Cohabiting with the Logical Paradoxes: A Negative Assessment and a Proposal

Riccardo Bruni and Francesco Orilia***

* *University of Florence*

** *University of Macerata*

Abstract

At the commonsensical level of the manifest image, we seem to take for granted logical laws of all sorts, including classical logic (CL) and naive principles of truth and predication (TP), which, however, generate logical paradoxes such as the liar, Russell's paradox and Curry's paradox. The formal logic of the scientific image comes to the rescue by proposing many competing formal systems that restore consistency, by sacrificing either principles of CL or principles of TP. We wish to explore a different path, which aims at saving both CL and TP, and deals with the paradoxes when they come to the fore, without swallowing contradictions or explosion. We consider the viability of Batens' Inconsistency-Adaptive Logic (IAL) to pursue this goal and we end up with a negative assessment. We then sketch an alternative proposal that incorporates IAL's distinction between provisional and final derivability.

Keywords: Truth, Predication, Logical paradoxes, Adaptive logics, Provisional derivability, Final derivability.

1. Introduction

There is a familiar clash between the scientific image offered by science and the manifest image emerging from common sense. This dichotomy is typically put forward at a metaphysical level, but it can be extended to logic by distinguishing between the formal logic of the scientific image and the informal logic of the manifest image. At the commonsensical level of the manifest image, we seem to take for granted logical laws and rules of all sorts, which taken together could be seen as constituting a logical system, which we may call the *Global Deductive System*, or GDS in short (following Orilia 2014). This presumably includes classical logic (CL), as well as naive principles of truth and predication (TP). However, as it is well known, CL and TP taken together generate logical paradoxes such as the liar, Russell's paradox and Curry's paradox. Russell's paradox and the liar gener-

Argumenta 8, 2 (2023): 357-371
ISSN 2465-2334

DOI 10.14275/2465-2334/202316.bru

ate contradictions, and, by assuming CL and its Ex Falso Quodlibet rule, explosion, i.e., that every proposition whatsoever can be proven. Curry's paradox, on the other hand, generate explosion directly.

The formal logic of the scientific image comes to the rescue by proposing many competing formal systems that restore consistency, by sacrificing in some way or another either principles of CL or principles of TP. For example, in the truth theory of Kripke 1975, or the property theory of Field 2004, the law of Excluded Middle of CL is not generally valid, and in the paraconsistent logic of Priest 1987 Modus Ponens and the principle of non-contradiction are rejected. In contrast, the truth theory of Gupta and Belnap 1993, and the property theories of Orilia 2000, and Orilia and Landini 2019, save CL, but circumscribe TP.

It is as if the scientific image prescribes a reform of the GDS based on one or the other of such proposals; any such reform in effect rejects as non-valid some *prima facie* valid logical rules. In this way however we do not do full justice to the manifest image, for it seems that ordinary reasoners are deeply committed to all the *prima facie* principles that are sacrificed in one or the other reformatory proposal, and thus it is difficult to admit that the GDS does not really include some of them. Moreover, the number of reformatory proposals is very large (see Cantini and Bruni 2021: §6), and it is hard to see how one can choose among them. Indeed, it is even hard to understand how there can be such radical disagreements among the experts about how the GDS should be reformed.

These problems suggest a different path. Perhaps, we should save the manifest image and make it compatible with the scientific image in a different way. That is, we should seek consistency, not by rejecting rules of the GDS, but by administering deductions by meta-deductive principles that allow one to deal with problematic inferences when they come to the fore, without swallowing contradictions or explosion, while otherwise enjoying the full deductive power of the GDS.

An approach of this sort has been informally outlined in Orilia 2014. In that work, it was also suggested that the adaptive logic framework put forward by Batens in many publications (e.g., Batens 1999, 2000, 2001) could provide a way to formalize the approach in question. In this paper, we explore this suggestion and we end up with a negative response. Unfortunately, adaptive logic does not seem fit for this goal. We then move forward by making some further steps toward an appropriate formalization, in a way that inherits some ideas from adaptive logic, in particular, a distinction between provisional and final derivability.

The paper is organized as follows. In §2, we make our desiderata more precise, introduce a formal language adequate to them, and explain how Russell's paradox and Curry's paradox arise, in a format useful to explain our proposal in the following. In §3, we discuss adaptive logic and explain why, despite its initial appeal, is not in fact satisfactory. In §4, we focus on our proposal, explain how it is meant to deal with the paradoxes, without sacrificing neither CL nor TP, and lay down directions for future research.

2. Desiderata, Formal Language and Paradoxes

The project of saving the manifest logical image, as we may call it, generates the following three desiderata for a GDS, $\mathcal{T}$.

(D1) $\mathcal{T}$ is a classical theory, in the sense that it contains the rules of inference of CL, and deductively grants all the laws of CL.[1]

(D2) $\mathcal{T}$ must be consistent, in the sense that it should not feature any logical contradictions among its theorems, and it must be non-trivial, in the sense that the set of its theorems does not have to coincide with the whole set of sentences of the language it is based upon.[2]

(D3) $\mathcal{T}$ contains TP, i.e., naive principles of truth and predication.

It is immediately clear for all those who are acquainted with the literature on logical paradoxes where the difficulty of keeping together these desiderata lies: for every realization of (D3) one can think of, the paradoxes make it impossible to think of keeping (D1) and (D2). The solution that we are going to explore here relies on approaching to the concept of "provability" in a way that departs significantly from the one in which this is usually implemented in formal logic. It is based on a distinction between *provisionally derivable* sentences, and *finally derivable* ones. This distinction leads to considering deductive processes that may be revised, so as to reject previously derived conclusions in order to avoid inconsistency. The family of logics of the adaptive type going under the name of inconsistency-adaptive logics may seem to provide what is needed here. In the next section, we shall explain the reason why these logical systems are natural candidates for trying to formally capture an approach of this sort, in the spirit of the proposal made in Orilia 2014. However, we shall also show that the adaptive mechanism is not fully satisfactory in this respect. In the light of the critical issues that we isolate, we shall then move on to our proposal in §4. Before proceeding we shall here make explicit a formal framework apt to implement our desiderata and review how it gives rise to logical paradoxes.

A formal language appropriate to desiderata (D1) and (D3) is the language $\mathcal{L}^+$ from Orilia 2000, which is also at use in Orilia and Landini 2019. $\mathcal{L}^+$ is based on a standard alphabet, with symbols for the logical connectives "$\neg$" (negation), "&" (conjunction), "$\vee$" (disjunction), "$\rightarrow$" (conditional), and "$\leftrightarrow$" (biconditional), as well as for the quantifiers "$\forall$" (universal) and "$\exists$" (existential), the abstraction operator "λ", a denumerable set $\mathcal{V}$ of individual variables, and a denumerable set $\mathcal{P}^+$ of predicate constants including the dyadic predicate constant "$=$" (the identity predicate), and, for any $n > 0$, the predicate constants "p^n" (the *predication* or *exemplification* predicates), plus additional punctuation signs, such as parentheses and brackets.[3]

[1] What we mean here by "logical theory" is an informal notion, basically referring to the collection of inferences that are accepted for the sake of producing an argument. Our project is to find formal candidates to play the role of such a collection, hence to substitute a candidate theory in this informal sense with a proper, formal one, even though this may turn out not to be a "standard" system of axioms (i.e., a theory whose set of theorems is a recursively enumerable collection of formulas).

[2] If $\mathcal{T}$ is a classical theory, i.e., it is based upon classical logic, then consistency is enough to entail non-triviality (and *vice versa*). So, if (D1) already holds, then (D2) simply amounts to requiring that $\mathcal{T}$ be consistent (or non-trivial). However, if one considers (D2) *per se*, as we are doing here, then consistency and non-triviality are not equivalent as, for example, a theory based on a non-classical logic, like, e.g., a paraconsistent logic, can be inconsistent but non-trivial (see §3).

[3] We have here retained the definition of the language $\mathcal{L}^+$ from Orilia 2000 for the sake of illustrating a general point. In particular, the language as it is here defined contains no

The collections $\mathcal{T}^+$ and $\mathcal{F}^+$ of expressions of $\mathcal{L}^+$, containing the *terms* and *well-formed formulas* of $\mathcal{L}^+$ respectively, are defined by simultaneous induction as follows:

Definition 1. *The sets $\mathcal{T}^+$, $\mathcal{F}^+$ are the smallest collections of expressions of $\mathcal{L}^+$ such that:*

(T1) all elements of $\mathcal{V}$ are elements of $\mathcal{T}^+$ as well;

(F1) if $t_1, \dots, t_n$ are elements of $\mathcal{T}^+$ and p^n is any n-adic predicate constant, then $p^n(t_1, \dots, t_n)$ is an (atomic) element of $\mathcal{F}^+$;

(F2) if x is an element of $\mathcal{V}$, and A and B are elements of $\mathcal{F}^+$, then $\neg A$, $(A \& B)$, $(A \vee B)$, $(A \rightarrow B)$, $(A \leftrightarrow B)$, $\exists x A$ and $\forall x A$ are elements of $\mathcal{F}^+$ as well;

(T2) if $x_1, \dots, x_n$ are elements of $\mathcal{V}$ and A is an element of $\mathcal{F}^+$, then $[\lambda x_1, \dots, x_2. A]$ is an element of $\mathcal{T}^+$.

We shall often omit superscripts from formulas $p^n(t_1, \dots, t_n)$, as they can be easily recovered from the context.

The reasons for choosing such a language have been extensively explained elsewhere, for example in Orilia and Landini 2014: §2, §4. Thus, for space consideration, we shall not dwell on them again, but the main idea is to have a type-free, first-order language for predication,[4] which could replicate and account for phenomena which seem to be present in natural languages.[5] In particular, the predicate p^1 can work as a truth predicate, when applied to vacuous lambda terms (as in $p^1([\lambda x. A])$), and the predicates p^n (for $n \geq 2$) can work as predication relations, when applied to non-vacuous lambda terms (as in $p^2([\lambda x. A(x)], t)$).

Clearly, to meet desideratum (D1), we need to associate to this language a complete system of rules for CL. Moreover, as explained in Orilia 2014, to meet desideratum (D3), one should also associate to this language the following generalized lambda-conversion schema:

(λ-conv) $p^n([\lambda x_1, \dots, x_n. A], t_1, \dots, t_n) \leftrightarrow A(t_1/x_1, \dots, t_n/x_n)$, for every formula A in $\mathcal{F}^+$ and terms $t_1, \dots, t_n$ in $\mathcal{T}^+$.

Let Γ^+ be the collection of instances of the lambda-conversion schema over formulas of $\mathcal{L}^+$, namely the whole set of instances (λ-conv). The combination of the expressive power of the language $\mathcal{L}^+$ with the deductive power of Γ^+ provides us with a very general theory of predication. However, the combination of Γ^+ with classical logic is explosive, as it allows one to reconstruct, for instance, the arguments by Russell and Curry leading to contradiction or explosion.

individual constants. Of course, should one have particular applications in mind (like, for instance, introducing a theory for doing arithmetic), this feature would have to be changed, but to get an extension of the given language by means of new constants is unproblematic, as is well known. Since we are not concerned with any such application in this paper, we have preferred to leave the definition of $\mathcal{L}^+$ as it was.

[4] As a referee has pointed out to us, the language $\mathcal{L}^+$ is, literally speaking, not "type-free" if by this we mean that there are no distinctions of expression types. The language is type-free however in the sense that its predication predicates allow one to view lambda terms as capable of occurring in both subject and predicate position, and thus as capable of giving rise to self-predication, as it will be illustrated in the discussion of Russell's and Curry's paradoxes below.

[5] For instance, this language makes it possible to foster a compositional approach to meaning *à la* Montague (see Montague 1974: Chapters 6-8, for instance), as explained by Orilia and Landini (2014: §4).

For Russell's paradox, let $R \equiv [\lambda x. \neg p(x,x)]$. Then the formula $p(R,R) \leftrightarrow \neg p(R,R)$ is in our set Γ^+. (We shall abbreviate the formulas $p(R,R)$ and $\neg p(R,R)$ of $\mathcal{L}^+$ with $R \in R$ and $R \notin R$, respectively.)

For Curry's paradox, let A be an arbitrary, but fixed formula of $\mathcal{L}^+$. Take t_A to be $[\lambda x. (p(x,x) \to A)]$. Call C_A the formula $p(t_A, t_A)$, the Curry formula for A. Then, $p(t_A, t_A) \leftrightarrow p(t_A, t_A) \to A$ is also in Γ^+.

The derivations of the paradoxes then go on according to the standard route, which is the following, as far as Russell's paradox is concerned:[6]

(1) $R \in R \leftrightarrow R \notin R$	λ-conv
(2) $R \in R \to R \notin R$	ConjEl
(3) $R \in R \to R \in R$	Ref
(4) $R \in R \to (R \in R \,\&\, R \notin R)$	ConjCons
(5) $\neg(R \in R \,\&\, R \notin R)$	LNC
(6) $R \notin R$	MT
(7) $R \notin R \to R \in R$	ConjEl
(8) $R \in R$	MP.

Curry's paradox is then obtained as follows:[7]

(1) $C_A \leftrightarrow (C_A \to A)$	λ-conv
(2) $C_A \to (C_A \to A)$	ConjEl
(3) $\big(C_A \to (C_A \to A)\big) \to (C_A \to A)$	Contr
(4) $C_A \to A$	MP
(5) $(C_A \to A) \to C_A$	ConjEl
(6) C_A	MP
(7) A	MP.

3. Adaptive Logic and the Paradoxes

Adaptive logics are deductive systems that have been introduced by Diderik Batens for the sake of recovering features that are commonly lacking in standard formal theories and that are typical of the natural way of arguing.[8] In particular, adaptive proofs exhibit two sorts of dynamics that are absent in standard formal systems:

(A) an *external dynamics* that may cause a conclusion to be withdrawn in view of new information;

[6] We use a simplified Hilbert-style presentation with abbreviations for the relevant applications of classical laws and rules of inference involved: "λ-conv" just indicate an instance of a formula contained in the set Γ^+; "ConjI" ("ConjEL") is an abbreviated reference to the derivation going from a premise of the form A, B ($A \,\&\, B$) to $A \,\&\, B$ (either A, or B); "Ref" indicates an instance of the classical law of reflexivity $A \to A$; "ConjCons" refers instead to the inference going from $A \to B$ and $A \to C$ to $A \to (B \,\&\, C)$; "LNC" indicates an instance of the classical law of non-contradiction $\neg(A \,\&\, \neg A)$; "MT" is an abbreviation for *modus tollens* and "MP" for *modus ponens*.

[7] In addition to the abbreviations used before and explained in Footnote 6, we here introduce "Contr" to indicate an instance of the classical law of contraction.

[8] The adaptive logic project is now huge. See the webpage illustrating all the directions of work it has given birth to at http://logica.ugent.be/adlog/al.html.

(B) an *internal dynamics* according to which a conclusion may be withdrawn in view of a better understanding of the premises provided by a continuation of the reasoning.

On the basis of the first feature, adaptive proofs are non-monotonic with respect to the set of their premises, since a conclusion drawn on the basis of a certain set of this sort might be withdrawn in a deduction from a larger set. On the basis of the second feature, adaptive proofs are also non-monotonic with respect to the length of the proof, since a conclusion drawn at a certain stage in a proof might be withdrawn at a later stage. So, every conclusion in an adaptive logic is provisional, as it can be revised in view of additional information. This is made possible by means of a *marking system* that keeps track of all those deductive steps that are candidate to be critical with respect to the adaptive proviso we have just described: a conclusion is marked whenever it is drawn under the assumption that some of the premises it relies upon are normal formulas (i.e., whenever it would require the application of an inference of the upper-limit logic—see below—that is not fully safe).

In short,[9] an adaptive logic AL can be presented as a triple whose components are: a *lower-limit logic*, a set of formulas representing *abnormalities*, i.e., formulas that exhibit a non-normal behaviour, that one would want to "block" deductively speaking, and a *strategy* which coincides with the rules governing the marking system and how it applies to proofs. The lower-limit logic reflects how prudent one is willing to be in this respect, as it contains all of the axioms and inferences that are regarded as safe. Therefore, all consequences drawn by means of the lower-limit logic are never retracted and they end up being finally derivable. The upper-limit logic of the systems we are going to consider for our own purposes is CL.

As we saw, the GDS must come to terms with the contradictions arising from the paradoxes. Therefore, adaptive logics that could work as the GDS-like theory we are looking for are those in which the set of abnormalities contains logical contradictions. This is the characteristic feature of the so-called *inconsistency-adaptive logics* from Batens 1999.

The three elements which characterize AL in general take the following shape in case AL is an inconsistency-adaptive logic (IAL, henceforth):

(IAL1) the lower-limit logic of the IAL is a paraconsistent logic;
(IAL2) the set of abnormalities are logical contradictions;
(IAL3) the adopted strategy may vary, with the so-called *reliability strategy* and the *minimal abnormality strategy* (both described in Batens 1999), being the most prominent examples.

The paraconsistent logic that serves the purpose of lower-limit logic is usually chosen in such a way to avoid explosion, so that it normally coincides with weak paraconsistent systems that somehow help blocking the deductive consequences of logical contradictions. Examples of IALs are the theories presented in Batens 1999, Batens 2000, and Verdée 2012. The choice of the lower-limit logic and the set of abnormalities determines, as we said, the upper-limit logic, which is CL.

[9] The informal presentation of adaptive logics that follows, as well as the slightly more detailed one that can be found right after it, is based upon some general characterization of adaptive systems available in the literature, e.g. in Batens 2001.

This justifies viewing derivability in the lower-limit logic as corresponding to derivability in CL with the additional assumption about certain abnormalities being false.[10] As far as the strategy is concerned, we assume here the reliability strategy, which is simpler to describe, since changing the strategy would be irrelevant with respect to the critical issues we raise in the following.

The nice aspect of IALs (as well as of ALs in general) is that they come equipped with a more or less standard proof theory. So, derivations in an IAL admit a simple description (despite their not being simple objects)[11], that can be informally explained as follows. Let $\mathcal{L}$ be a standard first-order language, and let Γ and A be a finite set of formulas of $\mathcal{L}$ and a formula of $\mathcal{L}$, respectively. Let **S** be an arbitrary, but fixed, IAL with LLL as its lower-limit logic, and let Ω be the set of its abnormalities. The derivation of A from Γ in **S** depends upon *conditions* Δ, where Δ is a subset of Ω, and it amounts to three cases: either (i) A is a premise, that is A is an element of Γ, and then it is provable in **S** under no condition (i.e., under conditions $\Delta = \emptyset$), or (ii) A is derivable in the LLL from formulas that have already been derived in **S** from Γ, and then it is derivable under all conditions upon which the derivations of these formulas depend, or (iii) A is derivable in the LLL from formulas that have already been derived in **S** from Γ under the assumption that a finite set Θ of abnormalities is false, and then A is derivable in **S** from Γ under all of the conditions on which the proofs of the previously derivable formulas depend, plus the formulas in Θ.

This informal description has the following formal counterpart, where we describe the proof procedure as being given in *stages*, to indicate which we use the Greek letters $\alpha, \beta, \ldots$ possibly with subscripts. The only assumptions that we make about stages in an **S**-proof is that these are numbers greater than 1.[12]

Definition 2. Let *Γ be a set of formulas of $\mathcal{L}$, $\Delta \subseteq \Omega$, A a formula of $\mathcal{L}$, and* $\alpha \geq 1$. *We say that A is provable in **S** from Γ at stage α under conditions Δ (in symbols: $\Gamma \vdash_{\alpha}^{\Delta} A$), if and only if:*

(P1) either $A \in \Gamma$, $\Delta = \emptyset$ and α is any number greater than 1,

(P2) or, for some formulas $B_1, \ldots, B_n$ of $\mathcal{L}$ we have $B_1, \ldots, B_n \vdash_{\mathrm{LLL}} A$, and, for some $\alpha_1, \ldots, \alpha_n$ with $1 \leq \alpha_i < \alpha$, and for some $\Delta_1, \ldots, \Delta_n \subseteq \Omega$ we have $\Gamma \vdash_{\alpha_j}^{\Delta_j} B_j$ for every $1 \leq j \leq n$, and $\Delta = \bigcup_{1 \leq j \leq n} \Delta_j$,

[10] As a matter of fact, one can show that, for every formula A of the language $\mathcal{L}$, A is provable in CL if and only if there exists a finite set Δ of abnormalities such that $A \vee \Delta$ is provable in the lower-limit logic, where $\vee\Delta$ is the disjunction of all the formulas in Δ (i.e., $\vee\Delta := (D_1 \vee D_2 \vee \ldots D_n)$ if $\Delta = \{D_1, \ldots, D_n\}$). See for instance Batens 1999, where both the case of a propositional IAL, as well as a first-order one are considered.

[11] The non-monotonic character of the process by means of which finally derivable formulas—see below—, are obtained in an AL makes proofs essentially infinite objects, as it is shown in Horsten and Welch 2007.

[12] In view of the result by Horsten and Welch (2007: §3) about adaptive derivations being infinite objects, stages should be indicated by ordinals, or by means of a suitable ordinal notation. Our own notation is chosen to remain coherent with this fact. However, since the matter is largely irrelevant for the purposes of the present paper, we have decided to avoid stressing this aspect any further here.

(P3) or, for some formulas $B_1, \ldots, B_n$ *of* $\mathcal{L}$ *and* $\Theta \subseteq \Omega$*, we have* $B_1, \ldots, B_n \vdash_{\mathrm{LLL}} A \vee \Theta$*, and, for some* $\alpha_1, \ldots, \alpha_n$ *with* $1 \leq \alpha_i < \alpha$*, we have* $\Gamma \vdash_{\alpha_j}^{\Delta_j} B_j, 1 \leq j \leq n$ *and* $\Delta = \bigcup_{1 \leq j \leq n} \Delta_j \cup \Theta$*.*

That a formula is derivable in an IAL **S** in the above sense is just a first step toward the complete definition of theoremhood for a theory like **S**. This is because, as we said, provability in an adaptive system is provisional and it depends upon the strategy adopted to distinguish final theorems from non—final ones. The reliability strategy we are using to illustrate how this distinction may actually apply, is based upon making a difference, as the name suggests, between "reliable" and "unreliable" conclusions. In turn, this is based upon stressing the role of *minimal abnormalities*. In short, a finite set Δ of abnormalities is *minimal* at a given stage α in an S-derivation (or, α*-minimal*), if and only if:

(M1) Δ is derivable in **S** under no condition at a prior stage β, i.e., $\vdash_{\beta}^{\emptyset} \vee\Delta$ for some $\beta < \alpha$;

(M2) no proper subset Σ of Δ is similarly derivable at a stage up to and including α under no conditions, i.e. $\nvdash_{\gamma}^{\emptyset} \vee\Sigma$ for every $\Sigma \subset \Delta$ and $\gamma \leq \alpha$.

Then, we put:

Definition 3. *Let* Γ *be a set of formulas of* $\mathcal{L}$*. Then, for every* $\alpha \geq 1$*, the set* $U_{\alpha}(\Gamma)$ *of the* Γ*-unreliable formulas at* α *is defined as:*

$$U_{\alpha}(\Gamma) := \bigcup\{\Delta : \Delta \text{ is } \alpha\text{-minimal}\}$$

Not surprisingly, "unreliability" in this calculus is a consequence of (provable) abnormality: at any stage of an adaptive proof, unreliable formulas are those that are proved to be abnormal and that are minimally so in the sense of the previous definition of the term.

The idea behind the marking system based on the reliability strategy reflects the attempt of keeping track of anything that is proved under unreliable conditions:

Definition 4. *Let* Γ *be as in the previous definition. Then, for every formula A of* $\mathcal{L}$*, and stages* $\alpha, \beta \geq 1$*, we say that the* ***S****-derivation of A from* Γ *at* α *is marked at stage* β *if and only if* $\Gamma \vdash_{\alpha}^{\Delta} A$ *and* $\Delta \cap U_{\beta}(\Gamma) \neq \emptyset$*, and we say that the derivation is not marked at* β *otherwise.*[13]

It must be noticed that the definition does not presuppose any previously established relation between the stages α and β that it involves. In particular, it might be that $\alpha > \beta$, meaning that the derivation of A from Γ under conditions Δ at α is marked because some of the formulas in Δ have *already* been proved to be unreliable. It might be that $\alpha \leq \beta$ instead, which means that the prior proof of A from Γ under conditions Δ is marked only later, when the information about the unreliability of some of the formulas included in the set of conditions is gained. So, the marking system has quite a dynamics, since markings can apply forward

[13] We are here tacitly incorporating into our definition the modified marking system based on the reliability strategy from Horsten and Welch 2007. The main difference with Batens' original definition (see Batens 2001: 60), lies in the fact that markings are never removed. The two marking systems give rise to equivalent definitions of finally derivable formulas (see Horsten and Welch 2007: §3.2).

(if $\beta < \alpha$), or backward (if $\alpha < \beta$). This dynamic reflects the one affecting the set of unreliable formulas, as this is based on minimally derivable abnormalities, which is something that can change along a proof. For, a given set of formulas Δ that turns out to be minimally derivable at a certain stage of an **S**-derivation, and which is unreliable at that given stage, might be partially "redeemed" at a later stage, due to a subset of it being equally derivable.

This very same dynamics is also responsible for the following definition, that introduces the crucial distinction concerning finally vs. provisionally derivable formulas of **S**:

Definition 5. *A formula A of $\mathcal{L}$ is finally derivable from Γ at α in **S** if and only if the **S**-derivation of A from Γ at α is not marked, and, for every $\beta > \alpha$, the derivation of A from Γ at stage α is not marked at β .*

This concludes our *excursus* on the adaptive model of proof. We now consider whether it could provide a formal way to capture the desiderata set up in §2.

Adaptive proofs are based on the idea of expanding the deductive strength of the LLL of a given AL with what follows from premises that are assumed to behave not abnormally, unless and until these are proven otherwise. So, the LLL represents the set of deductive principles that are regarded as trustworthy. Then, with a precise definition of the logical form of unreliable assumptions, one keeps working out the deductive consequences of formulas that do not turn out to have such a form. It may seem that in this way our desiderata can be met, because contradictions are handled without generating explosion. Nevertheless, there are problems, as we shall now see.

First of all, the idea that the lower-limit logic encapsulates trustworthy deductive rules causes the whole AL to be at least as strong, deductively speaking, as the LLL. For, as it turns out from the definitions we have spelled out in details in the previous section, formulas provable in the LLL, i.e., formulas that are provable under no conditions, are also finally derivable. This means that an AL is as tolerant as its own LLL is toward unconditionally provable abnormalities. If the AL is an IAL, this general fact has the following consequence: minimal abnormalities are finally derivable. In particular, provable logical contradictions are finally derivable. The marking system is made up in such a way as to avoid that the logical consequences of provable contradictions spread, but it does not block the derivability of contradictions themselves. So, IALs are not in general consistent theories (as all logical contradictions that are provable in their LLL are finally derivable), but they are not trivial theories (i.e., they are theories whose set of finally derivable formulas does not coincide with the whole set of formulas of the language).[14]

The previous observation can be made specific, by considering the logical paradoxes. Take the language $\mathcal{L}$ to be $\mathcal{L}^+$ from § 2 and let Γ to be Γ^+. This gives one the means required for carrying on Russell's proof. It follows that a logical

[14] A referee for this journal has pointed out to us that our analysis sounds like a criticism to the adaptive-logic project while neither ALs in general, nor IALs in particular have been conceived with the goal of meeting our desiderata (D1) — (D3). This is certainly correct and we agree with the referee that our criticism applies just to IALs and not to ALs in general. However, among the ALs that have been developed in detail, they seem the most plausible candidates for this goal. Moreover, there are aspects of our proposal from §4 that might recall features of ALs (see for instance the comment in Footnote 21) and we wanted to make sure that there was no possible confusion between the two lines of work.

contradiction involving R is finally derivable in a IAL featuring a fairly weak LLL.[15] So, the "hybrid" nature of an AL, which does not entirely coincide with the nature of its own LLL, is in fact very much conditioned by the latter. In particular, despite the fact that CL is "saved" in an IAL, since it coincides with its upper-limit logic, it is only the axioms and rules of the LLL that are unquestionably accepted, even if they lead to the derivation of a logical contradiction.[16] This obviously goes against our desiderata.

To further stress this conclusion, we can refer to a peculiar relation that may occur in an IAL between its LLL and CL. While the two systems are obviously different proof-theoretically speaking with respect to the whole language $\mathcal{L}$, there is a fragment $\mathcal{L}'$ of it and a "translation" τ of formulas of $\mathcal{L}$ into formulas of $\mathcal{L}'$ such that, for every set Γ of formulas of $\mathcal{L}$ and for every formula A of it, A is derivable from Γ in CL if and only if $\tau(A)$ is derivable from $\tau(\Gamma)$ in LLL (where $\tau(\Gamma)$ is the set of formulas of $\mathcal{L}'$ that one obtains by applying τ to each element of Γ).[17] As noted, formulas derivable in LLL are finally derivable. Hence, the IAL in question is also as deductively strong as CL in the sense that for every formula A that is provable in CL, $\tau(A)$ is finally derivable in it. In a sense, this is good news, for it gives us a way to recover the deductive power of classical logic in an IAL, and weakens the impression that a theory of this sort is thought of as deductively replicating the "nature" of its LLL.

However, it seems to us that this is not enough to declare the above shortcomings as dismissed. For (i) this does not change the fact that logical contradictions remain finally derivable in general; (ii) the previous result is metatheoretical in nature (i.e., the logical equivalence of A and $\tau(A)$ is not itself derivable in the LLL, while it is provable in CL instead[18]), so that, in order to see it as allowing one to recover the deductive power of CL, one should be able to recognize $\tau(A)$ as a translation of A. To make sense of (i) and (ii) it seems that one has to exploit what we have referred to before as the "hybrid nature" of IALs, and to retain any instance of them as being both committed to the view underlying its LLL, and to the one justifying CL (which is the upper-limit logic of a IAL in the most relevant cases). At the same time, this "hybrid" view explaining (i) and (ii), seems impossible to be reconciled with our desiderata, and with (D1) and (D2) in particular:

[15] For instance, the contradiction in question is finally derivable in the LLL of the theory APIL1 from Batens 1999, whose LLL is the paraconsistent logic PIL that is obtained by: (i) restricting the axioms and rules of CL to positive formulas of $\mathcal{L}$ (i.e., formulas that are not within the scope of a negation sign), (ii) by retaining as axioms and rules of PIL, all axioms and rules of CL but those applying to negated formulas, with the only exception of the law of excluded middle, $(\neg A \vee A)$, for each and every formula A of $\mathcal{L}$.

[16] In view of this, it is not surprising that ALs are listed among paraconsistent systems in Priest et al. 2022.

[17] To illustrate this situation, take LLL to be the system PIL from Batens 1999. Then, take τ to be the function recursively defined in such a way that, for every formula A of $\mathcal{L}$, $\tau(A)$ is the formula obtained by substituting every subformula $\neg B$ of A where B is positive (i.e., non-negated) with $B \rightarrow \perp$ (where $\perp$ is a propositional constant for falsity—i.e., characterized in CL by the axiom $\perp \rightarrow C$ for every formula C of $\mathcal{L}$). Notice that, for every A, $\tau(A)$ is a positive formula. Therefore, *all* axioms and rules of CL applied to τ-formulas are provable and valid in PIL.

[18] This is what happens for instance with the exemplification of τ from Footnote 17.

for, while (ii) seem to "push" us in the right direction, (i) necessarily requires to depart from it.[19]

To summarize: ALs regarded as theories allowing one to make use of an inconsistent set of premises as consistently as possible for deductive purposes, namely as IALs, give rise to non-trivial but logically inconsistent theories; theories of this sort featuring principles that would make them fit as candidate GDSs (since they feature naive abstraction principles and naive comprehension principles) are those from Verdée 2012 and Batens 2020; despite the fact that such theories are strong enough to recover the strength of classical logic (even if in the form of a "replica"), they require us to abandon either desideratum (D1), or (D2) and are therefore unfit for our project.[20]

4. Our Proposal

Let us give a look at the proof of A from $C_A \leftrightarrow (C_A \rightarrow A)$ first. What is striking in this construction is that there is nothing of it, neither in the construction of the term t_A and of the formula C_A, nor in the argument leading to the proof of A, that seems to be specific to A itself. The construction behind Curry's paradox is "modular" in the sense that it can be applied *mutatis mutandis* to *any given* A. This is what makes it a particularly malignant construction: A may correspond to a blantantly false pronouncement, like "2+2=5", and thus Curry's construction can be used to disprove a perfectly legitimate arithmetical sentence. Now, imagine to carry on the same construction by substituting everywhere A with $\neg A$. If you perform this substitution inside the term t_A, you obtain the term $[\lambda x.(p(x,x) \rightarrow$

[19] The specific situation involving PIL and CL depicted in Footnote 17, allows one to hint at another instance of this push-pull situation between the two "souls" of an IAL. For, the τ considered there allows one to define a negation symbol $\sim$ in addition to the one the language is already provided with by putting $\sim A := (A \rightarrow \perp)$ as suggested in Batens 2001: 54. This has the effect that logical contradictions that are already provable in their "standard" form, i.e., as $\neg$-contradiction, or formulas of the form $(A \,\&\, \neg A)$, can be shown to be also provable in PIL as $\sim$-contradictions, namely as formulas of the form $(A \,\&\, {\sim}A)$, (since $\sim A$ is a positive formula, as we noted). Of course, this can be recognized as a logical contradiction only "from the point of view" of CL (since $\sim A$ and $\neg A$ are logically equivalent in CL). However, if one feels that (ii) above pushes in the "right" direction, one should also feel that this further consideration pulls back in the direction of LLL, which allows one to argue that $(A \,\&\, {\sim}A)$ is not a logical contradiction in the end.

[20] There is another notable thread of research related to ALs, namely the one that goes back to Batens 2000 and which constitutes the basis for the theory proposed in Verdée 2013: §4, which one may think provides reasons for having second thoughts about this conclusion. In particular, Batens 2000 shows that it is possible to characterize reasoning from maximally consistent subsets of an inconsistent set of premises in terms of adaptive logics. That is: if Γ is an inconsistent set of premises, consider the set $\mathcal{T}$ of all formulas that are deducible from every maximally consistent subset of Γ; then, there exists a IAL **S** whose set of finally derivable formulas contains $\mathcal{T}$. This may lead to think that, despite being inconsistent, IALs can be reduced to consistent theories. However, by closely exploring Batens' result one sees that this is made possible by changing the definition of derivability of the logic and by restricting the admissible premises. This does not change the fact that the theory is inconsistent anyway (as $\mathcal{T}$ is only a proper subset of the set of finally derivable formulas of **S**). As to the theory considered in Verdée: §4, this is defined by using a "modal trick" that changes the set of abnormalities, which do not correspond to logical contradictions in the sense of CL, the view that we are favouring here.

$\neg A)]$, which we abbreviate as $t_{\neg A}$. Then, the formula $p(t_{\neg A}, t_{\neg A}) \rightarrow (p(t_{\neg A}, t_{\neg A}) \rightarrow \neg A)$, which is similarly obtained from C_A by substituting A with $\neg A$, is nothing but $C_{\neg A}$. This is the formula that can be used to derive $\neg A$ by means of Curry's argument. The latter, by the way, amounts to the following proof:

(1) $C_{\neg A} \leftrightarrow (C_{\neg A} \rightarrow \neg A)$	λ-conv
(2) $C_{\neg A} \rightarrow (C_{\neg A} \rightarrow \neg A)$	ConjEl
(3) $(C_{\neg A} \rightarrow (C_{\neg A} \rightarrow \neg A)) \rightarrow (C_{\neg A} \rightarrow \neg A)$	Contr
(4) $C_{\neg A} \rightarrow \neg A$	MP
(5) $(C_{\neg A} \rightarrow \neg A) \rightarrow C_{\neg A}$	ConjEl
(6) $C_{\neg A}$	MP
(7) $\neg A$	MP

Note that this is nothing but the same proof that we used before to prove A, where the same substitution of A with $\neg A$ has been performed. Nothing else is changed, as the steps that make up this argument have remained exactly the same ones. This is an effect of the "modularity" of Curry's construction, although it is a rather counter-intuitive effect. For, it is certainly odd that a proof works well to conclude that a certain formula holds *and* to conclude that its own logical negation also does. This is not what happens with "normal" proofs. These observations prompt the following definition of a *negation-symmetric* proof, which we shall use in our proposal in order to tame Curry's paradox.

Definition 6. *Let $\mathcal{D}$ be a classically valid proof of a formula A from Γ^+. Let $\mathcal{D}^N$ be obtained from $\mathcal{D}$ by substituting every occurrence of A with $\neg A$. We say that $\mathcal{D}$ is negation-symmetric if and only if $\mathcal{D}^N$ is a classically valid derivation of $\neg A$ from Γ^+.*

Let us now turn to Russell's proof. This is not negation-symmetric and we have to deal with it in a different way. We resort to the idea that proofs can have different degrees of validity, depending on the rules that they use. We can do this by placing all rules in a hierarchy with the understanding that a higher place in the hierarchy means greater validity. Thus, for example, the rules at the top of the hierarchy are given a validity of degree 1 and any rule lower in the hierarchy is given a lower degree of validity $1 - r$, where r is some positive real less than 1; the number $1 - r$ is, we may say, the distance that a "lower rule" has from the topmost rules. Once rules are so ordered, we can assign a degree of validity to an argument by subtracting from 1 the sum of all values d such that d is a distance that a lower rule used at least once in the argument has from the topmost rules. For simplicity's sake, we shall consider here only one simple way to assign degrees of validity, although other options could be considered (see Orilia 2014). The simple option that we consider here is this: all rules of CL have the topmost degree of validity, 1, whereas lambda-conversion has a lower degree of validity, inferior to 1 (it is not important to decide what this value is). Given this option, all proofs that use only rules of CL have a maximum degree of validity, whereas proofs that make use of lambda-conversion have a lower degree. Consider, for example, a proof in CL of the following instance of the principle of non-contradiction: $\neg(R \in R \,\&\, R \notin R)$. This proof is more valid than the proof of

$(R \in R \,\&\, R \notin R)$ exhibited above because the former makes use only of rules of CL, whereas the latter uses both rules of CL and lambda-conversion.[21]

Let us assume the following abbreviation: if $\mathcal{D}, \mathcal{D}'$ are any two derivations from Γ^+, $\mathcal{D} \succcurlyeq \mathcal{D}'$ means that the degree of validity of $\mathcal{D}$ is higher than, or equal to, the degree of validity of $\mathcal{D}'$.

We are now ready to offer our notions of "provisional" and "final derivability":

Definition 7. *1. A formula A of $\mathcal{L}^+$ is provisionally derivable from Γ^+ iff there is a derivation $\mathcal{D}$ of A such that $\mathcal{D}$ is not negation-symmetric (we say that $\mathcal{D}$ makes A provisionally derivable).*

2. A formula A of $\mathcal{L}^+$ is finally derivable from Γ^+ iff there exists a derivation $\mathcal{D}$ that makes A provisionally derivable and there is no derivation $\mathcal{D}'$ of $\neg A$ such that (i) $\mathcal{D}'$ is not negation-symmetric, and (ii) $\mathcal{D}' \succcurlyeq \mathcal{D}$.

Clearly, no formula can be even provisionally derived on the basis of the argument of Curry's paradox, since this is a negation-symmetric derivation.[22] As regards Russell's paradox, there is a provisional derivation of $(R \in R \,\&\, R \notin R)$ on the basis of the argument for Russell's paradox offered above, but it is ruled out that it is finally derivable, since, as noted above, there is a proof of $\neg(R \in R \,\&\, R \notin R)$ that is more valid than this argument.

These definitions might not suffice as they are. They may have to be equipped with a "backtrack procedure". To illustrate this need, consider the derivation of $R \notin R$ provided by lines 1-6 of the above proof of Russell's paradox. We can imagine that, in the light of this, one takes $R \notin R$ to be true, *qua* provisionally derived, and uses it in further derivations to prove additional formulas, e.g., P. However, one later realizes that, in a similar vein, $R \in R$ can also be provisionally derived (consider lines 1-8 of the above proof of Russell's paradox). The proof of $R \in R$ and the proof of $R \notin R$ are equally valid, for both make use of

[21] It should be clear that our proposal to introduce degrees of validity in the form of a partial ordering among proofs is not a replica of the adaptive logic scheme. For, it is true that by ordering proofs in the way we want we get as a consequence that there is a system of axioms that we regard as fully reliable (classical logic) and a proper extension of it that is not (classical logic plus lambda abstraction). However, it is not enough to "combine" two systems that may act as a lower-limit logic and as an upper-limit one to get an AL: an AL is determined by the choice of the LLL *and* the logical form of the abnormalities (the formulas that one wants to block, deductively speaking). The upper-limit system is a "consequence" of these choices: it is the logic related to the LLL "by means of" abnormalities in the sense of the theorem we hinted at in Footnote 10. In our case there is no obvious indication that a result like that can be found, although we cannot even exclude that as a fact. We thank a referee for this journal for pointing out to us the need to clarify this issue.

[22] Actually, we should also take care of contingent versions of Curry's paradox, based on contingent assumptions such as, e.g., the longest sentence written on the blackboard is "if the longest sentence on the blackboard is true then 2+2=5", which is true if it so happens that the only sentence on the blackboard is in fact "if the longest sentence on the blackboard is true then 2+2=5". By assuming this fact, a Curry-style argument allows us to derive 2+2=5. We do not deal with this issue here for simplicity's sake, but we assume that the approach we are pursuing here can be extended to cover such cases in the way suggested in Orilia 2014: 193.

lambda-conversion, and accordingly neither $R \in R$ nor $R \notin R$ are finally derivable, since the latter is the negation of the former. To reject both of them is fine, but it is not enough, since presumably we do not want to assume at this point that P is finally derivable, since it was established on the basis of a proof relying on $R \notin R$. We rather want to reject this proof and rule out P from what is finally derivable (unless of course P is established via other means). In general, once a provisionally derived formula A is rejected, we want to backtrack all formulas that were derived by means of A, and reject them as well, unless they were also derived independently of A.

It should be clear from the foregoing that there is something in common between our approach and adaptive logic, as they both make use of a distinction between provisional and final derivability. However, our approach departs from the paradigm of adaptive logics, for while adaptive logics are based on setting criteria on formulas (in particular, in choosing which logical form corresponds to an abnormality), our approach is based on setting criteria on proofs.

Some natural questions about our proposal arise and should be tackled in future research. In particular: is the set of finally derivable formulas actually consistent? What is its complexity, or what can be said about the theory they give rise to?

Adaptive logics are known for leading to theories that exhaust by far the complexity of formal theories in their standard shape (see Horsten and Welch 2007). And there have been other attempts to "relax" these limits in the past (see Magari 1974, Jeroslow 1975, Hájek 1977), some of which have been recently rediscovered within a trial-and-error approach to do mathematics (see Amidei et al. 2016a, 2016b). Especially if a positive answer as regards the consistency issue can be given, it would be interesting to look at this matter more closely and see if any precise relationship between some of these approaches and our proposal can be found.[23]

References

Amidei, J., Pianigiani, D., San Mauro, L., Simi, G., and Sorbi, A. 2016a, "Trial and Error Mathematics I: Dialectical and Quasidialectical Systems", *The Review of Symbolic Logic*, 9, 299-324.

Amidei, J., Pianigiani, D., San Mauro, L., Simi, G., and Sorbi, A. 2016b, "Trial and Error Mathematics II: Dialectical Sets and Quasidialectical Sets, Their Degrees, and Their Distribution within the Class of Limits Sets", *The Review of Symbolic Logic*, 9, 810-35.

Batens, D. 1999, "Inconsistency-adaptive logics", in Orlowska E. (ed.), *Logic at Work: Essays Dedicated to the Memory of Elena Rasiowa*, Heidelberg: Springer, 445-72.

Batens, D. 2000, "Towards the Unification of Inconsistency Handling Mechanisms", *Logic and Logical Philosophy*, 8, 5-31.

[23] This work was supported by the Italian Ministry of Education, University and Research through the PRIN 2017 project "The Manifest Image and the Scientific Image" prot. 2017ZNWW7F_004. It was presented at the conference *The manifest image and the scientific image: objects, properties and relations*, at the University of Florence on November, 29-30, 2021. We wish to thank two referees for this journal for their useful comments.

Batens, D. 2001, "A General Characterization of Adaptive Logics", *Logique et Analyse*, 44, 45-68.

Batens, D. 2020, "Adaptive Frege's Set Theory", *Studia Logica*, 108, 903-39.

Cantini, A. and Bruni, R. 2021, "Paradoxes and Contemporary Logic", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2021 edition.

Field, H. 2004. "The Consistency of the Naïve Theory of Properties", *The Philosophical Quarterly*, 54, 214, 78-104.

Gupta, A. and Belnap, N. 1993, *The Revision Theory of Truth*, Cambridge, MA: MIT Press.

Hájek, P. 1977, "Experimental Logics and Π^0_3 Theories", *The Journal of Symbolic Logic*, 42, 515-22.

Horsten, L. and Welch, P. 2007, "The Undecidability of Propositional Adaptive Logic", *Synthese*, 158, 41-60.

Jeroslow, R., 1975, "Experimental Logics and Δ^0_2-Theories", *Journal of Philosophical Logic*, 4, 253-67.

Kripke, S. 1975, "Outline of a Theory of Truth", *The Journal of Philosophy*, 72, 19, 690-716.

Magari, R. 1974, "Su certe teorie non enumerabili", *Annali di Matematica Pura ed Applicata*, 48, 119-52.

Montague, R. 1974, *Formal Philosophy*, New Haven: Yale University Press.

Orilia, F. 2000, "Property Theory and the Revision Theory of Definitions", *The Journal of Symbolic Logic*, 65, 212-46.

Orilia, F. 2014, "Degrees of Validity and the Logical Paradoxes", in Weber, D.W.E. and Meheus, J. (eds.), *Logic, Reasoning and Rationality*, Dordrecht: Springer, 176-96.

Orilia, F. and Landini, G. 2019, "Truth, Predication and a Family of Contingent Paradoxes", *Journal of Philosophical Logic*, 48, 113-36.

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

Priest, G., Tanaka, K., and Weber, Z. 2022, "Paraconsistent Logic", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2022 edition.

Verdée, P. 2012, "Strong, Universal and Provably Non-Trivial Set Theory by means of Adaptive Logic", *Logic Journal of the IGPL*, 21, 1, 108-25.

Verdée, P. 2013, "Non-Monotonic Set Theory as a Pragmatic Foundation of Mathematics", *Foundations of Science*, 18, 4, 655-80.