THE REVIEW OF SYMBOLIC LOGIC
Volume 11, Number 2, June 2018

MANY-VALUED LOGIC OF INFORMAL PROVABILITY: A NON-DETERMINISTIC STRATEGY

PAWEL PAWLOWSKI

Centre for Logic and Philosophy of Science, Ghent University
and

RAFAL URBANIAK

Centre for Logic and Philosophy of Science, Ghent University and
Institute of Philosophy, Sociology and Journalism, University of Gdansk

Abstract. Mathematicians prove theorems in a semi-formal setting, providing what we'll call *informal proofs*. There are various philosophical reasons not to reduce informal provability to formal provability within some appropriate axiomatic theory (Leitgeb, 2009; Marfori, 2010; Tanswell, 2015), but the main worry is that we seem committed to all instances of the so-called reflection schema: $B(\varphi) \rightarrow \varphi$ (where B stands for the informal provability predicate). Yet, adding all its instances to any theory for which Löb's theorem for B holds leads to inconsistency.

Currently existing approaches (Shapiro, 1985; Horsten, 1996, 1998) to formalizing the properties of informal provability avoid contradiction at a rather high price. They either drop one of the Hilbert-Bernays conditions for the provability predicate, or use a provability operator that cannot consistently be treated as a predicate.

Inspired by (Kripke, 1975), we investigate the strategy which changes the underlying logic and treats informal provability as a partial notion. We use non-deterministic matrices to develop a three-valued logic of informal provability, which avoids some of the above mentioned problems.

§1. Formal vs informal provability.

In common mathematical practice mathematical claims are justified or proven in an informal way. Informal proofs are not stated in a proper formal language, but rather in a mixture of a native language expanded with mathematical notation. They abide by a different canon of rigour than formal proofs. From the perspective of fully formalized proofs, in informal proofs some inference steps seem to be missing. It is not even clear what counts as an axiom and some simple facts are said to be justified merely on the basis of mathematical insight (or intuition). Yet, the existence of an informal proof of a mathematical statement is a very good reason to take the claim to be true (or established). Provability in the above sense will be called *informal provability*.

On the other hand, there exist *formal proofs*, given in a fully formalized axiomatic theory by means of a fully specified formal proof system. Formal provability in this sense is always relative to some axiomatic theory.

The relation between formal and informal provability is often explained by the so-called *standard view*. The proponents of this view argue that any informal proof is at

Received: February 8, 2016.
2010 *Mathematics Subject Classification*: 03A05, 00A30.
Key words and phrases: informal provability, many-valued logic, non-deterministic semantics, Löb's theorem, paradoxes of provability.

doi:10.1017/S1755020317000363

least in principle reducible to a proper proof in an appropriate axiomatic system (usually, ZFC). On this view, informal proofs are just sloppy, incomplete versions of formal proofs.

Yet, there are reasons to think that there is at least a conceptual difference between these notions. Some philosophers (Horsten, 2002; Leitgeb, 2009; Marfori, 2010) argue against the standard view. According to them, the standard view does not fully explain why informal proofs are quite good at convincing mathematicians, whereas formal ones are not. They also point out that the role of axioms and definitions is quite different in both kinds of proofs and that there is no clear procedure for converting an informal proof into a formal one or for associating informal proofs with their formal counterparts (Tanswell, 2015).

For us, the most important argument for the difference between formal and informal provability lies in general principles valid for informal provability. There is an agreement that principles of formal provability are satisfied for informal provability. Yet, those principles are not enough, they do not express the reliability of informal proofs. The additional principle, which is thought to be sound for informal provability is *reflection schema*. It roughly says that any informally provable sentence is also true.

Unfortunately, the language of any arithmetical theory T containing Peano arithmetic, cannot contain a formula for which the combination described above holds. We will elaborate on this in §7.

Current theories of informal provability (Horsten, 2002) have to face the cost of adding all the instances of the reflection schema for a new informal provability predicate. It is quite high: some other principles which intuitively hold for informal provability (such as some of Hilbert-Bernays derivability conditions) have to go.

Another strategy of constructing a theory of informal provability is to pay a different price for adding all the instances of the reflection schema for informal provability. In such systems, provability can only be treated as an operator, but, under the threat of inconsistency, not as a predicate (Shapiro, 1985).

By the end of the article the reader will notice that in the system we present all the instances of reflection for informal provability can be added, while (some variants of) other intuitive principles are preserved. Our current goal is only to discuss the propositional level of the inferential machinery, so showing that our strategy can be consistently extrapolated to informal provability *predicate* lies beyond the scope of this article. However, it will become clear that the reasons which blocked the move to the predicate level for other systems are not going to constitute a similar obstacle in the case at hand.

We would like to suggest an unexplored strategy out of these difficulties, which stems from the intuitions that some of the solutions proposed in Kripke's theory of truth can be used to approach provability.

Instead of dropping or restricting Hilbert-Bernays conditions we will change the underlying logic. Most notably, our goal is to explore the option of treating mathematical provability as a partial notion—after all, there is an intuitive division of mathematical claims into provable, refutable and undecidable.

In the standard Kripke construction we rely on the Strong Kleene logic to deal with the partial truth predicate. But Kleene logic is not appropriate for modelling informal provability. It seems that informal provability has not truth-functional nature. Generally it is not always the case that disjunctions of two independent sentences of a given theory are independent of that theory.

We'll limit our attention to the arithmetical setting since it is at the same time quite simple to handle and expressive enough. The logic developed in this article is propositional and it still needs to be further developed to the full first-order version. Yet, some properties of informal provability can be studied in the propositional setting, and doing so seems like a good place to start, especially as it will turn out on page 7 that the propositional level is where most of the action is.

§2. The non-deterministic strategy. Let $\mathcal{L}$ be a propositional language (understood as the set of all well-formed formulas) constructed from propositional variables $W = \{p_1, p_2, \dots\}$ and Boolean connectives ($\neg, \wedge, \vee, \rightarrow, \equiv$) in the standard manner. We will use Greek letters $\varphi, \psi, \dots$ as meta-variables for formulas. The language that results from extending the set of Boolean connectives with one unary operator B will be denoted $\mathcal{L}_{\mathcal{B}}$. We will use B to express provability within the object language.

By an *assignment* we mean any function $\upsilon : W \mapsto Val$, where Val is a set of values. By an *evaluation* e_υ built over an assignment υ we will mean a function assigning values to all well-formed formulas ($e_\upsilon : \mathcal{L} \mapsto Val$) agreeing with υ 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.

It is possible to construct sensible logics for which this uniqueness fails. One nice example is the propositional variant of paraconsistent logic CLuN (Batens & Clercq, 2004).[1] The standard semantics of CLuN is similar to the semantics of classical propositional logic with one difference: the truth conditions for negation are different.

Both for classical logic and for CLuN we have $Val = \{0, 1\}$. In classical propositional logic $e_\upsilon(\neg\varphi) = 1$ iff $e_\upsilon(\varphi) = 0$. In CLuN this equivalence is weakened to an implication: if $e_\upsilon(\varphi) = 0$, then $e_\upsilon(\neg\varphi) = 1$. (Clauses for the rest of connectives are the same as in classical propositional logic.) In other words, CLuN allows for gluts for negation: both φ and $\neg\varphi$ can be true in one and the same evaluation.

The standard semantics of CLuN has another interesting feature. It is non-deterministic: assignments of values to propositional variables do not uniquely determine evaluations of all formulas. One and the same assignment might be extended in different ways to different evaluations, as long as they obey classical clauses for connectives other than negation and the implication above for negation. For instance, if $\upsilon(p) = 1$, there is one evaluation e^1_υ such that $e^1_\upsilon(\neg p) = 0$ and there is another one e^2_υ such that $e^2_\upsilon(\neg p) = 1$.

§3. Non-deterministic matrices for provability. We apply a similar trick to develop a non-deterministic semantics for a logic which would help us model the notion of informal provability.

The logic will be three-valued: we take the set of values $Val = \{0, n, 1\}$. The intended interpretation of the values is as follows. 1 stands for (*informally*) *provable*, 0 represents (*informal*) *refutability* and n stands for *being neither* (*informally*) *provable, nor* (*informally*) *refutable*. This is the *synchronic* interpretation, on which whether something is informally provable or refutable doesn't depend on the stage of the development of mathematics or on anyone's state of knowledge.

[1] A general framework for non-deterministic logics can be found in (Avron & Zamanski, 2011). Particular systems discussed there have, however, quite different motivation from ours, and quite different matrices.

https://doi.org/10.1017/S1755020317000363 Published online by Cambridge University Press

We will develop a logic with provability values. One might think that this approach is strictly speaking antirealist (because the values aren't interpreted in terms of what happens in the "external world" but rather in terms of the properties of the system), but we are not deeply committed to this way of thinking about it. One can be a truth-value realist or an ontological realist while using our logic to reason about provability and at the same time being aware that provability and truth are quite different.

Perhaps, one can think of these values *diachronically* by assuming that what is informally provable changes through time as new proofs are developed. In this sense, 1 would stand rather for *being informally proven*, 0 for *being informally refuted* and n for *being neither*. While we conjecture that this interpretation should abide by the same intuitively valid inferential principles, due to the limited scope of this article we have to postpone a proper discussion of this reading aside.

Recall that $\mathcal{L}_{\mathcal{B}}$ is the propositional language with provability operator B. We now move to specifying the semantics for connectives of $\mathcal{L}_{\mathcal{B}}$ by means of non-deterministic matrices. Let's start with negation:

- $e_v(\varphi) = 1$ iff $e_v(\neg\varphi) = 0$.
- $e_v(\varphi) = 0$ iff $e_v(\neg\varphi) = 1$.
- $e_v(\varphi) = n$ iff $e_v(\neg\varphi) = n$.

A given formula is informally provable iff its negation is informally refutable. A given formula is informally refutable iff its negation is informally provable. A formula is undetermined iff its negation is.

For disjunction we introduce non-deterministic clauses. The equivalence

$$e_v(\varphi \vee \psi) = 1 \text{ iff } e_v(\varphi) = 1 \text{ or } e_v(\psi) = 1$$

is weakened to one direction only:

$$\text{If } e_v(\varphi) = 1 \text{ or } e_v(\psi) = 1 \text{ then } e_v(\varphi \vee \psi) = 1.$$

The full set of clauses for disjunction is:

- If $e_v(\varphi) = 1$ or $e_v(\psi) = 1$, then $e_v(\varphi \vee \psi) = 1$.
- $e_v(\varphi \vee \psi) = 0$ iff $e_v(\varphi) = e_v(\psi) = 0$.
- If $e_v(\varphi) = 0$, $e_v(\psi) = n$, then $e_v(\varphi \vee \psi) = n$.
- If $e_v(\varphi) = n$, $e_v(\psi) = 0$, then $e_v(\varphi \vee \psi) = n$.
- If $e_v(\varphi) = n$, $e_v(\psi) = n$, then $e_v(\varphi \vee \psi) = n$ or $e_v(\varphi \vee \psi) = 1$.

The intention behind the introduction of nondeterminism is this. We want to allow for the possibility of there being informally (absolutely) undecidable mathematical sentences (without saying that there are any). Yet, even for such sentences (if there are any), some disjunctions built from them might be informally undecidable, while some others will be informally provable. 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 φ not being decidable.

For instance, you might be inclined to think that Continuum Hypothesis (CH) is informally undecidable, while $CH \vee \neg CH$ is still informally provable, being a logical truth. This however, clearly doesn't mean that $CH \vee CH$ is provable, and so not every disjunction of undecidable sentences is decided.

Conjunction $\varphi \wedge \psi$ is taken to have the same matrix as $\neg(\neg\varphi \vee \neg\psi)$, and so:

- If $e_v(\varphi) = 0$ or $e_v(\psi) = 0$ then $e_v(\varphi \wedge \psi) = 0$.
- $e_v(\varphi \wedge \psi) = 1$ iff $e_v(\varphi) = e_v(\psi) = 1$.

- If $e_v(\varphi) = 1$, $e_v(\psi) = n$ then $e_v(\varphi \wedge \psi) = n$.
- If $e_v(\varphi) = n$, $e_v(\psi) = 1$ then $e_v(\varphi \wedge \psi) = n$.
- If $e_v(\varphi) = n$, $e_v(\psi) = n$ then $e_v(\varphi \wedge \psi) = n$ or $e_v(\varphi \wedge \psi) = 0$.

The idea for the indeterministic case for conjunction is following. For some informally undecidable sentences we may be able to prove that they are mutually contradictory, which makes their conjunction informally refutable. For some others it may be impossible, and so their conjunction remains informally undecidable.[2]

Implication is taken to have the same matrix as $(\neg\varphi \vee \psi)$, and so:[3]

- If $e_v(\varphi) = 0$ then $e_v(\varphi \to \psi) = 1$.
- $e_v(\varphi \to \psi) = 0$ iff $e_v(\varphi) = 1$ and $e_v(\psi) = 0$.
- If $e_v(\varphi) = n$, $e_v(\psi) = n$ then $e_v(\varphi \to \psi) = n$ or $e_v(\varphi \to \psi) = 1$.
- If $e_v(\varphi) = n$, $e_v(\psi) = 1$ then $e_v(\varphi \to \psi) = 1$.
- If $e_v(\varphi) = n$, $e_v(\psi) = 0$ then $e_v(\varphi \to \psi) = n$.
- If $e_v(\varphi) = 1$, $e_v(\psi) = n$ then $e_v(\varphi \to \psi) = n$.
- If $e_v(\varphi) = 1$, $e_v(\psi) = 1$ then $e_v(\varphi \to \psi) = 1$.

Equivalence has the same matrix as $((\varphi \to \psi) \wedge (\psi \to \varphi))$, and therefore:

- $e_v(\varphi \equiv \psi) = 1$ if $e_v(\varphi) = e_v(\psi) = 1$ or $e_v(\varphi) = e_v(\psi) = 0$.
- $e_v(\varphi \equiv \psi) = 0$ if ($e_v(\varphi) = 1$ and $e_v(\psi) = 0$) or ($e_v(\varphi) = 0$ and $e_v(\psi) = 1$).
- $e_v(\varphi \equiv \psi) = n$ if exactly one of ψ, φ has value n.

While this doesn't need to be stated and follows from the above, notice that if $e_v(\varphi) = e_v(\psi) = n$ then $e_v(\varphi \equiv \psi)$ is either $0, n, 1$.

The intended reading of $\mathtt{B}\varphi$ is 'φ is informally provable.' The matrix for $\mathtt{B}$ is non-deterministic:

- $e_v(\mathtt{B}\varphi) = 1$ iff $e_v(\varphi) = 1$.
- If $e_v(\mathtt{B}\varphi) = 0$, then $e_v(\varphi) = 0$ or $e_v(\varphi) = n$.
- If $e_v(\mathtt{B}\varphi) = n$, then $e_v(\varphi) = n$.

The intuition behind these conditions is the following.

If a formula is informally provable ($e_v(\varphi) = 1$), then giving its own proof is also a proof of its provability ($e_v(\mathtt{B}\varphi) = 1$), and the other way around. If a formula is informally refutable $e_v(\varphi) = 0$, then giving its own refutation is also a refutation of its provability ($e_v(\mathtt{B}\varphi) = 0$). If a formula is informally undecidable ($e_v(\varphi) = n$), 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 ($e_v(\mathtt{B}\varphi) = 0$). Second, it may be the case that its absolute informal undecidability is not informally provable, and so its absolute informal provability is informally undecidable ($e_v(\mathtt{B}\varphi) = n$).

---

[2] Notice that just because $\varphi \wedge \psi$ has the same truth table as $\neg(\neg\varphi \vee \neg\psi)$, it doesn't follow that the substitution of expressions of this form preserves the value under an interpretation. This will fail due to indeterminacy. (The substitutability will be regained once we move from BAT logic to CABAT logic.)

[3] There are other ways to introduce implication in many-valued contexts, but given how, as it will turn out, the behavior of implication deserves additional attention, we postpone the discussion of various ways it can or cannot be introduced to another paper.

All these conditions are captured by the following 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	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

Because we interpret value 1 as **B**eing an **A**bsolute **T**heorem (BAT), we call the logic thus obtained BAT-logic and we'll use the bat symbol 🦇 to denote its consequence relation, which we define as follows.

A BAT-assignment υ is a function from propositional variables W to $\{0, n, 1\}$. A BAT-evaluation over an assignment υ is a function which assigns values to all formulas of $\mathcal{L}_{\mathcal{B}}$, agrees with υ 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 Γ🦇φ, where Γ is a set of formulas, we will mean that any BAT-evaluation which assigns 1 to all formulas in Γ assigns 1 to formula φ. We say that φ is a BAT-tautology iff $\emptyset$🦇φ. We say that φ is a BAT-countertautology iff $\emptyset$🦇$\neg\varphi$.

§4. Properties of BAT.

First, note:

THEOREM 4.1. *BAT-logic has neither tautologies nor counter-tautologies.*

Proof. Because n is contagious, it is easy to see by induction on formula complexity that for the assignment υ which assigns n to all propositional variables and for any formula φ there will be a way of extending υ to e_υ such that $e_\upsilon(\varphi)$ will be n. □

THEOREM 4.2. *Let $\Gamma \subseteq L$ and $\varphi \in L$, then for any set of formulas Γ and any formula φ if Γ🦇φ then $\Gamma \models \varphi$, where $\models$ is the classical consequence relation (we'll use $\models$ in this sense throughout the article).*

Proof. By contraposition suppose that $\Gamma \not\models \varphi$. Then there is an evaluation over an assignment υ such that $e_\upsilon(\psi) = 1$ for all $\psi \in \Gamma$ and $e_\upsilon(\varphi) = 0$. It is easy to see that e_υ is also a BAT-evaluation. For the assignment υ is classical (it is into { 0,1 }) and BAT-evaluations behave in the same manner as classical evaluations over classical assignments. Hence, there is at least one BAT-assignment which makes all formulas of Γ true and φ false, which means that it is not the case that Γ🦇φ (that is, Γ/🦇φ). □

Quite expectedly, classical consequence is strictly stronger than BAT- consequence:

THEOREM 4.3. *There are some $\Gamma \subseteq L$ and $\varphi \in L$ such that $\Gamma \models \varphi$ but Γ/🦇φ.*

Proof. For instance, $\neg(\varphi \wedge \psi)$/🦇$\neg\varphi \vee \neg\psi$. Take any evaluation for which $e_\upsilon(\varphi) = n = e_\upsilon(\psi)$, $e_\upsilon(\varphi \wedge \psi) = 0$, $e_\upsilon(\neg\varphi \vee \neg\psi) = n$. □

The following table illustrates the assessment of some standard classically valid inference patterns in BAT.

Premises	Conclusion	🦇 ?
φ	$\neg\neg\varphi$	Yes
$\neg\neg\varphi$	φ	Yes
$\neg(\varphi \wedge \psi)$	$\neg\varphi \vee \neg\psi$	No
$\neg\varphi \vee \neg\psi$	$\neg(\varphi \wedge \psi)$	No
$\neg(\varphi \vee \psi)$	$\neg\varphi \wedge \neg\psi$	Yes
$\neg\varphi \wedge \neg\psi$	$\neg(\varphi \vee \psi)$	Yes
$\varphi \wedge \psi$	$\psi \wedge \varphi$	Yes
$\varphi \vee \psi$	$\psi \vee \varphi$	No
$\varphi \rightarrow \psi$	$\neg\psi \rightarrow \neg\varphi$	No
$\varphi \wedge \psi$	$\varphi \vee \psi$	Yes
φ	$\psi \rightarrow \varphi$	Yes
$\varphi \wedge (\psi \vee \chi)$	$(\varphi \vee \psi) \wedge (\varphi \vee \chi)$	No
$\varphi \vee (\psi \wedge \chi)$	$(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	No
$(\varphi \vee \psi) \wedge (\varphi \vee \chi)$	$\varphi \wedge (\psi \vee \chi)$	No
$(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	$\varphi \vee (\psi \wedge \chi)$	No
$\varphi \rightarrow \psi, \psi \rightarrow \chi$	$\varphi \rightarrow \chi$	No
$\neg\psi$	$\neg(\varphi \wedge (\varphi \rightarrow \psi))$	No
φ	$\neg[(\varphi \rightarrow \psi) \wedge \neg\psi]$	No
$\varphi \vee \psi, \neg\varphi$	ψ	Yes
$\varphi \rightarrow \psi, \neg\psi$	$\neg\varphi$	Yes
$\neg\psi \wedge (\varphi \rightarrow \psi)$	$\neg\varphi$	Yes
$\varphi \wedge (\varphi \rightarrow \psi)$	ψ	Yes
$\varphi, (\varphi \rightarrow \psi)$	ψ	Yes
$\varphi \rightarrow \psi$	$(\varphi \wedge \lambda) \rightarrow \psi$	No
$\varphi \rightarrow \psi$	$\varphi \rightarrow (\psi \vee \lambda)$	No
$\varphi \wedge (\psi \wedge \chi)$	$(\varphi \wedge \psi) \wedge \chi$	Yes
$\varphi \vee (\psi \vee \chi)$	$(\varphi \vee \psi) \vee \chi$	No
$(\varphi \wedge \psi) \wedge \chi$	$\varphi \wedge (\psi \wedge \chi)$	Yes
$(\varphi \vee \psi) \vee \chi$	$\varphi \vee (\psi \vee \chi)$	No

Notice that *Modus Ponens* works in both formulations, while its contraposed form fails. Similarly, *Modus Tollens* works in both forms, while its contraposed form fails. This entails:

THEOREM 4.4. *It is not generally the case that if* φ 🦇 ψ *then* $\neg\psi$ 🦇 $\neg\varphi$.

Observe that disjunction is neither commutative nor associative. Take the assignment υ 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:

$$e^1_\upsilon(\varphi) = n = e^1_\upsilon(\psi),$$
$$e^2_\upsilon(\varphi) = 1, e^2_\upsilon(\psi) = n,$$
$$e^3_\upsilon(\varphi) = n, e^3_\upsilon(\psi) = 1,$$
$$e^4_\upsilon(\varphi) = 1 = e^4_\upsilon(\psi).$$

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

§5. Strengthening BAT. Usually, to obtain a stronger logic from a logic with non-deterministic semantics we have to limit the range of available possible extensions of given assignments.[4] We would like to propose our own solution to this problem in terms of either enriching one logic by another one or by additional closure condition.

DEFINITION 5.1. *Let L be a logic. We say that a BAT-evaluation e belongs to the L-filtered set of BAT-evaluations just in case the following conditions hold*:

1. *For any two formulas* φ, ψ*, if* $\models_L \varphi \equiv \psi$ *then* $e(\varphi) = e(\psi)$*,*
2. *For any L-tautology* φ*,* $e(\varphi) = 1$*,*
3. *For any L-countertautology* φ*,* $e(\varphi) = 0$*.*

We will focus on the case where L is classical logic (L=CL), and we simply use $\models$ to denote the classical consequence relation. By Γ 🦇$_{CL}$ φ 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 CLBAT.

We may be also inclined to strengthen BAT-logic in a different manner. A quite intuitive way to go is to close BAT-logic under classical consequence. It can be done by the following condition:

DEFINITION 5.2 (Closure condition). *An extension of BAT (in* $\mathcal{L}_\mathcal{B}$*) satisfies the closure condition just in case for all* $\mathcal{L}_\mathcal{B}$*-formulas* $\varphi_1, \varphi_2, \dots, \varphi_k, \psi$ *such that*

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

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

The result of closing BAT-logic under the closure condition will be called CABAT logic and its consequence relation will be denoted by 🦇$_C$.

It turned out that the above conditions are equivalent, and so are the resulting logics:

THEOREM 5.3. Γ 🦇$_C$ φ *iff* Γ 🦇$_{CL}$ φ*.*

Proof. We will show that the set of CL-filtered BAT-evaluations respects the closure condition and that the set of evaluations for which the closure condition holds is exactly the set of CL-filtered BAT-evaluations.

$\Rightarrow$: Let $\Gamma = \{\varphi_1 \dots \varphi_n\}$ and φ be such that $\Gamma \models \varphi$. We have to show that for any CLBAT- evaluation e, if $e(\mathtt{B}\varphi_1) = e(\mathtt{B}\varphi_2) = \dots = e(\mathtt{B}\varphi_n) = 1$ then $e(\mathtt{B}\varphi) = 1$. Assume the antecedent. By the deduction theorem for classical propositional logic we know that $\models \bigwedge_{i=1}^{i=n} \varphi_i \rightarrow \varphi$. By the assumption and the definition of CLBAT-evaluation, $e(\bigwedge_{i=1}^{i=n} \varphi_i \rightarrow \varphi) = 1$. Since any CLBAT-evaluation which assigns 1 to all conjuncts has to assign 1 to the whole conjunction, we have $e(\bigwedge_{i=1}^{i=n} \varphi_i) = 1$. By the matrix for implication it follows that $e(\varphi) = 1$. By the matrix of $\mathtt{B}$, we have $e(\mathtt{B}\varphi) = 1$.

[4] The most common way to strengthen a non-deterministic logic is to use the level-evaluation method (Coniglio, Fariñas del Cerro, & Peron, 2015) Due to simplicity, we prefer our method.

$\Leftarrow$: To show that any CABAT-evaluation e is also a CLBAT-evaluation, since both sets are subsets of BAT-evaluation, we only need to check that CABAT-evaluations respect the filtration conditions.

We will start with the third condition. Let φ be a formula in $\mathcal{L}_{\mathcal{B}}$. First, suppose that for any classical evaluation e, $e(\varphi) = 0$. It follows that for any classical evaluation $e(\neg\varphi) = 0$, so $\models \neg\varphi$. We have to show that φ has value 0 in any CABAT-evaluation. By the closure condition $\text{🦇}_C\ \mathtt{B}\neg\varphi$. By the matrix for $\mathtt{B}$, $\text{🦇}_C\ \neg\varphi$. So any CABAT-evaluation assigns 0 to formula φ.

Next, consider the second condition. Let φ be a formula in $\mathcal{L}_{\mathcal{B}}$. First, suppose that for any classical evaluation e, $e(\varphi) = 1$. We have to show that φ has value 1 in any CABAT-evaluation. By the closure condition $\text{🦇}_C\ \mathtt{B}\varphi$. By the matrix for $\mathtt{B}$, $\text{🦇}_C\ \varphi$. So any CABAT-evaluation assigns 1 to formula φ.

Finally, consider the first condition. Suppose now that for any classical evaluation $e(\varphi) = e(\psi)$. In other words, $\models \varphi \equiv \psi$. By the deduction theorem, $\varphi \models \psi$, $\psi \models \varphi$, $\neg\psi \models \neg\varphi$ and $\neg\varphi \models \neg\psi$. We want to show that for any CABAT-evaluation e_c, $e_c(\varphi) = e_c(\psi)$. By the closure condition, $\mathtt{B}\varphi\, \text{🦇}_C\, \mathtt{B}\psi$, $\mathtt{B}\psi\, \text{🦇}_C\, \mathtt{B}\varphi$, $\mathtt{B}\neg\psi\, \text{🦇}_C\, \mathtt{B}\neg\varphi$ and $\mathtt{B}\neg\varphi\, \text{🦇}_C\, \mathtt{B}\neg\psi$. Thus, by the matrix for $\mathtt{B}$ we have $\varphi\, \text{🦇}_C\, \psi$, $\psi\, \text{🦇}_C\, \varphi$, $\neg\varphi\, \text{🦇}_C\, \neg\psi$ and $\neg\psi\, \text{🦇}_C\, \neg\varphi$.

We will consider three cases: $e_c(\varphi) = 1$, $e_c(\varphi) = 0$ and $e_c(\varphi) = n$. We will start with the first case. It follows from $\varphi\, \text{🦇}_C\, \psi$ that $e_c(\psi) = 1$, thus by the matrix for $\mathtt{B}$, $e_c(\mathtt{B}\psi) = 1$.

For the second case if $e_c(\varphi) = 0$, then $e_c(\neg\varphi) = 1$, thus $e_c(\neg\psi) = 1$, so $e_c(\psi) = 0$, which implies by the matrix for $\mathtt{B}$. that $e_c(\mathtt{B}\psi) = 1$.

In the third case note that if $e_c(\varphi) = n$ then $e_c(\psi) = n$ because otherwise by analogous argument to the ones above from $e_c(\psi) \neq n$ we would have that $e_c(\varphi)$ is either 1 or 0, which contradicts the assumption. □

Given that both CABAT and its internal logic are closed under classical consequence, all the worries about syntactic sensitivity that applied to BAT disappear.

§6. Properties of CABAT. Quite trivially, CABAT is strictly stronger than BAT. The first interesting thing to see is that the deduction theorem is not generally valid in CABAT:

THEOREM 6.1. *If* $\text{🦇}_C\ \varphi \to \psi$ *then* $\varphi\, \text{🦇}_C\, \psi$ *but it is not always the case that* $\varphi\, \text{🦇}_C\, \psi$ *implies* $\text{🦇}_C\ \varphi \to \psi$.

In CABAT implications are stronger than the corresponding consequence relation, simply because the consequence relation informs us only about those evaluations in which all the premises have value 1. For instance, the consequence relation $\varphi\, \text{🦇}_C\, \psi$ does not determine the value of implication $\varphi \to \psi$ when both ψ and φ have value n. On the other hand, $\text{🦇}_C\ \varphi \to \psi$ uniquely determines the value of the implication under the previous assignment.

Lack of the deduction theorem makes the difference when we look at inference patterns with provability operator. Usually, principles for provability are valid in CABAT as consequence relations whereas their implicational formulations may be invalid. We are not terribly worried about that, since given our reading $\varphi\, \text{🦇}_C\, \psi$ means that if φ is informally provable then φ is and this is the phrase we intended to formalize. On the other hand, $\text{🦇}_C\ \varphi \to \psi$ says if φ is informally provable, then so is ψ and *if the antecedent is undecidable then the consequent is either provable or independent*, which is a stronger claim than $\varphi\, \text{🦇}_C\, \psi$.

https://doi.org/10.1017/S1755020317000363 Published online by Cambridge University Press

Now we will take a look at some schemas involving provability predicate. Intuitively, informal provability commutes with conjunction but not with disjunction. The fact that either φ or ψ is informally provable does not imply that we can prove either the first or the second disjunct. Of course, the consequence relation in the opposite direction ($\mathtt{B}\varphi \vee \mathtt{B}\psi \text{🦇}_C \mathtt{B}(\varphi \vee \psi)$) should hold.

Provided our reading of the consequence relation, $\mathtt{B}\varphi \text{🦇}_C \varphi$ may be seen as a certain version of the reflection schema, which is definitely a sound principle for informal provability.

The fact that a certain statement is informally provable is itself an evidence for the informal provability of the provability of that statement. Thus, iterative principles allowing to either add or subtract the $\mathtt{B}$ operator from the beginning of a formula are also natural.

The table below summarizes which inference patterns are valid in CABAT and whether the principle, according to us, is intuitive or not:

Principle	Valid?	Intuitive?
$(\mathtt{B}\varphi \wedge \mathtt{B}\psi) \text{🦇}_C \mathtt{B}(\varphi \wedge \psi)$	Yes	Yes
$\mathtt{B}(\varphi \wedge \psi) \text{🦇}_C (\mathtt{B}\varphi \wedge \mathtt{B}\psi)$	Yes	Yes
$\mathtt{B}(\varphi \vee \psi) \text{🦇}_C (\mathtt{B}\varphi \vee \mathtt{B}\psi)$	No	No
$(\mathtt{B}\varphi \vee \mathtt{B}\psi) \text{🦇}_C \mathtt{B}(\varphi \vee \psi)$	No	?
$\varphi \text{🦇}_C \mathtt{B}\varphi$	Yes	Yes
$\mathtt{B}\varphi \text{🦇}_C \varphi$	Yes	Yes
$\mathtt{B}\varphi \text{🦇}_C \neg\mathtt{B}\neg\varphi$	Yes	Yes
$\mathtt{B}\varphi \text{🦇}_C \mathtt{B}\mathtt{B}\varphi$	Yes	Yes
$\mathtt{B}\mathtt{B}\varphi \text{🦇}_C \mathtt{B}\varphi$	Yes	Yes
$\mathtt{B}(\varphi \rightarrow \psi) \text{🦇}_C (\mathtt{B}\varphi \rightarrow \mathtt{B}\psi)$	No	?
$\mathtt{B}(\varphi \rightarrow \psi), \mathtt{B}\varphi \text{🦇}_C \mathtt{B}\psi$	Yes	Yes
$\mathtt{B}(\varphi \wedge \neg\varphi) \text{🦇}_C \mathtt{B}(\psi)$	Yes	Yes
$\mathtt{B}\varphi \vee \mathtt{B}\neg\varphi$	No	No
$\mathtt{B}\varphi \vee \neg\mathtt{B}\varphi$	Yes	Yes
$\neg\mathtt{B}\varphi \text{🦇}_C \mathtt{B}(\neg\varphi)$	No	No
$\mathtt{B}(\neg\mathtt{B}\varphi) \text{🦇}_C \mathtt{B}(\neg\varphi)$	No	No
$\mathtt{B}(\neg\mathtt{B}\neg\varphi) \text{🦇}_C \neg\mathtt{B}(\neg\mathtt{B}\varphi)$	No	No

One thing that might seem worrying is that

$$(\mathtt{B}\varphi \vee \mathtt{B}\psi) \not\text{🦇}_C \mathtt{B}(\varphi \vee \psi).$$

After all, if φ is informally provable, shouldn't $\varphi \vee \psi$ also be informally provable? This worry, however, stems from the fact that the provability of a disjunction in CABAT says something weaker than that one of its disjuncts is provable—after all, $\varphi \vee \neg\varphi$ is going to be informally provable without either φ or $\neg\varphi$ being informally provable. So, we submit, the intuition should be rather captured by requiring that the following should hold:

$$\mathtt{B}\varphi \text{🦇}_C \mathtt{B}(\varphi \vee \psi) \;\; \textit{and}$$
$$\mathtt{B}\psi \text{🦇}_C \mathtt{B}(\varphi \vee \psi)$$

and indeed, they do.

Another worry might be that the following assymetry between at least *prima facie* close cousins can be observed:

https://doi.org/10.1017/S1755020317000363 Published online by Cambridge University Press

$$\mathtt{B}(\varphi \to \psi) \not🦇_C (\mathtt{B}\varphi \to \mathtt{B}\psi), \tag{Fake K}$$

$$\mathtt{B}(\varphi \to \psi), \mathtt{B}\varphi 🦇_C \mathtt{B}\psi. \tag{Real K}$$

The answer is, however, that putting $\mathtt{B}\varphi \to \mathtt{B}\psi$ on the right-hand side of 🦇$_C$ doesn't adequately capture the intuition that *if φ is informally provable, then so is ψ*. For $\mathtt{B}\varphi \to \mathtt{B}\psi$ actually contains more information than that. *If φ is informally provable, then so is ψ* tells us only what happens when φ (and so, $B\varphi$) is informally provable, while the provablity of $\mathtt{B}\varphi \to \mathtt{B}\psi$ puts further constraints on what happens if φ is not informally provable (for instance, that if it is undecidable, ψ cannot be refutable). It's (Real K) and not (Fake K) that properly captures the underlying intuition.

§7. CABAT and provability. Now, let's see the difference between using 🦇$_C$ and its provability operator on the one hand, and using Peano Arithmetic and its standard provability predicate (or the modal logic of provability GL and its provability operator, for that matter) on the other.

Quite crucially, we may want to see which principles that hold for standard formal provability predicates hold for operator $\mathtt{B}$ as well.

First, recall Hilbert-Bernays conditions for PA:

$$\mathrm{PA} \vdash \varphi \Rightarrow \mathrm{PA} \vdash Bew(\ulcorner \varphi \urcorner), \tag{HB1}$$

$$\mathrm{PA} \vdash Bew(\ulcorner \varphi \to \psi \urcorner) \to (Bew(\ulcorner \varphi \urcorner) \to Bew(\ulcorner \psi \urcorner)), \tag{HB2}$$

$$\mathrm{PA} \vdash Bew(\ulcorner \varphi \urcorner) \to Bew(\ulcorner Bew(\ulcorner \varphi \urcorner) \urcorner). \tag{HB3}$$

In the arithmetical setting standard Hilbert-Bernays conditions allow one to prove Löb's theorem:

THEOREM 7.1. *If* $\mathrm{PA} \vdash Bew(\ulcorner \varphi \urcorner) \to \varphi$ *then* $\mathrm{PA} \vdash \varphi$.

Since we'll want to make a point about how the standard proofs of the theorems that we'll discuss proceed, we'll go over them quickly.

Proof. Suppose $\mathrm{PA} \vdash Bew(\ulcorner \varphi \urcorner) \to \varphi$. By the diagonal lemma there is a formula such that $\mathrm{PA} \vdash \lambda \equiv (Bew(\ulcorner \lambda \urcorner) \to \varphi)$. Now, arguing inside Peano Arithmetic we get:

$$\lambda \to (Bew(\ulcorner \lambda \urcorner) \to \varphi), \tag{1}$$

$$Bew(\ulcorner \lambda \to (Bew(\ulcorner \lambda \urcorner) \to \varphi) \urcorner), \tag{2}$$

$$Bew(\ulcorner \lambda \urcorner) \to (Bew(\ulcorner Bew(\ulcorner \lambda \urcorner) \urcorner) \to Bew(\ulcorner \varphi \urcorner)), \tag{3}$$

$$Bew(\ulcorner \lambda \urcorner) \to Bew(\ulcorner Bew(\ulcorner \lambda \urcorner) \urcorner), \tag{4}$$

$$Bew(\ulcorner \lambda \urcorner) \to Bew(\ulcorner \varphi \urcorner), \tag{5}$$

$$Bew(\ulcorner \lambda \urcorner) \to \varphi, \tag{6}$$

$$\lambda, \tag{7}$$

$$Bew(\ulcorner \lambda \urcorner), \tag{8}$$

$$\varphi. \tag{9}$$

□

(2) is obtained by neccesitation. Next, we use the second Hilbert-Bernay's condition to distribute provability over implication twice to obtain (3). Line (4) is the third Hilbert-Bernays condition thanks to which we obtain (5) from (3). By the assumption that

https://doi.org/10.1017/S1755020317000363 Published online by Cambridge University Press

PA $\vdash Bew(\ulcorner\varphi\urcorner) \rightarrow \varphi$ we obtain (6). Applying detachment to the sentence generated by the diagonal lemma we get (7). Then by necessitation and *modus ponens*, we obtain the last two lines.

Löb's theorem is not an intuitively sound principle for informal provability. There is no reason to suppose that only those instances of the reflection schema hold for which φ is already a theorem. The intuitions are rather clear that all instances of reflection are plausible.

In the arithmetical setting we get Löb's theorem as a side-effect of the diagonal lemma. It is not something that we would like to postulate as an interesting and independently motivated principle. Rather, it is an unwanted surprising consequence. It is also one of the reasons why we cannot consistently put all together instances of the reflection schema together with HB conditions in the classical setting.

In CABAT we have certain versions of HB conditions:

$$\varphi \mathrel{\text{🦇}}_C \mathtt{B}\varphi, \tag{HB1$'$}$$

$$\mathtt{B}(\varphi \rightarrow \psi), \mathtt{B}\varphi \mathrel{\text{🦇}}_C \mathtt{B}\psi, \tag{HB2$'$}$$

$$\mathtt{B}\varphi \mathrel{\text{🦇}}_C \mathtt{BB}\varphi. \tag{HB3$'$}$$

The first condition in CABAT is a bit stronger, since it is not restricted only to theorems. The condition starts to be intuitive as soon as you recall you interpretation of $\varphi \mathrel{\text{🦇}}_C \psi$, which says that if φ is informally provable then so is ψ. Some may be worried that the above formulation of HB1, in some sense, allows to go from premises which are true (and may not be theorems) to premises which are theorems. But as we explained, according to our reading formulas on the left hand side of $\mathrel{\text{🦇}}_C$ are not true but informally provable. So the principle allows only to go from informally provable premises to informally provable premises having informal provability expressed in the object language.

One interesting question is whether the above conditions are enough to prove Löb's theorem. The key observation, in the standard proof, is that once the premises, including the one produced by an application of the diagonal lemma, are listed, the theorem follows by classical propositional logic. So it seems that the issue can be handled at the propositional level.

The natural way to go about the translation is this. We translate both Bew and $\vdash$ as $\mathtt{B}$. It is a standard practice to translate them using a single symbol (see Boolos, 1993).

Slightly more challenging is the question how to translate implications from the language of PA. The straightforward approach is to translate them as material implications in $L_{\mathtt{B}}$.

But we think it will not do justice to the original theorem. The deduction theorem does not hold for CABAT. Implications are stronger claims than consequence claims and are much harder to prove. Thus, whenever possible, we will translate $\varphi \rightarrow \psi$ in the conclusions as $\varphi \mathrel{\text{🦇}}_C \psi$. We leave implications in the premises, especially within the scope of $\mathtt{B}$. But this is not a cheap way for us to avoid an undesired consequence: by leaving material implications in the premises we make them as strong as we can.

As for sentences produced by the application of Diagonal Lemma we will build them to the assumptions.

FACT 7.2 (Löb's theorem failure). $\mathtt{B}(\mathtt{B}\varphi \to \varphi), \mathtt{B}(\lambda \to (\mathtt{B}\lambda \to \varphi)), \mathtt{B}(\mathtt{B}(\lambda \to \varphi) \to \lambda) \not\models_C \mathtt{B}\varphi$.

Proof. Just take an assignment $v(\varphi) = v(\lambda) = n$ and extend it to an evaluation where for each implication if it is possible to choose n, it should be chosen. It is easily verifiable that all the premises have value 1, and yet the conclusion has value n, while all the constraints on valuations remain satisfied. □

Why doesn't the standard argument work? Suppose e_v gives value 1 to all the premises. Since $e_v(\mathtt{B}(\lambda \to (\mathtt{B}\lambda \to \varphi))) = 1$, it follows that $e_v(\mathtt{B}\lambda \to (\mathtt{BB}\lambda \to \mathtt{B}\varphi)) = 1$. Now, in the standard proof we use the fact that $e_v(\mathtt{B}\varphi \to \mathtt{BB}\varphi) = 1$, but we cannot do that here, since in general the previous formula is not a CABAT-tautology.

In other words, it is not the case that only those instances of the reflection schema are provable for which φ is already a theorem. The lack of Löb's theorem is rather promising since it leaves open the possibility for adding all the instances of the reflection schema consistently.

We will take a quick look at two other theorems related to provability and the reflection schema.

As we already stated: there is a problem with the reflection schema in the standard setting. It is impossible to add all instances of the schema and at the same time have all Hilbert-Bernays conditions. This is shown by the Montague's paradox:

THEOREM 7.3. *Peano Arithmetic, if consistent, cannot contain* (*or be consistently extended to contain*) *a* (*possibly complex*) *predicate for which all Hilbert-Bernays conditions and all instances of the reflection schema hold.*

Proof. Suppose that there is such a predicate, call it P. We will use natural deduction system. Argue inside the theory:

1. $\lambda \equiv P(\ulcorner \neg\lambda \urcorner)$	Diagonal lemma
1.1 λ	Hypothesis
1.2 $P(\ulcorner \neg\lambda \urcorner)$	equivalence elimination: 1,1.1
1.3 $\neg\lambda$	modus ponens and reflection schema: 1.2
2. $\neg\lambda$	reductio ad absurdum: 1.1 $\to$ 1.3
3. $P(\ulcorner \neg\lambda \urcorner)$	HB 1
4. $\neg P(\ulcorner \neg\lambda \urcorner)$	1, 2
5. contradiction	3, 4.

□

To rephrase the above theorem, it is impossible, given Hilbert-Bernays conditions, to extend the theory with the inverse of the implication corresponding to the necessitation rule. The same goes for the implication directly corresponding to the necessitation rule:

THEOREM 7.4. *Peano Arithmetic, if consistent, cannot contain* (*or be consistently extended to contain*) *a* (*possibly complex*) *predicate for which all Hilbert-Bernays conditions and all instances of* $\varphi \to P(\ulcorner\varphi\urcorner)$ (*Provabilitation*) *hold and is closed under the co-necessitation rule*: *if* $P(\ulcorner\varphi\urcorner)$ *then* φ.

Proof. Suppose that there is such a predicate, call it P. We will use natural deduction system. Argue inside the theory:

1. $\neg P(\ulcorner\kappa\urcorner) \equiv \kappa$	the diagonal lemma
1.1 κ	conditional assumption
1.2 $\neg P(\ulcorner\kappa\urcorner)$	equiv elimination: 1, 1.1
1.3 $\kappa \rightarrow P(\ulcorner\kappa\urcorner)$	instance of provabilitation for κ
1.4 $\neg\kappa$	MTT: 1.3, 1.2
2. $\neg\kappa$	conditional assumption discharge: 1.1 $\leftarrow$ 1.4
3. $P(\ulcorner\kappa\urcorner)$	equiv elimination: 1,2
4 κ	co-necessitation: 3
5. contradiction	2,4.

□

The moral is that the price for all Hilbert-Bernays conditions together with all instances of the reflection schema on the one hand, or all instances of to provabilitation one the other (assuming co-necessitation) is too high in the standard setting.

FACT 7.5 (Montague). $\mathtt{B}(\mathtt{B}\lambda \rightarrow \lambda), \mathtt{B}(\mathtt{B}(\neg\lambda) \rightarrow \lambda), \mathtt{B}(\lambda \rightarrow \mathtt{B}(\neg\lambda)) \not\models_C \lambda \wedge \neg\lambda$.

Proof. We can omit the initial $\mathtt{B}$ on the left side, since all evaluations which assign 1 to a formula $\mathtt{B}\varphi$ also assign 1 to φ.

Suppose $e_v(\lambda) = 1$. Then from $e_v(\lambda \rightarrow \mathtt{B}(\neg\lambda)) = 1 = e_v(\lambda)$ by *modus ponens*, $e_v(\mathtt{B}\neg\lambda) = 1$. Thus, $e_v(\neg\lambda) = 1$, contradiction.

Suppose $e_v(\lambda) = n$. Then it is possible to extend this assignment so that all the premises have value 1, and the conclusion has value n or 0. □

In other words, our provability conditions can be consistently extended with the reflection schema. So in a sense CABAT provability conditions are more appropriate for informal provability.

What is also interesting is the fact that the dual paradox in which we add provabilitation instead of reflection still works.

FACT 7.6 (Dual Montague). *The following consequence still holds:* $\mathtt{B}(\lambda \rightarrow \neg\mathtt{B}\lambda), \mathtt{B}(\neg\mathtt{B}(\lambda) \rightarrow \lambda), \mathtt{B}(\lambda \rightarrow \mathtt{B}\lambda), \mathtt{B}(\neg\lambda \rightarrow \mathtt{B}\neg\lambda) \models_C \lambda \wedge \neg\lambda$.

Proof. Similarly we can omit all occurrences of $\mathtt{B}$, in all formulas of the form $\mathtt{B}\varphi$. Let e_v be an evaluation under which all the premises have value 1. Then, since $e_v(\lambda \rightarrow \mathtt{B}\lambda) = 1 = e_v(\lambda \rightarrow \neg\mathtt{B}\lambda)$, it follows that $e_v(\neg\lambda) = 1$. Note that $e_v(\neg\mathtt{B}\lambda \rightarrow \lambda)$ implies, by *modus tollens*, $e_v(\neg\neg\mathtt{B}\lambda) = 1$. Thus, $e_v(\mathtt{B}\lambda) = 1 = e_v(\lambda)$. Contradiction. □

This is also a signal that our initial intuition behind CABAT is not completely insane: with CABAT in the background it is possible to add all the instances of the reflection schema, but not all the instances of provabilitation (which is less intuitive for informal provability).

Even if we add reflection for both λ and $\neg\lambda$, where λ states the provability of its own negation, CABAT proves that λ is informally undecidable.

FACT 7.7 (Reflection and provability). *The following consequence holds:* $\mathtt{B}(\mathtt{B}\lambda \rightarrow \lambda), \mathtt{B}(\mathtt{B}\neg\lambda \rightarrow \neg\lambda), \mathtt{B}(\mathtt{B}(\neg\lambda) \rightarrow \lambda), \mathtt{B}(\lambda \rightarrow \mathtt{B}(\neg\lambda)) \models_C \neg\mathtt{B}\neg\lambda \wedge \neg\mathtt{B}\lambda$.

Proof. As usual we omit the left-hand side $\mathtt{B}$. Take any evaluation for which $e_v(\varphi) = 1$ for all φ which are premises. The instance of the reflection schema $e_v(\mathtt{B}\neg\lambda \rightarrow \neg\lambda)$ combined with $\mathtt{B}\neg\lambda \rightarrow \lambda$ gives $e_v(\neg\mathtt{B}\neg\lambda) = 1$. This implies $e_v(\neg\lambda) = 1$, thus $e_v(\neg\mathtt{B}\lambda) = 1$. In other words every evaluation e_v which gives 1 to all the premises also gives 1 to $\neg\mathtt{B}\neg\lambda \wedge \neg\mathtt{B}\lambda$. □

That's for a sentence which says that its own negation is informally provable. Another interesting pet worth playing with is what we'll call *informal Gödel sentence*: a sentence which says of itself that it is not *informally* provable. The first thing to observe is that its formalization over CABAT doesn't lead to contradiction:

FACT 7.8 (Informal Gödel sentence). $\gamma \rightarrow \neg B\gamma, \neg B\gamma \rightarrow \gamma \not\models_C \gamma \wedge \neg\gamma$.

Proof. Because of the closure condition, $\models_C B(\neg(\gamma \wedge \neg\gamma))$, and so $\models_C \neg(\gamma \wedge \neg\gamma)$ and $e_v(\gamma \wedge \neg\gamma)$ has to be 0, independently of what $e_v(\gamma)$ is, in particular it is possible that $e_v(\gamma) = n$. So now the only thing that we need to check is if it's possible that all the premises have value 1. Indeed, if $e_v(\gamma) = n$, there is no problem with assuming that e_v assigns 1 to both premises. After all, they are just implications whose at least one argument has value n, and such implications can be assigned value 1 (see the matrix for $\rightarrow$). □

However, as soon as we add either provabilitation or reflection contradiction follows:

FACT 7.9 (Informal Gödel with provabilitation). *The following holds*:

$$\gamma \rightarrow \mathtt{B}\gamma, \gamma \rightarrow \neg B\gamma, \neg B\gamma \rightarrow \gamma \models_C \gamma \wedge \neg\gamma.$$

Proof. Again, the proof proceeds by showing that no evaluation can assing 1 to all the premises. For contradiction, suppose e_v is an evaluation which assigns 1 to all the premises. By the fact that $e_v(\gamma \rightarrow \mathtt{B}\gamma) = e_v(\gamma \rightarrow \neg B\gamma) = 1$, we have $e_v(\gamma) = 0$. Then, $e_v(\neg\gamma) = 1$. Apply *modus tollens* to the third premise, we have $e_v(\neg B\gamma) = 0$ and $e_v(B\gamma) = e_v(\gamma) = 1$, which is a contradiction. □

The above fact isn't too worrying, because provabilitation doesn't seem too plausible for informal provability to start with. Here's a more interesting case.

FACT 7.10 (Informal Gödel with reflection). *The following holds in CABAT*:

$$\mathtt{B}\gamma \rightarrow \gamma, \mathtt{B}\neg\gamma \rightarrow \neg\gamma, \mathtt{B}\gamma \rightarrow \neg\gamma, \neg\gamma \rightarrow \mathtt{B}\gamma \models_C \gamma \wedge \neg\gamma.$$

Proof. By the argument used in the previous proof, $\gamma \wedge \neg\gamma$ will always have value 0. So the only way of showing that the consequence holds is to prove that the premises cannot all have value one. For contradiction, assume e_v is an evaluation which assigns 1 to all the premises. Then, by the transitivity of implication in CABAT (guaranteed by the closure condition) we have $e_v(\neg\gamma \rightarrow \gamma) = 1$ By closure condition, $e_v(\neg\gamma \rightarrow \gamma) = e_v(\neg\neg\gamma \vee \gamma) = e_v(\gamma \vee \gamma) = e_v(\gamma) = 1$. Thus $e_v(\mathtt{B}\gamma) = 1$ and by *modus ponens* applied to the third premise $e_v(\neg\gamma) = 1$, which is a contradiction. □

One might be worried that the last fact gives rise to a paradox in the vein of (Priest, 1987; Beall, 1999), where the argument is put forward to the effect that informal mathematics is inconsistent, because one can take the sentence:

(γ) γ is not informally provable.

and both the assumption that γ is informally provable, and that it isn't informally provable lead to contradiction.

Paradoxical arguments in natural language aside, notice that given that CABAT consequence relation is defined in terms of informal provability preservation, Fact 7.10 is to be read: *if the formulae on the left-hand side of* $\models_C$ *are informally provable, then so is the formula on the right-hand side*. So, assuming reflection is informally provable, for the

https://doi.org/10.1017/S1755020317000363 Published online by Cambridge University Press

paradox to arise we actually have to assume not only that the following is true:

$$\gamma \equiv \neg B\gamma, \qquad (\gamma')$$

but also that it is *provable in informal mathematics*. This, however, is a stronger assumption that standard paradoxical arguments failed to establish: whether writing down (γ) constitutes an *informal mathematical proof* of (γ') is far from obvious and deserves a separate discussion. These and related fascinating issues, however, lie beyond the scope of this article.

§8. Conclusions. Once we intuitively divide mathematical claims into provable, refutable and independent, the question arises as to how these three classes interact with Boolean connectives. This interaction is not straightforward, because facts about whether certain claims are provable, refutable, or independent do not unambiguously determine the status of their Boolean combinations.

This obstacle, however, is not fatal. Once we move to indeterministic semantics, the basic constraints on how provability, refutability and independence behave with respect to Boolean connectives can be explicated by a formal system: BAT. The constraints captured by BAT matrices are a bit too basic, though. They don't give justice to the fact that informal mathematical provability is closed under classical consequence. Adding this requirement to BAT results in a stronger system, CABAT, which is studied in the remainder of the article.

CABAT, in contrast with BAT, doesn't fall prey to syntactic sensitivity. CABAT also validates many intuitively plausible and invalidates many intuitively implausible inference patterns for informal provability. Among the invalidated ones, we have Löb's theorem, which when applied to informal provability seems to be making the unintuitive claim that reflection holds only for those statements which are already informally provable. The failure of Löb's theorem makes all the instances of reflection schema consistent with CABAT.

§9. Acknowledgments. Research on this article has been funded by the Research Foundation Flanders (FWO) and National Science Centre (NCN, 2016/22/E/HS1/00304). The authors would like to express their gratitude to all those who commented on the earlier versions of this article or contributed to discussions about the topic when the material was presented: Cezary Cieśliński, Leon Horsten, Hannes Leitgeb, Frederik Van De Putte and Stanislav Speransky.

BIBLIOGRAPHY

Avron & Zamanski (2011). Non-deterministic semantics for logical systems. In Gabbay, D., and Guenthner, F., editors. *Handbook of Philosophical Logic*, Vol. 16. Netherlands: Springer, pp. 227–304.

Batens, D. & Clercq, K. (2004). A rich paraconsistent extension of full positive logic. *Logique Et Analyse*, **47**, 227–257.

Beall, J. (1999). Prom full blooded platonism to really full blooded platonism. *Philosophia Mathematica*, **7**(3), 322–325.

Boolos, G. (1993). *The Logic of Provability*. Cambridge: Cambridge University Press.

Coniglio, M. E., Fariñas del Cerro, L., & Peron, N. M. (2015). Finite non-deterministic semantics for some modal systems. *Journal of Applied Non-Classical Logics*, **25**(1), 20–45.

Horsten, L. (1996). Reflecting in epistemic arithemtic. *The Journal of Symbolic Logic*, **61**, 788–801.

Horsten, L. (1998). In defence of epistemic arithmetic. *Synthese*, **116**, 1–25.

Horsten, L. (2002). An axiomatic investigation of provability as a primitive predicate. In Halbach, V., and Horsten, L., editors. *Principles of Truth*. Hansel-Hohenhausen: Walter de Gruyter, pp. 203–220.

Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**(19), 690–716.

Leitgeb, H. (2009). On formal and informal provability. In Bueno, O., and Linnebo, Ø, editors. *New Waves in Philosophy of Mathematics*. New York: Palgrave Macmillan, pp. 263–299.

Marfori, M. A. (2010). Informal proofs and mathematical rigour. *Studia Logica*, **96**, 261–272.

Priest, G. (1987). *In Contradiction: A Study of the Transconsistent*. Oxford: Oxford University Press.

Shapiro, S. (1985). Epistemic and intuitionistic arithmetic. In Shapiro, S., editor. *Intensional Mathematics*. Amsterdam: North Holland, pp. 11–46.

Tanswell, F. (2015). A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica*, **23**(3), 295–310.

CENTRE FOR LOGIC AND PHILOSOPHY OF SCIENCE
GHENT UNIVERSITY
GHENT, BELGIUM
E-mail: haptism89@gmail.com

CENTRE FOR LOGIC AND PHILOSOPHY OF SCIENCE
GHENT UNIVERSITY
GHENT, BELGIUM
and
INSTITUTE OF PHILOSOPHY, SOCIOLOGY AND JOURNALISM
UNIVERSITY OF GDANSK
GDANSK, POLAND
E-mail: rfl.urbaniak@gmail.com