L&PS – Logic & Philosophy of Science
Vol. IX, No. 1, 2011, pp. 207-222

Paradoxes, circularity and learning processes

Salvatore Roberto Arpaia
Department of Human and Social Sciences
University of Bergamo (Italy)
roberto.arpaia@unibg.it

ABSTRACT. Considering Bateson's theory of learning and communication, it could be argued that creative change in cognitive systems (e.g., as described in relational-systemic psychology) is related to the existence of a set of Liar-type sentences in the communicative context of the systems (pragmatic paradox), and that the solution to pragmatic paradoxes are connected to the construction of a new meta-context in which the old messages have to be interpreted. A simple set theoretical model of this kind of processes could be defined, in the theory of non-well-founded sets, using the notion of *partial model* first introduced by Barwise.

KEYWORDS: Liar Paradox, Pragmatic Paradox, Meta-communication, Non-well-founded Semantics

1. Introduction: psychology and pragmatic paradoxes

The study of paradoxes in natural languages lead psychologists and cognitive scientists, since the publication of [7] and [21], to give an account of change in cognitive systems in terms of the existence of Liar-type sentences in the communicative (pragmatic) context of the system: in this account, a learning process for a cognitive system is deeply connected to the the construction of a new meta-communicative context in which the old messages have to be reinterpreted. Following this point of view, psychopathologies can be seen as the consequence of an undecidability situation in the cognitive system under analysis: these pathological states of mind could be bypassed by means of a learning

Published by L&PS – Logic and Philosophy of Science
http://www2.units.it/˜episteme/ – ISSN: 1826-1043

process which allows the cognitive system to decide, after the assumption of a revised set of cognitive premisses, the previous undecidable statement. Such a process of revision was deeply linked to the notion of *meta-context* after Gregory Bateson's researches on the nature of learning and change: in its initial formulation ([7, pp. 159-176]), Bateson's *theory of learning* consisted of the hypothesis that each instance of mechanical learning (Pavlovian learning, trial-and-error learning, etc.) involves as a "side effect" a process of reorganization of the psychological character of the learning subject, a process which Bateson called *deuterolearning* (or *learning to learn*). Thus, beside the traditional notion of *Proto-learning*, Bateson defines the *Deutero-learning* as the processes of cognitive reorganization through which an individual, who possesses the ability to solve a *single problem*, gains the ability is able to solve *classes of similar problems*, or in other words, the processes that enable an individual to switch from operating in a certain context to operating in a "context of contexts", that is, in a *meta-context*. In ([7, pp.279-308]), this analysis was refined by the description of almost three different kinds of learning processes, called *Learning I* (or *Proto-learning*), *Learning II* (or *Deutero-learning*) and *Learning III*. Roughly speaking, the three kinds of processes could be seen as three different steps in an ordered hierarchy of processes: (1) the level of learning how to *solve a single problem*, (2) the level of learning how to *solve a class of problems* and (3) the level of *creation of a new class of problems (and of related rules of solution)*. Note that the three levels represent, from a naïve set theoretical perspective, a sequence of objects obtained increasing the level of set theoretical complexity (hence, of abstraction). Moreover, we can observe that, while the passage from (1) to (2) is generally continuous, the passage from (2) to (3) involves a discontinuity (which we can refer to as *paradigmatic* or *trans-contextual* change). The feature of discontinuity of some change processes in learning lead some scholars (see for instance [17] and [16]) to the investigation of limitative theorems as a source of inspiration in the study of change in cognitive processes: the basic idea of these studies is that paradigmatic change within a cognitive system could be represented as an ordered sequence of cognitive states $\langle S_1, \dots, S_n \rangle$ where, for some i ($i < n$), there is a proposition u_i such that u_i is undecidable in S_i but it is decidable (is true or false) in S_n, for some n. In this way S_n is obtained from S_j ($j < n$) by some kind of revision of the set of axioms and rules of S_j.

In systemic psychology, the *undecidable sentence* which gave rise to pathology is referred to as a *double bind*, or *pragmatic paradox*. As an instance of communicative paradox which could be pathological, Bateson observed that among the general characteristics of the family situation in psychopathologies, there is often a communicative paradoxical circle between some of the members of the family:

> The necessary ingredients for a double bind situation, as we see it, are:
> 1. Two or more persons.
> 2. Repeated experience.
> 3. A primary negative injunction. This may have either of two forms: (a) "Do not do so and so, or I will punish you", or (b) "If you do not do so and so, I will punish you". Here we select a context of learning based on avoidance of punishment rather than a context of reward seeking.
> 4. A secondary injunction conflicting with the first at a more abstract level, and like the first enforced by punishments or signals which threaten survival. [...] Verbalization of the secondary injunction may, there-fore, include a wide variety of forms; for example, "Do not see this as punishment"; "Do not see me as the punishing agent; "Do not submit to my prohibitions"; and so on.
> 5. A tertiary negative injunction prohibiting the victim from escaping from the field. ([7, pp.206-207])

As pointed out in the classic book *Pragmatics of human communication* ([21, p.206]), the main feature of such a *pragmatic paradox* is related to the lack of the ability, for the subject, to change and react to a paradoxical context:

> The main distinction between contradictory and paradoxical injunctions is the following: when facing a contradictory injunction, one chooses one of the alternatives and loses the other alternative [...], contradictory injunction always offer the possibility of making a logical choice. Instead, paradoxical injunctions *make the choice itself fail*, nothing is possible and it give rise to an oscillating self-maintaining sequence.)

Even if the notion of *double bind* applies mainly to the analysis pathological states, there is a wider range of applications that allow the introduction of this notion in the study of human communication and natural language semantics. One of the main scientific contributions of Bateson in the study of natural language and communication is the notion of *meta-communication*, based on the *theory of logical types* ([18]) and on Tarski's account of truth and meaning in terms of hierarchy of languages:

> Earlier fundamental work of Whitehead, Russell, Wittgenstein, Carnap, Whorf, etc., as well as my own attempt to use this earlier thinking as an epistemological base for psychiatric theory, led to a series of generalizations: That human verbal communication can operate and always does operate at many contrasting levels of abstraction. These range in two directions from the seemingly simple denotative level ("The cat is on the mat"). One range or set of these more abstract levels includes those explicit or implicit messages where the subject of discourse is the language. We will call these metalinguistic (for example, "The verbal sound 'cat' stands for any member of such and such class of objects", or "The word, 'cat' has no fur and cannot scratch"). The other set of levels of abstraction we will call metacommunicative (e.g., "My telling you where to find the cat was friendly", or "This is play"). In these, the subject of discourse is the relationship between the speakers. It will be noted that the vast majority of both metalinguistic and metacommunicative messages remain implicit; and also that, especially in the psychiatric interview, there occurs a further class of implicit messages about how metacommunicative messages of friendship and hostility are to be interpreted. ([7, pp. 177-178])

Bateson extended Tarski's concept of *metalanguage* ([19]) with the notion of *metacommunication*, defining two different type-theoretical hierarchies which concur both in the construction of the meanings of the messages. In Bateson's theory of communication, the meaning of a sentence (belief, assertion, thought, etc.) depends strictly on the context in which the communication is situated. Metacommunicative messages build the context in which linguistic and metalinguistic messages have to be interpreted. In this perspective, a *double bind* is a particular communicative message that violates the (well-founded) type-theoretical hierarchy of communication. However, as said above, the presence of double binds in learning and communication processes could be a source of pathology, but could be as well a creative source of new meanings. As an instance of creative pragmatic paradox, Bateson considered the metacommunicative sequence involved when two animals are playing:

> Now, this phenomenon, play, could only occur if the participant organisms were capable of some degree of meta-communication, i.e., of exchanging signals which would carry the message "This is play". The next step was the examination of the message "This is play", and the realization that this message contains those elements which necessarily generate a paradox of the Russellian or Epimenides type (a negative statement containing an implicit negative metastatement). Expanded, the statement "This is play" looks something like this: "These actions in which we now engage do not denote what those actions for which they stand would denote". ([7, p. 180])

The main cognitive feature involved in the understanding of a *playing communication sequence* lies therefore in the subjects' ability in creating a set of *contextual indexicals* by which the trans-contextual violations of the logical types in communication could be arranged in a precise order and relativized to a determinated context (on this pint, see [9, p. 80]). The next section is devoted to illustrate a possible formal translation of this interpretation of pragmatic paradoxes.

2. From pragmatics to logic: situation theory and non-well-founded set-theory

It is interesting to note that, if in observing a double bind situation we limit ourselves to the consideration of the communicative interactions described in point (3) and (4) as the primary negative injunction and the secondary (metacommunicative) injunction, we get a pair of sentences which is very similar to the the pair of messages described in the case of playing. In particular, both the creative and pathological double binds described in the previous section share the presence, in their structure, of an ordered pair of sentences of the form

- The following sentence is true

- The previous sentence is false

In [2, pp.148-149], two such sentences are called a *Liar circle of length 2*. Starting from the publication of Jon Barwise's first work on situation theory ([3]) and of Peter Aczel's well known monograph on non-well-founded set theory ([1]), situation semantics (see e.g. [10], [11], [12], [13]) allowed an interpretation of Liar paradox, Liar circles and self-referential sentences in terms of semantical models (called by Barwise *partial models*) that give an account of the notion of *truth value of a formula in a context*, where the meaning of Liar-type sentences vary together with the variation of the values of some contextual indexicals:

> The idea, roughly, is that the fact of asserting the Liar about the whole world results in some sort of pragmatic shift of context, a shift that is overlooked in the reasoning that seems to lead to paradox. Put crudely, "here" (as referred to h) before the claim and "here" after the claim are slightly different.[...] Our discussion suggests that, contrary to appearances, the Liar Paradox does not force one into abandoning the intuitive idea that any claim is either true or it is not true. What it does is forces one to be extraordinarily sensitive to subtle shifts in context. This seems to us quite a plausible explanation for the intuitive reasoning behind the paradox: it just fails to be sufficiently sensitive to such subtleties. ([6, pp. 188-189])

The diagnosis of the existence of Liar-type sentences proposed by *partial model theory* (for a summary on this point, see [8]) is that the paradoxical character of those sentences depends on a some kind of misuse of the symbols denoting contexts in communication and in cognition. Barwise's semantical account of these kind of context-dependent sentences is based in an essential way to the possibility of defining set-theoretical structures involving at some point of the construction non-well-founded sets and classes, and it goes therefore in a different direction with respect of models based on the Tarskian definition of truth (based on the theory of types, and therefore on structure involving only well-founded sets). According to the ideas exposed in the previous section, in what follows I will give a representation of pragmatic paradox and of contextual change in a cognitive system in terms of an ordered sequence of partial models $\langle M_1, \dots, M_n \rangle$ where, for some i ($i < n$), there is a Liar sentence λ_i such that λ_i is paradoxical in M_i but it is decidable (is true or false) in a partial model M_k, for some $i < k < n$, defined from M_i by the switch of meaning, in the new model, of some context-indexical contained in λ_i.

Previously, let me recall some basic notions and definitions of partial model theory. Let *ZFA* be Zermelo-Fraenkel set theory with the antifoundation axiom and and axiom for Urelements. Let $\mathscr{U}$ represent the class of urelements, and assume there are disjoint sets *Rel* of relation symbols, *Const* of constant symbols, and *Var* of variables such that $Rel \cup Const \cup Var \subseteq \mathscr{U}$. To every relation symbol in *Rel* corresponds a natural number *n* called the *arity* of the symbol. The notion of sentence built up form these symbols by the first order connectives is intended as usual.

Definition 2.1. [6, p. 178] A *partial model* M is a tuple $\langle D_M, L_M, Ext_M, Anti_M, d_M, c_M \rangle$ satisfying the following conditions:

1. D_M is a non-empty set, called the *domain* of M.

2. $L_M \subseteq Rel \cup Const$ is a set of relation and constant symbols, called the *language* of M.

3. Ext_M and $Anti_M$ are functions with domain $L_M \cap Rel$ such that, for each n-ary relation symbol R, $Ext_M(\mathsf{R})$ and $Anti_M(\mathsf{R})$ are disjoint n-ary relations on D_M, called the *extension* and the *anti-extension* of R in M, respectively.

4. d_M is a function with domain $L_M \cap Const$ taking values in D_M; if $d_M(\mathsf{c}) = b$, then b is said to be the *denotation* of c in M.

5. c_M is a function with domain a subset of *Var* taking values in D_M, called the *context* of M; if $c_M(\mathsf{v}) = b$, then b is said to be the *denotation* of v in M.

Definition 2.2. A *total model* is a model M such that for each n-ary relation symbol R of L_M and every n-tuple $m_1, \ldots, m_n$ from the domain of M, either $\langle m_1, \ldots, m_n \rangle \in Ext_M(\mathsf{R})$ or $\langle m_1, \ldots, m_n \rangle \in Anti_M(\mathsf{R})$, and such that $d_M(\mathsf{c})$ is defined for every constant symbol c of L_M.

Definition 2.3. A sentence φ of L_M is *defined* in a model M if every constant and variable of φ has a denotation in M. $Def(M)$ is the set of sentences defined in M.

The notion of *extension* of a model M is defined without problems. It was first proved in [2] the following:

Theorem 2.1. *Every model M has an extension M_{tot} which is a total model.*

The following definition introduces the crucial notion of *Kleene evaluation* of a formula in a model ([6, pp.179-180])

Definition 2.4. Given a model M and a sentence $\varphi \in Def(M)$, $M \models \varphi$ and $M \models^{-} \varphi$ are defined as follows:

Evaluation of Atomic Sentences If φ is atomic, say $\mathsf{R(x,y,z)}$

$M \models \varphi$ iff $\langle den_M(\mathsf{x}), den_M(\mathsf{y})\rangle \in Ext_M(\mathsf{R})$

$M \models^{-} \varphi$ iff $\langle den_M(\mathsf{x}), den_M(\mathsf{y})\rangle \in Anti_M(\mathsf{R})$

(where $den_M(\mathsf{t})$ is defined to be $d_M(\mathsf{t})$ if t is a constant, and $c_M(\mathsf{t})$ if t is a variable in the domain of c_M).

Evaluation of Molecular Sentences Sentences built by application of first order primitive connectives $\wedge, \neg, \exists$, are defined in the expected way:

$M \models \neg\varphi$ iff $M \models^{-} \varphi$

$M \models^{-} \neg\varphi$ iff $M \models \varphi$

$M \models (\psi \wedge \xi)$ iff $M \models \psi$ and $M \models \xi$

$M \models^{-} (\psi \wedge \xi)$ iff $M \models^{-} \psi$ or $M^{-} \models \xi$

$M \models \exists \mathsf{x}\psi$ iff for some model M', $M' =_{\mathsf{x}} M$, $M' \models \psi$

$M \models^{-} \exists \mathsf{x}\psi$ iff for some model M', $M' =_{\mathsf{x}} M$, $M' \models^{-} \psi$

(where $M' =_{\mathsf{x}} M$ means that the model M' and M are identical with the possible expection as to the value of the contexts on the variable x).

A sentence φ is said to be *not true* in M if $M \not\models \varphi$, is said to be *false* in M if $M \not\models \neg\varphi$. If M is a total model and $v\varphi \in Def(M)$,we have $M \not\models \varphi$ iff $M \models \neg\varphi$. In order to give a formal definition of the *Liar sentence*, we have to assume that one of the predicates of L is a binary truth predicate $\mathsf{True(x,y)}$, which is intended to express the condition that the sentence denoted by x is true in the model y. Given a model M, *True*(a,b) will mean that $\langle a,b\rangle \in Ext_M(\mathsf{True})$, and *False*($a$,$b$) will mean that $\langle a,b\rangle \in Anti_M(\mathsf{True})$. $\mathsf{True_y(x)}$ stands for $\mathsf{True(x,y)}$.

Definition 2.5. Given a model M, consider the following conditions for every $\varphi, N \in D_M$):

- (T1) If *True*(φ,N), then N is a model, $\varphi \in Def(N)$ and $N \models \varphi$.

- (T2) If N is a model, $\varphi \in Def(N)$, and $False(\varphi,N)$, then $N \models^{-} \varphi$.
- (T3) If N is a model, $\varphi \in Def(N)$, and $False(\varphi,N)$, then $N \not\models \varphi$.

Then the model M is said to be *truth-correct* if, for all sentences $\varphi \in D_M$ and all models $N \in D_M$, M satisfies conditions (T1) and (T3). The model M is said to be *truth complete* if it is truth-correct and, for all sentences $\varphi \in D_M$ and all models $N \in D_M$, M satisfies the converses of conditions (T1) and (T3).

We now come to the formal definition of the Liar sentence in a model M: suppose the language of M contains a binary relation symbol True, a costant symbol this (intuitively, the constant this represent the demonstrative "this"). Then a *Liar sentence* is any sentence of the form

$$\neg \mathsf{True}_{\mathsf{h}}(\mathsf{this})$$

where h (h stays intuitively for the demonstrative "here") is a variable of the language.

In the book *Vicious circles* by Barwise and Moss, the formal counterpart of the Liar Paradox is contained in the following theorem ([6, Theorem 13.10, p.187]):

Theorem 2.2 (The Liar). . *Let λ be a Liar sentence* $\neg \mathsf{True}_{\mathsf{h}}(\mathsf{this})$. *If a M is a truth correct model then at least one the following must fail:*

- *1.* this *denotes λ in M.*
- *2.* h *denotes M in M.*
- *3. $M \models \lambda \vee \neg\lambda$.*

In particular, if (1) and (2) both hold, then M is not total.

Proof. (Hint) Assume (1), (2), (3), and derive the contradiction $M \models \lambda$ iff $M \models \neg\lambda$. □

The "traditional solutions" proposed for the Liar paradox, based on Tarski's account of truth, are constructed abandoning assumption (1) of the theorem: namely, in classical Tarskian semantics, there are no sentences which could speak about their own truth. Abandoning assumption (3) entails the consideration of "truth gaps", that is, the existence of sentences which lack of a truth value in the model. The original "solution" introduced by Barwise *et al.* ([4], [2]) suggests to quit assumption (2), that is, proposed to solve the Liar by introducing the notion of *context* of a sentence, and to refer the existence of a Liar-type paradox to a wrong use of contextual indexes such as h.

Now, in order to recapture the informal concept of *pragmatic paradox* in the formal theory of partial model, we need some further considerations. A pragmatic paradox of the type considered in the example of communication in the case of playing, is constituted by a logical part (the semantical structure instantiated by a Liar circle of length 2) and a pragmatic part (the pragmatic content of communication).

With regards to the logical part of a pragmatic paradox, the formalization *Liar circle of length 2* could be defined in the theory partial models as follows: if M is a model, $Def(M)$ must contain sentences λ_1, λ_2 that intuitively assert

(λ_1) The following sentence is true

(λ_2) The previous sentence is false

and the latter pair of sentences could be easily translate in formal language by setting:

(λ_1) $\mathsf{True_h}(\mathsf{that}_2)$

(λ_2) $\neg\mathsf{True_h}(\mathsf{that}_1)$

But in common life situations, sentences that occur in a pragmatic paradox are not only self-asserting (or self-denying) sentences, and in general are not only Liar-circles: they are meaningful in the precise sense that they contain set of sentences about the real world that are not in general self-contradictory or paradoxical, but rather they are so only *contingently*. So in order to express the pragmatic content of a pragmatic paradox, in our model we need to combine sentences like λ_1 and λ_2 with a pair of (non paradoxical) declarative sentences p and q. Consider the following example from clinical psychology : if a person assumes at same time in his cognitive world the validity of the two sentences $p =$"I feel myself like a knight in a shining armor" and $q =$"I feel myself like a small animal in need of protection", under certain circumstances the simultaneous belief in p and q could lead the subject in question to a pathological situation (for a detailed account of this therapeutic case, see [14]). Clearly, p and q do not form in themselves a Liar circle: however, in certain cases, they could be perceived by the subject as a pragmatic paradox and a source of pathology: if so, the possibility of a cognitive change is connected with the possibility of a shift in the contexts in which p and q are made meaningful, obtaining a new context and a new meaning of p and q.

In order to give a formal account of this kind of cognitive change, following the ideas and the results contained in [6], I will define a truth-correct model M_0 for the (logical) Liar circle of length 2, and to expand this model to a truth-correct total model M_1 obtained by a "context switch" of the indexical variable h, obtaining thus a hierarchy of models that could be intended as the formal counterpart of the change process in the cognitive system under analysis (for a hierarchical-contextual approach to Liar-type sentences, see [15]). First I will define M_0 and M_1 for the case of a Liar circle of length 2, and then I will modify the example in order to obtain similar models for a pragmatic paradox. Suppose our language contains a binary relation symbol True and two constants that_1 and that_2.

Proposition 2.1. *There exists a model M_0 which satisfies the following:*

1. that_1 *denotes* $\lambda_1 = \mathsf{True}_\mathsf{h}(\mathsf{that}_2)$ *in* M_0.
2. that_2 *denotes* $\lambda_2 = \neg\mathsf{True}_\mathsf{h}(\mathsf{that}_1)$ *in* M_0.
3. h *denotes* M_0 *in* M_0.
4. *M_0 is truth correct (therefore, by Theorem 3.2, M_0 is not total).*

Proof. (Hint) Let M_0 be defined as follows:
$D_0 = \{M_0, \lambda_1, \lambda_2\}$
$L_0 = \{\mathsf{True}, \mathsf{that}_1, \mathsf{that}_2\}$
$Ext_0 = \{\langle \mathsf{True}, \emptyset \rangle\}$
$Anti_0 = \{\langle \mathsf{True}, \emptyset \rangle\}$
$d_0 = \{\langle \mathsf{that}_1, \lambda_1 \rangle, \langle \mathsf{that}_2, \lambda_2 \rangle\}$
$c_0 = \{\langle \mathsf{h}, M_0 \rangle\}$ □

This model is truth-correct, but clearly, according with the Liar, neither $M_0 \models \lambda_1$ nor $M_0 \models \neg\lambda_1$ (and the same for λ_2). Moreover, note that M_0 is *reflexive* in the sense that $M_0 \in D_0$, and M_0 has a name for itself in M_0 (the variable h).

The following Proposition defines the model obtained form M_0 by a shift of context.

Proposition 2.2. *There is a model M_1 which satisfies the following:*

1. that_1 *denotes* $\lambda_1 = \mathsf{True}_{\mathsf{h}}(\mathsf{that}_2)$ *in* M_1.
2. that_2 *denotes* $\lambda_2 = \neg\mathsf{True}_{\mathsf{h}}(\mathsf{that}_1)$ *in* M_1.
3. h *denotes* M_0 *in* M_1.
4. M_1 *is truth correct and total.*

Proof. (Hint) Let M_1 be defined as follows:
$D_1 = \{\lambda_1, \lambda_2, M_0, M_1\}$
$L_1 = \{\mathsf{True}, \mathsf{that}_1, \mathsf{that}_2\}$
$Ext_1 = \{\langle \mathsf{True}, \emptyset \rangle\}$
$Anti_1 = \{\langle \mathsf{True}, D_1 \times D_1 \rangle\}$
$d_0 = \{\langle \mathsf{that}_1, \lambda_1 \rangle, \langle \mathsf{that}_2, \lambda_2 \rangle\}$
$c_0 = \{\langle \mathsf{h}, M_0 \rangle\}$ □

Note that M_1 is a total model, and that in M_1 the first sentence of the Liar sentence λ_1 is *false* (of M_0), and the second Liar sentence λ_2 is true, since we have $M_1 \models \neg\lambda_1$ and $M_1 \models \lambda_2$. The intuitive meaning of this example lies in the following crucial point: in M_0 the sentences λ_1 and λ_2 refer to the very same model M_0, in M_1 they both refer to a different model (namely, M_0). Let me try to make this remark clearer by introducing a new notation for denotation: given a model M, an element $a \in M$ is *named* in M if in L_M there is a symbol s such that $den_M(\mathsf{s}) = a$. If an element a of the domain D_M is named in M, let $\ulcorner a \urcorner$ denote an arbitrary constant or variable such that $den_M(\ulcorner a \urcorner) = a$. But then, since the Liar sentences have the form $\lambda_1 = \mathsf{True}_{\ulcorner M_0 \urcorner}(\mathsf{that}_2)$ and $\lambda_2 = \neg\mathsf{True}_{\ulcorner M_0 \urcorner}(\mathsf{that}_1)$, by Theorem 3.2 they are paradoxical in M_0, while they are not in M_1.

The last example concerns the model of a double bind situation: I define a pair of sentences π_1 and π_2 which are intended to model both the logical and the pragmatic part of a double bind situation. For simplicity's sake I will assume that the sentences p and q which represent the pragmatic part of the the pragmatic paradox (intuitively, the factual content of the metacommunicative sequence) are atomic sentences, built up from two unary relation symbols P and Q and constants b and c of the language.

Proposition 2.3. *There is a truth correct model M_3 which satisfies the following:*

1. that_1 *denotes* $\pi_1 = \mathsf{P}(b) \wedge \mathsf{True}_\mathsf{h}(\mathsf{that}_2)$ *in* M_3.
2. that_2 *denotes* $\pi_2 = \mathsf{Q}(c) \wedge \neg\mathsf{True}_\mathsf{h}(\mathsf{that}_1)$ *in* M_3.
3. h *denotes* M_3 *in* M_3.

Proof. (Hint) Let p , q be Urelements, and let M_3 be defined as follows:
$D_3 = \{p, q, \pi_1, \pi_2, M_3\}$
$L_3 = \{\mathsf{True}, \mathsf{P}, \mathsf{Q}, \mathsf{b}, \mathsf{c}, \mathsf{that}_1, \mathsf{that}_2\}$
$Ext_3 = \{\langle \mathsf{True}, \emptyset \rangle\}$
$Anti_3 = \{\langle \mathsf{True}, \emptyset \rangle\}$
$d_3 = \{\langle \mathsf{that}_1, \pi_1 \rangle, \langle \mathsf{that}_2, \pi_2 \rangle, \langle \mathsf{b}, p \rangle, \langle \mathsf{c}, q \rangle\}$
$c_3 = \{\langle \mathsf{h}, M_3 \rangle\}$ □

Intuitively, M_3 represents a double bind situation, where there is a pragmatic conflict between messages p and q. Observe that not only in M_3 we have neither $M_3 \models \pi_i$ nor $M_3 \not\models \pi_i$ ($i \in \{1,2\}$), but also neither $M_3 \models \mathsf{P}(\mathsf{b})$ nor $M_3 \not\models \mathsf{P}(\mathsf{b})$ (and the same holds for $\mathsf{Q}(\mathsf{c})$). The next example is the desired set theoretical structure for the construction of a model in which the previous double bind situation is solved (in the precise sense that the new model is truth-correct, *total* expansion of the old one).

Theorem 2.3. *Given a truth correct model M_3 which satisfies the conditions of the previous Proposition, there exist a model M_4 following:*

1. that_1 *denotes* $\pi_1 = \mathsf{P}(b) \wedge \mathsf{True}_\mathsf{h}(\mathsf{that}_2)$ *in* M_4.
2. that_2 *denotes* $\pi_2 = \mathsf{Q}(\mathsf{c}) \wedge \neg\mathsf{True}_\mathsf{h}(\mathsf{that}_1)$ *in* M_4.
3. h *denotes* M_3 *in* M_4.
4. *M_4 is a truth correct total, reflexive model.*

Proof. (Hint) Let p , q be Urelements, and let M_4 be defined as follows:
$D_4 = \{p, q, \pi_1, \pi_2, M_3, M_4\}$
$L_4 = \{\mathsf{True}, \mathsf{P}, \mathsf{Q}, \mathsf{b}, \mathsf{c}, \mathsf{that}_1, \mathsf{that}_2\}$
$Ext_4 = \{\langle \mathsf{Q}, \{q\} \rangle, \langle \mathsf{True}, \mathscr{S} \rangle\}$, where $\mathscr{S}^+ = \{\langle M_4, \pi_2 \rangle\}$
$Anti_4 = \{\langle \mathsf{P}, \{p\} \rangle, \langle \mathsf{True}, \mathscr{S}^- \rangle\}$ where $\mathscr{S}^- = (D_4 \times D_4) - \{\langle M_4, \pi_2 \rangle\}$
$d_4 = \{\langle \mathsf{that}_1, \pi_1 \rangle, \langle \mathsf{that}_2, \pi_2 \rangle, \langle \mathsf{b}, p \rangle, \langle \mathsf{c}, q \rangle\}$
$c_4 = \{\langle \mathsf{h}, M_3 \rangle\}$ □

Observe that $M_4 \models \pi_2$ and $M_4 \models \neg\pi_2$, and in particular $M_3 \models \mathsf{Q}(\mathsf{c})$ and $M_3 \models \neg\mathsf{P}(\mathsf{b})$. In the previous model, it could have been the case that both π_1 and π_2 were true, but not that they were both false: in other words, *at least* one of the two sentences have to be true. However, if we slightly modify the definition of the pragmatic Liar circle, setting

- $\pi_1 = \mathsf{P}(\mathsf{b}) \wedge \mathsf{True}_\mathsf{h}(\mathsf{that}_2)$.
- $\pi_2 = \neg\mathsf{Q}(\mathsf{c}) \wedge \neg\mathsf{True}_\mathsf{h}(\mathsf{that}_1)$,

it is easy to prove that, in M_4, *at most* one of two sentence could be true (indeed, we can define two models, one where $\mathsf{P}(\mathsf{b})$ is true and $\mathsf{Q}(\mathsf{c})$ is false, and another in which they are both false).

REFERENCES

[1] ACZEL, P. (1988), *Non-well-founded sets*, Stanford: CSLI Lectures Notes, 14.

[2] BARWISE, J. and ETCHMENDY, J. (1987), *The Liar: an essay on truth and circularity*, London-Oxford: Oxford University Press.

[3] BARWISE, J. and PERRY, J. (1983), *Situation and Attitudes*, Cambridge: MIT Press.

[4] BARWISE, J. (1989), "On the model theory of common knoledge", in: [5], pp. 201–220.

[5] BARWISE, J. (1989), *The situation in logic*, Stanford: CSLI Lectures Notes, 17.

[6] BARWISE, J. and MOSS L. (1996), *Vicious circles: on the mathematics of non-well-founded phenomena*, Stanford: CSLI Lectures Notes, 60.

[7] BATESON, G. (1972), *Steps to an ecology of mind*, New York: Paladin Book.

[8] CANTINI, A. (2008), "Paradoxes, self-reference and truth in the 20th century", in: GABBAY, D. and WOODS, J. (eds.), *Handbook of the*

History of Logic, Volume 7, Amsterdam: Elsevier-North Holland, pp. 875–1013.

[9] CERUTI, M. (1989), *La danza che crea. Evoluzione e cognizione nell'epistemologia genetica*, Milano: Feltrinelli.

[10] DEVLIN, K., "Jon Barwise's papers on natural language semantics", *Bulletin of Symbolic Logic*, 10 (1), 2004, pp. 54–85.

[11] DEVLIN, K. and ROSENBERG, D. (2008), "Information in the study of human interaction", in: JOHAN VAN BENTHEM ET AL. (eds.), Handbook of the Philosophy of Information, Amsterdam: Elsevier-North Holland, pp. 685–710.

[12] DEVLIN, K. (2008), "Situation theory and situation semantics", in: GABBAY, D. and WOODS, J. (eds.), *Handbook of the History of Logic*, Volume 7, Amsterdam: Elsevier-North Holland, pp.601–664.

[13] DEVLIN, K. (2009) "Modeling real reasonin", in: SOMMARUGA G. (ed.), *Formal Theories of Information*, Berlin: Springer-Verlag (Lecture Notes in Computer Science), 2009, pp.234–252.

[14] DI PASQUALE, R. and RIVOLTA, A. (2008), *Sentieri e mappe dell'identità narrante. Una proposta metodologica*, Bergamo: Sestante Edizioni.

[15] GLANZBERG, M. (2004), "A Contextual-Hierarchical Approach to Truth and the Liar Paradox", *Journal of Philosophy*, 33, pp. 27–88.

[16] HOFTSTADTER, D. (1979), *Gödel, Escher, Bach. An eternal golden braid*, New York: Basic Books.

[17] LADRIÈRE, J. (1957), *Les limitations internes des formalismes*, Paris: Gauthier-Villars.

[18] RUSSELL, B. (1908), "Mathematical logic as based on theory of types", *American Journal of Mathematics*, 30, pp.222–262.

[19] TARSKI, A. (1935), "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica*, 1, pp. 261-405.

[20] VON FOERSTER, H. (1987), *Sistemi che osservano*, Roma: Astrolabio.

[21] WATZLAWICK, P., BEAVIN, J. and JACKSON, D. (1967), *Pragmatics of human communication*, New York: W.W Norton and Co..

[22] WIENER, N. (1950), *The human use of human beings*, Boston: Houghton Mifflin Co..