Bulletin of the **S**ection of **L**ogic
Volume 53/4 (2024), pp. 419–453
https://doi.org/10.18778/0138-0680.2024.14

Piotr Łukowski
Konrad Rudnicki

EXPLORING THE DEFINITION OF NON-MONOTONICITY – LOGICAL AND PSYCHOLOGICAL CONSIDERATIONS

Abstract

When humans reason, they are able to revise their beliefs in light of new information and abandon obsolete conclusions. Logicians argued, that in some cases, such reasonings appear to be non-monotonic. Thus, many different, seemingly non-monotonic systems were created to formally model such cases. The purpose of this article is to re-examine the definition of non-monotonicity and its implementation in non-monotonic logics and in examples of everyday human reasoning. We will argue that many non-monotonic logics employ some weakened versions of the definitions of non-monotonicity, since in-between different steps of reasoning they either: a) allow previously accepted premises to be removed, or b) change the rules of inference. Of the two strategies, the second one seems downright absurd, since changing the rules of a given logic is a mere replacement of that logic with the rules of another. As a consequence we obtain two logics, whereas the definition of a non-monotonic logic is supposed to define one. The definition of non-monotonicity does not permit either of these cases, which means that such logics are monotonic.

Keywords: non-monotonicity, reasoning, belief revision.

Presented by: Andrzej Indrzejczak
Received: November 8, 2023
Published online: October 8, 2024

1. Introduction

Humans are relatively proficient at revising their beliefs and behavior based on incoming new information. They are able to abandon old conclusions while seemingly retaining all old premises. For example, I could easily say that *'My apartment is located 5 minutes from the main station by car, we will get there in that time,'* and then upon seeing the traffic correct myself by saying: *'Well, now it is going to take us at least 10 minutes.'* This appears to be a reasoning in which I have learned a new premise *'There is a heavy traffic'* and based on that I have rejected a previously accepted conclusion *'It will take us 5 minutes.'* Thus, I realize that the premise 'My apartment is located 5 minutes from the main station by car,' was actually: 'My apartment is located 5 minutes from the main station by car, with average traffic volume.' This is remarkable because in classical logic all reasonings are *monotonic*, which means that adding new premises to an existing set can never cause a previously accepted conclusion to be abandoned. If something was true based on past information, then adding new information (but not removing old information!) can never render it *untrue*. In fact, any deductive inference appears to be necessarily monotonic [24, p. 223]. Because of that, for some logicians it appeared that monotonicity is not a property of human everyday thinking. To address that, they created *non-monotonic* logics, which were, among other things, intended to model how humans adapt to new information.

The purpose of this article is to critically examine the claim that human reasoning is non-monotonic. In order to do that, we will analyze the relations between: the definition of non-monotonicity, the rules of some non-monotonic logics and psychological data about human belief revision and reasoning. In short, we will argue that: a) human reasoning could **appear to be** non-monotonic at a first glance, b) the analyzed non-monotonic logics do not satisfy the definition of non-monotonicity, c) in empirical studies it is extremely difficult to address the question if humans reason non-monotonically due to formal constraints and the requirements posed by the definition of non-monotonicity.

2. The understanding of 'reasoning' in logic and psychology

In this article we will analyze if it is justified to use the term *non-monotonic* when describing human reasoning.[1] Therefore, we must start from addressing what *reasoning* is. Both psychologists and logicians will agree that reasoning is a process of reaching conclusions from premises. Unfortunately, the details of what that definition entails are going to differ [40]. For logicians, that process is typically going to entail formal manipulations on the truth valuations of propositions. These truth valuations may be expressed just as simple '0's and '1,'s (in classical logic and most other logics), but also other values (e.g., probabilities) [12, 45]. All of these variants consider reasoning to be a manipulation of 'truth-functional' operators. In this sense, contemporary cognitive psychology differs from logic. Psychologist P. N. Johnson-Laird [14, p. 1], one of the most esteemed researchers of logic within human reasoning, wrote:

> *'Thirty years ago psychologists believed that human reasoning depended on formal rules of inference akin to those of a logical calculus. This hypothesis ran into difficulties, which led to an alternative view: reasoning depends on envisaging the possibilities consistent with the starting point – a perception of the world, a set of assertions, a memory, or some mixture of them. We construct mental models of each distinct possibility and derive a conclusion from them. (...) On this account, reasoning is a simulation of the world fleshed out with our knowledge, not a formal rearrangement of the logical skeletons of sentences.'*

In this quote, Johnson-Laird [14] signals that reasoning in cognitive science must necessarily entail processing entire contents of sentences and not just operations on their truth-values. In other words, truth-functional logics were ill-equipped to model human reasoning, because they stripped it of its essential property: processing information [37]. This shift in paradigm that Johnson-Laird [14] describes here resulted in cognitive science moving

[1] The word "reasoning" normally does not have a plural form in the English language. However, in this work "reasoning" is understood more broadly than the traditional use of the word in logic. Namely, we also refer to specific instances of human thought processes, which we call: "reasonings."

away from using truth-functional logics as the proper notation for mental models. In response to that, logicians started developing some non-truth-functional logics (for examples see: [21, 25]. Fortunately, even when reasonings are not truth-functional and cannot be easily expressed with formulaic relations between a handful of truth values, they can still be either monotonic or non-monotonic. Both logic and cognitive science usually agree that *reasoning* involves reaching conclusions from premises. The difference is, that in truth-functional logics this process will just involve applying rules to truth-values of sentences. In contrast, in cognitive science *reasoning* will involve exploration of the content and discovery of information entailed by it [37].

3. Fundamental problems with non-monotonicity

3.1. Definition of monotonicity and non-monotonicity

Inference $\vdash$ is monotonic if and only if for any $\alpha \in For$ and $X, Y \subseteq For$: if $X \vdash \alpha$, then $X \cup Y \vdash \alpha$, where For is the set of formulas of the language. Thus, an inference $\vdash$ is non-monotonic if and only if for some $\alpha \in For$ and $X, Y \subseteq For$: $X \vdash \alpha$ and $X \cup Y \nvdash \alpha$. Equivalently, an inference $\vdash$ is non-monotonic if and only if there are such $\alpha \in For$ and $X, Y \subseteq For$ that $X \vdash \alpha$ but $X \cup Y \nvdash \alpha$. In other words, in monotonic inference/reasoning, if a conclusion α follows from a set of premises X, then it must also follow from any superset of X. The addition of new premises is not able to invalidate anything in the set of the old conclusions. In contrast, in non-monotonic inference it is possible that a conclusion α follows from the set X but does not follow from some superset $X \cup Y$. In that case, adding a new premise may invalidate some of the previously accepted conclusions. Thus, two obvious facts are clear from the definition of non-monotonic inference [20].

The first fact is that the set of premises X must be identical at both of the steps: $X \vdash \alpha$ and $X \cup Y \nvdash \alpha$. It means that even after new premises are added and create a new superset $Z = X \cup Y$, the X symbol must denote exactly the same set of premises in both steps.

The second fact is that the inference '$\vdash$' must also be the same in both of the reasoning steps, which means it is defined by the same set of axioms and inference rules. Otherwise, the definition is not a definition of one inference, but two, since it contains symbols of two different inferences: $X \vdash_1 \alpha$ and $X \cup Y \nvdash_2 \alpha$.

In this work we will point out that most of the allegedly non-monotonic logics violate these conditions from a purely formal perspective [20]. For example, the fundamental works of Makinson [23] on non-monotonic extensions of classical logic have that problem, as well as the popular *System P* [12, 15] or adaptive logics [1]. The creation of these non-monotonic systems is often motivated by the desire to better capture the fact that human everyday reasonings are flexible and deal well with uncertainty [16]. For example, Makinson [23, p. 5] directly claims that: *'We are all non-monotonic'*. Therefore, these systems are good candidates for analyzing if non-monotonic logics really exist.

3.2. Is there such a thing as a 'non-monotonic consequence relation'?

The conditions of reflexivity, monotonicity and idempotence were used by Alfred Tarski (1935, 1936) to define the concept of a consequence relation. Let, as above, *For* be the set of all formulas of some formal language $\mathcal{L}$, then: $C : 2^{For} \to 2^{For}$[2] is a consequence operation if and only if for any $X \subseteq For$:

1. $X \subseteq C(X)$ **reflexivity** of operation C in language $\mathcal{L}$
2. $C(C(X)) \subseteq C(X)$ **idempotence** of operation C in language $\mathcal{L}$
3. if $X \subseteq Y$ then $C(X) \subseteq C(Y)$ **monotonicity** of operation C in language $\mathcal{L}$

X is a set of premises, while $C(X)$ is a set of consequences – conclusions following from X on the basis of operation C. The condition of reflexivity means that all the premises follow from themselves on the basis of C. The condition of idempotence means that repeated application of the operation C on the set $C(X)$ will not expand $C(X)$ with any new conclusions. The condition of monotonicity means that expanding the set of premises cannot shrink the set $C(X)$. Therefore, if someone would like to assume the condition of **non-monotonicity** then they negate the monotonicity condition: for some $X, Y \subseteq For$,

$$X \subseteq Y \text{ and it is not true that } C(X) \subseteq C(Y)$$

[2] 2^X denotes the set of all subsets of the set X.

Thus, the consequence operation by definition assumed monotonicity. As a result, any construction which abandons that condition is not a consequence operation in the Tarskian sense. Instead, it is some other kind of operation. By analogy, the equivalence relation is by definition reflexive, symmetric and transitive. It is not permissible to say that some equivalence relations could be non-transitive, because that would constitute a different type of relation, not equivalence. Similarly, the term *"non-monotonic consequence operation"* violates the basic definition of consequence operation. Non-monotonic operations are operations, but not consequence operations. However, as *'operations'* they can be further analyzed in good faith.

In formal logic it is more common to write about consequence relations than operations. A consequence relation $\vdash$ is equivalent to some consequence operation C: for any $\alpha \in For$ and $X \subseteq For$,

$$X \vdash \alpha \text{ iff } \alpha \in C(X)$$

In that notation, the condition of non-monotonicity takes the, already used above, form: for some $\alpha \in For$ and $X, Y \subseteq For$,

$$X \vdash \alpha \text{ and } X \cup Y \nvdash \alpha$$

4. Hunting for an example of a non-monotonic reasoning

4.1. Common examples from philosophical literature

After all the concepts have been properly defined we can move on to analyzing claims made about non-monotonicity. The inspiration for developing non-monotonic logics was human everyday reasoning. Some examples of such everyday reasoning have become classic in the literature. Namely, *"Tweety the Ostrich"*, *"medical diagnosis"* and *"meeting in the pub."*

Tweety the Ostrich. We know that *Tweety is a bird* (1^{st} premise). Based on the 2^{nd} premise: *birds fly* we conclude that Tweety flies. However, when we later learn that *Tweety is an ostrich* (new premise) we abandon the previous conclusion that Tweety flies.

Medical diagnosis. While observing a patient we notice symptoms a, b and c (1^{st} premise). Based on these symptoms and the medical knowledge about these symptoms (2^{nd} and other premises) we conclude that the patient suffers from the disease z. However, when we later learn that the patient suffers also from the symptom d (new premise) we abandon the previous diagnosis and decide that the patient suffers from the disease y and does not suffer from the disease z.

Meeting in the Pub. John has an appointment with Thomas at a certain time in the pub (1^{st} premise). When the time is near John leaves his house and goes to the pub. However, on his way he receives a message that Thomas had an accident and was taken to the hospital (new premise). Because of that John abandons his plan of going to the pub.

An analysis of these three examples reveals similarities between them. In each of them, reasoning leads to the most *probable* or *expected* conclusion, which could become false if something stands in the way of its truthfulness. There are many possible scenarios in which conclusions like the ones presented could become falsified. Tweety flies, unless it is not an ostrich, penguin, kiwi, has a broken wing, etc. Patient with symptoms a, b, c suffers from the disease z unless they do not also suffer from other symptoms d, e, f. John is going to meet Thomas in the pub unless one of them has an accident, one of them forgets about the meeting, etc. [19].

When formalizing the structure of each of these reasonings let us assume that X is the set of premises, α is a conclusion and β is a premise containing information that *nothing stood in the way.* β is true when every condition for the falsity of α is false, where: δ_i (for $i \in I$), are those conditions. We accept β as long as we do not accept any δ_i. This is necessary because the set $\{\beta, \delta_i\}$ is inconsistent. Therefore, each of the examples presented above can be represented with two steps:

Step one: $X \cup \{\beta\} \vdash \alpha$

Step two: $X \cup \{\delta_i\} \nvdash \alpha$

Such a reasoning cannot be considered non-monotonic because the set of premises $X \cup \{\beta\}$ is not a subset of $X \cup \{\delta_i\}$. The premise β only belongs to the first one and not to the second one. Therefore, the original set of premises was not really expanded with a premise, but instead replaced with a new one, that was not a superset of the old one. Unfortunately, "non-monotonic logicians do NOT want to explicitly represent the "nothing-stood-in-the-way" information given here by the formula β (nor do they believe it is possible)."[3] Meanwhile, it is not difficult to see

[3] The quoted sentence comes from one of the anonymous reviews of this article, hence the capitalization of the word "NOT." The inclusion of the enthymematic condition "unless..." is the essence of the so-called default logics proposed by Reiter [31]. Indeed,

that the enthymematic use of the aforementioned beta condition in the first step of reasoning is obvious and natural, as long as our goal is to represent human everyday thinking.

After taking a brief look at these common examples of alleged human non-monotonic reasonings, it is time to systematically analyze the problem.

4.2. In formal logic

Let us start with the most general issue – can a formal logical system be non-monotonic? Logic can be defined syntactically and semantically. Let us consider both cases. Let $\mathcal{L}$ be a formal language and For a set of all formulas of $\mathcal{L}$.

4.2.1. Logic in syntactic form

Let us consider a logic S with a formal language $\mathcal{L}$ given by a set of axioms A and a set of rules R. The relation $\vdash_S$ of syntactic derivability on the grounds of S we define in a standard way: for any $\alpha \in For$ and $X \subseteq For, X \vdash_S \alpha$ if and only if there exists $(\gamma_1, ..., \gamma_n)$ a finite sequence of language $\mathcal{L}$ formulas so that the last element of that sequence is α: $\gamma_n = \alpha$, and moreover, for any $i \in \{1, ..., n\}, \gamma_i$ is:

1. a formula from the set X (i.e., an assumption), or
2. a formula from the set A (i.e., an axiom), or
3. a formula which is the result of using a rule from the set R on formulas appearing earlier in the sequence than γ_i – which means that they belong to the subsequence $(\gamma_1, ..., \gamma_{i-1})$

The sequence of formulas $(\gamma_1, ..., \gamma_n)$ is the proof of α from set X on the grounds of S. It appears that the inference $\vdash_S$ must be monotonic. Let us assume that $\alpha \in For$ and $X, Y \subseteq For$ fulfill two criteria:

1. $X \subseteq Y$
2. $X \vdash_S \alpha$

---

the mentioned condition is not explicitly represented in default logics. Taking such enthymematic conditions into account in some hidden form is present in almost all supposedly non-monotonic systems.

From the second criterion it follows that there is a sequence $(\gamma_1, ..., \gamma_n)$ that is the proof of α from the set X on the grounds of S. From the first criterion it follows that every formula belonging to X which appears in the proof $(\gamma_1, ..., \gamma_n)$ must also belong to the set Y. That means that exactly the same sequence of formulas $(\gamma_1, ..., \gamma_n)$, that is the proof of α from X is also a proof of α from Y. Therefore, $Y \vdash_S \alpha$.

This demonstrates that the concept of non-monotonicity is incompatible with the standard Hilbertian concept of *proof* (see [21]).

4.2.2. Logic in semantic form

Let us consider logic S with a formal language $\mathcal{L}$ given with a set V of valuations: $v : For \to \{1, 0\}$. The relation $\models_S$ of semantic consequence on the grounds of S is defined in a standard way: for any $\alpha \in For$ and $X \subseteq For, X \models_S \alpha$ if and only if, for any valuation $v \in V$, if v fulfils the set X, then $v(\alpha) = 1$. The valuation v fulfils the set of formulas from X if and only if $v(\beta) = 1$, for any formula $\beta \in X$.

It appears that the inference $\models_S$ must be monotonic. Let us assume that $\alpha \in For$ and $X, Y \subseteq For$ fulfil the criteria: $X \subseteq Y, X \models_S \alpha$. Let us also assume that $X \neq \emptyset$.[4] Let $v(X) = \{v(\beta) : \beta \in X\}$. Then, because of the second condition, for any $v \in V, v(\alpha) = 1$, if $v(X) = \{1\}$. Now, let us assume that for some $v \in V, v(Y) = \{1\}$. Because of the first condition $v(X) = \{v(\beta) : \beta \in X\} \subseteq \{v(\beta) : \beta \in Y\} = v(Y) = \{1\}$. Therefore, $x(X) = \{1\}$, and because of the third condition: $v(\alpha) = 1$. This means that for any $v \in V, v(\alpha) = 1$, if $v(Y) = \{1\}$. Therefore, $Y \models_S \alpha$. When $X = \emptyset$, then α is a tautology of S, so it follows from any set, also from Y.

This demonstrates that the concept of non-monotonicity is incompatible with the concept of semantic consequence. Such proof can be replicated also for logics which semantic interpretation is given by the more general notion of models (see [21]).

The definition of non-monotonicity is extremely hard to fulfil, no matter if a logic is defined syntactically or semantically. Many of the so-called non-monotonic logics are *de facto* monotonic logics that tinker with some of the premises. This tinkering takes various forms, which sometimes gives the impression that old premises are not really removed. However, barring

[4] In other words, the value of a set is either a set with a singleton 1 or a singleton 0 or a set with 0 and 1.

us from using a premise, no matter the reason for it, is equivalent to removing it. Sometimes, instead of blocking premises, the rules of inference change or the "application of rules of inference" changes (e.g., adaptive logics). However, the two steps of non-monotonicity require us to keep all the old premises, as well as all the rules of inference intact. This fundamental incompatibility between the concept of logicality and the concept of non-monotonicity is sometimes acknowledged when researchers point out that the term *non-monotonic logic* should not be used since it is an oxymoron [7].

4.3. Everyday reasoning: the rationale $\implies$ succession conditionals

After discussing the most general formal systems, let us move on to everyday reasoning and consider if non-monotonicity is indeed present there. After all, the fact that non-monotonicity is not easily formalized is not an argument against its presence in real-life human reasonings.

Some fundamental everyday reasonings are traditionally classified into four types: inference, proving, explaining and verification [35, 5, 18]. They are based on the conditional "*rationale* $\implies$ *succession*", which expresses the empirical, analytical, structural, tetical, logical or mixed relations. The decision to accept or reject truthfulness of some of these conditionals is arbitrary, but usually they represent the most commonly accepted ways of forming beliefs. Let us analyze if any of these reasonings can be non-monotonic.

4.3.1. Inference

In the case of inference we accept some sentence R as true and we wonder about its consequences. We find a succession N of R so that the conditional: $R \implies N$ is true. As long as $R \implies N$ and R are accepted as true, we are forced to accept N as true too. However, expanding the set of premises could possibly make us determine that ¬N, in which case the set of conclusions would become inconsistent. Such inconsistency forces us to revise the set of premises and resign from the truth of either $R \implies N$ or R. Such a procedure does not violate monotonicity. In fact, staying with inconsistent conclusions also does not.

4.3.2. Proving

In the case of proving, we wonder if some sentence N is true. To find that out we look for some R which is a true rationale for N, so that the conditional $R \implies N$ is true. If we demonstrate the truth of R, then just like in inference, we are bound to accept N. Similarly, expanding the set of premises can produce inconsistency, but not abandonment of conclusions. After considering the infallible conditionals, let us move on to the fallible ones.

4.3.3. Verification

In the case of verification, we wonder if some sentence R is true. To find that out we look for some true succession N for our R so that the conditional: $R \implies N$ is true. Such a procedure is fallible, since the truth of $R \implies N$ and N does not guarantee the truth of R. Because of that fallibility verification can be an element of a two-step reasoning that generates the impression of non-monotonicity.

Example

Let us imagine that the teacher wants to check if Eve read the book assigned in class. In order to check that, she asks Eve three questions about the content of the book and Eve answers them all correctly. The teacher concludes that Eve read the book. However, guided by intuition, the teacher asks one more question about a very central point of the plot and Eve does not know the answer. The teacher changes her opinion and concludes that Eve did not read the book (maybe she just watched the TV adaptation that changed the plot).

Let us denote:

R = *Eve read the book*

N_1 = *Eve answered 3 questions correctly*

N_2 = *Eve answered 4 questions correctly*

The teacher accepts the truth of relations $R \implies N_1$ and $R \implies N_2$, and also knows N_1 to be true. In a fallible way of verification she concludes that R. However, after asking the fourth question, she learns that N_2 is false and changes her conclusion to $\neg R$. The trick is that in this second step she is no longer using verification, which is fallible, but instead infallible deduction. From $R \implies N_2$ and N_2 infallibly follows $\neg R$.

Step one: $\{R \implies N_1, R \implies N_2, N_1\} \approx_v R$.[5]
Step two: $\{R \implies N_1, R \implies N_2, N_1, \neg N_2\} \models \neg R$
As we see, such an example does not fulfil the criteria of non-monotonicity and the same principle can be applied to other examples of verification. Even though expanding the set of premises causes abandonment of previously accepted conclusions, the two steps of reasoning employ different relations of consequence – fallible verification $\approx_v$ versus infallible deduction $\models$.

4.3.4. Explaining

In the case of explaining we know that some sentence N is true and we wonder why. To explain the truth of N we look for some R so that the conditional $R \implies N$ is true. Knowing only the truth of the conditional and the truth of the succession we conclude that R is true too. However, this is a fallible reasoning. In most reasoning that employ explaining there are multiple rationales for a true succession. Most often there are more than one conditional, $R_1 \implies N, \ldots, R_n \implies N$ and we choose one.

Example

Let us imagine that we work at the office and our colleague John is still not at his desk. We wonder about his absence and conclude that he must be sick. After a while John calls us and tells us that he is stuck in traffic. After the phone call we abandon our earlier conclusion about his illness. Let us denote:

R_1 = *John is sick*
R_2 = *John is stuck in traffic*
N = *John is absent at work*

We accept the truth of relations $R_1 \implies N$ and $R_2 \implies N$. In this example, we also accept that $R_1 \implies \neg R_2$ and $R_2 \implies \neg R_1$. Moreover, we know N to be true. Because we have a choice between R_1 and R_2, we arbitrarily choose R_1 as the rationale for N. However, after talking to John we know R_2 to be true, which falsifies R_1. Just like in the previous example, in the first step we used fallible explaining, and in the second step infallible deduction: from R_2 it follows that R_2.

[5] Fallible reasonings are denoted with $\approx$, since the symbol $\models$ is used for infallible deduction.

Step one: $\{R_1 \implies N, R_2 \implies N, N, R_1 \implies \neg R_2, R_2 \implies \neg R_1\} \approx_E R_1$, (also $\neg R_2$).

Step two: $\{R_1 \implies N, R_2 \implies N, N, R_1 \implies \neg R_2, R_2 \implies \neg R_1, R_2\} \models R_2$, (also $\neg R_1$) and because we do not tolerate contradictions:

$$\{R_1 \implies N, R_2 \implies N, N, R_1 \implies \neg R_2, R_2 \implies \neg R_1, R_2\} \not\models R_1.$$

Similarly as in the case of verification, the condition of non-monotonicity cannot be fulfilled, because we are changing between different relations of consequence. That means that we still perform monotonic reasonings according to the well-known definition.

At their inception, non-monotonic logics were intended to encompass infallible deductive reasonings [11, 41]. These attempts were not as fruitful as originally envisioned, which led David Makinson – the progenitor of non-monotonicity to write that non-monotonicity may only hold in fallible reasonings [24, p. 223]. However, even that seems to encounter difficulties, given that verification and explaining seem to have no inherent need for claims of non-monotonicity. However, let us delve deeper into other types of reasoning in search for non-monotonicity.

4.4. Other types of reasoning

The abovementioned verification and explaining are fallible, but it is still difficult to formulate any example that demonstrates undisputable non-monotonicity. In both examples we presented, the fallible reasoning performed initially is replaced with infallible deduction when new premises arise. Because of that, even though fallible, they are subjected to some logical rigor, since they accept the rule governing an implication, that it is not possible for a true implication to have a true rationale and false succession. This separates them from even more loose types of reasoning, some of which are known as heuristics. In this sense heuristics are reasoning patterns based on fallible rules, which are otherwise known to *usually* provide accurate conclusions.[6] Here we will discuss one in particular, because it shows that non-monotonicity can actually be observed, if one loosens the rigor of reasoning enough.

[6] Rules of reasoning in living organisms are subjected to adaptive pressures, like every other biological trait those organisms may have. As a result, these rules reflect their *utility* for the organism – trade-off between accuracy, speed and resource consumption – not their accuracy alone.

4.4.1. Analogy

Reasoning with analogy is a heuristic based on perceived similarity. From logical perspective it entails ascribing two objects with a shared property and inductively deriving some other properties which these objects are supposed to share [22]. From cognitive perspective it is an act of comparing mental representations which involves their retrieval from long-term memory, identifying elements shared between those representations and inductively deriving new information [13]. Its fallibility is particularly high, even though it is an extremely widespread phenomenon. Therefore, its popularity is not caused by its accuracy, but rather its remarkable ability to start reasoning from a scratch when we have very little information about the subject at hand. In general analogy is a reasoning that takes the form:

It was the case in the past, in some situation S_1 that a, b, c, d.

It is the case now, in some situation S_2 that a, b, c.

Therefore I conclude that in the current situation S_2 it is d as well.

Due to the particularly loose structure of the analogy, we are able to construct reasoning that is indeed non-monotonic.

Example

I remember that in city A the town hall is in the city center and that A is an old city. Therefore, I conclude that in B, which is also an old city, the town hall also has to be in the city center. However, I then also remember that A was never damaged during the war, while B was. I also know that in C, which was damaged during the war, the town hall was moved away from the historical center, to a more modern area. As a result, by analogy between B and C I now conclude that B has its town hall in the modern area and abandon my previous conclusion.

Step one:
a = A is an old city.
b = Town hall in A is in the city center.
a' = B is an old city.
Therefore:
b' = Town hall in B is in the city center.

Step two:
a = A is an old city.
b = Town hall in A is in the city center.
c = A was not damaged during the war.

a' = B is an old city.
c' = B was damaged during the war.
Therefore:
No analogy is made, no conclusion is reached.

Because no analogy was found, step two arguably does not even exist. However, the next step does, given the new information that we learned about the city C. Steps two and three could me merged together, but we keep them separate here for clarity purposes.

Step three:
a = A is an old city.
b = Town hall in A is in the city center.
c = A was not damaged during the war.
e = C is an old city.
f = C was damaged during the war.
g = Town hall in C is in a modern district.
a' = B is an old city.
c' = B was damaged during the war.
Therefore:
g' = Town hall in B is in a modern district.

To summarize, in step one through analogy we have:

$$\{a, b, a'\} \approx_A b'$$

In step three through analogy we have:

$$\{a, b, c, a', e, f, g, e', c'\} \approx_A g'$$

Among the premises of step three there are all the premises from step 1. No premises are abandoned or blocked. Then, because b' and g' are contradictory, we also obtain:

$$\{a, b, c, a', e, f, g, e', c'\} \not\approx_A b'$$

as well as,

$$\{a, b, c, a', e, f, g, e', c'\} \approx_A \neg b'$$

It seems to be a reasonable assumption that heuristics are good candidates for possible reasonings that involve non-monotonicity.

4.5. The three constructions of David Makinson

4.5.1. The first construction

In the first example of allegedly non-monotonic reasoning in this paper we have said that seeing a lot of road traffic may change our assessment of the time needed to arrive somewhere. One could say that the original statement *'My apartment is located 5 minutes from the main station'* could be complemented with a hidden assumption: *'Unless something unusual happens.'* To account for such unspoken premises in reasonings, Makinson [23] proposed an additional set K of background assumptions, which he called *the set of expectations.* Such an idea was known since antiquity where philosophers worked with the concept of *enthymemes,* the premises that are not explicitly stated due to their obviousness.

The non-monotonicity of an inference that employs the set K of expectations was defined in the following way: first, we must define a new consequence relation. Let $\mathcal{L}$ be some language with For a set of all formulas,, where $K \subseteq For$, and Cn be the classical consequence operation. Then, C_K will be the *consequence relation of the axial assumptions* K and $\vdash_K$ the *relation of the axial assumptions* K, if for any $X \subseteq For, \alpha \in For$:

$$\alpha \in C_K(X) \text{ iff } \alpha \in Cn(K \cup X)$$
$$X \vdash_K \alpha \text{ iff } (K \cup X) \vdash \alpha.$$

Then:

$$Cn_K(X) = \{\cap Cn(K' \cup X) : K' \subseteq K \text{ and } K' \text{ is maximally consistent with } X\}$$
$$X \mathrel{|\!\!\sim}_K \text{ iff } (K' \cup X) \vdash \alpha, \text{ for any } K' \subseteq K, \text{ maximally consistent with } X$$

The relations $\mathrel{|\!\!\sim}_K$ are called the *background assumptions consequences.* Based on the way they were just defined, Makinson [23] argues that they are non-monotonic in the following way: let us assume the following set $K = \{p \to q, q \to r\}$. Because $K \cup \{p\}$ is consistent, then just the whole K is the only one maximally consistent with $\{p\}$ subset of K. Therefore, $r \in C_K(\{p\})$, because $r \in Cn(K \cup \{p\}) \neq L$, while at the same time: $r \notin C_K(\{p, \neg q\})$. In fact $K \cup \{p, \neg q\}$ is inconsistent. Moreover, there is only one subset of K, which is maximally consistent with $\{p, \neg q\}$. It is $K' = \{q \to r\}$. It is easy to notice that $r \notin Cn(\{q \to r, p, \neg q\})$. Thus, $r \in C_K(\{p\})$

and $r \notin C_K(\{p, \neg q\})$, although $\{p\} \subseteq \{p, \neg q\}$. As a result we obtain an apparently non-monotonic reasoning where:

$r \in C_K(\{p\})$ (i.e., Proposition r belongs to the set of conclusions following from $\{p\}$ by the rules of C_K)

But at the same time:

$r \notin C_K(\{p, \neg q\})$ (i.e., Proposition r does not belong to the set of conclusions following from $\{p, \neg q\}$ by the rules of C_K)

At first glance it appears that the defined consequence relation is non-monotonic. After all, expanding the set of premises with $\neg q$ just shrunk the possible list of conclusions. However, this construction unfortunately does not satisfy the definition of non-monotonicity. Namely, that the set of premises cannot be changed in other ways than adding new premises to it. In the procedure outlined above for every reasoning step a new *set of expectations* (i.e., the hidden assumptions) is selected. The used notation seems to suggest that the whole *set of expectations* K is used at every step by using the same: C_K everywhere, whereas in fact various subsets of the set K are used. This problem was pointed out by [20] and puts into question the usage of the term: *'non-monotonicity'* for this and similar constructs. However, despite terminological confusions, this logical construction is heavily grounded in our current understanding of human cognition. It was an attempt at capturing one of the many ways in which our everyday reasonings deviate from the predictions of classical logic. Namely, the fact that our beliefs, attitudes, memories and any other construct expressible with propositions does not form a single unified set, but is instead partitioned based on various criteria [32, 42].

The existence of such a partitioning mechanism in cognition is highly useful, since it conserves resources when communicating and when processing information on your own. It would be highly inefficient (if not impossible) from the point of energy expenditure if humans explicitly stated all the premises they used in every reasoning. In fact, cognitive scientists postulate the existence of a hierarchy of beliefs [10]. This hierarchy can take many forms. For example, in many models of thinking, conscious (or *language based*) information processing is considered to run *'on top of'* unconscious stimuli-based processing [38]. Despite being non-verbalized, all the information processed by those evolutionarily older systems can potentially be expressed in symbolic form compatible with formal logic and they certainly influence the way in which humans reason. As a result, any logic that intends to model human thinking should be aware of the existence of information that influences how we reason, but remains unspoken. However, even the hierarchies of beliefs within linguistic cognitive systems are enough to justify the utility of the *set of expectations* by [23], without having to rely on the stimuli-based ones.

In cognitive psychology, researchers use the term *cognitive schema* to describe *'the basic structural components of cognitive organization through which humans come to identify, interpret, categorize and evaluate their experiences'* [34, p. 129]. From a logical perspective, schemas can be seen as sets of beliefs expressed as propositions, partitioned on the basis of their utility in given situations. The same sentence can be understood completely differently in two different contexts, because different *enthymemes/expectations/schemas* are active in those contexts. If a friend calls us *'an idiot'* in a pub, we are significantly less likely to be offended than when a random person on the street does the same. That is because in the cognitive schema relevant for interpreting communications with friends, insults are considered playful and bonding, which is not the case with strangers.

Cognitive schemas are organized hierarchically [10]. At the top of the hierarchy are the *core beliefs*, which are the most basic, central and unquestionable convictions we hold about reality. Researchers believe that people very rarely articulate them, even to themselves. In simple words, these core beliefs describe how we think that the world really *'is'* [4]. From these beliefs, other, more detailed convictions and attitudes are derived and separated into schemas for different situations. Psychotherapists also tend to categorize them into: beliefs about the self, the others and the external world. As a result, it appears that the idea that different premises from the *set of expectations* K should be used depending on the situation, agrees with the current view on how unconscious cognitive schemas guide our reasoning. However, such a procedure/strategy has nothing to do with non-monotonicity.

4.5.2. The second construction

The second construction of Makinson [23] through which he intended to introduce non-monotonicity to modeling human thinking relies on selective usage of Boolean valuations (V). Let $W \subseteq V, X \subseteq For, \alpha \in For$. Then,

$$\alpha \in C_W(X)(X \vdash_W \alpha) \text{ iff for any } v \in W, \text{ if } v(X) = 1, \text{ then } v(\alpha) = 1.$$

Here, the consequence relations C_W (also $(\vdash_W)$) are called *axial-valuation consequences*. These are monotonic, and non-monotonicity is achieved through the introduction of the so-called *preferential model*, which is a set (W) ordered by $<$, an irreflexive and transitive relation on W. If we let $\langle W, <\rangle$ be a preferential model, then

$$X \mid\!\sim_< \alpha \text{ iff } v(\alpha) = 1, \text{ for any } v \in W \text{ minimal among all valuations from } W \text{ satisfying } X.$$

Here, '$\mid\!\sim_<$' are called *preferential consequences* or *default-valuations consequences*. They are shown to be non-monotonic in the following way: assume a language containing only three sentences: p, q, r and let $W = \{v_1, v_2\}$ such that $v_1(p) = v_2(p) = 1, v_1(q) = 0, v_2(q) = 1, v_1(r) = 1, v_2(r) = 0$. The relation $<$ orders the set W as follows: $v_1 < v_2$. Then, $\{p\} \mid\!\sim_< r$. That is because $\{v_1\}$ is the set of all elements minimal among all valuations satisfying $\{p\}$ and $v_1(r) = 1$. However, it is not the case that $\{p, q\} \mid\!\sim_< r$. That is because, it is $\{v_2\}$ that is the set of all elements minimal among all valuations satisfying $\{p, q\}$, and $v_2(r) = 0$. Thus, for some $\langle W, <\rangle, X, Y \subseteq For, \alpha \in For$, it is the case that: $X \mid\!\sim_< \alpha$, but not $X \cup Y \mid\!\sim_< \alpha$ (see: [20] for an in-depth analysis).

Within the sound and complete semantics designed by Makinson, the second construction allows the user of a language to select different rules of inference at different steps of reasoning. This is possible since the set W as well the order on that set is arbitrary. As a result, the second construction does not define one non-monotonic logic, but rather a whole class of them without stable rules and with varying forms of implication. Therefore, just like the first construction, it violates the definition of non-monotonicity because the rules of inference change. The attempt at modeling human thinking via dynamic, changeable sets of premises and rules of inference corresponds to how sensitive our cognition is to different contexts in which reasonings happen.

Context dictates which unspoken schema will guide our inferences and psychologists have shown that it can be easily influenced, creating an impression of non-monotonicity. In an extensive field of research, spanning decades, researchers have shown a robust *framing effect* in risky decision making [36, 6]. It is a phenomenon, where presenting people with logically equivalent information, but expressed in a slightly different way, may completely change the conclusions they derive from it. A famous example was given by Tversky and Kahneman [39] and dubbed *"the Asian disease problem."* In their experiment two groups of people were examined. Both groups were informed that due to an outbreak of a deadly disease 600 people may die. The task of the participants was to choose one of the treatment programs to combat the disease based on the expected number of saved lives. The first group had to make a choice between the following options:

A: *'200 people will be saved'*

B: *'There is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved'*

The second group of participants had to choose between:

C: *'400 people will die'*

D: *'There is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die'*

Despite the fact that options A and C are logically equivalent, presenting the treatment program in a positive language (A) makes 72% of participants choose it, while presenting it in a negative language (C) makes that number only 22%. This problem persists through rigorous methodological control of the ambiguity of the used language, to keep the options presented to participants as undeniably equivalent as possible [6]. According to the allegedly non-monotonic constructions of Makinson [23] as well as the schema theory [27] this *framing effect* and dynamically changing inference principles (i.e., the cognitive schema, for extensive examples see: [17, 9]) can be successfully modeled with a change in the underlying set of expectations K.

4.5.3. The third construction

The third construction of Makinson [23] is intended to capture the human ability of changing the understanding of a sentence at successive stages of reasoning. Such a change might be a minor correction to how we interpret a word, but it is always dictated by some previously accepted premises. The transformation of a proposition is achieved by applying the so-called *rules of sentence conversion*. Every rule has the form of $\langle \alpha, \beta \rangle$ and together they form a set $R \subseteq For^2$. Applying the rules of R to sentences in X yields an *image of X closed on R* set: $R(X) = \{\beta \in For : \langle\alpha, \beta\rangle \in R$ and $\alpha \in X\}$. Because applying some rules could potentially result in introducing inconsistency to the set of premises, they can be used selectively. That selectivity is expressed by ordering the set R and indexing every rule: $\langle R\rangle = \{\langle\alpha_i, \beta_i\rangle : i < \omega\}$. With the help of this set we can now define the *axial-rules consequence* relation for any $X \subseteq For$:

$$Cn_{\langle R\rangle}(X) = \cup\{X_n : n < \omega\}, \text{ where } X_0 = Cn(X) \text{ and}$$
$$X_{n+1} = Cn(X_n \cup \{\beta\}).$$

where $\langle\alpha, \beta\rangle$ is the first rule in $\langle R\rangle$ such that $\alpha \in X_n, \beta \notin X_n$ and β are not inconsistent with X_n (see: [20] for an in-depth analysis).

The fact that rules in $\langle R\rangle$ are ordered means that premises can be effectively changed at different steps in the reasoning. From a psychological perspective this could be another example of the *framing effect* mentioned above alongside the second construction. Furthermore, the fact that change in the interpretation of a proposition happens mid-reasoning reminds of a process known in psychology as *cognitive reappraisal*. Cognitive reappraisal is a *'flexible regulatory strategy that draws on cognitive control and executive functioning to reframe stimuli or situations within the environment to change their meaning and emotional valence'* [43, p. 390]. In other words, cognitive reappraisal happens when we consciously try to reinterpret a situation in the light of new information. For example, when a person is devastated after being fired from their job, they may reappraise the situation and instead of seeing it as a failure, see it as the beginning of a new opportunity to grow. Cognitive reappraisal is different from just simply changing our conclusions based on new information, because it necessarily entails changing the interpretation of some old information. The authors of the allegedly non-monotonic systems focused a lot on the conclusions that change in reasonings, but they failed to see the premises that also change with them. This does not mean their constructions are altogether wrong or useless. In fact, they are useful for modelling, for example, cognitive reappraisal. Something that more classical approaches could not do. However, they are not non-monotonic.

4.6. Adaptive logics

One of the examples of modern logics that were not afraid of addressing the reality of actual human reasonings are adaptive logics. Despite the fact that many logicians hold *psychologism* in low regard, Diderik Batens developed a whole family of logics that were *'intended to explicate actual forms of reasoning'* [1, p. 47] and *'both everyday reasoning and scientific reasoning'* [2, p. 222]. Adaptive logics are defined as logics that *adapt* specifically to the premises of reasonings. Adaptation, from a semantic perspective means that some models of the premises are selected preferentially, depending on the *abnormalities* of those premises. From a proof-theoretic perspective it means that some rules of inference apply depending on the presence or absence of some consequences derived from the set of premises [1]. Defining logic this way is potentially very useful from the perspective of its accuracy in representing human reasoning but raises doubts about meeting the definition of non-monotonicity. This is important, because adaptive logics were created with the intention of being non-monotonic and put non-monotonicity forward as one of their central concepts [1, 2].

The rationale for making adaptive logics allegedly non-monotonic is based on the existence of *external* and *internal* dynamics in reasonings [1]. External dynamics are the concept that has been discussed in this paper many times already: the fact that when new premises become known, old conclusions can be withdrawn. In contrast, internal dynamics describe that even if premises do not change, conclusions can change at different stages of reasoning.

Let us consider the way adaptive logics are formalized and then identify the specific points in which they become unwillingly monotonic. A so-called *flat* adaptive logic is characterized by:

1. A *Lower limit logic* – any monotonic logic
2. A *set of abnormalities* – a set of formulas characterized by a logical form
3. An *adaptive strategy* – a description of how to interpret the premises

The first part of the adaptive logic – its *lower limit logic* defines the part that does not adapt itself to the premises. It could be classical logic but also any other logic, for example, some paraconsistent logic. From a semantic perspective, the rules of the adaptive logic AL are therefore a superset of rules from the lower limit logic, $Cn_{LLL}(X) \subseteq Cn_{AL}(X)$. The *set of abnormalities* Ω *'comprises the formulas that are presupposed to be false, unless and until proven otherwise.'* [1, p. 48]. This is extremely reminiscent of the *set of expectations* in the first construction by Makinson [23]. Both contain formulas which are going to be blocked under some specific circumstances. However, instead of classically understood presuppositions, Ω deals with formulas that are in force if and only if they are not contradicted by the set of premises. This is explained by introducing another concept: the *upper limit logic*. The upper limit logic is obtained by extending the lower limit logic with the requirement that no abnormalities from the set Ω are logically possible. The upper limit logic requires premise sets to be free from *abnormalities* and if there are any, it trivializes the set of conclusions (i.e., the principle of explosion). For example, an adaptive logic can be constructed so that if the lower limit logic is set to be the classical logic and the set of abnormalities Ω contains formulas of the form: $\exists\alpha \wedge \exists\neg\alpha$ ($\exists\alpha$ is an abbreviation of the existential closure of α), then the upper limit logic is classical logic extended with the axiom: $\exists\alpha \supset \forall\alpha$ [1].

In consequence, if the set of premises does not contain any abnormality, then the conclusions derived with adaptive logic ($\vdash_{AL}$) are going to be identical to the conclusions derived with the upper limit logic ($\vdash_{ULL}$). However, as soon as a new premise is added, so that it satisfies one of the abnormality formulas from Ω, then adaptive logic is going to deviate from the upper limit logic. The author states that *'it avoids abnormalities 'in as far as' the premises permit'* [1, p. 49]. However, this means that if during reasoning we add a new premise that satisfies a formula from Ω, we change the rules of inference. The change in those rules follows a pattern, which is described by the adaptive strategy of the logic AL, but the change happens nonetheless.

For example, if the lower limit-logic is set to be some paraconsistent logic PL, which is a fragment of CL and the set Ω consists of formulas that take the form: $\exists(\alpha \wedge \neg\alpha)$, then the upper limit logic of that adaptive logic AL will be CL. This means that if the premise set X contains some formulas that take the form of some elements of set of abnormalities Ω then the adaptive logic AL will deliver more consequences than the lower limit logic. Namely, all the consequences from the upper limit logic that are not *blocked* by the abnormalities from Ω [3]. The key term here is the word *blocked*, because in order to block a consequence that was previously derived it necessarily means to block a rule of inference that was previously used. The claim to non-monotonicity in this case comes from the fact that the *adaptive strategy* of a given adaptive logic AL is specified upfront and defines how and when some rules of inference will be used or not. As a result, if we denote the dynamic, non-monotonic nature of adaptive logics by saying that if there are some: X, Y and α such that: $X \vdash_{AL} \alpha$ and $X \cup Y \nvdash_{AL} \alpha$, then it is questionable whether indexing the $\vdash$ with AL means the same thing at both stages. Thanks to the fact that we specify Y we can reconstruct which rules of inference does $\vdash_{AL}$ use at each stage[7], but they are going to be different depending on the stage. In fact, while at a first glance Y appears to be merely a set of premises added to the X, in reality it also alters the logic operating behind $\vdash_{AL}$. By *alter* we mean here that it defines a selection of the rules of inference that are allowed or disallowed. It selects them in accordance with the *adaptive strategy*, but the resulting set of rules is different nonetheless. **We understand that barring a rule from being applied or premise from being used is the same as removing that rule or premise altogether.** The relevance of that postulate is most visible when confronted with the way adaptive logics describe the effects of using the set of abnormalities. Rule of inference can be barred from applying: *"Put differently, that the premises have certain consequences may prevent a rule of inference to be applicable to some other consequences of the premises"* [1, p. 46], and premises can be removed: *"The set of abnormalities (...) Ω comprises the formulas that are presupposed to be false, unless and until proven otherwise"* [1, p. 48]. If a formula is *"presupposed"* to be either true or false, then that formula is effectively a premise in reasoning, even if not explicitly named that way. If the original presupposition changes at some later stage of reasoning, then that premise has changed and the reasoning cannot satisfy the definition of non-monotonicity.

This means that the definition of monotonicity is satisfied here, because dynamically changing rules of inference and/or dynamically changing premises are still used in a monotonic manner – every new rule defines a new logic which is clearly monotonic. As a result, any given AL can be seen not as a single logic, but as a formal system of some kind, from which one logic

[7] In adaptive logics, the term "stage" has a strict meaning as "stage of proof," which is different from the common language "stage of reasoning," see: *"(...) if A is "derivable at a stage" from Y, there is a proof from Γ and a stage s such that A is derived at stage s of that proof"* [1, p. 58].

is being selected based on Y. Both upper and lower-limit logics are monotonic, but the impression of non-monotonicity is created via the addition of the *adaptive strategy*. However, it is important to note that this issue is merely terminological and does not question neither the utility of adaptive systems nor their construction. Instead, we argue that adaptive systems achieve the appearance of non-monotonicity using only monotonic logics.

4.7. System P

One of the most widely endorsed allegedly non-monotonic systems is the *System P* proposed by Kraus et al. [15]. System P is thought to achieve non-monotonicity through the use of preferential models, which provide a formal framework for reasoning about plausibility and normality among possible worlds. In this framework, conclusions are drawn based not on the entirety of possible worlds that satisfy a given set of premises, but rather on a dynamically determined subset of these worlds – the so-called "preferred" worlds. The preference ordering among worlds, denoted as $\prec$, captures the relative normality or plausibility of different scenarios.

In preferential models, a conditional assertion $\alpha \sim \beta$ is interpreted to mean that in all of the most preferred worlds satisfying α, the formula β holds. The key feature of this system is that the set of most preferred worlds satisfying α may shift when additional premises are introduced. This mechanism permits the invalidation of previously drawn conclusions, seemingly enabling non-monotonic reasoning. Specifically, while $\alpha \sim \beta$ may hold under a given set of premises, the introduction of new information, such as γ, can alter the set of preferred worlds satisfying α and, consequently, the validity of β in this revised context.

To illustrate this formally, let $W = \langle S, l, \prec \rangle$ denote a preferential model, where S is the set of states (each corresponding to a possible world), $l : S \to U$ is a labeling function mapping states to worlds, and $\prec$ is a strict partial order representing the preference relation. A state $s \in S$ is said to satisfy a formula α if and only if the world $l(s)$ satisfies α. The conclusion $\alpha \sim \beta$ holds if and only if for all s that are minimal with respect to $\prec$ in the set of states satisfying α, $l(s)$ also satisfies β. When an additional premise γ is introduced, the set of minimal states satisfying $\alpha \wedge \gamma$ may exclude some states that were previously minimal for α. If these excluded states were crucial for supporting the conclusion β, the inference $\alpha \sim \beta$ will no longer hold.

Similarly to adaptive logics and to the three constructions of David Makinson, the dynamic nature of the preferential consequence relation hides the fact that it should be indexed with the underlying preferential model. A conclusion $\alpha \mid\!\sim_{W_1} \beta$ is valid only within the specific preferential model W_1. When new premises lead to a shift in the model (e.g., from W_1 to W_2), the consequence relation should change accordingly (i.e., to $\mid\!\sim_{W_1}$), and previously valid conclusions may no longer hold. Thus, the apparent non-monotonicity is, in fact, a result of changing the rules of inference via a shift in the model, rather than a violation of monotonicity within any fixed model.

The correspondence between this approach and our everyday reasoning is expressed in the way authors interpret the non-monotonic inference relation $'\mid\!\sim'$. If we write: $a \mid\!\sim c$ then we read it as: 'if a then *normally* c.' In our example, 'if my apartment is reachable in 5 minutes from the station by car then we *normally* will get there in time.' On a practical note, the probability semantics by Gilio [12] interpret the conditional $'\mid\!\sim'$ with probability intervals, making the phrase *'normally'* easier to study empirically. *'Normally'* then means *with high probability*. The required probability (x) is arbitrary and expressed with an interval $[x_*, x^*]$, creating a probabilistic consequence relation:

$$a \mid\!\sim_x c \text{ is interpreted as } P(c \mid a) \in [x_*, x^*]$$

For an extensive overview of the probabilistic interpretations of non-monotonic logics see: [26, 29].

5. Discussion

So far we have established several clues on the way to determining if the label of non-monotonicity can be assigned to human everyday reasoning. We know that numerous inferences that humans routinely perform violate the predictions of classical logic. Humans are context sensitive and within one reasoning they are able to switch between different sets of premises, different rules of inference and revise previously accepted conclusions. As a result, logics that call themselves non-monotonic perform well at predicting the outcomes of some human inferences. For example, their rules make it so that the conjunction fallacy stops seeming like a fallacy and presents itself as a rational decision making process in an uncertain environment [28]. However, we have also established that removing or changing premises as well as changing rules of inference cannot be a part of a non-monotonic system. Monotony is violated if and only if addition of a new premise invalidates past conclusions.[8]

When summarizing the selected non-monotonic systems it is important to note that many of them were created with the intention of modeling deductive reasonings [15, 23]. However, due to frequent problems these systems have with following the definition of non-monotonicity [20] and their use of defeasible inference, it appears that they are better suited to model solely abductive reasonings and resign from the ambition of deductivity. For example, Makinson [24, p. 223] consciously noted that: *'While monotony holds for deductive inference, (...) it is quite unacceptable for non-deductive reasoning, whether probabilistic or expressed in qualitative terms,'* thus admitting that monotony holds for deduction.

In the face of these issues with allegedly non-monotonic systems, how can we answer the question: is human reasoning non-monotonic? To make a claim: *'We are non-monotonic'* we cannot just rely on the fact that some allegedly non-monotonic logics are better than monotonic logics at predicting some heuristics. Science already knows many examples of phenomena that are convieniently modeled with some paradigm, even though we know that its rules do not correspond well with reality. For example, Newtonian physics is still the most useful way of predicting physical phenomena on medium-size scale, even though our understanding of physics has moved way past beyond them. Given that many allegedly non-monotonic logics struggle to satisfy the definition of non-monotonicity, we are facing a very difficult conundrum in trying to answer if humans reason non-monotonically. Some non-monotonic logicians are very aware of that fact. For example, Pfeifer and Douven [30, p. 108] summarized their experimental results that showed agreement between System P and empirical data by saying: *'It would be misleading, though, to speculate that our subjects have a 'nonmonotonic inference engine' in their minds that processes incomplete uncertain information. Even if human subjects were perfect in handling the axioms and some elementary theorems of System P, they would not necessarily be able to handle more complex tasks.'*

[8]Abandoning a conclusion after nothing new was added would also constitute a violation of monotony as it can be expressed via an addition of an empty set.

Despite those issues, in this article we have identified a very promising candidate for strict non-monotonicity: reasoning with analogy. However, a question remains: is such a reasoning typical or rather an outlier? This question is very difficult to answer even when using the state-of-the-art neuroscientific tools which are able to track information processing when it unfolds (i.e., functional magnetic resonance, electroencephalography, magnetoencephalography). These techniques are not able to track the neuronal symbolic representation well enough to say which premise was used and which is not when people reach a conclusion. That means that we would not be able to tell if a reasoning analyzed with these techniques satisfied the definition of non-monotonicity. Naturally, by saying that, we admit that we consider reasoning to be a phenomenon of brain activity and that the structure of reasoning is represented by patterns of that activity.

It might seem suspicious that we say that neuroscientific tools are ill-equipped to answer if humans reason non-monotonically. After all, are not most studies in cognitive science and experimental philosophy performed by analyzing participants' responses to carefully crafted questions or stimuli [44]? Why would non-monotonicity be different? The answer lies once again in the fact that we are now concerned with the definition of non-monotonicity. We are specifically interested in identifying a reasoning where absolutely all original premises are fixed throughout the whole reasoning. That means that our neuroscientific tools would need spatio-temporal resolution high enough to locate and track every single individual premise, as they are represented by the brain. Such technology does not exist yet. Unfortunately, including traditional methods and just asking people about their reasoning does not help either. The *'hidden/enthymematic'* premises that people use at different stages of reasoning are mostly unconscious and unspoken. As a result, to identify them we cannot just rely on what people say, but have to analyze the neurophysiological trace of their unconscious reasoning.

Not every property of reasoning has to fulfill such steep empirical requirements to be falsified. For example, paraconsistency (i.e., a property characterizing reasonings that tolerate contradictory premises) is easier to investigate because it only requires a single pair of inconsistent premises to exist within one reasoning [33]. Examining a single pair of premises gives experimentalists the ability to forcibly present them to research participants as experimental stimuli. Then, the brain activity in response to these two particular inconsistent premises may be examined. In the case of non-monotonicity such an approach is impossible, because we need to track every single premise that may or may not have been used in a reasoning.

However, despite the fact that we are currently unable to investigate the question if our reasoning is non-monotonic, we are able to investigate some predictions of allegedly non-monotonic logics. For example, Da Silva et al. [8, p. 110] justify their empirical investigations of the rules of system P by saying: *'No current experimental device can provide relevant and direct observation of the human inferential system 'at work'. Yet, we are able to observe the conclusions derived by human participants in the context of a given set of premises.'* This is true because we do not have to keep track of all the premises to test the effects of some particular logical rules. Only monotonicity itself poses an exceptional challenge. In fact, Da Silva et al. [8, p. 110] were already partially aware of that problem since they remarked immediately afterwards: *'In other words, we do not see these patterns as direct inference rules(...), but as general emerging properties of the inferential apparatus.'*, indicating that they are not studying how the human reasoning *really* works, but instead what patterns emerge from the answers of participants.

The answer to the question: *do humans reason non-monotonically?* is entirely open. It will remain that way until our neuroscientific tools become even more accurate. Despite that, the existing allegedly non-monotonic systems have done an excellent work at pointing out the differences between the classical logic and the way in which humans reason. However, the definition of non-monotonicity appears to be so fundamental that creating a logic which would completely satisfy it, is hard to achieve.

6. Conclusion

There are many very interesting formal systems that have been/are being developed under the banner of non-monotonicity. There are many examples that illustrate peoples' abandonment of previously derived conclusions and are also classified as examples of non-monotonicity in our thinking. The purpose of this work is not to question the value of the systems discussed here – their value is indisputable. Nor is the purpose of this paper to question the fact that people sometimes reject beliefs that they themselves once arrived at. The purpose of this paper is to show that both systems considered non-monotonic and examples of supposed non-monotonicity in human thinking do not satisfy the definition of non-monotonicity. It seems that the condition of the monotonicity expresses such a fundamental property of our thinking that:

1. We still do not have any formal system satisfying the definition of non-monotonicity.

2. The examples of human reasoning widely cited in logical literature as being non-monotonic also do not satisfy the definition of non-monotonicity.

Naturally, the fact that we "still" do know neither a non-monotonic system nor a well-established case of non-monotonic thinking does not mean that we will never know one. However, the scale of attempts to construct non-monotonic logics, as well as the multitude of examples of allegedly non-monotonic human thinking, may suggest that monotonicity is an unassailable principle of our thinking. Perhaps we should start using a more appropriate term to describe reasoning that abandons previously deduced conclusions. Such a change in nomenclature would be advisable, as the current common use of the term "non-monotonic" is misleading in suggesting something that might be not realizable. Perhaps "self-corrective" would be a good candidate to replace the unfortunate "non-monotonic"?

Acknowledgements. We would like to extend thanks to Bert Leuridan for his valuable input and help with writing the manuscript.

References

[1] D. Batens, *A general characterization of adaptive logics*, **Logique et Analyse**, vol. 44(173/175) (2001), pp. 45–68.

[2] D. Batens, *A universal logic approach to adaptive logics*, **Logica Universalis**, vol. 1(1) (2007), pp. 221–242, DOI: https://doi.org/10.1007/s11787-006-0012-5.

[3] D. Batens, *The need for adaptive logics in epistemology*, [in:] S. Rahman, J. Symons, D. M. Gabbay, J. P. van Bendegem (eds.), **Logic, Epistemology, and the Unity of Science**, Springer, Dordrecht (2009), pp. 459–485, DOI: https://doi.org/10.1007/978-1-4020-2808-3.

[4] A. T. Beck, D. A. Clark, *Anxiety and depression: An information processing perspective*, **Anxiety Research**, vol. 1(1) (1988), pp. 23–36, DOI: https://doi.org/10.1080/10615808808248218.

[5] J. M. Bocheński, S. Judycki, **Współczesne metody myślenia (Contemporary ways of thinking)**, W drodze, Poznań (1992).

[6] C. F. Chick, V. F. Reyna, J. C. Corbin, *Framing effects are robust to linguistic disambiguation: A critical test of contemporary theory*, **Journal of Experimental Psychology: Learning, Memory, and Cognition**, vol. 42(2) (2016), p. 238, DOI: https://doi.org/10.1037/xlm0000158.

[7] A. M. Colman, **A Dictionary of Psychology**, Oxford University Press (2016).

[8] N. R. Da Silva, J. F. Bonnefon, E. Raufaste, *An empirical test of patterns for nonmonotonic inference*, **Annals of Mathematics and Artificial Intelligence**, vol. 34(1) (2002), pp. 107–130, DOI: https://doi.org/10.1023/A:1014473706150.

[9] T. De Haan, R. Van Veldhuizen, *Willpower depletion and framing effects*, **Journal of Economic Behavior & Organization**, vol. 117 (2015), pp. 47–61, DOI: https://doi.org/10.1016/j.jebo.2015.06.002.

[10] D. J. Dozois, A. T. Beck, *Cognitive schemas, beliefs and assumptions*, [in:] K. S. Dobson, D. J. A. Dozois (eds.), **Risk Factors in Depression**, Elsevier Academic Press (2008), pp. 119–143.

[11] D. M. Gabbay, *Intuitionistic basis for non-monotonic logic*, [in:] D. W. Loveland (ed.), **6th Conference on Automated Deduction. CADE 1982**, vol. 138 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (1982), pp. 260–273, DOI: https://doi.org/10.1007/BFb0000064.

[12] A. Gilio, *Probabilistic reasoning under coherence in System P*, **Annals of Mathematics and Artificial Intelligence**, vol. 34(1) (2002), pp. 5–34, DOI: https://doi.org/10.1023/A:1014422615720.

[13] K. J. Holyoak, *Analogy and relational reasoning*, [in:] K. J. Holyoak, R. G. Morrison (eds.), **The Oxford Handbook of Thinking and Reasoning**, Oxford University Press, New York (2012), pp. 234–259.

[14] P. N. Johnson-Laird, *Mental models and human reasoning*, **Proceedings of the National Academy of Sciences**, vol. 107(43) (2010), pp. 18243–18250, DOI: https://doi.org/10.1073/pnas.1012933107.

[15] S. Kraus, D. Lehmann, M. Magidor, *Nonmonotonic reasoning, preferential models and cumulative logics*, **Artificial Intelligence**, vol. 44(1–2) (1990), pp. 167–207, DOI: https://doi.org/10.1016/0004-3702(90)90101-5.

[16] G. Kuhnmünch, M. Ragni, *A Formal Non-monotonic Systems Properly Describe Human Reasoning?*, [in:] **Proceedings of the Annual Meeting of the Cognitive Science Society**, vol. 36 (2008), pp. 1222–1227.

[17] S. Lerner, *A non-monotonic intensional framework for framing effects*, **Journal of Economic Methodology**, vol. 21(1) (2014), pp. 37–53, DOI: https://doi.org/10.1080/1350178X.2014.886469.

[18] J. Łukasiewicz, *O nauce (About science)*, [in:] **Poradnik dla Samoukow. Wskazowki metodyczne dla studiujacych poszczegolne nauki (Handbook for the self-taught. Methodical tips for those who study science)**, vol. 1, Wydaw. A. Heflicha i St. Michalskiego (1915), pp. 15–39.

[19] P. Łukowski, *Is human reasoning really nonmonotonic?*, **Logic and Logical Philosophy**, vol. 22 (2013), pp. 63–73, DOI: https://doi.org/10.12775/LLP.2013.004.

[20] P. Łukowski, *Logical problems with nonmonotonicity*, **Logic and Logical Philosophy**, vol. 23(2) (2014), pp. 171–188, DOI: https://doi.org/10.12775/LLP.2014.004.

[21] P. Łukowski, **Logika praktyczna z elementami wiedzy o manipulacji**, Wolters Kluwer (2021).

[22] F. Macagno, *The logical and pragmatic structure of arguments from analogy*, **Logique et Analyse**, vol. 240 (2017), pp. 465–490, DOI: https://doi.org/10.2143/LEA.240.0.3254093.

[23] D. Makinson, *Bridges between classical and nonmonotonic logic*, **Logic Journal of IGPL**, vol. 11(1) (2003), pp. 69–96, DOI: https://doi.org/10.1093/jigpal/11.1.69.

[24] D. Makinson, **Sets, logic and maths for computing**, Heidelberg, Berlin: Springer, Cambridge (2008), DOI: https://doi.org/10.1007/978-3-030-42218-9.

[25] J. Marcos, *What is a non-truth-functional logic?*, **Studia Logica**, vol. 92(2) (2009), pp. 215–240, DOI: https://doi.org/10.1007/s11225-009-9196-z.

[26] M. Oaksford, N. Chater, **Bayesian Rationality: The Probabilistic Approach to Human Reasoning**, Oxford University Press (2007).

[27] H. Okder, *The illusion of the framing effect in risky decision making*, **Journal of Behavioral Decision Making**, vol. 25(1) (2012), pp. 63–73, DOI: https://doi.org/10.1002/bdm.715.

[28] N. Pfeifer, **Psychological investigations on human nonmonotonic reasoning with a focus on System P and the conjunction fallacy**, Master's thesis, Institut für Psychologie, Universität Salzburg, Salzburg (2002).

[29] N. Pfeifer, I. Douven, *Formal epistemology and the new paradigm psychology of reasoning*, **Review of Philosophy and Psychology**, vol. 5(2) (2014), pp. 199–221, DOI: https://doi.org/10.1007/s13164-013-0165-0.

[30] N. Pfeifer, G. D. Kleiter, *Coherence and nonmonotonicity in human reasoning*, **Synthese**, vol. 146(1) (2005), pp. 93–109, DOI: https://doi.org/10.1007/s11229-005-9073-x.

[31] R. Reiter, *A Logic for Default Reasoning*, **Artificial Intelligence**, vol. 13 (1980), pp. 81–132, DOI: https://doi.org/10.1016/0004-3702(80)90014-4.

[32] N. Rescher, **Introduction to Value Theory**, University Press of America (1969).

[33] K. Rudnicki, *The psychological aspects of paraconsistency*, **Synthese**, vol. 199 (2021), pp. 4393–4414, DOI: https://doi.org/10.1007/s11229-020-02983-8.

[34] N. B. Schmidt, K. L. Schmidt, J. E. Young, *Schematic and interpersonal conceptualizations of depression: An integration*, [in:] T. Joiner, J. C. Coyne (eds.), **The Interactional Nature of Depression**, American Psychological Association, Washington, DC (1999), pp. 127–148, DOI: https://doi.org/10.1037/10311-000.

[35] B. Sobociński, *Kazimierz Ajdukiewicz. Klasyfikacja rozumowań (Classification of reasonings)*, **The Journal of Symbolic Logic**, vol. 21(3) (1956), pp. 311–312, DOI: https://doi.org/10.2307/2269111.

[36] A. Steiger, A. Kühberger, *A meta-analytic re-appraisal of the framing effect*, **Zeitschrift für Psychologie**, vol. 226(1) (2018), p. 45, DOI: https://doi.org/10.1027/2151-2604/a000321.

[37] K. Stenning, M. van Lambalgen, *Interpretation, representation, and deductive reasoning*, [in:] J. E. Adler, L. J. Rips (eds.), **Reasoning: Studies of human inference and its foundations**, Cambridge University Press (2019), pp. 223–248.

[38] F. Toates, *A model of the hierarchy of behaviour, cognition, and consciousness*, **Consciousness and Cognition**, vol. 15(1) (2006), pp. 75–118, DOI: https://doi.org/10.1016/j.concog.2005.04.008.

[39] A. Tversky, D. Kahneman, *The framing of decisions and the psychology of choice*, **Science**, vol. 211(4481) (1981), pp. 453–458, DOI: https://doi.org/10.1126/science.74556.

[40] D. N. Walton, *What is reasoning? What is an argument?*, **The Journal of Philosophy**, vol. 87(8) (1990), pp. 399–419, DOI: https://doi.org/10.2307/2026735.

[41] H. Wansing, *Semantics-based nonmonotonic inference*, **Notre Dame Journal of Formal Logic**, vol. 36(1) (1995), pp. 44–54, DOI: https://doi.org/10.1305/ndjfl/1040308828.

[42] M. Werning, S. Cheng, *Taxonomy and unity of memory*, [in:] S. Bernecker, K. Michaelian (eds.), **The Routledge Handbook of Philosophy of Memory**, Routledge (2017), pp. 7–20, DOI: https://doi.org/10.4324/9781315687315.

[43] P. B. Williams, H. C. Nusbaum, *Toward a neuroscience of wisdom*, [in:] J. R. Absher, J. Cloutier (eds.), **Neuroimaging Personality, Social Cognition, and Character**, Academic Press (2016), pp. 383–395, DOI: https://doi.org/10.1016/C2013-0-18623-0.

[44] J. T. Wixted, S. L. Thompson-Schill, **Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Language and Thought**, vol. 3, John Wiley & Sons (2018), DOI: https://doi.org/10.1002/9781119170174.

[45] L. A. Zadeh, *Fuzzy logic*, **Computer**, vol. 21(4) (1988), pp. 83–93, DOI: https://doi.org/10.1109/2.53.

Piotr Łukowski
Jagiellonian University
Department of Logic
Kraków, Poland
e-mail: piotr.lukowski@uni.lodz.pl

Konrad Rudnicki
University of Antwerp
Department of Communication Science
Antwerp, Belgium
e-mail: konrad.rudnicki@uantwerpen.be