to a theory preserves previous inferences. That is,

Definition 1.1 (McDermott and Doyle 1980; McCarthy 1981). *A logic is called globally monotonic, if the following statement holds.*

> *Let A and B be two theories, such that $A \subseteq B$ then $Th(A) \subseteq Th(B)$.*

Furthermore, $Th(S)$ is defined as the set of sentences derivable from a given theory S by the means of inference provided by the given logic, i.e. $Th(S) = \{p \mid S \vDash p\}$ (or syntactically $Th(S) = \{p \mid S \vdash p\}$).

This means that, since $A \subseteq B$ every model of B must also be model of A. Therefore, $A \vDash p$ implies $B \vDash p$ for all models of B. Secondly, *local monotonicity* ensures that adding formulas to a specific inference preserves the conclusion. That is,

Definition 1.2 (Bochman 2005). *A logic is called locally monotonic, if it upholds the notion of* Strengthening the Antecedent.

This means that if A implies C then one should be able to expand the statement to $A \wedge B$ implies C. Lastly, an example where both notion of monotonicity can be observed is classical logic (Bochman 2005, 2007).

2 Motivating Problems

The development of non-monotonic formalisms was heavily influenced by the field of artificial intelligence. That is, the pragmatic solution developed to cope with certain problems served as a strong incentive for the development of a more systematic foundation for non-monotonic reasoning. Two common motivating problems will be presented. Namely, the canonical *Tweety*-Problem, depicting the limits of monotonicity, and the general form of the *Frame*-Problem, i.e. the *Temporal Projection Problem*, depicting issues of practicality with respect to monotonicity (Bochman 2007).

2.1 Tweety-Problem and defeasible reasoning

One of the most prominent examples in the literature displaying why non-monotonic reasoning would be desirable is the *Tweety*-Problem, introduced in Reiter 1980. Here, a slight variant of the problem will be presented.

Here the desire is to formalise the fundamental principle of *Innocent until proven guilty*, i.e. if x is a person one should assume that x is innocent. Naively, this may result in

$$\mathcal{T}_1 \coloneqq \{\forall x\, Person(x) \to Innocent(x), Person(Tweety)\}$$

allowing the safe inference of $Innocent(Tweety)$. However, if additional information is added to the theory,

$$\mathcal{T}_2 \coloneqq \mathcal{T}_1 \cup \{Murderer(Tweety), \forall x\, Murderer(x) \to Person(x), \forall x\, Murderer(x) \to \neg Innocent(x)\}$$

this will result, given the formalisation presented, in a inconsistent theory. To solve this the newly introduced exception has to be accounted for.

$$\mathcal{T}_3 \coloneqq \mathcal{T}_2 \smallsetminus \{\forall x\, Person(x) \to Innocent(x)\} \cup \{\forall x\, Person(x) \wedge \neg Murderer(x) \to Innocent(x)\}$$

As originally intended it is now possible to infer $\neg Innocent(Tweety)$. Unfortunately, if the theory is expanded further, i.e. $\mathcal{T}_4 \coloneqq \mathcal{T}_3 \cup \{Person(Polly)\}$ the desired assumption of innocence is no longer possible, as $\neg Murderer(Polly)$ blocks the application of the rule. Hence, $\neg Murderer(Polly)$ has to be proven before innocence can be proclaimed, which is a dramatic departure from the notion of *Innocent until proven guilty*.

On a formal level the issue here is twofold. Firstly, in general it is impossible to account for all possible exceptions, thus there is always the danger of producing an inconsistent theory, if new information is obtained. Secondly, in practice it is highly impractical to block any inference until every exception is checked. Hence, it would be desirable to have a formalism, which captured the notion of normality. Thus allowing for the making of assumptions, as well as the retraction of conclusion in light of new information. That is, a formalism that captures *defeasible reasoning*. Lastly, the relevance of defeasible reasoning is not restricted to artificially constructed problems such as the Tweety problem, but it occurs frequently in the everyday reasoning of humans. For example, in medical diagnosis or even scientific reasoning, e.g. revising the hypothesis in light of contradicting measurements. (Reiter 1980; Bochman 2007; Koons 2017; Strasser and Antonelli 2018).

2.2   The Temporal Projection Problem

The *temporal projection problem* is a pervasive problem within the fiel of artificial intelligence. It encompasses the persistence, ramification and qualification problems. Among those three subproblems it is the persistence or frame problem, initially proposed by McCarthy, which strongly motivated the development of non-monotonic formalisms (Bochman 2007).
The problem stated in Shanahan 2016 (adapted based on Ginsberg and Smith 1987) will be used to present all three subproblems. The theory

$$\forall x \, (PAINT(x,c) \to COLOUR_{n+1}(x,c))$$
$$\forall x \, (MOVE(x,p) \to POSITION_{n+1}(x,p))$$

expresses that if in state n something (x) is painted ($PAINT$) the colour c, then in state $n+1$ the object x will have the colour ($COLOUR$) c. Similarly, if in state n something (x) is moved ($MOVE$) to the position p, then in state $n+1$ the object x will be at position ($POSITION$) p.

The persistence problem / frame problem: The general problem of predicting the properties that remain the same as actions are performed, i.e. stating what remains unaffected by said action. Given the example above in state 0 one has $POSITION_0(Book, Desk)$, if $PAINT(Book, Red)$ is invoked, only $COLOUR_1(Book, Red)$ exist, while $POSITION_1(Book, Desk)$ does not. In the absence of inertia, the book lost its position in space. Ideally inertia could be assumed, i.e. if not explicitly specified everything remains the same, as an axiomatic approach will result in a theory of immense scope (Bochman 2007; Lifschitz 2015).

The ramification problem: The general problem of predicting the properties that do change as actions are performed. Hereby, the difficulty lies within modelling implicit changes. Given the example, assume that inertia axioms where added to the theory, that is

$$\forall x \, ((MOVE(x,p) \wedge COLOUR_n(x,c)) \to COLOUR_{n+1}(x,c))$$
$$\forall x \, ((PAINT(x,p) \wedge POSITION_n(x,p)) \to POSITION_{n+1}(x,p))$$

Now given $COLOUR_0(Book, Red)$, if $MOVE(Book, Black\_Paint)$ is invoked it will be concluded that $POSITION_1(Book, Black\_Paint)$ and $COLOUR_1(Book, Red)$. However, by moving the book its colour was implicitly changed, thus $COLOUR_1(Book, Black)$ should hold. Hence, if inertia axioms are used, exceptions have to be considered. Hence, it would be desirable to assume that no implicit changes occur, with the possibility of retraction in light of additional information (Bochman 2007; Ginsberg and Smith 1987).

The qualification problem: The general problem of predicting the properties that have to hold such that an action has its intended effects. Given the example, there can be certain conditions prohibiting $MOVE(x,y)$. For example, assume $POSITION_0(Book, Blackhole)$, thus any application of $MOVE(Book, y)$ will result in $POSITION_0(Book, Blackhole)$. That is, $MOVE$ has no longer its intended effect. Ideally, one could simply assume that an action has its intended effects (Ginsberg and Smith 1987; Bochman 2007).

The problems described above where simply motivation factors for the development of non-monotonic formalisms, thus it does not follow that these problems can only be accounted for by such formalisms. In fact the non-monotonic formalisms discussed below failed to solve the frame problem (see Yale-Shooting Problem). Moreover, there are also some monotonic approaches,

that provide adequate solutions for the frame problem. Lastly, on a technical level for applicative purposes the frame problem is solved to a satisfying degree (Shanahan 2016).

Given the problems presented within this section one can conclude that it would be desirable to have a formalism, which captures statements such as "Normally X holds", "Typically X is the case" or "Assume X as a default". Hence, it should be able to model some notion of normality and thus also abnormality. Ideally, it would allow for a reasoned use of assumptions, which includes the retraction of inferences based on these assumptions. Therefore enabling safe reasoning within a dynamic environment. Moreover, it should be possible to distinguish between assumptions based on preference. Hence, the expectations with non-monotonic reasoning is that by rejecting the notion of monotonicity and by embracing defeasible reasoning rather than just relying deductive reasoning, some of these desires may be satiated.

3 Approaches to Non-Monotonic Reasoning

Generally speaking there are two main approaches for capturing non-monotonic reasoning, namely *Explanatory Non-Monotonic Reasoning* (or *Consistency-Based Logics*) and *Preferential Non-Monotonic Reasoning*. However, other modes of categorisation exist. For instance, while most non-monotonic formalisms extend classical logic in order to obtain non-monotonicity, this can be done either by means of adding additional logical formulas (e.g. Circumscription or Modal Non-Monotonic Logic) or by adding additional rules of inference (e.g. Default Logic). Additionally, how conflicting conclusions are handled, could also be used for categorisation. That is, do they employ sceptical reasoning and simply prohibit the inference of contradicting conclusions, or do they employ credulous reasoning and account for all possible sets of consistent of beliefs. Lastly, another possible distinction can be made based on the property of cumulativity, a property essential for the save use of lemmas. Incidentally, most formalisms in the preferential approach are sceptical and cumulative in nature, while most in the explanatory one are credulous and non-cumulative in nature. (Strasser and Antonelli 2018; Bochman 2007; Brewka, Dix, and Konolige 1997; Antoniou and Wang 2007).

3.1 Preferential Non-Monotonic Reasoning

Preferenctial non-monotonic reasoning is historically speaking younger than *Explanatory Non-Monotonic Reasoning*, yet some of its core principals are already present in the introductory paper of predicate circumscription, i.e. McCarthy 1981. Some of the formalisms subsumed by this approach are *Circumscription*, *Closed World Assumption* and *Conditional Logics*. In general preferential non-monotonic reasoning is locally non-monotonic, possibly globally monotonic, cumulative and better suited for sceptical reasoning. Furthermore, it is characterised by the notion preference between interpretation, which introduces non-monotonicity as adding new information may result in a shift of preference. This notion of preference can, in its most general form, be captured by the model theoretical notion of *Cumulative Models*, which if restricted properly can be used to capture any preferential formalism on a model theoretic level. Moreover, it is possible to capture the behaviour of cumulative models by imposing certain rules onto an inference relation. Hence, establishing a strong connection between the meta-theoretic analysis of inference relations and the preferential approach to non-monotonic reasoning. In this subsection cumulative models as well as two important restrictions are presented (Bochman 2007; Brewka, Dix, and Konolige 1997; Bochman 2005; Kraus, Lehmann, and Magidor 1990).

These cumulative models mentioned above can be defined as follows.

Definition 3.1 (Kraus, Lehmann, and Magidor 1990)**.** $\langle S, l, \prec \rangle$ *where, S is a set of states, $l : S \to \wp(\mathcal{U})$ is a function with $\mathcal{U}$ set of all interpretations and $\prec$ is a strict partial order on S satisfying the smoothness condition.*

with the smoothness condition being

Definition 3.2 (Kraus, Lehmann, and Magidor 1990)**.** $\langle S, l \prec \rangle$ *satisfies smoothness condition if $\forall \alpha \in \mathcal{L}$ the set $\{s \in S \mid \forall m \in l(s) \ m \vDash \alpha\}$ is smooth, i.e. every state is either minimal or is in relation to a minimal state.*

and can be captured on a meta-theoretical level (i.e. system **C**, see Kraus, Lehmann, and Magidor 1990). Moreover, the by restricting cumulative models one obtains *preferential models*. That is,

Definition 3.3 (Kraus, Lehmann, and Magidor 1990)**.** *A cumulative model $\langle S, l, \prec \rangle$ where $|l(s)| = 1$, is called preferential model.*

which can also be captured on a meta-theoretical level (i.e. system **P**, see Kraus, Lehmann, and Magidor 1990). Lastly, if preference models are further restricted model preference logics can be obtained. The core idea behind them is, to impose a well-founded strict partial order onto all possible interpretations, thus restricting the set of possible interpretations to those minimal with respect to the established preference. Formally, they are constructed as follows.

Definition 3.4 (Kraus, Lehmann, and Magidor 1990)**.** *Model preferential logics is a preference model $W \coloneqq \langle S, l, \prec \rangle$, such that S is a subset of all interpretations, i.e. $S \subseteq \mathcal{U}$, l is the identity and $\prec$ is well-founded.*

These models are not only more expressive than the others, they are also the most relevant for this text, as if further restricted model preference logic can be used to capture circumscription on a model theoretical level (Brewka, Dix, and Konolige 1997; Bochman 2007).

3.2 Explanatory Non-Monotonic Reasoning

Explanatory Non-Monotonic Reasoning (or Consistency-Based Logics) the older and more 'mainstream' approach to non-monotonic reasoning. Some of the formalisms subsumed by this approach are default logic and modal non-monotonic logic. Moreover, it is also strongly connected with logic programming, as well as abductive and causal reasoning. In general explanatory non-monotonic reasoning is globally non-monotonic, possibly locally monotonic, not cumulative and well suited for credulous reasoning. Being not cumulative makes the formalisation of lemmas difficult, e.g. in default logic it is only possible to express lemmas, if the context of its derivation is accounted for. If contrasted with the preferential approach, it is difficult to formally define the core concept of this approach as it does not have such a clear underlying model theoretic notion. Hence, they will only be discussed informally (Bochman 2007; Brewka, Dix, and Konolige 1997; Bochman 2005).

Firstly, central to explanatory non-monotonic reasoning is the notion *explanatory closure*, which expresses that any fact holding in a model should be explained by the rules that describe the domain. This is critical, as future inference can invalidate previously made assumptions, thus it may be possible that previous conclusions my loose their justification. The necessity of this principal can be observed in the default logic example below.

Secondly, explanatory closure requires that any fact used for inference has to be explained as well, resulting in possible self-referential behaviour. Therefore, many formalisms of this approach use a fixed-point operator to generate a set of all provable sentences, the names of which vary based on the formalism. Here they will be referred as *stable extensions*. A stable extension represents one set of consistent beliefs. Since it is possible to have several consistent, but contradicting sets of beliefs, there are multiple possible stable extensions, see *Nixon-Diamond* in Strasser and Antonelli 2018, thus explaining the connection with credulous reasoning. Lastly, it has to be noted that Bochman 2005 attempts to create a unifying semantics for capturing all formalisms subsumed by explanatory non-monotonic reasoning by means of describing its inference via a biconsequence relation combined with a four-valued logic. (Bochman 2007; Brewka, Dix, and Konolige 1997; Bochman 2005)

4 Specific Formalisms

There are a variety of different formalisms that try to capture non-monotonic reasoning, each of which contains a plethora of additional variants. Hence, it is not within the scope of this work to present all formalisms and their variants. Therefore, only two specific formalisms will be discussed. Namely, *Circumscription* and *Default Logic*. On the one hand, circumscription was chosen as an representative for a formalism that can be understood as part of the preferential approach, as well as a formalism that introduces monotonicity by adding additional logic formulas. On the other hand, default logic was chosen, as it can be understood as a formalism adhering to the explanatory approach, as well as a formalism that uses inference rules to obtain non-monotonicity.

4.1 Circumscription

This variant is called *Predicate Circumscription* and was introduced by McCarthy 1981. At its core it can be understood as a method of adding additional formulas to a theory, such that the set of possible interpretations is restricted to those models that are minimal in their extensions of the circumscribed predicate. That is, the circumscribed predicate is only positive, if absolutely necessary. Hence, some notion of preference among models is expressed, thus tying it to the preferential approach of non-monotonic reasoning (Bochman 2007; Brewka, Dix, and Konolige 1997).

This form of circumscription can be defined by the following second-order formula.

Definition 4.1 (McCarthy 1981). *Let S FOL-sentence containing $P(x_1, \ldots, x_n)$, short $P(\overline{x})$. Let $S(\Phi)$ replaces P with Φ.*
Then the schema

$$\forall \Phi \left((S(\Phi) \wedge \forall \overline{x} \left(\Phi(\overline{x}) \to P(\overline{x}) \right)) \to \forall \overline{x} \left(P(\overline{x}) \to \Phi(\overline{x}) \right) \right)$$

is called the circumscription of P.

That is, during process of circumscription this formulas ensures that P is minimal with respect to its extensions, by means of adding formulas. Therefore, while circumscription was initially designed on a purely syntactic level, it is possible to capture circumscription on a model theoretic level, by imposing the following restriction onto a model preference logic.

Definition 4.2 (Brewka, Dix, and Konolige 1997). *Let $\langle S, l, \prec \rangle$ be a model preference logic, s.t.*

$$s_1 \prec s_2 \text{ iff } l(s_1) \vDash P(\overline{x}) \text{ implies } l(s_2) \vDash P(\overline{x}) \text{ and not } l(s_2) \vDash P(\overline{x}) \text{ implies } l(s_1) \vDash P(\overline{x})$$

with P being circumscribed.

That is, interpretations are preferred based on their extensions of P.

Before moving on, it has to be mentioned that in order to model the previously mentioned notion of normality *abnormality theory* was introduced. Herby one specifies abnormality predicates which will be circumscribed, thus minimising abnormalities (McCarthy 1981).
In order understand how one may reason with circumscription, a small example is required. Hence, $S = Guilty(A) \wedge Guilty(B) \wedge Guilty(C)$ and let $Guilty$ the predicate to be circumscribed. Firstly, how does the restricted form of model preference logic behave. (For sake of simplicity simply pretend that the following interpretations are the only in existence.)

For $x \in \{0, 1, 2, 3\}$ let $\mathcal{I}_x := (D_{\mathcal{I}_x}, I_{\mathcal{I}_x})$ s.t. $I_{\mathcal{I}_0}(Guilty) := \{\}$, $I_{\mathcal{I}_1}(Guilty) := \{\delta\}$, $I_{\mathcal{I}_2}(Guilty) := \{\delta, \sigma, \eta\}$ and $I_{\mathcal{I}_3}(Guilty) := \{\delta, \sigma, \eta, \gamma\}$ with $D_{\mathcal{I}_x} := \{\delta, \sigma, \eta, \gamma\}$. Hence based on the size of the extensions of $Guilty$ one obtains $\mathcal{I}_0 \prec \mathcal{I}_1 \prec \mathcal{I}_2 \prec \mathcal{I}_3$. Now given S the preferred model is $\mathcal{I}_1$, because $\mathcal{I}_0$ is not a model of S. If for example $(A \neq B \wedge A \neq C \wedge C \neq B)$ is added, now the preferred model will be $\mathcal{I}_2$.

Secondly, the application of the circumscription formula. In this example, let Φ be $\Phi(x) = (x = A \vee x = B \vee x = C)$, thus by substituting everything into the circumscription formula one obtains.

$$(A = A \vee A = B \vee A = C) \wedge (B = A \vee B = B \vee B = C) \wedge (C = A \vee C = B \vee C = C)$$
$$\wedge \, \forall \overline{x} \, ((\overline{x} = A \vee \overline{x} = B \vee \overline{x} = C) \rightarrow Guilty(\overline{x}))) \rightarrow \forall \overline{x} \, (Guilty(\overline{x}) \rightarrow (\overline{x} = A \vee \overline{x} = B \vee \overline{x} = C))$$

This equation ensures that there are at most three objects that satisfy the theory, i.e. the set of possible interpretations was restricted to those interpretations having at least one and at most three elements for which $Guilty$ holds. If the same outcome as the model theoretic approach is desired a different Φ has to be chosen (recall $\forall \Phi$). In order to observe the non-monotonic behaviour of circumscription, one simply has to extend S by adding $Guilty(D)$. By doing so it is no longer possible to derive $\forall \overline{x} \, (Guilty(\overline{x}) \rightarrow (\overline{x} = A \vee \overline{x} = B \vee \overline{x} = C))$, due to $(D = A \vee D = B \vee D = C)$ (McCarthy 1981).

4.2   Default Logic

Default Logic, introduced in Reiter 1980, is one of the more expressive and popular formalisms of non-monotonic reasoning. At its core, default logic relies on certain inference rules called *defaults*. These defaults can be understood as rules of thumb that enable the reasoner to jump to conclusions given incomplete information. Moreover, reasoning within default logic is done, according to the initial definition, by means of a fixed-point opertion, thus tying it to the explanatory approach of non-monotonic reasoning (Bochman 2007; Antoniou and Wang 2007).

Firstly, a default is defined as

Definition 4.3 (Reiter 1980). *A default δ has the form*

$$\frac{\varphi : \psi_1, \ldots, \psi_n}{\chi}$$

with $\varphi, \chi, \psi_1, \ldots, \psi_n$ being closed propositional formulas for $n > 0$. Moreover, φ is called the prerequisite, $\psi_1, \ldots, \psi_n$ are the justifications and χ is the consequent.

Moreover, a default containing variables is just a compact representation of the set of defaults containing all its grounded instances. In order to reason within default logic the notion of a default theory and the fixed-point operator required for generating stable extensions have to be introduced. That is,

Definition 4.4 (Reiter 1980). *$\Delta = (D, W)$ is a default theory. With W a set of predicate formulas and D a set of defaults. For any $S \subseteq \mathcal{L}$ (first order logic), let $\Gamma(S)$ be the smallest set satisfying*

- *D1: $W \subseteq \Gamma(S)$*

- *D2: $Th_{\mathcal{L}}(\Gamma(S)) = \Gamma(S)$*

- *D3: if $(\varphi : \psi_1, \ldots, \psi_n / \chi) \in D$ and $\varphi \in \Gamma(S)$ and $\neg\psi_1, \ldots, \neg\psi_n \notin S$ then $\chi \in \Gamma(S)$.*

That is, any stable extension generated by Γ has to include the set of facts W, has to be deductively closed, must be maximal with respect to the application of defaults and must not contain any ungrounded beliefs (Brewka, Dix, and Konolige 1997).

In order to understand how one reasons within default logic, two examples are presented. Firstly, given $\Delta \coloneqq (W, D)$

$$W \coloneqq \{Murderer(Tweety), Person(Polly), Murderer(x) \to \neg Innocent(x), Murderer(x) \to Person(x)\}$$
$$D \coloneqq \{Person(x) : Innocent(x) / Innocent(x)\}$$

There are two possible sets to be generated:

$W \cup \{Innocent(Polly), Person(Tweety), \neg Innocent(Tweety)\}$: Here $Innocent(Polly)$ is obtained by applying the default, while the other two are obtained by closing the set deductively. Moreover, due to $\neg Innocent(Tweety)$ the application of the default is blocked. Since every fact is explained and no additional inferences can be made, a stable extension is obtained.

$W \cup \{Innocent(Polly), Person(Tweety), \neg Innocent(Tweety), Innocent(Tweety)\}$: Here $Innocent(Polly)$ is obtained by applying the default. Now first $Person(Tweety)$ is derived, thus enabling the application of the default resulting in $Innocent(Tweety)$, but now one has to close the set deductively adding $\neg Innocent(Tweety)$. However, now the default can no longer serve as an explanation for $Innocent(Tweety)$. Thus it is not a stable set and will be discarded.

Secondly, given $\Delta \coloneqq (W, D)$

$$W \coloneqq \{Murderer(Tweety), Person(Polly), Murderer(x) \to Person(x)\}$$
$$D \coloneqq \{Person(x) : Innocent(x) / Innocent(x), Murderer(x) : \neg Innocent(x) / \neg Innocent(x)\}$$

There are two possible sets to be generated:

$W \cup \{Innocent(Polly), Person(Tweety), \neg Innocent(Tweety)\}$: Here $Innocent(Polly)$ is obtained by applying the first default and $\neg Innocent(Tweety)$ by applying the second. Furthermore, $Person(Tweety)$ is obtained by deduction. Lastly, due to $\neg Innocent(Tweety)$ the application of the first default is blocked. Since every fact is explained and no additional inferences can be made, a stable extension is obtained.

$W \cup \{Innocent(Polly), Person(Tweety), Innocent(Tweety)\}$: Here after obtaining $Person(Tweety)$ by deduction, $Innocent(Polly)$ and $Innocent(Tweety)$ are obtained by applying the first default. Lastly, due to $Innocent(Tweety)$ the application of the second default is blocked. Since every fact is explained and no additional inferences can be made, a stable extension is obtained.

Lastly, it has to be mentioned that this is a good example of credulous reasoning as there are two possible sets of consistent beliefs. If a sceptical point of view should be entertained, one would simply have to take the intersection, thus obtaining $W \cup \{Innocent(Polly), Person(Tweety)\}$ (Antoniou and Wang 2007; Brewka, Dix, and Konolige 1997).

5 Applications

5.1 Law

Logic and law are tightly interconnected due to the interdisciplinary field of *Artificial Intelligence and Law*. This interconnection is mutually beneficial, as it enables a more formal approach to reasoning with a legal context, while at the same time providing a vast test bed consisting of actual problems and their corresponding solutions for applications in artificial intelligence and logic. Moreover, reasoning within a legal context is divers and can not be reduced to formalisation and automated deduction. That is, there are various different areas of reasoning within a legal context, some of which will be teased below.

Firstly, *reasoning with the law* is an area of legal reasoning that is mainly concerned with formalising legal rules and reasoning with them on an axiomatic basis, without questioning the facts or the legal rules. For example, how to resolve conflicts between conflicting legal rules. Secondly, *reasoning about the law* is an area that opens up the possibility to reasoning about legal rules, their interpretation, their applicability and validity, while at the same time accepting the facts as a given. For example, how are legal rules to be interpreted? Thirdly, *reasoning about the facts* this area of legal reasoning is concerned with how to reason about facts and how they may be proven. For example, which source is given the context more reliable. Fourthly, *reasoning about interactions* this area of legal reasoning is concerned with the dynamic multi-agent environment that arises in legal disputes. One task within this field could be to model how facts are "discovered" during the back and forth of arguments and claims that occur during the dispute. Lastly, *reasoning about legal action* this area is of legal reasoning is concerned with helping an agent to determine the legal status future actions. This area is fairly new and is heavily influenced by autonomous agents. For example, whether driving over a yellow traffic light is legal or not? (Prakken and Sartor 2015; Prakken 2017)

Within *reasoning with the law* several areas of applicability for non-monotonic reasoning exist. Firstly, *rules with exceptions*, i.e. it is common that legal rules contain exception. For example, property damage will be prosecuted *unless* it was acted upon *force majeure*. Here standard non-monotonic techniques are sufficient that is one could even use abnormality theory or default logic. Secondly, *rules with conflicting conclusions*, i.e. it is common that legal rules have conflicting conclusions. Hence, these conflicts have to be resolved. This is accomplished by means of prioritisation based on specificity (*Lex Specialis Derogat Legi Generali*), authority (*Lex Superior Derogat Legi Inferiori*) or recency. In order to express these preferences, rules have to be prioritised, one example of such prioritisation is *preference default logic* (Prakken and Sartor 2015; Prakken 2017; Antoniou and Wang 2007).

While "traditional" non-monotonic, as presented above, are useful in the case of reasoning with the law as it were an axiomatic system, it is limited with respect to other areas of legal reasoning. This is due to the fact that the central notion in law is not necessary deduction, which is still an major part of non-monotonic reasoning, but argument. Therefore, *Argumentation Theory* is widely applied and of elevated importance when it comes to legal reasoning. Formal argumentation theory is also used in non-monotonic reasoning, e.g. as a semantic for default logic and some logic programs. One example of is Dung's *abstract argumentation theory*, which relies on concepts such as arguments as well as the notion of attack and defeat between arguments, in order to generate extensions. Within law it can be used to capture the argumentation behind whether a law should be applied, e.g. *force majeure* defeats the argument for property damage. Moreover, it can also be used to model legal proof, i.e. which source of information defeat other sources, as well as to model the shifting burden of proof, i.e. in the case of a murder accusation, the accuse has to prove intent, then the defendant has to attack this argument by proving self-defence and so on (Prakken and Sartor 2015)

Lastly, it has to be mentioned that this is just a surface level investigation of a topic, providing a vast range of possible applications for non-monotonic reasoning and logic in general. Some actual application of non-monotonic formalisms are *normal conditional logic*, *Defeasible Logic*, *Reason-Based Logic* and *logic programming* (Prakken and Sartor 2015).

5.2   Other Applications

As already eluded non-monotonic reasoning was heavily influenced by problems within the field of artificial intelligence, see the frame problem. Hence, only a few specific applications of non-monotonic reasoning in computer science are discussed. Firstly, *Truth Maintenance Systems* can be seen as one of the first formal approaches trying to capture non-monotonic reasoning, which is closely related to the explanatory approach of non-monotonic reasoning. In fact, default logic can be seen as formalisation of such systems. Secondly, *normal logic programming*, also part of the explanatory approach, uses the concept of negation as failure in order to achieve non-monotonicity and is mostly presented as a formalism for non-monotonic reasoning in itself. Thirdly, as far as databases are concerned one is mostly confronted with a finite universe and unique names, thus the *Closed World Assumption* is especially suitable for the field of *Database Theory*. This allows the answering of queries such as "Does there exists a flight to the moon." with no. (Brewka, Dix, and Konolige 1997; Bochman 2007)

However, there are a variety of other field where non-monotonic reasoning could be beneficial. One example of a field where non-monotonic reasoning is useful is, *cognitive science*, since day human reasoning seems to be defeasible in nature. Combined with the fact that deduction in human reasoning is heavily influenced by biases and context, a strong argument for preferring non-monotonic formalisms over classical logic when attempting formalise parts of human reasoning. McCarthy even claimed that non-monotonicity is essential for intelligent behaviour. One highly interesting result is that it is possible to construct relationships between certain types of neural networks and certain non-monotonic formalisms, thus bringing symbolic and connectionist approaches closer together (Isaac, Szymanik, and Verbrugge 2014; Evans 2002).

Another example is *biology*, a field where categorisation hierarchies with exceptions, e.g. Penguin is a flightless bird, platypus a mammal that lays eggs, are commonly used. Hence, making it necessary to deal with exceptions in a formal manner. This may be achieved by building exception

hierarchies based on default logic (Antoniou and Wang 2007)

6 Conclusion

Non-monotonic reasoning is a diverse and somewhat messy field consisting of various different formalisms with different degrees of material adequacy. While there are two general approaches to non-monotonic reasoning, the distinction between them is somewhat fuzzy. Moreover, given the range of possible applications, it seems to be an important field of study especially since it seems to be a better approximation of human reasoning than classical logic. Lastly, it has to be mentioned that this text is restricted to the classical approaches of non-monotonic reasoning, thus serving only as a small glimpse into the complexity of the field. Yet, hopefully, it provided some context and understanding of how one may achieve non-monotonicity in a formal setting.

References

Antoniou, Grigoris, and Kewen Wang. 2007. "Default Logic". In *The Many Valued and Nonmonotonic Turn in Logic*, ed. by Dov M. Gabbay and John Woods, 8:517–555. Handbook of the History of Logic. North-Holland. doi:`https://doi.org/10.1016/S1874-5857(07)80011-2`. `http://www.sciencedirect.com/science/article/pii/S1874585707800112`.

Bochman, Alexander. 2005. *Explanatory nonmonotonic reasoning*. Vol. 4. World scientific.

– . 2007. "Nonmonotonic Reasoning". In *The Many Valued and Nonmonotonic Turn in Logic*, ed. by Dov M. Gabbay and John Woods, 8:557–632. Handbook of the History of Logic. North-Holland. doi:`https://doi.org/10.1016/S1874-5857(07)80012-4`. `http://www.sciencedirect.com/science/article/pii/S1874585707800124`.

Brewka, Gerhard, Jürgen Dix, and Kurt Konolige. 1997. *Nonmonotonic reasoning: an overview*. Vol. 73. CSLI publications Stanford.

Evans, Jonathan St BT. 2002. "Logic and human reasoning: An assessment of the deduction paradigm." *Psychological bulletin* 128 (6): 978.

Ginsberg, Matthew L, and David E Smith. 1987. "Reasoning about action I: A possible worlds approach". In *The frame problem in artificial intelligence*, 233–258. Elsevier.

Isaac, Alistair MC, Jakub Szymanik, and Rineke Verbrugge. 2014. "Logic and complexity in cognitive science". In *Johan van Benthem on logic and information dynamics*, 787–824. Springer.

Koons, Robert. 2017. "Defeasible Reasoning". In *The Stanford Encyclopedia of Philosophy*, Winter 2017, ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Kraus, Sarit, Daniel Lehmann, and Menachem Magidor. 1990. "Nonmonotonic reasoning, preferential models and cumulative logics". *Artificial intelligence* 44 (1-2): 167–207.

Lifschitz, Vladimir. 2015. "The dramatic true story of the frame default". *Journal of Philosophical Logic* 44 (2): 163–176.

McCarthy, John. 1981. "Circumscription—a form of non-monotonic reasoning". In *Readings in Artificial Intelligence*, 466–472. Elsevier.

McDermott, Drew, and Jon Doyle. 1980. "Non-monotonic logic I". *Artificial intelligence* 13 (1-2): 41–72.

Prakken, Henry. 2017. "On the problem of making autonomous vehicles conform to traffic law". *Artificial Intelligence and Law* 25, no. 3 (): 341–363. ISSN: 1572-8382. doi:`10.1007/s10506-017-9210-0`. `https://doi.org/10.1007/s10506-017-9210-0`.

Prakken, Henry, and Giovanni Sartor. 2015. "Law and logic: a review from an argumentation perspective". *Artificial Intelligence* 227:214–245.

Reiter, Raymond. 1980. "A logic for default reasoning". *Artificial intelligence* 13 (1-2): 81–132.

Shanahan, Murray. 2016. "The Frame Problem". In *The Stanford Encyclopedia of Philosophy*, Spring 2016, ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Strasser, Christian, and G. Aldo Antonelli. 2018. "Non-monotonic Logic". In *The Stanford Encyclopedia of Philosophy*, Summer 2018, ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.