Australasian Journal of Logic

Two-valued Logics for Transparent Truth Theory

Lucas Rosenblatt

University of Buenos Aires - Conicet

Abstract

It is part of the current wisdom that the Liar and similar semantic paradoxes can be taken care of by the use of certain non-classical multivalued logics. It is also well-known that some of these logics can be characterized by means of two-valued semantics. An immediate consequence of this is that there are two-valued logics that support a transparent truth predicate. In this paper I want to suggest that these logics are not just interesting from a formal point of view but also from a philosophical perspective. In particular, I will argue that the two-valued presentation of these logics has a number of advantages over the more usual presentations.

1 Introduction

It is part of current wisdom that the Liar and similar semantic paradoxes can be taken care of without compromising the transparency of the truth predicate[1] by the use of certain non-classical multivalued logics. This much was shown by Kripke in his classical paper [14] on truth, where he uses three-valued interpretations with $\{1, \frac{1}{2}, 0\}$ as the set of semantic values. His fixed-point construction starts from a classical interpretation for a first-order base language $\mathcal{L}$ without a truth predicate $Tr(x)$ and provides a way to generate an interpretation for the language $\mathcal{L}_{Tr}$ containing such a predicate. Although the fixed-point construction can be carried out with several valuation schemata, here I will only focus on Kleene's strong valuation schema. According to this schema, negation $\neg$ is defined as 1 minus the value of the negated formula

[1] By a transparent truth predicate I mean a predicate $Tr(x)$ such that for every formula ϕ, $Tr\ulcorner\phi\urcorner$ and ϕ are everywhere intersubstitutable (where $\ulcorner\phi\urcorner$ is a name for the sentence ϕ).

and disjunction $\vee$ is defined as the maximum of the values of the disjuncts. It is easy to define the other logical connectives in terms of these two[2]. Since I want to discuss semantic paradoxes, I assume, as usual, that $\mathcal{L}_{Tr}$ has some way to talk about itself. More specifically, for each formula ϕ of $\mathcal{L}_{Tr}$ there is a term $\ulcorner\phi\urcorner$ that is the name of that formula. Later on I will make this more precise. Kripke's insight is that we can construct different strong Kleene interpretations with the additional feature that the value assigned to any formula $Tr\ulcorner\phi\urcorner$ is the same as the value assigned to ϕ itself. That is, it is possible to provide theories based on strong Kleene interpretations where truth behaves as a transparent predicate.

The logics K_3 and LP are notorious examples of this (for an overview of these logics see [16]). The only important (non-philosophical) difference between K_3 and LP has to do with the consequence relation defined by each of these logics. In K_3 an argument is valid if it preserves the value 1, while in LP an argument is valid if it preserves the non-0 values. In other words, whereas K_3 takes 1 as the only designated value, LP takes both 1 and $\frac{1}{2}$. This difference has a major impact on the set of valid inferences and formulas. Crucially, both the Law of Excluded Middle and *Reductio ab Absurdum* fail in K_3 but not in LP[3]. And both Explosion and Disjunctive Syllogism fail in LP but not in K_3.

However, they have an important feature in common. Consider a Liar sentence λ saying of itself that it is false[4]. Both in K_3 and LP λ is categorized by means of the intermediate value $\frac{1}{2}$. This gives a consistent assignment to λ since the value of $\neg\phi$ is defined as 1 minus the value of ϕ for any formula ϕ. So we can let the values of $\lambda, Tr\ulcorner\neg\lambda\urcorner, \neg Tr\ulcorner\lambda\urcorner$ and $Tr\ulcorner\lambda\urcorner$ be all $\frac{1}{2}$.

Even though this idea works quite well, it is well-known that these logics can also be characterized by means of two-valued semantics. The four ways of doing this I know of are: Routley star semantics, relational semantics, partial semantics and bivaluation semantics (see [16] for the first two, [13] for the third and [7] for the fourth[5]). In this paper I will present a formal semantics based on a more recent approach developed by Arnon Avron and Iddo Lev in [2]. The idea, roughly, is to define negation by means of a non-deterministic matrix[6]. This will allows us to obtain a transparent truth

[2]I will ignore the quantifiers for now. For my purposes they bring extra complications without adding interesting insights.

[3]Moreover, given that $\frac{1}{2}$ is not designated and that for each formula there is an assignment that gives it $\frac{1}{2}$, K_3 has no tautologies at all.

[4]I won't concern myself at this point on how to get self-reference. If the reader prefers, she can take λ to be *equivalent* to the sentence asserting its own falsehood.

[5]Actually, this type of semantics has been known since [11], where a treatment of Da Costa's C_n-systems in terms of bivaluations is offered.

[6]Some of the points I am going to argue for below could have been made with the

predicate and, crucially, there will be no need to introduce a third semantic category[7]. As I will later show, the idea that the semantic paradoxes do not require the postulation of a third semantic category raises a number of interesting philosophical issues. The main contribution of this paper is to discuss some of those issues and to argue for the claim that the two-valued presentation of these logics has a number of advantages over the more usual presentations.

Its structure is as follows. The next section gives a brief overview of the idea of a non-deterministic matrix. Section 3 shows in which sense some of these non-deterministic matrices are compatible with a transparent truth predicate. More specifically, it is observed that by making negation non-deterministic we can obtain two-valued versions of the theories K_3 and LP. In section 4 it is shown how to give a Kripke-style definition of the truth predicate in these two-valued logics. In section 5 I consider a number of philosophical issues with the non-deterministic account of negation and I briefly sketch another way in which a non-deterministic account of the connectives might be helpful with semantic paradoxes. Section 6 contains some concluding remarks.

2 Non-deterministic matrices

Intuitively, in a non-deterministic framework there is at least one connective such that you cannot completely determine the value of a compound formula involving that connective even if you know the values of all the atomic formulas of the language. In other words, you need to make a non-deterministic choice between the values in a certain set. This idea can be spelled out rigorously (a more detailed account can be found in [1]):

Definition ($\mathcal{NDM}atrix$) A *non-deterministic matrix* for a language $\mathcal{L}$ is a tuple $\mathcal{M} = < \mathcal{V}, \mathcal{D}, \mathcal{O} >$, where:

- $\mathcal{V}$ is a non-empty set of values,
- $\mathcal{D}$ is a non-empty proper subset of $\mathcal{V}$, and
- $\mathcal{O}$ is a set of functions such that for every n-ary connective $\diamond$ in $\mathcal{L}$, there is a corresponding function $\diamond^{\mathcal{M}}$ in $\mathcal{O}$ such that $\diamond^{\mathcal{M}}$: $\mathcal{V}^n \longrightarrow 2^{\mathcal{V}} - \varnothing$.[8]

framework of bivaluations as well. But for reasons that will become clear in section 5, I prefer to use the non-deterministic account.

[7] I'll have more to say on the idea of a 'semantic category' in section 5. For now, I'll use the term somewhat informally.

[8] The reason for excluding the empty set is that it is not straightforward how to compute

Of course, $\mathcal{V}$ is meant to be a set of truth-values and $\mathcal{D}$ a set of designated values. The interesting part of the definition has to do with the set $\mathcal{O}$ of functions for the non-deterministic connectives. In a deterministic matrix, for each n-ary connective $\diamond$ in $\mathcal{L}$ there is a corresponding function $\diamond^{\mathcal{M}}$ such that $\diamond^{\mathcal{M}}$: $\mathcal{V}^n \longrightarrow \mathcal{V}$. The function takes a certain n-tuple of values in $\mathcal{V}^n$ and outputs a value in $\mathcal{V}$. In the case of non-deterministic connectives, the co-domain of the corresponding function is the set of sets of values $2^{\mathcal{V}} - \varnothing$, rather than the set of values $\mathcal{V}$.

Also notice that deterministic matrices are a special case of non-deterministic matrices. More specifically, for each n-ary connective $\diamond$ in a deterministic matrix $\mathcal{M}$ which is interpreted as a function $\diamond^{\mathcal{M}}$: $\mathcal{V}^n \longrightarrow \mathcal{V}$, we can build a non-deterministic matrix $\mathcal{M}'$ where that connective can be taken as a function that only outputs singleton values, that is, $\diamond^{\mathcal{M}'}$: $\mathcal{V}^n \longrightarrow \{\mathcal{A} \subseteq \mathcal{V}$: $|\mathcal{A}| = 1\}$. By doing this we obtain a non-deterministic matrix with connectives that mimic the behavior of the deterministic connectives.

It is straightforward to characterize the usual notions of *valuation*, *satisfaction*, *validity*, and so on, for non-deterministic matrices. For example, a valuation is defined in the following way:

Definition ($Valuation$) Let $Form_{\mathcal{L}}$ denote the set of formulae of the language $\mathcal{L}$. A *valuation* in $\mathcal{M}$ is a function I: $Form_{\mathcal{L}} \longrightarrow \mathcal{V}$ such that for each n-ary connective $\diamond$ of $\mathcal{L}$, the following holds for all $\phi_1, ..., \phi_n \in Form_{\mathcal{L}}$: $I(\diamond(\phi_1, ..., \phi_n)) \in \diamond^{\mathcal{M}}(I(\phi_1), ..., I(\phi_n))$

Notice that since $\diamond^{\mathcal{M}}(I(\phi_1), ..., I(\phi_n))$ gives a set of values rather than a single value, we use '$\in$' instead of '$=$' in the previous definition. With this new notion of valuation, the concepts of $Satisfaction$ and $Validity$ can be defined as usual.

3 Two-valued non-deterministic logics

It is clear from the definition of a non-deterministic matrix that there are many ways in which a matrix can be non-deterministic. For my purposes, it is enough to consider only two-valued matrices where every connective but negation is deterministic.

Let $\mathcal{L}$ be a propositional language with one unary connective $\neg$ and two binary connectives $\vee$ and $\wedge$. Let $\mathcal{M}_1 = < \mathcal{V}_1, \mathcal{D}_1, \mathcal{O}_1 >$, where:

- $\mathcal{V}_1 = \{1, 0\}$,

the value of compound formulae where at some step of the computation we have as input the empty set.

- $\mathcal{D}_1 = \{1\}$, and
- $\mathcal{O}_1 = \{\neg^{\mathcal{M}_1}, \vee^{\mathcal{M}_1}, \wedge^{\mathcal{M}_1}\}$ is defined in the following way:

	$\neg^{\mathcal{M}_1}$
1	{0}
0	{1,0}

		$\vee^{\mathcal{M}_1}$
1	1	{1}
1	0	{1}
0	1	{1}
0	0	{0}

		$\wedge^{\mathcal{M}_1}$
1	1	{1}
1	0	{0}
0	1	{0}
0	0	{0}

The matrix $\mathcal{M}_1$ characterizes a non-deterministic negation. In particular, it is compatible with the existence of valuations I such that for some formula ϕ, $I(\phi) = I(\neg\phi) = 0$. This matrix corresponds to the conditional-free fragment of the logic $CLaN$, developed in [5].

Of course, negation can be modified in yet another way. Let $\mathcal{L}$ be as before and let $\mathcal{M}_2$ be just as $\mathcal{M}_1$ except for the negation connective, which is now defined in the following way:

	$\neg^{\mathcal{M}_2}$
1	{1,0}
0	{1}

$\mathcal{M}_2$ also characterizes a non-deterministic negation. But this time there can be valuations I such that for some formula ϕ, $I(\phi) = I(\neg\phi) = 1$. As it is pointed out in [2], this matrix corresponds to the conditional-free fragment of the well-known logic $CLuN$, also developed in [5].

It is straightforward to check that $\mathcal{M}_1$ is a paracomplete logic and that $\mathcal{M}_2$ is a paraconsistent logic, given that $\not\models_{\mathcal{M}_1} \phi \vee \neg\phi$, and $\phi \wedge \neg\phi \not\models_{\mathcal{M}_2} \psi$. What I find interesting about these matrices is that they are compatible with a transparent truth predicate, even though they are two-valued. I have already mentioned that both the paracomplete three-valued logic K_3 and the paraconsistent three-valued logic LP can support a transparent truth predicate. But it is not hard to show that every K_3-countermodel can be turned into an $\mathcal{M}_1$-countermodel, and also that every LP-countermodel can be turned into an $\mathcal{M}_2$-countermodel[9].

The proofs of these facts are straightforward modifications of the proofs offered in [5] for $CLaN$ and $CLuN$. To prove the first fact, the idea is to replace all assignments of the value $\frac{1}{2}$ by 0, and leave everything else

[9]Just for the sake of completeness, there is a third two-valued matrix $\mathcal{M}_3$ with a non-deterministic negation which is also compatible with a transparent truth predicate. $\mathcal{M}_3$ is given by

untouched. For the second, we need to replace all assignments of the value $\frac{1}{2}$ by 1, and leave everything else untouched.

Fact 3.1 If $\Gamma \models_{\mathcal{M}_1} \Delta$, then $\Gamma \models_{K_3} \Delta$[10].

Proof sketch. If $\Gamma \not\models_{K_3} \Delta$, then there is a K_3-valuation $I^{\mathcal{A}}$ such that $I^{\mathcal{A}}$ assigns every γ in Γ the value 1 and assigns every δ in Δ either 0 or $\frac{1}{2}$. Now construct a valuation $I^{\mathcal{B}}$ which is exactly as $I^{\mathcal{A}}$ except that it assigns 0 whenever $I^{\mathcal{A}}$ assigns $\frac{1}{2}$. More specifically, $I^{\mathcal{B}}$ is such that for each formula ϕ:

- if $I^{\mathcal{A}}(\phi) = \frac{1}{2}$, then $I^{\mathcal{B}}(\phi) = 0$, and
- if $I^{\mathcal{A}}(\phi) \neq \frac{1}{2}$, then $I^{\mathcal{B}}(\phi) = I^{\mathcal{A}}(\phi)$.

Clearly, $I^{\mathcal{B}}$ is an $\mathcal{M}_1$-valuation and $I^{\mathcal{B}}$ assigns every γ in Γ the value 1 and every δ in Δ the value 0. So $\Gamma \not\models_{\mathcal{M}_1} \Delta$. □

This means that $\mathcal{M}_1$ is a sublogic of K_3. Hence, $\mathcal{M}_1$ is a paracomplete two-valued! consistent (non-deterministic) matrix that supports a transparent truth predicate.

Fact 3.2 If $\Gamma \models_{\mathcal{M}_2} \Delta$, then $\Gamma \models_{LP} \Delta$.

Proof sketch. Similar to the proof of Fact 3.1. □

This shows that $\mathcal{M}_2$ is a sublogic of LP. Therefore, $\mathcal{M}_2$ is a paraconsistent two-valued! non-trivial (non-deterministic) matrix that supports a transparent truth predicate.

It has been proved that there are soundness and completeness results for these matrices. Actually, this follows from a more general fact proved by Avron and Zamansky in [3]. In a multiple conclusion sequent calculus setting, the logic of $\mathcal{M}_1$ is sound and complete with respect to (propositional) classical logic *minus* the following rule:

	$\neg_{\mathcal{M}_3}$
1	{1,0}
0	{1,0}

For those interested, it turns out that $\mathcal{M}_3$ is the conditional-free fragment of the (paracomplete and paraconsistent) logic $CloN$ developed in [5] and that it is a subtheory of the four-valued logic of first degree entailment FDE. I should also note that there is a fourth logic in this family, sometimes called S_3. Accordingly, there is a fourth matrix, which I'll call $\mathcal{M}_4$, which can be obtained from $\mathcal{M}_3$ by admitting only those $\mathcal{M}_3$-valuations where negations behaves non-deterministically either for input 0 or for input 1, but not both. It turns out that $\mathcal{M}_4$ is a subtheory of the three-valued logic S_3. See [12], p.81 for more details on S_3.

[10]The reader can take single conclusions if she likes. At this point nothing important depends on this, except that the author likes multiple conclusions better.

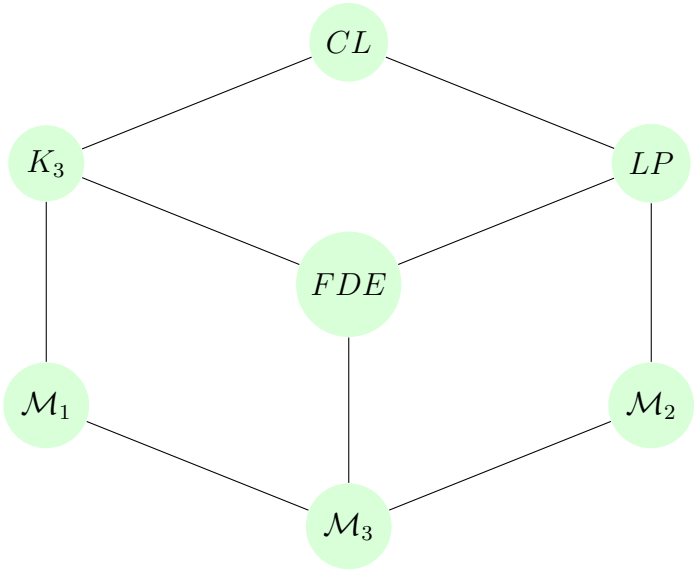
$$\text{Right}\neg\ \frac{\Gamma, \phi \vdash \Delta}{\Gamma \vdash \Delta, \neg\phi}$$

And the logic of $\mathcal{M}_2$ is sound and complete with respect to (propositional) classical logic[11] *minus* the rule:

$$\text{Left}\neg\ \frac{\Gamma \vdash \Delta, \phi}{\Gamma, \neg\phi \vdash \Delta}$$

This is no surprise, as each of these rules correspond to a row in the matrix for $\neg$. If we take Right$\neg$ away, $\neg$ "doesn't know" what to do with false formulas, while if we take Left$\neg$ away, $\neg$ "doesn't know" what to do with true formulas.

An obvious problem with these logics is that they are too weak. $\mathcal{M}_1$ is a *proper* sublogic of K_3 and that $\mathcal{M}_2$ is a *proper* sublogic of LP, as we can see below (CL stands for 'classical logic'):



In fact, these logics are much weaker than their three-valued cousins. For example, we cannot define conjunction (disjunction) in terms of disjunction (conjunction) and negation. Moreover, there is no interaction at all[12] between disjunction and conjunction since *all* the de Morgan Laws fail in these logics.

[11]Since one of the negation rules will not be available, in both cases we need to make a minor adjustment: $\phi \vdash \phi$ is an initial sequent *for all formulas*, and not only for all atomic formulas.

[12]Except for some negation-free inferences like the one from $\phi \wedge \psi$ to $\phi \vee \psi$ and the one from $\phi \wedge (\psi \vee \chi)$ to $(\phi \wedge \psi) \vee (\phi \wedge \chi)$.

There is a well-known way to fix this (see again [5]). Here we show, once again, that thanks to the aid of non-deterministic matrices, this maneuver does not require the postulation of a third semantic category. Let $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ be as the matrices $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, plus the following extra requirements:

- For each formula ϕ and every valuation I:
  1. $I(\phi) = I(\neg\neg\phi)$.
  2. $I(\neg(\phi \wedge \psi)) = I(\neg\phi \vee \neg\psi)$.
  3. $I(\neg(\phi \vee \psi)) = I(\neg\phi \wedge \neg\psi)$.[13]

The matrix $\mathcal{M}_1^+$ corresponds to the conditional-free fragment of the logic $CLaNs$ and the matrix $\mathcal{M}_2^+$ corresponds to the conditional-free fragment of the logic $CLuNs$, both also developed in [5]. It is possible to prove that $\mathcal{M}_1^+$ and K_3 characterize the same set of valid inferences and that $\mathcal{M}_2^+$ and LP also characterize the same set of valid inferences.

Fact 3.3 $\Gamma \models_{\mathcal{M}_1^+} \Delta$ if and only if $\Gamma \models_{K_3} \Delta$.

Proof sketch. The proof of the left-to-right direction is similar to the proof of Fact 3.1. For the other direction, assume that $\Gamma \not\models_{\mathcal{M}_1^+} \Delta$. Then there is an $\mathcal{M}_1^+$-valuation $I^{\mathcal{A}}$ such that $I^{\mathcal{A}}$ assigns every γ in Γ the value 1 and assigns every δ in Δ the value 0. Now construct a valuation $I^{\mathcal{B}}$ which is exactly as $I^{\mathcal{A}}$ except that it assigns $\frac{1}{2}$ to a formula ϕ whenever $I^{\mathcal{A}}$ assigns 0 to both ϕ and $\neg\phi$. More specifically, $I^{\mathcal{B}}$ is such that for any *atomic* formula ϕ:

- $I^{\mathcal{B}}(\phi) = \frac{1}{2}$ whenever $I^{\mathcal{A}}(\phi) = I^{\mathcal{A}}(\neg\phi) = 0$, and
- $I^{\mathcal{B}}(\phi) = I^{\mathcal{A}}(\phi)$ otherwise.

It is not hard to see that $I^{\mathcal{B}}$ is a K_3-valuation and that for every formula ϕ it holds that if $I^{\mathcal{A}}(\phi) = 1$, then $I^{\mathcal{B}}(\phi) = 1$, and that if $I^{\mathcal{A}}(\phi) = 0$, then either $I^{\mathcal{B}}(\phi) = 0$ or $I^{\mathcal{B}}(\phi) = \frac{1}{2}$. It follows that for every $\gamma \in \Gamma$, $I^{\mathcal{B}}(\gamma) = 1$ and for every $\delta \in \Delta$, $I^{\mathcal{B}}(\delta) = 0$ or $I^{\mathcal{B}}(\delta) = \frac{1}{2}$. This means that $\Gamma \not\models_{K_3} \Delta$. □

As usual, an analogous result can be proved for the dual LP.

[13] If quantifiers were also available, we would need to stipulate in addition that for each formula ϕ and every valuation I:

- $I(\exists x\neg\phi) = I(\neg\forall x\phi)$.
- $I(\forall x\neg\phi) = I(\neg\exists x\phi)$.

Fact 3.4 $\Gamma \models_{\mathcal{M}_2^+} \Delta$ if and only if $\Gamma \models_{LP} \Delta$.

Proof sketch. For the right-to-left direction, the only relevant difference with respect to the previous proof is that we let $I^{\mathcal{B}}$ be exactly as $I^{\mathcal{A}}$ except that for each atomic formula ϕ, $I^{\mathcal{B}}(\phi) = \frac{1}{2}$ whenever $I^{\mathcal{A}}(\phi) = I^{\mathcal{A}}(\neg\phi) = 1$, and $I^{\mathcal{B}}(\phi) = I^{\mathcal{A}}(\phi)$ otherwise. □

These results show that $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ are two-valued versions of K_3 and LP, respectively. The third value need not be there if negation is characterized by a non-deterministic matrix.

These matrices also enjoy soundness and completeness results. However, the usual proof-theoretic presentations of the logics K_3 and LP use three-sided sequents, which are fairly natural when the matrices are three-valued. However, for our two-valued logics $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ a different presentation seems more appropriate. Again using sequents, it can be proved that the $\mathcal{M}_1^+$-matrices ($\mathcal{M}_2^+$-matrices) are sound and complete with respect to the calculus for $\mathcal{M}_1$ ($\mathcal{M}_2$) plus the following rules (see [8] for a more detailed presentation):

$$\mathrm{L}\neg\neg\ \frac{\Gamma, \phi \vdash \Delta}{\Gamma, \neg\neg\phi \vdash \Delta}$$

$$\mathrm{R}\neg\neg\ \frac{\Gamma \vdash \phi, \Delta}{\Gamma \vdash \neg\neg\phi, \Delta}$$

$$\mathrm{L}\neg\wedge\ \frac{\Gamma, \neg\phi \vee \neg\psi \vdash \Delta}{\Gamma, \neg(\phi \wedge \psi) \vdash \Delta}$$

$$\mathrm{R}\neg\wedge\ \frac{\Gamma \vdash \neg\phi \vee \neg\psi, \Delta}{\Gamma \vdash \neg(\phi \wedge \psi), \Delta}$$

$$\mathrm{L}\neg\vee\ \frac{\Gamma, \neg\phi \wedge \neg\psi \vdash \Delta}{\Gamma, \neg(\phi \vee \psi) \vdash \Delta}$$

$$\mathrm{R}\neg\vee\ \frac{\Gamma \vdash \neg\phi \wedge \neg\psi, \Delta}{\Gamma \vdash \neg(\phi \vee \psi), \Delta}$$

What happens to λ and other problematic sentences in these logics? While in three-valued logics these sentences are dealt with by the introduction of a new semantic category, the logics presented in the previous section require no such thing. In particular, $I^{\mathcal{M}_1^+}(\lambda) = 0$ and $I^{\mathcal{M}_2^+}(\lambda) = 1$. These assignments are unproblematic. On the one hand, since in $\mathcal{M}_1^+$ we can have $I^{\mathcal{M}_1^+}(\phi) = I^{\mathcal{M}_1^+}(\neg\phi) = 0$ for some formulas ϕ, the following holds if a truth

predicate is available:
$I^{\mathcal{M}_1^+}(\lambda) = I^{\mathcal{M}_1^+}(Tr^{\ulcorner}\lambda^{\urcorner}) = I^{\mathcal{M}_1^+}(\neg Tr^{\ulcorner}\lambda^{\urcorner}) = 0$.
Analogously, since in $\mathcal{M}_2^+$ we can have $I^{\mathcal{M}_2^+}(\phi) = I^{\mathcal{M}_2^+}(\neg\phi) = 1$ for some formulas ϕ, the following holds:
$I^{\mathcal{M}_2^+}(\lambda) = I^{\mathcal{M}_2^+}(Tr^{\ulcorner}\lambda^{\urcorner}) = I^{\mathcal{M}_2^+}(\neg Tr^{\ulcorner}\lambda^{\urcorner}) = 1$.
So, once a non-deterministic negation is present, there is no need to have a third truth-value to make truth a transparent predicate.

4 A definition of truth

This section is devoted to show that we can provide a two-valued fixed-point semantics for a language containing a truth predicate. Although given the results of the previous section, this fact should be immediate, it is still instructive to see how exactly the construction can be carried out. This will help us understand how the truth predicate can be appropriately interpreted in a two-valued setting and, specially, how paradoxical sentences can be handled.

Here is a sketch of how the construction works for $\mathcal{M}_1^+$. Let $\mathcal{L}$ be the language of Peano arithmetic and let $Tr(x)$ be $\mathcal{L}$ plus a truth predicate $Tr(x)$. In what follows we will only consider interpretations I^δ (where δ is an ordinal number) for $\mathcal{L}_{Tr}$ where the arithmetic literals (i.e. atomic formulas and their negations) have their usual truth-value and the domain is exactly ω[14]. Let $^{\ulcorner}\phi^{\urcorner}$ denote the (Gödel code of the) formula ϕ and let $I^\delta(Tr^+)$ and $I^\delta(Tr^-)$ be the extension and the antiextension of the predicate $Tr(x)$ at I^δ, respectively.

We need to impose an additional condition on the interpretations:

- For all formulas ϕ and all interpretations I^δ, it holds that $I^\delta(\neg Tr(^{\ulcorner}\phi^{\urcorner})) = I^\delta(Tr(^{\ulcorner}\dot{\neg}\phi^{\urcorner}))$[15], where, as usual, $\dot{\neg}$ is a function that outputs (the code of) the formula $\neg\phi$ if it is given (the code of) the formula ϕ as input.

[14]Observe that because negation is non-deterministic, it is not enough to stipulate that the arithmetic part of the vocabulary has its intended meaning in all interpretations, we also need to fix the values of the negations of the atomic formulas at every interpretation. Otherwise there will be interpretations where atomic arithmetic sentences will have the same value as their negation.

[15]Notice that in this respect, the truth predicate is no different from the connectives. In fact, if we were to provide a sequent calculus for the theory extended with the truth predicate, in addition to the usual rules for formulas of the form $Tr^{\ulcorner}\phi^{\urcorner}$, we would need left and right rules for formulas of the form $\neg Tr^{\ulcorner}\phi^{\urcorner}$. More specifically, we would need:

$$\mathrm{L}\neg Tr\ \frac{\Gamma, Tr\langle\neg\phi\rangle \vdash \Delta}{\Gamma, \neg Tr\langle\phi\rangle \vdash \Delta} \qquad \mathrm{R}\neg Tr\ \frac{\Gamma \vdash Tr\langle\neg\phi\rangle, \Delta}{\Gamma \vdash \neg Tr\langle\phi\rangle, \Delta}$$

As usual we define a *jump* operator $\mathcal{J} : \mathcal{P}\omega \to \mathcal{P}\omega$ on interpretations for $\mathcal{L}_{Tr}$. This operator is defined as follows:

$$\mathcal{J}(I^\delta(Tr^+)) = \{\ulcorner\phi\urcorner : I^\delta(\phi) = 1\}.$$

So $\mathcal{J}$ yields, if applied to a set which is the extension of the truth predicate, another set which is the new extension of the truth predicate. Since the matrix is two-valued, we can define the antiextension of the truth predicate at any interpretation I^δ as follows:

$$I^\delta(Tr^-) = \omega - I^\delta(Tr^+).$$

We will say that the operator $\mathcal{J}$ is *monotone* if and only if $I^\alpha(Tr^+) \subseteq I^\beta(Tr^+)$ implies $\mathcal{J}(I^\alpha(Tr^+)) \subseteq \mathcal{J}(I^\beta(Tr^+))$. The monotonicity of $\mathcal{J}$ entails the existence of fixed points for $\mathcal{J}$, i.e., interpretations $I^\delta(Tr^+)$ such that $\mathcal{J}(I^\delta(Tr^+)) = I^\delta(Tr^+)$.

Intuitively, the idea is that we start with an interpretation I^0 that assigns a consistent (perhaps empty) extension to the truth predicate. Since negation is non-deterministic, the construction is such that some formulas and their negations will be in the antiextension of the truth predicate. Although some true formulas might still be in the antiextension of the truth predicate at I^0, the jump operator successively fixes this situation by including more and more Gödel codes of true formulas in the extension of the truth predicate in a monotonic way. This means that the sets
$I^0(Tr^+), \mathcal{J}(I^0(Tr^+)), \mathcal{J}(\mathcal{J}(I^0(Tr^+))), ...$
form an increasing sequence[16], whereas the sets
$I^0(Tr^-), \mathcal{J}(I^0(Tr^-)), \mathcal{J}(\mathcal{J}(I^0(Tr^-))), ...$
form a decreasing sequence.

To show how this works properly, we first need the following lemma:

Lemma 4.1. *If $I^\alpha(Tr^+) \subseteq I^\beta(Tr^+)$, then for every formula ϕ of $\mathcal{L}_{Tr}$, if $I^\alpha(\phi) = 1$, then $I^\beta(\phi) = 1$, and if $I^\beta(\phi) = 0$, then $I^\alpha(\phi) = 0$.*

Proof. Since there are just two truth-values, we only need to establish the first claim, from which the second one follows. The claim is proven by induction on the complexity of ϕ, but we have to consider each kind of positive formula

[16]More formally, we obtain an increasing sequence $I^0(Tr^+)$, $I^1(Tr^+)$, $I^2(Tr^+)$,...,$I^\omega(Tr^+)$, $I^{\omega+1}(Tr^+)$,.... in this way:

- For successor ordinals $\alpha + 1$, $I^{\alpha+1}(Tr^+) = \mathcal{J}(I^\alpha(Tr^+)) = \{\ulcorner\phi\urcorner : I^\alpha(\phi) = 1\}$, and
- For limit ordinals Λ, $I^\Lambda(Tr^+) = \bigcup_{\beta<\Lambda} I^\beta(Tr^+)$.

and each kind of negated formula. Assume that $I^\alpha(Tr^+) \subseteq I^\beta(Tr^+)$. Then we have the following cases:

- *ϕ is an atomic arithmetical formula or the negation of an atomic arithmetical formula.* Since the truth of purely arithmetical formulas is not affected by what is assigned to $Tr(x)$, for all α and β, $I^\alpha(\phi) = I^\beta(\phi)$.

- *ϕ is of the form $Tr(\ulcorner\psi\urcorner)$.* Assume that $I^\alpha(Tr\ulcorner\psi\urcorner) = 1$. Then $\ulcorner\psi\urcorner \in I^\alpha(Tr^+)$. Since $I^\alpha(Tr^+) \subseteq I^\beta(Tr^+)$, $\ulcorner\psi\urcorner \in I^\beta(Tr^+)$. Hence, $I^\beta(Tr\ulcorner\psi\urcorner) = 1$.

- *ϕ is of the form $\neg Tr(\ulcorner\psi\urcorner)$.* Here we need to use the fact that for each interpretation I^δ, $I^\delta(\neg Tr(\ulcorner\psi\urcorner)) = I^\delta(Tr(\ulcorner\dot{\neg}\psi\urcorner))$. Assume that $I^\alpha(\neg Tr\ulcorner\psi\urcorner) = 1$. Then $I^\alpha(Tr\ulcorner\dot{\neg}\psi\urcorner) = 1$, and so $\ulcorner\dot{\neg}\psi\urcorner \in I^\alpha(Tr^+)$. Since $I^\alpha(Tr^+) \subseteq I^\beta(Tr^+)$, we can infer that $\ulcorner\dot{\neg}\psi\urcorner \in I^\beta(Tr^+)$, which in turn gives $I^\beta(Tr(\ulcorner\dot{\neg}\psi\urcorner)) = 1$ and then $I^\beta(\neg Tr\ulcorner\psi\urcorner) = 1$.

- *ϕ is of the form $\neg\neg\psi$.* We use the fact that for each interpretation I^δ, δ, $I^\delta(\neg\neg\psi) = I^\delta(\psi)$. Assume that $I^\alpha(\neg\neg\psi) = 1$. Hence, $I^\alpha(\psi) = 1$. By the inductive hypothesis, $I^\beta(\psi) = 1$, and so $I^\beta(\neg\neg\psi) = 1$.

- *ϕ is of the form $\psi \wedge \chi$.* Straightforward.

- *ϕ is of the form $\neg(\psi \wedge \chi)$.* We use the fact that for each interpretation I^δ, $I^\delta(\neg(\psi \wedge \chi)) = I^\delta(\neg\psi \vee \neg\chi)$. Assume that $I^\alpha(\neg(\psi \wedge \chi)) = 1$. It follows that $I^\alpha(\neg\psi \vee \neg\chi) = 1$, which means that $I^\alpha(\neg\psi) = 1$ or $I^\alpha(\neg\chi) = 1$. By the inductive hypothesis we can infer that $I^\beta(\neg\psi) = 1$ or $I^\beta(\neg\chi) = 1$. So it follows that $I^\beta(\neg\psi \vee \neg\chi) = 1$, and therefore $I^\beta(\neg(\psi \wedge \chi)) = 1$.

- *ϕ is of the form $\psi \vee \chi$.* Straightforward.

- *ϕ is of the form $\neg(\psi \vee \chi)$.* We use the fact that for each interpretation I^δ, $I^\delta(\neg(\psi \vee \chi)) = I^\delta(\neg\psi \wedge \neg\chi)$. Assume that $I^\alpha(\neg(\psi \vee \chi)) = 1$. It follows that $I^\alpha(\neg\psi \wedge \neg\chi) = 1$, which means that $I^\alpha(\neg\psi) = 1$ and $I^\alpha(\neg\chi) = 1$. By the inductive hypothesis we can infer that $I^\beta(\neg\psi) = 1$ and $I^\beta(\neg\chi) = 1$. So it follows that $I^\beta(\neg\psi \wedge \neg\chi) = 1$, and therefore $I^\beta(\neg(\psi \vee \chi)) = 1$.

This completes the proof[17]. □

[17]Well, not quite. We still need to consider the quantifiers. For formulas of the form $\neg\exists x\phi$, we use its equivalence with $\forall x\neg\phi$, and for formulas of the form $\neg\forall x\phi$, we use its equivalence with $\exists x\neg\phi$. Formulas of the form $\exists x\phi$ and $\forall x\phi$ present no additional complications.

As a corollary we can obtain the following:

Lemma 4.2 (*Monotonicity of $\mathcal{J}$*)**.** *The sequence enjoys the following monotonicity property: If $I^{\alpha}(Tr^{+}) \subseteq I^{\beta}(Tr^{+})$, then $\mathcal{J}(I^{\alpha}(Tr^{+})) \subseteq \mathcal{J}(I^{\beta}(Tr^{+}))$ and $\mathcal{J}(I^{\beta}(Tr^{-})) \subseteq \mathcal{J}(I^{\alpha}(Tr^{-}))$.*

Proof. As before, it is enough to prove only one of the claims. So assume that $\ulcorner\phi\urcorner \in \mathcal{J}(I^{\alpha}(Tr^{+}))$. This means that $I^{\alpha}(\phi) = 1$. By Lemma 4.1, we can infer that $I^{\beta}(\phi) = 1$. Therefore, $\ulcorner\phi\urcorner \in \mathcal{J}(I^{\beta}(Tr^{+}))$. □

Because of the usual cardinality considerations, at some point the antiextension of the truth predicate stops decreasing and its extension stops increasing. And so we reach a fixed point.

Theorem 4.3 (*Existence of fixed point for $\mathcal{J}$*)**.** *The construction has the fixed point property. That is, there are interpretations I^{δ} such that:*

- $\mathcal{J}(I^{\delta}(Tr^{+})) = I^{\delta}(Tr^{+})$, *and*
- $\mathcal{J}(I^{\delta}(Tr^{-})) = I^{\delta}(Tr^{-})$.

Such interpretations deal with $Tr(x)$ in the desired way. If I^{δ} is a fixed point for $\mathcal{J}$, then for all sentences ϕ of $\mathcal{L}_{Tr}$ it holds that:

$I^{\delta}(\phi) = I^{\delta}(Tr\ulcorner\phi\urcorner)$.

What about $\mathcal{M}_2^{+}$? For $\mathcal{M}_2^{+}$ the idea is similar, but instead of assigning a consistent extension to the truth predicate at I^{0}, we assign it a consistent antiextension at I^{0} and this time the jump operator includes more and more Gödel codes of false formulas in the antiextension of the truth predicate in a monotonic way[18].

5 Making sense of two-valued non-deterministic logics

Although technically attractive, it might be argued that these logics have no real philosophical interest. In this section I'll show why I disagree. The absence of a third truth-value gives us a number of very desirable features. Firstly, there is a nice symmetry between truth and falsity. A valid argument is both truth-preserving from premises to conclusions and falsity-preserving

[18]We can deal with natural numbers that do not code formulas in both $\mathcal{M}_1^{+}$ and $\mathcal{M}_2^{+}$ by putting them in the antiextension of the truth predicate.

from conclusions to premises. I consider this to be an advantage over three-valued approaches like K_3 and LP, in which either falsity-preservation fails or truth-preservation fails[19]. In [21], discussing single-premiss, single-conclusion arguments, Alan Weir claims that

> (...) I take it as constitutive of the notion of logical consequence that if the premiss is true the conclusion is true and if the conclusion is false the premiss is false. The second, upwards falsity-preservation direction is often omitted, probably because in classical bivalent semantics it follows from the first, but I see no reason at all for an asymmetrical treatment of downwards truth-preservation and upwards falsity-preservation.

Although I wouldn't go as far as saying that the symmetry between truth and falsity is constitutive of the notion of logical consequence, I do think that the asymmetry is an odd feature that should be avoided whenever possible. As he goes on to point out, in a multiple conclusion setting, this amounts to the requirement that if all the premises are true, then at least one of the conclusions has to be true *and* if all the conclusions are false, then at least one of the premises has to be false. Since bivalence holds, this requirement is fully satisfied in $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$, even with the transparent truth predicate around.

Secondly, a defect usually attributed to three-valued approaches is that they cannot appropriately specify the semantic status of those sentences that obtain the third truth-value. In K_3 it is not possible to truly express that the Liar sentence is neither true nor false, while in LP the sentence expressing that the Liar is both true and false is not only true, but also false. In $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ there is no such problem. In $\mathcal{M}_1^+$ the sentence saying that the Liar is true is in fact false, and in $\mathcal{M}_2^+$ the sentence saying that the Liar is true is in fact true. It might be argued that there is still a problem because in $\mathcal{M}_1^+$ the sentence saying that the Liar is false is itself false, and in $\mathcal{M}_2^+$ the sentence saying that the Liar is false is itself true. However, the sentence saying that the Liar is false is just the Liar sentence, so to demand that $\mathcal{M}_1^+$ ($\mathcal{M}_2^+$) should categorize this sentence as true (false) is just the same as to require that it categorizes the Liar itself as true (false).

[19]Two remarks are in order. First, in the three-valued paracomplete and paraconsistent logic S_3 we do have both truth-preservation and falsity-preservation. But the problem is that the resulting consequence relation is extremely weak. Second, there is a sense in which there is truth- and falsity-preservation in LP, given that the value $\frac{1}{2}$ is to be interpreted as both-true-and-false. However, this is not enough to address the symmetry issue: although valid LP-arguments preserve *strict* falsity, they might not preserve strict truth.

A related issue is whether there is a strengthened version of the Liar affecting these theories. In many-valued theories, such as K_3, there cannot be a predicate expressing the concept of not being true. And similarly, in theories like LP, there cannot be a predicate expressing the concept of strict falsehood. If there were such predicates, there would be strengthened forms of the Liar, which would make these theories trivial. However, this problem is simply dissolved in $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$, where being not true is the same as being strictly false. This is not to say that these theories are semantically closed, in the sense that they can express every intelligible semantic concept. For in $\mathcal{M}_1^+$ there cannot be an exhaustive negation, and in $\mathcal{M}_2^+$ there cannot be an exclusive negation. But nevertheless, there is no strengthened Liar different from the original Liar. To put it in a slogan: in these theories all Liars are the same[20].

Finally, a methodological advantage of presenting K_3, LP and similar logics by means of two-valued matrices (or two-sided sequent calculi) is that we can see in a very precise way what the difference is between these logics and classical logic. And the difference comes to this: negation behaves differently. It behaves non-deterministically. In some cases, it leaves open what the semantic value of the negated formula is. However, in the three-valued (-sided) matrix (proof-theoretic) presentation, the difference is cashed out in terms of the definition of validity. There is nothing particularly special about negation[21].

Let us move on to the potential problems these logics face. First, the nice thing about logics such as K_3 and LP is that they give us a conceptual story as to why the Liar and similar sentences are special. In the first, we reason in the following way. If we assume that the Liar is true, we can infer a contradiction, and if we assume that it is false we can infer a contradiction again. So it must be neither true nor false and hence the theory assigns it the value $\frac{1}{2}$. In the second, we reason in the following way. If we assume that the Liar is true, we can infer that it is false, and if we assume that it is false we can infer that it is true. Hence, it must be both true and false and so the theory assigns it the value $\frac{1}{2}$.

[20]One related problem that cannot be dissolved is posed by sentences such as 'there is a sentence that is neither true nor false' and 'there are no sentences that are true and false'. The first will be false according to $\mathcal{M}_1^+$ and the second will be true according to $\mathcal{M}_2^+$, which seems quite unpleasant. However, in this regard $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ are no better or worse than their three-valued versions, which evaluate the first sentence as neither true nor false and the second as both true and false, respectively.

[21]As an anonymous referee correctly pointed out, this is only an advantage with respect to the three-valued presentation of K_3, LP and similar logics, so this criticism does not affect the relational semantics nor the Routley star semantics, where the crucial change has to do with the behavior of negation as well.

What about $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$? In the first one the Liar receives the value 0, while in the second one it receives the value 1. Is there something interesting to say about why this is so? I think there is. A plausible and very straightforward interpretation of these logics is in terms of acceptance and rejection. In particular, we can make sense of $\models_{\mathcal{M}_1^+}$ and $\models_{\mathcal{M}_2^+}$ along the lines of [17]. $\Gamma \models_{\mathcal{M}_i^+} \Delta$ (for $i = 1, 2$) amounts to the claim that we ought not to accept each member of Γ and reject each member of Δ. For sentences, we can say that if a sentence has the value 1, then that means that we ought to accept it, and if a sentence has the value 0, that means that we ought not accept it or, what amount to the same thing here, that we ought to reject it. So, in the case of $\mathcal{M}_1^+$ we accept neither the Liar sentence nor its negation, and in the case of $\mathcal{M}_2^+$ we accept them both. I think that, in this sense, $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ are "more honest" than their three-valued counterparts, in which bivalence is tacitly reinstated in terms of the dichotomy between being designated and being undesignated.

A second issue is that it is unclear to what extent $\neg^{\mathcal{M}_1^+}$ and $\neg^{\mathcal{M}_2^+}$ represent "real" negations. A nice feature of most multivalued logics is that for formulas that receive a classical value, the negation symbol behaves just as classical negation. In $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ no such thing happens.

I think it is quite natural to understand non-deterministic connectives as *ambiguous* expressions. In the case at hand, the non-deterministic character of negation reflects the fact that negation is used ambiguously in philosophical theorizing about truth. In particular, negation behaves in one way in contexts where paradoxical sentences are involved, and it behaves in a different way in paradox-free situations. For example, a paracomplete (paraconsistent) theorist claims that negation behaves as a non-exhaustive (non-exclusive) operator when it occurs in a Liar sentence, but that it behaves like boolean negation when it occurs in non-paradoxical sentences. In other words, negation is ambiguous between two readings, an exhaustive (exclusive) reading and a non-exhaustive (non-exclusive) reading. On one of those readings, negation is always is able to form contradictories, on the other sometimes it is not. One nice feature of the non-deterministic framework is that it captures this ambiguity in a very neat way.

Moreover, we only need to worry about the ambiguity to a certain extent. For one thing, the non-truth-functionality of negation only affects negations of atomic formulas. Once the truth-values of all the literals are determined, we can calculate the values of the rest of the formulas compositionally. But more importantly, once we recognize that we are reasoning in a complete (consistent) context, negation behaves just like boolean negation. More precisely, we can "recapture" classical reasoning in certain contexts, just as in K_3

and LP (see [9]). In $\mathcal{M}_1^+$ classical reasoning holds exactly when the atomic formulas in the conclusions respect excluded middle, and in $\mathcal{M}_2^+$ classical reasoning holds exactly when the atomic formulas in the premises are contradictory. More rigorously, we can prove the following claim (where $\models_{CL}$ is the classical entailment relation, $\{\gamma_1^{At}, ..., \gamma_m^{At}\}$ is the set of atomic formulas occurring in Γ and $\{\delta_1^{At}, ..., \delta_n^{At}\}$ is the set of atomic formulas occurring in Δ):

Fact 5.1 $\Gamma \models_{CL} \Delta$ if and only if $\delta_1^{At} \vee \neg\delta_1^{At}, ..., \delta_n^{At} \vee \neg\delta_n^{At}, \Gamma \models_{\mathcal{M}_1^+} \Delta$.

Proof sketch. The right-to-left direction is straightforward. For the other direction, assume that $\delta_1^{At} \vee \neg\delta_1^{At}, ..., \delta_n^{At} \vee \neg\delta_n^{At}, \Gamma \not\models_{\mathcal{M}_1^+} \Delta$. This means that there is an $\mathcal{M}_1^+$-valuation $I^{\mathcal{A}}$ such that $I^{\mathcal{A}}$ assigns every γ in Γ the value 1, every δ in Δ the value 0 and $\delta_i^{At} \vee \neg\delta_i^{At}$ the value 1 for every i such that $1 \leqslant i \leqslant n$. Now construct a valuation $I^{\mathcal{B}}$ which is exactly as $I^{\mathcal{A}}$ except that it assigns 1 to a formula $\neg\phi$ whenever $I^{\mathcal{A}}$ assigns 0 to ϕ. More specifically, $I^{\mathcal{B}}$ is such that for each atomic formula ϕ:

- $I^{\mathcal{B}}(\phi) = I^{\mathcal{A}}(\phi)$, and
- $I^{\mathcal{B}}(\neg\phi) = 1$ whenever $I^{\mathcal{A}}(\phi) = 0$.

It is not hard to see that $I^{\mathcal{B}}$ is a classical valuation and that for every $\gamma \in \Gamma$, $I^{\mathcal{B}}(\gamma) = 1$ and for every $\delta \in \Delta, I^{\mathcal{B}}(\delta) = 0$. This means that $\Gamma \not\models_{CL} \Delta$. □

As usual, there is an analogous "recovering" result for $\mathcal{M}_2^+$ (where $\models_{CL}, \Gamma$ and Δ are as before):

Fact 5.2 $\Gamma \models_{CL} \Delta$ if and only if $\Gamma \models_{\mathcal{M}_2^+} \Delta, \gamma_1^{At} \wedge \neg\gamma_1^{At}, ..., \gamma_m^{At} \wedge \neg\gamma_m^{At}$.[22]

Proof sketch. Similar to the proof of Fact 5.1. □

The first fact amounts, roughly, to the idea that in situations where the Law of Excluded Middle holds, that is, in complete situations, $\mathcal{M}_1^+$ collapses with classical logic. The second fact shows, again roughly, that in situations where the Law of Non-Contradiction holds, that is, in consistent situations,

[22] In [6], Batens proved the following result (I've modified his notation slightly):

$\vdash_{CL} \phi$ if and only if $\vdash_{PIz} (\psi_1 \wedge \neg\psi_1) \vee ... \vee (\psi_n \wedge \neg\psi_n) \vee \phi$, for some $\psi_1, ..., \psi_n$

where PIz is any paraconsistent extension of $CLuN$ (which is actually PI in the notation of [6]). If we take into account of the existence of disjunctive normal form in $CLuNs$, then we can take $\psi_1, ..., \psi_n$ to be atoms and thus obtain Fact 5.2 for $CLuNs$ (and hence its conditional-free fragment, $\mathcal{M}_2^+$). I'm grateful to an anonymous referee for this finding.

$\mathcal{M}_2^+$ collapses with classical logic. These facts should be sufficient to dispel the suspicion that negation behaves oddly even in non-paradoxical situations[23].

Of course, as an anonymous referee observed, the fan of truth-functionality might still be unhappy. But it is important to notice that the results above show that once we find out whether the sentence being negated is paradoxical or not, negation does behave truth-functionally. As I already pointed out, in a paradox-free context, negation is just boolean negation. But if we are reasoning with a paradoxical sentence, negation behaves just like the null operator, leaving the value of the sentence being negated intact. In the case of the Liar paradox and similar sentences, this is well motivated: the Liar can be identified, roughly, with its own negation, so no change of truth-value should be expected. So, once we find out whether the sentence being negated is paradoxical or not, truth-functionality is restored.

A third worry is that there are other ways to provide two-valued formal semantics for K_3, LP and similar logics. As we mentioned in the introduction, Routley star semantics, relational semantics, partial semantics and bivaluation semantics have been given for these logics. However, in the first three frameworks there are more than two *semantic categories*. By a semantic category I mean, roughly, a way of semantically evaluating a sentence. Notice that this may or may not coincide with the truth-values that we can assign to a sentence in a given model-theoretic framework. For instance, although partial semantics for K_3 uses only two truth-values, it introduces a third semantic category, that of lacking a truth-value. For another example consider the relational semantics for LP, which also uses only two truth-values but assigns them both to certain sentences. Once again, those sentences having both truth-values belong to a third semantic category. The case of Routley star semantics is harder to analyze, since the presence of possible worlds obscures things a bit. The present framework, on the other hand, only requires the postulation of two semantic categories.

The issue with bivaluations is different. As with the two-valued nondeterministic matrices, bivaluations only require the postulation of two semantic categories. In fact, it is known that any consequence relation that is reflexive, monotonic, and obeys cut (i.e. any *Tarskian* consequence relation) is exactly the consequence relation determined by some set of bivaluations. Now, any matrix logic of the sort that I have been considering obeys these principles. Because of this, its consequence relation has a bivaluational model

[23]There are other ways to recapture classical reasoning within these logics. A popular strategy is to introduce a determinateness operator in the case of paracomplete theories, and a consistency operator, in the case of paraconsistent theories. However, the introduction of these operators is by no means trivial in the presence of paradoxical sentences.

theory[24]. So, from a technical standpoint, I could have used bivaluations instead of non-deterministic matrices.

The reason why I prefer to use the latter framework and not the former is that, from a more conceptual perspective, defining negation by means of a non-deterministic matrix brings out one key aspect of negation that I've mentioned before: negation behaves ambiguously. While non-deterministic matrices capture this ambiguity very explicitly, this fact is somewhat concealed in the framework of bivaluations.

Another worry -one that has been considered crucial- is that $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ are still too weak. For instance, $\mathcal{M}_1^+$ *cannot* have a classical conditional. For suppose the conditional is defined in the following way:

		$\supset^{\mathcal{M}_1^+}$
1	1	$\{1\}$
1	0	$\{0\}$
0	1	$\{1\}$
0	0	$\{1\}$

Then *classical* negation $\neg_C$ becomes definable:

$$\neg_C\phi =_{def} \phi \supset \neg\phi$$

Something similar holds for $\mathcal{M}_2^+$. This time the conditional is unproblematic[25], but there cannot be, among other things, an exclusive disjunction:

		$\nleftrightarrow^{\mathcal{M}_2^+}$
1	1	$\{0\}$
1	0	$\{1\}$
0	1	$\{1\}$
0	0	$\{0\}$

In the presence of this connective, once again, *classical* negation $\neg_C$ becomes definable:

$$\neg_C\phi =_{def} (\phi \vee \neg\phi) \nleftrightarrow \phi$$

[24]A version of this result was used by Roman Suszko in the 70's [20] to motivate the claim that there are only two truth-values. This is sometimes called Suszko's Thesis. A proof of this result as well as some discussion can be found in [19].

[25]Well, this is true as long as there is no absurdity constant $\perp$ available in the language nor any sentence that is false in every valuation. Notice that without such things a Curry sentence cannot be constructed. However, if there is an absurdity constant or a sentence that is always false, the conditional can still be used to define classical negation: $\neg_C\phi =_{def} \phi \supset \perp$.

So the theories we have considered, even in their strengthened versions, are no stronger than K_3 and LP. Any existing criticisms related to the weakness of these logics apply to $\mathcal{M}_1^+$ and $\mathcal{M}_2^+$ as well. And there have been many such criticisms, specially regarding the conditional of these theories. K_3 does not validate Identity (i.e. $\phi \supset \phi$), while LP does not validate Modus Ponens. For example, [12] and [10] are attempts to add a strong conditional connective to K_3 and (a theory similar to) LP, respectively. I know of no similar attempts for $\mathcal{M}_1^+$ or $\mathcal{M}_2^+$, but we can safely claim that if a nice conditional can be added to the former theories, it can also be added to the latter theories. In any case, I will not go into the details here, since the jury is still out on whether these attempts are successful or not.

Instead, I will finish by mentioning another strategy to cope with the problem of the conditional. It should be clear by now that by taking a matrix and making it non-deterministic, the original logic is (possibly) weakened. This might not be a good strategy for matrices which are already charged of being too weak, such as K_3 and LP, but it can be a good idea for matrices that are too strong.

One interesting example of this is the continuum-valued Łukasiewicz logic Ł_∞ (see [16] for details), which is known to be too strong for transparent truth in the sense that adding a transparent truth predicate to it produces an ω-inconsistency[26]. So an interesting project is to see whether we can obtain strong subtheories of Ł_∞ which are not ω-inconsistent by making its conditional non-deterministic[27]. In the logic Ł_∞ the conditional is defined in the following way:

$$v(\phi \to \psi) = \begin{cases} 1 & \text{if } v(\phi) \leqslant v(\psi) \\ 1 - (v(\phi) - v(\psi)) & \text{otherwise} \end{cases}$$

However, we can also define a non-deterministic version of Łukasiewicz conditional as follows:

$$v(\phi \to \psi) \in \begin{cases} \{1\} & \text{if } v(\phi) \leqslant v(\psi) \\ \mathcal{V} - \mathcal{D} & \text{otherwise} \end{cases}$$

Although this conditional is rather weak, it is possible to set up several restrictions on the set of admissible valuations in much the same way as we

[26]Semantically speaking, we say that a theory $\mathcal{T}$ is *ω-inconsistent* if and only if for some formula ϕ and for each term t, $\mathcal{T} \models \phi[t/x]$ but $\mathcal{T} \models \exists x \neg \phi(x)$. A proof that L_∞ plus a transparent truth predicate is ω-inconsistent can be found in [18].

[27]Relevant to this sort of project is [4], where some (strong) subtheories of Ł_∞ are identified axiomatically.

did for the weak logics $\mathcal{M}_1$ and $\mathcal{M}_2$. By doing so we can identify a number of strong non-deterministic subtheories of $Ł_\infty$ which are not obviously ω-inconsistent[28].

6 Concluding remarks

In this paper I have argued that there are two-valued logics that can support a transparent truth predicate as long as the matrix by which negation is defined is non-deterministic. Moreover, we have seen how to obtain logics that are two-valued versions of the well-known theories K_3 and LP by the use of such matrices. I have argued that the friends of paracomplete and paraconsistent logics have reasons to prefer the two-valued presentations of these logics (specially the presentation in terms of non-deterministic matrices) over other formal frameworks where an extra semantic category is overtly or tacitly introduced. The extra semantic category brings with it some unwanted consequences which are avoided in the non-deterministic setting. I have also sketched a very natural way of conceptually understanding this setting and I have argued that a number of criticisms that are sometimes presented against these logics are not too damaging.

These considerations can be used to show the power of non-deterministic matrices. In fact, it remains to be explored whether this kind of matrices can be useful in other philosophically relevant issues.

Acknowledgements

Earlier versions of this paper were presented at conferences in Buenos Aires and Santiago de Chile. I am grateful to the audiences of those conferences for their comments and suggestions. I would specially like to thank José Tomás Alvarado, Eduardo Barrio, Natalia Buacar, Eleonora Cresto, Nicolás Loguercio, Lavinia Picollo, Juan Redmond, Thomas Schindler, Damián Szmuc, Diego Tajer and Paula Teijeiro. I owe a special thanks to Federico Pailos who went through several drafts of this paper. I am also very grateful to Dave Ripley for spoting an ugly mistake, and to an anonymous referee of AJL for his/her comments which -I think- have made the paper a lot better.

[28] I have explored this idea in much more detail in [15].

References

[1] Avron, A. and Zamansky, A. (2011) "Non-deterministic Semantics for Logical Systems", in Gabbay, D. and Guenthner, F. (eds.), *Handbook of Philosophical Logic*, Vol.16: 227-304.

[2] Avron, A. and Lev, I. (2005) "Non-Deterministic Multiple-valued Structures", *Journal of Logic and Computation* 15(3): 241-261.

[3] Avron, A. and Konikowska, B. (2005) "Proof Systems for Logics Based on Non-deterministic Multiple-valued Structures", *Logic Journal of the IGPL* 13: 365-387.

[4] Bacon, A. (2013) "Curry's Paradox and ω-Inconsistency", *Studia Logica* 101: 1-9.

[5] Batens, D., De Clercq, K. & Kurtonina, N. (1999) "Embedding and Interpolation for some paralogics. The propositional case", *Reports on Mathematical Logic* 33: 29-44.

[6] Batens, D. (1989) "Dynamic dialectical logics", in Priest, G., Routley, R. and Norman, J. editors, *Paraconsistent Logic. Essays on the Inconsistent*, München, Philosophia Verlag: 187-217.

[7] Batens, D. (1980) "Paraconsistent extensional propositional logics, *Logique et Analyse* 90-91: 195-234.

[8] Beall, J.C. (2013) "LP+, K3+, FDE+, and their 'classical collapse'" , *The Review of Symbolic Logic* 6(4): 742-754.

[9] Beall, J.C. (2011) "Multiple-conclusion LP and default classicality", *The Review of Symbolic Logic*, 4(2): 326-336.

[10] Beall, J.C. (2009) *Spandrels of Truth*, New York, Oxford University Press.

[11] Da Costa, N. and Alves, E. (1977) "A Semantical Analysis of the Calculi C_n", *Notre Dame Journal of Formal Logic*, 18: 621-630.

[12] Field, H. (2008) *Saving Truth from Paradox*, Oxford University Press, Oxford.

[13] Halbach, V. & Horsten, L. (2006) "Axiomatizing Kripke's Theory of Truth", *The Journal of Symbolic Logic* 71 (2): 677-720.

[14] Kripke, S. (1975) "Outine of a Theory of Truth", *The Journal of Philosophy* 72 (19): 690-716.

[15] Pailos, F. & Rosenblatt, L. (2014) "Non-deterministic conditionals and transparent truth", forthcoming in *Studia Logica*.

[16] Priest, G. (2008) *An introduction to non-classical logics*, Cambridge University Press, Cambridge.

[17] Restall, G. (2013) "Assertion, Denial and non-classical theories", in K. Tanaka, F. Berto, E. Mares, and F. Paoli (eds.) *Paraconsistency: Logic and Applications*, Dordrecht, Springer, 81-100.

[18] Restall, G. (1992) "Arithmetic and truth in Łukasiewicz infinitely valued logic", *Logique et Analyse* 140:303-312.

[19] Shramko, Y. & Wansing, H. (2011) *Truth and Falsehood: An Inquiry into Generalized Logical Values*, Trends in Logic Vol. 36, Dordrecht, Heidelberg, London, New York: Springer.

[20] Suszko, R. (1977) "The Fregean axiom and Polish mathematical logic in the 1920's", *Studia Logica* 36: 373–380.

[21] Weir, Alan (2013) "A Robust Non-transitive Logic", forthcoming in *Topoi*.