"Can possible worlds be used for analyzing counterfactual conditionals?"

Dr. Jesús Jasso Méndez¹

Abstract: The goal of this paper is not to elaborate an elucidation of the notion of possible world, but to show some of the difficulties the usage of this notion yields if it is considered to be the ontological basis of a truth-conditional semantics for counterfactuals.

Key words: *Possible worlds; Conditionals; Counterfactual conditionals; Semantic; Lewis; Kripke.*

¹ Doctor en Filosofía de la Ciencia, área Filosofía de las Matemáticas y Lógica de la Ciencia, IIFs – FFyL, UNAM. Estancia de Investigación doctoral: Grupo LOGOS, Universidad de Barcelona, España. Profesor-investigador, Academia B de Filosofía e Historia de las Ideas, UACM-Tezonco, Área: Lógica y Epistemología.

Thanks to Carlos Jimenez for his valuable assistance in Spanish to English translation of this paper.

The concept of possible world is one of the most broadly used concepts in contemporary Philosophy of Logic. After Kripke's work, this notion is usually thought to be an appropriate basis for the building of semantic theories which interpret modal systems —in particular, systems including propositions which have the form of counterfactual conditionals.

The goal of this paper is not to elaborate an elucidation of the notion of possible world, but to show some of the difficulties the usage of this notion yields if it is considered to be the ontological basis of a truth-conditional semantics for counterfactuals.

Our strategy to achieve this goal will be as follows:

Introducing an standardized analysis of the form of a counterfactual conditional.

Enumerating three basic requirements a semantic interpretation of counterfactual conditionals must not only consider but satisfy.

Analyzing some of the problems which arise when (ii) are attempted to be satisfied using the notion of possible world as the ontological basis of semantic interpretation —for this we will consider David Lewis' realistic approach as it was put forward in Counterfactuals (1973)² and Kripke's conceptualist version of possible worlds found in Naming and Necessity (1972)³.

² In particular, we are refering to 4. Foundations: 4.1 Possible Worlds and 4.2 Similarity of Lewis, D (1973), Counterfactuals, B. Blackwell, Oxford.

³ Kripke, S (1972), Naming and Necessity, Harvard University Press. Spanish versión by Margarita M. Valdés (1980), Instituto de Investigaciones Filosófica-UNAM, México.

Finally, we will try to explain in which sense, the difficulties which arose in (iii) haven't been successfully dealt by these proposals.

This paper has been divided in two main sections: 1. Counterfactual Conditionals: A Semantic Interpretation, and 2. Possible Worlds and Counterfactuals: Semantic Problems. The former deals with (i) and (ii), while the latter deals with (iii) and (iv).

Counterfactual Conditionals: A Semantic Interpretation.

A counterfactual conditional has the following form:

If it had been p, then it would have been q

This kind of conditionals express a non-actual state of affairs. In other words, the antecedent of (a)⁴ in the actual world is always false. What the antecedent of a counterfactual could express ranges from any possible⁵ fact —i.e. stating a non-real fact of the actual world— to any logically impossible fact. Our argument is focused only on the first kind of counterfactuals, that is, the ones which have the form of (a) and express a possible fact. The justification for our choice is simple: Regarding natural language, this kind of <u>counterfactuals are</u> the most frequently used⁶. Current assessments ⁴ The antecedent of (a) "If it had been p, then it would have been q" is "If it had been p,"

194

⁵ A possible fact is one which is non-contradictory, logically speaking, e.g. Today could be cold, even though it is actually quite hot. However, it could not be the case that today were cold and not cold at the same time.

⁶ We must be aware that not everything that can be said about counterfactuals of the form (a) which express possible facts could be extended to all kind of counterfactuals. Some semantic generalizations could be made, but dealing with counterfactuals that have an antecedent expressing both logical and factual impossibilities comprises semantic singularities that require an analysis of a different sort.

"Can possible worlds be used for analyzing counterfactual conditionals?"

of such counterfactuals —i.e. giving a semantic interpretation of them in truth-conditional terms— systematically appeal to possible worlds discourse. From our point of view, this seems to lead to certain non-solvable problems. Let us see.

Let us assume there is a close relationship between the concept of possible world and an interpretation of counterfactual conditionals. Considering the primary object of our analysis, this relationship will be expressed as follows: if the antecedent of a counterfactual conditional of the type (a) is always false in the actual world, we should find at least a possible world where this antecedent is true and assess afterwards if the consequent in that world is also true. In other words, we should find a world where p and then q are true —according to (a)—, i.e. a possible world where not only p is true, but also the material conditional "if p then q" is.

If we could obtain these results, it should be assumed our semantic theory satisfies certain requirements such that truth conditions for counterfactuals of type (a) could be formulated within our theory. We will consider three basic requirements:

Number and Kinds of Objects:

i.i The set of objects of different possible worlds should not necessarily be the same of the actual world.

i.ii It is certainly possible to have worlds with no individuals in common, but if the objects of a possible world are the same than the ones of the actual world, then the theory should allow their properties and relations among them to be different from the ones that hold in the actual world. i.iii Either the set of objects of the actual world could include the set of objects of a possible world as a subset, or the former set could be included as a subset of the set of objects of a possible world

Identity Criterion: If the actual world must be related to at least a possible world for interpreting a counterfactual of type (a), a criterion for deciding when an object in the actual world is the same than an individual in a possible world must be used.

Criterion for Choosing Possible Worlds: If the antecedent of a counterfactual of type (a) is always false in the actual world, we must find at least a world where such antecedent is true ¿Which conditions should be observed to select one possible world over another if the set of possible worlds is infinite? We must have a criterion at hand, not only for making choices possible, but for having them grounded on a certain order —e.g. degrees of similarity or closeness to the actual world.

Having presented these conditions, we will see what kind of problems the realist approach —defended by Lewis— and the conceptualist approach —defended by Kripke— must cope with.

Possible worlds and counterfactuals: semantic problems

Regarding (i), we think most people will accept its content. In fact, when we refer to the set of objects of all possible worlds related to our actual world, we are obviously including the objects which are a product of our imagination. We have the capability of imagining that the actual world's state of affairs could have been different, even though some of the objects that exist in our imagination are actually considered just fiction. Precisely, the only way to deny (1) is avoiding the inclusion of these latter conditions, which seems to be far too counter-intuitive. In many occasions, we just simply consider the idea that things could have been otherwise.

Regarding requirements (ii) and (iii) we must have at hand an explanation of the notion of possible world in order to understand properly what such semantic requirements demand and therefore be able to assess the different proposals being made. We will try to clear this explanation up after Kripke's and Lewis' proposals.

Kripke maintains a conceptualist interpretation of 'possibe worlds'. In Naming and Necessity (1972), he sates that possible worlds constitute different ways of imagining or conceiving counterfactual situations, i.e. non-actual situations:

Further, if one wishes to avoid the Weltangst and philosophical confusions that many philosophers have associated with the 'worlds' terminology, I recommended that possible state (or history) of the world', or 'counterfactual situation' might be better (Kripke, 1972, p. 15).

From a realist approach David Lewis in Counterfactuals (1973) calls for the existence of possible worlds. Possible worlds are isolated entities which have an ontological status of its own and are not related to language or thought at all. They are exactly the same kind of entity that our actual world is. They have a real and concrete character⁷. So, for an individual of another world, our actual world is just another possible world:

I emphatically do not identify possible worlds in any way with respectable linguistic entities; I take them to be respectable entities

⁷ The expression 'concrete world', according to Lewis, refers to a world comprised of real objects such as the ones which exist in our actual world, e.g. building, people, landscapes, animals, ecosystems, etc.

in their own right. When I profess realism about possible worlds, I mean to be taken literally. Possible worlds are what they are, and not some other thing... (Lewis, 1973, p. 85).

Our actual world is only one world among others. We call it alone actual not because it differs in kind from all the rest but because it is the world we inhabit (Lewis, Ibidem, p. 85)

How do Lewis conditions satisfy requirement (ii)? Even though Lewis maintains that possible worlds are isolated entities and that, therefore, an individual x does not exist in different worlds, but just in the actual world —i.e. the world where x exists—, requirement (ii) points toward a problem of trans-world identity that not only Kripke's conceptualist version, but Lewis realist approach must consider and overcome.

The problem of trans-world identity can be put forward in the following terms: In principle, we have an infinite number of possible worlds. If we have exhaustive knowledge of an object x in the actual world (AW), i.e. we know its intrinsic conditions of existence and the extrinsic relations it holds with other objects, we can play with its descriptions and make some variations on them. For modifying x, we should also consider making some changes in the objects which x is actually related to. So, when adjusting certain sort of relation of x into a new description, x is now considered within the context of a possible world (PW)⁸. If this is the case, how can the following question be answered: Is the object x of PW the same object x we start from in AW?

⁸ It is important not to lose sight of the fact that the relations an object x has, are elements which distinguish it from other objects. So, when modifying these relations, in certain sense, we're modifying not only x, but the set of objects related to it.

What we hold in this paper is that the answer to this question is contextually dependent on the conception of possible worlds we maintain within a theoretical framework. In what follows we will give a look at the answers that Kripke and Lewis would provide within their own frameworks.

Even though Kripke accepts that possible worlds follow from our faculty for conceiving certain situations, he rejects the idea that we are able of conceiving everything that is true or false of such situations, except for what concerns the counterfactual situation. Actually, Kripke believes it is possible to legitimately stipulate that an object in a possible world is the same than another object in the actual world, even if its properties and relations have been modified. According to Kripke, this is the behavior speakers of most linguistic communities usually have when they face or have to use a counterfactual conditional within natural language context. e.g.

A speaker says:

(b) "If Nixon hadn't won the elections, then he would have been a soccer player⁹"

According to Kripke, the speaker uses (b) to express a belief about a particular situation, i.e. the person who says (b) is imagining a situation where the man, Nixon, who won the elections and the non-winner Nixon who is a soccer player are the same individual. Considerations of this sort are not only in the basis of the thesis of proper names conceived as rigid designators, but the reason why descriptions are not conceived as such:

⁹ Let us imagine and grant the Nixon of our actual world loved to play soccer.

In these lectures, I will argue, intuitively, that proper names are rigid designators, for although the man (Nixon) might not have been the President, it is not the case that he might not have been Nixon (though he might not have been called 'Nixon' (Kripke, Ibidem, p. 49).

This way of interpreting counterfactuals on the basis of the thesis of rigid designation and the idea of possible worlds as "counterfactual situation" entails a rejection of the various problems which follow from "trans-world identification" (Cfr. Kripke, Ibidem, pág. 49-50). However, on Kripke's view, the issue of trans-world identity turns into the problematic issue of essential properties. A property of an object is essential to it, if and only if, that property forms part of the object in any place where the object exists or could exist. The proper name 'Nixon' is a rigid designator, and even if in a counterfactual situation we could conceive him as the one who did not win the elections, it could not be the case that he were not Nixon. It is precisely because we rigidly designate Nixon that we could stipulate that we are talking about Nixon in a counterfactual situation, even though we attribute him different properties from the ones he has in the actual world.

Now then, with these resources, Kripke must answer our question: Is the object x in (PW) the same object than the object x we start from in (AW)? Kripke, seems to offer an affirmative answer by appealing both to the thesis of rigid designation —as we have mentioned above— and to the idea of possible world as a "counterfactual situation".

Año I Num. 2 Julio - Diciembre 2012

However, although we could grant Kripke that the behavior of a speaker who states a counterfactual of type (b) entails that the object x in (PW) is the same object x in (AW) even though some of its properties have been changed, this —we think— does not solve the problem. ¿Can we keep saying the same about x while modifying gradually its properties and relations in different possible worlds ranging from (W1, W2 ... Wn)? ¿How far can we go maintaining we are dealing with the same object x in such a situation? If we reach a point in which the variation of properties and relations turns into a complete exchange of properties between the object x and y, including essential and accidental properties (such that it is possible to say, for instance, that Nixon is, let us say, Benjamin Franklin), is x still the same object?

From our point of view, Kripke must cope with some problems which can be better appreciated after considering the following:

if x and y do only exchange non-essential properties, there are no problems on maintaining the kripkeanian interpretation in which Nixon is the same object in different worlds (W1, W2 ... Wn),

if the set of essential and accidental properties of x are exchanged for the set of essential and accidental properties of y, it is possible to say that Nixon is not Nixon anymore, but, let us say, Benjamin Franklin,

if x and y have exchanged all their properties, but not their names, 'Nixon' and 'Benjamín Flanklin' respectively, it is possible to say that, now, the name 'Nixon' designates Benjamin Franklin and the name 'Benjamín Franklin' designates Nixon. but, according to Kripke, if 'Nixon' and 'Benjamín Flanklin' are rigid designators, these names must designate the same individual in all possible worlds —point which is in doubt after considering (iii),

if (iii) and (iv) are the case, then 'Nixon' and 'Benjamín Flanklin' are not rigid designators anymore, but secondary properties of x and y respectively, the properties of "being named Nixon" or "being named Benjamín Franklin". So, proper names will designate different objects in different possible worlds.

The point we simply want to make with the above is that defending Kripke's reduction of trans-world identity to essential properties determination is highly problematic.

If a theory deems essential the property of an object so that a world where the object does not have this property cannot be conceived, the problem comes down to a problem about the capability of a subject for conceiving possible words from her beliefs system. Therefore, a conceptualist version of possible worlds, at least Kripke's, has this burden to deal with: identifying what is possible with what is conceivable¹⁰.

Another objection is concerned with the Identity Axiom: If two individuals are identical, then they have the same properties. According to Kripke's considerations, this axiom will stop being relevant, since we can talk about the same object even when it had different properties in different worlds. So we have two options:

¹⁰ In the history of science we can find cases where this identification has shown to be inappropriate. There had been scientific facts which were real, even though, at a certain moment, they were unconceivable, and vice-versa.

"Can possible worlds be used for analyzing counterfactual conditionals?"



Or, if we refuse leaving this axiom, we must leave aside the thesis of rigid designation and with it the belief that trans-world identity of objects could be determined by having proper names working as rigid designators.

Let us talk now about Lewis. At first sight, it seems David Lewis' realist version of possible worlds, tries to brush off all the problems that Kripke's conceptualist approach has to face. Lewis supplies his possible world notion with a matching a counterpart theory. An object x exists in just one and the same world, in other worlds, there only are objects highly similar to x. Lewis calls these objects -highly similar to x- x counterparts. The objects of a particular world can have at least one counterpart in other possible worlds. If this is the case, assessing the truth value of a conterfactual of type (b) requires choosing not a possible world from the ones where the actual world Nixon had not won the elections, but one world where Nixon's counterpart has not won the elections and is a soccer player. Trans-world identity in the realist approach acquires a different shade from the one it had in the conceptualist version. Within Lewis' proposal, the identity relation turns to be a similarity relation between an object and its possible counterparts in other possible worlds.

¿Have we got any benefits from this change? ¿Is the realist version actually immune to all the problems the conceptualist version has been pointed out? Lewis admits the similarity relation between an object and its counterpart does not have the logical properties entailed by the identity criterion. The similarity relation between an object and its counterparts, and the similarity relation between worlds this former relation relies on are vague links. If a counterpart is an object y that, in an specific world, resembles an object x of another world to the highest degree regarding its extrinsic and intrinsic features, then there is no other object of that world which resembles x more than y:

...something has for counterparts at a given world those things existing there that resemble it closely enough in important respects of intrinsic quality and extrinsic relations, and that resemble it no less closely than do other things existing there. Ordinary something will have one counterpart or none at world, but ties in similarity may give it multiple counterparts (Lewis, 1973, p. 34)

Let us take a look at the following example to see how much this conception resist:

In (AW) Carnap is a philosopher who died in 1970,

In (PW1) Carnap has a counterpart who is a violinist and died in 1970,

In (PW2) Carnap has a counterpart who is a philosopher and died in 2002,

After considering (i), (ii) y (iii) the expected question is ¿which of these two counterparts resembles Carnap the most? We have at least two options:

The counterpart we are looking for is (ii) because he does not only have the same name, but lived at the same Age and died in the same year Carnap did.

204

Año I Num. 2 Julio - Diciembre 2012

The counterpart we are looking for is (iii) because he does not only have the same name, but the same profession Carnap had.

Now, If we provide our example with the following counterfactual: (c) "If Carnap had died in 2002, he would have been the best violinist of the London School of Music". ; Which counterpart -a or b- satisfies this counterfactual better? We think neither (a) nor (b) satisfy (c) properly. Anyway, we should be able to find out a counterpart in a specific world where an object had died in 2002 and at the same time had been the best violinist of The London School of Music. If we find out and object with those features in a possible world and, besides, this object has some other features relevant to the actual world Carnap such as being the son of a man and a woman who had the same name, and having written in 1947 the book Meaning and Necessity: A Study in Semantics and Modal Logic we would effectively think that object, in a possible world, is our actual world Carnap's counterpart. However, under what criteria can we tell Carnap's counterpart resembles more, or less, our actual world Carnap? ; What kind of features and how many should an object of a possible world have so we could effectively decide we are dealing with the counterpart of an object of the actual world? Where should the limits be drawn?

Lewis does not solve this problem. The inaccuracies of the realist approach regarding the object-counterpart relation are the case. And even though we could grant a degree of vagueness when interpreting counterfactual conditionals using a notion of possible world contextually dependent on particular theories, the inaccuracies maintained by Lewis seem far too much. It seems to be a conflict tough to go out from, because if Lewis effectively drew limits and fixed certain criteria for deciding which and what kind of features must and object have in order to be the counterpart of an object of the actual world in a possible world, we would fall So, as we see, neither Kripke's conceptualist version of possible worlds, nor Lewis' realist proposal deal succesfully with the semantic problem entailed by requirement (ii). Regarding this aspect, within none of these theoretical frameworks it is possible to actually assess the truth value of a counterfactual of type (a).

Let us talk about requirement (iii) now. ¿Under what criteria can we select a world —from the set of all possible worlds— for evaluating a counterfactual? What we are looking for is a possible world where the antecedent of the counterfactual of the form (a) "If it had been p, then it would have been q" is true; so we can grant that not only the antecedent, but the conditional "p, then q" is true.

According to Kripke —as it has been mentioned above possible worlds represent different worlds of imagining or conceiving counterfactual situations. Therefore, Kripke would accept possible worlds are built up from the real world, by introducing different hypothetical modifications. If this is the case, we could also say Kripke would grant the difference between close and distant worlds in terms of variation degrees, i.e., the closest worlds to the actual would be the ones whose objects have been conceived after variations of real world objects, while the most distant would be the ones which include strictly fictional objects, or fictional objects build up from reality. It is from this point of view that counterfactual conditionals are assessed.

If the antecedent of a counterfactual expresses variations from the actual, then these variations work as admissible hypotheses which yield a sort of classification of possible worlds in terms of

206

their short or long distance to the actual world. Therefore, the degree of variation depends on the acceptability of the hypothesis comprised by a particular counterfactual. Considering these results, imagining or conceiving a set of possible worlds is a constructive process: We conceive a world where the antecedent is true and then adjust the characteristics of the objects related to the contents of the antecedent accordingly.

If we grant the above, the constructive process of possible worlds is a semantic issue which relies on what an individual could conceive or imagine after the actual world.

Conceiveability, then, is a criterion which permits the building up of a possible world or a "counterfactual situation". So, what can be conceived is not any situation or any product of our imagination, but what is admissibly conceivable, i.e. if the contents of the antecedent of a counterfactual is an acceptable hypothesis, such hypothesis is only possible if it satisfies the conceiveability criterion which determines what hypotheses are admissibly conceivable. If this is the case, then, Kripke faces a new problem $\frac{1}{2}$ How can we sharply distinguish when variations of an object are admissible and when they are not? If the criterion of worlds conceiveability follows a logic of acceptability $\frac{1}{2}$ how do we conceive the most acceptable world for deciding on the truth of both the antecedent and the whole conditional —if p then q— comprised by a counterfactual of type (a)? By not offering an answer, this becomes a latent problem in Kripke's conceptualist proposal.

Jus as it happens with the conceptualist point view, from Lewis's realist assumptions regarding possible worlds ontology we could yield an order among worlds. This order depends on comparative similarity between possible worlds and the actual world. Within a scheme, the actual world is in the centre. The closest worlds would be the ones which resemble the actual the most, while the most distant would be the ones which resemble the actual the least. So, if this representation accurately mirrors Lewis proposal, we can say that if the antecedent of a counterfactual of type (a) is true in a world close to the actual, it will be true in all the worlds which are at the same distance of comparative similarity to the actual world.

Regarding this, the main problem the realist approach must face will be not only how can we determine the degree of similarity of a possible world related to the actual, but how can we order the set of possible worlds in terms of comparative similarity? As we have seen above, Lewis says that possible worlds are not related to linguistic or mental considerations, possible worlds just have the same ontological status than the actual world. They exist in the same way. However this ontological consideration does not entail an established fixed order among worlds, and this is a requirement [modal] realism must consider. Let us recall Lewis, himself, grants the vagueness of the notion of comparative similarity:

Overall similarity consist of innumerable similarities and differences in innumerable respects of comparison, balanced against each other according to the relative importance we attach to those respects of comparison. Insofar as these relative importance differ from one person to another, or differ from one occasion to another, or are indeterminate even for a single person on a single occasion, so far is comparative similarity indeterminate (Lewis, 1973, p.91)

There could be worlds which have the same degree of similarity regarding one feature, but different regarding some other. If the worlds scheme put forward by Lewis' modal realism is not fixed, then similarity variation will depend on the aspects being taken into account. In principle, this appears to be such an amazing result for a realist version of possible worlds. However, Lewis holds that even though the notion of comparative similarity is vague, we have enough elements at hand for deciding when the antecedent of a counterfactual is true in the set of worlds which share the same range of similarity related to the actual world, i.e. there is range which will vary according to the aspects under consideration and even though such a range is not fixed, it let us decide not only when a counterfactual is true in a set of worlds and not in some other set, but establish when some counterfactuals are true or false in every possible world where the antecedent is true.

Our criticism to this realist version is that there are possible worlds for which the range used for measuring the distance of possible worlds related to the actual seems to fade, i.e. it seems impossible to decide which world, from at least two, has a higher degree of comparative similarity related to the actual world.¹¹

Actually, besides having examples of quite weird worlds whose degrees of similarity related to the actual world cannot be assed, we still do not have a clear example which let us illustrate this point within the framework of counterfactuals analysis. However, intuitively, this seems to be related to the truth indeterminacy of the conditional 'if p then q' when it comprises a reference to specific times in the antecedent and in the consequent., e.g. (d) "If Laura had studied math, at a certain moment, she would not have failed the course later on". If we can certainly grant that the relation between a particular event and a particular time —in the antecedent and in the consequent— is fully determined in the actual world, such a relation between events and particular times might not the be same in other possible worlds.

¹¹ Let us imagine worlds inhabit by quite weird objects such as talking monkeys, flying tigers, flying pigs, green elephants, etc. How can we decide how far, regarding comparative similarity, are these worlds from the actual world? ¿Where are the limits between what is similar and what is not?

We know Lewis' stance regarding this would be something similar to this: the difference between the particular facts expressed by the counterfactual is more important in the antecedent than it is in the consequent. So, the antecedent of a counterfactual of type (b) will be true in those worlds where there is a similarity relation closer to the actual world in terms of a state of affairs pretty similar to the one expressed by this antecedent leaving aside the consequences the referred state of affairs could have yielded which are expressed by the consequent of the counterfactual.

However the problem is still open. There is no way to determine the truth value of a counterfactual within a vagueness framework where decisions about similarity and closeness between worlds depend, either on the belief systems of a speaker, or —if we are realists à la Lewis— on the basis of a discourse which insufficiencies are left aside.

As a conclusion: Both the realist approach and the conceptualist point of view go far away from a conception of possible worlds as ways of conceiving or imagining the different forms the actual world could vary or could have varied. Their versions of what possible worlds are, at the end, comprise features such as the conceiveability criterion for determining the acceptability of worlds, a continuous reference to the belief systems of speakers for determining degrees of comparative similarity between worlds, or for determining -- in its case- if one and the same name designates the same object in different counterfactual situations which entail certain variation of properties and relations. Actually, we think a notion of what possible world is must be given in terms of belief systems regulated by a theoretical semantic framework which attempts to avoid the problems we have mentioned, i.e. a framework which could solve the vagueness features which block the appropriate assessment of counterfactuals in truth-conditional terms. The possibility of

210



Bibliography

Lewis, D (1973), Counterfactuals, B. Blackwell, Oxford.

Kripke, S (1972), *Naming and Necessity*, Blackwell, Oxford. Versión española de Margarita M. Valdés (1980), Instituto de Investigaciones Filosófica-UNAM, México.

R. M. Chisholm (1967), "*Identity Trough Possible Worlds: some questions*", in Nous, Vol. I.