Bivalence, Excluded Middle and Non Contradiction

Jean-Yves Béziau

« Ou il est faux, ou il est vrai, que ma femme me trompe, mais si le tiers n'est pas exclu, alors il se peut qu'il soit faux que ma femme me trompe et qu'elle ne me trompe pas, et dans ce cas, sans ambivalence, je ne suis point cocu. »

Baron de Chambourcy (1887, p.231)

0. Introduction

The relations between the principles of bivalence, excluded middle and non contradiction are not necessarily clear. One reason is that there are various formulations of these principles, not always equivalent. In this paper, we try to formulate precisely these principles in order to discuss their relationships and to answer the question: is the conjunction of the principles of excluded middle and non contradiction equivalent to the principle of bivalence?

1. The principle of bivalence

1.1. Formulations of the principle of bivalence

The principle of bivalence can be formulated as:

(B) A proposition is either true or false.

The "or" has to be understood here as exclusive so that (B) has to be understood as the conjunction of

(B1) A proposition cannot be neither true nor false.
(B2) A proposition cannot both be true and false.

In a mathematical semantic, true and false are considered as objects, called truth-values, and a relation is established between the set of propositions and the set of truth-values. Let us call such a relation, an *evaluation relation*.

We can reformulate the principle of bivalence as the conjunction of the four following principles:

(Ba) The set of truth-values is limited to two-values.
(Baa) These truth-values are true and false.

(Bf) The evaluation relation is a function.
(Bt) The evaluation relation is a total relation.

These four principles can be put together as follows:

(BM) The evaluation relation is a total function whose domain is the set of propositions and whose codomain is a set of two truth-values, true and false.

1.2. Bivaluations and bivalent semantics

A function obeying (BM) is called a *bivaluation* and a *bivalent semantics* is defined as a set of bivaluations. So a bivalent semantics obeys the principle of bivalence in the sense that any of its constituents obeys it. What is a semantics which does not obey the principle of bivalence is not necessarily clear as we will see later on.

Now what does mean the expression “a logic obeys the principle of bivalence” or “a bivalent logic”? This is quite ambiguous, since a logic can have different semantics, some obeying the principle of bivalence, some other not. A reasonable definition is the following: a *logic is bivalent* iff it has at least one (sound and complete) semantics which obeys this principle.

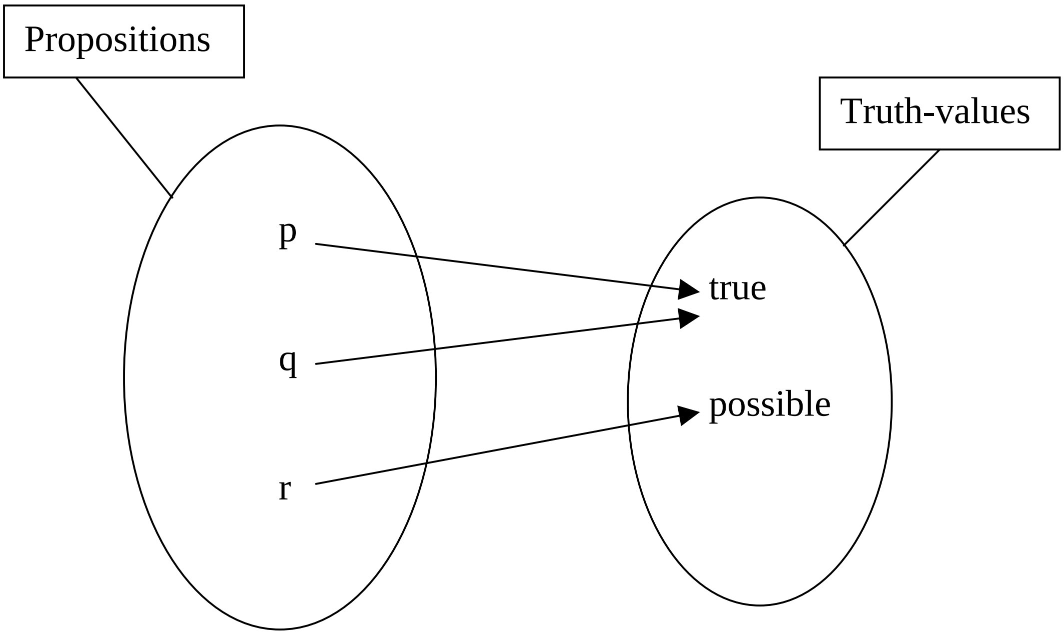
It has been shown that, if we consider a logic as a consequence relation, obeying some elementary properties (identity, monotony and generalized transitivity), then it always has a (sound and complete) semantics. You just have to take the characteristic functions of closed theories. If moreover the consequence relation is finite, you can take the characteristic functions of relatively maximal theories (cf. da Costa & Béziau 1994b).

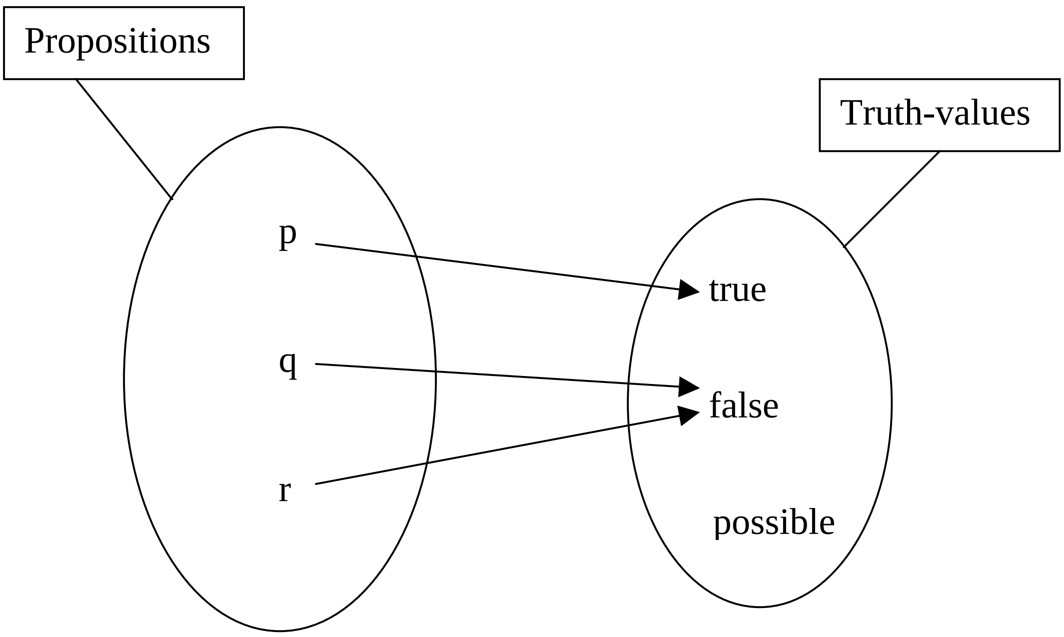
For these reasons, some people have claimed that “every logic is bivalent”. In fact standard modal logics, intutionistic logic, polar logic and many many-valued logics are all bivalent in this sense (cf. Béziau 1997).

1.3. Violation of the principle of bivalence

1.3.1. Pathological cases

There are many cases of pathological cases where the the principle of bivalence does not hold. Sometimes a good picture is better than a long discourse. So we give two pictures below before examining more serious cases.





1.3.2. Many-valued semantics

In many-valued semantics, there are more than two values. For example, Lukasiewicz introduced a third value, added to “true” and “false”, that he called “possible” or “undetermined” (Symbolically denoted by ½; 0 and 1 being used as usual to denote falsity and truth). Each proposition is either true, false or possible, but cannot have more than one value, e.g., cannot be true and possible at the same time. So, apparently, this 3-valued semantics violates (Ba) and (Baa), but not (Bf) and (Bt).

However it seems that the principle of bivalence reappears at another level, since this set of three values is divided in two in order to define the notion of logical truth. The values “false” and “possible” are considered as *non-distinguished* values and the value “true” is considered as *distinguished*. A proposition is said to be a *logical truth* iff its value is distinguished for every valuation. This is in fact the standard procedure used in any many-valued semantics, whatever the number of values is: the values are divided into two sets and this division is the cornerstone of the definition of logical truth (and logical consequence).

Using the valuations of a many-valued semantics, one can define a evaluation relation obeying (BM): by defining true as the set of distinguished values and false as the set of non distinguished values, and by stating that a proposition is true iff its value is 1, a proposition is false iff its value is ½ or 0. This forms indeed a bivalent semantics for L3, the usual one, which is formed by the set of characteristic functions of relatively maximal theories.

One can argue that it is misleading to call some of the values of a many-valued semantics true and false, and that these words should be used indeed to denote the sets of distinguished and non distinguished values (cf. Suszko 1977, Tsuji 1998).

1.3.3. Gaps and Gluts

Consider a semantics in which we have evaluation relations obeying (Ba), (Baa), (Bf) but not (Bt). Such evaluation relations are functions from propositions into the set of two objects, true and false, but are not total. In this case some propositions may have no values. We have something called *truth-value gaps*. Obviously this is not very different from having a three-valued semantics in which all these “gappy” propositions have a truth-value called, for example, “undetermined”. In fact, mathematically it is exactly the same, in the sense that for any gappy semantics, we can construct a gapless semantics which defines the same logic. However many-valued semantics are a lot more sophisticated because different values can represent different degrees of gapness. So that one can express the fact that a proposition has absolutely no truth-value or is not so dramatically gappy (cf *fuzzy logic*).

The point is that the same thing that has been said about many-valued semantics applies to partial semantics. It is not necessary correct to still call the truth-values, “true” and “false”, when there are gaps. And even with gaps, a principle of bivalence is still present and is used in a crucial way to define the notion of logical truth. Moreover gappy logics are bivalent in the sense that it is possible to find some bivalent semantics for them.

Now consider a semantics in which we have evaluation relations obeying (Ba), (Baa), (Bt) but not (Bf). Such evaluation relations are not functions and a proposition can both be true and false. This is a kind of dual situation of the situation of partial semantics and people sometimes say that we have *gluts*. We can also here, given any glutty semantics, reconstruct a many-valued semantics, where an additional value will play the role of these gluts. And what we have said about partial semantics applies here again. It applies also to semantics with gaps *and* gluts. The four-valued semantics of Belnap (1977) can be seen in fact as a representation of a gappy and glutty semantics without real gaps and gluts. And Belnap logic is bivalent in the sense that one can provide a bivalent semantics for it.

1.3.4. Do Kripke semantics violate the principle of bivalence?

If one considers a bivalent semantics as a set of bivaluations, one may wonder if Kripke semantics are bivalent semantics and if the principle of bivalence holds in a Kripke semantics. A standard Kripke structure can easily be considered as a set of bivaluations, with an accessibility relation defined between these bivaluations (cf. Béziau & Sarenac 2004). This is a good reason to say that Kripke semantics obey the principle of bivalence and that logics defined by Kripke semantics are bivalent.

Of course, it is also possible to consider Kripke structures which are sets of trivaluations or n-valuations, representing for example incomplete and/or inconsistent worlds (cf. Béziau 2001 and 2004). What we have said about many-valued semantics applies to such semantics.

1.4. Bivalence and proposition

One can consider that the principle of bivalence (B) is a definition of the notion of proposition. In this case objects to which values other than true or false are attributed or to which no values are attributed should not be called propositions, but formulas, or whatever. In fact if one considers (B) as a definition of the notion of proposition, he should consider semantics obeying other principles as defining other notions whose nature he must specify if he doesn’t want to only play formal games.

Admitting (B) as a definition of proposition, in case of a bivalent logic, even if not bivalent at first sight, one can call the objects of this logic, propositions (which is a better name that the meaningless terminology “formulas”).

1.5. Bivalence and negation

A striking feature of the principle of bivalence is that no negations appear in its formulation, this is in fact a important difference with the principles of excluded middle and non contradiction as we will see.

However someone may say that false means “not true”, so that the principle of bivalence should be formulated as

(B’) A proposition is either true or not true.

But still the principle is not about the negation of propositions, it is not about negation as a connective, but negation as a truth-function, as a function defined on truth-values. On the contrary, the principles of excluded middle and non contradiction can be viewed as principles about negation as a connective.

Unfortunately most of the introductory books of logics make the confusion between negation as a connective and negation as a truth-function (see Béziau 2002a). This confusion tends in turn to generate a confusion between the principle of bivalence and the principles of excluded middle and non contradiction.

2. The principles of excluded middle and non contradiction

2.1. The principle of excluded middle

We state the principle of excluded middle as follows:

(EM) A proposition p and its negation ~p cannot be false together.

If we have a bivalent semantics, this can be expressed equivalently, using 0 for false and 1 for true, in the two following ways:

For every bivaluation b, if b(p)=0, then b(~p)=1.
For every bivaluation b, if b(~p)=0, then b(p)=1.

Sometimes people say that the principle of excluded middle corresponds to:

(B1) A proposition cannot be neither true nor false.

Consider a bivalent semantics in which there exists a bivaluation b such that b(p)=0 and b(~p)=0. Then (B1) holds, but not (EM). This is a good reason not to call (B1) the principle of excluded middle. (B1) is (half) of the principle of bivalence. There are numerous logics in which this half of the principle of bivalence holds but not the principle of excluded middle.

What about the converse? It all depends what we admit as a semantics not obeying the principle of bivalence. If one sustains that a semantics like Lukasiewicz’s three-valued semantics is violating the principle of bivalence, calling 1, true, 0, false and the third value ½, possible, then (EM) holds but not the principle of bivalence.

This is not necessarily very convincing, since when we have the standard condition for disjunction, it follows that in a bivalent semantics (EM) is also equivalent to

For every bivaluation b, b(pv~p)=1 (i.e. pv~p is a logical truth).

In Lukasiewicz's logic L3, p and ~p can both be possible. Since possible is non distinguished, they can therefore both be non distinguished, hence pv~p is not a logical truth of L3. In this case, we cannot equate "pv~p is a logical truth" with (EM), which seems unfortunate. So this is a further reason to say that Lukasiewiz's 3-valued semantics does not violate the principle of bivalence and that in this semantics false corresponds not to 0 but to the set of non distinguished values {0, ½} and true corresponds not to 1 but to the set of disguished value {1}.

In case of a many-valued semantics, we can reformulate the principle of excluded middle as follows:

(EMM) A proposition p and its negation ~p cannot be non distinguished together.

2.2. The principle of non contradiction

The problem of the principle of non contradiction is quite similar to the problem of the excluded middle, a little trickier though.

We state the principle of non contradiction as follows:

(NC) A proposition p and its negation ~p cannot be true together.

If we have a bivalent semantics, this can be expressed equivalently, using 0 for false and 1 for true, in the two following ways:

For every bivaluation b, if b(p)=1, then b(~p)=0.
For every bivaluation b, if b(~p)=1, then b(p)=0.

Sometimes people says that the principle of non contradiction corresponds to:

(B2) A proposition cannot both be true and false.

Consider a bivalent semantics in which there exists a bivaluation b such that b(p)=1 and b(~p)=1. Then (B2) holds, but not (NC). This is a good reason not to call (B2) the principle of non contradiction. (B2) is (half) of the principle of bivalence. There are numerous logics in which this half of the principle of bivalence holds but not the principle of non contradiction.

What about the converse? It all depends what we admit as a semantics not obeying the principle of bivalence. If one sustains that a semantics like Lukasiewicz's three-valued semantics is violating the principle of bivalence, calling 1, true, 0, false and the third value ½, possible, then (NC) holds but not the principle of bivalence.

So far, so identical.

~(p&~p) is not a logical truth in L3, because when p and ~p are both possible, p&~p is possible and also ~(p&~p), therefore non distinguished. We cannot equate "~(p&~p) is a logical truth" with (NC), but this is not necessarily unfortunate, because in a bivalent semantics (NC) is not necessarily equivalent to "~(p&~p) is a logical truth", even with the standard condition for conjunction.

If we have the standard definition of semantical consequence, in any bivalent semantics (NC) is equivalent to:

For any proposition q, q is a consequence of p and ~p.

Paraconsistent logics are logics in which this does not hold, and there are a lot of paraconsistent logics in which ~(p&~p) is a logical truth. On the other hand, L3 is not paraconsistent but ~(p&~p) is not a logical truth of L3.

If we equate (NC) with "for any proposition q, q is a consequence of p and ~p", then we have the same situation as when we equate (EC) with "pv~p is a logical truth" and we reach the same conclusions.

In particular, we can reformulate the principle of non contradiction as follows:

(NCM) A proposition p and its negation ~p cannot be distinguished together.

2.3. Are the principles of excluded middle and non contradiction together equivalent to the principle of bivalence?

We can find thousands of logicians, men or women, Aristotelian or Fregean, stupid or clever, who have said that the conjunction of the principle of excluded middle and the principle of non contradiction is equivalent to the principle of bivalence (samples are given in da Costa, Béziau & Bueno 1996).

But if we put together the principles of excluded middle (EM) and non contradiction (NC), we have good reasons to consider that they are not equivalent to (B1) plus (B2), i.e. to the principle of bivalence.

There are a lot of logics which are bivalent but do not obey neither (EM) nor (EC), this is the cases of logics which are both paraconsistent and paracomplete, like Belnap's so called four-valued logic.

Once one admits the possibility of constructing logics in which there are negations not obeying (EM) or (EC), one must reject the equation:

(B) = (EM)+(NC)

unless he would like to change radically his conception of logic. The reason is that, as recent works have shown, a logic can always be considered as bivalent.

What is not very clear is a situation where the principle of excluded middle and non contradiction hold but not the principle of bivalence. This does not mean that it is obvious that (EM)+(NC) implies (B). The problem is that we don't have good examples of logics which are not bivalent in a strong sense. Before such examples are presented, this implication is simply true by default.

3. Conclusion

We can say that it is either true or (understood as exclusive) false that snow is white, but "snow is white" and "snow is not white" can both be false, can both be true. "snow is not white" does not mean necessarily that "snow is white" is false. « "snow is not white" is true » is not necessarily equivalent to « "snow is white" is false », or « "snow is white" is not true ».

Note that this does not contradict Tarski's T schema. This may seem strange to someone who doesn't know the elementary basis of modern logic, and in particular who doesn't make a difference between *not* as a connective and *not* as a truth-function.

"snow is white" and "snow is not white" can both be true means that we can find a logic in which there is a relatively maximal set of propositions, not trivial, containing these two propositions. This logic is bivalent, since it has a semantics made of relatively (non trivial) maximal sets of propositions. Of course one can wonder if in this logic, *not* is still a negation, but this is a question we have dealt with elsewhere (cf. Béziau 2002b). Anyway, if one considers that the principle of non contradiction should always hold for a negation, then the principle of bivalence trivially implies it, the same for the principle of excluded middle.

The construction of logics having negations not obeying the principle of non contradiction and/or not obeying the principle of excluded middle, together with the development of a theory of bivaluations showing how to construct bivalent semantics for a wide class of logics, have led clearly to the rejection of the equation:

$$(B) = (EM)+(NC)$$

More precisely, they have led to the rejection of the belief that the principle of excluded middle or the principle of non contradiction are consequence of the principle of bivalence. These constructions have been carried out by Newton da Costa and his school (cf. e.g. Loparic & da Costa 1984) and this consequence is an important philosophical aspect of their work.

References

N.Belnap, 1977, "A useful four-valued logic" in *Modern uses of multiple-valued logic (Proceedings of the Fifth International Symposium on Multiple-Valued - Bloomington 1975),* Reidel, Dordrecht, pp.5-37

J.-Y.Béziau, 1997, "What is many-valued logic?" in *Proceedings of the 27th International Symposium on Multiple-Valued logic – Antigonish 1997,* IEEE Computer Society Press, Los Alamitos, pp.117-121.

J.-Y.Béziau, 2000, "What is paraconsistent logic?" in *Frontiers of paraconsistent logic (Proceeding of the 1st World Congress on Paraconsistency - Ghent 1998),* Research Studies Press, Baldock, pp.95-111.

J.-Y.Béziau, 2001, "The logic of confusion" in *Proceedings of the International Conference on Artificial Intelligence IC-AI'2001 – Las Vegas 2001*, CSREA Press, Las Vegas, pp.821-826

J.-Y.Béziau, 2002a, "The philosophical import of Polish logic" in *Logic, Methodology and Philosophy of Science at Warsaw University*, M.Talasiewicz (ed), Semper, Warsaw, pp.109-124.

J.-Y.Béziau, 2002b, "Are paraconsistent negations negations?" in *Paraconsistency, the logical way to the inconsistent (Proceeding of the 2nd World Congress on Paraconsistency - Juquehy 2001)*, W.Carnielli et al. (eds), Marcel Dekker, New-York, pp.465-486.

J.-Y.Béziau, 2004, "Many-valued and Kripke semantics" in *The age of alternative logics*, Kluwer, Dordrecht, forthcoming.

J.-Y.Béziau & D.Sarenac, 2004, *Possible worlds: a fashionable nonsense?*, forthcoming.

L. de Chambourcy (Baron de Chambourcy), 1887, *Les aventures d'un logicien au pays des femmes*, Bourisse, Paris.

N.C.A da Costa, 1997, *Logiques classiques et non classiques* (Translated from the Portuguese and with appendices by Jean-Yves Béziau), Masson, Paris.

N.C.A da Costa & J.-Y.Béziau, 1994a, "La théorie de la valuation en question" in *Proceedings of the Ninth Latin American Symposium on Mathematical Logic – Bahia Blanca 1992*, Universidad del Sur, Bahia Blanca, pp.95-104.

N.C.A da Costa & J.-Y.Béziau, 1994b, "Théorie de la valuation", *Logique et Analyse*, 37, 95-117.

N.C.A da Costa, J.-Y.Béziau & O.A.S.Bueno, 1996, "Malinowski and Suszko on many-valued logics: on the reduction of many-valuedness to two-valuedness", *Modern Logic* 6, 272-300.

A.Loparic & N.C.A. da Costa, 1984, "Paraconsistency, paracompleteness, and valuations", *Logique et Analyse* 27, 119-131.

R.Suszko, 1977, "The Fregean axiom and Polish mathematical logic in the 1920s", *Studia Logica* 36, 377-380.

M.Tsuji, 1998, "Many-valued logics and Suszko's Thesis revisited", *Studia Logica* 60, 299-309.

Jean-Yves Béziau
Professor SNF at the Institute of Logic, University of Neuchâtel, Switzerland
Member of the Locia Project - CNPq, Brazil
jean-yves.beziau@unine.ch
www.unine.ch/unilog

Thanks to Alexandre Costa Leite for comments.
Work supported by a grant of the Swiss National Science Foundation.