Logic and Logical Philosophy
Volume 31 (2022), 501–527
DOI: 10.12775/LLP.2021.016

Pawel Pawlowski and Rafal Urbaniak

Informal Provability, First-Order BAT Logic and First Steps Towards a Formal Theory of Informal Provability

Abstract. **BAT** is a logic built to capture the inferential behavior of informal provability. Ultimately, the logic is meant to be used in an arithmetical setting. To reach this stage it has to be extended to a first-order version. In this paper we provide such an extension. We do so by constructing non-deterministic three-valued models that interpret quantifiers as some sorts of infinite disjunctions and conjunctions. We also elaborate on the semantical properties of the first-order system and consider a couple of its strengthenings. It turns out that obtaining a sensible strengthening is not straightforward. We prove that most strategies commonly used for strengthening non-deterministic logics fail in our case. Nevertheless, we identify one method of extending the system which does not.

Keywords: informal provability; non-deterministic logic; BAT logic

1. Motivations

1.1. Informal provability vs. formal provability

BAT and its extension **CABAT** are logics that have been developed to model the inferential behavior of the informal notion of provability in classical mathematics. Roughly speaking, informal provability is associated with mathematical practice and so-called informal proofs, and it is usually contrasted with formal proofs understood as syntactical derivations in an axiomatic system. Informal proofs are those which are actually used in mathematical practice. Mathematicians spell out such proofs in a mixture of formal and natural languages. Each inference

Received October 22, 2019. *Accepted* June 25, 2021. *Published online* November 29, 2021

step in such a proof is made by means of widely accepted mathematical means.[1]

On the other hand, *formal proofs* are always relative to a formal system. A sentence φ in the language of the system is formally provable iff there is a proper derivation (formal proof) in this system whose last element is φ. It's widely accepted that derivations in some formal systems (usually, some axiomatic versions of set theory) can be taken as informal proofs. The question whether every informal proof can be somehow treated as a formal proof is a bit more controversial.

The proponents of the so-called *standard view*[2] claim that, at least in principle, every informal proof can be translated into a fully formal proof in a preferred axiomatic theory (usually a variation of set theory). There are good reasons to think that this claim is at least not obvious:

1. Antonutti Marfori [2010] claims that there is no clear algorithm for converting a given informal proof into a proper proof in a relevant axiomatic system.

2. Tanswell [2015] claims that it is not obvious how we can identify different informal proofs with their translations.

3. Rav [1999, 2007] discusses the epistemological and explanatory superiority of informal proofs over formal ones, arguing that this supe-

[1] We are well aware of the fact that mathematics as a whole is not a unified discipline. Different mathematical sub-disciplines may rely on specific mathematical methods unavailable in the others. Still, we think that there is a particular commonly accepted core of mathematical techniques. This core may change over time. Usually, Usually, a change to the core happens through an increase in the rigor and the level of precision in proofs, or through the addition of new mathematical axioms. In both cases, the extension of provability is preserved, since for any mathematical claim whose prove does not meet the current standards of precision, one can formulate a new informal proof with a sufficient level of rigor. As for computer-assisted proofs, we do not have a strong position. They can simply be accepted as a way of available proving methods.

[2] This view is usually shared by mathematicians. For instance Enderton [1977, 10–11] says:

> It is sometimes said that "mathematics can be embedded in set theory." This means that mathematical objects (such as numbers and differentiable functions) can be defined to be certain sets. And the theorems of mathematics (such as the fundamental theorem of calculus) then can be viewed as statements about sets. Furthermore, these theorems will be provable from our axioms. Hence our axioms provide a sufficient collection of assumptions for the development of the whole of mathematics — a remarkable fact. (In Chapter 5 we will consider further the procedure for embedding mathematics in set theory.)

riority is not convincingly explained by the proponents of the standard view.

4. Leitgeb [2009] observed that these concepts of proofs are different. While in formal proofs, the language is precisely defined and divided according to logical order, informal proofs are stated in a natural language expanded with additional mathematical vocabulary. Moreover, the connection between steps in an informal proof has a different nature than in the formal one. The former often employs steps that are supposed to be intuitively seen as truth-preserving, without explicitly following syntactically formulated rules of inference and the latter is based purely on syntactical proofs forming rules.[3]

For our purpose the crucial difference lies within the logic of informal provability. An important inference pattern for informal provability is the *reflection schema*.[4] Roughly, it says that whatever is provable, is true. It is a well-known fact that there is no consistent formal theory extending Peano arithmetic in which all instances of the reflection schema for its own formal provability predicate are provable [Montague, 1963; Myhill, 1960]. So, it seems that the informal notion of provability cannot be formally represented in the standard setting.

To be fair, some attempts to capture informal provability have been made. They can be divided into two groups: those which treat informal provability as a **S4** modality,[5] and those that treat it as a predicate [see Horsten, 2002]. Theories of the first type severely limit the expressive power, whereas the second type of theories starts from technical solutions that seems to be a bit philosophically unmotivated.

The main aim of this paper is to build a first-order version of logic **BAT** where informal provability is a predicate, and not an operator. This was the main motivation behind the construction of **BAT** systems. In our paper, the first-order version of **BAT** is in a language without the provability operator. It is possible to add this operator, but this is not our aim. Our aim is to have a predicate whose behavior is analogous

[3] To be fair, Leitgeb's contribution to the topic is much wider but we have only mentioned the observation he made that is relevant for the current paper.

[4] This schema was thoroughly studied in [Arai, 1998; Beklemishev, 1997, 2003].

[5] See [Alexander, 2013; Antonutti Marfori and Horsten, 2016, 2018; Bellantoni and Hofmann, 2002; Carlson, 2016, 2000; Flagg, 1985; Flagg and Friedman, 1986; Friedman and Sheard, 1989; Goodman, 1984, 1986; Halbach and Horsten, 2000; Heylen, 2013; Horsten, 1994, 1996, 1997, 2006; Koellner, 2016; Reinhardt, 1985, 1986; Rin and Walsh, 2016; Shapiro, 1985].

to the behavior of the operator on the propositional level. This is why we start with the standard first-order language and then we introduce a special predicate for informal provability.

The paper is structured as follows. First, in Section 2 we remind the reader what propositional logics of provability are. Next, in Section 3, we show how to lift **BAT** up to a full first-order version. In this section, we also study and prove some theorems about **BAT** models and compare them with those related to some other many-valued logics. Section 4 is devoted to various strategies for strengthening the **BAT** framework. This section consists mostly of proofs of negative results. Section 5 is devoted to using the **BAT** framework within the context of arithmetic. Section 6 offers a summary of the work and spells out some ideas about the future work related to this framework.

2. Propositional BAT and CABAT

One way to dodge the problem of the inconsistency of the principles of informal provability is to use a non-standard setting and to see how far we can go. Pawlowski and Urbaniak [2018] developed logics of informal provability **BAT** and **CABAT**. The authors treat mathematical informal provability as partially defined.[6] On their view, mathematical claims can be either informally provable, informally refutable or informally undecidable. So, their logics are three-valued: 1 (informally provable), 0 (informally refutable) and n (informally undecidable).[7]

In the standard Kripke construction one relies on the Strong Kleene logic to deal with the partial truth predicate. However, Kleene logic does not seem to be appropriate for modeling the notion of informal provability. If we take a closer look at the behavior of complex sentences and their provability status, it seems that the behavior of complex sentences and their provability status is not truth-functional. Say φ and ψ are informally undecidable (and therefore, so is $\neg\varphi$). Then, while we might think that $\varphi \vee \psi$ is informally undecidable, we might be inclined to think that $\varphi \vee \neg\varphi$ is informally provable, despite both disjuncts being undetermined on both occasions.

[6] This idea comes from [Kripke, 1975].

[7] There are some alternative approaches to construct theories of informal provability [Flagg and Friedman, 1986; Horsten, 1994, 1998; Reinhardt, 1986; Shapiro, 1985; Stern, 2015]. We will not discuss them.

Let $\mathcal{L}$ be a propositional language (understood as the set of all well-formed formulas) constructed from propositional variables $W = \{p_1, p_2, \ldots\}$ and Boolean connectives $(\neg, \wedge, \vee, \rightarrow, \equiv)$ in the standard manner. We will use Greek letters φ, ψ, ... as meta-variables for formulas. The language that results from extending the set of Boolean connectives with one unary operator B will be denoted by $\mathcal{L}_{\mathcal{B}}$. We will use B to express provability within the object language. This operator will allow us to express sentences containing informal provability within the object level and not only on the meta level.

By an *assignment* we mean any function $a\colon W \rightarrow Val$, where Val is any set of values. By an *evaluation* e_a built over an assignment a we will mean a function assigning values to all well-formed formulas, $e_a\colon \mathcal{L} \rightarrow Val$, agreeing with a on W (propositional variables), and satisfying some additional constraints determined by a given logic.

In the case of standard classical propositional logic, evaluations are unambiguously determined by assignments. For each assignment there is exactly one evaluation extending it. In **BAT** we will use the following deterministic and non-deterministic "truth-tables" for connectives.[8]

$\neg$	φ
0	1
n	n
1	0

$\vee$	0	n	1
0	0	n	1
n	n	n/1	1
1	1	1	1

$\wedge$	0	n	1
0	0	0	0
n	0	0/n	n
1	0	n	1

$\rightarrow$	0	n	1
0	1	1	1
n	n	n/1	1
1	0	n	1

$\equiv$	0	n	1
0	1	n	0
n	n	0/n/1	n
1	0	n	1

B	φ
1	1
n/0	n
0	0

Consider a disjunction of two sentences whose value is n. One of two things may happen: it may be possible to informally prove the disjunction (for instant by *reductio ad absurdum*) and the disjunction would have value 1; or it's impossible, and the disjunction remains informally undecided. For instance, consider a formula $p \vee \neg p$. It seems that even if both p and $\neg p$ are informally undecided, the claim is informally provable in virtue of being a substitution of commonly accepted classical tautology. For a more interesting example consider two claims: $\varphi := \mathbf{Con}(\mathbf{PA})$ and $\psi := \mathbf{Con}(\mathbf{PA} + \mathbf{G})$. The first one is a statement saying that Peano arithmetic is consistent. The other one asserts that

[8] The truth conditions can be immediately read-off from the "truth-tables".

not only Peano arithmetic is consistent but Peano arithmetic together with its own Gödelian sentence **G** is consistent. Assuming that Peano arithmetic is indeed consistent, both of these claims are not decided by Peano arithmetic. On the other hand, it's easy to see that $\mathbf{PA} \vdash \psi \rightarrow \varphi$, which by simple propositional reasoning gives $\mathbf{PA} \vdash \neg\psi \vee \varphi$. Thus, both of the claims are undecided, yet their disjunction can be treated as informally provable, since it's provable in Peano arithmetic. So, we cannot limit our attention to those combinations of informally undecided sentences which are substitutions of classical tautologies. For the rest of the cases the truth-table works as a *max* function defined according to the following ordering of information: 0, n, 1.

Similar considerations apply to a conjunction of two informally undecided sentences: either we can prove that they cannot hold together (resulting in a conjunction having value 0) or we cannot do that (so the value remains n). If a formula is informally provable ($e_a(\varphi) = 1$), then giving its own proof is also a proof of its provability ($e_a(\mathtt{B}\varphi) = 1$), and the other way around. If a formula is informally refutable $e_a(\varphi) = 0$, then giving its own refutation is also a refutation of its provability ($e_a(\mathtt{B}\varphi) = 0$).

If a formula is informally undecidable ($e_a(\varphi) = \mathrm{n}$), then one of two things may happen. First, it may be the case that the undecidability of that formula is informally provable, and so its informal provability is refutable ($e_a(\mathtt{B}\varphi) = 0$). Second, it may be the case that its absolute informal undecidability is not informally provable, and so its absolute informal provability is informally undecidable ($e_a(\mathtt{B}\varphi) = \mathrm{n}$).

A **BAT** assignment a is a function from propositional variables W to $\{0, \mathrm{n}, 1\}$. A **BAT** evaluation over an assignment a is a function which assigns values to all formulas of $\mathcal{L}_{\mathtt{B}}$, agrees with a on W and obeys the constraints we gave for the connectives. Notice that due to non-deterministic clauses, one and the same assignment might underlie multiple evaluation functions.

By $\mathit{\Gamma}$🦇φ, where $\mathit{\Gamma}$ is a set of formulas, we will mean that any **BAT** evaluation which assigns 1 to all formulas in $\mathit{\Gamma}$ assigns 1 to the formula φ. We say that φ is a **BAT** tautology iff $\emptyset$🦇φ. We say that φ is a **BAT** countertautology iff $\emptyset$🦇$\neg\varphi$.

Unfortunately, **BAT** is a bit too weak to be treated as the logic of informal provability.[9] In order to fix that, Pawlowski and Urbaniak [2018] introduced additional constraints.

DEFINITION 1. Let **CL** be a classical propositional logic.[10] We say that a **BAT** evaluation e belongs to the **CL**-filtered set of **BAT** evaluations just in case the following conditions hold:

1. For any two formulas φ, ψ, if $\models_{\mathbf{CL}} \varphi \equiv \psi$ then $e(\varphi) = e(\psi)$,
2. For any propositional tautology φ, $e(\varphi) = 1$,
3. For any propositional counter-tautology φ, $e(\varphi) = 0$.

By $\Gamma \models_{\mathbf{CL}} \varphi$ we will mean that for any evaluation e in the **CL**-filtered set of **BAT** evaluations if $e(\psi) = 1$ for all $\psi \in \Gamma$ then $e(\varphi) = 1$. The resulting logic is called **CABAT**.

An equivalent characterization of the logic **CABAT** can be obtained by starting with **BAT** and closing it under the following condition:

DEFINITION 2 (Closure condition). An extension of **BAT** (in $\mathcal{L}_{\mathtt{B}}$) satisfies the closure condition just in case for all $\mathcal{L}_{\mathtt{B}}$-formulas $\varphi_1, \varphi_2, \ldots, \varphi_k, \psi$ such that

$$\varphi_1, \varphi_2, \ldots, \varphi_k \models \psi,$$

where $\models$ is the classical consequence relation for $\mathcal{L}_{\mathtt{B}}$, for any **BAT** evaluation e, if $e(\mathtt{B}\varphi_i) = 1$ for any $0 < i \leqslant k$, then $e(\mathtt{B}\psi) = 1$.

The ultimate goal of this research is to develop an arithmetical theory based on some version of **CABAT**. To do this, we need to extend the whole framework to a first-order version. In the next section we'll see how to do it.

---

[9] Observe that disjunction is neither symmetric nor associative. Take the assignment v where all propositional variables have value n and consider two formulas: $\varphi = p \vee q$ and $\psi = q \vee p$. As far as φ and ψ are concerned, there are four possible ways to extend this assignment:

$$\begin{aligned} &e_v^1(\varphi) = \mathrm{n} = e_v^1(\psi) \\ &e_v^2(\varphi) = 1, \ e_v^2(\psi) = \mathrm{n} \\ &e_v^3(\varphi) = \mathrm{n}, \ e_v^3(\psi) = 1 \\ &e_v^4(\varphi) = 1 = e_v^4(\psi). \end{aligned}$$

BAT is too weak to eliminate extensions (e_v^1, e_v^2, e_v^3), in which φ and ψ obtain different values, and which show that neither φ 🦇 ψ, nor ψ 🦇 φ. Thus, it needs to be strengthened.

[10] In the language extended by a modal operator.

3. First-order BAT

This section is the meat of the paper. First, we will define the notion of a **BAT** model. This will be done by adapting the standard construction of three-values models. The main difference between the standard and **BAT** three-valued models is that the latter does not partition the set of all sentences according to their truth-valued. Instead, the model interprets atomic sentences and complex sentences that are composed of sentences whose values are classical according to the model. To obtain a full interpretation of the language we introduce the notion of evaluation based on a model. This results in non-deterministic setting since a single model may have multiple evaluations based on it.

In the second part we show how those models and evaluations behave. We prove some of the briefly-mentioned results in order to proceed to the next subsection where we focus on the structure of the set of evaluations based on a model. We show that they have some interesting algebraic properties. Next, we define the validity and the consequence relation and compare **BAT** with some other well-known many-valued logics. Last, we comment on possible strategies of strengthening **BAT** since the system itself is quite weak. We show that the standard strategies for strengthening non-deterministic logics are a bit too strong and can't actually be used.

3.1. Three-valued non-deterministic models

In order to construct a first-order version of **BAT**, we need to start with a couple of definitional and notational conventions. First, let $\mathcal{L}$ be a first-order language understood as a set of formulas built in the standard way. We use $Var = \{x_1, x_2, \ldots\}$ to denote the set of variables, Con for constants, $Term$ for the set of terms.[11] Sometimes, we'll be interested in the language $\mathcal{L}^+$ defined as $\mathcal{L}$ plus constants for all elements in the universe. Usually, to define a three-valued first-order logic, the notion of a three-valued structure is used. This notion is pretty standard [Halbach, 2011]. In our case, we use a slight variation of this notion, since the logic is non-deterministic.

A *three-valued structure* is a tuple $\langle \mathtt{M}, \mathtt{i} \rangle$, such that:

1. $\mathtt{M} \neq \emptyset$ is a domain of quantification (sometimes called the universe of the structure).

[11] The set may contain function symbols.

2. $\mathtt{i}$ is an interpretation of $\mathcal{L}$ in $\mathtt{M}$:
   - To every n-ary predicate P, $\mathtt{i}$ ascribes a triple $\langle E(P), A(P), F(P)\rangle$ such that:

$$E(P), A(P), F(P) \subseteq \mathtt{M}^n$$
$$E(P) \cap A(P) = E(P) \cap F(P) = A(P) \cap F(P) = \emptyset$$
$$E(P) \cup F(P) \cup A(P) = \mathtt{M}^n$$

     $E(P)$ is called the *extension* of a predicate P, $A(P)$ stands for the *antiextension* of P and $F(P)$ is called the *fringe* of P. Intuitively, $E(P)$ corresponds to the things that are P, $A(P)$ to the things that are not P and the fringe correspond to those elements of the domain to which a predicate is not applicable. In the classical context we always assume that the fringe is empty. In other words, the interpretation of each predicate P is a partition of the domain into the things that are P, are not P, and those for which P doesn't apply.
   - $i(c) \in \mathtt{M}$ for every constant c.
   - For any n-ary function symbol $\circ$, $i(\circ) : \mathtt{M}^n \to \mathtt{M}$.
   - Identity is classical: $i(=)$ is $\langle E(=), A(=), F(=)\rangle$ such that $E(=)$ is $\{\langle x, x\rangle \mid x \in \mathtt{M}\}$, $A(=)$ is $\mathtt{M}^2 \setminus E(=)$ and $F(=)$ is empty.

Now, a *three-valued* ***BAT*** *model* $\mathcal{M}$ is a triple $\langle \mathtt{M}, \mathtt{i}, v\rangle$, where $\langle \mathtt{M}, \mathtt{i}\rangle$ is a three-valued structure and $v\colon \mathit{Var} \to \mathtt{M}$ is a valuation function. Relative to a valuation we can define the interpretation of terms:

- $t^{\mathcal{M}}(\tau) = \mathtt{i}(\tau)$ if τ is a constant,
- $t^{\mathcal{M}}(x) = v(x)$ if $x \in Var$,
- $t^{\mathcal{M}}(\circ(\tau_1, \dots \tau_n)) = (\mathtt{i}(\circ))(t^{\mathcal{M}}(\tau_1), \dots t^{\mathcal{M}}(\tau_n))$.

For a moment, let's focus on atomic formulas. In the classical context, each atomic formula $P(a)$ is either true (if $i(a) \in E(P)$) or false (if $i(a) \in A(P)$). In our case, since we have three values, we are not going to use a classical satisfaction relation. Instead, we use a satisfaction triple $\langle \Vdash_1, \Vdash_{\mathrm{n}}, \Vdash_0\rangle$ defined as:

- $\mathcal{M} \Vdash_1 P(\tau_1, \dots \tau_n)$ iff $\langle \mathtt{i}(\tau_1), \dots \mathtt{i}(\tau_n)\rangle \in E(P)$
- $\mathcal{M} \Vdash_0 P(\tau_1, \dots \tau_n)$ iff $\langle \mathtt{i}(\tau_1), \dots \mathtt{i}(\tau_n)\rangle \in A(P)$
- $\mathcal{M} \Vdash_n P(\tau_1, \dots \tau_n)$ iff $\langle \mathtt{i}(\tau_1), \dots \mathtt{i}(\tau_n)\rangle \in F(P)$

Notice that the *satisfaction triples* essentially partition the set of atomic sentences into three classes. At this stage, the framework is perfectly deterministic, but it does not tell us how to interpret complex formulas. This is the moment when non-determinism kicks in. To provide the interpretation of Boolean connectives, we extend, not necessarily in a unique way, the notion of the satisfaction triple to all complex formulas. *An evaluation* is a total function $e_{\mathcal{M}} \colon \mathcal{L}^{+} \to \{0, \mathrm{n}, 1\}$ such that for atomic formulas φ we have:

- $e_{\mathcal{M}}(\varphi) = 1$ iff $\mathcal{M}$ φ,
- $e_{\mathcal{M}}(\varphi) = \mathrm{n}$ iff $\mathcal{M}$ φ,
- $e_{\mathcal{M}}(\varphi) = 0$ iff $\mathcal{M}$ φ.

There are two reasons to introduce evaluations in this manner. First, they relate satisfaction triples with logical values that we use in the system. Second, they allow us to express certain things more easily (for instance, to talk about the set of all sentences whose values are n). In order to cope with quantifiers, we treat them as "infinite" conjunctions and infinite disjunctions.

DEFINITION 3 (**BAT** evaluation). Let $\mathcal{M}$ be a three-valued model. We say that an evaluation $e_{\mathcal{M}}$ is **BAT** evaluation iff for all formulas φ, ψ:

1. Negation:
   (a) $e_{\mathcal{M}}(\neg\varphi) = 1$ iff $e_{\mathcal{M}}(\varphi) = 0$,
   (b) $e_{\mathcal{M}}(\neg\varphi) = 0$ iff $e_{\mathcal{M}}(\varphi) = 1$,
   (c) $e_{\mathcal{M}}(\neg\varphi) = \mathrm{n}$ iff $e_{\mathcal{M}}(\varphi) = \mathrm{n}$,
2. Disjunction:
   (a) If $e_{\mathcal{M}}(\varphi) = 1$ or $e_{\mathcal{M}}(\psi) = 1$, then $e_{\mathcal{M}}(\varphi \vee \psi) = 1$,
   (b) $e_{\mathcal{M}}(\varphi \vee \psi) = 0$ iff $e_{\mathcal{M}}(\varphi) = 0$ and $e_{\mathcal{M}}(\psi) = 0$,
   (c) If $e_{\mathcal{M}}(\varphi) = 0$ and $e_{\mathcal{M}}(\psi) = \mathrm{n}$, then $e_{\mathcal{M}}(\varphi \vee \psi) = \mathrm{n}$,
   (d) If $e_{\mathcal{M}}(\varphi) = \mathrm{n}$ and $e_{\mathcal{M}}(\psi) = 0$, then $e_{\mathcal{M}}(\varphi \vee \psi) = \mathrm{n}$,
   (e) If $e_{\mathcal{M}}(\varphi) = \mathrm{n}$ and $e_{\mathcal{M}}(\psi) = \mathrm{n}$, then $e_{\mathcal{M}}(\varphi \vee \psi) = \mathrm{n}$ or $e_{\mathcal{M}}(\varphi \vee \psi) = 1$,
3. The case for the remaining Boolean connectives are exactly the same as on the propositional level.
4. Quantifiers:
   (a) $e_{\mathcal{M}}(\forall x \varphi(x)) = 1$ iff for any constant $c \in \mathcal{L}^{+}$, we have $e_{\mathcal{M}}(\varphi(c)) = 1$,
   (b) If for any constant $c \in \mathcal{L}^{+}$, we have $e_{\mathcal{M}}(\varphi(c)) = \mathrm{n}$, then $e_{\mathcal{M}}(\forall x \varphi(x)) = 0$ or $e_{\mathcal{M}}(\forall x \varphi(x)) = \mathrm{n}$,
   (c) If there is a constant $a \in \mathcal{L}^{+}$ such that $e_{\mathcal{M}}(\varphi(a)) = \mathrm{n}$, and for any other constant $c \in \mathcal{L}^{+}$ $e_{\mathcal{M}}(\varphi(c)) \neq 0$, then $e_{\mathcal{M}}(\forall x \varphi(x)) = \mathrm{n}$,

(d) If there is a constant $a \in \mathcal{L}^+$ such that $e_{\mathcal{M}}(\varphi(a)) = 0$, then $\forall x\varphi(x)) = 0$,
(e) $e_{\mathcal{M}}(\exists x\varphi(x)) = 0$ iff for all constants $c \in \mathcal{L}^+$, $e_{\mathcal{M}}(\varphi(c)) = 0$,
(f) If there is a constant $c \in \mathcal{L}^+$ such that $e_{\mathcal{M}}(\varphi(c)) = 1$, then $e_{\mathcal{M}}(\exists x\varphi(x)) = 1$,
(g) If for all constants $a \in \mathcal{L}^+$ we have $e_{\mathcal{M}}(\varphi(a)) = \mathrm{n}$ or $e_{\mathcal{M}}(\varphi(b)) = 0$ (and we have at least one witness for either option), then $e_{\mathcal{M}}(\exists x\varphi(x)) = \mathrm{n}$.
(h) If for all constants $c \in \mathcal{L}^+$, we have $e_{\mathcal{M}}(\varphi(c)) = \mathrm{n}$, then $e_{\mathcal{M}}(\exists x\varphi(x)) = \mathrm{n}$ or $e_{\mathcal{M}}(\exists x\varphi(x)) = 1$.

Since **BAT** evaluations are defined relative to a model $\mathcal{M}$, we call them evaluations based on a model $\mathcal{M}$. The set of all **BAT** triples based on $\mathcal{M}$ is denoted as $Str_{\mathcal{M}}$.

It is quite easy to see that in general $\mathcal{M}$ does not decide the logical values of all complex formulas. For instance, consider a disjunction of two atoms whose value is n. Then, there is an evaluation for which the disjunction's value is 1 and a different one, where the value remains n. Complex formulas whose all sub-formulas have classical values (so either 1 or 0) have the same values in all **BAT** evaluations based on a single model.

3.2. Properties of non-deterministic models

Let's start with the definitions of satisfaction, dissatisfaction, and neutral set.

DEFINITION 4 (Satisfaction, dissatisfaction, neutral set). Let $\mathcal{M}$ be a **BAT** model and let $\langle \Vdash_1, \Vdash_{\mathrm{n}}, \Vdash_0 \rangle$ a **BAT** triple based on it. We will use the following abbreviations: $S = \{\varphi \mid \mathcal{M} \Vdash_1 \varphi\}, D = \{\varphi \mid \mathcal{M} \Vdash_0 \varphi\}, N = \{\varphi \mid \mathcal{M} \Vdash_n \varphi\}$. We will refer to these sets respectively as *satisfaction, dissatisfaction and neutral* sets corresponding to a given **BAT** triple. Those sets give a partition of the set of all sentences.

FACT 1. *Let $\langle \Vdash_1, \Vdash_{\mathrm{n}}, \Vdash_0 \rangle$ be a* **BAT** *triple based on $\mathcal{M}$. If either S or D is empty, then N contains all the formulas of the language.*

PROOF. Suppose that S is empty and that there is a formula $\varphi \notin N$. Then, since $S \cup N \cup D$ exhaust all the formulas of the language, $\varphi \in D$. From this follows that $\neg\varphi \in S$, which contradicts the assumption. The case for D is symmetric. $\square$

FACT 2. *If N does not contain atoms, then N is empty.*

PROOF. Suppose N does not contain atoms. It means that all the atoms are either in S or D. Let's focus on quantifier-free formulas. It is easy to see, that if all sub-formulas of a given formula have classical values only, the whole formula has to have a classical value. This means that N cannot contain a quantifier-free formula.

For quantifiers, suppose that $\exists x\varphi(x) \in N$, where $\varphi(x)$ is a quantifier-free formula. By the satisfaction clauses for the existential quantifier we have to consider two cases. The first one, where for some constant $c \in \mathcal{L}^+$ we have $\varphi(c) \in N$, which implies that at least one atom has a non-classical value, which contradicts the assumption. The second case is where for all constants $c \in \mathcal{L}^+$ either $\varphi(c) \in D$ or $\varphi(c) \in N$ and both options have witnesses. This means that for some $a \in \mathcal{L}^+$, we have $\varphi(a) \in N$, which again is impossible. Analogous reasoning can be applied to the universal quantifier. □

THEOREM 1 (Identity criterion 1). *Let* $\langle \Vdash^1_1, \Vdash^1_{\mathrm{n}}, \Vdash^1_0 \rangle$, $\langle \Vdash^2_1, \Vdash^2_{\mathrm{n}}, \Vdash^2_0 \rangle$ *be* **BAT** *triples based on* $\mathcal{M}$. *If either* $S_1 = S_2$ *or* $D_1 = D_2$, *then the triples are identical.*

PROOF. Suppose that $S_1 = S_2$. We will show that from this it follows that $D_1 = D_2$. From this we may infer that $N_1 = N_2$, since the sum of S, D, N exhaust all the formulas. Next, we assume $D_1 = D_2$ and argue that then $S_1 = S_2$.

Let's start with the first case, and assume for contradiction that $D_1 \neq D_2$. This means that there is a formula $\varphi \in D_1$ and $\varphi \notin D_2$. It follows that $\neg\varphi \in S_1$. We assumed that $S_1 = S_2$ so $\neg\varphi \in S_2$, thus $\neg\neg\varphi \in D_2$ and $\varphi \in D_2$— contradiction.

In the second case, $D_1 = D_2$ and $S_1 \neq S_2$. Again, there is a formula $\varphi \in S_1, \varphi \notin S_2$. So by the conditions for negation $\neg\varphi \in D_1$, so $\neg\varphi \in D_2$. By the conditions for negation, $\varphi \in S_2$ resulting in a contradiction. □

THEOREM 2 (Satisfaction inclusion). *Let* $\langle \Vdash^1_1, \Vdash^1_{\mathrm{n}}, \Vdash^1_0 \rangle$, $\langle \Vdash^2_1, \Vdash^2_{\mathrm{n}}, \Vdash^2_0 \rangle$ *be* **BAT** *triples based on* $\mathcal{M}$. $D_1 \subseteq D_2$ *iff* $S_1 \subseteq S_2$.

PROOF. Left to right: assume $S_1 \subseteq S_2$. Take any $\varphi \in D_1$, it follows that $\neg\varphi \in S_1$. From the assumption, $\neg\varphi \in S_2$, and from satisfaction clauses for negation $\varphi \in D_2$. For the implication in the other direction assume $D_1 \subseteq D_2$. Take any $\varphi \in S_1$, it follows that $\neg\varphi \in D_1$. By the assumption $\neg\varphi \in D_2$, giving by satisfaction clauses for negation $\varphi \in S_2$. □

THEOREM 3 (Identity criterion 2). *Let* $\langle \Vdash^1_1, \Vdash^1_{\mathrm{n}}, \Vdash^1_0 \rangle$, $\langle \Vdash^2_1, \Vdash^2_{\mathrm{n}}, \Vdash^2_0 \rangle$ *be* **BAT** *triples based on* $\mathcal{M}$. *If* $N_1 = N_2$ *then the triples are identical.*

Proof. Take two **BAT** triples $\langle \Vdash_1^1, \Vdash_n^1, \Vdash_0^1 \rangle$, $\langle \Vdash_1^2, \Vdash_n^2, \Vdash_0^2 \rangle$ based on $\mathcal{M}$. Assume that $N_1 = N_2$. By induction on the complexity of φ we show that $\varphi \in S_1$ iff $\varphi \in S_2$ which, by Theorem 1, implies $\langle \Vdash_1^1, \Vdash_n^1, \Vdash_0^1 \rangle = \langle \Vdash_1^2, \Vdash_n^2, \Vdash_0^2 \rangle$.

For atomic formulas the claim holds by the fact that both triples are based on the same model. Assume that the claim holds for χ, ψ. We will show that it works for all formulas.

$\neg$: Assume $\varphi = \neg\psi$ and $\varphi \in S_1$. It follows that $\psi \in D_1$ and by the induction hypothesis $\psi \notin S_2$. Thus, either $\psi \in N_2$ or D_2. The first option is not possible, so $\psi \in D_2$, resulting in $\varphi \in S_2$.

$\vee$: Let $\varphi = \psi \vee \chi$, and assume $\varphi \in S_1$. We have three options to consider: either $\psi \in S_1$, or $\chi \in S_1$, or $\psi, \chi \in N_1$. For the first two options, by the induction hypothesis we have $\psi \in S_2$ or $\chi \in S_2$, resulting in both cases in $\varphi \in S_2$.

For the third option, we have $\chi, \psi \in N_1$. By the initial assumption $\chi, \psi \in N_2$. Assume for contradiction that the claim fails, so $\varphi \notin S_2$. Now, this splits our case in two: either $\varphi \in N_2$, or $\varphi \in D_2$. For the former, since $N_1 = N_2$ and so $\varphi \notin S_1$, which contradicts the assumption. So it has to be the case that $\varphi \in D_2$. By conditions for disjunction we have $\psi, \chi \in D_2$ which together with our assumption that $\chi, \psi \in N_1$ and $N_1 = N_2$ results again in contradiction.

$\wedge$: Let $\varphi = (\psi \wedge \chi)$ and $\varphi \in S_1$. It follows $\psi \in S_1, \chi \in S_1$ so by the induction hypothesis, $\psi, \chi \in S_2$ resulting in $\varphi \in S_2$.[12]

$\exists$: Let $\varphi = \exists\, x\psi(x)$ and $\varphi \in S_1$. We have two cases. The first one, there is an element $a \in \mathcal{L}^+$ such that $\psi(a) \in S_1$. By the induction hypothesis $\psi(a) \in S_2$ and by satisfaction clauses for quantifiers, $\exists\, x\psi(x) \in S_2$.

For the second case, for all constants $a \in \mathcal{L}^+$, we have $\psi(a) \in N_1$. By the initial assumption, this means that for any constant $a \in \mathcal{L}^+$, we have $\psi(a) \in N_2$. By the semantics for the (existential) quantifier, either $\exists\, x\psi(x) \in N_2$ or $\exists\, x\psi(x) \in S_2$. Since $\exists\, x\psi(x) \in S_1$ and $N_1 = N_2$ the first option is impossible so the second option has to be the case.

$\forall$: Let $\varphi = \forall\, x\psi(x)$ and $\varphi \in S_1$. This means that for any constant $a \in \mathcal{L}^+$, $\psi(a) \in S_1$. By the induction hypothesis, for any a, $\psi(a) \in S_2$ so $\forall\, x\psi(x) \in S_2$. □

[12] Notice that here we do not use the main assumption that $N_1 = N_2$.

Definition 5 (Classical formula). Let $\mathcal{M}$ be a model. We say that a formula φ is classical in $\mathcal{M}$ iff the fringes of all predicates used in φ are empty.

Theorem 4 (Model and classical formulas). *Let $\mathcal{M}$ be a model and φ a formula that is either classical in $\mathcal{M}$ or has the form $\neg \cdots \neg\psi$ for some atomic ψ. Then all* **BAT** *triples based on $\mathcal{M}$ agree on φ, i.e. in all of them, either $\varphi \in S$, or $\varphi \in D$, or $\varphi \in N$.*

Proof. Let's start with atomic formulas. It's quite easy to see that by the definition of an evaluation, evaluations have to agree with the model on atomic formulas. Notice that the matrix for negation in **BAT** is deterministic, the value of a formula of the form $\neg \cdots \neg\psi$ for some atomic ψ is uniquely determined by the model itself. For the classical formulas it is a bit more tricky. First, recall that evaluations on classical predicates work classically. It follows that these evaluations uniquely determine the value of all such formulas, so it's impossible that some **BAT** triples disagree with them. $\square$

As already mentioned, the next section aims at structuring the set of evaluations based on a single model. The main motivation here is to allow us to mimic the move from propositional **BAT** to **CABAT**. In order to do so, one has to have a good procedure for narrowing down evaluations. One way of doing this is to consider only evaluations which maximize their own satisfaction.

3.3. The structure of BAT evaluations

Up till now the picture has been quite clear. We have **BAT** models that assign values to some formulas of the language and we extend them with **BAT** triples to obtain the satisfaction relation for the complex formulas of the language. One of the problems with this approach is a feature of propositional **BAT**. On the propositional level the logic is too weak (for instance, **BAT** disjunction is not symmetric).

This is also the case for the first-order version. In other words, we have too many extensions based on a single model and not all of them are interesting. The idea now is to provide some philosophically justified criteria to single out a class of interesting evaluations based on a given model. To do that, one has to know what relations between different triples based on the same model are, and this is the topic of current section. We'll prove that the set of all triples indeed has an interesting structure and a natural ordering between triples can be defined.

Naturally, in our case, we will be mostly interested in sentences whose value, is 1 — those in the satisfaction set. They can be used to define an ordering on the triples. So the ordering $\leq$ on $Str_{\mathcal{M}}$ is given by

$$\langle \Vdash_1^1, \Vdash_{\mathrm{n}}^1, \Vdash_0^1 \rangle \leq \langle \Vdash_1^2, \Vdash_{\mathrm{n}}^2, \Vdash_0^2 \rangle \text{ iff } S_1 \subseteq S_2,$$

(we define $<$ in the usual way). Equivalently, we can phrase the ordering directly by using evaluations. Let e_1, e_2 be two evaluations. We say that $e_1 \leq e_2$ iff $S_{e_1} \subseteq S_{e_2}$.

First, notice that the relation is indeed an ordering relation.

FACT 3 ($<$ is a strict partial order). *Let $\mathcal{M}$ be a model. The relations $\leq$ and $<$ partially and strictly partially order the set $Str_{\mathcal{M}}$, respectively.*

DEFINITION 6 (Evaluations product). Let $\mathcal{M}$ be a model and let e_1 and e_2 be an evaluations based on it. By $e_1 \otimes e_2 : \mathcal{L} \to \{0, \mathrm{n}, 1\}$ we mean the following function:

$$e_1 \otimes e_2(\varphi) = \begin{cases} 1, e_1(\varphi) = e_2(\varphi) = 1 \\ 0, e_1(\varphi) = e_2(\varphi) = 0 \\ \mathrm{n}, \text{ otherwise} \end{cases}$$

THEOREM 5 (Triple product is **BAT** triple). *If if e_1 and e_2 are evaluations so is $e_1 \otimes e_2$.*

PROOF. We need to show that $\otimes$ respects all 27 conditions put on being a **BAT** evaluation. This proof is straightforward. □

We have a way of comparing **BAT** evaluations over a given model. We have also defined one way of combining evaluations to the effect that the combined evaluation is "smaller" than both evaluations used in the combination. This ordering can be used for further study of the algebraic properties of triples. Note that for any sentence φ that is not decided by the model,[13] there is an evaluation based on that model which puts φ in N. This follows directly from the way we have defined tables for connectives and quantifiers. So, there is a **BAT** evaluation that puts all formulas undecided by the model into N. Such a triple will be called

[13] A formula is decided by the model iff it has the same value in all evaluations based on this model.

a *Fully non-deterministic triple* (FND triple). Equivalently we can talk about FND-evaluation. Algebraically speaking, this triple is important because it is one of the minimal triples over a given model. Even more, if the model is fully classical,[14] this triple together with the model gives the model of classical logic.

THEOREM 6. *There is no evaluation e such that $e < e_{\mathrm{FND}}$.*

PROOF. Suppose for a contradiction that the theorem does not hold, so $e < e_{\mathrm{FND}}$. So there is φ such that $e(\varphi) = 1$ and $e_{\mathrm{FND}} \in \{0, n\}$. By table for negation we have $e(\neg\varphi) = 0$, and so $e_{\mathrm{FND}}(\neg\varphi) = \mathrm{n}$. Now, since FND-evaluation puts as many formulas as possible into N, we know that $e_{\mathrm{FND}}(\varphi) = \mathrm{n}$, which again leads to inconsistency with the main assumption. □

So a formula is true in a model iff it is true in all possible evaluations based on this model. Equivalently we may think of it as being true in the smallest evaluation, namely the FND triple. Now we know what it means for a formula to be true, let's define the notion of **BAT** tautology:

DEFINITION 7 (**BAT** tautology). We say that φ is a **BAT** tautology, $\models_{\mathsf{B}} \varphi$, iff for every model $\mathcal{M}$, and every evaluation based on it we have $\mathcal{M} \Vdash_1 \varphi$.

DEFINITION 8 (Local consequence relation). Let Γ be a finite set of formulas, we say that φ is a local **BAT** semantic consequence of Γ, $\Gamma \models_{\mathsf{B}} \varphi$, iff for any model $\mathcal{M}$ and for any evaluation e based on this model, if for all $\psi \in \Gamma$ we have $\psi \in S$, then $\varphi \in S$.

Before delving into extensions of **BAT**, let's see how it compares against other well known three-valued logics.

3.4. BAT vs. Kleene and Łukasiewicz three-valued logics

Since non-deterministic logics have not been adequately studied up to now, we will consider only well-known deterministic many-valued logics. The main aim is to show that for any **BAT** model there is a triple that gives a model of the desired three-valued logics. We will take a look at the familiar candidates:

[14] All predicates have a classical value 0 or 1.

DEFINITION 9 (Strong Kleene-valuation schema). Let $\mathcal{M}$ be a model. By a *Kleene-evaluation* we mean the following evaluation $e_{\mathcal{M}}$ given by the tables:

$\neg$	φ
0	1
n	n
1	0

$\vee$	0	n	1
0	0	n	1
n	n	n	1
1	1	1	1

$\wedge$	0	n	1
0	0	0	0
n	0	n	n
1	0	n	1

$\rightarrow$	0	n	1
0	1	1	1
n	n	n	1
1	0	n	1

and for quantifiers:

$$v(\forall x\varphi(x)) = \begin{cases} 1, \text{if for all } a \in \mathcal{M}, v(\varphi(a)) = 1 \\ 0, \text{if there is } a \in \mathcal{M}, v(\varphi(a)) = 0 \\ \text{n, otherwise} \end{cases}$$

$$v(\exists x\varphi(x)) = \begin{cases} 1, \text{if there is } a \in \mathcal{M}, v(\varphi(a)) = 1 \\ 0, \text{if for all } a \in \mathcal{M}, v(\varphi(a)) = 0 \\ \text{n, otherwise} \end{cases}$$

Quite expectedly, the consequence relation is defined as usual as preservation of the value 1, $\Gamma \models_{\mathrm{K}} \varphi$ iff for any model $\mathcal{M}$ and for any formula $\psi \in \Gamma$, $e(\psi) = 1$ implies $e(\varphi) = 1$, where e is the Kleene valuation based on the model $\mathcal{M}$.

Now we proceed to the other well-known three-valued logic, Łukasiewicz's three-valued logic. The difference between Kleene's and this logic is the truth-table for the material conditional:

$\rightarrow$	0	n	1
0	1	1	1
n	n	1	1
1	0	n	1

The rest of the Boolean connectives and quantifiers are defined as in Kleene's logic. Similarly, the consequence relation $\Gamma \models_{\mathrm{L}} \varphi$ is defined as the preservation of the value 1.

The first thing to notice is that Kleene's and Łukasiewicz's valuation schemata agree with **BAT** on every deterministic case. For the non-deterministic case, quite easily one can find an evaluation that corresponds to either Kleene's or Łukasiewicz's schema. So for any **BAT** model, it's possible to find an evaluation whose behavior is exactly as those of Kleene or Łukasiewicz valuations schemata. This means that if an inference is locally valid in **BAT** it is valid in both Kleene's and Łukasiewicz's logics:

Theorem 7. *If* $\Gamma \models_{\mathrm{B}} \varphi$ *then* $\Gamma \models_{\mathrm{K}} \varphi$ *and* $\Gamma \models_{\mathrm{L}} \varphi$.

4. Strengthening BAT

Consider a **BAT** model $\mathcal{M}$. Intuitively, the maximal evaluation based on this model would be one that contains the most information. This mean that this evaluation is supposed to minimize the number of sentences whose values are n (equivalently they belong to N). So in principle, one would want to consider only the maximal valuations and see how strong the resulting logic is. This can be achieved by the following straightforward procedure. Let's start with a model $\mathcal{M}$ and take the set of all evaluations $Str_{\mathcal{M}}$. Consider the following procedure of filtration: for any φ:

- If there are two evaluations e_1, e_2 such that $e_1(\varphi) = 1$, and $e_2(\varphi) \neq 1$, delete the evaluation.
- If there are two evaluations e_1, e_2 such that $e_1(\varphi) = 0$, and $e_2(\varphi) \neq 0$, delete the second extension.

Although the filtration seems like a plausible strengthening, for some model the set of extensions after the filtration has been applied is empty.

Fact 4. *For any model* $\mathcal{M}$ *that is not fully classical, its filtrated set of extensions is empty.*

Proof. If the model is not classical then there is an atomic formula $P(a)$ such that $P(a) \in N$. Now, consider the formula $\psi = P(a) \to P(a) \wedge P(a)$. By truth-tables, all extensions have to put the formula either in N or in S. According to the filtration procedure, extensions where $\psi \in N$ are filtered out. This means that $P(a) \wedge P(a) \in N$, so $\neg(P(a) \wedge P(a)) \in N$ for all filtrated extensions. On the other, there is an extension of the model where $\neg(P(a) \wedge P(a))\neg S$. This implies that the extension where $\psi \in S$ is also filtered out, leaving us with the empty set of extensions. □

The other filtration procedure that immediately comes to mind is to remove only those valuations which gives the least amount of information: for any φ, if there is an extension where $\varphi \in N$ and there is an extension where $\varphi \notin N$, remove the former extension. The idea here is analogous to super-valuation. Call this procedure an s-filtration. The set of all valuations based on $\mathcal{M}$ after s-filtration is $\mathcal{F}_{\mathcal{M}}$. Unfortunately, the trick will not work.

Fact 5. *There is a model $\mathcal{M}$ whose $\mathcal{F}_{\mathcal{M}}$ is empty.*

Proof. Take a model $\mathcal{M}$ that is not purely classical i.e. there is an atomic sentence $P(a) \in N$. Consider two formulas $\varphi = \neg P(a) \vee P(a) \wedge P(a)$ and $\psi = \neg(P(a) \wedge P(a))$. Notice, that for any extension, if $\varphi \in S$, then $\neg(P(a) \wedge P(a)) \in N$. But on the other hand, there is an extension where $\neg(P(a) \wedge P(a)) \in S$. So, after s-filtration both types of extensions have to be removed, leaving us with the empty set of extensions. $\square$

Yet another procedure is to consider a procedure that associates sets of classical models with each particular extension. Take a model $\mathcal{M}$ and an extension E, by a classical extension $\mathcal{M}^{c}$ of $\mathcal{M}, E$ we mean any classical model for which we have:

1. For any atomic φ, if $\varphi \in S_E$, then $\mathcal{M}^{c} \models \varphi$.
2. For any atomic φ, if $\varphi \in D_E$, then $\mathcal{M}^{c} \models \neg\varphi$.

The idea here is quite simple. With each non-deterministic model we associate a set of classical models, where each classical model in the set is one particular way of deciding all formulas that belong to N, according to the non-deterministic model. Yet again, the process will not help here. This is because some extensions of non-classical models are not classical. Consider an extension where $P(a) \in N$, $P(a) \vee \neg P(a) \in N$, $\neg P(a) \vee P(a) \in N$, and $(\neg P(a) \vee P(a)) \wedge (P(a) \vee \neg P(a)) \in D$. Clearly this extension does not have any classical models associated with it, since in any classical model $\mathcal{M}^{c}$ we have $\mathcal{M}^{c} \models (\neg P(a) \vee P(a)) \wedge (P(a) \vee \neg P(a))$.

The moral is quite straightforward. The usual method of strengthening a non-deterministic logic is not the way to go for **BAT**. These types of filtration won't work because the interaction between non-deterministic functors is a bit more complex.[15]

We conclude with a proof of a negative result, that most of those strategies that are usually used to strengthen a non-deterministic logic fail in this context. This happens because the interplay between non-deterministic first-order operators in **BAT** is a bit more complex than in the rest of well-studied non-deterministic logics.[16] The negative result

[15] This is probably why in non-deterministic logics, usually we want to have only one purely non-deterministic functor.

[16] Mostly, because some non-deterministic cases combine designated values with non-designated values. This seems to generate technical problems.

is quite important because it shows that in order to move from **BAT** to some stronger system at the first-level one needs to come up with new ways of strengthening. We solve this issue by using the same formal trick that, on the propositional level, gives **CABAT**. The main drawback of this solution is that it does not give easy access to the models. This makes the theory quite unpleasant to work with.

The other strategy is to use the same move that on the propositional level allowed us to move from **BAT** to **CABAT**. This move is based on two assumptions.[17]

The first one says that classically equivalent formulas should have the same provability status. It is quite obvious that if a mathematical claim is informally provable or refutable all its logically equivalent formulations are as well. A similar argument holds for sentences that are neither. In more abstract setting this condition goes as follow:

Definition 10 (L-Equivalence condition). Let L be a logic. We say that a given **BAT** triple respects the L-equivalence condition iff for all formulas φ, ψ if the formulas are L-equivalent then either both formulas belong to the satisfaction set, or both are in dissatisfaction set, or both are in non-satisfaction set of the triple.

So this condition enables one not to take into account those tricky extensions where trivially equivalent formulas are put in different sets. Does this solve all the problems with **BAT** validity? It helps, but it is not sufficient. The reason is as follows. Sane mathematicians do not question inference steps in an informal proof that are correct form the point of view of classical logic. They believe that axioms of classical logic are true and that the rules are truth-preserving. So far, even with L-equivalence with respect to classical logic, we do not do justice to this postulate. Still, some classically valid inferential moves are not valid according to this condition. In order to preserve the intuitive validity of classical inference steps we introduce an additional condition:

Definition 11 (L-respect). If L is a logic, then we say that a **BAT** triple satisfies the L-respect condition iff any formula φ that is L-tautology belongs to the satisfaction set of this triple.

[17] Similar moves have been done in the context of non-deterministic semantics [Kearns, 1981; Omori and Skurt, 2016].

The above conditions are justified and do the job that they were supposed to do. We restrict our attention to the case where L is classical logic. Then, we can define fully-fledged first-order **CABAT**:

DEFINITION 12 (**CABAT** triple). Let $\mathcal{M}$ be a **BAT** model. **CABAT** satisfaction triple is any triple ⟨🦇, 🦇, 🦇⟩ that satisfies both L-equivalence and L-respect, for classical logic. We use $\Gamma \models_{\mathrm{C}} \varphi$ to denote the **CABAT** local consequence relation.

First, notice that **CABAT** triples are not uniquely determined by the model. Consider a model $\mathcal{M}$ and assume that $P(a), P(b)$ both have value n. Consider $P(a) \vee P(b)$ and $P(b) \vee P(a)$. By CL-equivalence, both of them have to be either in S, or in N, so there are two **CABAT** triples one which puts both formulas in S and one which puts them in N.

FACT 6. *Classical contradictions are refutable according to* **CABAT** *triples.*

PROOF. First, any classical tautology φ is informally provable due to the respect condition. This means that they belong to the satisfaction set. Any negation of classical contradiction is a classical tautology, so it has to belong to the satisfaction set as well. Which means, by the clauses for negation, that the negation of negation of classical contradiction belongs to the satisfaction set, so the contradiction is in the dissatisfaction. $\square$

FACT 7. *For any sentence φ, $\models_{\mathrm{C}} \varphi$ iff $\models \varphi$.*

PROOF. $\Rightarrow$: If $\not\models \varphi$, then there is a classical model $\mathcal{M}$ such that $\mathcal{M} \not\models \varphi$. This model can be seen as **BAT** model where all predicates are classical, which shows that $\not\models_{\mathrm{C}} \varphi$, since the classical model preserves both CL-equivalence and CL-respect.

$\Leftarrow$: If $\models \varphi$, then we know that φ is a classical tautology and as such has to have value 1 in every **BAT** model in every **CABAT** triple. $\square$

This can be quite easily extended to a stronger result.

FACT 8. *For any finite set of sentences Γ, and a sentence φ, $\Gamma \models_{\mathrm{C}} \varphi$ iff $\Gamma \models \varphi$.*

PROOF. $\Rightarrow$: by contraposition, suppose $\Gamma \not\models \varphi$. This means that there is a classical model $\mathcal{M}$, where all elements of Γ have a designated value and $\mathcal{M} \not\models \varphi$. Quite trivially, this is also a **BAT** model with a **CABAT** triple, and so it means that there is a **CABAT** triple which shows that the whole inference is not valid. $\Leftarrow$: since Γ is finite and $\Gamma \models \varphi$, by the deduction theorem we know that $\models \bigwedge_{\varphi_i} \to \varphi$, where $\varphi_i \in \Gamma$. By the L-respect condition, we have $\models_{\mathrm{C}} \bigwedge_{\varphi_i} \to \varphi$. This implies that $\Gamma \models_{\mathrm{C}} \varphi$. $\square$

5. How to get arithmetical models?

So far, we have extended the **BAT** setting to the first-order level. The next step is to actually show how to use the logic within an arithmetical setting. One straightforward solution is to relativize the construction to a particular basic theory **T**, whose inferences we intuitively perceive as valid. This assumption is not controversial as it simply says that certain basic facts are informally provable. So, we will start with an arithmetical theory that extends theory **Q**, which is an arithmetical theory without the schema of induction.[18]

The next question is how to extend the language with a provability predicate `B` whose behavior is similar to the behavior of the provability operator in the propositional case. After all, this is the main task that we want to accomplish. As is usual, first we extend the language of arithmetic with a unary predicate `B` and we provide an interpretation of it. This means that we are interested in tuples $(\mathcal{M}, A)$, where $\mathcal{M}$ is a **BAT** model in the arithmetical language extended with `B` and $A = (E, A, F)$ is an interpretation of `B`.[19] Next we put some additional conditions on the sets of admissible evaluations. The straightforward condition is the following:

Definition 13 (Provability **BAT** evaluations). Let **T** be a consistent arithmetical theory extending **Q** such that $\mathbb{N} \models \mathbf{T}$[20]. We say that an evaluation e based on a model $\mathcal{M}$ is faithful with respect to **T**, if the following conditions are satisfied:

1. All φ that are theorems of **T**, we have $e(\varphi) = 1$.
2. If $\mathbf{T} \vdash \varphi$, then $\ulcorner \varphi \urcorner \in E$.
3. If $\mathbf{T} \vdash \neg\varphi$, then $\ulcorner \varphi \urcorner \in A$.

[18] So, there are axioms governing the multiplication, addition and the successor functions [see Halbach, 2011].

[19] So the triple satisfies the usual conditions put on the interpretation: sets are disjoint and mutually exhaust the domain.

[20] $\mathbb{N}$ is the standard model of natural numbers.

where $\ulcorner\varphi\urcorner$ is the arithmetical code of the sentence φ. We will use $\mathcal{M} \models_i \varphi$ to mean that any provability extension of model $\mathcal{M}$ assigns i to φ, where $i \in \{0, \mathrm{n}, 1\}$.

This is still too weak to actually be used as a theory of informal provability. For instance, it does not validate the iterations of the provability predicate. There are at least three ways of solving this problem. We will go through all of them in the next subsection.

Here we present a way of strengthening the logic. This is done by adapting the technique developed by Kripke to handle the partial truth predicate. The idea is to have an infinite sequence of models in which at any given stage the extension and anti-extension of provability predicate is enlarged. We start with the empty sets and at each subsequent level we put into the extension of B codes of formulas valid at the previous level, and into the anti-extension those whose value was 0 at the previous level and into F codes of formulas whose values were n.

Definition 14 (Recursive strengthening). Let $\mathcal{M}_0$ be an arithmetical **BAT** model and let e be a provability evaluation. Consider the following procedure:

- $\mathcal{M}_{n+1} = \langle \mathcal{M}, \langle E = \{\varphi \mid \mathcal{M} \models_1 \varphi\}, A = \{\varphi \mid \mathcal{M}^B_n \models_0 \varphi\}, F = \{\varphi \mid \mathcal{M} \models_{\mathrm{n}} \varphi\}\rangle\rangle$.
- If λ is a limit ordinal, then $\mathcal{M}_\lambda = \langle \mathcal{M}, \langle E = \{\varphi \mid \bigcup_{\kappa<\lambda} \mathcal{M}_\kappa \models_1 \varphi\}, A = \{\varphi \mid \bigcup_{\kappa<\lambda} \mathcal{M}_{\kappa n+1} \models_0 \varphi\}, F = \{\varphi \mid \bigcup_{\kappa<\lambda} \mathcal{M}^B_\kappa \models_{\mathrm{n}} \varphi\}\rangle\rangle$.

The last part is to take a look at the fixed point of this construction. As soon as we have an interesting arithmetical theory we can use the following procedure to get to the fixed-point models.

Unfortunately, this is not enough. There are a few problems still to be solved. The first problem is with sentences that are not decided by a model. They do not behave nicely in the provability models, which results in the external logic not being classical. The second problem is caused by the first one. **BAT**, as already noted, is a bit too weak to offer an interesting theory of informal provability. This is why on the propositional level **BAT** it is strengthened to **CABAT**. In the next section, we consider a couple of ways of strengthening first-order **BAT**. One of them is similar to the strengthening used on the propositional level to get **CABAT**.

6. Comments and future work

The main aim of the paper was to construct the first-order **BAT** logic. We achieved this by distinguishing between a model and its extensions. The intuition was that a model interprets only some formulas of the language, and extensions were responsible for providing an interpretation of sentences of the full language. The non-deterministic nature of the semantics then made it possible for a single model to have multiple extensions. Next, we sketched how one can adapt **BAT** to an arithmetical setting. The construction was a variation of Kripke's fixed-point construction and as such is quite general and can be used for other non-deterministic logics. The construction started with an arbitrary **BAT** model and by the construction we ended up in a unique fixed point. The other option would have been to start with an arbitrary model supplemented by the evaluation and then to apply the fixed-point construction. By doing so, we would have ended up in a model where the fixed-point construction would have been total in the sense that the interpretation of $\mathtt{B}$ would have exhausted the whole domain.

Acknowledgements. We would like to thank to reviewers for their thoughtful comments. The work is supported by Polish National Science Centre (NCN) grant SONATINA 2 number 2018/28/C/HS1/00251 (Pawel Pawlowski) and grant SONATA BIS number 2016/22/E/HS1/00304 (Rafal Urbaniak).

References

Alexander, S., 2013, "A machine that knows its own code", *arXiv preprint arXiv:1305.6080*. DOI: 10.1007/s11225-013-9491-6

Antonutti Marfori, M., 2010, "Informal proofs and mathematical rigour", *Studia Logica* 96: 261–272. DOI: 10.1007/s11225-010-9280-4

Antonutti Marfori, M., and L. Horsten, 2016, "Epistemic Church's thesis and absolute undecidability", page 254 in *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*. DOI: 10.1093/acprof:oso/9780198759591.003.0011

Antonutti Marfori, M., and L. Horsten, 2018, "Human-effective computability", *Philosophia Mathematica* 27 (1): 61–87. DOI: 10.1093/philmat/nky011

Arai, T., 1998, "Some results on cut-elimination, provable well-orderings, induction and reflection", *Annals of Pure and Applied Logic* 95 (1–3): 93–184. DOI: 10.1016/s0168-0072(98)00020-7

Beklemishev, L., 1997, "Induction rules, reflection principles, and provably recursive functions", *Annals of Pure and Applied Logic* 85 (3): 193–242. DOI: 10.1016/s0168-0072(96)00045-0

Beklemishev, L., 2003, "Proof-theoretic analysis by iterated reflection", *Archive for Mathematical Logic* 42 (6): 515–552. DOI: 10.1007/978-3-319-22156-4_9

Bellantoni, S., and M. Hofmann, 2002, "A new 'feasible' arithmetic" *The Journal of Symbolic Logic* 67 (1): 104–116. DOI: 10.2178/jsl/1190150032

Carlson, T., 2016, "Collapsing knowledge and epistemic Church's thesis", pages 129–147 in L. Horsten and P. Welch (eds.), *Gödel's Disjunction: The scope and limits of mathematical knowledge*, Oxford Scholarship Online. DOI: 10.1093/acprof:oso/9780198759591.003.0006

Carlson, T. J., 2000, "Knowledge, machines, and the consistency of Reinhardt's strong mechanistic thesis", *Annals of Pure and Applied Logic* 105 (1–3): 51–82. DOI: 10.1016/s0168-0072(99)00048-2

Enderton, H., 1977, *Elements of Set Theory*, Academic Press, New York. DOI: 10.1016/C2009-0-22079-4

Flagg, R., 1985, "Church's thesis is consistent with epistemic arithmetic", pages 121–172 in S. Shapiro (ed.), *Intensional Mathematics*, Studies in Logic and the Foundations of Mathematics, Vol. 113, North-Holland. DOI: 10.1016/s0049-237x(08)70142-3

Flagg, R., and H. Friedman, 1986, "Epistemic and intuitionistic formal systems", *Annals of Pure and Applied Logic* 32 (1): 53–60. DOI: 10.1016/0168-0072(86)90043-6

Friedman, H., and M. Sheard, 1989, "The equivalence of the disjunction and existence properties for modal arithmetic", *The Journal of Symbolic Logic* 54 (4): 1456–1459. DOI: 10.2307/2274825

Goodman, N. D., 1984, "Epistemic arithmetic is a conservative extension of intuitionistic arithmetic", *Journal of Symbolic Logic* 49 (1): 192–203. DOI: 10.2307/2274102

Goodman, N. D., 1986, "Flagg realizability in arithmetic", *The Journal of Symbolic Logic* 51 (2): 387–392. DOI: 10.2307/2274062

Halbach, V., 2011, *Axiomatic Theories of Truth*, Cambridge University Press.

Halbach, V., and L. Horsten, 2000, “Two proof-theoretic remarks on EA + ECT”, *Mathematical Logic Quarterly* 46 (4): 461–466. DOI: 10.1002/1521-3870(200010)46:4<461::aid-malq461>3.0.co;2-i

Heylen, J., 2013, “Modal-epistemic arithmetic and the problem of quantifying in”, *Synthese* 190 (1): 89–111. DOI: 10.1007/s11229-012-0154-3

Horsten, L., 1994, “Modal-epistemic variants of Shapiro’s system of epistemic arithmetic”, *Notre Dame Journal of Formal Logic* 35 (2): 284–291. DOI: 10.1305/ndjfl/1094061865

Horsten, L., 1996, “Reflecting in epistemic arithmetic”, *The Journal of Symbolic Logic* 61: 788–801. DOI: 10.2307/2275785

Horsten, L., 1997, “Provability in principle and controversial constructivistic principles”, *Journal of Philosophical Logic* 26 (6): 635–660. DOI: 10.1023/a:1017954806119

Horsten, L., 1998, “In defence of epistemic arithmetic”, *Synthese* 116: 1–25. DOI: 10.1023/A:1005016405987

Horsten, L., 2002, “An axiomatic investigation of provability as a primitive predicate”, pages 203–220 in V. Halbach and L. Horsten (eds.), *Principles of Truth*, Hansel-Hohenhausen. DOI: 10.1515/9783110332728.203

Horsten, L., 2006, “Formalizing Church’s thesis”, *Church’s Thesis After* 70: 253–267. DOI: 10.1515/9783110325461.253

Kearns, J. T., 1981, “Modal semantics without possible worlds”, *The Journal of Symbolic Logic* 46 (1): 77–86. DOI: 10.2307/2273259

Koellner, P., 2016, “Gödel’s disjunction”, in L. Horsten and P. Welch (eds.), *Gödel’s Disjunction: The Scope and Limits of Mathematical Knowledge*, Oxford Scholarship Online. DOI: 10.1093/acprof:oso/9780198759591.003.0007

Kripke, S. A., 1975, “Outline of a theory of truth”, *Journal of Philosophy* 72 (19): 690–716. DOI: 10.2307/2024634

Leitgeb, H., 2009, “On formal and informal provability”, pages 263–299 in O. Bueno and Ø. Linnebo (eds.), *New Waves in Philosophy of Mathematics*, New York: Palgrave Macmillan. DOI: 10.1057/9780230245198_13

Montague, R., 1963, “Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability”, *Acta Philosophica Fennica* (16): 153–167. DOI: 10.2307/2271809

Myhill, J., 1960, “Some remarks on the notion of proof”, *Journal of Philosophy* 57 (14): 461–471. DOI: 10.2307/2023664

Omori, H., and D. Skurt, 2016, "More modal semantics without possible worlds", *IfCoLog Journal of Logics and their Applications* 3 (5): 815–845.

Pawlowski, P., and R. Urbaniak, 2018, "Many-valued logic of informal provability: A non-deterministic strategy", *The Review of Symbolic Logic* 11 (2): 207–223. DOI: 10.1017/S1755020317000363

Rav, Y., 1999, "Why do we prove theorems?", *Philosophia Mathematica* 7 (1): 5–41. DOI: 10.1093/philmat/7.1.5

Rav, Y., 2007, "A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices", *Philosophia Mathematica* 15 (3): 291–320. DOI: 10.1093/philmat/nkm023

Reinhardt, W. N., 1985, "The consistency of a variant of, Church's thesis with an axiomatic theory of an epistemic notion", *Revista Colombiana de Matemáticas* 19 (1-2): 177–200.

Reinhardt, W. N., 1986, "Epistemic theories and the interpretation of Gödel's incompleteness theorems", *Journal of Philosophical Logic* 15 (4): 427–74. DOI: 10.1007/bf00243392

Rin, B. G., and S. Walsh, 2016, "Realizability semantics for quantified modal logic: Generalizing Flagg's 1985 construction", *The Review of Symbolic Logic* 9 (4): 752–809. DOI: 10.1017/S1755020316000095

Shapiro, S., 1985, "Epistemic and intuitionistic arithmetic", in S. Shapiro (ed.), *Intensional Mathematics*, Studies in Logic and the Foundations of Mathematics, Vol. 113, North-Holland. DOI: 10.1016/s0049-237x(08)70138-1

Stern, J., 2015, *Toward Predicate Approaches to Modality*, Trends in Logic, Springer. DOI: 10.1007/978-3-319-22557-9

Tanswell, F., 2015, "A problem with the dependence of informal proofs on formal proofs", *Philosophia Mathematica* 23 (3): 295–310. DOI: 10.1093/philmat/nkv008

Pawel Pawlowski and Rafal Urbaniak
Institute of Philosophy, Sociology and Journalism
University of Gdańsk, Poland
{haptism89,rfl.urbaniak}@gmail.com