L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006

ARTICLES

DIDERIK BATENS, *On a Logic of Induction* p. 3

THEO A. F. KUIPERS, *A Brand New Type of Inductive Logic. Reply to Diderik Batens* 33

ALESSANDRO BECCHI, *Łukasiewicz e il determinismo logico* 39

ROBERT G. HUDSON, *Howson on Novel Prediction* 91

FRANCESCO PAOLI, *Comparing Two Views of Comparison: Peña and Casari on Vagueness and Comparatives* 105

CRITICAL NOTICE

ROBERTO FESTA, *Philosophy, Science, and (Anti-)Communism: The Two Lives of Imre Lakatos* 123

Information on the Journal 133

L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/
ISSN: 1826-1043

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 3-32

On a Logic of Induction*

Diderik Batens
Centre for Logic and Philosophy of Science,
University of Ghent (Belgium)
e-mail: diderik.batens@UGent.be

1. Prelude
2. Aim of this paper
3. Tinkering with the dynamic proof theory
4. The dynamic proof theory
5. The semantics
6. Heuristic matters and further comments
7. Further research

ABSTRACT. In this paper I present a simple and straightforward logic of induction: a consequence relation characterized by a proof theory and a semantics. This system will be called **LI**. The premises will be restricted to, on the one hand, a set of empirical data and, on the other hand, a set of background generalizations. Among the consequences will be generalizations as well as singular statements, some of which may serve as predictions and explanations.

1. Prelude

I published my first paper in English a long time ago. In the paper (Batens 1968) I compared Carnap's and Popper's approach to induction, and basically assigned each approach a context of application, except that a modification was proposed for Popper's corroboration function. I had sent the paper to Carnap, Popper, Hempel, Kemeny, and several other famous people. With one exception, all had returned a few encouraging lines. Not long thereafter, I received a letter, in Dutch, by someone I immediately recognized as Dutch be-

* This paper appeared in R. Festa, A. Aliseda and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation* (*Poznań Studies in the Philosophy of the Sciences and the Humanities,* vol. 83), Amsterdam/New York, NY: Rodopi, 2005, pp. 221-247. As a result of an unhappy mistake, the book contains the uncorrected version, which is hardly understandable at some points. For this reason the corrected version is made available here.

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

cause he used an impressive number of middle initials – the Flemish use them in official documents only. The letter contained some questions and suggestions; a brief correspondence ensued.

I left the field later. However, for the sake of an old friendship, I dedicate this first logic of induction to Theo.

2. Aim of this paper

It is often said that there is no logic of induction. This view is mistaken: this paper contains one. It is not a contribution to the great tradition of Carnapian inductive logic – see Kuipers (2000, Ch. 4); it is a logic of induction in the most straightforward sense of the term, a logic that, from a set of empirical data and possibly a set of background generalizations, leads to a set of consequences that comprises generalizations and their consequences. Incidentally, the underlying ideas oppose the claims that were widespread in Carnap's tradition – see, for example, Bar-Hillel (1968).

LI is characterized by a proof theory and a semantics. Some people will take these properties to be insufficient for calling **LI** a logic. I shall not quarrel about this matter, which I take to be largely conventional. As far as I am concerned, any further occurrence of 'logic' may be read as 'giclo'. The essential point is that **LI** is characterized in a formally decent way, that its metatheory may be phrased in precise terms, and, most importantly, that **LI** may serve to explicate people's actual inductive reasoning.

LI takes as premises descriptions of empirical data as well as background generalizations that are formulated in the language of standard predicative logic. Its consequences follow either deductively or inductively from the premises. By deductive consequences I mean statements that follow from the premises by Classical Logic (**CL**). The main purpose of **LI** obviously concerns the inductive consequences. In this respect the proof of the pudding will be in the eating: the reader will have to read this paper to find out whether he or she considers **LI** as sensible with respect to the intended domain of application. For now, let me just mention that the inductive consequences of a set of empirical data and a set of background knowledge will, first and foremost, be empirical generalizations, and next, the deductive consequences of the empirical generalizations and the premises, including singular statements that may serve the purposes of prediction and explanation.

LI is only one member of a family of logics. It is severely restricted by the standard predicative language. This rules out statistical generalizations as well as quantitative predicates (lengths, weights, etc.). **LI** will not take account of degrees of confirmation or the number of confirming (and disconfirming) instances. **LI** will not deal with serious problems, usually connected to discovery and creativity, such as the genesis of new concepts and other forms of conceptual dynamics. Nor will **LI** deal with the historically frequent case of inconsistent background knowledge – see Brown (1990), Norton (1987; 1993), Smith (1988), Nersessian (2002), Meheus (1993; 2002), … **LI** is a bare backbone, a starting point.

More sophisticated inductive logics may be designed by modifying **LI**. Some of the required modifications are straightforward. But given the absence of any logic of induction of the kind, it seems advisable to present a simple system that applies in specific (although common) contexts. Incidentally, I shall also keep my remarks in defense and justification of **LI** as simple as possible. As most people reading the present book will be familiar with the literature on induction, they will easily see further arguments. It also seems wise, in defending a logic of induction, to refrain from siding with one of the many parties or schools in the research on induction. The logic **LI** is intended to please most of these parties. It should serve as a point of unification: this bit at least we all agree about, even if each explains it in his or her own way.

When working on this paper I wondered why a system as simple and clarifying as **LI** had not been presented a long time ago.[1] However, although **LI** is simple and straightforward to understand, its formulation presupposes familiarity with the adaptive logic programme. I shall not summarize this programme here because several easy introductions to its purpose and range are available, such as Batens (2000; 2004). Rather, I shall introduce the required adaptive elements as the paper proceeds. However, it is only fair to the reader to mention that the ideas underlying adaptive logics and dynamic proof theories have some pedigree and are by no means the outcome of the present research.

3. Tinkering with the dynamic proof theory

Children have a profound tendency to generalization. This tendency has a clear survival value. In a sense, our present scientific (and other) knowledge is the result of a sophistication of this tendency. Of course, we know today that all

[1] In the form of a formal logic, that is. Mill's canons come close. There are also clear connections with Reichenbach's straight rule, if restricted to general hypotheses, and with Popper's conjectures and refutations. Articulating the formal logic is worthwhile, as we shall see.

simple empirical generalizations are false – compare Popper (1973, p. 10). This insight is a result of experience, of systematization, of free inquiry, and of systematic research. Our present knowledge, however, is neither the result of an urge that is qualitatively different from children's tendency to systematization, nor the outcome of a form of reasoning that is qualitatively different from theirs.

Let us for a moment consider the case in which only a set of empirical data is available – I shall remove this utterly unrealistic supposition in the present section. Where these empirical data are our only premises, what shall we want to derive from them? Apart from the **CL**-consequences of the premises, we shall also want to introduce some general hypotheses. Only by doing so may we hope to get a grasp of the world – to understand the world and to act in it. And from our premises and hypotheses together we shall want to derive **CL**-consequences (to test the hypotheses, to predict facts, and to explain facts).

LI should define a consequence relation that connects the premises with their **CL**-consequences, with the generalizations, and with their common **CL**-consequences. Is there such a consequence relation? Of course there is. The consequence relation is obviously non-monotonic[2] – inductive reasoning is the oldest and most familiar form of non-monotonic reasoning.

Generalizations that are inductively derived from the set of premises, Γ, should be compatible with Γ. A further requirement on inductively derived statements is that they should be *jointly* compatible with Γ. The latter requirement is the harder one. The logic of compatibility – see Batens and Meheus (2000) – provides us with the set of all statements that are compatible with Γ. The problem of induction is, in its simplest guise, to narrow down this set in such a way that the second requirement is fulfilled. And yet, as I shall now explain, this problem is easy to solve.

Consider an (extremely simple) example of a **CL**-proof of the usual kind – for the time being, just disregard the $\emptyset$s at the end of the lines. As stated before, all premises will be singular statements.

1	$(Pa \wedge Pb) \wedge Pc$	PREM	$\emptyset$
2	$Rb \vee {\sim}Qb$	PREM	$\emptyset$
3	$Rb \supset {\sim}Pb$	PREM	$\emptyset$

[2] A consequence relation '$\vdash$' is non-monotonic iff a consequence of a set of premises need not be a consequence of an extension of this set. Formally: there is a Γ, a Δ, and an A such that $\Gamma \vdash A$ and $\Gamma \cup \Delta \nvdash A$.

4	$(Sa \wedge Sb) \wedge Qa$		PREM	$\emptyset$
5	Pa	1	RU	$\emptyset$
6	Pb	1	RU	$\emptyset$
7	Qa	4	RU	$\emptyset$
8	Sa	4	RU	$\emptyset$
9	Sb	4	RU	$\emptyset$

The rule applied in lines 5-9 is called **RU**. This name refers to the generic "unconditional rule". For the moment, just read it as: formula 5 is **CL**-derivable from formula 1, etc.

Suppose that our data comprise 1-4, and that we want to introduce an empirical generalization, for example $(\forall x)(Px \supset Sx)$. Obviously, this formula is not **CL**-derivable from 1-4. However, we may want to accept it *until and unless* is has been shown to be problematic – for example, because some P are not S. In other words, we may want to consider $(\forall x)(Px \supset Sx)$ as conditionally true in view of the premises. By a similar reasoning, we may want to consider $(\forall x)(Px \supset Qx)$ as conditionally true. This suggests that we add these universally quantified formulas to our proof, but attach a condition to them, indicating that the formulas will not be considered as derived if the condition shows false. So we extend the previous proof as follows:[3]

10	$(\forall x)(Px \supset Sx)$	RC	$\{(\forall x)(Px \supset Sx)\}$
$11^{L_{14}}$	$(\forall x)(Px \supset Qx)$	RC	$\{(\forall x)(Px \supset Qx)\}$

The set $\{(\forall x)(Px \supset Sx)\}$ will be called the *condition* of line 10. If some member of this set is contradicted by the data, the formula derived at line 10, which happens to be $(\forall x)(Px \supset Sx)$, should be withdrawn (considered as not derived). Conditionally derived formulas may obviously be combined by **RU**. As expected, the condition of the derived formula is the union of the conditions of the formulas from which it is derived. Here is an example:

$12^{L_{14}}$	$(\forall x)(Px \supset (Qx \wedge Sx))$	10, 11 RU	$\{(\forall x)(Px \supset Sx), (\forall x)(Px \supset Qx)\}$

The interpretation of the condition of line 12 is obviously that $(\forall x)(Px \supset (Qx \wedge Sx))$ should be considered as not derived if either $(\forall x)(Px \supset Sx)$ or $(\forall x)(Px \supset Qx)$ turns out to be problematic.

[3] The superscript L_{14} on line 11 is explained below.

Logicians not familiar with dynamic proofs will complain that the negation of 11 *is* derivable from 1-4. Let me first show them to be right:

13	$\sim Qb$	2, 3, 6	RU	$\varnothing$
14	$\sim(\forall x)(Px \supset Qx)$	6, 13	RU	$\varnothing$

As $(\forall x)(Px \supset Qx)$ is shown to be contradicted by the data, lines 11 and 12, which rely on the presupposition that $(\forall x)(Px \supset Qx)$ is not problematic, have to be *marked.* Formulas that occur in marked lines are considered as not being inductively derivable from the premises.[4]

Some logicians may still complain: 14 is **CL**-derivable from 1-4, and hence, they might reason, it was simply a mistake to add lines 11 and 12 to the proof. Here I strongly disagree. Moreover, the point touches an essential property of dynamic proofs; so let me explain the matter carefully.

Suppose that Γ is a finite set. In view of the restrictions on generalizations and on Γ, it is decidable whether a generalization (in the sense specified below) is or is not derivable, and hence it is decidable whether some singular statement is or is not derivable. So, indeed, one may avoid applications of RC that are later marked (if Γ is finite). However, nearly any variant of **LI** that overcomes some of the restrictions on **LI** – see earlier as well as subsequent sections – will be undecidable and, moreover, will lack a positive test for derivability.[5]

In view of this, and in preparation for those more fascinating variants, it seems rather pointless to try circumventing a dynamic proof theory for **LI**. There is a second argument and it should not be taken lightly. It is the purpose of the present paper to explicate actual inductive reasoning. Quite obviously, humans are unable to see all the relevant consequences of the available information. Given our finite brains it would be a bad policy to make inductive hy-

[4] When a line is marked I shall sometimes say that the formula that is its second element is marked. We shall see later that there are two kinds of marks, L and B. At stage 14, lines 11 and 12 have to be L-marked. Normally, one would just add an L to those lines. In order to avoid repeating the proof at each stage, I add L_{14} to indicate that the lines are L-marked at stage 14 of the proof.

[5] A logic is decidable iff there is an algorithm to find out, for any finite set of premises and for any formula, whether the formula is derivable from the premises. There is a positive test for derivability iff there is an algorithm that leads, for any recursive set of premises and for any formula, to the answer 'Yes' in case the formula is derivable from the premises. **CL**-derivability is decidable for the propositional fragment of **CL** and undecidable for full **CL**. Nevertheless, there is a positive test for **CL**-derivability. A standard reference for such matters is Boolos and Jeffrey (1989).

potheses contingent on complete deductive certainty. To do so would slow down our thinking and often paralyse it. This does not mean that we neglect deductive logic. It only means that we often base decisions on incomplete knowledge, including incomplete deductive knowledge – see Batens (1995) for a formal approach to the analysis of deductive information. The third (and last) argument is of a different nature. I shall show in this paper that the dynamic proof theory of **LI** is formally sound and leads, within the bounds described in Section 2, to the desired conclusions. All this seems to offer a good reason to continue our journey.

To find out whether the sketched proof procedure holds water, we should also have a look at its weirder applications. Let us consider a predicate that does not occur in our premises, and see what happens to generalizations in which it occurs.

$15^{L_{17}}$	$(\forall x)(Px \supset Tx)$		RC	$\{(\forall x)(Px \supset Tx)\}$

Obviously, 1-4 do not enable one to contradict $(\forall x)(Px \supset Tx)$. However, we may moreover add:

$16^{L_{17}}$	$(\forall x)(Px \supset {\sim}Tx)$		RC	$\{(\forall x)(Px \supset {\sim}Tx)\}$

And now we see that we are in trouble, as the proof may obviously be continued as follows:

17	${\sim}(\forall x)(Px \supset Tx) \vee {\sim}(\forall x)(Px \supset {\sim}Tx)$	5	RU	$\emptyset$

Although neither 15 nor 16 is contradicted by the empirical data, their conjunction is. The thing to do here is obvious (and well known from the Reliability strategy of adaptive logics). As 15 and 16 are on a par, both of them should be considered as *unreliable*, and hence lines 15 and 16 should both be marked in view of their conditions.[6]

Let me straighten this out and introduce some useful terminology. We suppose that generalizations are not problematic until and unless they are shown to be contradicted by the empirical data. So the normal case will be that a generalization is compatible with the data. In view of this, the (derivable) negation of a generalization will be called an *abnormality.* Sometimes abnormali-

[6] In comparison to the Reliability strategy, the Minimal Abnormality strategy leads to a slightly richer consequence set in some cases. I return to this point later.

ties are connected. Line 17 is a good example: the disjunction of two abnormalities is derivable, but neither of the abnormalities is. Derivable abnormalities and derivable disjunctions of abnormalities will be called *Dab*-formulas – an abnormality is itself a disjunction with one disjunct only. Where Δ is a finite set of generalizations, $Dab(\Delta)$ is a handy abbreviation for $\bigvee(\{\sim A \mid A \in \Delta\})$ (the disjunction of the negations of the members of Δ).

In view of the derivability of 17, both $(\forall x)(Px \supset Tx)$ and $(\forall x)(Px \supset \sim Tx)$ are unreliable. But the fact that $Dab(\Delta)$ is derivable does not indicate that all members of Δ are unreliable. Indeed,

$$\sim(\forall x)(Px \supset Qx) \vee \sim(\forall x)(Px \supset Sx)$$

is derivable from 14, but adding this formula to the proof does not render $(\forall x)(Px \supset Sx)$ unreliable. The reason is that, even if the displayed formula were added to the proof, it would not be a minimal *Dab*-formula in view of 14 (in the sense that a formula obtained by removing one of its disjuncts has been derived). A is unreliable at some stage of a proof, iff there is a Δ such that $A \in \Delta$ and $Dab(\Delta)$ is a *minimal Dab*-formula that is unconditionally derived in the proof at that stage.[7] Here is a further illustration:

18	$\sim(\forall x)(Px \supset Sx) \vee \sim(\forall x)(Px \supset \sim Sx)$	5	RU	$\varnothing$
19	$\sim(\forall x)(Px \supset \sim Sx)$	5, 8	RU	$\varnothing$

At stage 18 of the proof, $(\forall x)(Px \supset Sx)$ is unreliable, and hence line 10 is marked. However, at stage 19, $(\forall x)(Px \supset Sx)$ is again reliable – 19 is a minimal *Dab*-formula at this stage, whereas 18 is not – and hence line 10 is unmarked.[8] This nicely illustrates both sides of the dynamics: formulas considered as derived at one stage may have to be considered as not derived at a later stage, and *vice versa.* All this may sound unfamiliar, or even weird. And yet, as we shall see in subsequent sections, everything is under control: ultimately the dynamics is bound to lead to stability, the stable situation is determined only by the premises (as the semantics illustrates), and there are heuristic means to speed up our journey towards the stable situation.

[7] The reason for considering only unconditionally derived formulas is straightforward. Indeed, from 17 one may derive $\sim(\forall x)(Px \supset \sim Tx)$ on the condition $\{(\forall x)(Px \supset Tx)\}$, but this obviously does not render $(\forall x)(Px \supset Tx)$ reliable. The disjunction 17 follows from the premises by **CL**, and neither of its disjuncts does.

[8] So line 10 is L-marked at stage 18 but not at stage 19.

Having made this precise – formal definitions follow in Section 4 – I leave it to the reader to check that the introduction of 'wild' hypotheses leads nowhere. As the predicates U and V do not occur in the premises 1-4, applying RC to add formulas such as $(\forall x)(Ux \supset Vx)$ to the proof, will lead to lines that are bound to be marked sooner or later – and very soon if some simple heuristic instructions are followed.

Before moving on to background knowledge, let me add some important comments. We have seen that $(\forall x)(Px \supset Qx)$ was not inductively derivable from 1-4. However, $(\forall x)((Px \wedge Rx) \supset Qx)$ is. Indeed, line 20 below is not marked in the present proof. In some artificial and clumsy extensions of the proof, line 20 may be marked. But it is easy enough to further extend the proof in such a way that line 20 is unmarked. This is an extremely important remark to which I return later.

20	$(\forall x)((Px \wedge Rx) \supset Qx)$	RC	$\{(\forall x)((Px \wedge Rx) \supset Qx)\}$

The next comment concerns the *form* of formulas derived by RC. All that was specified before is that these formulas should be universally quantified. However, a further restriction is required. Suppose that it is allowed to add

[21]	$(\forall x)((Qx \vee {\sim}Qx) \supset {\sim}Sc)$	RC	$\{(\forall x)((Qx \vee {\sim}Qx) \supset {\sim}Sc)\}$

to the proof. As

$${\sim}(\forall x)(Px \supset Sx) \vee {\sim}(\forall x)((Qx \vee {\sim}Qx) \supset {\sim}Sc)$$

is derivable from 1, not only line [21] but also line 10 would be marked in view of this formula. Similar troubles arise if it is allowed to introduce such hypotheses as $(\forall x)((Qx \vee {\sim}Qx) \supset (\exists x)(Px \wedge {\sim}Sx))$.

The way out of such troubles is simple enough. RC should not allow one to introduce singular statements or existentially quantified statements in disguise. Hence, we shall require that the generalizations introduced by RC consist of a sequence of universal quantifiers followed by a formula of the form $A \supset B$ in which no constants, propositional letters or quantifiers occur. From now on, '*generalization*' will refer to such formulas only.[9] Some people will raise a his-

[9] It is possible to devise a formal language in which the effect of this restriction is reduced to nil. This is immaterial if such languages are not used in the empirical sciences to which we want to apply **LI**. But indeed, the formal restriction hides one on content: all predicates should be well entrenched, and not abbreviate identity to an individual constant.

torical objection to this restriction. Kepler's laws explicitly refer to the sun, and Galileo's law of the free fall to the earth. This, however, is related to the fact that the earth, the sun, and the moon had a specific status in the Ptolemaic worldview, and were slowly losing that status in the days of Kepler and Galileo. In the Ptolemaic worldview, each of those three objects was taken, just like God, to be the only object of a specific kind. So those generalizations refer to kinds of objects, rather than to specific objects – by Newton's time, any possible doubt about this had been removed.[10]

Any *generalization* may be introduced by RC. This includes such formulas as 21 and 22, that are **CL**-equivalent to 23 and 24 respectively. So the implicative form of generalizations may be circumvented.

21	$(\forall x)((Qx \vee {\sim}Qx) \supset Px)$		RC	$\{(\forall x)((Qx \vee {\sim}Qx) \supset Px)\}$
22	$(\forall x)(Rx \supset (Qx \wedge {\sim}Qx))$		RC	$\{(\forall x)(Rx \supset (Qx \wedge {\sim}Qx))\}$
23	$(\forall x)Px$	21	RU	$\{(\forall x)((Qx \vee {\sim}Qx) \supset Px)\}$
24	$(\forall x){\sim}Rx$	22	RU	$\{(\forall x)(Rx \supset (Qx \wedge {\sim}Qx))\}$

Is the dynamics of the proofs bound to stop at some finite point? The answer to this question is not simple, but nevertheless satisfactory. However, I postpone the related discussion until we have gained a better grasp of **LI**.

Let us now move on to situations in which *background knowledge* is available. Clearly, background knowledge cannot be considered as unquestionable. For one thing, the empirical data might contradict it. If they do, we face an inconsistent set of premises, which leaves us nowhere on the present approach.[11] So we shall consider background knowledge as defeasible. It is taken for granted *unless and until* it is shown to be problematic.

This being settled, it is simple enough to integrate background knowledge in the dynamic proof format. Background knowledge is the result of inductive inferences made in the past, by ourselves our by our predecessors.[12] For this

[10] Even in the Ptolemaic era, those objects were identified in terms of well entrenched properties – properties relating to their kind, not to accidental qualities. The non-physical example is even more clear: God has *no* accidental properties.

[11] Scientists may justifiably stick to hypothetical knowledge that is falsified by the empirical data, for example because no non-falsified theory is available. Including such cases in the picture requires that we move to a paraconsistent logic. Although I have worked in this domain for some thirty years now, I fear that I would lose most of the readers if I were to open that Pandora's box. So let me stress that there is absolutely no problem in including the paraconsistent case in the logic of induction, but that I leave it out for reasons of space as well as for pedagogical reasons.

[12] Or rather, background knowledge is so interpreted for present purposes. This is a simpli-

reason, I shall restrict background knowledge to background generalizations – another simplification – and introduce them as *conditional premises.* Here is an example:

25	$(\forall x)(Qx \supset Rx)$		BK	$\{(\forall x)(Qx \supset Rx)\}$
26	Ra	7, 25	RU	$\{(\forall x)(Qx \supset Rx)\}$

The central difference between background generalizations and other generalizations – the latter will be called *local generalizations* from now on – is that the former are retained whenever possible. If $Dab(\Delta)$ is unconditionally derived, and each member of Δ is a background generalization, then, in the absence of further information, we have to consider all members of Δ as unreliable. So we shall mark all lines the condition of which overlaps with Δ. This includes the lines on which the background generalizations are introduced as conditional premises.[13]

If, however, we unconditionally derive $\sim A_1 \vee ... \vee \sim A_n \vee \sim B_1 \vee ... \vee \sim B_m$, and each A_i is a reliable background generalization (in the sense of the previous paragraph), then we should consider the local generalizations $B_1, ..., B_m$ as unreliable, and retain the background knowledge $A_1, ..., A_n$. Here is a simple example:

27^{L29}	$(\forall x)(Qx \supset \sim Rx)$		RC	$\{(\forall x)(Qx \supset \sim Rx)\}$
28^{L29}	$\sim Ra$	7, 27	RC	$\{(\forall x)(Qx \supset \sim Rx)\}$
29	$\sim(\forall x)(Qx \supset Rx) \vee \sim(\forall x)(Qx \supset \sim Rx)$	7	RU	$\varnothing$

$(\forall x)(Qx \supset Rx)$ is a background generalization and has not been shown to be an unreliable background generalization.[14] But the local generalization $(\forall x)(Qx \supset \sim Rx)$ is unreliable in view of 29. Hence, lines 27 and 28 are marked, but lines 25 and 26 are not, as desired.

In view of the asymmetry between background hypotheses and local hypotheses, **LI** is a *prioritized* adaptive logic. This means that the members of one set of defeasible formulas, the background hypotheses, are retained at the expense of the members of another set, the local generalizations.

fication. Humanity did not start from scratch at some point in time, not even with respect to scientific theories – see also Section 7.

[13] Here is a simple example. If $(\forall x)(Px \supset Qx)$ is a background generalization, one may introduce it by the rule BK. However, this background generalization (and anything derived from it) would be B-marked in view of line 14. See the next section for the precise definition.

[14] This agrees with the above criterion: there is no set of background generalizations Δ such that $(\forall x)(Qx \supset Rx) \in \Delta$ and $Dab(\Delta)$ is a minimal *Dab*-formula at stage 29 of the proof.

Before moving on to the precise formulation of the dynamic proof theory, let me intuitively explain some peculiarities of the proof format. Traditionally, a *proof* is seen as a list of formulas. This is not different for **LI**-proofs: the line numbers, the justification of the line (a set of line numbers and a rule), the conditions, and the marks are all introduced to make the proof more readable, but are not part of the proof itself. However, there is a central difference in this connection. In the dynamic case, one writes down a list of formulas, but the proof consists only of the *unmarked* formulas in the list. This does not make the marks part of the proof itself: which formulas are marked is determined by the empirical data, the background generalizations, and the list of formulas written down. Let us now continue to speak in terms of the annotated proof format.

What we are interested in are formulas that are *finally* derivable. On our way toward them, we have to go through the stages of a proof. Some formulas derived *at a stage* may not be finally derivable. As formulas that come with an empty condition (fifth element of the line) cannot possibly be marked at a later stage, they are sometimes called *unconditionally* derived. These formulas are deductively derived (by **CL**) from the empirical data. Formulas that have a non-empty condition are called *conditionally* derived. These formulas are inductively derived *only*. Of course, the interesting formulas are those that are inductively derived only, but nevertheless finally derived. In the present paper I offer a correct definition of final derivability, but cannot study the criteria that are useful from a computational point of view.

A last comment concerns the rules of inference. The *unconditional rules* of **LI** are those of Classical Logic, and they carry the conditions from their premises to their conclusion. The *conditional rules* BK and RC add a *new* element to the condition, and hence start off the dynamics of the proofs. As far as their structure is concerned, however, they are of the same type as the standard premise and axiom rules.

4. The dynamic proof theory

Our language will be that of predicative logic. Let $\forall A$ denote A preceded by a universal quantifier over any variable free in A. A *generalization* is a formula of the form $\forall(A \supset B)$ in which no individual constant, sentential letter or quantifier occurs in either A or B.

A dynamic proof theory consists of (i) a set of unconditional rules, (ii) a set of conditional rules, and (iii) a definition of *marked* lines. The rules allow one to add lines to a proof. Formulas derived on a line that is marked at a stage of the proof are considered as not inductively derived at that stage (from the premises and background generalizations).[15]

Lines in an annotated dynamic proof have five elements: (i) a line number, (ii) the formula derived, (iii) a set of line numbers (of the lines from which the formula is derived), (iv) a rule (by which the formula is derived), and (v) a set of conditions.

The logic **LI** operates on ordered sets of premises, $\Sigma = \langle \Gamma, \Gamma^* \rangle$, in which Γ is a set of *singular formulas* (the empirical data) and Γ^* is a set of *generalizations* (the background generalizations).

The rules of **LI** will be presented here in generic format. There are two unconditional rules, PREM and RU, and two conditional rules, BK and RC:

PREM — If $A \in \Gamma$, one may add a line comprising the following elements: (i) an appropriate line number, (ii) A, (iii) –, (iv) PREM, and (v) $\varnothing$.

BK — If $A \in \Gamma^*$, one may add a line comprising the following elements: (i) an appropriate line number, (ii) A, (iii) –, (iv) BK, and (v) $\{A\}$.

RU — If $A_1, ..., A_n \vdash_{\mathbf{CL}} B$ and each of $A_1, ..., A_n$ occur in the proof on lines $i_1, ..., i_n$ that have conditions $\Delta_1, ..., \Delta_n$ respectively, one may add a line comprising the following elements: (i) an appropriate line number, (ii) B, (iii) $\mathrm{i}_1, ..., i_n$, (iv) RU, and (v) $\Delta_1 \cup ... \cup \Delta_n$.

RC — Where A is a *generalization*, one may add a line comprising the following elements: (i) an appropriate line number, (ii) A, (iii) –, (iv) RC, and (v) $\{A\}$.

A proof constructed by these rules will be called an **LI**-proof from Σ. In such a proof, a formula is *unconditionally derived* iff it is derived at a line of which the fifth element is empty. It is conditionally derived otherwise.

An *abnormality* is the negation of a generalization. *Dab*-formulas are formulas of the form $Dab(\Delta) = \bigvee\{\sim A \mid A \in \Delta\}$, in which Δ is a finite set of generalizations.[16] $Dab(\Delta)$ is a *minimal Dab*-formula at stage s of a proof iff $Dab(\Delta)$

[15] That background generalizations may be B-marked themselves illustrates that they are defeasible premises.

[16] Note that $Dab(\Delta)$ refers to any formula that belongs to an equivalence class that is closed under permutation and association.

is unconditionally derived in the proof at stage s and there is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ is unconditionally derived in the proof at that stage.

DEFINITION 1

Where $Dab(\Delta_1)$, ..., $Dab(\Delta_n)$ *are the minimal Dab-formulas at stage s of a proof from* $\Sigma = \langle \Gamma, \Gamma^* \rangle$, $U^*_s(\Gamma) = \bigcup\{\Delta_i \subseteq \Gamma^* \mid 1 \leq i \leq n\}$.

DEFINITION 2

Where Δ *is the fifth element of line i, line i is B-marked iff* $\Delta \cap U^*_s(\Gamma) \neq \varnothing$.

$U^*_s(\Gamma)$ comprises the background generalizations that are *unreliable* at stage s of the proof. As lines that depend on unreliable background generalizations are B-marked, these generalizations are themselves removed from the proof. This is interpreted by not considering them as part of the background knowledge at that stage of the proof. What remains of the background knowledge at stage s will be denoted by $\Gamma^*_s = \Gamma^* - U^*_s(\Gamma)$.

Now we come to an important point. In order to determine which local generalizations are unreliable, we have to take the reliable background knowledge for granted. A *Dab*-formula *Dab* will be called a *minimal local Dab-formula* iff no formula $Dab(\Delta')$ occurs in the proof such that $(\Delta' - \Gamma^*_s) \subset (\Delta - \Gamma^*_s)$.

DEFINITION 3

Where $Dab(\Delta_1)$, ..., $Dab(\Delta_n)$ *are the minimal local Dab-formulas at stage s of a proof from* $\Sigma = \langle \Gamma, \Gamma^* \rangle$, $U^o_s(\Gamma) = \bigcup\{\Delta_\mathrm{i} - \Gamma^*_s \mid 1 \leq i \leq n\}$.

DEFINITION 4

Where is the fifth element of a line i that is not B-marked, line i is L-marked iff $\Delta \cap U^o_s(\Gamma) \neq \varnothing$.

$U^o_s(\Gamma)$ comprises the unreliable local generalizations at stage s. These generalizations may have been introduced by RC, they may be unreliable background generalizations, or they may be generalizations that do not occur in the proof (or occur as derived formulas only). Let me briefly clarify Definition 3.

Given the B-marks, we have to assess the hypotheses introduced by RC. Which of these are unreliable at stage s of the proof? The key to the answer to this question lies in the following theorem, the proof of which is obvious:

THEOREM 1

$Dab(\Delta \cup \Delta')$ is a minimal Dab-formula at stage s of a proof, iff a line may be added that has $Dab(\Delta)$ as its second, RC as its fourth, and Δ' as its fifth element.

Suppose that, in a proof at stage s, Δ' contains only reliable background generalizations, whereas no such background generalization is a member of Δ – that is, $\Delta' \subseteq \Gamma^*_s$ and $\Delta \cap \Gamma^*_s = \varnothing$. That $Dab(\Delta)$ is derivable on the condition Δ' indicates that some member of Δ is unreliable *if* the background generalizations in Δ' are reliable. Moreover, we consider the background generalizations to be more trustworthy than the local generalizations. So from the occurrence of the minimal local *Dab*-consequence $Dab(\Delta \cup \Delta')$ we should conclude that the members of Δ are unreliable.

Incidentally, an equivalent (and also very intuitive) proof theory is obtained by defining $U^o_s(\Gamma)$ in a different way. Let $Dab(\Delta_1), ..., Dab(\Delta_n)$ be the minimal (in the usual, simple sense) *Dab*-formulas that have been derived at stage s on the conditions $\Theta_1, ..., \Theta_n$ respectively, and for which $(\Delta_1 \cup ... \cup \Delta_n) \cap \Gamma^*_s = \varnothing$ and $\Theta_1 \cup ... \cup \Theta_n \subseteq \Gamma^*_s$. $U^o_s(\Gamma)$ may then be defined as $\Delta_1 \cup ... \cup \Delta_n$.[17] But let us stick to Definition 3 in the sequel.

DEFINITION 5

A formula A is derived at stage s of a proof from Σ, iff A is the second element of a non-marked line at stage s.

DEFINITION 6

$\Sigma \vdash_{\mathbf{LI}} A$ *(A is finally* **LI***-derivable from Σ) iff A is derived at a stage s of a proof from Σ, say at line i, and, whenever line i is marked in an extension of the proof, it is unmarked in a further extension of the proof.*

This definition is the same as for other dynamic proof theories. The following theorem is helpful to get a grasp of **LI**-proofs. The formulation is somewhat clumsy because the line may be marked, in which case A cannot be said to be derivable.

[17] This alternative definition need not lead to the same results with respect to a specific proof at a stage, but it determines the same set of finally derivable consequences (see below in the text) in view of Theorem 1.

THEOREM 2

To an **LI***-proof from* $\Sigma = \langle \Gamma, \Gamma^* \rangle$ *a (marked or unmarked) line may be added that has* A *as its second element and* Δ *as its fifth element, iff* $\Gamma \vdash_{\mathbf{CL}} A \vee Dab(\Delta)$.

The proof of the theorem is extremely simple. Let the **CL**-transform of an **LI**-proof from $\Sigma = \langle \Gamma, \Gamma^* \rangle$ be obtained by replacing any line that has B as its second and Θ as its fifth element, by an unconditional line that has $B \vee Dab(\Theta)$ as its second element. To see that this **CL**-transform is a **CL**-proof from Γ, it is sufficient to note the following: (i) the **CL**-transform of applications of PREM are justified by PREM, (ii) the **CL**-transform of applications of BK and RC are justified in that they contain a **CL**-theorem of the form $A \vee {\sim}A$, and (iii) the **CL**-transform of applications of RU are turned into applications of the **CL**-derivable (generic) rule "If $A_1, ..., A_n \vdash_{\mathbf{CL}} B$, then from $A_1 \vee C_1, ..., A_n \vee C_n$ to derive $B \vee C_1 \vee ... \vee C_n$". This establishes one direction of the theorem. The proof of the other direction is immediate in view of the **LI**-derivable rule: "Where all members of Δ are generalizations, to derive A on the condition Δ from $A \vee Dab(\Delta)$".

So, in a sense, **LI**-proofs are **CL**-proofs in disguise. We interpret them in a specific way in order to decide which generalizations should be selected.

In order to obtain a better grasp of final derivability, I first define the sets of unreliable formulas with respect to Γ, independently of the stage of a proof. First we need: $Dab(\Delta)$ is a *minimal Dab-consequence* of Γ iff $\Gamma \vdash_{\mathbf{CL}} Dab(\Delta)$ and, for no $\Delta' \subset \Delta$, $\Gamma \vdash_{\mathbf{CL}} Dab(\Delta')$.

DEFINITION 7

Where Ω^* *is the set of all minimal Dab-consequences of* Γ *in which occur only members of* Γ^*, $U^*(\Gamma) = \bigcup(\Omega^*)$.

This defines the set of background generalizations that are unreliable with respect to the empirical data Γ. The set of *retained* background generalizations is $\Gamma^*_\Gamma = \Gamma^* - U^*(\Gamma)$.

$Dab(\Delta)$ is a *minimal local Dab-consequence* of Γ iff $\Gamma \vdash_{\mathbf{CL}} Dab(\Delta)$ and, for no Δ', $\Gamma \vdash_{\mathbf{CL}} Dab(\Delta')$ and $(\Delta' - \Gamma^*_\Gamma) \subset (\Delta - \Gamma^*_\Gamma)$.[18]

DEFINITION 8

Where Ω *is the set of all minimal local Dab-consequences of* Γ, $U^\circ(\Gamma) = \bigcup(\Omega) - \Gamma^*_\Gamma$.

[18] As $\Gamma \vdash_{\mathbf{CL}} Dab(\Delta)$, $\Delta \neq \varnothing$ and, in view of Definition 7, $\Delta' - \Gamma^*_\Gamma \neq \varnothing$; similarly for Δ'.

This defines the set of local generalizations that are unreliable with respect to the empirical data Γ.

Given that **LI**-proofs are **CL**-proofs in disguise, the proofs of the following theorems can safely be left to the reader:

THEOREM 3

Where $\Sigma = \langle\Gamma, \Gamma^*\rangle$, $\Gamma^*_\Gamma = \{A \in \Gamma^* \mid \Sigma \vdash_{\mathbf{LI}} A\}$.

THEOREM 4

Where $\Sigma = \langle\Gamma, \Gamma^*\rangle$, $\Sigma \vdash_{\mathbf{LI}} A$, *A is finally* **LI**-*derivable from* Σ, *iff there is a* (*possibly empty*) Δ *such that* (*i*) $\Gamma \cup \Gamma^*_\Gamma \vdash_{\mathbf{CL}} A \vee Dab(\Delta)$, *and* (*ii*) $(\Delta - \Gamma^*_\Gamma) \cap U^\circ(\Gamma) = \varnothing$.

This sounds much simpler in words. A is an **LI**-consequence of Σ iff A is **CL**-derivable from Γ together with the reliable background generalizations, or, for some set Δ of reliable local generalizations,[19] $A \vee Dab(\Delta)$ is **CL**-derivable from Γ together with the reliable background generalizations.

The **LI**-consequence relation may be characterized in terms of compatibility – where is compatible with Δ' iff $\Delta \cup \Delta'$ is consistent (iff no inconsistency is **CL**-derivable from this set).[20] The characterization is remarkably simple, as appears from the following three theorems. The proof of the theorems is obvious in view of Definition 7 and Theorem 4.

THEOREM 5

$A \in \Gamma^*_\Gamma$ *iff* $A \in \Gamma^*$ *and* $\Delta \cup \{A\}$ *is compatible with* Γ *whenever* $\Delta \subseteq \Gamma^*$ *is compatible with* Γ.

A background generalization A is retained iff, whenever a set of background generalizations is compatible with the data, then A and Δ are jointly compatible with the data.

THEOREM 6

Where $\Sigma = \langle\Gamma, \Gamma^*\rangle$ *and A is a generalization*, $\Sigma \vdash_{\mathbf{LI}} A$ *iff* $\Delta \cup \{A\}$ *is compatible with* $\Gamma \cup \Gamma^*_\Gamma$, *whenever a set of generalizations* Δ *is compatible with* $\Gamma \cup \Gamma^*_\Gamma$.

[19] Remark that $\Gamma \cup \Gamma^*_\Gamma \vdash_{\mathbf{CL}} A \vee Dab(\Delta)$ iff $\Gamma \cup \Gamma^*_\Gamma \vdash_{\mathbf{CL}} A \vee Dab(\Delta - \Gamma^*_\Gamma)$.

[20] This definition presupposes that nothing is compatible with an inconsistent set – see also Batens and Meheus (2000), for an alternative.

A generalization A is inductively derivable iff, whenever a set Δ of generalizations is compatible with the data and retained background generalizations, then A and Δ are jointly compatible with the data and retained background generalizations. Let Σ^G be the set of generalizations that are inductively derivable from Σ.

THEOREM 7
Where $\Sigma = \langle \Gamma, \Gamma^* \rangle$, $\Sigma \vdash_{\mathbf{LI}} A$ *iff* $\Gamma \cup \Gamma^*_\Gamma \cup \Sigma^G \vdash_{\mathbf{CL}} A$.

A is inductively derivable from a set of data and background generalizations iff it is **CL**-derivable from the data, the reliable background generalizations, and the inductively derivable generalizations.

Let me finally mention, without proofs, some properties of the **LI**-consequence relation: Non-Monotonicity, Proof Invariance (any two proofs from Γ define the same set of final consequences), **CL**-Closure ($Cn_{\mathbf{CL}}(Cn_{\mathbf{I}}(\Sigma)) = Cn_{\mathbf{I}}(\Sigma)$),[21] Decidability of $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} A$ whenever Γ and Γ^* are finite and A is either a generalization or a singular formula. Cautious cut with respect to facts: where A is a singular statement, if $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} A$ and $\langle \Gamma \cup \{A\}, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$, then $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$. Cautious monotonicity with respect to facts: where A is a singular statement, if $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} A$ and $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$, then $\langle \Gamma \cup \{A\}, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$. By the last two: that inductively derivable predictions are verified, does not lead to new inductive consequences. Cautious cut with respect to generalizations: where A is a generalization, if $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} A$ and $\langle \Gamma, \Gamma^* \cup \{A\} \rangle \vdash_{\mathbf{LI}} B$, then $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$. Cautious monotonicity with respect to generalizations: where A is a generalization, if $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} A$ and $\langle \Gamma, \Gamma^* \rangle \vdash_{\mathbf{LI}} B$, then $\langle \Gamma, \Gamma^* \cup \{A\} \rangle \vdash_{\mathbf{LI}} B$. By the last two: if inductively derivable generalizations are accepted as background knowledge, no new inductive consequences follow.

5. The semantics

The previous sections merely considered the dynamic proof theory of **LI**. This proof theory is extremely important, as it enables us to explicate actual inductive reasoning – humans reach conclusions by finite sequences of steps. A logical semantics serves different purposes. Among other things, it provides insights into the conceptual machinery. Such insights increase our understanding of a logic, even if they are not directly relevant for the computational aspects.

Let $\mathcal{M}_\Gamma$ denote the set of **CL**-models of Γ. The **LI**-models of $\Sigma = \langle \Gamma, \Gamma^* \rangle$,

[21] $\mathrm{C}n_{\mathbf{L}}(\Gamma) = \{A \mid \Gamma \vdash_{\mathbf{L}} A\}$ as usual.

will be a subset of $\mathcal{M}_\Gamma$. This subset is defined in terms of the abnormal parts of models – see Batens (1986) for the first application of this idea (to a completely different kind of logic). The abnormal part of a model (the set of abnormalities verified by a model) is defined as follows. Let $\mathcal{G}$ denote the set of generalizations.

DEFINITION 9
$Ab(M) = \{\forall(A \supset B) \mid M \nvDash \forall(A \supset B); \forall(A \supset B) \in \mathcal{G}\}$.

In words: the abnormal part of a model is the set of generalizations it falsifies. Obviously, $Ab(M)$ is not empty for any model M. For example, either $(\forall x)((Px \vee {\sim}Px) \supset Qx) \in Ab(M)$ or $(\forall x)((Px \vee {\sim}Px) \supset {\sim}Qx) \in Ab(M)$. And if $M \vDash Pa$, then either $(\forall x)(Px \supset Qx) \in Ab(M)$ or $(\forall x)(Px \supset {\sim}Qx) \in Ab(M)$. However, in some models of Pa both $(\forall x)(Px \supset Qx)$ and $(\forall x)(Px \supset {\sim}Qx)$ belong to $Ab(M)$, whereas in others only one of them does.

Given that **CL** is sound and complete with respect to its semantics, $Dab(\Delta)$ is a *minimal Dab-consequence* of Γ iff all $M \in \mathcal{M}_\Gamma$ verify $Dab(\Delta)$ and no $\Delta' \subset \Delta$ is such that all $M \in \mathcal{M}_\Gamma$ verify $Dab(\Delta')$.

This semantic characterization of the minimal *Dab*-consequences of Γ immediately provides a semantic characterization of $U^*(\Gamma)$, of $U^\circ(\Gamma)$, and of Γ^*_Γ. This is sufficient to make the first required selection. The proof of Theorem 8 is obvious.

DEFINITION 10
$M \in \mathcal{M}_\Gamma$ *is* background-reliable *iff* $(Ab(M) \cap \Gamma^*) \subseteq U^*(\Gamma)$.

THEOREM 8
$M \in \mathcal{M}_\Gamma$ is background-reliable iff $M \vDash \Gamma^*_\Gamma$.

In words, the retained background knowledge consists of the members of Γ^* that are verified by all background-reliable models of Γ. So a model M of Γ is background-reliable iff it verifies all reliable background generalizations. For any consistent Γ and for any set of background generalizations Γ^*, there are background-reliable models of Γ.[22] This is warranted by the compactness of **CL**: Γ is compatible with Γ^*_Γ iff it is compatible with any finite subset of Γ^*_Γ.

I now proceed to the second selection of models of Γ.

[22] In the worst case, all background generalizations are unreliable, and hence all models of Γ are background-reliable.

DEFINITION 11

$M \in \mathcal{M}_\Gamma$ *is* reliable (*is an* **LI**-*model of* Σ)[23] *iff* $Ab(M) \subseteq U^\circ(\Gamma)$.

Since, in view of Definitions 7 and 8, $U^*(\Gamma) = U^\circ(\Gamma) \cap \Gamma^*$, it follows that:

THEOREM 9

All reliable models of Σ *are background reliable.*

One should not be misled by this. $Ab(M) \subseteq U^\circ(\Gamma)$ only warrants $(Ab(M) \cap \Gamma^*) \subseteq U^*(\Gamma)$ because the definition of $U^\circ(\Gamma)$ refers to the definition of $U^*(\Gamma)$.

DEFINITION 12

Where $\Sigma = \langle \Gamma, \Gamma^* \rangle$, $\Sigma \vDash_{\mathbf{LI}} A$ *iff all reliable models of* Γ *verify* A.

THEOREM 10

$\Sigma \vdash_{\mathbf{LI}} A$ *iff* $\Sigma \vDash_{\mathbf{LI}} A$. (*Soundness and Completeness*)

The proof is longwinded, especially its right-left direction, but follows exactly the reasoning of the proofs of Theorems 5.1 and 5.2 from Batens (1999). The present proof is simpler, however, as it concerns **CL**.

Some further provable properties: Strong Reassurance (if a **CL**-model M of Γ is not an **LI**-model of Σ, then some **LI**-model M' of Σ is such that $Ab(M') \subset Ab(M)$), and Determinism of final derivability (the co-extensive semantic consequence relation defines a unique consequence set for any Σ).

Although it is important to semantically characterize final **LI**-derivability in terms of a set of models of Σ, some might complain that the dynamics of the proofs does not appear in the semantics. However, there is a simple method to obtain a dynamic semantics for adaptive logics. This method, exemplified in Batens (1995), offers a dynamic semantics that is characteristic for derivability at a stage.

A slightly different (and richer) result would be obtained by applying the Minimal Abnormality strategy. I skip technicalities and merely mention the central difference from the Reliability strategy. In the presence of an instance[24] of Px and in the absence of instances of both $Px \wedge Qx$ and $Px \wedge {\sim}Qx$, the Reliability strategy leads to the rejection of both $(\forall x)(Px \supset Qx)$ and $(\forall x)(Px \supset {\sim}Qx)$ – if any

[23] As **LI** is an adaptive logic, it does not make sense to say that M is or is not an **LI**-model, but only that M is or is not an **LI**-model of some Σ.

[24] An instance of the open formula A, is any closed formula obtained by replacing each variable free in A by some constant.

of these formulas occurs in the fifth element of a line, the line is marked. It follows that even the disjunction of both generalizations will be marked. On the Minimal Abnormality strategy, both generalizations are marked but their disjunction is not. This supplementary consequence seems weak and pointless. Moreover, the Minimal Abnormality strategy, while leading to a very simple semantics, terribly complicates the proof theory. For this reason I shall not discuss it further here.

6. Heuristic matters and further comments

Some people think that all adaptive reasoning (including all non-monotonic reasoning) should be explicated in terms of heuristic moves rather than in terms of logic proper. For their instruction and confusion, I shall first spell out some basics of the heuristics of the adaptive logic **LI**. I leave it to the reader to compare both conceptions.

Suppose that one applies RC to introduce, on line i, the generalization $\forall(A \supset B)$ on the condition $\{\forall(A \supset B)\}$. As (1) is a **CL**-theorem, it may be derived in the proof and causes $\forall(A \supset B)$ to be L-marked.

$$\sim\forall(A \supset B) \vee \sim\forall(A \supset \sim B) \vee \sim\forall(\sim A \supset B) \vee \sim\forall(\sim A \supset \sim B) \qquad (1)$$

So, in order to prevent $\forall(A \supset B)$ from being L-marked, one needs to unconditionally derive

$$\sim\forall(A \supset \sim B) \vee \sim\forall(\sim A \supset B) \vee \sim\forall(\sim A \supset \sim B)$$

or a "sub-disjunction" of it. How does one do so? An instance of A enables one to derive

$$\sim\forall(A \supset B) \vee \sim\forall(A \supset \sim B) \qquad (2)$$

whereas an instance of $\sim A$ enables one to derive

$$\sim\forall(\sim A \supset B) \vee \sim\forall(\sim A \supset \sim B) \qquad (3)$$

An instance of $A \wedge B$ enables one to derive

$$\sim\forall(A \supset \sim B) \qquad (4)$$

and so on.

In view of this, it is obvious how one should proceed. Suppose that one is interested in the relation between A and B. It does not make sense to introduce by RC, for example, the generalization $\forall(A \supset B)$, if falsifying instances (instances of $A \wedge {\sim}B$) are derivable – if there are, the generalization is marked and will remain marked forever. Moreover, in order to prevent $\forall(A \supset B)$ from becoming marked in view of (1) or (2), one needs a confirming[25] instance (an instance of $A \wedge B$) and one needs to derive (4) from it. So two aims have to be pursued: (i) search for instances of $A \wedge {\sim}B$ in order to make sure that one did not introduce a falsified generalization, and (ii) search for instances of $A \wedge B$ in order to make sure that the generalization is not marked.

To see that the matter is not circular, note that it does not make sense, with respect to (ii) from the previous paragraph, to derive, say $B(a)$ from $A(a)$ together with the generalization $\forall(A \supset B)$ itself. Indeed, $B(a)$ will then be derived on the condition $\{\forall(A \supset B)\}$. (4) is derivable from $B(a)$, but again only on the condition $\{\forall(A \supset B)\}$. The only *Dab*-formula that can be unconditionally derived from (4) on the condition $\{\forall(A \supset B)\}$ is (2) – compare Theorem 2. In view of this, the line at which $\forall(A \supset B)$ was introduced by RC will still be marked.

But suppose that $A(a)$ and $C(a)$ occur unconditionally in the proof and that the generalization $\forall(C \supset B)$ was introduced by RC. If we derive $B(a)$ from these, it will be derived on the condition $\{\forall(C \supset B)\}$. So we are not able to unconditionally derive (4) from $A(a)$ and $B(a)$. All we can unconditionally derive along this road is

$$\sim\forall(A \supset {\sim}B) \vee {\sim}\forall(C \supset B) \tag{5}$$

and, in view of this, both $\forall(C \supset B)$ and $B(a)$ will be marked.

The reader might find this weird. There may be unconditional instances of $C \wedge B$ in the proof, and hence ${\sim}\forall(C \supset {\sim}B)$ may be unconditionally derived. This seems to warrant that $\forall(C \supset B)$ is finally derived, but obviously it does not. If such unexpected dependencies between abnormalities obtain, are we not losing control? Nothing very complicated is actually going on here. Control is provided by the following simple and intuitive fact:

(†) If the introduction of a local generalization G entails a falsifying instance of another generalization $\forall(A \supset B)$, and no falsifying instance of the latter is derivable from the empirical data together with the re-

[25] Obviously, 'confirming' is meant here in the qualitative sense – see Kuipers (2000, Ch. 2).

liable background knowledge, then $\sim G \vee \sim\forall(A \supset \sim B)$ is unconditionally derivable.

What does all this teach us? If we introduce a generalization, we want to find out whether it is finally derived in view of the present data. In order to do so, we should look for falsifying as well as for confirming instances, and we should look for falsifying instances of other generalizations, as specified in (†).[26] These instances may be derived from the union of the empirical data, the reliable background generalizations, and the reliable local generalizations. There is a clear bootstrapping effect here. At the level of the local generalizations the effect is weak, in that wild generalizations will not be finally derivable. At the level of the background generalizations, the effect is very strong – it is only annihilated by falsifying instances. However, at the level of the local generalizations, the bootstrapping effect does *not* reduce to a form of circularity.

So in order to speed up our journey towards the stable situation we need to look for the instances mentioned in the previous paragraph. As this statement may easily be misunderstood let me clarify it. Let the generalization introduced by RC be $\forall(A \supset B)$. (i) We need to find a confirming instance – if there are none, the generalization is bound to be marked.[27] (ii) We need to search for falsifying instances of the generalization and for falsifying instances of other generalizations that are novel with respect to the empirical data and reliable background generalizations – if there are falsifying instances of either kind, the generalization is bound to be marked. As a result of the search for falsifying instances (of either kind), we may find more confirming instances as well as a number of undetermined cases – individual constants for which there is an instance of A but not of either B or $\sim B$. When new empirical data become available, objects about which we had no information, or only partial information, may turn out to be falsifying, and so may objects about which we can only derive *conditionally* that they are confirming. So, (iii) we need to collect further data, by observation and experiment. At this point, confirmation theory enters the picture. Although **LI** does not take into account the number of confirming instances, only well-established hypotheses will convincingly eliminate potential falsifiers. Incidentally, I tend to side with Popper in this re-

[26] It is unlikely that effects like the one described by (†) will be discovered if one does not handle induction in terms of a logic. I have never seen such effects mentioned in the literature on induction, and they certainly are not mentioned in Kuipers (2000).

[27] This should be qualified. If there are instances of $\sim A$ and none of A, then $\forall(A \supset B)$ may be derivable and unmarked because $\forall\sim A$ is so.

spect: what is important is not the number of confirming instances, but rather the strength of the tests to which a generalization has been subjected. Whether this concept may be explicated within the present qualitative framework is dubious.

Although the heuristics of **LI** depends on confirmation theory in the sense described above, **LI** in itself enables us to spell out quite interesting heuristic maxims. Given a set of empirical data and a set of background generalizations, it is clear how we should proceed. Most of what was said above relates to that. If the given data and background knowledge do not allow one to finally derive any generalization concerning the relation between A and B because there is insufficient information, **LI** clearly instructs one about the kind of data that should be gathered to change the situation. In this sense, **LI** *does* guide empirical research. This guidance may be considered somewhat unsophisticated, but it is the basic guidance, the one that points out the most urgent empirical research.

I now turn to a different kind of heuristic maxims. In order to speed up our journey towards stability with respect to *given* empirical data and background generalizations, it is essential to derive as soon as possible the minimal *Dab*-consequences of Γ and to derive as soon as possible the minimal local *Dab-consequences* of Γ. Some **LI**-derivable rules are extremely helpful in this respect, and are related to deriving inconsistencies – the techniques to do so are well-known from the **CL**-heuristics. I mention only two examples. Suppose that, in an **LI**-proof from Σ, A is unconditionally derived, and that $\sim A$ is derived on the condition Δ. Then $Dab(\Delta)$ is unconditionally derivable in the proof. Similarly, if an inconsistency is derived on the condition Δ, $Dab(\Delta)$ is unconditionally derivable in the proof.

An equally helpful derivable rule was exemplified before (and is warranted by Theorem 2). If a *Dab*-formula $Dab(\Delta)$ is derived on the condition Δ', then $Dab(\Delta \cup \Delta')$ is unconditionally derivable. Similarly, if an instance of A is derived on the condition Δ and an instance of B is derived on the condition Δ', then $\sim\forall(A \supset \sim B) \vee Dab(\Delta \cup \Delta')$ is unconditionally derivable – either or both of Δ and Δ' may be empty.

A very rough summary reads as follows: derive all singular statements that lead to instances of formulas no instances of which have been derived, and derive *Dab*-formulas that change either the minimal *Dab*-formulas or the minimal local *Dab*-formulas. The first instruction requires the derivation of a few formulas only. The second may be guided by several considerations, (i) Whenever $Dab(\Delta)$ has been derived, one should try to unconditionally derive $Dab(\Delta')$ for all $\Delta' \subset \Delta$. This is a simple and decidable task. (ii) One should only try to derive $Dab(\Delta)$ when Δ consists of background generalizations, generalizations introduced by the rule RC, or "variants" of such generalizations – the variants of $\forall(A \supset B)$ being the four generalizations that occur in (1). This instruction may be further restricted. Given a background generalization or local generalization $\forall(A \supset B)$, one should first and foremost try to derive the $Dab(\Delta)$ for which Δ contains variants of $\forall(A \supset B)$. The only cases in which it pays to consider other *Dab*-formulas is the one described in (†).

Up to now I have considered the general heuristic maxims that apply to **LI**. However, **LI** has distinct application contexts, in which different aims are pursued and specific heuristic maxims apply. I shall consider only two very general application contexts.

If one tries to derive *Dab*-formulas that result in some lines being marked or unmarked, one basically checks whether the introduced generalizations are compatible with and confirmed by the available empirical data. However, one might also, after introducing a generalization, concentrate on its consequences by deriving singular statements from it. These singular statements will be derived conditionally. As said before, this may be taken to be a good reason to invoke observation and experiment in order to test them. This leaves room for a "Popperian" application of **LI**. Even if a generalization may be marked in view of derivable *Dab*-formulas, and even if it *is* marked in view of derived *Dab*-formulas, we may try to gather novel data that cause the generalization to be unmarked. Incidentally, the "stronger" generalizations in the sense of Popper (1935; 1963) are those from which a larger number of weaker generalizations are derivable, and hence have more potential falsifiers. Popper was quite right, too, to stress that it is advisable to infer the most general (the bolder) generalizations first. If they become marked, we may still retract to less general generalizations. As long as these are not marked, the less general generalizations are available for free because they are **CL**-consequences of the more general ones.

A distinction is useful in the present context. If an instance of Px is derivable from the empirical data together with the reliable background knowledge, but no instances of either $Px \wedge Qx$ or $Px \wedge {\sim}Qx$ are so derivable, then both $(\forall x)(Px \supset Qx)$ and $(\forall x)(Px \supset {\sim}Qx)$ may be marked *because we have no means to choose between them.* If instances of both $Px \wedge Qx$ and $Px \wedge {\sim}Qx$ are **CL**-derivable from the empirical data together with the reliable background knowledge, then both $(\forall x)(Px \supset Qx)$ and $(\forall x)(Px \supset {\sim}Qx)$ may be marked *because both are falsified.* The transition from the first situation to the second clearly indicates an increase in knowledge. Moreover, in the second situation it does not make sense to look for further confirming instances of either generalization. What does make sense in the second situation, and not in the first, is that one looks for less general hypotheses, for example $(\forall x)((Px \wedge Rx) \supset Qx)$ that may still be derivable.

This at once answers the objection that **LI** too severely restricts a scientist's freedom to launch hypotheses. **LI** does not in any way restrict the freedom to introduce generalizations. Rather, **LI** points out, if sensibly applied, which generalizations cannot be upheld, and which empirical research is desirable. A scientist's "freedom" to launch hypotheses is not a permission for dogmatism – to make a claim and stick to it. If it refers to anything, then it is to the freedom to break out of established conceptual schemes. Clearly, the introduction of new conceptual schemes goes far beyond the present simple logic of induction – I return to this in Section 7. Given the limits of **LI**, the set of **LI**-consequences of a given Σ should be determined by Σ and should be independent of any specific line of reasoning. In this respect the rule RC differs drastically from such rules as Hintikka's bracketing rule – see, for example, Hintikka (1999; forthcoming).

A very different application context concerns predictions derived in view of actions. It makes sense, in the Popperian context, to derive predictions from a generalization A, even before checking whether the proof can be extended in such a way that A is marked. In the present context, it does not. It would be foolish to act on the generalization $(\forall x)(Px \supset Qx)$ in the absence of confirming instances – such actions would be arbitrary. In action contexts, one should play the game in a safer way by introducing only well-confirmed generalizations, not bold ones. Thus $(\forall x)(Px \supset Qx)$ should be *derived* from safe generalizations, for example, $(\forall x)((Px \wedge Rx) \supset Qx)$ and $(\forall x)((Px \wedge {\sim}Rx) \supset Qx)$ if both of these happen to be safe.

In both contexts,[28] **LI** suggests a specific heuristic procedure. This procedure differs from one context to the other, and may be justified in view of the specific aims.

Some people may find it suspect that applications of the rule RC do not require the presence of any formulas in the proof. RC is a positing rule rather than a deduction rule. This is no reason to worry. **LI** has a dynamic proof theory. A proof at a stage should not be confused with a proof of a logic that has a (static) proof theory of the usual kind. The central question in an **LI**-proof is not whether a generalization can be introduced, but whether it can be retained – the aim is final derivability, not derivability at a stage. The preceding paragraphs make it sufficiently clear that final derivability is often difficult to reach, and that one needs

[28] This distinction between action contexts and contexts concerning theoretical inquiry was one of the points made in my (1968).

to follow a set of heuristic rules in order even to obtain a sensible estimate of final derivability – see also below. In this connection, it is instructive to see that $Cn_{\mathbf{I}}(\langle\varnothing, \varnothing\rangle) = Cn_{\mathbf{CL}}(\varnothing)$, that $Cn_{\mathbf{I}}(\langle\{Pa, {\sim}Pa\}, \varnothing\rangle) = Cn_{\mathbf{CL}}(\{Pa, {\sim}Pa\})$, and hence that neither of these comprises a non-tautological generalization.

A final comment concerns the nature of an adaptive logic. It would be foolish to build a logic that allows for some mistakes. Obviously, adaptive logics do not allow for mistakes: $Cn_{\mathbf{I}}(\Sigma)$ is a well-defined set that leaves no room for any choice or arbitrariness. The dynamic proof theory constitutes a way to search for $Cn_{\mathbf{I}}(\Sigma)$. A proof at a stage merely offers an estimate of $Cn_{\mathbf{I}}(\Sigma)$ – an estimate that is determined by the insights in the premises that are provided by the proof. We have seen that there are heuristic means to make these insights as rich and useful as possible. There also are criteria to decide, in some cases, whether a formula is finally derived in a proof – see Batens (2002). In the absence of a positive test, that is the best one can do in a computational respect.

For large fragments of the language, **LI**-derivability is decidable. This includes all generalizations, and hence all predictions and explanations. But even for undecidable fragments of the language, dynamic proofs at a stage offer a sensible estimate of $Cn_{\mathbf{I}}(\Sigma)$, the best estimate that is available from the proof – see Batens (1995). This means that an **LI**-proof at a stage is sufficient to take justified decisions: decisions that may be mistaken, but are justified in terms of our present best insights.

7. Further research

As announced, **LI** is very simple – only a starting point. In the present section I briefly point to some open problems. Some of these relate to alternatives for **LI**, others to desirable sophistication.

With respect to background generalizations, an interesting alternative approach is obtained by not introducing members of Γ^* but rather generalizations that belong to $Cn_{\mathbf{CL}}(\Gamma^*)$. Suppose that $(\forall x)(Px \supset Qx) \in \Gamma^*$, and that Pa, Ra and ${\sim}Qa$ are **CL**-consequences of Γ. According to **LI**, $(\forall x)(Px \supset Qx)$ is falsified, and hence not retained. According to the alternative, $(\forall x)((Px \wedge {\sim}Rx) \supset Qx)$ would, for all that has been said, be a retained background generalization. This certainly deserves further study, both from a technical point of view and with respect to application contexts.

LI is too empiricist, even too positivistic. Let me just mention some obvious sophistication that is clearly desirable. Sometimes our background knowledge is inconsistent and sometimes falsified generalizations are retained. As there is room for neither in **LI**, this logic is unfit to explicate certain episodes from the history of the sciences. It is not difficult to modify **LI** in such a way that both inconsistent background knowledge and the application of falsified generalizations are handled. Available (and published) results on inconsistency adaptive logics make this change a rather easy exercise.

Another weakness of **LI**, or rather of the way in which **LI** is presented in the present paper, is that there seems to be only room for theories in the simple sense of the term: sets of generalizations. This weakness concerns especially background theories – the design of new theories is not a simple inductive matter anyway. Several of the problems listed above are solved in Batens and Haesaert (2001); this paper contains also a variant of **LI** that follows the standard format for adaptive logics.

LI does not enable one to get a grasp of conceptual change or of similar phenomena that are often related to scientific creativity and discovery. This will be the hardest nut to crack. That it is not impossible to crack it will be obvious to readers of such papers as Meheus (1999a; 1999b; 2000).

Let me say no more about projected research. The basic result of the present paper is that there is now a logic of induction. It is simple, and even a bit old-fashioned, but it exists and may be applied in simple circumstances.[29]

Acknowledgments
Research for this paper was supported by subventions from Ghent University and from the Fund for Scientific Research – Flanders, and indirectly by the Flemish Minister responsible for Science and Technology (contract BIL98/37). I am indebted to Atocha Aliseda, Theo Kuipers, Dagmar Provijn, Ewout Vansteenkiste, and Liza Verhoeven for comments on previous drafts.

REFERENCES

BAR-HILLEL, Y. 1968: "The Acceptance Syndrome", in I. Lakatos (ed.), *The Problem of Inductive Logic*, Amsterdam: North-Holland, pp. 150-161.

BATENS, D. 1968: "Some Proposals for the Solution of the Carnap-Popper Discussion on Inductive Logic", *Studia Philosophica Gandensia*, 6, pp. 5-25.

——— 1986: "Dialectical Dynamics within Formal Logics", *Logique et Analyse*, 114, pp. 161-173.

[29] Unpublished papers by members of our research group are available from the Internet address http://logica.UGent.be/centrum/writings/.

——— 1995: "Blocks. The Clue to Dynamic Aspects of Logic", *Logique et Analyse*, 150-152, pp. 285-328.

——— 1999: "Inconsistency-Adaptive Logics", in E. Or³owska (ed.), *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*, Heidelberg/New York: Physica Verlag Springer, pp. 445-472.

——— 2000: "A Survey of Inconsistency-Adaptive Logics", in D. Batens, C. Mortensen, G. Priest, and J.P. Van Bendegem (eds.), *Frontiers of Paraconsistent Logic*, Baldock, UK: Research Studies Press, pp. 49-73.

——— 2002: "On a Partial Decision Method for Dynamic Proofs", in H. Decker, J. Villadsen, and T. Waragai (eds.), *PCL 2002. Paraconsistent Computational Logic* (*Datalogiske Skrifter*, vol. 95), pp. 91-108. Also available as cs.LO/0207090 at http://arxiv.org/ archive/cs/intro.html.

——— 2004: "Extending the Realm of Logic. The Adaptive-Logic Programme", in P. Weingartner (ed.), *Alternative Logics. Do Sciences Need Them?*, Berlin-Heidelberg: Springer Verlag, pp. 149-164.

BATENS, D., HAESAERT, L. 2001: "On Classical Adaptive Logics of Induction", *Logique et Analyse*, 173-175, pp. 255-290.

BATENS, D., MEHEUS, J. 2000: "The Adaptive Logic of Compatibility", *Studia Logica*, 66, pp. 327-348.

BOOLOS, G.S., JEFFREY, R.J. 1989: *Computability and Logic*, 3rd edition, Cambridge: Cambridge University Press.

BROWN, B. 1990: "How to Be Realistic about Inconsistency in Science", *Studies in History and Philosophy of Science*, 21, pp. 281-294.

HINTIKKA, J. 1999: *Inquiry as Inquiry: A Logic of Scientific Discovery*, Dordrecht: Kluwer.

——— (forthcoming), "Argumentation in a Multicultural Context".

KUIPERS, T. 2000: *From Instrumentalism to Constructive Realism*, Dordrecht: Kluwer.

MEHEUS, J. 1993: "Adaptive Logic in Scientific Discovery: The Case of Clausius", *Logique et Analyse*, 143-144, pp. 359-389.

——— 1999a: "Deductive and Ampliative Adaptive Logics as Tools in the Study of Creativity", *Foundations of Science*, 4, pp. 325-336.

——— 1999b: "Model-Based Reasoning in Creative Processes", in L. Magnani, N. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, Dordrecht: Kluwer/Plenum, pp. 199-217.

——— 2000: "Analogical Reasoning in Creative Problem Solving Processes: Logico-Philosophical Perspectives", in F. Hallyn (ed.), *Metaphor and Analogy in the Sciences*, Dordrecht: Kluwer, pp. 17-34.

——— 2002: "Inconsistencies in Scientific Discovery. Clausius's Remarkable Derivation of Carnot's Theorem", in H. Krach, G. Vanpaemel, and P. Marage (eds.), *History of Modern Physics*, Brepols, Brepols, pp. 143-154.

NERSESSIAN, N. 2002: "Inconsistency, Generic Modeling, and Conceptual Change in Science", in J. Meheus (ed.), *Inconsistency in Science*, Dordrecht: Kluwer, pp. 197-211.

NORTON, J. 1987: "The Logical Inconsistency of the Old Quantum Theory of Black Body Radiation", *Philosophy of Science*, 54, pp. 327-350.

——— 1993: "A Paradox in Newtonian Gravitation Theory", in *PSA 1992*, vol. II, pp. 421-420.

POPPER, K.R. 1935: *Logik der Forschung*, Wien: Verlag von Julius Springer.

——— 1959: *Logic of Scientific Discovery*, London: Hutchinson. English translation with new appendices of Popper 1935.

——— 1963: *Conjectures and Refutations*, London: Routledge & Keagan.

——— 1973: *Objective Knowledge*, Oxford: Clarendon.

SMITH, J. 1988: "Inconsistency and Scientific Reasoning", *Studies in History and Philosophy of Science*, 19, pp. 429-445.

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 33-38

A Brand New Type of Inductive Logic
Reply to Diderik Batens*

Theo A. F. Kuipers
Department of Theoretical Philosophy,
University of Groningen (The Netherlands)
e-mail: T.A.F.Kuipers@philos.rug.nl

The correspondence to which Diderik Batens refers dates from the autumn of 1971, and resulted in my very first publication in English, albeit a very short one (Kuipers 1972). Ever since, he has been for me one of the few role models as a philosopher trying to bridge the gap between logic and philosophy of science. Although he certainly is much more of a logician than I am, in many cases, as in the present one, he remains driven by questions stemming from philosophy of science. I am not the only Dutch speaking philosopher influenced by this role model. In Belgium, notably Ghent, he shaped the interests of Jean Paul Van Bendegem, Erik Weber, Helena de Preester and Joke Meheus, to mention only those who have contributed to one of the present two volumes. Certainly the great example in the Netherlands is Evert Willem Beth. Unfortunately I was too young to ever meet him. Although Beth exerted a powerful influence on a whole generation of Dutch philosophers, their emphasis was even more on (mathematical or philosophical) logic and, later, its computational and linguistic applications. Happily enough, Hans Mooij is one of the few exceptions. He was the first supervisor of my dissertation and has now contributed to the present volume. At one time, Johan van Benthem, Beth's indirect successor, seemed to become the great example from and for my own generation. However, after his review-like programmatic paper "The Logical Study of Science" (1982) on general philosophy of science, he, unfortunately for my field, directed his logical skills to other areas. But times seem to change, witness his contribution to the present volume.

Batens' contribution is a typical example of doing logic in the service of philosophy of science. Since his contribution is already an impressive logical sys-

* This paper appeared in R. Festa, A. Aliseda and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation* (*Poznań Studies in the Philosophy of the Sciences and the Humanities,* vol. 83), Amsterdam/New York, NY: Rodopi, 2005, pp. 248-252.

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

tem, it may be seen as the idealized point of departure for a really rich logic of induction and so I would like to focus my reply on some points that may be relevant for further concretization. However, before doing that, I would like to situate Batens' project in the realm of different approaches to inductive logic.

1. Kinds of inductive logic

It is interesting to see how Batens deviates from the old approaches to a logic of induction or an inductive logic. Basically, I mean the two approaches initiated by Carnap, the first being based on the idea of first assigning degrees of inductive probability to hypotheses, prior and posterior relative to the evidence, and then basing rules of inference on them that avoid paradoxes, notably the lottery paradox. Hintikka and Hilpinen made serious progress along these lines, although at the price of assigning non-zero prior probabilities to genuine generalizations. Carnap was not willing to pay this price, which makes him a dogmatic skeptic, to use Niiniluoto's (1999) apt phrase for this attitude. Be that as it may, Carnap made the decision-theoretic move by restricting the task of inductive logic to the probability assignments to be used in decisions, taking relevant utilities into account. As can be derived from Ch. 4 of my (2000), even this restricted program of inductive logic, despite its dogmatic skeptic nature, was certainly successful, internally and externally, falsifying Lakatos' premature claim that it was a degenerating program.

It is true that the general idea of an "inductive logic" has several other elaborations. Bayesian philosophy of science is sometimes described this way. As a matter of fact, its standard version can be seen as one of the three basic approaches in the second sense indicated above (see Kuipers 2005, Section 4; and more extensively 2001, Section 7.1.2), viz. the one rejecting dogmatic skepticism, that is, by taking "inductive priors" into account, but also rejecting "inductive (or adaptive) likelihoods". Carnap, in contrast, rejected inductive priors in favor of inductive likelihoods. Finally, Hintikka has chosen the "double inductive" approach, that is, inductive priors and inductive likelihoods. The common feature of these three approaches is that they aim at realizing the property of instantial confirmation or positive instantial relevance: another occurrence of a certain outcome increases its probability for the next trial.

Besides these (restricted or unrestricted) probabilistic approaches to inductive logic, there are a number of totally different approaches. Besides that of Batens, three of them should be mentioned, all of a computational nature. The first one is that of Thagard c.s. (Holland *et al.* 1986; Thagard 1988), leading to the computer program **PI** (Processes of Induction). The second operates under the heading of "inductive logic programming" (see Flach and Kakas 2000) and the third under "abductive logic programming" (see Kakas *et al.* 1998). Whereas the first is not so much logically inspired, but connectionistic, the other two typically are. Batens' approach is, at least so far, a purely logical one and hence is rightly called a "logic of induction". It is a specialization of his own adaptive version of dynamic logic aiming at deriving (inductive) generalizations of the type: for all *x*, if *Ax* then *Bx*.

2. Points for concretization

I shall not concentrate on technical matters regarding Batens' logic of induction. Although it is presented in a very transparent way by first giving a more informal description of the main means and ends, I do not want to suggest that I have grasped all the details. Incidentally, readers will find in Meheus' paper another nice entry into adaptive logic. Although Batens writes of modifications rather than concretizations, his contribution, like several others, nicely illustrates that not only the sciences but also philosophy can profit greatly from the idealization & concretization (I&C) strategy.[1] I shall concentrate on some points of concretization that are desirable from the point of view of philosophy of science.

A first point is the restriction to generalizations not referring to individual constants. In my opinion Batens defends this idealization in Section 3 too strongly by referring – as such correctly – to the history of the laws of Galileo and Kepler according to which the reference to the earth and the sun, respectively, disappeared in a way in light of Newton's theory (see also his Notes 9 and 10). Typically of inductive methods, rather than hypothetico-deductive ones, I would suggest that in particular in the heuristic phase of inductive research programs (see Kuipers 2000, Section 7.5.4) reference to individual objects seems very normal. Indeed, the work of Galileo and Kepler may well be seen from this perspective, whereas Newton indeed saw earth and sun merely as objects of a kind. Moreover, in many areas, e.g. in the humanities, many (qua-

[1] In Kuipers (forthcoming) I illustrate this conceptual version of I&C, as a variant of the empirical version, in terms of the theory of (confirmation, empirical progress, and) truth approximation presented in my (2000). In this illustration the two versions of I&C meet each other: revised truth approximation is a conceptual concretization of basic truth approximation, accounting for empirical concretization, e.g. the transition from the ideal gas law to the law of Van der Waals.

si-) generalizations seem only to make sense when linked to individuals. More precisely, dispositions of human beings are frequently bound to one individual. People may have more or less unique habits. Hence, a realistic logic of induction should be able to deal with generalizations that merely hold for individual objects. Happily enough, Batens claims, also in his Note 9, that it is at least possible to reduce the effect of the relevant restriction to zero.

A second possible concretization is leaving room for falsified background knowledge. In Note 11 Batens explains that it would be possible to do so by moving to paraconsistent logic. To be sure, Batens is the leading European scholar is this enterprise. Although his formulation might suggest otherwise, I am fairly sure that he does not want to suggest that this paraconsistent move requires a complete departure from the present adaptive dynamic approach. What is at stake here seems to be a matter of the order of concretization. The concretization to paraconsistent adaptive logic is a general concretization of that logic, not specifically related to inductive ends. Hence, the question that intrigues me is how important the concretization to paraconsistency is from my philosophy of science point of view. In this respect it is important to note first that I fully subscribe to Batens' first sentence of Note 11: "Scientists may justifiedly stick to hypothetical knowledge that is falsified by the empirical data, for example because no non-falsified theory is available" (p. 12). In a way, this sentence could be seen as the main point of departure of my (2000). However, my book develops an explication of this observation that, at least at first sight, completely differs from the paraconsistent move. In this respect it may be interesting to note that paraconsistent logic is still very much "truth/falsity" oriented, whereas my book is basically "empirical progress and truth approximation" oriented. (See Kuipers 2000, Ch. 1, for this distinction.) The strange thing, however, is that although "being falsified" of a theory becomes from my perspective a meaningful but non-dramatic event for a theory, the falsification of a hypothetical inductive generalization (or a first order observational induction, Kuipers 2000, p. 65) is a crucial event. Since the data at a certain moment (t) are composed of (partial) descriptions of realized possibilities $R(t)$ and inductive generalizations based on them, summarized by $S(t)$, a falsification of one of the latter means that the "correct data" assumption is no longer valid. In other words, we have to weaken $S(t)$ in a sufficient way, preferably such that it is just sufficient. Note that this not a concretization move. Note moreover, that it not only holds for the basic approach but also for the refined approach (Kuipers 2000, Ch. 10). To be sure, one may argue in particular that taking falsifications of $S(t)$ into account in some sophisticated way might further concretize the refined approach. However, I submit that scientists will be more inclined to adapt $S(t)$ as suggested. Hence, from my point of view, the concretization to paraconsistency is not particularly urgent or even relevant for the role of inductive generalizations in aiming at empirical progress and truth approximation. This attitude seems to be supported by Batens and Haestert (forthcoming) where they extend and improve upon Batens' present contribution. Of course, when genuinely inconsistent theories are at stake the paraconsistent move may become unavoidable.

Another possibility for concretization intrigues me very much. Batens argues at the beginning of Section 6 that it becomes relevant to search for confirming and falsifying instances of "for all x if $A(x)$ then $B(x)$" of the type $A(x)$ & $B(x)$ and, of course, $A(x)$ & *non-*$B(x)$, respectively. Although he refers in Note 25 to qualitative confirmation in the sense of Ch. 2 of my (2000), it remains unclear whether my analysis of kinds of non-falsifying instances in terms of two types of confirming instances ($A(x)$ & $B(x)$ and *non-*$A(x)$ & *non-*$B(x)$) and one type of neutral instances (*non-*$A(x)$ & $B(x)$) plays any role. More specifically, from that perspective one would expect, in line with general dynamic logic intuitions, that one starts either with A-cases, and finds out whether they are B or *non-B*, or with *non-B*-cases, and find out whether they are A or *non-A*. All this in order to avoid searching for neutral cases. If I am right that this selective search does not yet play a role, a concretization in this direction would certainly lead to a more realistic and more efficient logic.[2]

Let me conclude with a point that has nothing to do with concretization, but that puzzles me a lot. Although I think I can follow why (†) holds in the logic, I do not understand why it is a "simple and intuitive fact" (p. 24) of which it is "unlikely that [it] will be discovered if one does not handle induction in terms of logic" (p. 25, Note 26). The combination seems implausible, but knowing Batens, he must have something serious in mind.

REFERENCES

BATENS, D., HAESERT, L. (forthcoming): "On Classical Adaptive Logics of Induction", *Logique et Analyse*.

BENTHEM, J. van 1982: "The Logical Study of Science", *Synthese*, 51, pp. 431-472.

FLACH, P.A., KAKAS, A.C. (eds.) 2000: *Abduction and Induction*, Dordrecht: Kluwer.

[2] Unfortunately, I had difficulties in understanding precisely the core of the paragraph starting with "So in order to speed up our journey towards the stable situation..." (p. 25). Maybe this paragraph entails selective search.

HOLLAND, J., HOLYOAK, K., NISBETT, R., and THAGARD, P. 1986: *Induction. Processes of Inference, Learning and Discovery*, Cambridge MA: The MIT Press.

KAKAS, A., KOWALSKI R., and TONI, F. 1998: "The Role of Abduction in Logic Programming", in D. Gabbay, C. Hogger, and J. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, Oxford: Oxford University Press, vol. V, pp. 235-324.

KUIPERS, T. 1972: "A Note on Confirmation", *Philosophica Gandensia*, 10, pp. 76-77.

——— 2000: *From Instrumentalism to Constructive Realism*, Kluwer: Dordrecht.

——— 2001: *Structures in Science: Heuristic Patterns Based on Cognitive Structures. An Advanced Textbook in Neo-Classical Philosophy of Science*, Kluwer: Dordrecht.

——— 2005: "The Threefold Evaluation of Theories: A Synopsis of *From Instrumentalism to Constructive Realism. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation*", in R. Festa, A. Aliseda, and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation*, Amsterdam/New York, NY: Rodopi, pp. 21-85.

——— (forthcoming): "Empirical and Conceptual Idealization and Concretization. The Case of Truth Approximation", in English and Polish editions of *Liber Amicorum for Leszek Nowak*.

NIINILUOTO, I. 1999: *Critical Scientific Realism*, Oxford: Oxford University Press.

THAGARD, P. 1988: *Computational Philosophy of Science*, Cambridge MA: MIT Press.

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 39-89

Łukasiewicz e il determinismo logico

Alessandro Becchi
Dipartimento di filosofia, Università degli Studi di Firenze (Italy)
e-mail: notung@unifi.it

1. Introduzione
2. Logica bivalente e modalità aletiche
3. Il terzo valore logico e la definizione di possibilità
4. In che senso la logica trivalente è una logica delle modalità?
5. Terzo escluso, determinismo, causalità
6. Indeterminatezza ontologica e proposizioni possibili
7. Conseguenze filosofiche e problemi aperti
Appendice. Portatori di verità e determinismo logico: un approccio deflazionistico al problema

SOMMARIO. Questo lavoro ha lo scopo di illustrare le motivazioni filosofiche che furono all'origine dell'elaborazione del primo sistema di logica trivalente per il calcolo proposizionale ad opera di Łukasiewicz. In particolare, attraverso una analisi critica dei due saggi "Osservazioni filosofiche sui sistemi polivalenti della logica proposizionale" e "On Determinism", è mia intenzione mostrare come Łukasiewicz sia stato spinto a mettere in discussione il principio di bivalenza sulla base di due questioni che egli vedeva strettamente legate tra loro, ossia: (1) l'importo ontologico in senso deterministico del principio di bivalenza; (2) la possibilità di un approccio estensionale alla logica delle modalità. Oltre ad illustrare le due questioni in dettaglio e separatamente, cerco di mostrare come la seconda di esse possa essere correttamente compresa solo se inserita nel quadro della prima; in altri termini, che è possibile fornire una "lettura modale" adeguata (rispetto a certe nostre intuizioni) della logica trivalente solo se teniamo presente il problema originario dal quale essa ha avuto origine: quello del valore logico degli enunciati al futuro. Dopo una breve analisi critica degli argomenti avanzati da Łukasiewicz contro la bivalenza e il determinismo, concludo il lavoro proponendo un approccio deflazionistico al problema del determinismo logico, argomentando come una analisi attenta dei pos-

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

sibili portatori di verità renda possibile una "dissoluzione" del problema stesso.

PAROLE CHIAVE: bivalenza, logica trivalente, possibilità, logica modale, determinismo, causalità, terzo escluso, portatore di verità, proposizione, enunciato, indeterminatezza ontologica.

1. Introduzione

Nel presente lavoro è mia intenzione mettere in luce alcune questioni di carattere filosofico che furono all'origine dell'elaborazione del primo sistema trivalente per il calcolo proposizionale ad opera di Jan Łukasiewicz (si veda Łukasiewicz 1920), accennando inoltre ad alcune problematiche che a mio avviso restano aperte e proponendo (nell'appendice) un approccio "deflazionistico" al problema del determinismo logico. Le motivazioni che indussero il filosofo e logico polacco ad abbandonare il principio di bivalenza, nonché l'interpretazione intuitiva del sistema trivalente nato sulla base di tale abbandono, sono state ampiamente illustrate dallo stesso autore in una serie di articoli di carattere sia logico in senso stretto che filosofico (si veda Łukasiewicz 1970). Come cercherò di mostrare in quel che segue, considerazioni di carattere modale (inerenti in modo specifico la nozione di "fatto contingente") furono alla base dell'introduzione di un terzo valore logico delle proposizioni, accanto al "vero" e al "falso".

Tuttavia il sistema a tre valori del calcolo proposizionale è sempre stato considerato, contrariamente alle intenzioni esplicite di Łukasiewicz, come del tutto inadeguato per una trattazione delle modalità aletiche (possibilità, necessità ecc.), e le motivazioni che furono all'origine della nascita di quel sistema logico sono state spesso trascurate o fraintese. In quel che segue prenderò in esame sostanzialmente due lavori dell'autore in questione, ossia: "Osservazioni filosofiche sui sistemi polivalenti della logica proposizionale" e "On Determinism".[1] Tali articoli costituiscono infatti a mio avviso i contributi più completi e significativi di Łukasiewicz se vogliamo collocare all'interno della corretta cornice concettuale il genere di argomenti avanzati dall'autore contro il principio di bivalenza. Oltre a gettare luce sul significato della logica trivalente e sulla sua specificità rispetto a quella classica, tali lavori lasciano tuttavia

[1] Entrambi ristampati in traduzione inglese in Łukasiewicz (1970). Del primo esiste anche una traduzione italiana a cura di Giovanna Corsi, in Casari (a cura di) (1979, pp. 241-264).

aperte alcune questioni che richiedono una risposta e che mi propongo di affrontare nel corso del presente lavoro.

Nonostante i forti punti di continuità presenti tra i due articoli appena citati (in merito ai quali entreremo fra breve), è opportuno sottolineare anche gli elementi di peculiarità che li caratterizzano. Il primo articolo si propone infatti di mostrare come i principi modali accettati come evidenti dalla tradizione logico-filosofica non possano trovare un'espressione adeguata nel linguaggio della logica proposizionale bivalente, pena collasso modale o contraddizione. La logica trivalente, al contrario, fornirebbe secondo Łukasiewicz gli strumenti semantici adeguati per l'elaborazione di un sistema logico proposizionale in cui, data un'opportuna caratterizzazione (verofunzionale)[2] degli operatori modali, tali principi risultano verificati. L'interesse per la logica modale da un punto di vista sistematico non è invece al centro dell'articolo "On Determinism", nel quale Łukasiewicz si propone in sostanza di mettere a fuoco il nesso di carattere logico che secondo lui vi sarebbe tra il principio di bivalenza da una parte e una certa formulazione della tesi determinista dall'altra, elaborando quindi una critica del principio di bivalenza sulla base di alcune considerazioni di carattere filosofico-linguistico *e* di una visione del mondo indeterministica. Risulta del tutto chiaro tuttavia come il "filo conduttore" che attraversa questi (e altri) lavori di Łukasiewicz e che sta all'origine del primo sistema di logica trivalente sia costituito dall'idea che all'interno della logica bivalente classica non vi sia alcuno spazio per la *contingenza* e che ciò abbia delle immediate ripercussioni non solo di carattere ontologico (il determinismo), ma anche filosofico-morale (*in primis*, la negazione del libero arbitrio).

2. Logica bivalente e modalità aletiche

Venendo dunque al primo dei due articoli in questione, "Osservazioni filosofiche sui sistemi polivalenti della logica proposizionale", osserva Łukasiewicz in apertura:

> Il sistema trivalente del calcolo proposizionale deve la sua origine ad alcune ricerche da me compiute riguardo alle cosiddette "proposizioni modali" e ai concetti con esse strettamente connessi di possibilità e necessità (1930; trad. it. p. 241).

[2] Come è noto, Łukasiewicz ha sempre cercato di difendere un approccio *estensionale* alla logica modale.

In quella sede Łukasiewicz cerca di mostrare come non sia possibile una trattazione (estensionale) adeguata delle nozioni modali all'interno del linguaggio del calcolo proposizionale bivalente, ma si rendano necessari un rifiuto della bivalenza e l'introduzione di un terzo valore logico al fine di giustificare in modo rigoroso certe intuizioni che stanno alla base di quei principi modali tradizionalmente accettati nella storia della logica. Per "nozioni modali" egli intende, conformemente alla tradizione medievale, i quattro "modi" in cui una data proposizione (e quindi il fatto che essa esprime) può presentarsi: come *possibile*, *impossibile*, *contingens*, *necessarium*.[3] Data una proposizione "p", un operatore di possibilità "M" e la negazione proposizionale "¬" possiamo caratterizzare:

(1) "è possibile che p" = Mp;[4]
(2) "è impossibile che p" = ¬Mp;
(3) "è possibile che non-p" = M¬p;
(4) "è necessario che p" = ¬M¬p.

La contingenza può essere caratterizzata come la congiunzione di (1) e (3), ossia: Mp ∧ M¬p (una proposizione è detta "contingente" quando sia essa sia la sua negazione sono possibili). Tra le quattro proposizioni modali valgono – secondo il tradizionale "quadrato modale" delle opposizioni – i seguenti rapporti: contrarietà tra (2) e (4), contraddittorietà tra (1) e (2) e tra (3) e (4), implicazione tra (4) e (1) e tra (2) e (3), subcontrarietà tra (1) e (3).

Łukasiewicz fa notare che guardando alla storia della logica possono essere enucleati tre gruppi di principi relativi alle proposizioni modali, ognuno dei quali farebbe riferimento ad una *peculiare* accezione di possibilità o necessità. Esempi di principi modali del primo gruppo sono:

(a) *Ab oportere ad esse valet consequentia*;
(b) *Ab esse ad posse valet consequentia*;
(c) *Ab non posse ad non esse valet consequentia.*

[3] Egli aggiunge che i logici medievali indicavano anche altri due modi che una proposizione può assumere, e cioè *verum* e *falsum*. Tuttavia tali modi non furono oggetto di indagini approfondite al pari degli altri, poiché le proposizioni modali costruite mediante essi, "È vero che p" ed "È falso che p", venivano considerate equivalenti alle proposizioni "p" e "non-p".

[4] Osserva Łukasiewicz: "L'espressione 'è possibile che' non è qui definita; il suo significato emerge dai teoremi che valgono per le proposizioni modali" (1930; trad. it. p. 242).

Questo genere di principi, che ci sono stati trasmessi dalla logica classica, vengono considerati verità autoevidenti senza bisogno di dimostrazione (benché non ricavabili direttamente dal quadrato modale delle opposizioni, essendo in essi coinvolta la nozione di esistenza). Un secondo gruppo di principi modali, che fa riferimento ad un genere diverso di necessità, che potremmo chiamare "necessità temporale" (o *ex hypothesi*), trova un suo rappresentante nel seguente teorema:[5]

(d) *Unumquodque, quando est, oportet esse.*

Ossia, "qualunque cosa, quando è, è necessaria". Tale teorema modale, forse non immediatamente intuitivo, ha tuttavia una lunga storia che risale almeno al *De Interpretatione* di Aristotele, dove si afferma (19a23):

> Che ciò che è sia, *quando* è, e che ciò che non è non sia, *quando* non è, risulta certo necessario; non è però necessario che tutto ciò che è sia, né che tutto ciò che non è non sia. In effetti, l'essere per necessità di tutto ciò che è, *quando* è, non equivale all'essere per necessità, assolutamente, di tutto ciò che è [corsivi aggiunti].

Łukasiewicz fa notare come l'espressione *quando* presente in (d) e il corrispettivo greco όταν della citazione aristotelica non sia una particella condizionale, bensì *temporale*. La necessità cui si fa riferimento in (d) è dunque una necessità che riguarda il non poter essere altrimenti di qualcosa nel momento in cui è quel che è. Volendo fare qualche esempio, se prendiamo gli enunciati (contenenti un riferimento temporale) "Io sarò a casa stasera" e "Adesso non ho soldi in tasca" e supponiamo che siano veri,[6] allora ne segue, rispettivamente, la necessità del mio essere a casa stasera e l'impossibilità del mio avere soldi in tasca adesso. In questa accezione, il necessario e l'impossibile equivalgono semplicemente all'*assenza di alternative* rispetto ad un certo stato di cose dato o pensato come (già) esistente. Vedremo in quel che segue quanto peso abbia *questa* accezione della necessità nel modo in cui Łukasiewicz guarda al nesso tra logica bivalente e determinismo e alle modalità in genere. Un terzo gruppo di principi modali è basato sul concetto aristotelico di "possibilità bilaterale". Secondo Aristotele infatti vi sono molte cose che possono essere e possono anche non essere; ossia

[5] Citato da Leibniz nella *Teodicea*; si veda *Die Philosophisce Schriften von Gottfried Wilhelm Leibniz*, a cura di C. I. Gerhard, Berlin: Weidmannsche Buchhandlung, 1875-1890, vol. VI, p. 131.

[6] Per il momento lasciamo da parte il problema dei cosiddetti "portatori di verità" (in inglese *truth-bearers*), nonostante tale problema risulti a mio avviso – come vedremo più avanti – centrale per la questione del nesso tra principi logici e tesi determinista.

che possono essere ma che non è necessario che siano – che sono cioè *contingenti*.[7] Afferma Aristotele nel *De Interpretatione* (19a10-13):

> in linea generale agli oggetti che non sempre sono in atto tocca indifferentemente il potere di essere o di non essere; per tali oggetti entrambe le cose sono possibili, sia l'essere che il non essere, cosicché risultano possibili sia il divenire che il non divenire. E molti oggetti si comportano evidentemente a questo modo.

Un mantello, per esempio, ha la possibilità di venire tagliato (anche se di fatto non verrà tagliato ma si logorerà col tempo); e allo stesso tempo ha la possibilità di non venire tagliato (anche se di fatto verrà tagliato).

Questi tre gruppi di principi modali, che secondo Łukasiewicz sono sufficientemente evidenti all'intuizione, trovano dei loro rappresentanti nei seguenti "teoremi":

(I) Se non è possibile che p, allora non-p.
(II) Se è supposto che non-p, allora non è possibile (sotto questa supposizione) che p.
(III) Per un certo p: è possibile che p ed è possibile che non-p.

Il primo teorema corrisponde al principio (c) visto sopra, che è sempre stato accettato come immediatamente evidente. Il secondo al principio (d), riguardante la "necessità temporale". Il terzo corrisponde alla nozione aristotelica appena illustrata di "possibilità bilaterale". Ora, se vogliamo rendere questi teoremi nel linguaggio del calcolo proposizionale (esteso)[8] bivalente, possiamo trascrivere (I)-(III) come segue, facendo uso dell'operatore di possibilità introdotto in precedenza:[9]

(i) $\neg Mp \rightarrow \neg p$;
(ii) $\neg p \rightarrow \neg Mp$;
(iii) $\exists p(Mp \wedge M\neg p)$.

[7] Se infatti definiamo "contingente" come "possibile e non necessario" le nozioni di "contingenza" e di "possibilità bilaterale" risultano equivalenti.

[8] Ossia un linguaggio in cui si ammette la quantificazione su variabili proposizionali (logica enunciativa del secondo ordine).

[9] Qui e in quel che segue non faccio uso della notazione logica (cosiddetta) polacca, in cui i connettivi verofunzionali vengono sempre posti *davanti* alle funzioni proposizionali su cui operano.

Riguardo a (ii) osserva Łukasiewicz che questo è *l'unico* modo di rendere (II) nel linguaggio del calcolo proposizionale bivalente (1930; trad. it. pp. 244 e 246-247).[10] È degno di nota che quando (II) verrà reso nel linguaggio del calcolo proposizionale trivalente, esso assumerà una forma diversa rispetto a (ii), come vedremo tra breve.

Łukasiewicz mostra come i teoremi (ii) e (iii) presi sia singolarmente sia assieme portino a conseguenze del tutto indesiderabili all'interno del calcolo proposizionale bivalente. Più precisamente egli dimostra (*ivi*, pp. 244-247), sulla base di alcuni lemmi logici proposizionali e delle regole di sostituzione e separazione, che mentre dal teorema (i) possono essere derivati i lemmi logici modali accettati dalla tradizione logica come evidenti e non problematici,[11] dal teorema (ii) può essere derivato esattamente il converso di tali lemmi logici modali;[12] ciò significa che l'accettazione di entrambi i teoremi (i) e (ii) comporta come conseguenza immediata un *collasso modale*. Quanto al teorema (iii), Łukasiewicz mostra, sulla base delle regole suddette unitamente alla regola di introduzione dell'universale, che da esso è derivabile la proposizione Mp (*ivi*, pp. 247-248).[13] Commenta Łukasiewicz su questo punto:

> Liberamente parlando, in base al teorema III siamo stati condotti a riconoscere ogni cosa come possibile. Ma, se ogni cosa è possibile, allora niente è impossibile e niente è necessario (*ivi*, p. 248).

Infatti, essendo Mp un lemma logico, esso vale per ogni p, dunque anche per ¬p. Ma allora non possono mai valere ¬Mp (p è impossibile) e ¬M¬p (p è necessario), essendo queste proposizioni contraddittorie, rispettivamente, a Mp e M¬p.

---

[10] Dichiara Łukasiewicz a questo proposito: "Se infatti una proposizione 'β' vale sotto l'assunzione 'α', questo non significa niente di più che 'β' è vera *se* 'α' è vera. L'implicazione 'se α, allora β' vale, quindi, se 'α' è vera. Poiché questa implicazione deve valere anche se 'α' è falsa, vale in entrambi i casi". Con ciò Lukasiewicz intende fornire una giustificazione della validità classica della "legge di contrazione" $(\alpha \rightarrow (\alpha \rightarrow \beta)) \leftrightarrow (\alpha \rightarrow \beta)$; tale formula non costituisce tuttavia una legge logica del calcolo proposizionale trivalente.

[11] Per esempio: $p \rightarrow Mp$, $\neg M\neg p \rightarrow p$, $\neg Mp \rightarrow M\neg p$.

[12] Ossia: $Mp \rightarrow p$, $p \rightarrow \neg M\neg p$, $M\neg p \rightarrow \neg Mp$.

[13] In realtà Łukasiewicz, nella dimostrazione di Mp a partire dal teorema (iii) fa anche uso di un lemma logico della "prototetica" elaborata da Leśniewski, ossia la tesi: $(\varphi(p) \wedge \varphi(\neg p)) \rightarrow \varphi(q)$, dove "φ" denota un operatore logico *variabile* cui appartiene una sola proposizione come argomento. Tale lemma rappresenta una sorta di generalizzazione della tesi proposizionale classica conosciuta anche come "legge di Duns Scoto" (*ex absurdo sequitur quodlibet*), legge che non vale nel calcolo trivalente.

Se presi singolarmente i teoremi (ii) e (iii) danno luogo a risultati indesiderabili da un punto di vista di una trattazione adeguata delle modalità, la loro congiunzione risulta essere addirittura *inconsistente*. Infatti dal teorema (ii) è possibile derivare – per contrapposizione e separazione – la proposizione $Mp \rightarrow p$, e dal teorema (iii) è possibile derivare la proposizione Mp; ma allora è possibile derivare, per separazione, la proposizione (qualsiasi) p. E ciò equivale all'inconsistenza del sistema in cui vengono accettati entrambi i teoremi (ii) e (iii). Questo fatto può essere illustrato intuitivamente[14] anche nel modo seguente: in conformità al teorema (iii) c'è almeno una proposizione p tale che risultano vere simultaneamente sia Mp che $M\neg p$. Ma dal teorema (ii) sono derivabili le proposizioni $Mp \rightarrow p$ e $M\neg p \rightarrow \neg p$. Ciò implica che, per una certa p, le proposizioni p e $\neg p$ sono entrambe vere simultaneamente, il che è contraddittorio.

I teoremi (ii) e (iii) non possono essere dunque mantenuti entrambi all'interno di un quadro logico bivalente, pena contraddizione. D'altronde, anche la scelta di abbandonare soltanto uno di tali teoremi dà luogo a conseguenze indesiderabili: se abbandoniamo il teorema (ii), siamo costretti ad ammettere, sulla base del teorema (iii), che *ogni* cosa è possibile. Ma, come abbiamo visto, in questa circostanza risulta insensato introdurre nella logica le proposizioni modali. Se invece decidiamo di mantenere il teorema (ii) andiamo incontro ad un collasso delle modalità, ed in più dobbiamo rinunciare alla nozione estremamente intuitiva di "possibilità bilaterale", pena contraddizione. In effetti, osserva Łukasiewicz,

> Non ci si poteva aspettare un risultato diverso. Ciò diventa particolarmente chiaro se il sistema del calcolo proposizionale a due valori è definito dal cosiddetto metodo matriciale. Sulla base di questo metodo si assume che tutte le variabili proposizionali possano prendere solo due valori costanti, cioè "0" o "il falso" e "1" o "il vero" (*ivi*, p. 250).

All'interno del quadro bivalente, infatti, un qualunque operatore proposizionale monadico φ coincide con una delle seguenti quattro funzioni:

$\varphi(0) = 0$ e $\varphi(1) = 0$;
$\varphi(0) = 1$ e $\varphi(1) = 1$;
$\varphi(0) = 1$ e $\varphi(1) = 0$;
$\varphi(0) = 0$ e $\varphi(1) = 1$.

[14] E senza fare uso del lemma logico della prototetica indicato sopra.

La prima funzione, che può essere detta “funzione Falso” (Fp), associa il valore falso ad ogni argomento. La seconda, che può essere detta in modo analogo “funzione Vero” (Vp), associa il valore vero ad ogni argomento. La terza coincide con la consueta negazione proposizionale ($\neg$p). La quarta può essere detta “funzione Identica” (Kp), poiché il valore che essa assume è identico al suo argomento. Ora, poiché abbiamo finora trattato l’operatore modale M come un operatore proposizionale monadico, all’interno del quadro bivalente esso deve coincidere con una delle quattro funzioni appena viste, ossia deve valere uno dei seguenti quattro casi:

(a) Mp = Fp;
(b) Mp = Vp;
(c) Mp = $\neg$p;
(d) Mp = Kp.

Tuttavia, nessuno dei casi (a)-(d) verifica *tutti e tre* i teoremi modali (i)-(iii). Di ciò è facile convincersi se passiamo in rassegna le condizioni di verità di tali teoremi:

(i)	$\neg$Mp $\to$ $\neg$p	vale solo per	Mp = Kp o Mp = Vp;
(ii)	$\neg$p$\to$ $\neg$Mp	vale solo per	Mp = Kp o Mp = Fp;
(iii)	$\exists$p(Mp $\wedge$ M$\neg$p)	vale solo per	Mp = Vp.[15]

Risulta altresì evidente come i teoremi (ii) e (iii) risultino nel quadro bivalente incompatibili tra di loro, non essendovi alcuna interpretazione dell’operatore M che li verifichi entrambi (laddove tali teoremi risultano, singolarmente presi, quantomeno compatibili con il primo teorema). Da tutto ciò ricaviamo che *all’interno del quadro bivalente non può esistere nessun operatore proposizionale monadico che soddisfi i teoremi modali presi in esame*, teoremi che Łukasiewicz vede come l’espressione più completa del modo in cui concepiamo le nozioni modali. La morale che egli trae dalle osservazioni di cui sopra consiste nell’idea che il quadro logico bivalente non sia conciliabile con le intuizioni che soggiacciono ai concetti modali:

> Si ha infatti l’impressione che le nostre intuizioni connesse con i concetti di possibilità e di necessità facciano riferimento a un sistema logico che è fonda-

[15] Abbiamo infatti: $\exists$p(Mp $\wedge$ M$\neg$p) è logicamente equivalente a $\neg\forall$p$\neg$(Mp $\wedge$ M$\neg$p). Ma in generale, nel quadro bivalente, la proposizione $\forall$pα(p) vale *sse* la proposizione α(o) $\wedge$ α(1) vale. Ne segue che $\neg\forall$p$\neg$(Mp $\wedge$ M$\neg$p) vale *sse* $\neg$($\neg$(M0 $\wedge$ M1) $\wedge$ $\neg$(M1 $\wedge$ M0)) *sse* $\neg\neg$(M0 $\wedge$ M1) *sse* M0 $\wedge$ M1 *sse* Mp = Vp.

mentalmente diverso dalla logica ordinaria basata sulla legge di bivalenza (*ivi*, pp. 258-259).

3. Il terzo valore logico e la definizione di possibilità

Il motivo che spinge Łukasiewicz a mettere in discussione la legge di bivalenza[16] e ad assumere un terzo valore logico accanto al "vero" e al "falso" consiste dunque innanzitutto nel fatto che, secondo lui, all'interno della logica a due valori non trovano alcuno spazio i concetti modali, così come essi vengono intuitivamente intesi. In particolare, il concetto di "possibilità bilaterale" – o contingenza – espresso dal teorema (iii) visto sopra risulta inconciliabile con il quadro bivalente. Una attenta riflessione sulle nozioni modali ci spingerebbe perciò a cambiare radicalmente la nostra logica; per meglio chiarire questo punto, Łukasiewicz argomenta come segue:

> Posso assumere senza contraddizione che la mia presenza a Varsavia a un certo istante del prossimo anno, per esempio a mezzogiorno del 21 dicembre, non sia in questo momento decisa né in senso positivo, né in senso negativo. Quindi è *possibile* ma *non necessario* che io sarò presente a Varsavia in quel dato momento. Sotto questa ipotesi la proposizione "Sarò a Varsavia a mezzogiorno del 21 dicembre dell'anno prossimo" non può essere oggi né vera né falsa. Infatti, se fosse vera oggi, la mia futura presenza a Varsavia dovrebbe essere necessaria, il che è in contraddizione con l'ipotesi. Se, d'altra parte, essa fosse oggi falsa, la mia presenza futura a Varsavia dovrebbe essere impossibile, il che pure è in contraddizione con l'assunzione. Quindi la proposizione considerata non è oggi *né vera né falsa* e deve possedere un terzo valore, diverso da "0" o falsità e da "1" o verità. Possiamo denotare questo valore con "½". Esso è proprio "il possibile", che si affianca al "vero" e al "falso" come terzo valore. [...] Il sistema trivalente della logica proposizionale deve la sua origine a questa linea di pensiero (*ivi*, pp. 252-253).[17]

---

[16] Łukasiewicz osserva come il principio di bivalenza (secondo cui ogni proposizione è o vera o falsa) venga talvolta chiamato anche "legge del terzo escluso". Egli preferisce tuttavia distinguere, intendendo per "legge del terzo escluso" il noto principio della logica classica stando al quale due proposizioni (mutuamente) contraddittorie non possono essere entrambe false; si veda (1930; trad. it. p. 252). Questa distinzione ha senso se si accetta che una data proposizione possa essere né vera né falsa; in questo modo possiamo coerentemente rifiutare la bivalenza e mantenere il terzo escluso (come sembrerebbe sostenere Aristotele nel noto cap. 9 del *De Interpretatione*).

[17] La strategia argomentativa adottata da Łukasiewicz nel saggio "On Determinism" per introdurre il terzo valore logico è sostanzialmente arricchita, rispetto a quella adottata qui, da con-

Quello che subito colpisce il lettore in questo passo di Łukasiewicz, e che rappresenta il cuore della sua argomentazione, è il "salto" che egli opera dalla verità alla necessità e, rispettivamente, dalla falsità all'impossibilità. Ma che cosa giustifica tale salto? L'autore non dice niente in proposito, e questo ci lascia presumere che in tale argomento egli stia facendo implicitamente uso di quella che abbiamo sopra chiamato "necessità temporale" (o *ex hypothesi*), che è espressa dal principio modale (d): *Unumquodque, quando est, oportet esse*. Sotto l'assunzione che la proposizione "Sarò a Varsavia a mezzogiorno del 21 dicembre dell'anno prossimo" sia vera (oggi), il mio essere a Varsavia in quel dato momento futuro è già deciso. Analogamente, sotto l'ipotesi che la stessa proposizione sia falsa (oggi), il mio non essere a Varsavia in quel dato momento futuro è ugualmente già deciso. Ma all'interno del quadro bivalente non vi sono altre alternative. Dunque in entrambi i casi il fatto futuro concernente la mia presenza o meno a Varsavia in quel dato momento non è un fatto contingente, bensì determinato già oggi, contro la ragionevole ipotesi che *vi siano* fatti contingenti.[18] Di fronte all'alternativa tra abbandonare tale ipotesi e abbandonare il principio di bivalenza, Łukasiewicz opta per la seconda soluzione, introducendo così un terzo valore logico, il "possibile" appunto, accanto al "vero" e al "falso"; ossia, il valore che possiedono attualmente tutte quelle proposizioni che non risultano ancora[19] decise rispetto alla loro verità o falsità. D'ora in poi, seguendo Łukasiewicz, indicheremo il terzo valore con "½".

Senza entrare in merito alle regole semantiche che stanno alla base del sistema trivalente per il calcolo proposizionale,[20] vediamo adesso come Łukasiewicz fornisce all'interno del quadro logico trivalente una definizione (vero-funzionale) dell'operatore di possibilità M in grado di soddisfare i teoremi modali (I)-(III). Tale definizione, che a quanto ci dice lo stesso Łukasiewicz gli fu

siderazioni di carattere causale e filosofico-linguistico (vedi paragrafo 5). Tale genere di considerazioni permettono a Łukasiewicz di introdurre il terzo valore logico *senza* far ricorso al teorema modale (d) impiegato in questa sede in modo essenziale per passare dalla verità (rispettivamente, falsità) alla necessità (rispettivamente, impossibilità) di una data proposizione.

[18] Per una critica di questo argomento di Łukasiewicz e in particolare del teorema modale (d) da egli impiegato in questo contesto, si veda Woleński (2003). Woleński propone di abbandonare il teorema (d) e di sostituirlo con il principio, a suo avviso "più ragionevole", (d*): *unumquodque, quando est, reale est*.

[19] Da quanto emerso fin qui risulta chiaro che il terzo valore ha a che fare con proposizioni che esprimono fatti contingenti futuri. Come vedremo più avanti, nel lavoro "On Determinism" Łukasiewicz si spinge a sostenere che il "possibile" investe non solo il futuro ma anche (sorprendentemente!) il passato.

[20] Esposte in modo sistematico in Łukasiewicz (1920; 1930). Qui e in quel che segue darò per scontate le regole semantiche del calcolo proposizionale trivalente.

suggerita dal giovane Alfred Tarski durante alcune esercitazioni seminariali all'Università di Varsavia, consiste nel caratterizzare l'operatore di possibilità M nel modo seguente:

(*) $Mp = \neg p \rightarrow p$.

Tale definizione di possibilità dà luogo, nella semantica trivalente, alle seguenti valutazioni di Mp sulla base dell'attribuzione di un certo valore di verità a p:

se p = 1, allora Mp = 1;
se p = ½, allora Mp = 1;
se p = 0, allora Mp = 0.

La prima e la seconda attribuzione ci dicono che la proposizione "È possibile che p" risulta vera sia nel caso in cui p è vera, sia nel caso in cui p è possibile.[21] La terza attribuzione ci dice che se p è falsa, allora è falsa anche la proposizione "È possibile che p". Quest'ultimo fatto equivale a dire che se p è falsa, allora p è impossibile, e ciò stride con le nostre intuizioni, stando alle quali l'impossibile implica il falso ma *non* viceversa. Se p risulta di fatto falsa, da ciò non segue, in generale, che quanto espresso da p non possa (o non avrebbe potuto) in linea di principio accadere.[22] È quindi abbastanza sorprendente che Lukasiewicz, nel commentare le valutazioni di Mp appena illustrate, affermi: "Ciò si accorda molto bene con le nostre intuizioni" (*ivi*, p. 255).

A quali "intuizioni" si sta dunque riferendo qui Łukasiewicz? Dalla risposta a questa domanda dipende la stessa possibilità di interpretare il sistema lo-

---

[21] Il fatto che Mp = 1 quando p = ½ dipende dal modo in cui viene definita, nella matrice del calcolo proposizionale a tre valori, la verità di una implicazione: quando sia l'antecedente che il conseguente di una implicazione sono possibili, l'implicazione è vera. Se tuttavia interpretiamo il "possibile" come "non ancora vero o falso" allora non è affatto chiaro perché l'implicazione debba avere, in generale, la proprietà appena descritta; tipicamente, infatti, l'antecedente ed il conseguente di una implicazione sono proposizioni *diverse*, per cui non si vede perché, in questo caso, sia l'antecedente che il conseguente debbano passare *entrambi* dal valore "possibile" al valore "vero" o al valore "falso". È quindi naturale pensare che Łukasiewicz, nel fornire le condizioni di verità dell'implicazione per il calcolo a tre valori, faccia una scelta che risulta funzionale proprio alla definizione di Mp come $\neg p \rightarrow p$ (nella quale è coinvolta *una* sola variabile proposizionale).

[22] Questa difficoltà fu segnalata già nel 1932 da Lewis e Langford (1959, p. 215, nota 4): "L'asserzione: 'Quando p è falso, allora non è possibile che p sia vero' è altamente ambigua: se essa si-

gico a tre valori come *un certo* sistema di logica modale, secondo le intenzioni esplicite dello stesso Łukasiewicz. Nel prossimo paragrafo affronteremo specificamente questo problema, cercando di avanzare una risposta a tale domanda; per il momento ci basti osservare che il genere di "intuizioni" cui fa riferimento l'autore nel passo citato vanno inquadrate nel contesto del *problema generale* che fa da sfondo all'elaborazione del sistema a tre valori proposto da Łukasiewicz: quello del valore di verità delle proposizioni che esprimono fatti futuri contingenti, problema che Łukasiewicz ha sempre visto come strettamente connesso a quello di una trattazione logica delle nozioni modali (come appare chiaramente dall'ultima frase dell'argomento di Łukasiewicz citato poco sopra).

Ma quale è l'idea che sta dietro alla definizione (*)? Seguendo quanto sostenuto su questo punto da Arthur N. Prior (1953, p. 321), l'idea è che siamo autorizzati a considerare una proposizione p come "possibile" quando la sua contraddittoria ¬p non è più vicina al vero di quanto lo sia p. Dietro questa idea gioca un ruolo essenziale il concetto di "possibilità bilaterale" discusso poco sopra: date due proposizioni mutuamente contraddittorie riguardanti un certo fatto futuro contingente, è ragionevole supporre che allo stato attuale le due proposizioni si trovino in una sorta di "equilibrio" per quanto concerne il loro valore di verità.[23] Ora, nel quadro bivalente p e ¬p → p sono logicamente equivalenti, dunque l'*unico* caso in cui ¬p non è più vicina al vero rispetto a p è quello in cui p è vera e ¬p è falsa. Questo è il motivo per cui nel quadro bivalente non sussiste alcuna distinzione tra p e Mp. Nel quadro trivalente, al contrario, vi è un *ulteriore* caso in cui ¬p non risulta più vicina al vero rispetto a p, ossia quello in cui p assume valore ½; in questo caso anche ¬p assume valore ½ (vi è quindi una sorta di "equidistanza" delle due proposizioni rispetto al vero) e inoltre ¬p → p assume valore 1. Dunque le proposizioni p e Mp risultano, nel quadro trivalente, logicamente distinte.[24] Łukasiewicz, da parte sua,

gnifica 'Non è possibile che p sia vero quando esso è falso' allora è ovviamente valida; ma se essa significa: 'Quando p è falso, p è logicamente inconcepibile, inconsistente', allora essa è scorretta, poiché identifica la concepibilità con la verità di fatto". Per una discussione riguardo al modo in cui l'impossibilità (¬Mp) e la necessità (¬M¬p) siano suscettibili di una interpretazione adeguata all'interno del quadro trivalente adottato da Łukasiewicz si veda il paragrafo seguente.

[23] Cfr. anche Aristotele, *De Interpretatione*, 19a36-37.

[24] Questo è anche il motivo per cui nella logica trivalente non vale la nota legge proposizionale classica detta *consequentia mirabilis*: (¬p → p) → p. Nel caso appena discusso, quello in cui p = ½, abbiamo infatti: (½ → ½) → ½, ossia 1 → ½, che nel calcolo trivalente equivale a ½. Resta valido, invece, il converso di tale implicazione, ossia: p → (¬p → p), che è una istanza della legge logica dell'*a fortiori*. Dal punto di vista modale (nel senso di Łukasiewicz) è proprio la non validità della *consequentia mirabilis* che permette di distinguere, nel sistema a tre valori, la possibilità (Mp) dalla verità (p). Se tale legge classica risultasse valida, il sistema a tre

cerca di chiarire l'idea di "possibile" che sta dietro alla definizione (*) dichiarando quanto segue:

> Questo solo è certo: se un proposizione *può* essere inferita dalla sua contraddittoria, è certamente non falsa, quindi neppure impossibile. È appunto *possibile*, come stabilisce la definizione di Tarski (1930; trad. it. p. 279).

In effetti, se assumiamo che $\neg p \rightarrow p$ valga, allora p non può essere falsa, poiché allora p risulterebbe vera; ma dal vero non può seguire il falso. E se essa non è falsa, *a fortiori* non è impossibile, dunque è (almeno) possibile.

Dalla definizione (*) di "possibile" è facilmente ricavabile la definizione di "necessario" nel modo seguente:

(**) $Np = \neg M \neg p = \neg(\neg\neg p \rightarrow \neg p) = \neg(p \rightarrow \neg p)$.

Volendo dare una lettura intuitiva di tale funzione proposizionale, osserva Łukasiewicz, essa asserisce che una data proposizione p non "contiene" (nel senso che non implica) la propria negazione: in *questo* senso essa è necessaria (*ivi*, p. 256). Analogamente a quanto osservato poco sopra a proposito della possibilità, anche p e $\neg(p \rightarrow \neg p)$ nel quadro bivalente risultano logicamente equivalenti. Non vi è modo, dunque, in tale quadro, di distinguere le proposizioni p e Np. Al contrario, tali proposizioni non sono equivalenti nel quadro a tre valori, per $p = ½$; in questo caso abbiamo infatti che Np è falsa e ciò si accorda bene con le nostre intuizioni, stando alle quali ciò che è (meramente) possibile non è necessario. Allo stesso modo, dalla definizione (*) ricaviamo la definizione di "impossibile":

(***) $Ip = \neg Mp = \neg(\neg p \rightarrow p)$.

Esattamente per la stessa ragione per cui all'interno del calcolo bivalente le funzioni proposizionali p, Np, Mp risultano logicamente equivalenti, sono – all'interno di tale quadro – equivalenti anche $\neg p$, Ip, $M\neg p$, che risultano invece distinte nel quadro trivalente per $p = ½$ (in questo caso abbiamo: $\neg p = ½$, $Ip = 0$, $M\neg p = 1$).

valori "collasserebbe" in quello a due valori (si veda Lewis e Langford, 1959, p. 220). Nel calcolo proposizionale a tre valori vale tuttavia la seguente legge, che asserisce qualcosa di strettamente analogo alla *consequentia mirabilis*: $(M\neg p \rightarrow p) \rightarrow p$, ossia: se la mera possibilità di non-p implica p, allora p vale. Tale legge, nella forma non abbreviata $((p \rightarrow \neg p) \rightarrow p) \rightarrow p$, costituisce un'istanza del quarto schema assiomatico del sistema di Wajsberg per il calcolo trivalente).

Vediamo adesso come, secondo Łukasiewicz, i teoremi modali (I)-(III) possano essere coerentemente verificati tutti assieme all'interno del calcolo proposizionale trivalente. Il primo di essi, che avevamo formalizzato come segue,

(i) $\neg Mp \rightarrow \neg p$

segue immediatamente dalla legge logica proposizionale dell'*a fortiori* nel modo seguente:

1.	$\alpha \rightarrow (\beta \rightarrow \alpha)$	[*a fortiori*];
2.	$p \rightarrow (\neg p \rightarrow p)$	[istanza di 1];
3.	$p \rightarrow Mp$	[def. (*) / 2];
4.	$(p \rightarrow Mp) \rightarrow (\neg Mp \rightarrow \neg p)$	[contrapposizione, 3];
5.	$\neg Mp \rightarrow \neg p$	[separazione, 3-4].

Inoltre, dal teorema (i) otteniamo facilmente tutti i principi modali appartenenti al primo gruppo, che risultano evidenti e non problematici.[25] Da un punto di vista semantico, il teorema (i) è una legge logica del calcolo trivalente, in quanto esso risulta vero anche per $p = ½$; abbiamo infatti: $\neg Mp \rightarrow \neg p = \neg M½ \rightarrow \neg ½ = \neg(\neg ½ \rightarrow ½) \rightarrow \neg ½ = \neg(½ \rightarrow ½) \rightarrow ½ = \neg(1) \rightarrow ½ = 0 \rightarrow ½ = 1$.

Il teorema (II) – riguardante la necessità temporale – risulta verificato nel calcolo trivalente, ma *non* nella forma data a tale teorema precedentemente, ossia:

(ii) $\neg p \rightarrow \neg Mp$.

In questa forma abbiamo infatti, per $p = ½$: $\neg ½ \rightarrow \neg(\neg ½ \rightarrow ½) = ½ \rightarrow \neg(½ \rightarrow ½) = ½ \rightarrow \neg(1) = ½ \rightarrow 0 = ½$. Il teorema (II) risulta tuttavia verificato nel calcolo trivalente se espresso nella forma:

(ii*) $\neg p \rightarrow (\neg p \rightarrow \neg Mp)$.

Se prendiamo $p = \neg q$, (ii*) assume la forma:

[25] La riga 3 della derivazione costituisce ad esempio uno di tali principi (vale l'implicazione dalla verità alla possibilità). Sostituendo $\neg q$ a p in 5 abbiamo inoltre: 6. $\neg M\neg q \rightarrow \neg\neg q$, ossia 7. $Nq \rightarrow q$, che è un altro teorema modale del primo gruppo (dalla necessità segue la verità). Per contrapposizione (8) e separazione da 7 otteniamo poi: 9. $\neg q \rightarrow \neg Nq$ (dalla falsità di fatto di una proposizione segue che essa non è necessaria).

(ii**) $q \to (q \to Nq)$,

che risulta più aderente al modo in cui il principio *unumquodque, quando est, oportet esse* è formulato. Per $q = ½$ abbiamo dunque da (ii**): $½ \to (½ \to \neg(½ \to \neg½)) = ½ \to (½ \to \neg(½ \to ½)) = ½ \to (½ \to 0) = ½ \to ½ = 1$.

D'altronde, come abbiamo visto poco prima, è proprio la formulazione (ii) del teorema (II) che genera un collasso delle modalità, cosa che non accade se il medesimo teorema viene formulato come (ii*). Il fatto che nel calcolo trivalente (ii*) risulta valida ma (ii) non lo è, dipende dalla circostanza generale che in tale calcolo *non* vale (per $\alpha = ½$, $\beta = 0$) la legge di "contrazione":

(LC) $(\alpha \to (\alpha \to \beta)) \leftrightarrow (\alpha \to \beta)$

per cui non si può passare dalla verità di (ii*) a quella di (ii). Questo è anche il motivo per cui, come osserva Łukasiewicz, il teorema (II) può essere espresso, nel calcolo bivalente, *soltanto* nella forma (ii), poiché all'interno di tale calcolo vale l'equivalenza tra $\alpha \to (\alpha \to \beta)$ e $\alpha \to \beta$. Dopo aver mostrato che il teorema (ii*) è derivabile sulla base di alcuni lemmi logici proposizionali (*ivi*, pp. 257-258) osserva Łukasiewicz a questo proposito:

> È stato dimostrato che il teorema II intuitivamente evidente è valido e inoltre in modo tale da rispettare la massima aristotelica secondo la quale non ogni cosa che è, è necessaria, e non ogni cosa che non è, è impossibile. Infatti le espressioni "α" e "$\neg M \neg \alpha$" così come "$\neg \alpha$" e "$\neg M \alpha$" non sono fra loro equivalenti (*ivi*, p. 258).

Tali espressioni sarebbero infatti equivalenti se valesse il teorema (II) nella forma (ii), ma come abbiamo visto non è questo il caso.

Infine, il teorema (III) riguardante il concetto di "possibilità bilaterale" (o contingenza) risulta verificato nella forma:

(iii) $\exists p(Mp \wedge M\neg p)$;

infatti esiste un valore di p per il quale risulta valida: $Mp \wedge M\neg p$. Tale valore è proprio ½. Per $p = ½$ abbiamo infatti: $M½ \wedge M\neg½ = (\neg½ \to ½) \wedge (½ \to \neg½) = (½ \to ½) \wedge (½ \to ½) = 1 \wedge 1 = 1$. Tutte le proposizioni che esprimono fatti futuri contingenti assumono il valore ½ e verificano il teorema in esame. Inoltre, nel calcolo a tre valori *non* può essere derivata dal teorema (iii) la proposizione Mp, come accade invece nel calcolo bivalente; in quella derivazione si fa infatti uso della tesi: $(p \wedge \neg p) \to q$, che non è una legge del calcolo trivalente (per $p = ½$, $q = 0$ abbiamo infatti: $(½ \wedge \neg½) \to 0 = (½ \wedge ½) \to 0 = ½ \to 0 = ½$).

A conclusione generale di quanto esposto finora Łukasiewicz osserva:

> Tutti i tradizionali teoremi per le proposizioni modali sono stati provati esenti da contraddizione nel calcolo proposizionale trivalente sulla base della definizione $Mp = \neg p \rightarrow p$ (*ibid.*).

A ciò egli aggiunge (*ivi*, p. 259) come tale definizione (verofunzionale) di "possibile" dovuta a Tarski sia l'unica capace di soddisfare, nel sistema trivalente, le richieste espresse dai teoremi modali (I)-(III).

4. In che senso la logica trivalente è una logica delle modalità?

L'interpretazione modale del calcolo trivalente presenta indubbiamente alcuni problemi, problemi che hanno indotto molti studiosi a sostenere che tale calcolo sia del tutto inadeguato per una trattazione delle modalità che voglia soddisfare alcuni requisiti intuitivi minimali associati alle nostre idee di "possibile", "necessario" ecc. Accettando infatti la definizione verofunzionale di "possibile" adottata da Łukasiewicz

(*) $Mp = \neg p \rightarrow p$

andiamo incontro ad alcune conseguenze quantomeno "problematiche", ossia:

A. La falsità di una qualsiasi proposizione implica la sua impossibilità; per $p = 0$ abbiamo infatti $Mp = 0$, ossia $\neg Mp$ (p è impossibile). Ciò stride fortemente con le nostre intuizioni, poiché ciò che risulta falso *de facto* non è per ciò stesso impossibile.[26]
B. La verità di una qualsiasi proposizione implica la sua necessità; per $p = 1$ abbiamo infatti $Np = 1$, dove Np è definito, conformemente a (*), come $\neg M\neg p = \neg(\neg\neg p \rightarrow \neg p) = \neg(p \rightarrow \neg p)$. Anche questo fatto contrasta le nostre intuizioni, stando alle quali la verità *de facto* di una qualsiasi proposizione non è sufficiente ad stabilirne la necessità.[27]

[26] Nei casi in cui $p = 1$ oppure $p = \frac{1}{2}$ abbiamo che $Mp = 1$, il che si accorda bene con le nostre intuizioni: se p è vera oppure p è possibile allora (in entrambi i casi) è vero che p è possibile.

[27] Nel casi in cui $p = 0$ oppure $p = \frac{1}{2}$ abbiamo che $Np = 0$. Nel primo caso si asserisce che la falsità di una data proposizione è sufficiente ad escluderne la necessità, il che si accorda bene con le nostre intuizioni; nel secondo caso si asserisce che se una certa proposizione è (meramente) possibile, allora essa non è per ciò stesso necessaria, il che risulta altrettanto intuitivo.

C. Da quanto sopra segue immediatamente che se una qualsiasi proposizione è vera oppure è falsa, allora essa non è contingente, bensì – come abbiamo appena visto – rispettivamente necessaria o impossibile. Conformemente a (*) possiamo infatti definire la contingenza di una proposizione p come: $Cp = Mp \wedge M\neg p = (\neg p \rightarrow p) \wedge (p \rightarrow \neg p)$; e quest'ultima proposizione risulta falsa sia per $p = 1$ che per $p = 0$. Anche ciò è in contrasto con il modo in cui comunemente pensiamo alla contingenza: il fatto che una qualsiasi proposizione sia vera oppure falsa non esclude di per sé la sua contingenza.[28]

Da tutto ciò emerge un quadro delle modalità abbastanza singolare, in cui la verità e la falsità di una proposizione implicano, rispettivamente, la necessità e l'impossibilità della stessa e nel quale nessuna proposizione che sia (anche solo *di fatto*)[29] vera o falsa può dirsi contingente. Ciò ha senz'altro alimentato tra i critici l'idea che, sebbene il sistema di logica trivalente elaborato da Łukasiewicz abbia il merito di essere stato uno dei primi sistemi (coerenti) di logica "non classica" ed abbia aperto la strada a numerosi sviluppi successivi, tuttavia esso non è suscettibile di un'interpretazione e di un impiego soddisfacenti dal punto di vista delle modalità.

È tuttavia possibile argomentare – come è stato fatto ad esempio da Arthur N. Prior (1953) – che il sistema di logica trivalente elaborato da Łukasiewicz è suscettibile di un'interpretazione sensata dal punto di vista modale se lo mettiamo in relazione al problema originario che il logico polacco intendeva affrontare e risolvere. Tale problema è quello dei cosiddetti "futuri contingenti" ed ha una storia lunga quasi quanto la storia della filosofia occidentale. In breve, il problema consiste nel chiederci se gli enunciati (o proposizioni)[30] che esprimono fatti futuri contingenti (ossia fatti che possono sia accadere che non accadere) posseggano già al momento presente un valore logico determinato, "vero" o "falso".

---

[28] Chiaramente, data la definizione di contingenza appena illustrata, risulta che l'unico caso in cui $Cp = 1$ è quello in cui $p = ½$. Ossia, una proposizione è detta "contingente" quando (e solo quando) il suo valore di verità è il "possibile".

[29] Intendo qui proposizioni che esprimono fatti empirici, in quanto contrapposte alle proposizioni che esprimono fatti di carattere ideale o astratto (per esempio le proposizioni della matematica).

[30] Per il momento non distinguo tra enunciati e proposizioni, sebbene la questione relativa ai diversi tipi di "portatori di verità" (*truth-bearers*) sia a mio avviso assai rilevante nell'avanzare una qualsiasi soluzione del problema in esame (il rapporto tra bivalenza e determinismo); a questo proposito si veda l'Appendice.

Tale problema fu posto esplicitamente e affrontato per la prima volta (almeno nella letteratura pervenutaci) da Aristotele, nel cap. 9 del *De Interpretatione*. Quanto alla soluzione avanzata da Aristotele per tale problema non vi è mai stato accordo unanime tra gli studiosi. Quello che comunque risulta del tutto chiaro è che egli (e non solo lui) individuò in modo consapevole un problema autentico connesso al principio di bivalenza, ossia: in che misura è possibile sostenere al tempo stesso (i) una certa indeterminatezza ontologica riguardo a certi fatti futuri e (ii) la validità universale del principio secondo cui ogni enunciato è vero o falso? Al di là delle varie interpretazioni che sono state date della soluzione avanzata da Aristotele, è certo che egli colse in modo chiaro la *tensione* tra (i) e (ii); la stessa tensione che spinse Łukasiewicz molti secoli dopo ad elaborare il suo sistema di logica trivalente. Come rileva quest'ultimo:

> La disputa intorno alla legge di bivalenza ha uno sfondo metafisico: i suoi difensori erano decisi deterministi, mentre i suoi oppositori propendevano per una *Weltanschauung* indeterministica. Con ciò siamo ritornati ancora una volta nell'ambito dei concetti di possibilità e necessità (1930; trad. it. p. 252).

Tra i difensori della legge di bivalenza nel mondo antico sono certamente da annoverare gli stoici, con in testa Crisippo,[31] mentre tra gli oppositori sono da includere gli epicurei e lo stesso Aristotele nella misura in cui egli sospende la validità della legge di bivalenza per quegli enunciati che esprimono fatti futuri contingenti. È rilevante, come fa notare lo stesso Łukasiewicz, che gli opposti atteggiamenti rispetto alla legge di bivalenza siano stati dettati da diverse visioni del mondo; ma *tutti* i protagonisti coinvolti nel dibattito *accettavano* il nesso tra legge di bivalenza e determinismo (o fatalismo), al pari dello stesso Łukasiewicz.[32]

---

[31] Tanto che lo stesso Łukasiewicz fa notare (1930; trad. it. p. 262) come il sistema di logica trivalente da lui elaborato dovrebbe essere propriamente detto "non crisippeo", essendo stato Crisippo il più tenace difensore della validità universale della legge di bivalenza, laddove Aristotele si sarebbe reso conto del fatto che tale legge potesse non valere per certi asserti riguardanti il futuro.

[32] Una delle fonti principali del dibattito sul rapporto tra principi logici e determinismo nell'antichità e sulle varie posizioni in campo è rappresentato dall'operetta di Cicerone *De Fato*. Lo stesso Łukasiewicz fa esplicito riferimento a tale lavoro (per esempio in 1930; trad. it. pp. 263-264). Da tale fonte risulta che l'unico filosofo coinvolto nel dibattito che *rifiutava* il nesso tra bivalenza e determinismo fosse Carneade, primo scolarca della Nuova Accademia, la cui posizione sul tema in esame viene abbracciata dallo stesso Cicerone. Per un quadro ampio e dettagliato del complesso problema del rapporto tra principi logici e determinismo nella filosofia antica (in particolare con riferimento alla Scuola stoica) si veda Bobzien (1988).

Come Aristotele, Łukasiewicz sostiene che vi sono fatti futuri attualmente indeterminati, ossia fatti in relazione ai quali sussiste al momento presente una alternativa reale (una alternativa di carattere *ontologico*, non meramente epistemico); gli enunciati che esprimono tali fatti risultano attualmente "possibili", poiché non vi è niente che al momento presente renda veri tali enunciati oppure li renda falsi (ossia, renda vere le loro negazioni). Qui gioca un ruolo centrale, come vedremo meglio nel prossimo paragrafo, l'idea che se un enunciato è attualmente vero (rispettivamente, falso), allora c'è attualmente qualcosa che lo *rende* vero (rispettivamente, falso); ossia gioca un ruolo determinante la dimensione dei *truth-makers* (usando l'odierna terminologia anglosassone). Lo stesso Aristotele nel *De Interpretatione* (19b) osserva su questo punto:

> non sempre, riguardo ad una affermazione e ad una negazione contrapposte, sarà necessario che una di esse sia vera e l'altra invece falsa: in effetti, ciò che vale per gli oggetti che sono [o che sono stati] non vale allo stesso modo per quelli che non sono [ancora], ed hanno la possibilità di essere o di non essere.

Appare dunque chiaro come in questo contesto la contrapposizione fondamentale concerne (i) ciò che attualmente risulta (ancora) indeterminato ed ha la possibilità di accadere o non accadere e (ii) ciò che attualmente risulta (già) deciso, determinato e non ha alcuna possibilità di accadere o non accadere diversamente.[33] La logica trivalente di Łukasiewicz intende dar conto di *questa* contrapposizione, che rispecchia i diversi *modi* in cui gli eventi possono "presentarsi" nella realtà. Volendo far uso di una certa terminologia metafisica,[34] potremmo dire che un evento può "presentarsi" come *determinato positivamente* (ossia esistente), *determinato negativamente* (ossia non esistente), oppure *indeterminato* (ossia che ha tanto la possibilità di esistere quanto quella di non esistere). Se una proposizione esprime un evento del primo tipo sarà vera, se esprime un evento del secondo tipo sarà falsa, se esprime un evento del terzo tipo sarà possibile. All'interno di questo quadro, nel quale il "regno della possibilità" viene contrapposto al "regno di ciò che non può essere alterato" (sia esso determinato positivamente o negativamente), risulta più facile com-

[33] Ho messo le espressioni "ancora" e "già" tra parentesi per indicare che, sebbene il caso tipico di un fatto possibile riguardi il futuro e il caso tipico di un fatto determinato riguardi il passato, secondo Łukasiewicz vi sono tuttavia molti fatti futuri che risultano già determinati al momento attuale e fatti passati che risultano invece, al presente, meramente possibili (su quest'ultimo punto si veda il paragrafo seguente).

[34] Che Łukasiewicz non utilizza (almeno in questi termini), ma che mi sembra utile per chiarire la distinzione presa qui in esame.

prendere in che senso la verità di una proposizione implichi la sua necessità: ciò significa semplicemente che ciò che viene espresso dalla proposizione (un certo fatto passato, presente o futuro) non è aperto a possibilità alternative, bensì è deciso per l'eternità. Un discorso analogo vale per il rapporto tra falsità ed impossibilità. Verità e falsità comportano in altri termini una situazione di "chiusura" rispetto a possibilità alternative ed in *questo* senso esse implicano, rispettivamente, la necessità e l'impossibilità.[35] È all'interno di questo quadro che i tre valori logici che una proposizione può assumere devono essere interpretati. Osserva Prior a questo proposito:

> Il valore "1" viene assegnato, naturalmente, ad asserzioni che sono vere senza alcun dubbio, o perché si riferiscono a relazioni atemporali (per esempio "2 + 2 = 4") o perché ciò di cui esse parlano è già accaduto oppure il suo accadere è già determinato; il valore "0" è assegnato ad asserzioni che sono certamente false per ragioni analoghe; e il valore "½" ad asserzioni riguardanti il futuro indeterminato. Data questa interpretazione, vi è un chiaro senso in cui ciò che è certamente falso è sempre "impossibile" [...] e ciò che è certamente vero è sempre "necessario" [...]. Poiché abbiamo una verità determinata o una falsità determinata soltanto quando la possibilità di accadere in un modo o nell'altro che viene attribuita a certi eventi futuri è, per una ragione o per l'altra, assente (1953, p. 323).

Per le stesse ragioni appena viste *nessuna* proposizione che sia vera oppure falsa esprime un fatto contingente: un fatto si dice "contingente" proprio quando ha sia la possibilità di accadere che quella di non accadere. Ma ciò che viene asserito secondo verità o falsità non è aperto a possibilità alternative (essendo già deciso) e non può dunque dirsi "contingente". I fatti autenticamente contingenti devono dunque corrispondere ad un *terzo* genere di proposizioni (accanto a quelle vere e quelle false), ossia le proposizioni aventi come valore logico il "possibile".

---

[35] Il fatto che la verità di una proposizione implichi la sua necessità e che la falsità di una proposizione implichi la sua impossibilità, non deve essere confuso con l'equivalenza logica tra verità e necessità da una parte e falsità e impossibilità dall'altra. La circostanza che $p = 1$ implica infatti che $Np = 1$, poiché $Np = \neg M \neg p = \neg(p \to \neg p)$; e la circostanza che $p = 0$ implica che $Ip = 1$, poiché $Ip = \neg Mp = \neg(\neg p \to p)$. Ma per $p = ½$ abbiamo che $Np = 0$ e che $Ip = 0$. Dunque le proposizioni p e Np, così come $\neg p$ e Ip, non hanno nel quadro trivalente le stesse condizioni di verità e non possono perciò dirsi logicamente equivalenti. Le proposizioni "$p \to Np$" e "$\neg p \to Ip$" non sono dunque leggi logiche del calcolo trivalente. Nel quadro bivalente, al contrario, le proposizioni in questione risultano, per la mancanza del terzo valore ½, tautologiche, il che genera un autentico "collasso modale".

Da tutto ciò risulta chiaro che le modalità di cui si occupa Łukasiewicz in questo contesto non coincidono con le modalità "logiche" che usualmente vengono indicate attraverso i simboli "$\Diamond$" (possibile) e "$\Box$" (necessario). Tali modalità stanno infatti ad indicare della proprietà di "livello superiore" di certe funzioni enunciative; se per esempio scriviamo "$\Box(p \rightarrow p)$" intendiamo asserire che la funzione enunciativa "$(p \rightarrow p)$" risulta vera per ogni possibile interpretazione di p, oppure che "$(p \rightarrow p)$" può essere derivata da un dato insieme di assiomi logici mediante l'applicazione di certe regole di inferenza. Analogamente, scrivendo "$\Diamond(p \wedge q)$" vogliamo asserire che c'è almeno un'interpretazione di p e di q tale che la funzione enunciativa "$(p \wedge q)$" risulta vera, ossia che l'espressione in esame è (quantomeno) logicamente coerente. Le modalità logiche non sono dunque esse stesse delle funzioni enunciative, bensì esprimono certe caratteristiche possedute dalle funzioni enunciative. Diversamente stanno le cose nel sistema trivalente elaborato da Łukasiewicz, in cui le modalità vengono intese come i diversi *modi di esistenza* dei fatti; esse sono concetti che si riferiscono a certi aspetti della realtà, non esprimono delle proprietà di "livello superiore" possedute dalle funzioni enunciative. In tale sistema infatti le proposizioni modali *sono* esse stesse delle funzioni enunciative, contrariamente a quanto accade nel caso delle proposizioni che esprimono delle modalità logiche. L'espressione "Mp" (abbreviazione di "$\neg p \rightarrow p$"), nell'ottica di Łukasiewicz, non asserisce la coerenza logica della proposizione p, il suo essere vera in almeno una interpretazione; bensì asserisce che il *fatto* espresso da p è attualmente indeterminato quanto al suo accadere o meno. Analogamente, l'espressione "Np" (abbreviazione di "$\neg(p \rightarrow \neg p)$") non esprime la necessità logica della proposizione p, bensì asserisce che tale proposizione non è soggetta a mutare il proprio valore di verità (il "vero"), ossia che il sussistere del *fatto* da essa espresso risulta deciso per sempre.[36]

[36] Volendo fornire un esempio, l'elemento che distingue i due generi di necessità consiste essenzialmente in questo: se prendiamo l'enunciato [e] "Se Socrate è morto allora Socrate è morto", abbiamo che esso è logicamente necessario in virtù del fatto che ogni possibile sostituzione di p con un qualsiasi enunciato dell'italiano nella funzione enunciativa "$p \rightarrow p$" che esemplifica [e] dà luogo ad un enunciato vero; in *questo* senso asseriamo che [e] è logicamente necessario. Al contrario, il motivo per cui [e] risulta necessario nel senso di Łukasiewicz consiste *semplicemente* nel fatto che [e] è vero. Esso risulta necessario esattamente per la *stessa* ragione per cui l'enunciato atomico [e*] "Socrate è morto" è necessario, essendo quest'ultimo (adesso) vero. In generale, dunque, nel sistema di Łukasiewicz, la ragione per cui un enunciato atomico risulta necessario è la stessa per cui risulta necessario un enunciato di complessità qualunque. La necessità in questo senso non ha dunque a che fare con la forma logica degli enunciati, bensì con lo *status* ontologico dei fatti che ad essi corrispondono.

Per queste ragioni, come ha osservato Prior (*ivi*, p. 324), le proposizioni logicamente necessarie formano una sottoclasse delle proposizioni necessarie nel senso del sistema di Łukasiewicz, proposizioni che potremmo chiamare (seguendo Prior) “vero-funzionalmente necessarie”; le prime, infatti, al pari delle seconde, non sono soggette a mutare il proprio valore di verità (il “vero”) ed esprimono fatti determinati,[37] decisi per sempre. Ciò ha certamente il vantaggio di mettere in luce un elemento di continuità tra la semantica verofunzionale che soggiace al calcolo enunciativo classico e la logica modale (secondo una certa accezione delle modalità).

La distinzione tra questi due generi di necessità ci aiuta inoltre a gettare qualche luce sul significato del principio modale *unumquodque, quando est, oportet esse*, impiegato da Łukasiewicz nell’argomento contro la bivalenza e discusso nel paragrafo precedente. Abbiamo visto come l’unico modo di rendere tale principio in simboli all’interno del quadro bivalente sia: (ii) $p \rightarrow Np$, e che tale trascrizione formale, oltre a non essere suscettibile di un’interpretazione sensata in termini di necessità logica, genera, assieme al principio non problematico (i) $Np \rightarrow p$, un collasso modale. Nel quadro trivalente, al contrario, è possibile[38] rendere il principio modale in questione attraverso la funzione enunciativa (ii*) $p \rightarrow (p \rightarrow Np)$, in cui “N” non va inteso come un operatore di necessità logica, bensì come il non poter essere altrimenti di qualcosa che è nel momento in cui (*quando*) è. Riguardo a quest’ultima accezione abbiamo parlato di “necessità temporale”, un genere di modalità che risulta assai rilevante quando abbiamo a che fare con enunciati che esprimono fatti futuri. Se p è l’enunciato: “Domani alle 12.00 sarò a Varsavia”, allora, *nel momento* in cui tale enunciato è vero, la mia presenza a Varsavia alle ore 12.00 di domani è necessaria, nel senso che essa non può non aver luogo (ossia, è determinata), pena contraddizione. Il genere di necessità che soggiace al detto latino appena citato, *non* è dunque la necessità logica e di ciò sembra essere stato consapevole lo stesso Aristotele nel passo 19a23 del *De Interpretatione* citato nel paragrafo 2.

Sulla base delle considerazioni precedenti possiamo dunque affermare che la logica a tre valori elaborata da Łukasiewicz *sia* suscettibile, sulla base della definizione (verofunzionale) di “possibile” come “$\neg p \rightarrow p$”, di un’interpretazione modale adeguata rispetto a certe intuizioni le quali trovano espressione in alcuni fondamentali principi modali comunemente accettati dalla tradizione logico-filosofica e illustrati in precedenza. In particolare, la semantica a

[37] Ammesso che sia legittimo parlare di “fatti” in riferimento alle verità logiche.

[38] Data la non validità della “legge di contrazione”: $(p \rightarrow q) \leftrightarrow (p \rightarrow (p \rightarrow q))$.

tre valori costituisce il quadro di riferimento entro cui è possibile dar conto di quelle proposizioni che esprimono fatti futuri contingenti, cosa che – nell'ottica di Łukasiewicz – è preclusa all'interno del quadro bivalente, per le ragioni viste nel paragrafo 2. È dunque all'interno di questa problematica, quella concernente il valore logico delle proposizioni al futuro, che la logica trivalente di Łukasiewicz trova la sua genesi e il suo senso come base di una logica modale estensionale.[39]

5. Terzo escluso, determinismo, causalità

Come è stato accennato in apertura al presente lavoro, un secondo contributo fondamentale per mettere a fuoco le motivazioni filosofiche che indussero Łukasiewicz a rifiutare il principio di bivalenza in favore di una logica a tre valori è rappresentato dal suo articolo "On Determinism" (1946).[40] L'aspetto più originale di tale contributo, rispetto all'altro già preso in esame, consiste essenzialmente nel tentativo di mostrare: (1) come la tesi determinista sia de-

---

[39] Per una dettagliata caratterizzazione *sintattica* del "contenuto modale" della logica a tre valori di Łukasiewicz, si veda Minari (2003). In tale lavoro si mostra come l'assiomatizzazione W ad opera di Wajsberg (1931) del calcolo proposizionale a tre valori risulti equivalente ad un sistema assiomatico modale $W^{\Box}$ nel quale gli operatori di possibilità "$\Diamond$" e necessità "$\Box$" vengono definiti al modo di Łukasiewicz come le particolari funzioni enunciative unarie viste sopra e per il quale valgono (o sono derivabili) gli schemi di assiomi caratteristici del sistema modale classico S5 con *in più* lo schema del "collasso parziale": $(\alpha \to (\alpha \to \Box\alpha))$, che abbiamo visto essere una legge logica del calcolo trivalente. In virtù di quest'ultimo fatto, risulta che il sistema assiomatico $W^{\Box}$ non coincide con nessuna delle logiche modali "normali" comprese tra i sistemi S1 ed S5. Ciò ci permette di rilevare come il fondamentale risultato di James Dugundji sulle logiche modali (1940) *non* sia applicabile all'interpretazione modale della logica trivalente di Łukasiewicz. Il teorema di Dugundji asserisce infatti che non esiste alcuna matrice polivalente finita in grado di caratterizzare i sistemi modali classici compresi tra S1 ed S5; ma poiché, come ha mostrato Minari, (1) il sistema $W^{\Box}$ esplicita a livello sintattico il "contenuto modale" del calcolo trivalente di Łukasiewicz e (2) tale sistema non coincide con nessuno dei sistemi modali classici S1-S5 (sebbene sia in grado di provare tutti gli assiomi modali caratteristici di S5), l'impatto del risultato di Dugundji sull'interpretazione modale di tale calcolo è minimo. Di fatto Łukasiewicz ha sempre cercato di difendere il proprio approccio (estensionale) alla logica modale, ben sapendo che esso era in contrasto con i sistemi modali classici come quelli elaborati da Lewis e Langford, che notoriamente trattano le modalità in come operatori intensionali.

[40] L'articolo in questione è una revisione del discorso inaugurale che Łukasiewicz tenne, in qualità di rettore, all'Università di Varsavia nell'anno accademico 1922/1923. Esso assunse forma definitiva nel 1946 e fu pubblicato per la prima volta in polacco con il titolo *O Determinizmie*, J. Łukasiewicz, *Z zagadnień logiki i filozofii*, a cura di J. Słupecki, Warsaw, 1961.

rivabile, attraverso passaggi puramente logici, dal principio del terzo escluso (data un'opportuna formulazione sia dell'una che dell'altro); (2) che il principio del terzo escluso in riferimento ad enunciati al futuro ha senso solo se accettiamo un certo principio (forte) di causalità; (3) che non vi sono ragioni logicamente cogenti per accettare un tale principio di causalità,[41] e quindi che è legittimo sospendere la validità generale del principio del terzo escluso e, con esso, della bivalenza. Quello che qui propongo al lettore è un tentativo di ricostruzione, il più possibile lineare, del complesso (e non sempre chiarissimo) argomento di Łukasiewicz contro la tesi determinista; da tale ricostruzione dovrebbero risultare evidenti alcuni nessi problematici che legano la logica a questioni metafisiche tradizionali, come quella della libertà del volere. Concludo il paragrafo mettendo in luce alcuni problemi sollevati, a mio avviso, dalla soluzione che Łukasiewicz propone.

Secondo un noto detto latino, *Facta infecta fieri non possunt*: ciò che è accaduto non può divenire non accaduto; se qualcosa è accaduto, resta sempre vero affermare che ciò è accaduto; ossia, la verità riguardante fatti passati è in qualche modo eterna, stabile, si conserva. Detto in modo più preciso, se un certo fatto ha luogo all'istante t, allora resta vero in ogni istante *successivo* a t che quel fatto ha avuto luogo a t.[42] Tutto ciò sembra essere intuitivamente accettabile. Sorge a questo punto però una questione, alla quale l'intuizione da sola non riesce a rispondere: vale anche che se un certo fatto ha luogo all'istante t, allora è vero in ogni istante *precedente* a t che quel fatto avrà luogo a t? Il determinista, secondo Łukasiewicz, è colui che risponde affermativamente a questa domanda (l'indeterminista è colui che risponde negativamente). In quel che segue cercherò – ai fini di una maggiore chiarezza – di dare una veste semiformale ad alcune tesi ed argomenti avanzati da Łukasiewicz, il quale, pur esprimendosi lungo tutto il saggio in linguaggio naturale, non disdegna – e an-

[41] È opportuno notare che quando l'autore tenne la lezione da cui il saggio in questione fu tratto (1921), le teorie di fisica atomica che hanno messo in crisi il determinismo erano ancora sconosciute. Ciò significa che negli anni in cui "On Determinism" fu elaborato poteva avere senso sostenere una certa forma di determinismo in ambito fisico. Quando nel 1946 Łukasiewicz pubblica per la prima volta l'articolo in questione, dichiara di trascurare intenzionalmente i risultati ormai già noti relativi al principio (fisico) di indeterminazione, al fine di non modificare nella sostanza l'argomento originario, di carattere essenzialmente logico-filosofico, attraverso considerazioni tratte da altre discipline scientifiche.

[42] Tale idea era già stata ampiamente teorizzata da Tadeusz Kotarbiński nel saggio "The Problem of the Existence of the Future" (1913). È piuttosto naturale pensare che Łukasiewicz avesse letto quel saggio e ne avesse fatte proprie alcune tesi, o che comunque vi avesse trovato un riscontro delle proprie, vista la vicinanza di vedute dei due autori su diversi punti.

zi, auspica – l'eventualità che quanto egli argomenta possa essere esposto in modo più "rigoroso".[43] Sono tuttavia del tutto consapevole del fatto che proponendo una certa formalizzazione di alcuni passi del lavoro di Łukasiewicz vengono con ciò accettate alcune assunzioni che potrebbero essere messe in discussione da parte dell'autore stesso.

In generale, stabiliamo di relativizzare sempre ad un dato istante temporale il possesso di una proprietà da parte di un oggetto. Scrivendo "$P^2(x,t)$" intenderemo dunque asserire: "l'oggetto x possiede la proprietà P all'istante t". Per maggiore concisione, conveniamo di indicizzare la variabile temporale nel modo seguente:

$$P^2(x,t) =_{df} P_t(x).$$

Se intendiamo "vero" come una relazione binaria $T^2(p,t)$ che intercorre tra proposizioni[45] ed istanti temporali, allora abbiamo:

[1] $T_t(p) = 1$ se e solo se è vero all'istante t che p;
[2] $T_t(p) = 0$ se e solo se è vero all'istante t che $\neg p$;

dove t è un istante di tempo qualsiasi e p (in corsivo) è la proposizione espressa da un qualunque enunciato p (1 e 0 stanno per i valori logici "vero" e "falso"). Possiamo adesso esprimere formalmente la tesi determinista come segue:

$$[D]\ P_t(x) \rightarrow \forall s < t.T_s(P_t(x)).$$

Alla lettera [D] dice: se un qualche oggetto x possiede la proprietà P all'istante t, allora, per ogni istante s precedente a t è vero all'istante s che x possiede P all'istante t. In altre parole, se un dato fatto occorre ad un certo istante di

[43] Dichiara Łukasiewicz a questo proposito: "Vorrei confessare fin d'ora che non sono in grado di esaminare questo problema [quello del determinismo], in ogni suo dettaglio, con quella precisione scientifica che pretendo da me stesso. Quello che io offro qui è un saggio assai imperfetto, del quale forse un giorno qualcuno potrà fornire, sulla base di queste considerazioni preliminari, una sintesi più precisa e matura" (1946; trad. ingl. p. 112).

[44] Qui e in ciò che segue con "proposizione" intenderò il *contenuto* espresso da un dato enunciato (proferito in un dato contesto, se l'enunciato contiene espressioni indicali). Così, se proferisco l'enunciato "Sto vedendo una macchia rossa" mentre sto guardando una macchia rossa, tale enunciato esprime la proposizione (vera) che *in quell'esatto istante e luogo* io sto vedendo una macchia rossa. E tale proposizione è un oggetto completamente determinato, libero da ogni ambiguità che caratterizza gli enunciati come quello appena menzionato.

tempo, la verità della proposizione che esprime quel fatto è decisa dall'eternità. Ma ciò sembrerebbe implicare che il fatto espresso da tale proposizione debba accadere necessariamente. In questo senso, il determinismo è secondo Łukasiewicz una tesi essenzialmente *semantica*: esso coincide con l'idea stessa che verità (e falsità) sono caratteristiche "sempiterne" delle proposizioni (ossia, caratteristiche immutabili che ogni proposizione ha *da* sempre e *per* sempre).[45] Ciò naturalmente sembra avere immediate ricadute su questioni di carattere metafisico, prima fra tutte quella della libertà dell'agire umano; osserva Łukasiewicz su questo punto:

> Il determinista guarda agli eventi che hanno luogo nel mondo come se essi costituissero un film drammatico prodotto in qualche studio cinematografico nell'universo. Noi siamo nel mezzo dell'azione e ma non conosciamo la sua fine, sebbene ognuno di noi non sia soltanto uno spettatore, ma anche un attore del dramma. Ma la fine è là, essa esiste fin dall'inizio della performance poiché l'intero quadro è completato dall'eternità (1946; trad. ingl. p. 113).

Il determinismo, nell'accezione in cui lo intende Łukasiewicz, è qualcosa *di più* rispetto alla negazione del libero arbitrio; quest'ultima tesi può esser sostenuta sulla base di argomenti che mettano in luce le coercizioni (esterne ed interne) cui siamo continuamente sottoposti: coercizioni di carattere fisico o psicologico. Ma la tesi [D] è una tesi che – secondo Łukasiewicz – può essere derivata da un principio *logico*, e secondo il nostro autore la coercizione di tipo logico (derivante cioè da principi autoevidenti) è assai più forte di quella fisica.[46] Il determinismo in questa accezione "implica" il rifiuto della libertà del volere, ma non vale il viceversa.

Storicamente, vi sono stati (almeno) due tipi di argomento addotti a sostegno della tesi determinista: un primo tipo fa leva sul principio aristotelico del terzo escluso, e già lo stesso Aristotele prende in esame un argomento di questo tipo; l'altro tipo (che risale almeno agli stoici) è basato sul "principio di causalità". La strategia argomentativa di Łukasiewicz è la seguente: egli espone separatamente i due argomenti a sostegno della tesi determinista; elabora poi una

[45] L'idea che verità e falsità siano caratteristiche sempiterne delle proposizioni è stata difesa con forza da Stanisław Leśniewski, il quale tuttavia – contro quanto sostenuto dai colleghi Łukasiewicz e Kotarbiński – ha sempre negato l'esistenza di un nesso evidente tra questo fatto e la tesi determinista; a questo proposito si veda Leśniewski (1913).

[46] Sulla distinzione tra coercizione di tipo fisico e coercizione di tipo logico si veda Łukasiewicz (1918), che risulta interessante soprattutto per farsi un'idea dello sfondo "ideologico" o "etico" sul quale si muoveva Łukasiewicz.

(acuta) critica del secondo di essi. Poi mostra come vi sia un forte nesso tra i due argomenti, in modo che la critica del secondo si ripercuota sul primo.

Nel corso della storia della logica sono state date formulazioni diverse del principio del terzo escluso; in questa sede, noi adottiamo con Łukasiewicz la seguente: due enunciati mutuamente contraddittori non sono entrambi falsi, ossia, uno di essi deve essere vero. In questa formulazione il principio del terzo escluso è equivalente al principio della bivalenza, secondo cui ogni enunciato (dichiarativo) è o vero o falso.[47] Quest'ultimo principio, sembra implicare il determinismo quando consideriamo enunciati al futuro. Prendiamo ad esempio l'enunciato "Domani a mezzogiorno Bin Laden sarà catturato dagli americani" proferito oggi; stando al principio del terzo escluso, un tale enunciato esprime una proposizione che è o vera o falsa. Ma se essa è vera già oggi, allora sembrerebbe che il fatto in questione *debba* aver luogo domani in modo determinato (un discorso analogo vale nel caso in cui la proposizione sia falsa). A ben vedere, non ha realmente importanza se un tale enunciato venga oggi proferito o pensato da qualcuno oppure no; sembrerebbe essere nella "natura" stessa delle cose che sussista o meno il fatto da esso espresso. E un tale fatto sarebbe "deciso" da sempre. Al di là delle nostre (imprecise) intuizioni, è tuttavia possibile fornire un argomento deduttivo rigoroso a favore della tesi determinista. Quanto segue è – come già accennato – un tentativo di esposizione semiformale dell'argomento proposto da Łukasiewicz a favore della derivazione della tesi determinista [D] dal principio del terzo escluso. Siano A e B due enunciati atomici qualsiasi aventi un riferimento temporale al futuro (essi avranno dunque struttura interna $P_t(x)$, conformemente alla nostra assunzione), e sia s un istante di tempo qualunque. Łukasiewicz afferma che siamo allora disposti ad accettare come valide intuitivamente le seguenti due tesi:

1. $T_s(A) \vee T_s(\neg A)$;
2. $T_s(B) \rightarrow B$.

[47] Tale formulazione del principio del terzo escluso è equivalente alla bivalenza solo se intendiamo la "contraddittorietà" nei termini della negazione enunciativa classica. È infatti impossibile accettare la bivalenza e rifiutare il principio del terzo escluso nella formulazione appena data se intendiamo la negazione come il consueto operatore enunciativo verofunzionale. Tuttavia è possibile, all'interno del quadro bivalente, rifiutare il principio del terzo escluso nella formulazione proposta se intendiamo la negazione diversamente; per esempio, come un "modo della predicazione" *accanto* all'affermazione (come negazione "interna"). Una esposizione sistematica di quest'ultimo punto di vista si trova in Wessel (1976). Per una ricostruzione storico-teoretica del modo in cui Łukasiewicz tratta la negazione si veda Betti (2002).

La tesi 1 asserisce, conformemente al principio del terzo escluso,[48] che dato un enunciato A, al tempo s è vera la proposizione *A* oppure è vera la sua negazione $\neg A$. La tesi 2 asserisce che se una proposizione *B* è vera ad un istante s qualsiasi, allora il fatto espresso dall'enunciato B sussiste. Sulla base delle assunzioni 1 e 2 è possibile derivare la tesi determinista [D] applicando alcuni lemmi logici, nel modo seguente:

3.	$T_s(\neg A) \rightarrow \neg A$	[da 2, sostituendo B con $\neg$A];
4.	$A \rightarrow \neg T_s(\neg A)$	[da 3, applicando $(\alpha \rightarrow \neg\beta) \rightarrow (\beta \rightarrow \neg\alpha)$ e MP];
5.	$\neg T_s(\neg A) \rightarrow T_s(A)$	[da 1, applicando $(\gamma \vee \alpha) \rightarrow (\neg\alpha \rightarrow \gamma)$ e MP];
6.	$A \rightarrow T_s(A)$	[da 4, 5, applicando $(\alpha \rightarrow \beta) \rightarrow ((\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \gamma))$ e MP];
7.	$A \rightarrow \forall s.T_s(A)$	[da 6, per GP];
8.	$P_t(x) \rightarrow \forall s.T_s(P_t(x))$	[da 7, sostituendo A con $P_t(x)$];
9.	$P_t(x) \rightarrow \forall s<t.T_s(P_t(x))$	[da 8, limitando $\forall$].

La tesi determinista [D], quindi, può essere derivata dal principio del terzo escluso sulla base di trasformazioni di carattere puramente logico; ciò significa che il principio del terzo escluso e la tesi determinista (così com'è intesa da Łukasiewicz) esprimono in fondo lo stesso "contenuto". Vediamo adesso l'argomento in favore del determinismo che si basa sul principio di causalità.

Łukasiewicz caratterizza nel modo seguente il concetto di "causa": il fatto F che accade all'istante s è detto causa del fatto G che accade all'istante t, e il fatto G è detto effetto del fatto F se e solo se (i) l'istante s *precede* l'istante t; (ii) F e G sono connessi in modo tale che è possibile *inferire* l'asserzione del fatto G dall'asserzione del fatto F sulla base della conoscenza delle leggi (fisiche) che sussistono tra i rispettivi stati di cose. Tale definizione implica che la relazione di causalità è *transitiva*: dati tre fatti F, G, H, se F è causa di G e G è causa di H, allora F è causa di H (tale proprietà, come vedremo subito, risulta fondamentale per sostenere la tesi determinista). Sulla base della precedente definizione, possiamo formulare il *principio di causalità* in questo modo:

[48] Sotto l'ipotesi che valga $T(\neg A) \leftrightarrow \neg T(A)$, l'assunzione 1 ha la stessa forma di $\alpha \vee \neg\alpha$, che è il modo usuale in cui il principio del Terzo escluso viene espresso.

(1) ogni fatto G che accade al tempo t ha la propria causa in qualche fatto F che accade ad un tempo $s < t$;
(2) per ogni istante n compreso tra s e t occorre un fatto N che è sia effetto di F che causa di G.

Il primo punto afferma che niente accade senza una causa, e che tale causa precede inevitabilmente l'effetto, sebbene talvolta noi percepiamo causa ed effetto come simultanei (vedi il caso in cui la pressione di un pulsante causa il suono di un campanello). Il secondo punto afferma che nell'insieme di fatti che si succedono l'un l'altro, ordinati nella relazione causale, non ci sono "salti"; ossia, l'insieme di tali fatti costituisce un ordine *denso*. Dato un qualsiasi fatto F che accade al tempo t_0, abbiamo quindi una sequenza infinita di fatti che si estende all'indietro nel tempo:

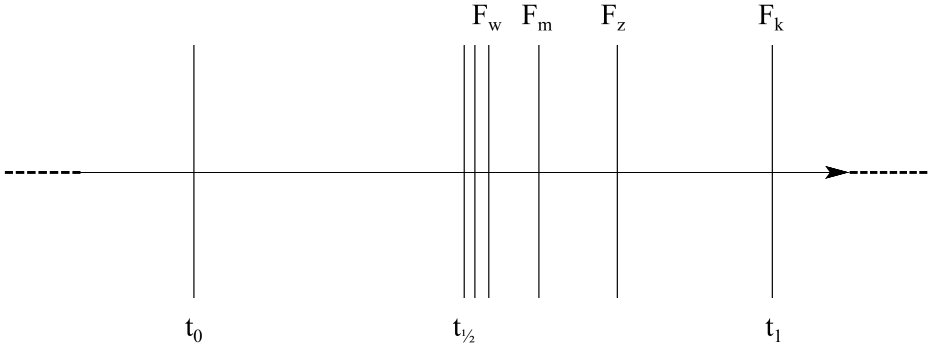
$$\ldots F(t_n), F(t_{n-1}), \ldots, F(t_2), F(t_1), F(t_0)$$

ognuno dei quali è sia causa di quello successivo che effetto di quello precedente, data la transitività della relazione causale.[49] Utilizzando la definizione di "causa" data più sopra e il principio di causalità è possibile fornire un secondo argomento a favore del determinismo. Infatti, preso un qualsiasi enunciato al futuro, per esempio "Domani a mezzogiorno Bin Laden sarà catturato dagli americani", se si dà il caso che esso esprima una proposizione vera, questo significa che il fatto in questione è l'effetto di una serie infinita di cause che lo precedono nel tempo (altri fatti) e dalle quali esso segue necessariamente. Se esso al contrario esprime una proposizione falsa, ossia se non sussiste il fatto in questione, ciò significa che nel tempo non esiste alcuna serie infinita di cause come quella che avrebbe portato al realizzarsi del fatto descritto. In entrambi i casi, la cattura o meno di Bin Laden da parte degli americani a mezzogiorno del giorno in questione risulta decisa dall'eternità in modo determinato.

[49] Nei lavori giovanili Łukasiewicz si è occupato a fondo del concetto di "causa" e del principio di causalità, nell'ambito dell'elaborazione di una "metafisica scientifica" intesa come teoria generale degli oggetti; quello di "causa" è per lui un "concetto astratto reale", ossia un concetto che descrive un aspetto generale della realtà. Il lavoro che affronta in modo specifico questa tematica è (1906). Per avere un'idea dei lavori filosofici giovanili di Łukasiewicz e dell'ambiente intellettuale nel quale si sono sviluppati può essere utile vedere Woleński (1989), in particolare il cap. 3.

6. Indeterminatezza ontologica e proposizioni possibili

L'argomento a favore del determinismo fondato sul principio di causalità che abbiamo appena visto è viziato, secondo Łukasiewicz, da un errore concettuale: ossia, il credere che la serie infinita di cause da cui dipenderebbe ogni fatto futuro debba necessariamente raggiungere l'istante presente, e dunque ogni istante passato. Infatti, la serie (infinita) delle cause di un qualche fatto futuro potrebbe avere un suo "limite inferiore" in un istante di tempo successivo all'istante presente, limite oltre il quale tale serie non si spingerebbe. Se così fosse, essendo tale limite collocato ad un tempo successivo rispetto all'istante presente, al momento attuale non sussisterebbe *nessuna* causa dell'evento futuro in questione. In effetti, tutto ciò è possibile se ci rappresentiamo il tempo come costituito da un insieme denso di istanti, i quali stanno in una relazione biunivoca con i fatti appartenenti ad una data serie causale, nel modo seguente:



Supponiamo che l'istante presente sia t_0, che un certo fatto futuro F_k accada all'istante t_1, e che la serie delle cause di F_k abbia un limite inferiore in $t_{½}$; se ad ogni istante di tempo corrisponde un fatto nella serie causale, le cause dell'evento F_k sono una serie infinita di fatti (F_z, F_m, F_w, ...) che hanno luogo in istanti t_n per $½ < n < 1$. Tale serie non ha un inizio, ossia non c'è una "causa prima". Infatti una tale causa, per dirsi "prima", dovrebbe aver luogo nell'istante corrispondente al più piccolo numero razionale maggiore di ½, ma un tale numero non esiste. Infatti nella serie ordinata dei numeri razionali, presi due numeri qualsiasi, tra di essi ne esiste sempre un altro (proprio questo fatto caratterizza, in generale, gli insiemi densi); dunque tra due numeri razionali ve ne sono infiniti. Analogamente, se ci rappresentiamo il tempo come un ordine denso di istanti, non vi sono due istanti temporali che si succedono l'un l'altro immediatamente, e di conseguenza presi due istanti di tempo fra di essi ve ne sono infiniti. Osserva Łukasiewicz:

> Questo ragionamento mostra che possono esistere infinite sequenze causali che non hanno avuto ancora inizio e che appartengono interamente al futuro. Tale idea è non solo logicamente possibile, ma sembra essere anche più ragionevole rispetto alla credenza che ogni evento futuro, anche il più insignificante, abbia le proprie cause che agiscono fin dall'inizio dell'universo. [...] Nessuno è in grado di predire oggi che una mosca che ancora non esiste ronzerà nel mio orecchio a mezzogiorno del 7 settembre del prossimo anno. La credenza che tale comportamento futuro di tale mosca futura abbia le proprie cause già oggi e le abbia avute fin dall'eternità sembra essere una fantasia, piuttosto che una tesi supportata da un barlume di conferma scientifica (*ivi*, p. 121).

Una prima importante conclusione cui perveniamo con quest'ultimo argomento, è che è possibile sostenere che niente accade senza una causa e che ogni fatto ha la propria causa in qualche fatto precedente, *senza* essere con ciò deterministi; in altre parole, il principio di causalità non implica, di per sé, il determinismo.

Ora, secondo Łukasiewicz l'argomento che deriva la tesi determinista dal principio del terzo escluso diviene pienamente comprensibile solo se si assume l'argomento che deriva la tesi determinista dal principio di causalità, ossia se si assume l'idea che ogni fatto sia il prodotto (necessario) di cause che agiscono dall'eternità. Infatti nel primo di questi argomenti facevamo uso dell'ipotesi (assunzione 1) che un enunciato al futuro proferito all'istante attuale esprima una proposizione vera oppure una proposizione falsa; per esempio, "Domani a mezzogiorno Bin Laden sarà catturato dagli americani" proferito oggi esprime una proposizione vera se domani accade il fatto in questione. Ma affinché tale proposizione sia vera – sostiene Łukasiewicz – vi deve essere *oggi* un "correlato reale" che la rende vera; dunque dobbiamo assumere che la catena causale da cui dipende il verificarsi del fatto futuro (ossia la cattura di Bin Laden da parte degli americani a mezzogiorno di domani) raggiunga l'istante presente, e che vi sia adesso un fatto, appartenente a tale catena, che rende vera la proposizione espressa dall'enunciato (che tale fatto attuale sia la causa del fatto espresso dalla proposizione ci è assicurato dalla transitività della relazione causale).[50]

[50] Questo stesso genere di considerazioni volte a spiegare in termini causali il valore di verità di un enunciato attuale concernente fatti futuri si trova esattamente negli stessi termini già in Crisippo, esponente della Scuola stoica (autore che Łukasiewicz aveva ben presente). Anche

Ma come abbiamo visto dalla critica del secondo argomento, non è affatto necessario che una tale catena causale raggiunga l'istante presente. Infatti la serie infinita di cause che determina un tale fatto potrebbe aver luogo interamente nel futuro. Di conseguenza, possiamo coerentemente sostenere che al momento attuale non c'è niente che renda vera la proposizione espressa dall'enunciato in esame. Così come è del tutto consistente sostenere che al momento attuale non vi sia alcun correlato reale che rende vera tale proposizione, possiamo sostenere – esattamente per le stesse ragioni – che non c'è niente che randa vera la sua negazione, ossia, che renda falsa la proposizione di partenza. Tale proposizione è quindi, al momento attuale, semplicemente "indeterminata", "possibile". Con ciò viene negato il principio logico della bivalenza, secondo il quale ogni proposizione è o vera o falsa. È opportuno sottolineare che l'indeterminatezza che (in generale) caratterizza le proposizioni al futuro è di carattere *ontologico*, non semplicemente *epistemico*; ossia, essa non dipende da un "limite" della nostra conoscenza, ma dal fatto che un certo evento futuro risulta attualmente non deciso nella "realtà delle cose". Così, se p è la proposizione espressa da un qualunque enunciato al futuro[51] ed s è il momento presente, risulta che sia p che $\neg p$ sono (in linea di principio) indeterminate ad s. Osserva Łukasiewicz a questo proposito:

> Io sostengo che vi sono proposizioni che non sono né vere né false, bensì *indeterminate*. Tutti gli enunciati riguardanti fatti futuri che non sono ancora decisi appartengono a questa categoria. Tali enunciati non sono veri al momento presente, poiché essi non hanno un correlato reale [che li renda veri], né sono falsi, poiché neanche le loro negazioni hanno un correlato reale. Facendo uso di una certa terminologia filosofica non particolarmente chiara, potremmo affermare che, da un punto di vista ontologico, a tali enunciati non corrisponde né l'essere né il non essere, bensì la possibilità (*ivi*, p. 126).

E quale valore logico dovremmo assegnare a $T_s(p)$? Poiché né p né $\neg p$ sono vere all'istante s, anche la proposizione espressa da $T_s(p)$ non sarà ad s né vera né falsa – si vedano le definizioni [1] e [2] date sopra – bensì indetermina-

Crisippo sostiene la tesi che affinché un enunciato al futuro possa dirsi vero (o falso) adesso, vi deve essere adesso una causa dalla quale il fatto futuro espresso dall'enunciato segue necessariamente (oppure necessariamente non segue). Crisippo tuttavia, sostenitore di una visione deterministica della realtà, accetta, a differenza di Łukasiewicz, la validità generale del principio di bivalenza; si veda Cicerone, *De Fato*, X, 20-21.

[51] Indipendentemente dal fatto che qualcuno asserisca effettivamente tale enunciato, oppure lo pensi.

ta anch'essa. Indicando questo *terzo* valore logico con ½, e con t un istante qualsiasi abbiamo dunque:

[3] $T_t(p) = ½$ se e solo se non è vero a t che p e non è vero a t che $\neg p$.

Nel caso in cui $T_t(p)$ è indeterminato, anche la sua negazione risulta dunque indeterminata, cosicché abbiamo che lo stesso principio del terzo escluso, nella forma:

[TE] $T_t(p) \vee \neg T_t(p)$

è indeterminato. Esso non può dunque essere più accettato come una verità logica.[52]

Ma cosa ne è della tesi determinista [D]? Si potrebbe argomentare che, essendo stata derivata sulla base di una premessa indeterminata, tale tesi non può che essere essa stessa indeterminata. In effetti possiamo constatarne l'effettiva indeterminatezza sulla base di una semplice considerazione. Dando un'occhiata alla formulazione della tesi,

[D] $P_t(x) \rightarrow \forall s < t.T_s(P_t(x))$

vediamo subito che possiamo scegliere opportunamente un t e un s, in modo tale che l'antecedente dell'implicazione risulti vero ed il conseguente indeterminato; tornando al nostro esempio, supponiamo che sia vero oggi (t_0) che domani a mezzogiorno (t) Bin Laden (x) sarà catturato dagli americani (P);[53] abbiamo dunque che $P_t(x) = 1$. Tuttavia, abbiamo visto che è del tutto coerente sostenere che la catena causale che determina il fatto in questione non si spinga all'indietro nel passato *per sempre*; dunque possiamo sostenere che ieri (s) non era né vero né falso che oggi è vero che domani a mezzogiorno Bin Laden sarà catturato dagli americani, ossia possiamo asserire che

$\exists s < t.T_s(P_t(x)) = ½$.

[52] Risulta abbastanza chiaro da quanto detto, che il rifiuto del principio del terzo escluso nella forma [TE], *dipende* da un rifiuto del principio logico della bivalenza. È quest'ultimo principio, dunque, quello che di fatto è alla radice del problema del determinismo, così com'è inteso da Łukasiewicz.

[53] Assumendo dunque che *oggi* la catena causale che porterà domani alla cattura di Bin Laden da parte degli americani sia già in atto.

Ma allora abbiamo un caso in cui il conseguente dell'implicazione della tesi [D] non è vero, ma indeterminato, cosicché tutta l'implicazione risulta indeterminata.[54] La tesi determinista non è dunque falsa, ma *indeterminata* essa stessa.

Una conseguenza di tutto ciò è che l'idea di poter intervenire sul corso degli eventi futuri mediante delle libere scelte si rivela un'idea (almeno) logicamente sostenibile:

> Il dramma universale non è un'immagine completa fin dall'eternità; più ci spostiamo lontano dalle parti del film che vengono mostrate all'istante presente, tanto più numerosi sono le lacune e gli spazi vuoti che il film include. [...] Tra le possibilità che ci aspettano possiamo scegliere il percorso migliore ed evitare il peggiore. Possiamo in una certa misura modellare il futuro del mondo secondo i nostri progetti (*ivi*, p. 127).

7. Conseguenze filosofiche e problemi aperti

Tra le conseguenze filosoficamente rilevanti che sembrano discendere dalle considerazioni sopra esposte, Łukasiewicz segnala una certa simmetria tra enunciati al futuro attualmente indeterminati ed enunciati al passato che esprimono fatti che non hanno più alcun "impatto causale" sul presente: anche questi ultimi enunciati sarebbero da considerare, al pari di quelli al futuro, indeterminati (o possibili). Infatti, al pari degli enunciati indeterminati al futuro, vi sarebbero molti enunciati concernenti il passato che al momento presente non hanno *più* alcun correlato reale che li renda veri (o che renda vere le loro negazioni); enunciati del tipo "Nel giorno dell'omicidio di Giulio Cesare uno scalpellino a Roma si è schiacciato un dito", anche se sono stati veri o falsi nel tempo cui si riferiscono, hanno avuto tuttavia con ogni probabilità una "storia causale" il cui "limite superiore" è da collocarsi a un istante di tempo di poco successivo al verificarsi dell'evento in questione. Enunciati come quello citato, in altri termini, descrivono dei fatti che sono stati, sì, causa di una serie infinita di effetti (allo stesso modo in cui sono stati effetto di una serie infinita di cause), ma tale serie avrebbe incontrato molto presto un suo limite superiore, limite che non avrebbe consentito a tale serie di raggiungere l'istante pre-

[54] Tra le leggi dell'implicazione per l'indeterminato abbiamo infatti: $(1 \rightarrow \frac{1}{2}) = \frac{1}{2}$. Ciò è piuttosto intuitivo, se pensiamo al significato classico dell'implicazione materiale e al fatto che l'indeterminato può "divenire" o vero o falso.

sente. Tutto ciò contraddice il detto latino da cui avevamo preso le mosse nel paragrafo 5: *Facta infecta fieri non possunt*; a quanto pare, molti fatti riguardanti il passato, al pari di molti riguardanti il futuro, appartengono al "regno del possibile". Una tesi, questa, che sembra avere una certa rilevanza sia come tesi di filosofia della storia, che nell'ambito della filosofia morale; commenta a questo proposito Łukasiewicz:

> Vi sono duri momenti di sofferenza nella vita di ognuno, e momenti ancora più duri di colpevolezza. Dovremmo essere lieti di poterli cancellare, non solo dalla nostra memoria, ma anche dall'esistenza. Possiamo credere che allorquando tutti gli effetti di quei momenti fatali si sono esauriti, dovesse ciò accadere anche soltanto dopo la nostra morte, anche le loro cause saranno cancellate dal mondo dell'attualità e passeranno nel regno della possibilità. Il tempo calma le nostre preoccupazioni e ci porta il perdono (*ivi*, p. 128).

Da un punto di vista epistemologico, è possibile sostenere che dal momento che vi sono catene causali che sono "collocate" interamente nel futuro, vi saranno infiniti fatti indeterminati al momento attuale, *fatti che neanche una mente onnisciente potrebbe predire*. Infatti l'indeterminatezza attuale di tali fatti futuri non è una questione legata ai limiti della conoscenza, ma dipende dalla circostanza che infinite catene causali hanno un loro limite inferiore che giace interamente nel futuro; ciò significa che anche una mente dotata di infinita memoria, infinita capacità di calcolo, e che conosca esattamente lo stato attuale dell'universo e le sue leggi fisiche, al momento attuale non sarebbe in grado di predire tutti quei fatti futuri la cui "storia causale" non è ancora iniziata. Una tale mente potrebbe così predire un numero sempre minore di fatti, man mano che volesse spingersi sempre più in profondità nel futuro.

Al di là del sicuro interesse filosofico degli argomenti addotti da Łukasiewicz a sostegno di una certa forma di indeterminismo, vi sono in essi, a mio avviso, alcuni nodi problematici che meriterebbero un'analisi e una discussione più ampie, cosa che in questa sede non mi è possibile portare avanti e che necessiterebbe di una trattazione separata. Per il momento mi limito a segnalarne brevemente alcuni.

Una questione non del tutto chiara riguarda la richiesta, avanzata da Łukasiewicz, che una proposizione, per essere vera al tempo t, debba avere al tempo t un correlato reale che la rende vera. Se la proposizione concerne il futuro, infatti, si potrebbe richiedere che la proposizione, per essere vera al tempo t, abbia un correlato reale nel futuro, in particolare nel momento cui essa fa riferimento. La mia impressione è che sotto questa scelta di Łukasiewicz vi sia l'idea che il futuro esista soltanto negli effetti del presente, e che il passato esista soltanto nelle cause del presente; laddove tali effetti non si sono ancora generati, o tali cause si sono ormai perdute, non esiste più né futuro né passato, per cui l'unico "luogo" in cui cercare un *truth-maker* è il presente. Tale idea risulta legata a quella concezione metafisica nota come "presentismo", stando alla quale soltanto ciò che è presente esiste. Se ci muoviamo in questa prospettiva, dunque, ciò che rende vera o rende falsa una proposizione al futuro deve essere per forza di cose qualcosa di presente. Resta il problema di spiegare fino a che punto abbia senso parlare delle proposizioni in relazione al tempo, dato il carattere ideale che comunemente è attribuito a tale genere di oggetti (per una discussione di questo punto si veda l'Appendice). Lo stesso Łukasiewicz, utilizzando indifferentemente i termini "enunciato" e "proposizione", non contribuisce a fare chiarezza riguardo a tale problema.

Altra questione riguarda l'assunzione che al continuo degli istanti temporali debba corrispondere un "continuo di fatti", ogni fatto essendo effetto di quelli precedenti e causa di quelli successivi. L'idea di un "continuo di fatti" sembra infatti andare incontro allo stesso problema sollevato dal celebre paradosso di Zenone. Se si assume che un fatto F che accade al tempo t_0 è causa di un fatto G che accade al tempo t_1, vi sarà un fatto H che accade a $t_{½}$ e che è a sua volta causa del fatto G. D'altronde, al tempo $t_{¼}$ vi sarà un fatto M che è causa del fatto H, e così via all'infinito; in questo modo non si vede come la serie delle cause possa raggiungere, nella realtà, il fatto G dal fatto F (nonostante la transitività della relazione causale).

Infine, è da segnalare un problema che riguarda il rapporto tra la libertà del volere e l'idea che vi siano infinite catene causali che appartengono interamente al futuro. In breve: secondo Łukasiewicz io ho la facoltà di scegliere liberamente tra il compiere o il non compiere, all'istante t, una certa azione X; tuttavia una tale azione, nel caso in cui io la compia, in quanto evento *fisico* avrà una serie (infinita) di effetti. Ma quale è la causa di tali effetti? Stando alla teoria delineata da Łukasiewicz, tali effetti non hanno una "causa prima", ma si perdono all'indietro nel tempo fino ad incontrare un loro limite inferiore, collocato ad un istante di tempo successivo a t; la causa di tali effetti, non può essere stata la mia libera scelta, altrimenti, in quanto *causa*, non sarebbe stata un'azione libera, bensì a sua volta effetto di una causa precedente (per definizione di "causa"). Ci troviamo dunque nella situazione paradossale in cui una libera scelta origina e non origina una serie di effetti. In altri termini, risulta assai problematico capire in che modo una *libera* scelta possa entrare in relazione con una serie *causale* infinita da cui dipende il realizzarsi del fatto futuro che è oggetto della scelta stessa.

Appendice
Portatori di verità e determinismo logico: un approccio deflazionistico al problema

Una qualunque risposta al problema del rapporto tra principio bivalenza e determinismo non può prescindere a mio avviso da precise assunzioni concernenti la questione del genere di oggetti che diciamo *essere* veri o falsi, ossia la questione dei cosiddetti "portatori di verità" (*truth-bearers*). Tali assunzioni sono presenti in modo più o meno esplicito negli argomenti proposti da Łukasiewicz contro la bivalenza e il determinismo e giocano un ruolo essenziale nella posizione da lui assunta rispetto al problema in esame. Tuttavia in molti passi dei due saggi presi in esame non è del tutto chiaro a *quale* genere di portatori di verità Lukasiewicz si riferisca, e ciò rende difficile fornire un'interpretazione chiara e univoca di alcune tesi da lui avanzate. In questa appendice mi propongo di mostrare (1) come una qualunque risposta al problema generale del rapporto tra bivalenza e determinismo non possa in alcun modo prescindere da una qualche assunzione riguardo a ciò di cui si predica la verità; e (2) come una analisi attenta e sistematica della natura dei diversi portatori di verità ci induca a mettere in discussione il modo stesso in cui il problema del determinismo logico è stato usualmente *posto*, suggerendoci, più che una soluzione dello stesso, una sua dissoluzione.

Fin qui si è parlato indifferentemente di verità e falsità come proprietà di enunciati o proposizioni, senza prestare troppa attenzione alle conseguenze che una determinata scelta riguardo a ciò di cui si predica la verità può avere rispetto al problema del determinismo logico. Per prima cosa, dunque, passiamo in rassegna in modo schematico i "candidati" *standard* al ruolo di portatori di verità.[55]

Un primo genere di oggetti che possono sensatamente essere detti veri o falsi sono gli *enunciati*, intesi come stringhe di simboli (espressioni) appartenenti a un dato linguaggio.[56] Per esempio sono enunciati dell'italiano:

(a) "Mi fa male un dente";
(b) "Berlusconi è l'attuale presidente del Consiglio dei ministri in Italia";

[55] In questa sede scartiamo subito – per brevità – l'idea che "vero" e "falso" possano essere legittimamente predicati di oggetti extralinguistici, come quando diciamo: "Questo oro è vero", oppure: "Quell'uomo è falso". Tale uso di "vero" e "falso" – detto anche "uso ontologico" – richiederebbe una critica a parte, che però esula dall'obiettivo della presente discussione.

[56] Diamo per scontato che tale linguaggio sia *interpretato*, ossia che alle sue espressioni sia associato un preciso significato – altrimenti il problema della verità/falsità non si porrebbe affatto.

(c) "L'acqua bolle a cento gradi centigradi";
(d) "L'acqua bolle";
(e) "Non esiste il più piccolo numero reale maggiore di uno";
(f) "Ogni cosa è identica a se stessa";
(g) "2 più 2 è uguale a 5";
(h) "Piove a Firenze alle ore 17.30 del 13 luglio 1993";
(i) "Piove".

Per prima cosa va osservato che se ci chiediamo quali degli enunciati (a)-(i) sono veri e quali sono falsi, ci accorgiamo che alcuni sono veri o falsi indipendentemente dal contesto di asserzione, mentre il valore logico degli altri dipende da tale contesto. Per "contesto di asserzione" intendiamo qui (i) il parlante che proferisce l'enunciato, (ii) le circostanze in cui si trova quando lo proferisce (compresi eventuali gesti suoi ed altrui) e (iii) il tempo del proferimento. Per esempio, gli enunciati (c), (e), (f) sono veri indipendentemente dal contesto di asserzione, mentre l'enunciato (g) è falso indipendentemente dal contesto di asserzione. L'enunciato (h) è – nonostante la nostra eventuale ignoranza in proposito – o vero o falso indipendentemente dal contesto di asserzione (esso esprime infatti un evento perfettamente determinato che o ha avuto luogo oppure no). Ci si riferisce di solito a tale classe di enunciati parlando di "enunciati eterni": con ciò si vuole sottolineare il carattere "eterno" del loro valore logico (o sempre vero o sempre falso). Al contrario di questi ultimi, gli enunciati che restano, ossia (a), (b), (d), (i), possono dirsi veri o falsi *soltanto* in riferimento ad un contesto di asserzione. Ciò dipende dal fatto che tali enunciati contengono quelle espressioni che sono usualmente chiamate "indicali", ossia espressioni la cui denotazione varia in funzione del contesto di proferimento dell'enunciato in cui occorrono.[57] Per esempio, (a) è vero se e solo se viene asserito da qualcuno nel momento in cui egli ha mal di denti; (b) è vero se e solo se viene asserito da qualcuno nell'arco di tempo in cui Berlusconi è presidente del Consiglio in Italia; (d) è vero se e solo se viene proferito da qualcuno in un contesto tale che fa sì che egli si riferisca a dell'acqua che sta bollendo; (i) è vero se e solo se viene proferito da qualcuno nel momento in cui sta piovendo nelle sue immediate vicinanze o nel luogo in cui il contesto di asserzione fa sì che egli si riferisca. Se le condizioni di verità specifica-

[57] Tra le tipiche espressioni indicali vengono comunemente annoverati i pronomi personali, i pronomi possessivi, gli avverbi di luogo, gli avverbi di tempo. Tuttavia, anche in assenza di tali particelle, gli stessi tempi verbali possono – in alcuni casi – svolgere funzione indicale, quando l'istante o arco di tempo cui si fa riferimento nell'enunciato dipende dal momento in cui questo viene proferito (si vedano per esempio gli enunciati (a), (d), (i)).

te non sussistono, gli enunciati appena indicati risultano falsi. Dunque a proposito di questi ultimi non ha senso parlare di verità o falsità "eterne". Ma cosa significa che un enunciato è vero o falso indipendentemente dal contesto di asserzione? Ciò può essere inteso come il fatto che tale enunciato è vero o falso "in sé", indipendentemente dall'essere concretamente impiegato da qualcuno in particolari circostanze. Oppure, se quest'ultima idea sembra bizzarra, l'indipendenza del valore di verità di un enunciato dal contesto di asserzione può essere vista come la proprietà di tale enunciato di mantenere lo stesso valore di verità in *qualunque* circostanza venga proferito. Gli enunciati che possono risultare ora veri ora falsi, al contrario, non hanno "in sé" alcun valore logico, ma ne acquisiscono uno soltanto "in relazione ad altro", ossia in rapporto al contesto di asserzione nel quale vengono impiegati.

Ciò suggerisce una ulteriore (ormai tradizionale) distinzione all'interno degli enunciati, ossia quella tra "tipi" di enunciati (*types*) ed occorrenze concrete degli stessi in particolari circostanze di uso (*tokens*).[58] Ogni enunciato – sia esso un "enunciato eterno" o meno – può esser considerato sia come "tipo" che come "occorrenza concreta". Per esempio, ognuno degli enunciati (a)-(i) può essere considerato sia nella sua dimensione "astratta" (come espressione sintatticamente ben formata della lingua italiana) che nella sua dimensione "concreta" (cioè nel suo uso in particolari circostanze). Gli enunciati che abbiamo caratterizzato come "eterni" risultano o veri o falsi in modo determinato *sia* nella loro dimensione astratta (come tipi), *sia* nella loro dimensione concreta (come occorrenze). Al contrario, gli altri enunciati risultano o veri o falsi in modo determinato *soltanto* nella loro dimensione concreta, ossia in riferimento a certe circostanze in cui vengono usati.[59]

Se dunque consideriamo i *tipi* di enunciati come portatori di verità, ne segue che vi sono enunciati eternamente veri (come (c), (e), (f)), enunciati eter-

[58] La distinzione tra tipi di enunciati e loro occorrenze concrete risulta dunque "trasversale" a quella tra enunciati "eterni" (sempre veri o sempre falsi) ed enunciati veri in certi contesti e falsi in altri e non va perciò confusa con quest'ultima suddivisione.

[59] Se questa idea della verità (falsità) di un *tipo* di enunciato appare discutibile o artificiosa, è possibile sostituirla con l'idea della verità (falsità) dell'occorrenza corrispondente in *ogni* possibile circostanza d'uso. È tuttavia indubbio che quando parliamo ci riferiamo alle espressioni (semplici o complesse) del nostro linguaggio sia nella loro dimensione astratta che nella loro dimensione concreta, e di entrambe predichiamo spesso delle proprietà. Si pensi ai seguenti enunciati: "La 'a' è la prima lettera dell'alfabeto italiano"; "Questa 'a' che hai scritto qui sembra una 'o'". Oppure, con riferimento agli enunciati, si pensi a: "La frase: 'La legge è uguale per tutti' è presente in ogni aula di tribunale"; "La frase scritta su quel muro è di inchiostro indelebile". Non si vedono ragioni stringenti del perché, accanto a molte altre proprietà, anche il vero e il falso non possano essere predicati degli enunciati tipo.

namente falsi (come (g)), ed enunciati eternamente né veri né falsi – come (a), (b), (d), (i)): ossia, quelli che acquistano un valore logico determinato soltanto in riferimento ad un contesto di asserzione; dunque come occorrenze concrete e *non* come tipi.[60] Si potrebbe obiettare che l'enunciato tipo (b) è falso finché Berlusconi non è presidente del Consiglio, diventa vero quando Berlusconi è presidente del Consiglio, e poi torna ad essere falso quando viene nominato un nuovo presidente.[61] Tuttavia tale posizione non sembra essere coerente. Infatti gli enunciati tipo, in quanto oggetti astratti, ideali, sono *fuori* dal tempo. Pertanto essi non possono possedere e non possedere una data proprietà (in questo caso la verità), pena contraddizione. Ciò ci fornisce anche l'occasione di precisare che quando parliamo di enunciati tipo veri o falsi "eternamente" quello che in realtà intendiamo esprimere mediante tale avverbio è il carattere atemporale o assoluto della loro verità o falsità. Al fine di escludere tra i portatori di verità enunciati tipo che non sono né veri né falsi, si può sostenere che – in linea di principio – ogni enunciato tipo di questo genere può essere "disambiguato", dando così luogo a una verità o una falsità assolute. Ciò potrebbe essere ottenuto eliminando le espressioni indicali presenti nell'enunciato in questione a favore di espressioni denotanti in modo univoco (ossia denotanti in modo indipendente da ogni contesto di asserzione). Se per esempio prendiamo l'enunciato tipo né vero né falso:

(a) "Mi fa male un dente",

otteniamo un numero *infinito* di enunciati tipo veri o falsi in modo assoluto se al pronome personale "mi" sostituiamo un nome proprio ed eliminiamo l'elemento indicale presente nel modo indicativo presente del verbo "fa" attraverso una precisa indicazione temporale; per esempio:

(a*) "A Giulio Cesare fa male un dente alle 13.45 del 5 luglio 40 a.C.";
(a**) "A Stalin fa male un dente alle 23.15 del 2 settembre 1940";
(a***) "A Karol Wojtyła fa male un dente alle 01.25 del 16 ottobre 1982";
[...]

[60] A ciò si possono aggiungere anche gli enunciati tipo che non sono né veri né falsi perché contengono termini vuoti (pur essendo privi di elementi indicali). Per esempio: "Il quadrato rotondo è una figura geometrica a due dimensioni". Se invece diamo di tali enunciati una lettura "esistenziale" al modo di Russell, essi risultano falsi.

[61] Questo è il modo in cui i filosofi antichi guardavano usualmente alle proposizioni (*λόγοι*) aventi un riferimento temporale. Si veda per esempio Aristotele, *Categorie*, 4a24.

In questo modo possiamo sostenere che i portatori di verità sono gli *enunciati tipo disambiguati* che, in quanto tali, sono veri o falsi in modo assoluto; e inoltre possiamo sostenere che a ogni fatto passato, presente, futuro risulta associato un corrispondente enunciato tipo vero e viceversa.[62] Va tuttavia rilevato che l'idea di potere in linea di principio disambiguare ogni enunciato tipo, sebbene intuitivamente plausibile, può presentare comunque delle difficoltà di carattere teorico. Siamo certi che enunciati come (a*)-(a***) siano del tutto privi di qualsiasi elemento indicale? In essi infatti facciamo riferimento ad istanti temporali che riteniamo essere identificabili in modo *assoluto*; tuttavia, se prestassimo attenzione alla teoria della relatività e a quanto ci dice la fisica contemporanea riguardo al tempo, potremmo nutrire seri dubbi a proposito del carattere assoluto della "retta temporale" cui implicitamente facciamo riferimento nell'asserire (a*)-(a***).[63]

In alternativa, possiamo sostenere che i portatori di verità sono le *singole occorrenze* (*tokens*) degli enunciati tipo usate in un determinato contesto. Tali occorrenze possono essere considerate, nella loro dimensione più materiale, come suoni che si propagano nell'aria, macchie di inchiostro, incisioni sulla pietra, fluorescenze su uno schermo ecc. Oppure, in modo un po' più raffinato, possiamo rappresentarci i *tokens* come funzioni che vanno da coppie formate da un enunciato tipo e un contesto di proferimento ai valori logici "vero" e "falso". In questa seconda accezione, l'occorrenza dell'enunciato tipo "Io ho fame" è una *funzione* che ha per argomenti tale enunciato tipo e un dato contesto di proferimento, e che assume come valore il "vero" se e solo se colui che proferisce tale enunciato tipo in tale contesto ha fame. In ogni caso, al di là di stabilire *che cosa* esattamente siano le singole occorrenze d'uso di un enunciato tipo, appare chiaro che esse hanno almeno due importanti caratteristiche, e cioè: (i) non possono mutare il loro valore logico, il quale è determinato in modo univoco dal contesto di proferimento; e (ii) sono evidentemente in numero finito – a differenza degli enunciati tipo che sono in numero infinito (è la stessa sintassi della lingua che non pone dei limiti al numero di nuovi enunciati che si possono costruire in linea di principio).[64]

---

[62] Ciò sotto l'assunzione (non banale) che la quantità dei fatti passati, presenti e futuri è un'infinità dello *stesso ordine* di quella degli enunciati tipo.

[63] Anche senza richiamarci alla teoria della relatività potremmo sollevare *altre* difficoltà a proposito della possibilità di disambiguare ogni enunciato. Tali difficoltà hanno a che fare, per esempio, con il carattere *ambiguo* dei nomi (comunemente) propri e con la *vaghezza* di molti predicati che usiamo.

[64] Anche le occasioni d'uso (*tokens*) possono essere prive di valore logico (né vere né false), nel caso in cui contengano termini vuoti, come abbiamo già rilevato nella nota precedente

Una terza ipotesi riguardante i portatori di verità è quella secondo cui essi non sarebbero gli enunciati (intesi come tipi o come occorrenze), bensì sarebbero ciò che viene espresso mediante gli enunciati: *proposizioni* o *contenuti*. In questa terza accezione i portatori di verità non sono dunque – propriamente – oggetti linguistici (stringhe di caratteri o suoni dotati di una loro forma, di una loro sintassi ecc.), bensì oggetti ideali, fuori dal tempo e dallo spazio, al pari delle entità matematiche.[65] La proposizione espressa (o il contenuto espresso) dall'enunciato:

(c) “L'acqua bolle a cento gradi centigradi”

è *che* l'acqua bolle a cento gradi centigradi. Il “che” in questi casi ci permette dunque di costruire *nomi* per proposizioni, ossia ci mette in grado di costruire espressioni linguistiche complesse che denotano proposizioni. Tuttavia, mentre enunciati come (c) esprimono sempre la *stessa* proposizione (vera o falsa in modo assoluto) in qualunque contesto vengano proferiti (o indipendentemente da qualunque contesto), enunciati come:

(a) “Mi fa male un dente”

esprimono proposizioni *diverse* in relazione ai diversi contesti nei quali vengono proferiti – ed esprimo una proposizione *soltanto* in relazione a *qualche* contesto di proferimento, ossia soltanto come occorrenze e non come enunciati tipo. Più precisamente, all'enunciato tipo (a) potrebbero corrispondere le seguenti occorrenze (indico tra parentesi quadre il contesto di proferimento dell'enunciato):[66]

(a*) “Mi fa male un dente [Giulio Cesare, Foro romano, 13.45 del 5 luglio 40 a.C.]”;
(a**) “Mi fa male un dente [Stalin, Cremlino, 23.15 del 2 settembre 1940]”;
(a***) “Mi fa male un dente [Karol Wojtyła, Vaticano, 01.25 del 16 ottobre 1982]”;

---

a proposito degli enunciati tipo. In tal caso avremmo una sorta di funzione parziale, ossia non definita per certe coppie di enunciati tipo e contesti di proferimento.

[65] Qui uso dunque le espressioni “proposizione” e “contenuto” in modo analogo all'uso (tecnico) che fa Frege dell'espressione “pensiero” (*Gedanke*).

[66] Chiaramente, eccetto forse Karol Wojtyła, nessuno degli altri due personaggi citati avrebbe proferito l'enunciato in questione in italiano; cio tuttavia è del tutto irrilevante per il punto che stiamo discutendo qui.

le quali esprimono, rispettivamente le seguenti proposizioni: *che* a Giulio Cesare fa male un dente alle 13.45 del 5 luglio 40 a.C.; *che* a Stalin fa male un dente alle 23.15 del 2 settembre 1940; *che* a Karol Wojtyła fa male un dente alle 01.25 del 16 ottobre 1982. Esse sono diverse proposizioni (diversi contenuti) aventi un loro valore di verità – o "vero" o "falso" – immutabile. Assumiamo inoltre che soltanto gli enunciati che siano o veri o falsi esprimono proposizioni; queste ultime saranno di conseguenza in ogni caso o vere o false (non vi sono proposizioni prive di valore di verità). Avendo caratterizzato le proposizioni come ciò che viene espresso dagli enunciati, come i loro contenuti, c'è da chiedersi se e come tali contenuti possano preesistere al proferimento effettivo degli enunciati (nel caso si tratti di occorrenze di enunciati tipo).[67] In questo contesto assumiamo che – conformemente alla loro natura astratta, ideale – le proposizioni espresse dalle varie occorrenze di enunciati tipo esistano anche *prima* del proferimento di tali enunciati (proferimento che ha sempre luogo in un determinato spazio e tempo).

Questi tre sembrano dunque essere i principali candidati al ruolo di portatori di verità: enunciati tipo disambiguati, occorrenze di enunciati tipo, proposizioni. Spesso vengono annoverati tra i portatori di verità anche le asserzioni e i giudizi; essi costituiscono gli *atti* mediante i quali ci si impegna nei confronti della verità di una data proposizione (o contenuto). Essi risultano dunque veri o falsi *perché* sono vere o false le proposizioni asserite o giudicate: quando la proposizione è vera (rispettivamente, falsa), la corrispondente asserzione risulta vera (rispettivamente, falsa); e ugualmente stanno le cose con i giudizi. Dunque, poiché la verità delle asserzioni e dei giudizi viene *spiegata* mediante la verità delle proposizioni, escludiamo che asserzioni e giudizi possano essere detti portatori di verità in senso proprio (o comunque in senso primario).

Alla luce delle caratterizzazioni date fin qui, vediamo adesso come le caratteristiche dei portatori di verità appena delineate possano avere rilevanza per il modo di porre il problema generale del rapporto fra bivalenza e determinismo. La questione del determinismo logico – ossia l'idea che certi principi logici abbiano implicazioni metafisiche in senso deterministico – si pone, abbiamo visto, allorquando consideriamo la verità o la falsità *attuale* di un dato portatore di verità che esprime un certo fatto futuro (in particolare, un fatto futuro contingente). Più precisamente, abbiamo visto che tale questione si pone nella misura in cui accettiamo la validità generale del principio di bivalenza e consideriamo delle istanze di esso in cui si fa riferimento: (i) al momento

[67] O all'atto di pensare o scrivere tali enunciati.

presente n; (ii) ad un dato portatore di verità *b* che esprime un fatto futuro contingente. Ossia, quando accettiamo la validità senza restrizioni di:

[BIV*] $V_n(b) \vee F_n(b)$.

Infatti, secondo diversi autori, la verità (rispettivamente, la falsità) attuale di un dato portatore che esprime un certo fatto futuro ci costringerebbe ad ammettere la necessità (rispettivamente, l'impossibilità) dell'accadere di tale fatto – secondo un'accezione di "necessità" come "inevitabilità". Risulta dunque chiaro che una condizione necessaria per *porre* il problema del determinismo logico consiste nel considerare la verità di un dato portatore in relazione ad un certo tempo. È dunque legittimo chiedersi *quali* siano i portatori per cui abbia senso porci il problema della loro verità o falsità in una prospettiva temporale.

Come abbiamo visto poco sopra, uno dei possibili candidati al ruolo di portatori di verità sono le proposizioni, intese come i contenuti espressi dagli enunciati. Abbiamo anche sottolineato il carattere immutabile di tali contenuti, che risultano o sempre veri o sempre falsi; volendo fare di nuovo un esempio, l'enunciato:

(a) "Mi fa male un dente $_{[x, y, z]}$"

proferito dal parlante x, nella circostanza y, al tempo z, esprime la proposizione che il parlante x ha mal di denti nella circostanza y, al tempo z. Ora, il contenuto espresso da (a), ossia la proposizione, *non muta* con il trascorrere del tempo,[68] poiché essa non ha al suo interno – se così possiamo dire – elementi indicali, a differenza dell'enunciato che la esprime, il quale possiede tali elementi (il pronome personale "mi", il tempo verbale indicativo presente). Di conseguenza tale proposizione possiede eternamente un valore di verità stabile. Se quindi assumiamo le proposizioni come portatori di verità, non ha senso parlare della verità o falsità di tali portatori in relazione *a un dato tempo*, bensì soltanto della loro verità o falsità *simpliciter*. Ma allora risulta scorretto, in relazione a tale genere di portatori, porre il problema del determinismo logico facendo riferimento alla loro verità o falsità *attuale*; ossia, porre tale problema in termini di [BIV*]. Non ha senso dire, rispetto ad una data proposizione che esprime un certo fatto futuro, che essa risulta vera o risulta falsa *prima* dell'accadere o meno di tale fatto, poiché il valore di verità delle proposizioni è una caratteristica che esse possiedono in modo atemporale. Più precisamen-

[68] Una volta che opportuni valori siano stati assegnati alle variabili x, y, z.

te, dato un enunciato E proferito al tempo n, supponiamo che esso esprima la proposizione che un certo fatto (futuro) F accade al tempo k > n, e sia p tale proposizione; abbiamo allora:

(i) p è vera$_{[simpliciter]}$ se e solo se al tempo k accade F;
(ii) p è falsa$_{[simpliciter]}$ se e solo se al tempo k non accade F.

Chiediamoci adesso per un attimo se è legittimo inferire da (i) e (ii) che al tempo n in cui l'enunciato E viene proferito risulta *predeterminato* l'accadere o meno del fatto F al tempo k, sulla base della verità o falsità di p. Si potrebbe sostenere che tale inferenza *non* è legittima, poiché non è la verità di p che determina l'accadere di F, né la falsità di p che determina il non accadere di F, bensì *viceversa*: le proposizioni sono vere o false *perché* i fatti stanno in un certo modo. Tuttavia si potrebbe obiettare che in quest'ottica non si riesce a spiegare come la proposizione p possa essere vera o falsa *prima* dell'accadere o meno del fatto F. Ma a ciò risponderemmo facendo semplicemente notare che – per come intendiamo le proposizioni – è *scorretto* parlare di "prima" o "dopo" in relazione ad esse, poiché le proposizioni possiedono un valore di verità in modo atemporale. In altri termini: data la proposizione p che esprime un dato fatto futuro contingente F, vale senza restrizioni:

[BIV] $V(p) \vee F(p)$

poiché quanto espresso da p necessariamente accadrà o non accadrà. Tuttavia è proprio l'accadere o meno del fatto F che rende p vera o che la rende falsa, *sebbene non la renda tale ad un determinato istante di tempo*, poiché le proposizioni non sono nel tempo. Il fatto F accade nel tempo, ma il suo rendere vera p oppure il suo renderla falsa non è un evento che ha luogo nel tempo. Tale proposta può suonare controintuitiva, ma se accettiamo il carattere ideale delle proposizioni non vedo come possa essere incoerente. Una prima conclusione parziale che possiamo dunque trarre riguardo al rapporto tra il problema del determinismo logico e quello dei portatori di verità è la seguente: se assumiamo che i portatori di verità siano le proposizioni intese come oggetti ideali, allora (i) non si vede un modo sensato di *porre* il problema del determinismo logico, il quale richiede di "indicizzare" la verità attraverso parametri temporali; e inoltre (ii) non si vedono ragioni stringenti per passare dall'idea di verità o falsità *simpliciter* di una data proposizione all'idea che il corso degli eventi sia interamente predeterminato.

Un discorso del tutto analogo con una analoga conclusione rispetto a quella appena vista vale anche se consideriamo gli enunciati tipo disambiguati come portatori di verità. In quanto oggetti astratti essi non hanno a che fare con la dimensione temporale; non ha dunque senso, rispetto ad un enunciato tipo, parlare della sua verità o falsità "al tempo t". Di conseguenza, anche in questo caso viene meno la stessa possibilità di porre il problema del determinismo logico nei termini del rapporto fra la verità *attuale* di certo un portatore e l'accadimento necessario del fatto in questione. Anche per quanto riguarda la relazione di "rendere vero" tra un certo fatto e un dato enunciato tipo valgono le considerazioni fatte sopra a proposito delle proposizioni: non è la verità di un enunciato tipo che determina l'accadere del fatto da esso espresso, bensì viceversa; e tuttavia non ha senso dire che tale enunciato tipo è vero *prima* dell'accadere del fatto in questione.

Diversamente stanno le cose con le singole occasioni di uso di un enunciato (*tokens*). Come abbiamo visto, l'uso effettivo che viene fatto di un dato enunciato tipo è un evento "concreto", spaziotemporalmente determinato. In relazione alle singole occorrenze di enunciati tipo *ha* dunque senso il parlare di "vero al tempo t". Se per esempio consideriamo la seguente occorrenza:

(a*) "Mi fa male un dente $_{\text{[Giulio Cesare, Foro romano, 13.45 del 5 luglio 40 a.C.]}}$"

e assumiamo che questo sia il genere di oggetti di cui si predicano il vero e il falso, allora siamo autorizzati a sostenere che (a*) è o vera o falsa *al momento* in cui viene proferita (o pensata, o scritta...). Inoltre, se (a*) è vera in quel momento, essa *resta* tale per l'eternità, e analogamente nel caso in cui sia falsa. Infatti, anche quando (a*) è passata, non si vedono ragioni forti per negare che possiamo continuare a predicare sensatamente di tale oggetto qualunque genere di proprietà, dunque anche il valore di verità che esso possedeva nel momento in cui è stato proferito. A ciò si potrebbe obiettare che un oggetto può possedere una qualsiasi proprietà soltanto nel momento in cui esso è presente, e dunque che anche una certa occasione d'uso di un dato enunciato può dirsi vera o falsa soltanto nell'arco di tempo in cui viene proferita (o pensata, o scritta...); al contrario, dopo tale tempo essa – non essendo più presente – non può più possedere alcuna proprietà, dunque non può neanche dirsi vera o falsa. Tale forma di "presentismo" (in breve, l'idea che soltanto ciò che è presente esiste) sembra tuttavia eccessiva, poiché trascura il fatto che generalmente per gli oggetti che sono passati continuiamo ad avere *criteri di identità* sufficientemente precisi al fine di attribuire o negare loro – al presente – certe proprietà. E ciò dipende almeno in parte dal nostro avvertire ciò che è passato come

– in qualche modo – “persistente” (ossia come *ancora* esistente): si pensi a tutte le circostanze in cui parliamo di eventi passati usando il tempo verbale presente. Assumeremo dunque che le occorrenze passate che erano vere o false al momento del proferimento mantengano per sempre il loro valore di verità, ossia che se una data occorrenza O è vera (rispettivamente, falsa) al tempo t, allora essa è vera (rispettivamente, falsa) ad ogni tempo s > t. Quindi, in riferimento alle occorrenze, ha senso predicare di esse la verità o la falsità al momento in cui vengono proferite e ad ogni momento successivo. Tuttavia non ha senso predicarne la verità o la falsità ad un istante di tempo *precedente* a quello in cui vengono proferite, per il semplice motivo che ad un tale istante esse non esistono ancora. La questione della verità o falsità delle occorrenze ha dunque un senso definito soltanto in relazione alle occorrenze presenti o passate.[69] Ciò comporta un elemento di asimmetria rispetto ai casi che abbiamo già discusso delle proposizioni e degli enunciati tipo. Ma l’insieme di tutte le occorrenze – presenti e passate – è certamente una quantità finita;[70] non solo, ma esso *rimane* sempre (ad ogni istante) una quantità finita, e di conseguenza tale insieme può esprimere al massimo una quantità finita di fatti. Dunque, sotto la ragionevole assunzione che il futuro contenga una quantità infinita di fatti,[71] possiamo concludere che *non per ogni fatto futuro esiste o è esistita un’occorrenza che esprime quel fatto*. Ne ricaviamo, quindi, che il principio:

[BIV*] $V_n(b) \vee F_n(b)$

riferito alle occasioni d’uso come portatori di verità risulta essere *inadeguato* ad esprimere la tesi del determinismo logico, stando alla quale *ogni* fatto futuro (rispetto all’istante attuale n) sarebbe determinato. L’inadeguatezza deriva

69 Si noti che qui non stiamo parlando di occorrenze che *si riferiscono* al presente o al passato, ma di occorrenze che *esistono* nel presente o che *sono esistite* in passato. Il tempo a cui tali occorrenze si riferiscono può indifferentemente essere precedente, contemporaneo, o successivo all’esistenza delle occorrenze stesse.

70 Le occorrenze avranno avuto origine, presumibilmente, con la nascita dei primi linguaggi umani, salvo non voler considerare l’ipotesi di una infinità di mondi popolati da esseri parlanti oppure l’ipotesi metafisica delle eventuali infinite occorrenze nella mente di Dio; ma in questa sede prescindiamo da ipotesi così “forti”.

71 Il “presentista” potrebbe sostenere che attualmente nessun fatto futuro esiste. In questo caso l’argomento del determinista logico sarebbe bloccato “sul nascere”, poiché verrebbe a mancare ciò che rende vere o false le occorrenze degli enunciati al futuro, cosicché risulterebbe illegittima l’applicazione ad esse del principio di bivalenza.

appunto dalla circostanza che la variabile enunciativa *b* può assumere, ad ogni istante di tempo, soltanto un numero finito di valori, mentre la quantità di fatti futuri che secondo il determinismo logico sarebbero attualmente già decisi sulla base di [BIV*] è una quantità infinita.[72] Un discorso del tutto analogo vale a proposito delle asserzioni o dei giudizi come portatori di verità. Anche in quest'ultimo caso si va incontro alla stessa difficoltà che abbiamo illustrato nel caso delle occorrenze d'uso, trattandosi sempre di quantità finite di portatori di verità, spaziotemporalmente determinate e come tali soggette a tutte le limitazioni di cui sopra.

Riassumendo, vediamo che una attenta analisi del problema dei portatori di verità ci induce a mettere in discussione il modo stesso in cui il problema del determinismo logico viene usualmente posto, che riposa in modo essenziale su certe relazioni – non sempre esplicitate – tra portatori, verità e tempo. Abbiamo visto che mettendo a fuoco ognuno dei più plausibili candidati al ruolo di portatori di verità, l'idea di derivare l'accadere necessario di ogni fatto futuro dal principio di bivalenza va incontro a delle difficoltà: nel caso delle proposizioni e degli enunciati tipo non si vede un modo sensato di temporalizzare la verità, mentre nel caso delle occasioni d'uso il principio di bivalenza risulta inadeguato agli scopi del determinismo logico. Personalmente, sono convinto che gran parte delle difficoltà interpretative che si incontrano nella lettura di testi filosofici sul determinismo logico (da quelli antichi a quelli contemporanei) nascano dal fatto che gli autori coinvolti non chiariscono mai – o solo raramente – qual è il genere di oggetti che diciamo essere veri o falsi; ciò comporta grossi problemi nel tentativo di ricostruzione degli argomenti che tali autori avanzano pro o contro il determinismo logico.

RIFERIMENTI BIBLIOGRAFICI

ARISTOTELE: *Categories*, *De Interpretatione*, trad. ingl. con note e glossario di J. L. Ackrill, Oxford: Clarendon Press, 1963.

——— : *Categorie*, *De Interpretatione*, in *Opere*, Roma-Bari: Laterza 1973, vol. I.

BECCHI, A. 2004: *Verità e determinismo: implicazioni metafisiche del principio di bivalenza*, Tesi di Dottorato, Università degli Studi di Firenze.

[72] Il determinista logico potrebbe sostenere che al momento presente risultano predeterminati tanti fatti futuri quante sono le occasioni d'uso (presenti o passate) esistenti fino a quel momento e concernenti tali fatti. Tuttavia questa limitazione della tesi determinista ad un numero finito di fatti futuri fa violenza alla natura stessa del determinismo logico, che (storicamente e teoricamente) pretende di essere una tesi generale concernente *ogni* fatto.

BECCHI, A., GUATELLI, F. 2000: “Di che cosa si predica la verità?”, *Kykéion*, 4, pp. 27-40.

BETTI, A. 2002: “The Incomplete Story of Łukasiewicz and Bivalence”, in T. Childers e O. Majer (a cura di), *The Logica Yearbook 2001*, Praha: Filosofia, pp. 21-36.

BOBZIEN, S. 1988: *Determinism and Freedom in Stoic Philosophy*, Oxford: Clarendon Press.

BRADLEY, R. D. 1959: “Must the Future Be What is Going to Be?”, *Mind*, 67, pp. 193-208.

CASARI, E. 1979 (a cura di): *Dalla logica alla metalogica. Scritti fondamentali di logica matematica*, Firenze: Sansoni.

CICERONE: *De Fato*, a cura di F. Antonini, Milano: Biblioteca Universale Rizzoli 1993.

DALLA CHIARA, M. L., GIUNTINI, R. 1999: *Łukasiewicz's Theory of Truth, from the Quantum Logical Point of View*, in J. Woleński ed E. Köhler (a cura di), *Alfred Tarski and the Vienna Circle*, Dordrecht: Kluwer, pp. 127-134.

DUGUNDJI, J. 1940: “Note on a Property of Matrices for Lewis and Langford’s Calculi of Propositions”, *The Journal of Symbolic Logic*, V, n. 4, pp. 150-151.

JORDAN, Z. 1963: “Logical Determinism”, *Notre Dame Journal of Formal Logic*, IV, 1, pp. 1-38.

KOTARBIŃSKI, T. 1913: “The Problem of the Existence of the Future”, trad. ingl. di R. Rand, *The Polish Review*, XIII, n. 3, 1968, pp. 7-22.

LEIBNIZ, G. W.: *Die Philosophische Schriften von Gottfried Wilhelm Leibniz*, 7 voll., a cura di C. I. Gerhardt, Berlin: Weidmannsche Buchhandlung, 1875-1890.

LEŚNIEWSKI, S. 1913: “Is all Truth Only True Eternally or Is It also True without a Beginning?”, in Id., *Collected Works*, Dordrecht: Kluwer, vol. I, pp. 86-114.

LEWIS, C. I., LANGFORD, H. C. 1959: *Symbolic Logic*, II ed., New York.

ŁUKASIEWICZ, J. 1906: “Analiza i konstrukcja pojęcia przyczyny (Analisi e costruzione del concetto di causa)”, *Przegląd Filozoficzny*, 9, pp. 105-179.

——— 1918: “Farewell Lecture by Professor Jan Łukasiewicz, delivered in the Warsaw University Lecture Hall on March 7, 1918”, in Łukasiewicz 1970, pp. 84-86.

——— 1920: “On Three-valued Logic”, in Łukasiewicz 1970, pp. 87-88 (trad. it. “Sulla logica trivalente”, in Casari, a cura di, 1979, pp. 213-214).

——— 1930: “Philosophische Bemerkungen zu den mehrwertigen Systemen des Aussagenkalküls”, *Monatshefte für Mathematik und Physik*, XXXVII, pp. 51-77 (trad. ingl. “Philosophical Remarks on Many-valued Systems of Propositional Logic”, in Łukasiewicz 1970, pp. 153-178; trad. it. “Osservazioni filosofiche sui sistemi polivalenti della logica proposizionale”, in Casari, a cura di, 1979, pp. 241-264).

——— 1946: “O Determinizmie”; trad. ingl. “On Determinism”, in Łukasiewicz 1970, pp. 110-130.

——— 1970: *Selected Works*, Amsterdam-London: North-Holland Publishing Company.

MINARI, P. 2003: “A Note on Łukasiewicz’s Three-valued Logic”, *Annali del Dipartimento di Filosofia* (Nuova Serie), Firenze University Press, pp. 163-190.

PRIOR, A. N. 1953: “Three-valued Logic and Future Contingents”, *Philosophical Quarterly*, III, n. 13, pp. 317-326.

WESSEL, H. 1976: *Logik und Philosophie*, Berlin: Logos Verlag.

WOLEŃSKI, J. 1989: *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.

——— 2003: *Determinism and Logic*, trad. ingl. non pubblicata di “Dieterminism i łogika”, *Voprosy Filosofii*, 5, pp. 71-81.

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 91-104

Howson on Novel Prediction

Robert G. Hudson
Department of Philosophy, University of Saskatchewan
(Saskatoon, Saskatchewan, Canada)
e-mail: r.hudson@usask.ca

1. Howson and Franklin's (1991) formalism
2. The Mendeleyev case
3. Howson's argument for predictivism
4. Conclusion

ABSTRACT. Bayesian discussions of the value of novel predictions have become moribund since the early 1990's. The last major discussion occurs in Howson and Urbach (1993), and the Bayesian position on novel prediction presented there is largely negative. Howson and Urbach (1993) contains a defense of predictivism utilizing an argument found in Howson (1984; 1990). However, I show this defense to be unsatisfactory. Instead, by deploying a formalism found in Howson and Franklin (1991), and despite their disavowal of predictivism in that paper, I demonstrate in Bayesian fashion the value of novel predictions. The elegance of this formalism bypasses the interesting but rather complicated, Bayesian defense on predictivism found in Maher (1988; 1990).

KEYWORDS: Bayesian confirmation theory, Colin Howson, Allan Franklin, predictivism, probability theory, novel predictions, Mendeleyev.

Colin Howson, in various publications (1984; 1990; Howson and Franklin 1991; Howson and Urbach 1993), addresses the question whether a hypothesis is better supported when it makes a novel prediction than when it accommodates evidence already known. His conclusions are as follows: 1) where one deliberately constructs a hypothesis to capture a certain piece of evidence, it is false to claim that, necessarily, such evidence does not support the hypothesis (this claim Howson 1990 calls the 'Null-Support Thesis'); and 2) al-

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

though in the usual case accommodated evidence supports a hypothesis just as well as predicted (synonymously, novel) evidence, there are nevertheless certain circumstances in which predicted evidence supports a hypothesis better. In this paper I am less interested in the accommodation thesis 1) than I am with the novel prediction thesis 2). In particular, I wish to claim that Howson's stated support for 2) is quite weak and has practically nothing to do with novel predictions; and that in criticizing Patrick Maher's (1988; 1990) support for the predictivist thesis, Howson along with Allan Franklin (1991) surprisingly creates an ideal formalism for defending 2), despite openly disavowing predictivism. Let me begin with the latter claim.

1. Howson and Franklin's (1991) formalism

In Howson and Franklin (1991), we are asked to compare two situations: first, a coin tossing case, and second, Mendeleyev's prediction of three new elements using his Periodic Table. Maher had first suggested this comparison in his (1988) and, though his formalism for handling these cases is interesting, I omit discussion of it here since I think Howson and Franklin (1991) offer a more approachable, straightforward formalism. We briefly review this formalism before turning to evaluating it.

In the first case, a subject predicts the outcomes of 100 tosses of a coin. e is the result of the first 99 tosses, h is the result of the 100^{th} toss plus e, and m is "the hypothesis that the subject has reliable advance information about the outcomes of the 100 flips of the coin" (Howson and Franklin 1991, p. 576), where such advance information might include such things as the precise bias of the coin, the subtle environmental conditions that could influence the flight of the coin, the relevant laws of physics that govern the coin's movement, and so on. By probability theory,

(1) $P(h/e) = P(h/e\&m)P(m/e) + P(h/e\&\text{-}m)P(\text{-}m/e)$.

The case is now sub-divided into two different scenarios: SCENARIO (A) the subject accurately predicts the outcomes of the first 99 flips without being informed about the results beforehand; SCENARIO (B) the subject is informed about the outcomes of the 99 flips beforehand.

SCENARIO (A): $P(h/e\&m) = 1$ (since if the subject has reliable advance information about the tosses, she will certainly know h). Thus,

(2) $P(h/e) = P(m/e) + P(h/e\&\text{-}m)P(\text{-}m/e)$.

Following Howson and Franklin, since $P(e/m) = 1$ and $P(e/\text{-}m)$ is very small, $P(m/e) \approx 1$, and so $P(h/e) \approx 1$; in all likelihood, *e* provides excellent support for *h*.

Analysis: Howson and Franklin here demonstrate that the subject's track record as an excellent predictor supplies confirmatory weight to the hypothesis that she correctly predicts the 100th toss. According to predictivism, this support should be better than the support provided for *h* in a case where we assume that the subject was told about the 99 tosses beforehand and was unable to make a successful prediction of such tosses otherwise. This is exactly what we find when we consider SCENARIO (B).

SCENARIO (B): Unfortunately, Howson and Franklin's discussion of this case is confused. They comment:

> the background information does not specify what the outcomes of the first 99 flips were, and so *m* does not entail *h* or *e* relative to that information (although *e&m* entails *h*). (1991, p. 576)

Yet, in their presentation of SCENARIO (A), there is no mention of what the background specifies as to the exact outcomes predicted by the subject – and still *m* is taken to entail *h* (and thus *e*). Indeed, given how Howson and Franklin define and use the symbols *m*, *h* and *e*, it does not matter whether the background information specifies what the outcomes are; *m* entails *h* (and so *e*), in any case, for given that the subject has reliable advance information about the outcomes of the 100 flips she will correctly predict *h*, whether in SCENARIO (A) or SCENARIO (B), that is, whether she was informed about the outcomes of the first 99 flips or not.

Thus, we have $P(h/e\&m) = 1$, and again (2), as above – an equation Howson and Franklin agree with (but I believe derive in a confusing way in their discussion of SCENARIO (B)). They then argue as follows:

> the probability of *e* conditional on *m* is now plausibly the same as its probability on *-m*. [Thus] [...] by Bayes' Theorem [...] $P(m/e) = P(m)$. (1991, p. 576)

But again given the meaning assigned to *m*, $P(e/m)$ is surely much larger than $P(e/\text{-}m)$. However, it is still the case than $P(m/e) = P(m)$ in the circumstances described in SCENARIO (B); for the fact that the subject correctly predicts *e* is irrelevant to the claim that she has advance reliable information about the outcomes of the 100 flips of the coin, given that she was informed about the outcomes of the first 99 flips beforehand. (And similarly for $-m$; accordingly, P($-m/e$) = $P(-m)$). Thus,

(3) $P(h/e) = P(m) + P(h/e\&-m)P(-m)$

which is the same equation Howson and Franklin arrive at, derived in more obvious fashion, I believe.

Given (3), the value of $P(h/e)$ will vary depending on $P(m)$. If $P(m) = 1$, $P(-m) = 0$ and so $P(h/e) = 1$. And if $P(-m) = 1$, $P(m) = 0$, and since $P(h/e\&-m) \approx .5$ (that is, the subject has a 50/50 chance of correctly predicting the result of the 100^{th} toss, given a complete lack of advance information about the tosses), $P(h/e) \approx .5$ as well. In brief, if $P(m)$ is high, so is $P(h/e)$; if $P(m)$ is low, $P(h/e)$ is about .5.

Analysis: It follows that the probative significance of evidence e in the 'accommodation' case (where the subject is informed about the outcome of the 99 flips beforehand) varies with the value of $P(m)$. Where $P(m)$ is low, $P(h/e) \approx .5$ in the accommodation case whereas $P(h/e) \approx 1$ in the prediction case. In other words, where $P(m)$ is low, predicted e better supports h than accommodated e. Conversely, where $P(m) \approx 1$, $P(h/e) \approx 1$ in both the prediction and the accommodation case. In other words, where $P(m) = 1$, there is no special confirmatory benefit in predicting rather than accommodating e.

What sense can we give to these results? It turns out that when the subject, to begin with, is unlikely to have advance knowledge about the coin tosses ($P(m)$ is low), she acquires significant confirmation for this claim by successfully predicting the first 99 tosses ($P(m/e) \approx 1$ with prediction). Conversely, she acquires no extra confirmation for the claim of advance knowledge if she is preinformed about the results about the tosses ($P(m/e) = P(m)$). In this way, the hypothesis that she correctly picks the outcome of the 100^{th} toss receives better confirmation in the prediction case than in the accommodation case. (If $P(m/e) \approx 1$, $P(h/e) \approx 1$; if $P(m/e)$ is low, $P(h/e) \approx .5$.) On the other hand, if the hypothesis of advance knowledge is maximally probable from the start ($P(m) = 1$), no extra confirmatory benefit is endowed on the hypothesis that the subject correctly discerns the 100^{th} toss when the subject is successful at predicting e as opposed to just accommodating e (for in either case, $P(h/e) = 1$).

My belief is that we have derived in a fairly straightforward Bayesian fashion the essential issue in adjudicating the benefit of prediction over accommodation. And Howson and Franklin apparently concur with this assessment in the coin-tossing case. Yet they claim that the scientific case, as exemplified by

Mendeleyev's prediction of three new chemical elements using his Periodic Table, is not amenable to the same sort of analysis. I believe their reasoning on this issue is flawed, a flaw connected to the confusion I cited above regarding their analysis of SCENARIO (B).

2. The Mendeleyev case

On the basis of his Periodic Table, Mendeleyev famously predicted the existence of three elements, scandium, gallium and germanium. The question is, does Mendeleyev's hypothesis that the third element, germanium, exists receive better support given that he correctly predicted the first two elements than if he had known about the two elements previously and simply accommodated their existence into the Periodic Table? If this case is analogous to the coin-tossing case above, then the answer here should be yes. However, Howson and Franklin do not think the cases to be analogous. They make the following comments: where *e* is the observation of the first two elements, and *h* is *e* plus the hypothesis that the third element is germanium,

> the Mendeleyev and the coin-tossing examples are not isomorphic precisely because there is no asymmetry in the Mendeleyev example in the relation between *m* and *e* in the predictive and accommodative cases. For what is *m* after all in the Mendeleyev example? It is the hypothesis that Mendeleyev's 'method' of prediction, that is, his theory of the Periodic Table, is true. *But m entails the truth of h and e independently of whether or not Mendeleyev had arrived at his theory before learning e*. Thus [...] the support of *h* by *e* should be independent of whether or not Mendeleyev had known about *e* before advancing his theory. (1991, p. 577; their italics)

It should be clear where Howson and Franklin go wrong in this quote. In their confused rendering of SCENARIO (B), discussed above, they claim that *m* does not entail *h* or *e*, and attribute the diminished support *e* provides for *h* in the accommodation case to this lack of entailment. However, I have argued that the asymmetry in the prediction and accommodation cases holds, *given that m entails h and e in both the prediction and accommodation cases*. Thus, the coin-tossing and 'science' cases are analogous, or at any rate Howson and Franklin have not shown how they are disanalogous.[1]

[1] As such, their final statement in their 1991, that the moral of their paper is that "science really is not coin-flipping" (p. 584), is gratuitous.

Now whether or not in some case prediction has value over accommodation depends, as I have noted, on $P(m)$, that is, on how likely the subject has reliable advance knowledge. For instance, in the Mendeleyev case, if from the perspective of the scientific community wherein Mendeleyev works the theory of the Periodic Table is thought to have a low probability of being true, it follows that Mendeleyev's hypothesis that the third element is germanium is better confirmed if he can predict the existence of the first two elements, scandium and gallium, than if he learns about them beforehand and accommodates them into his theory. On the other hand, if Mendeleyev's theory of the Periodic Table is considered plausible and likely to be true, there is no special benefit to prediction. These conclusions follow from the Bayesian analysis given above. However, Howson and Franklin provide exactly the opposite assessment. In response to Maher who provides a Bayesian defense of predictivism, they speculate on why Maher thinks prediction carries extra weight in the Mendeleyev case; "the answer", they submit,

> [is] very simple, and carries no comfort for Maher's view. It is that there was at the time no chemically plausible theory of the elements which explained [the original] 62 elements and e other than Mendeleyev's. This being the case, $P(e/\text{-}m)$ was small and $P(m/e)$ correspondingly large. [...] [We] feel secure in denying that the evidential force of the new evidence was due to the evidence's being *predicted*; rather it was due to the evidence's being *explained* by plausible theory. (1991, p. 578; their italics)

I find these comments entirely mysterious. Perhaps they are suggesting that Maher and other predictivists are fooled by a confirmation bias: a strongly held theory appears to get an extra confirmatory boost by making a prediction, just because it is a strongly held theory. To further support their anti-predictivist stance, they ask us to consider (a fictional) Mendeleyev who has formulated in his mind the Periodic Table prior to the discovery of the three elements. We suppose that these elements are discovered independently of Mendeleyev's knowledge of the Table. Mendeleyev then does two things: first, he publishes his Table, making clear to the scientific community how the Table coheres with the newly discovered elements (the non-predictive case). Then, after letting his first pronouncement settle in, he reveals that his Table had, in fact, anticipated the existence of these elements beforehand (the predictive case). From here Howson and Franklin make the following remarks:

> our thesis is that this autobiographical revelation should have made no difference to the evidential status of e. According to the predictivist thesis, on the

> other hand, it should have made every difference. But this seems absurd: the theory and the statement of the relevant physical facts are both independent of this information about Mendeleyev's personal history. (1991, pp. 578-579)

But there is no argument here. Instead of an argument, they provide an appeal to authority: they quote the particle physicist Yuval Ne'eman who makes the following comment about the successful prediction of omega minus by the eight-fold way:

> the importance attached to a successful prediction is associated with human psychology rather than scientific methodology. It would not have detracted at all from the effectiveness of the eightfold way if the omega minus had been discovered *before* the theory was proposed. (Quoted in Howson and Franklin 1991, pp. 579-580; Ne'eman's italics)

The last two quotes illustrate a potential confusion in discussions of novel predictions, one resulting from a failure to distinguish between two different kinds of novelty – temporal novelty and heuristic novelty. This distinction arose in the philosophic literature almost twenty years before Howson and Franklin published their paper (see Zahar 1973, p. 101; Musgrave 1974, p. 11), and it is a distinction that is glossed over by them in their paper. The difference between these two notions of novelty is this: evidence is *temporally novel* if no one had any knowledge of its occurrence prior to experiment or observation; and evidence is *heuristically novel* if it was not used in the construction of the hypothesis under test. Why this distinction matters is because one can be a predictivist and contend that heuristically novel evidence has extra confirmatory value, while still denying that temporally novel evidence has this value. That is, one can be a predictivist and *agree* with Howson and Franklin that an autobiographical revelation that Mendeleyev had not known of the elements before proposing his Table is irrelevant to the confirmatory worth of observations of these elements (so long as, had he known of the elements, this knowledge would not have played a role in his construction of the Table). And again, one can be a predictivist and *agree* with Ne'eman that it would not have detracted at all from the effectiveness of the eight-fold way if the omega minus had been discovered *before* the theory was proposed (so long as such an observation of omega minus would not have played a role in the construction of the eight-fold way).

Of course, Howson and Franklin might still contend that, whether understood temporally or heuristically, novelty is evidentially irrelevant. But a Bayesian analysis of the issue, using their own formalism (understood properly), leads to the opposite conclusion. Moreover, such a Bayesian analysis also leads to the surprising result that it is not the *plausibility* of the hypotheses supported that is at issue with prediction, but the *implausibility* of the hypothesis that the subject has reliable advance knowledge. For, as I have argued, where $P(m)$ is low $P(h/e) \approx .5$ in the accommodation case and $P(h/e) \approx 1$ in the prediction case, whereas $P(h/e) \approx 1$ in either case if $P(m)$ is high.

3. Howson's argument for predictivism

Despite denying that, in the usual case, prediction has special evidential value (and correlatively affirming that accommodation *does* have evidential value, contrary to what he takes some philosophers, such as John Worrall, seem to have claimed), Howson in various places (1984, pp. 249-250; 1990, pp. 236-237; Howson and Urbach 1993, pp. 411-412) allows there to be particular circumstances where prediction *does* have such special value. These are circumstances involving parameter adjustment with a suitably general hypothesis. We follow Howson and Urbach's 1993 presentation of this contention (though practically identical arguments occur in the other two cited papers). Consider a hypothesis h containing underdetermined parameters. Observation generates evidence e which suffices to fix these parameters, leading to the constructed hypothesis h'. However, we also suppose there is a hypothesis h'' that predicts e directly, without any parameter adjustment. Finally we assume that h and h'' have equal prior probabilities, and that e is logically implied by both h'' and h'. By probability theory, $P(h''/e) - P(h'') = P(h'')(1 - P(e))/P(e)$ and $P(h'/e) - P(h') = P(h')(1 - P(e))/P(e)$. Clearly, then, the level of support e provides for h'' as compared to h$'$ depends on $P(h'')$ and $P(h')$; but $P(h) > P(h')$ (h' is a special case of h, and so is less likely); and since $P(h'') = P(h)$ (by hypothesis), $P(h'') > P(h')$. Thus, under these circumstances, a hypothesis that predicts some evidence is better supported than a hypothesis that is constructed solely to capture this evidence.

This argument of course provides cold comfort for predictivists. The extra support garnered by h'' is not in fact the result of it having made a prediction. For if $P(h)$ were larger than $P(h'')$, large enough so that $P(h') > P(h'')$, it would follow that making a prediction decreases the support garnered from the evidence and, conversely, accommodation (parameter adjustment) increases this support. Thus, the benefit we have found with prediction here is not a benefit accruing to prediction *per se*, but to prediction given a certain prior distribution of probabilities; as such, it is a benefit that could just as well accrue to accommodation as to prediction.

Recall from above Howson and Franklin's discussion of the Mendeleyev case in which they claim that the support for a hypothesis increases with the prior plausibility of the hypothesis (again, this explains the greater support allotted to Mendeleyev's prediction of germanium). What we have just seen is a Bayesian formalization of this insight applied to the case of parameter adjustment. I claim we should adopt some caution at this stage if we take Bayesianism to provide a normative theory of evidential support, for it is surely excessively conservative to suppose that hypotheses about which we are confident are thereby, by that very reason, granted enhanced evidential support. Admittedly, if Bayesianism is understood as a descriptive theory of scientific rationality then Howson's result can be understood as capturing a typical though probably unfortunate feature of human reasoning, that we tend to see those hypotheses we are already sure about as better confirmed by evidence than hypotheses about which we are dubious. In other words, the common person is subject to a confirmation bias. But a descriptive theory is not what we are after here. So, confronting Howson, we have two problems: first, he hasn't adequately explained what evidential value there is in prediction *per se*; and secondly, his Bayesian analysis unduly legitimizes a conservative hypothesis-testing strategy.

However, with respect to the second objection, doesn't my approach to novelty commit a similar mistake? For I have said that the benefit of prediction over accommodation depends on the prior (im)plausibility of m, on the (im)plausibility of the hypothesis of advance knowledge. Again, if $P(m)$ is high, there is no benefit to prediction over accommodation, whereas there is such a benefit if $P(m)$ is low. How can I legitimately authorize this dependence of evidential support on the value of $P(m)$ in light of my criticism of Howson's view, according to which the level of evidential support depends on $P(h)$?

As a response, let me emphasize that there are some important differences between Howson's approach and mine. First of all, on my account, accommodation never generates *better* support than prediction (i.e., keeping the meanings of h, e and m fixed), whereas it sometimes will with Howson's approach (as we saw earlier). For either $P(m) < 1$ or $P(m) = 1$, and in the former case prediction has special evidential value whereas in the latter prediction and accommodation have the same value. In this regard, my account coheres better with our common understanding of the value of prediction: prediction, possibly, has an evidential advantage over accommodation, but not vice versa. Secondly, the benefit to prediction on my approach is not perpetual. As $P(m)$ approaches 1, prediction and accommodation come to have the same evidential value. Yet with Howson's strategy, given that $P(h'') \geq P(h)$, one could repeat his result again and again, using the same e. In each case, h'' still predicts e directly, and h' is still h with its parameters adjusted to accommodate e; moreover, as h'' is confirmed by e, the value of $P(h'')$ increases relative to $P(h)$, and so the beneficial, predictive effect is magnified. Of course, a halt to this iterative confirmation would stop if evidence e were to become 'old' – but Howson believes he has a response to Glymour's the old evidence problem (1990, p. 238ff), so this option is not open to him.

Finally, as we indicated earlier, the benefit of prediction on Howson's approach seems entirely spurious – it all rests on the values of the prior probabilities allotted to the hypotheses under test, and why these values should matter essentially to the benefit of prediction is left unexplained. On the other hand, my approach redirects focus from the plausibility of the hypotheses under test to the implausibility of m, the hypothesis that the subject has reliable advance knowledge about the observed phenomena under consideration. In the coin-tossing case, for example, m includes awareness of such things as the bias of the coin, the environmental conditions that could influence the flight of the coin and so on. In the Mendeleyev case, m stands for (following Howson and Franklin 1991, p. 577) Mendeleyev's theoretical understanding of the chemical elements, as expressed in the theory of the Periodic Table. On the approach I am suggesting the key is that by making a successful prediction, m is confirmed (that is, the posterior probability of m increases). For example, the coin-tosser's successful prediction of the first 99 tosses confirms the claim that she has precise knowledge about the coin, the relevant ambient environmental conditions, and other relevant factors. Similarly, Mendeleyev's prediction of the first two elements demonstrates his understanding of the chemical elements in the context of the Periodic Table. In both these cases, a successful prediction as opposed to an accommodation demonstrates the predictor's possession of knowledge regarding the observed phenomenon – as we have symbolized this fact, $P(m/e) \approx 1$.[2] And it is this demonstration that leads us to allot enhanced support to the subject's future predictions. Thus, I have not only shown there to be a link between making a prediction and enhanced support for a hypothesis in probabilistic terms; I have provided a link that makes intuitive sense of what value there might be in prediction. That is, if there is any value to prediction, I think my account is much better at explaining this value than Howson's 'parameter adjustment' account.

[2] In the accommodation case the subject may also possess such knowledge, but this fact is not *demonstrated* by having accommodated past results.

Nonetheless, one might raise the following objection to my approach: I have, as one critic puts it, "conflated and confused credit for the scientist with confirmation of the hypothesis". The ability to make successful predictions, supposedly, only exhibits a cognitive virtue in the scientist making the prediction (this is perhaps one way of interpreting an increase in $P(m/e)$) but says nothing, or should say nothing, about the degree of confirmation the scientist's hypothesis receives. However, I think it is arguable that by demonstrating a scientist's cognitive virtue, one thereby provides extra support for this scientist's predictions. For it is true that in assessing the quality and relevance of observational evidence we often refer to the intelligence and integrity of the individual producing this evidence. Evidence in support of a hypothesis generated by someone who is deadly honest is more probative than similar evidence produced by a liar; here, surely, credit leads to confirmation. Moreover, consider the further important feature of my Bayesian analysis, that where $P(m) =1$ to begin with there is *no* benefit with prediction over accommodation. In other words, if we interpret m as symbolizing a scientist's 'credit', we have put a point to the above critic's complaint regarding "conflating and confusing credit [...] with confirmation": sometimes credit for the scientist does have confirmatory impact and sometime it does not – what my account does is to provide a way of adjudicating this issue. In particular, where a scientist's credit is firm there is no further boon to her credit by making a successful prediction, and so on my analysis there is no enhanced confirmation for her hypotheses. Conversely, where a scientist's credit is under dispute, making a successful prediction *does* enhance her credit and so does provide extra support for her hypotheses.

Yet, I believe, the 'confusing confirmation with credit' issue is a complete red herring. For m is not the hypothesis that the scientist is 'virtuous' or 'has epistemic credit' or 'is an authority'. Rather, m is a set of claims to knowledge – it is a set of hypotheses made by the subject about the phenomena under inspection. For example, in the coin-tossing case, it expresses the subject's detailed understanding of the nature of the coin, the ambient environment, and so on. Accordingly, by successfully predicting the first 99 tosses, the coin-tosser's possession of such understanding is confirmed. Similarly, by successfully predicting the first two elements, scandium and gallium, Mendeleyev confirms his hypothesis of the theory of the Periodic Table. In each case, we represent probabilistically the effect of this confirmation by assigning a value of 1 to $P(m/e)$. An extra, confirmatory effect on h is then inferred by deducing that $P(h/e) = 1$. Along these lines I believe we are able to explain the benefit accruing to a successful prediction, without mentioning at all the 'virtue' of the scientist, her 'epistemic credit', or whatever. On my approach, credit is not "conflated and confused with confirmation" simply because credit never makes an appearance as an issue at all.

4. Conclusion

The Bayesian formalism I am recommending (which I believe captures the evidential significance of prediction over accommodation) has a precedent in the non-technical literature, to wit, in the work of John Worrall who is a long-time defender of heuristic novelty as a valuable criterion for assessing the worth of evidence. I want to conclude by invoking his approach since he is one of Howson's prime targets as an example of someone who has a faulty understanding of novel evidence. (See, e.g., Howson 1984, p. 250; 1990, pp. 225, 236-237.)

Worrall in his (1985) discusses a case of what one might presume to be a regrettable, ad hoc adjustment of a theory to accommodate evidence. The case involves the discovery by the 19th century optical theorist, Thomas Young, that light is subject to the principle of interference. Subsequent to his discovery, Young was confronted with the problem of explaining why two candles near one another did not produce illumination exhibiting light and dark fringes, as predicted by the interference principle. He responded by constructing a hypothesis: his interference principle, he claimed, applies only when "the two interfering 'portions' of light originate in the same source" (Worrall 1985, p. 311). Now, as Howson understands Worrall's position, such an ad hoc maneuver is completely unacceptable. To illustrate this reading of Worrall, he quotes the following passage from Worrall (1978): "of the empirically accepted logical consequences of a theory those, and only those, used in the construction of the theory fail to count in its support" (Howson 1984, p. 250; 1990, p. 233; originally in Worrall 1978, p. 48). In other words, on Howson's view, Worrall is a supporter of the 'Null-Support Thesis', the thesis that where one deliberately constructs a hypothesis to capture a certain piece of evidence, such evidence, necessarily, does not support the hypothesis; as such, Worrall is presumably compelled to reject Young's ad hoc theoretical maneuver. Conversely, Howson rejects the Null-Support Thesis: for him, where one deliberately constructs a hypothesis to capture a certain piece of evidence, it is false to claim that, necessarily, such evidence does not support the hypothesis. So on Howson's account and apparently contra Worrall, it is possible for Young' adjusted interference principle to receive support from the absence of an interference pattern found with two candles.

However, Worrall's approach is more subtle than Howson thinks. On Worrall's view, are we compelled to discard Young's constructed hypothesis without further notice? Not as such, for as he makes clear what actually went wrong with Young's construction is that "he failed to give the scientific community of his time any reason for the restriction of his principle to the case of 'portions' of light originating from the same source" (Worrall 1985, p. 321) – any reason, that is, other than the fact that this hypothesis 'saves the phenomena'. In other words, there is a tacit acknowledgement here by Worrall that if Young *had* shown the plausibility of the assumption he was using, that is, if he had provided some reason for thinking that interference only applies when "the two interfering 'portions' of light originate from the same source", then Young's maneuver would have been acceptable and his interference principle vindicated. As it happens, Young provided no such defense. By contrast (as Worrall points out) Augustin Fresnel did have a story that could account for the non-occurrence of interference fringes in the two candle case. Fresnel argued that the relevant, two-candle interference patterns would be continuously changing, indeed changing so fast that our eyes are unable to pickup these patterns, thereby creating the illusion of constant illumination. Fresnel's argument, because it is theoretically motivated and not simply ad hoc – that is, not thought up simply to save the phenomena – partially accounts on Worrall's view for the more favorable reception accorded to Fresnel's wave theory over Young's.

Consequently, the moral I think we should take from Worrall's work (*pace* its misrepresentation by Howson) is that when evaluating a hypothesis on the basis of evidence one needs to consider closely the plausibility of the auxiliary assumptions used in connecting the evidence to the hypothesis. Where these assumptions are highly probable, the question whether a hypothesis predicts evidence or is constructed using this evidence is irrelevant. To put this point in terms of the formalism provided earlier, here letting m stand for these auxiliary assumptions, we found that where $P(m) = 1$ there is no benefit to be found in prediction as opposed to accommodation. My contribution in this paper, then, is to go one step further: I am claiming that where such auxiliaries have *low* initial probability, then in fact we do have a benefit with prediction over accommodation. For it is by making a successful prediction that one confirms these less-than-certain auxiliaries, a confirmation which thereby (as we have demonstrated probabilistically) leads to heightened support for the hypothesis given the evidence.

REFERENCES

HOWSON, C. 1984: "Bayesianism and Support By Novel Facts", *British Journal for the Philosophy of Science*, 35, pp. 245-251.

———— 1990: "Fitting Your Theory to the Facts: Probably Not Such a Bad Thing After All", in C. W. Savage (ed.), *Scientific Theories*, Minneapolis: University of Minnesota Press.

HOWSON, C. and URBACH, P. 1993: *Scientific Reasoning: The Bayesian Approach*, 2nd Edition, Chicago: Open Court.

HOWSON, C. and FRANKLIN, A. 1991: "Maher, Mendeleev and Bayesianism", *Philosophy of Science*, 58, pp. 574-585.

MAHER, P. 1988: "Prediction, Accommodation and the Logic of Discovery", *PSA 1988*, vol. 1, pp. 273-285.

———— 1990: "How Prediction Enhances Confirmation", in M. Dunn and A. Gupta (eds.), *Truth or Consequences*, Dordrecht: Kluwer Academic Publishers.

MUSGRAVE, A. 1974: "Logical vs. Historical Theories of Confirmation", *British Journal for the Philosophy of Science*, 25, pp. 1-23.

WORRALL, J. 1978: "The Ways in Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology", in G. Radnitzky and G. Anderson (eds.), *Progress and Rationality in Science*, Dordrecht: Reidel.

———— 1985: "Scientific Discovery and Theory-Confirmation", in Joseph C. Pitt (ed.), *Change and Progress in Modern Science*, Dordrecht: Reidel.

ZAHAR, E. 1973: "Why Did Einstein's Programme Supersede Lorentz'?", *British Journal for the Philosophy of Science* 24, pp. 95-123.

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 105-121

Comparing Two Views of Comparison: Peña and Casari on Vagueness and Comparatives

Francesco Paoli
Department of Education, University of Cagliari (Italy)
e-mail: paoli@unica.it
website: http://www.unica.it/~paoli/index.htm

1. Introduction
2. The shared general framework
3. Vagueness, fuzziness, gradability
4. The rules of endorsement and maximality
5. The structure of truth degrees
6. The semantics of logical connectives
7. The sorites
8. Conclusion

ABSTRACT. This paper aims at a comparative assessment of two degree-theoretical views of vagueness and comparison – Ettore Casari's comparative logic and Lorenzo Peña's transitive logic. Although both approaches cope better than most rival theories with the thorny challenges posed by such issues, in the author's opinion Casari's perspective seems superior under a number of respects.

KEYWORDS: vagueness; comparatives; sorites paradox; comparative logic; transitive logic; Casari, Ettore; Peña, Lorenzo.

1. Introduction

After having been working for a number of years on Casari's comparative logic and its applications to the issues of vagueness and comparison, I have recently come across a closely related approach to these problems which has been developed over the last decades, unbeknownst to me, by the Spanish philosopher Lorenzo Peña and his coworkers. Both research streams started at

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

about the same time (the beginnings of Peña's work on transitive logic date back to the mid-seventies, while Casari's earliest published article on comparative logic was originally presented at a conference in 1979; see Casari 1981) but have hitherto evolved – to the best of my knowledge and as far as I can infer from the lack of mutual references in the writings of both schools – in a completely independent way. Yet, besides sharing a common underlying view of the phenomena under investigation, these approaches seem to show striking similarities both in their general philosophical outlooks and in their technical underpinnings, as well as a number of differences. It seems therefore useful to have a closer look at such similarities and differences. The aim of the present paper will be, if you let me put it this way, to compare the above-mentioned views of comparison.

I will not try to give a self-contained account either of comparative logic or of transitive logic, nor will I aim at a detailed exposition either of Casari's or of Peña's views on gradability and comparison: the interested reader should consult Casari (1989; 1997), Paoli (1999; 2003), Peña (1984; 1987a; 1987b; 1990; 1992; 1993; 1995; 1996), and Vásconez, Peña (1996) for more systematic information of this kind. Rather, I will discuss in a quite haphazard way some specific philosophical features of both approaches, trying to single out where they agree with each other and where they are at variance. Moreover, the reader is warned that my discussion will not be symmetric and neutral, but rather biased in favour of comparative logic (the term "bias" is taken here, of course, in its positive meaning of a rationally and critically supported inclination, which must not be confused with prejudice!). Due to this, I will mainly focus on a series of remarks on – often also objections to – Peña's approach as seen against the background provided by Casari's system.

Finally, let me frankly confess that my knowledge of the work of Peña and his disciples is utterly partial, as it mainly results from the items referenced in the bibliography below, for many of which I have been kindly provided with copies by the author himself, and from some stimulating conversations with Peña's coworkers Marcelo Vásconez and Txetxu Ausín (whom, by the way, I thank for stirring my interest into transitive logic and its underlying philosophy). Whether in Peña's extensive bibliography there are further writings which could substantially modify the discussion provided below, I do not dare to say.

2. The shared general framework

For a start, I will try to underscore the most apparent similarities between the above-mentioned perspectives. Before doing this, however, some terminological remarks are in order. Whenever one's talk is about vagueness lexical quarrels are quite likely to arise, since the very definition of such a term is a contentious matter. In particular, it seems as though authors belonging to the streams discussed here use this term in different meanings. At the present stage, however, I propose to stay with a rather inexpressive definition: let us call "vagueness" the phenomenon which gives rise to slippery slope arguments of soritical type, whatever its nature may be. We will see below to what extent the comparative-logical and the transitive-logical approaches make different uses of this word; until then, I beg the reader to accept my verbal convention.

The main similarities between both approaches can be summarized as follows:

A) Ontological view of vagueness. According to both perspectives, vagueness is neither an epistemic phenomenon of ignorance, as claimed by Williamson and other epistemicists, nor a semantic phenomenon of ambiguity, as maintained by supervaluationists like Fine; rather, it is an *ontological* phenomenon. Each vague predicate does not admit of borderline cases of application because it denotes a sharp property of whose extension we are in principle ignorant, or because it ambiguosly refers to several sharp properties, but because the unique property it refers to is itself *fuzzy*, in that it applies to some objects only to some extent. This stance is explicitly upheld in the writings of Peña:

> La aplicación de predicados difusos no se debe a alguna aberración de nuestro pensamiento o de nuestro lenguaje, sino que está basada en el carácter objetivamente difuso de ciertos cúmulos o propiedades, a saber aquellos que abarcan a alguno de sus respectivos miembros en una medida no total (1996, p. 146).

Casari is not just as outspoken on this aspect, but I think that the ontological view of vagueness is the one which best accords with his degree-theoretical approach to the issue.

B) Degree-theoretical attitude. According to both perspectives, thus, the possession of a property by an object, or the membership of an element in a set, is not an all-or-nothing issue, but a matter of *degree*. Degrees of memberships are paralleled, on the level of sentences, by *degrees of truth*: if it is a matter of degree whether or not John belongs to the set of tall people, it cannot be but a matter of degree whether or not the sentence "John is tall" is true. Also, a proper logical treatment of such kind of sentences demands that degrees of truth be *infinitely* *many*. Of these degrees, some will be positive (true) and some negative (false) – and here similarities on this aspect come to a sudden stop, as we will see.

C) Reductionist approach to comparatives. An adequate theory of adjectival comparison must be wide-ranging enough as to make room at least for cross-comparisons of the form

(1) *x* is at most as *P* as *y* is *Q*,

and the likes where "at most as... as" is replaced by "at least as... as", "more... than", "less... than". But this can be accomplished, together with a reduction of comparative adjectives to the corresponding positive adjectives, by resorting to the apparatus of truth degrees. In fact, the following *reduction principle* is endorsed within both perspectives:

(2) A sentence of the form specified in (1) is true iff "*x* is *P*" implies "*y* is *Q*"

where implication is an inherently comparative notion whose behaviour is governed by the following *simplification principle*:

(3) *A* implies *B* iff the truth degree of *A* is smaller than or equal to the truth degree of *B*.[1]

(D) Real fuzziness principle. All the features discussed under the previous headings are common not only to Casari's and Peña's approaches, but virtually to any fuzzy, degree-theoretical perspective on vagueness and comparison (where "fuzzy" is meant here in a broad, non-technical sense). Our two approaches, however, are quite serious about fuzziness, which is really taken at face value. Remember that, according to the received view on the subject, vague predicates are *tolerant* – in other words, it is possible for *x* and *y* to differ (if only very slightly) in their degrees of possession of the property *P* while the truth values of the sentences "*x* is *P*" and "*y* is *P*" do not differ at all (Wright 1975). As I illustrated in Paoli (2003), the standard fuzzy approaches, like the one based on Łukasiewicz logic, reject this assumption within the area of borderline cases, but not outside it – i.e., as far as definitely positive or negative *P*-cases are concerned. A really fuzzy semantics of

[1] A principle encompassing our simplification and reduction principles is called *principio de desincrustación* in Peña (1987b, p. 342).

vagueness and comparison, instead, must take into account all the relevant differences:

> Any difference, however small, in the possession of a property by an object does affect the justice with which the corresponding predicate can be applied to it (Vásconez 2002, p. 39).

3. Vagueness, fuzziness, gradability

So much for common ground. Let us now come to review the main points of disagreement between the theories under scrutiny.

We hinted above at a terminological discrepancy concerning the term "vagueness". According to Peña, in fact, vagueness is a pragmatic phenomenon that plays no role in soritical paradoxes – where it is the property of *gradability*, or of coming in degrees, which is really at issue:

> Vagueness is not the same as graduality. A statement may be vague because of pragmatic considerations concerning the context of utterance: the utterer is expected to convey more specific information and instead he contents himself with general remarks [...]. Terms can be said to be vague in so far as they are used in vague statements. Yet on their own they are not vague. "Tall" is not vague. It is a term denoting a property which comes in degrees. The issue is not vagueness but graduality (1993, pp. 403-404).

About such a pronouncement I have two comments to do. On the one hand, this usage of the word "vague" does not tally with the established usage to be found in substantial portions of the current scientific literature on the argument, where the phenomenon just described by Peña – failure to provide information which is specific enough – is labelled *generality* and is sharply distinguished from vagueness (see e.g. Sorensen 2002; Tye 1994). It must be said that Peña advances some appealing arguments to the effect that his terminology is more consistent with the everyday meaning of the term (see e.g. Peña 1996; Vásconez, Peña 1996); however, I believe that consistency with established philosophical use gives me at least some right to employ the word "vagueness" whenever Peña would resort to "fuzziness" (which I prefer to avoid in this case since it is by now too loaded with technical connotations: over the years, it has come to refer to a *particular theoretical approach* to vagueness, rather than to vagueness in itself).[2]

[2] At the beginning of Section 5 I will expand a little bit on this issue.

On the other hand, and apart from such terminological controversies, I have a more conceptual kind of qualm regarding the identification of vagueness (Peña's "fuzziness") with gradability. I argued elsewhere (Paoli 1999) that gradability is a necessary, but not a sufficient condition for vagueness.[3] In fact, there are predicates – like "acid" as a predicate of chemical substances, or "acute" as a predicate of angles – which seem to come in degrees and admit a nontrivial comparative: it seems perfectly all right to say, e.g. that a substance with PH3 is more acid than a substance with PH5, or that a 25° angle is more acute than a 65° one. Nonetheless, these predicates are not vague at all and can by no means give rise to sorites of any kind, because there are sharp cut-off points separating their domains of application from the domains of their antonyms (in our examples: PH7 for "acid", 90° for "acute").[4]

Examples like these lead us to surmise that vagueness is not just gradability, but a *special* kind of gradability, arising whenever a given predicate P admits of borderline cases of application. Casari's semantics – where one allows for *intermediate* degrees of truth, which are neither definitely true nor definitely false, as well as for degrees of definite truth and falsity – yields a satisfactory treatment of the distinction between gradability and vagueness. Generally speaking, a one-place property can be identified with a function from individuals of an appropriate domain to degrees of truth. In particular, a sharp gradable property (like acidity) is a function which takes up only positive ("true") or negative ("false") degrees, but never intermediate ones; whereas a vague property is a function which can take up positive, as well as negative and intermediate degrees. Whether such a distinction can be reproduced in the semantics for transitive logics, it is unclear to me.

4. The rules of endorsement and maximality

Although several different systems coexist in the family of transitive logics, they all can be endowed with an infinite characteristic matrix whose set of truth values is the closed real unit interval [0,1] and whose set of designated values is the semiopen interval]0,1] – i.e., the whole set of truth degrees ex-

[3] Other authors, like Bierwisch (1989), claim that it is not even a necessary condition, but their contentions need not concern us here.

[4] Interestingly enough, Engel – an author quoted by Peña as being somewhat sympathetic with his approach – discusses such examples (Engel 1989), to which however Peña does not seem to pay special attention.

cept 0, absolute falsity (differences among logics in the family mainly amount, if I am not mistaken, to differences in expressive power). Degrees in the open interval]0,1[– all the truth degrees except absolute truth and absolute falsity – are to some extent both true and false. Also, Peña espouses a rather uncontroversial (in degree-theoretical terms) semantics for negation, according to which the amount of truth in the negation of *A* increases as the amount of falsity in *A* increases, and vice versa. Examples from natural language clearly show that intermediate degrees are inhabited, i.e. there are sentences that can be assigned such degrees – hence, also their negations will be assigned intermediate truth degrees. Given the above choice of designated degrees, this means that there must be true contradictions (where it is essential to remark that "true", here, means "true to some extent").

From a philosophical viewpoint, this somewhat unusual selection of designated values (unusual at least as far as fuzzy logics are concerned) is bolstered by an assumption which Peña terms *rule of endorsement* (*regla de apencamiento*) and which is formulated and justified in the following terms:

> La *regla de apencamiento* nos permite pasar de "*p* es en alguna medida verdadero" a la conclusion de que "*p* es verdadero". Esta regla ha venido reconocida como legitima en la tradición logica y filosofica. La base de ese reconocimiento es que no puede ocurrir que la conclusion sea totalmente falsa, en el supuesto de que la premisa sea, en uno u otro grado, verdadera (Vásconez, Peña 1996).
>
> Since what is required in order for a statement to be right is nothing else but its being true (just true, without further qualifications), and any statement is (to some extent or other, however small) true unless it's wholly untrue, we can safely state any sentence provided we are convinced it's not altogether false (Peña 1984).

Upholding the rule of endorsement goes hand in hand with refusing the incompatible assumption named *rule of maximality*:

> What definitely must be waived is the maximality rule, viz.: $p \vdash Hp$.[5] [...] The purported rationale for it is that nothing can be rightly asseverated unless it's quite true, true without mixture of falsity. So, "true"' *tout court* would be equivalent to "utterly true". But to my mind such a reason doesn't carry conviction.

[5] "*Hp*" means "it is definitely the case that *p*" in Peña's notation, and is true just in case $v(p) = 1$.

> For, when we assert some sentence, we to be sure regard it as true; but why on earth should we regard it as wholly true? When I say that I'm hungy, I'm not saying that I'm altogether hungry, but just hungry [...]. The maximality rule thus lends to relinquishing what is of importance in fuzzy logic, namely: truth-nuances, in virtue of which not all that is true is entirely so (Peña 1984).

The standpoint of comparative logic is, here, completely different. Not only is the choice of the interval [0,1] as a system of truth degrees rejected, for reasons to be seen in the next Sections, but also contradictory degrees and the rule of endorsement are rebutted, while a properly understood maximality rule is accepted. In fact, comparative logic acknowledges that some sentences containing vague predicates can be approximately true and at the same time approximately false, but equates truth *tout court* not with approximate truth, but with *definite* truth. The important difference with respect to the standard fuzzy approaches, which permits a satisfactory account of vagueness and comparison, is the idea that *not only approximate truth (falsity), but also definite truth and definite falsity come in degrees*.

In my view, it is important to keep in mind some conceptual distinctions that are somehow overlooked by Peña. In particular, we must distinguish two senses of "possessing a property" and just as many senses of "definite truth". "Possessing a property *P*" can mean either of two different things for an individual *a*:

(a) That *a* has *P* to some non-null, however small, degree;
(b) That *a* has *P* to as high a degree as to warrant applicability of the predicate "*P*" to the name "*a*". It seems to me reasonable enough to maintain that the sentence *Pa* must be considered true just in case *a* has *P* in this second sense.

Likewise, "being definitely true" can mean either of two different things for a sentence *A*:

(c) That *A* is true "without mixture of falsity", i.e. it is not even approximately false;
(d) That no sentence *B* can be strictly truer than *A*.

In my opinion, the rule of endorsement covertly presupposes the identity of (a) of (b); but this causes no end of trouble, as the next example will show. On the other side, I concede that the rule of maximality is blatantly invalid if "definitely true" is taken in sense (d) – more than that, I even deny that such a notion makes any sense at all – yet think that there is nothing to blame in it if "definitely true" is correctly understood in sense (c).

With an eye on these distinctions, I will now argue that accepting the rule of endorsement would lead us in any case to unpalatable conclusions: that is, either to a clash with ordinary linguistic usage and with some naive intuitions concerning the meaning of predicates, or else to a violation of the real fuzziness principle and the semantics of comparison, i.e. of what we identified as two of the strong points of both Casari's and Peña's approaches.

Let John be an European adult who is just 1.20m tall. Should we say that John possesses to some extent the property of tallness? In other words, should we say that it is to some extent true that John is tall? Either way we go, we are in trouble. Suppose we do. Then, by the rule of endorsement, it is true that John is tall. Well, this would seem to contradict linguistic practice: nobody in his right mind would agree that John is tall (for an European adult). If the envisaged measure of 1.20 still seems controversial, just tweak the example to taste replacing it with as small a measure as you wish.

On the other hand, one might say that John completely lacks the property of tallness – after all, he's not even remotely a borderline case for a tall person. But this is at odds with our beloved real fuzziness principle, as well as with the account of comparison seen above. For suppose that Bill is 1.18m tall. Bill is strictly less tall than John, and thus should possess the property of tallness to a strictly smaller degree than John does. But how can that be, if John completely lacks the property?

Comparative logic provides a cheap way out from such puzzles. Under the given circumstances, it would seem reasonable to most of us to say that "John is tall" is definitely false, i.e. false and not even approximately true. However, since definite falsity also comes in degrees, "Bill is tall" will have a strictly falser degree of truth. This seems to reconcile our everyday understanding of sentences like the above with the principle according to which even minute differences in the possession of an attribute should be mirrored in one's semantics.

5. The structure of truth degrees

Comparative logic is not a fuzzy logic in the technical, narrow sense of the word, since it has no characteristic matrix semantics on the real unit interval [0,1]. I believe that the choice of such interval as a system of truth degrees carries two main shortcomings: this set, in fact, is *bounded* and *linearly ordered*.

I am going to be very cursory as regards these questions, since I discussed them rather extensively elsewhere (Paoli 1999; 2003). I just recall that a bounded set of truth values does not allow a satisfactory reconstruction of comparative sentences at least in three cases: when cases of definite applicability of a predicate are at issue, where comparative constructions are nested, and when the predicates involved are gradable but sharp. On the other hand, linear ordering seems too rigid a constraint for comparisons involving evaluative predicates and semantically anomalous comparative sentences.

To be sure, Peña occasionally seems to acknowledge the insufficiency of the linearity constraint (see e.g. Peña 1995). His hints at a "tensorial semantics" could be seen as a first attempt to tackle the issues just alluded to.

6. The semantics of logical connectives

So far we have not been concerned with the logical structure of sentences as respectively analyzed by comparative and transitive logics. It is about time we come deeper into this interesting topic.

As I fleetingly recalled above, transitive logics differ from one another mainly in having more or less rich stocks of logical constants. Here, my chief interest will lie in *propositional* logics – thus, I will leave aside issues related to quantification, even though, of course, I am most ready to acknowledge their crucial role for a proper reconstruction of the domains of discourse we are treating. The connectives of transitive logics which I intend to discuss (or, in any case, which I need for my discussion) are:

- weak negation (symbolized as N), which can be evaluated in a number of ways; an especially convenient one, also consistent with current fuzzy semantical practice, is taking $v(NA) = 1\text{-}v(A)$;
- strong negation (symbolized as $\neg$, which should be read as "It is absolutely not the case that..." and is thereby a sort of Stonean negation (Ovchinnikov 1983): $v(\neg A) = 1$ if $v(A) = 0$, $v(\neg A) = 0$ otherwise;
- conjunction (symbolized as $\wedge$, which obeys the clause $v(A \wedge B) = \min(v(A),v(B))$;
- disjunction (symbolized as $\vee$, which obeys the clause $v(A \vee B) = \max(v(A),v(B))$;
- the conditional (symbolized as $\supset$, definable in terms of strong negation and disjunction: $v(A \supset B) = v(\neg A \vee B)$;

– finally, implication (symbolized as $\rightarrow$), whose evaluation clause is given by $v(A \rightarrow B) = ½$ if $v(A) \leq v(B)$, $v(A \rightarrow B) = 0$ otherwise.

From the above clauses, it is easy to see that while *overcontradictions* (sentences of the form $A \wedge \neg A$) are always false, *simple contradictions* (sentences of the form $A \wedge NA$) may well be true. This, however, does not amount to a waiver of the law of noncontradiction, which – as it is immediate to check on the basis of the given semantics – is indeed valid. Likewise, the law of excluded middle is also valid even if the negation therein is weak negation:

> No hay por qué sacrificar el PTE [the excluded middle] en presencia de lo difuso. Todo lo contrario, de hecho, se puede probar el PTE [...] no sólo lo difuso no se opone al PTE, sino que lo entraña. [...] El PTE es verdadero y falso a la vez. Nada de extraño (Vásconez, Peña 1996).

Peña remarks that, if one accepts involutivity of negation and De Morgan laws, the excluded middle and the principle of noncontradiction stand or fall together. Therefore, any argument for the former also counts as an argument for the latter. Here is a possible argument in defence of excluded middle, to some extent independent of the particular valuation clauses for connectives chosen above (see Peña 1984).

If any sentence of the form $Pa \vee NPa$ has to be a counterexample to the excluded middle, it must perforce be one in which a is a borderline case for a P; in fact, if a is either a definitely positive or a definitely negative P-case, that instance of the excluded middle will be unquestionably true. However, if a is a borderline case for a P, both Pa and NPa will be true to some extent or other – hence true, by the rule of endorsement. By adjunction, $Pa \wedge NPa$ will be true; then, so will be Pa by simplification and $Pa \vee NPa$ by addition. If there are no counterexamples of the envisaged form, there seems to be no reason why there should be counterexamples of a more complex logical form. Thus, there are no counterexamples to the excluded middle.

It is not hard to see where the previous argument breaks down, according to the comparative-logical perspective: the mentioned application of the rule of endorsement, in fact, does not appear to be warranted. Borderline cases for vague properties do indeed provide counterexamples to the excluded middle in that both disjuncts are only approximately true, but not definitely true; hence there is no reason to suppose that the disjunction itself be definitely true.

Of course, the comparative logician now owes an explanation of the widespread credit given to such a law by the community of logicians. Why is the excluded middle generally regarded as valid? Here is a possible answer.

Its good reputation hinges on an unappreciated equivocation lurking behind the word “or”. According to comparative logic, in fact, there are at least two kinds of inclusive disjunction – which collapse onto each other in classical logic, but are to be kept distinct in a degree-theoretical perspective. The first one is what we might call a *parallel* disjunction: each disjunct is evaluated separately in order to ascertain its degree of truth, and the whole disjunction is evaluated as true if at least one disjunct turns out to have a positive (“true”) degree of truth. On the other side, we have a *comparative* disjunction: the degrees of truth of the disjuncts are not assessed independently, but compared to each other, and the whole disjunction comes out true just in case the amount of falsity in each disjunct does not exceed the amount of truth in the other; in other words, just in case each disjunct is at most as false as the other is true.

Therefore, even if negation is taken to mean weak negation and not strong negation, not all ambiguity is dissolved: the principle of excluded middle can still be understood in two different ways. The *comparative excluded middle* is undoubtedly valid, because it falls straight out of harmless semantical stipulations concerning negation that each one of *A*, *NA* is exactly as false as the other one is true. Nonetheless, the *parallel excluded middle* could admit of counterexamples, because both disjuncts might well fail, on their own, to meet the standards of truth. The outcome of all this is that there is a sense in which the excluded middle is indeed valid, whereby its popularity among logicians can be at least partly accounted for.

Another less than satisfactory aspect of transitive semantics is the non-gradability of implicational sentences, which can receive just *two* degrees of truth – a designated and an undesignated one. Either an implicational sentence is half-true, or else it is absolutely false. I argued elsewhere (Paoli 1999; 2003) that degree-theoretical approaches which endorse our principles of reduction and simplification but assign the same degree of truth to all true implicational sentences (for example, approaches based on Łukasiewicz logic) are bound to fail whenever nested comparative sentences are at issue.

Beyond that, this feature of Peña’s theory seems to contravene once again the real fuzziness maxim. In fact, let John be 1.70m tall, Bill be 1.72m tall, and Rick be 2.00m tall. It does no violence to our customary linguistic habits to say that Bill is only *slightly* taller than John, whereas Rick is *much* taller than John. The use of hedges points toward a property, that of being taller than John, that Bill and Rick possess to different degrees – and it falls out of the meanings of those hedges that Rick possesses it to a *higher* degree than Bill does. All these ruminations can be rephrased without undergoing much conceptual change (maybe only at the expense of intuitive perspicuity) if we replace “is taller than..." by "is at least as tall as...". However, given our simplification and reduction principles, "Bill is at least as tall as John" means nothing but "That John is tall implies that Bill is tall", and the same applies to the sentence where "Bill" is replaced by "Rick". Both such sentences are true implicational sentences, hence they get value ½. But then transitive semantics fails to reflect the difference in the respective degrees of possession of the property "being at least as tall as John" by Bill and Rick.

7. The sorites

Any purported theory of vagueness cannot claim to be tenable unless it provides a satisfactory account of the most debated puzzle in this whole area, the sorites paradox. Indeed, Peña attempts such an explanation. Let us now try to give the gist of it.

Consider the following version of the paradox. Suppose, for the sake of simplicity, that the sole measure of one's wealth is given by the amount of currency in one's bank account. Whoever has just one dollar in his account is poor. Moreover, suppose the properties of two people differ just by one dollar: since such a gap is too slight to make any difference as regards the application of the predicate "poor", this means that either the people at issue are both poor or they are both non-poor. From these premisses it follows that whoever has two dollars in his account is poor; but also that whoever has three dollars is such... Sliding down the slippery slope, we are little by little drawn to the unescapable, but absurd conclusion that whoever has 100,000,000 dollars in his account is poor.

If we let $P(n)$ stand for "Whoever has n dollars in his account is poor", the argument rests on a categorical premiss which can be represented as

(4) $P(1)$

and, for each n, on a disjunctive premiss of the form

(5) Either it is not the case that $P(n)$, or $P(n + 1)$.[6]

I purposely refrained from stating (5) in a formalized guise. Indeed, the first step towards a proper understanding of the argument amounts to determining the

[6] This is exactly what our above formulation says, if pruned of uncontroversial information.

actual logical form of (5). For example: does "it is not the case that..." express a weak or a strong negation? According to Peña, the negation at issue cannot be strong: suppose, in fact, that n is such that $P(n + 1)$ is absolutely false, but $P(n)$ is – if only to an extremely small degree – true. Then $\neg P(n) \vee P(n + 1)$ is false, and the argument is unsound. On the other hand, if the negation is weak, each disjunctive premiss must be true: the only possible way for $NP(n) \vee P(n + 1)$ to be false is just in case $P(n)$ is absolutely true, while $P(n + 1)$ is absolutely false – which cannot be, since adjacent objects in a soritical sequence cannot differ so much in their respective possessions of the relevant property as to make such an assignment of values possible. Therefore, (5) must be formalized using a weak negation (and a parallel disjunction). Yet Peña claims that the argument, if so understood, is *invalid* because disjunctive syllogism does not hold for weak negation. Suppose in fact that n is a borderline case for a P; then both $P(n)$ and $NP(n)$ will be true to some degree, and the inference from $P(n)$ and $NP(n) \vee P(n + 1)$ to $P(n + 1)$ will be unwarranted ($P(n + 1)$, for all that we know, might be absolutely false). Summing up:

> Los sorites se resuelven admitiendo la premisa mayor solo en la versión disyuntiva [...]: o uno de dos términos consecutivos en la cadena carece de la propiedad en cuestión, o el siguiente la tiene. Hay que rechazar la formulación condicional – aquella según la cual, si el uno la tiene, el otro también. No valiendo, en general, el silogismo disyuntivo, no nos veamos llevados a la conclusión desastrosa de que todo posee la propiedad en cuestion (Peña 1996, p. 146).

We just saw that, in the opinion of Peña, the connectives involved in the disjunctive premisses of the sorites are weak negation and *parallel* disjunction. Although I agree with the claim about negation, I believe that understanding disjunction as a parallel connective misrepresents the logical structure of the argument. Suppose in fact that someone, at first sight quite plausibly, were to contend that the disjunctive assumptions of this piece of reasoning are true. Would such a contention be based on a separate evaluation of each disjunct? Quite otherwise: the very fact that our proponent maintains that *every* disjunctive premiss is true, independently of the value of n, implies that she is deeming each disjunction true neither out of the truth of its negated disjunct, nor out of the truth of its atomic one: rather, such an assessment is based on a *comparison* of their respective degrees of truth. What gives the argument its bite is the fact that each disjunctive premiss is somehow meant as instantiating the principle of tolerance: all it says is that $P(n)$, whatever its degree of truth may be, is just as true as $P(n + 1)$ is, whence $NP(n)$ is just as false as $P(n + 1)$ is true – and thus, in particular, is at most as false as $P(n + 1)$ is true. The disjunction we are encountering, therefore, is a comparative disjunction and not a parallel one.

Does Peña's system possess the expressive means to define a comparative disjunction? Very much so indeed: such a disjunction is definable as $NA \rightarrow B$. And would the paradox be solved all the same if its premisses were so rephrased? From a purely technical viewpoint, it would. Each premiss would simply be equivalent to $P(n) \rightarrow P(n + 1)$ and, since the degree of truth of $P(n)$ would have to be strictly greater than the degree of truth of $P(n + 1)$ by the real fuzziness principle, Peña's semantics for implication would dictate that $P(n) \rightarrow P(n + 1)$ be absolutely false. The argument, in sum, would be valid – disjunctive syllogism holds for comparative disjunction – but unsound.

However, a proper solution to a paradox must not only explain what goes wrong in the chain of deductions, but also account for the *prima facie* appeal of the argument. If the disjunctive premisses are evaluated as absolutely false, the task is not accomplished: why do we feel so tempted to endorse what, after all, is just an absolute falsity? Remark that pragmatic grounds can be of no avail here: were we to claim that the disjunctive premisses are false but "highly assertable", according to some extra-logical notion of assertability, why not stay with classical logic and apply our pragmatic theory to it instead?

The solution provided by comparative logic, on the other hand, is along the lines of the standard degree-theoretical replies to the sorites: the disjunctive premisses are not definitely true, but fall barely short of definite truth. This explains their appeal, while leaving room for a dismissal of the argument as unsound. The additional gain, with respect to the standard approach, is that the recourse to the flexible framework of comparative logic makes it possible to evaluate all disjunctive premisses in the same way and to properly account for the *uniformity* of any soritical accumulative process (see Paoli 2003 for a more detailed discussion).

8. Conclusion

Let me conclude with a couple of remarks concerning the paraconsistent character of the logics under examination, and their overall bearings on philosophy.

Transitive logics are both paraconsistent and dialethic, i.e. not only they rebut the *ex absurdo quodlibet* but also they contain some contradictions as theses. On the other hand, comparative logic is not a dialethic logic but is paraconsistent – it contains non-trivial and well-motivated inconsistent extensions, such as Abelian logic.

Finally, I would like to conjecture that some of the differences between transitive and comparative logics could perhaps be explained in terms of the different interests of their respective founders, which probably determined a divergence in the intended philosophical applications of the logics themselves. Comparative logic originated, in Casari's intentions, as an attempt to reconstruct the theory of comparison advanced by Aristotle in his *Topics* and elsewhere, but later such historical motivations were outdone by purely mathematical stimuli on the one side, and by the desire to offer significant contributions to the semantics of natural language on the other. If I am not mistaken, transitive logics are more conspicuosly driven toward applications to other branches of philosophy – like metaphysics, theology, philosophy of law or epistemology – which at least so far have not been the primary concerns for comparative logicians.

It may well be possible that comparative logic cannot withstand objections arising from such areas of philosophical debate. What I would like to stress, however, is that neither Casari nor I have ever taken comparative logic to be an all-purpose logic which is appropriate to solve each and every philosophical problem. Rather, comparative logic should be understood as a *task-oriented logic*: a framework which perhaps could prove more adequate than its rival approaches to formally account for a limited fragment of natural language, which is however rich enough to contain gradable and vague predicates – as well as a good deal of adjectival and nominal comparative constructions.

Acknowledgements
I warmly thank Lorenzo Peña for sending me a voluminous parcel with some of his writings; this allowed me to have a less partial knowledge of his fascinating philosophical perspective. I am especially grateful to Marcelo Vásconez for some lively discussions we had on these topics. Finally, I thank a referee of L&PS for his precious suggestions.

REFERENCES

BIERWISCH, M. 1989: "The Semantics of Gradation", in M. Bierwisch and E. Lang (eds.), *Dimensional Adjectives*, Berlin: Springer, pp. 71-262.

CASARI, E. 1981: "Remarks on Comparison and Superlation", in *Atti del Convegno Nazionale di Logica 1979*, Napoli: Bibliopolis, pp. 261-271.

——— 1989: "Comparative Logics and Abelian *l*-groups", in A. Valentini *et al.* (eds.), *Logic Colloquium '88*, Amsterdam: North Holland, pp. 161-190.

——— 1997 "Conjoining and Disjoining on Different Levels", in M. L. Dalla Chiara *et al.* (eds.), *Logic and Scientific Methods*, Dordrecht: Kluwer, pp. 261-288.

ENGEL, E. 1989: "On Degrees", *Journal of Philosophy*, 86, pp. 23-37.

OVCHINNIKOV, S. V. 1983: "General Negation in Fuzzy Set Theory", *Journal of Mathematical Analysis and Applications*, 92, pp. 234-239.

PAOLI, F. 1999:"Comparative Logic as an Approach to Comparison in Natural Language", *Journal of Semantics*, 16, pp. 67-96.

——— 2003: "A Really Fuzzy Approach to the Sorites Paradox", *Synthese*, 134, pp. 363-387.

PEÑA, L. 1984: "Identity, Fuzziness and Noncontradiction", *Noûs*, 18, pp. 227-259.

——— 1987a: *Fundamentos de Ontología Dialéctica*, Madrid: Siglo XXI.

——— 1987b: "Contribución a la lógica de los comparativos", in C. M. Vide (ed.), *Lenguajes naturales y lenguajes formales*, Barcelona: University of Barcelona Press, pp. 335-350.

——— 1990: "Partial Truth, Fringes and Motion: Three Applications of a Contradictorial Logic", *Studies in Soviet Thought*, 39, pp. 283-312.

——— 1992: "Nuevas aplicaciones filosóficas de lógicas multivalentes", *Theoria*, 16-17-18, pp. 141-163.

——— 1993: "Review of Douglas Walton, *Slippery Slope Arguments*", *Pragmatics and Cognition*, 1, pp. 401-410.

——— 1995: "Lógicas multivalentes", *Enciclopedia IberoAmericana de Filosofía*, Madrid: Trotta-CSIC, pp. 323-349.

——— 1996: "Grados, franjas y líneas de demarcación", *Revista de Filosofía*, 9/16, pp. 121-149.

SORENSEN, R. 2002: "Vagueness", *Stanford Encyclopaedia of Philosophy*, http://plato.stanford.edu/entries/vagueness/vagueness.html.

TYE, M. 1994: "Vagueness: Welcome to the Quicksand", *Southern Journal of Philosophy*, 33 (Suppl.), pp. 2-22.

VÁSCONEZ, M. 2002: *The Sorites Paradox: General Survey*, Licentiate Thesis, Katholieke Universiteit Leuven.

VÁSCONEZ, M., PEÑA, L. 1996: "¿Qué es una ontología gradual?", *Agora*, 15, pp. 29-48.

WRIGHT, C. 1975: "On the Coherence of Vague Predicates", *Synthese*, 30, pp. 325-365.

L&PS – Logic & Philosophy of Science
Vol. IV, No. 1, 2006, pp. 123-132

Philosophy, Science, and (Anti-)Communism: The Two Lives of Imre Lakatos*

Critical notice of George Kampis, Ladislav Kvasz, and Michael Stöltzner (eds.), *Appraising Lakatos. Mathematics, Methodology and the Man*, Kluwer: Dordrecht - Boston - London, 2002

Roberto Festa
Department of Philosophy, University of Trieste (Italy)
e-mail: festa@units.it

Thirty years after the untimely death of Imre Lakatos, the enduring influence of several Lakatosian issues and ideas asks for a systematic appraisal of his work. A significant step in this direction has been made in two workshops, organized by the Institute Vienna Circle and the Institute for History and Philosophy of Science of Eötvös University, which were held in Vienna and Budapest in autumn 1997. The workshops resulted in the present substantial (382 pages) volume, which includes seventeen papers dealing with Lakatos' philosophy and biography and several documents and photographs about his work and life. The papers have been divided into three sections, concerning Lakatos' philosophy of science (Section I) and mathematics (Section II), and his intellectual and political biography (Section III).

The contributions collected in Section I deal especially with Lakatos' methodology of scientific research programmes – henceforth, MSRP. The conceptual skeleton of MSRP can be described as follows. A scientific research programme takes the form of a series of successively developed theories sharing a hard core and a positive heuristic. The hard core consists of hypotheses that are irrefutable within the programme, while the residual hypotheses, i.e.

* This paper is an extended version of a review published in *The British Journal for the Philosophy of Science*, 57 (2006), pp. 247-253.

Published by L&PS – Logic and Philosophy of Science
http://www.units.it/~episteme/ – ISSN: 1826-1043

the 'protective belt' of the programme, can be modified in response to the experimental outcomes; the search of appropriate modifications is guided by the so-called positive heuristic. If a programme generates new predictions, and such predictions are confirmed, the programme is progressive. A scientific community is rational only if it embraces the most progressive research programmes. The main issues about MSRP discussed in the contributions to Section I are the following: (1) the MSRP-solution to the Duhem problem and Lakatos' analysis of the role of crucial experiments; (2) the notion of positive heuristic and, more generally, the possibility and the nature of a genuine logic of scientific discovery; (3) the adequacy and the justification of the MSRP-rules governing theory choice decisions.

According to the Duhem thesis, the experimental observations made to test a theoretical hypothesis presuppose the acceptance of a network of auxiliary hypotheses. The Duhem thesis seems to imply, among other things, that the falsification of an isolated hypothesis is impossible, that Popper's falsifiability criterion is inadequate, and that falsifying crucial experiments are impossible. Moreover, the Duhem thesis poses the following further problem: given an experimental outcome conflicting with a group of hypotheses, which of them should be changed? Since MSRP is, to a great extent, motivated by Lakatos' criticisms of Popper's falsificationism, and such criticisms are based fundamentally on the Duhem thesis and the related Duhem problem, one might see MSRP as an elaborated solution of the latter. The MSRP-solution to the Duhem problem amounts, roughly speaking, to the claim that, while the hypotheses included in the hard core of a research programme should be considered as experimentally irrefutable, any hypothesis belonging to the protective belt is a good candidate to refutation. A discussion of the MSRP-solution to the Duhem problem, and of Lakatos' views about crucial experiments, is provided in the contributions by Donald Gillies and Péter Szegedy. Gillies agrees with Lakatos that Popper's philosophy needs to be modified in the light of the Duhem thesis, but argues that the Duhem problem should be solved in a way different from that suggested in MSRP. Szegedy analyses the role of a series of crucial experiments, such as the famous Aspect experiment, in the competition among different interpretations of quantum mechanics, where such interpretations are construed as research programmes characterized by a permanent hard core. His conclusion, that none of the mentioned experiments is a genuine crucial experiment, supports Lakatos' general claim that there are no such things as crucial falsifying experiments.

Lakatos' work on the philosophy of mathematical and empirical sciences is inspired by the assumption that there is a genuine, rationally reconstructible logic of discovery. Although the fortune of Lakatos' philosophy and, in particular, of his notion of positive heuristic, allows to consider him as one of the fathers of contemporary logic of discovery, one might wonder whether positive heuristic is something more than a fascinating metaphor. In his essay, John Worrall remarks (p. 88) that, "on reviewing Imre's published work, it is difficult to see how [his] reputation could have been based soundly on anything that found its way into print", since even his notion of positive heuristic "is very sketchily presented in his famous papers on MSRP". On the other hand, Worrall shares Lakatos' view that the process of discovery can be rationally reconstructed; indeed, his paper provides a constructive contribution to logic of discovery. After examining three historical episodes, related to the theoretical breakthroughs made in optics in the early 19th Century by Augustin Jean Fresnel, Worrall argues that such episodes can be reconstructed by the method of 'deduction from the phenomena,' which allows to deduce hypotheses from the phenomena plus background knowledge.

The debate on theory change and scientific progress, enfolded by Thomas Kuhn's work, dominated philosophy of science in the 1960s and 1970s. Most of the participants in such a debate shared the idea, which is constitutive of Lakatos' approach, that rational theory choice and scientific progress are intimately related; in fact, scientific progress is given by a series of rational theory choice decisions. Theory choice decisions were at the focus of the debate; in particular, the essays by Martin Carrier, Gábor Forrai, Matteo Motterlini, and John Watkins focus on the problem of the adequacy and the justification of the MSRP-rules governing theory choice decisions.

Since it provides, among other things, a rational reconstruction of some of Kuhn's allegedly descriptive generalizations about scientific change, Lakatos' analysis of scientific enterprise can be considered as a response to Kuhn's challenge. Starting from Lakatos' assumption that methodologies have a historical bearing, so that their performance is subject to quasi-empirical scrutiny, Carrier compares Kuhn's and Lakatos' approaches, by reconstructing them as methodological research programmes. More precisely, Carrier tests the consequences of Kuhn's and Lakatos' methodologies in terms of the corresponding theory choice decisions, and concludes (p. 68) that, although it cannot explain the existence of Kuhn losses, in any other respect Lakatos' methodology is better than Kuhn's approach "as an account of scientific change by its own methodological lights".

According to Lakatos, a historically-based normative methodology should satisfy two conditions: (i) methodological rules should be objective, universal, and explicit; (ii) they should be identified and justified by looking at the history of science. Given a historical justification of methodological rules, one may wonder whether they can also be justified in purely epistemological terms. A justification of this kind might consist in showing that methodological rules are functional to the achievement of the epistemic goals of science. For instance, Lakatos suggests that the MSRP-rules for the acceptance of scientific theories can be justified by showing that the growing empirical success, or corroboration, of the accepted theories is a sign of their increasing verisimilitude, i.e., of their increasing approximation to objective truth.

The neo-Hegelian Lakatos portrayed by Hacking (1979) develops a methodological account where the notion of truth as correspondence to the facts is replaced by scientific method, and the real aim of science is not (the approach to) the truth but, rather, the growth of knowledge, guaranteed by the application of the MSRP-rules. Motterlini convincingly argues (p. 44) that Hacking's influential interpretation of Lakatos' methodology is, to say the least, one-sided, since "the late Lakatos could not have regarded methodology as a substitute for truth because he explicitly demanded a connection between the 'game of science' (method) and its 'rational' end (truth)".

The relations between corroboration, verisimilitude and induction were the main issue of the famous Popper-Lakatos rift, which started as an intellectual disagreement but very soon escalated into a bitter quarrel. Although this dispute may be legitimately considered from the perspective of the World 2 of philosophical debates, Watkins examines it within the framework of the World 3 of conceptual relations, in order "to find out what revisions, if any, to the propositional content of Popper's philosophy of science are called by Lakatos' challenge" (p. 5). (Watkins suddenly passed away on July 26th, 1999, about two years after Vienna and Budapest workshops; to my knowledge, the article included here is his last published paper.) The late Lakatos was convinced that the corroboration-based estimates of the verisimilitude of scientific hypotheses had an inductive character, due to the fact that "a corroboration appraisal as understood by Popper sums up how a theory has performed so far and says nothing about future performance", while "verisimilitude appraisals have no temporal restriction" (p. 7). More specifically, he claimed that the link between corroboration and verisimilitude could be adequately described by some sort of synthetic inductive postulate. Although Popper was understandably reluctant to explicitly acknowledge the methodological necessity of some inductive principles in science, he seems to have understood the role of conjectural estimates of verisimilitude as early as between the 1950s and 1960s, when he was working to the paper later published in *Conjectures and Refutations*. Indeed, in his answer to Lakatos, Popper (1974, p. 1011) claims: "I *did* suggest in *Con-*

jectures and Refutations, Chapter 10, that the degree of corroboration may be taken as an indication of verisimilitude". According to Watkins (p. 7), "that little sentence means that the propositional content of Popper's earlier publications already contained the inductive postulate called for by Lakatos", and that "the whiff of induction which Lakatos invited Popper to introduce into his philosophy was already there". Watkins concludes (p. 10) that "very little revision to the propositional content of Popper's philosophy seems to be obligated by the criticisms by Lakatos"; in fact, "their main result was to provoke Popper into that veiled admission that he *had* introduced an inductive assumption".

Although Watkins' essay sheds light on several important features of the Popper-Lakatos rift on corroboration, verisimilitude, and induction, it seems to me that it underestimates its propositional content. A better understanding of the dispute can be achieved by integrating Watkins' account with the remarks on the same subject made by Niiniluoto (1989) on August 1986, at the International Conference on *Imre Lakatos and Theories of Scientific Change* held in Thessaloniki (Greece). (Strangely enough, Niiniluoto's paper is not mentioned by Watkins, although he is the only one, among the contributors to the volume reviewed here, who participated to the Thessaloniki Conference.) Niiniluoto, indeed, considers two substantial, and related, points on which Lakatos disagreed from Popper: (i) According to Lakatos, Popper's admission that the degree of corroboration of a theory may be an (inductive) indicator of its degree of verisimilitude was dramatically insufficient for the formulation of an adequate fallibilistic methodology. Indeed he believed that the link corroboration-verisimilitude had to be precisely reconstructed by appropriate inductive principles, providing plausible corroboration-based estimates of the verisimilitude of the scientific hypotheses, *even if already falsified*. (It should be recalled that, in practice, Lakatos never specified such principles; perhaps, as Forrai suggests (p. 82), "he regarded this task as formidable – or he simply died before he could get down to it".); (ii) Lakatos realised that the inductive principles of the requested kind could not be stated in terms of Popper's notions of corroboration and verisimilitude. More precisely, he was aware that Popper's measure of corroboration was not a good indicator of verisimilitude; in fact, if such a measure is adopted, any experimentally falsified hypothesis receives its minimum value. Apart from making impossible any corroboration-based choice among falsified hypotheses, Popperian corroboration violates a natural condition, taken from scientific practice, according to which it may happen that a hypothesis is falsified by evidence e but, at the same time, is considered to be highly verisimilar on the basis of e. For these reasons, Lakatos (1968, pp. 384-385) introduced a corrected version of Popper's corroboration which assigns positive degrees of corroboration even to refuted theories (see also Lakatos 1974, p. 270, note 122).

It is not clear whether Lakatos realized that also Popper's notion of verisimilitude was seriously inadequate. In any case, he died too soon to appreciate the consequences of the theorem discovered, in 1974, by David Miller and Pavel Tichý, who independently proved that, according to Popper's definition, a false theory can never be closer to the truth than another one. This negative outcome opened the road to the post-Popperian theories of verisimilitude, emerged since 1975. A large part of the work in this field can be seen as a response to the late Lakatos' challenge to identify inductive principles linking corroboration and verisimilitude in the appropriate way. For instance, Niiniluoto (1987) formulates a comprehensive view of scientific inference where a non-Popperian approach to verisimilitude is combined with the theory of inductive probabilities developed within the Carnap-Hintikka tradition. Also the so-called evaluation methodology, recently proposed by Kuipers (2000), where (also falsified) theories are judged in terms of their successes and problems, is explicitly motivated by Lakatos' idea that theory evaluation has a primarily comparative character, leaving falsified theories in the game as long as there are no better alternatives. Although the influence of Lakatos' ideas on the post-Popperian theories of verisimilitude is not examined in the contributions to the present volume, it is an important part of the Lakatosian philosophical legacy, which seems to deserve further research.

After working, between the end of the 1950s and the early 1960s, on *Proofs and Refutations* – henceforth PR – Lakatos came back on philosophy of mathematics in many occasions, until the last years of his life, when he was seeking to apply MSRP to mathematics, and to integrate it with the conceptual framework of PR. The seven papers collected in Section II can be seen as a sign of the recently revived interest in Lakatos' philosophy of mathematics.

Lakatos opposes to the view that mathematical concepts should be conceived of as immutable Platonic entities, and puts a strong emphasis on conceptual change in mathematics. By neglecting any speculation about the hidden nature of mathematical objects, he focusses on mathematical practice as such, in the attempt to develop a methodology of mathematical sciences able to shed some light on the processes underlying conceptual change and discovery in mathematics. Several issues concerning Lakatos' views about this subject are discussed in the articles by David Corfield, Thomas Mormann, Olga Kiss, and Ladislav Kvasz.

According to Lakatos, the most important mathematical knowledge is embodied in the theoretical statements accepted as a result of the process of mathematical discovery, and the axiomatisation of such statements is the end of the creative process in mathematical research. Corfield considers this view as a residual streak of logical empiricism, and argues (p. 117) that "axiomatisation is not the end of the road", since it allows "plenty of room for further disagreement" about "the direction of research, [...] which way to think about a theory, which way to generalise a concept, and so on". Corfield's interest in conceptual change 'after axiomatization' is shared by Mormann. While Lakatos applies his ideas of concept-formation and concept-stretching only to informal mathematics, Mormann contends that the role of conceptual variation is even more important in axiomatic mathematics. Indeed, one of the main purposes of his paper is to generalise Lakatos' analysis of conceptual variation, in order "to sketch an evolutionary theory of mathematical knowledge, which takes axiomatic variation of concepts as the fundamental driving force of the ongoing evolution of mathematics" (p. 139). Lakatos' idea that mathematical discovery is driven by a rationally reconstructible heuristic is borrowed from the Hungarian mathematician George Pólya. Pólya and Lakatos' approaches to mathematical heuristic are compared by Kiss. He points out (p. 248) that Pólya's *Induction and Analogy* (1954) and Lakatos' PR "can be read together, as if they were two parts of the same story": while Pólya is concerned with how mathematicians *find* plausible mathematical conjectures, Lakatos is concerned with how they *prove* them. In terms of discovery, Pólya focusses on the discovery of mathematical conjectures, and Lakatos focusses on the discovery of rigorous proofs of such conjectures. Kvasz argues (p. 211) that, if dialectic is construed "very broadly, as a current of philosophical thought which tries to interpret the growth of knowledge using a prescribed pattern of stages, methods, or laws of development of knowledge", then Lakatos' methodology is properly dialectical. More precisely, it exhibits all the three main "dialectical faults" (p. 212), i.e., the existence of prescribed patterns of conceptual change, their logical nature, and their universal character. Kvasz attacks the dialectical side of Lakatos' approach by showing that the basic methodological patterns outlined in PR and MSRP are far from being universal, and can be supplemented by further ones.

Section II is completed by three fine essays dealing with the implications of Lakatos' work for the foundations of mathematics (Teun Koetsier), the interactions between mathematics and modern mathematical physical science (Michael Stöltzner), and the relevance of philosophy of mathematics to mathematical teaching (Christian Reichel).

Lakatos' intense life appears as the combination of two completely different lives. He spent his first 34 years almost entirely in Hungary, until November 1956, when he left the country, following the Soviet crushing of the short-lived Revolution. After two months he was in Cambridge pursuing a Ph. D. in philosophy of mathematics with R. D. Braithwaite, and within three years he was working with Karl Popper in London. The papers collected in Section III, which were presented in the Budapest workshop by Lászlo Ropolyi, Jancis Long, and Lee Congdon, provide an excellent occasion for assessing Lakatos' intellectual formation and his human and political biography in the two halves, the Hungarian and the English, of his life.

The authors dealing with the dialectical ingredients of Lakatos' philosophy often construe 'dialectical' in a rather broad sense, by neglecting the details of Lakatos' dialectical apprenticeship, within the Hegelian-Marxist tradition. Such apprenticeship is the main subject of the essay by Ropolyi, who traces Lakatos' dialectical views back to the philosophical and political ideas he embraced in the 1940s and 1950s in Hungary, when he was heavily influenced by Georg Lukács. After arguing that a distinctive feature of Lukács' version of the Marxist dialectic is his rationalist view of the coincidence of rationality and progress, according to which rationality is the systematic attempt to choose progressive solutions to problems, Ropoly claims that Lukács' view of rationality was completely absorbed into MSRP. This allows him to conclude (p. 334) that, in spite of his dramatic political changes and his rebuttal of communist ideology, "with regard to methodology [...] Lakatos moved his whole life in a Lukácsian framework".

The Hungarian part of Lakatos' life was largely unknown until the Budapest workshop (30-31 October 1997), which provided a unique occasion to reconstruct some disturbing aspects of his political activities. In fact, since Lakatos was born on November 9th, 1922, the workshop was very close to what would have been his 75th birthday; thereby many of his contemporaries were still alive at that time, including some who suffered from his political deeds. In a detailed reconstruction of Lakatos' life in Hungary, Long tries to explain why Lakatos was 'the unforgiven' for almost all his Hungarian acquaintances. Lakatos' activities as a warm Stalinist are not a sufficient explanation for such a virulently bad reputation in Hungary so many years after his death, since many people were engaged in the arising of a Stalinist system in Hungary and they *were* forgiven; on the other hand, a number of stories which seem to reveal Lakatos' inclination to do harm to other people suggest a plausible explanation.

When, on March 1944, Germany invaded Hungary, the Marxist study group organized by Lakatos during 1943 became a tightly organized communist cell, where Lakatos assumed the right to control group members. In May a 19-year-old Jewish girl, named Éva Iszák, joined the group. Following some difficulties in finding her an adequate clandestine lodging, Lakatos came to the conclusion that she would very likely fall into Nazi hands and be forced to betray the rest of the group; thereby, he decided that her duty to the group and to communism was to commit suicide. In some impressive pages, Long tells us how Lakatos proposed the suicide and pressured all parties into carrying it out. Between May 1945 and December 1948, Lakatos helped establish Hungary's communist state. In spite of his young age, his contribution was so effective that many people considered him as a sort of grey eminence of the Communist Party in the field of culture and education. A salient enterprise carried out by Lakatos in this role is his contribution to the destruction of Eötvös College. In February 1947 Lakatos, who was student at Eötvös, opened a systematic attack against his College, by arguing that the Director and most teachers were unable to train the new intelligentsia. In the end the Director was replaced and the teachers who appeared on Lakatos' list were fired. Lakatos could not enjoy much the result of his action, since during the years 1949-1953 Hungary was plunged into a Stalinist nightmare and he shared the destiny of many other enthusiastic comrades. On September 1950 he was lead to Recsk, the infamous labor camp which reproduced the gulag in Hungary's Matras hills. Although the reasons of his staying there are still obscure, it seems likely that his bad luck under the new system was at least partially due to his ultra-Stalinist fanaticism. In any case, there is no doubt that the worst aspect of Lakatos' experience in those years is the circumstance that, since his discharge from prison, on September 1953, he had an agreement with the police to provide information on his best friends.

The renewed interest for politics exhibited by Lakatos about ten years after his arrival to England is the main subject of Congdon's article. Lakatos discovered that he was still excitable politically during the summer of 1966, in response to the student revolt at the LSE. In this occasion he strongly opposed the recommendation that students help to direct the academic policy, by arguing that what the radicals really wanted was "the winning of power that would enable them first to destroy universities as centers of learning and then, in their place, to create centers for the dissemination of social and political propaganda" (p. 342). Being determined to save the LSE from a fate similar to that he had imposed upon Eötvös College, he effectively helped "stiffen the authorities' backbone" (p. 344). Lakatos' views about international politics were characterized by a sharp anti-Sovietism; he was in favour of a strong Western initiative, also on military ground, and warmly supported U.S. actions in Vietnam. In 1972-1973 Lakatos enunciated the project to work on political philosophy, but he died before he could realize it. Congdon suggests, quite reasonably, that Lakatos, in formulating his own political philosophy, intended to adapt the basic ideas of MSRP, and guesses (p. 347) that he would have developed a theory not very distant from “the slightly left of center ‘Open Society’ liberalism of his quondam friend Karl Popper”.

REFERENCES

HACKING, I. 1979: “Imre Lakatos’ Philosophy of science”, *British Journal for the Philosophy of Science*, 30, pp. 381-410.

KUIPERS, T. A. F. 2000: *From Instrumentalism to Constructive Realism*, Dordrecht: Kluwer.

LAKATOS, I. 1968: “Changes in the Problem of Inductive Logic”, in I. Lakatos (ed.), *The Problem of Inductive Logic*, Amsterdam: North-Holland Publishing Company, pp. 315-417.

——— 1974: “Popper on Demarcation and Induction”, in P. A. Schilpp (ed.) 1974, pp. 241-73.

NIINILUOTO, I. 1987: *Truthlikeness*, Dordrecht: Reidel.

——— 1989: “Corroboration, Verisimilitude, and the Success of Science”. in K. Gavroglu, Y. Goudaroulis and P. Nicolacopoulos (eds.), *Imre Lakatos and Theories of Scientific Change*, Dordrecht: Reidel, pp. 229-43.

PÓLYA, G. 1954: *Induction and Analogy* (Vol. 1 of *Mathematics and Plausible Reasoning*), Princeton: Princeton University Press.

POPPER, K. R. 1974: “Replies to My Critics”, in P. A. Schilpp (ed.) 1974, pp. 961-1197.

SCHILPP, P. A. (ed.) 1974: *The Philosophy of Karl Popper*, La Salle, Illinois: Open Court.

L&PS – Logic & Philosophy of Science
Information on the Journal

Aims and contents

L&PS – Logic and Philosophy of Science is an on-line philosophical journal sponsored by the Department of Philosophy of the University of Trieste (Italy). The journal promotes both theoretical and historical research in the philosophy of science and logic, without excluding any particular cultural perspective.

Topics welcomed by the journal include:

- the theory of scientific knowledge and the analysis of the general methodological problems of science (such as scientific discovery, causation, scientific inference, induction and probability, the structure of scientific theories and their relations with empirical data);
- the methodological and foundational problems of the different sciences, from the natural, to the biomedical, to the social sciences;
- the problems related to the historical development of logic, in all its branches, and to the role of logical methods both in the general methodology of science and in the foundations of empirical and mathematical sciences;
- the philosophical problems raised by the development of the cognitive sciences and the philosophy of mind, with particular attention to those results that are relevant for the analysis of scientific practice;
- the epistemological problems related to Artificial Intelligence, robotics, virtual reality, and artificial life;
- the problems in the sociology and the history of science that are relevant to the philosophical investigation of science;
- the problems related to the ethics of science;
- the questions related to the historical and conceptual development of the philosophy of science and logic;
- the problems of the philosophy of language, with particular attention to those results that are relevant for logic and philosophy of science.

The journal will appear twice a year (in March and in September), and one of these issues will be usually devoted to a special topic.

Information for the authors

Papers submitted to the journal must be written either in Italian or in English, and must be accompanied by a short summary in English (and also in Italian for the articles written in Italian). All papers will be evaluated by anonymous referees.

In order to promote critical discussion and exchange among scholars, the journal is willing to publish reports on work in progress, to be submitted and evaluated according to the criteria already mentioned above.

Typesetting, graphical advice and technical assistance

Gustavo Cevolani (g.cevolani@gmail.com)

Luca Tambolo (l_tambolo@hotmail.com)

Webmasters

Gaia Serafini (gaia.sera@libero.it)

Alessio Lorenzi (ale.lorenzi@libero.it)