International Journal of Theoretical Physics, Vol. 44, No. 4, April 2005 (© 2005)
DOI: 10.1007/s10773-005-3853-4

Modeling Modal Talk in Quantum Mechanics

Thomas Müller[1]

Received October 13, 2004; accepted March 1, 2005

In this paper, modal and counterfactual logical connectives are defined in an extended framework of branching space-time (Belnap, N. D. (1992). Branching space-time. *Synthese* **92**, 385–434). It is shown that a variety of definitions of the counterfactual can be given. The validity of certain modal statements occurring in quantum mechanics depends on the choice of definition. These considerations can be applied to an analysis of Stapp's premises LOC1 and LOC2 from his purported proof of non-locality (Stapp, H. P. (1997). Nonlocal character of quantum theory. *American Journal of Physics* **65**, 300–304). It is shown that while the validity of LOC1 depends on the choice of the definition of the counterfactual, LOC2 is absolutely invalid.

KEY WORDS: branching space-times; modality; quantum mechanics; locality; counterfactuals.

1. INTRODUCTION

Derivations of Bell-type or GHZ-type theorems have to appeal to modal notions. This appeal to modality is often not made explicit. In order to assess the implications of the mentioned theorems, their inherent modality should be acknowledged both syntactically, through the use of modal symbolism, as well as semantically, through the use of appropriate models. Recently, Henry Stapp has given a proof that uses modal symbolism to argue that quantum mechanics of itself is nonlocal (Stapp, 1997). While that proof is based on modal syntax, it is not based on any modal semantics. This is problematic, since the details of the workings of modal operators depend to a large extent on the semantics that one has chosen to use. Especially in the case of the counterfactual conditional "if . . . were the case, . . . would be the case" (also used by Stapp), a clear account of the semantics is vital for correct formal reasoning. In this paper, I give a rigorous formal semantics for the modal operators "possibly" and "necessarily" and for the counterfactual conditional. The semantics is offered as part of a critique of

[1] Philosophisches Seminar, LFB III, Universität Bonn, 53113 Bonn, Germany; e-mail: Thomas.Mueller@uni-bonn.de.

Stapp's purported proof, but also in the hope that it should be suitable for analyzing quantum correlations in general.

The paper is organized as follows: In Section 2, I will briefly introduce the formal branching models on which the formal semantics is based. In Section 3, I will use these models to give a semantics for the modal operators and the counterfactual conditional. Finally, in Section 4, I will analyze some aspects of Stapp's proof using the semantics.

2. MODELS

The models I will use are derived from Belnap's *Branching space-time* (BST) (Belnap, 1992), recently augmented to the framework of *Stochastic outcomes in branching space-time* (SOBST) (Placek, 2000; Müller and Placek, 2001). In the SOBST models that I will employ, *histories* (also called *chronicles*) are Minkowski space-times. As the models exhibit branching, the future light cone above any point may contain alternative courses of events. These alternative courses of events are to be the basis for modal talk.

Both BST and SOBST are described in detail elsewhere (Belnap, 1992; Placek, 2000). Here, I will only give a brief outline of those aspects of the model framework that will be used later on.

2.1. Our World as a Partial Ordering

Our World is pictured as a nonempty partial ordering $\langle W, \leq \rangle$, where the elements of W are understood to be space-time points viewed as concrete particulars. For two elements $x, y \in W$, $x \leq y$ can be read "x is in the causal past of y," or "y is in the future of possibilities of x." $x < y$ is defined, as usual, as $x \leq y \wedge x \neq y$. The ordering is taken to be dense and without maximal elements. (Alternatively, a somewhat weaker condition may be imposed; cf. (Placek, 2000, p. 142).)

A *history* σ in W is a maximal upward directed subset of W, where "upward directed" means that for all $e_1, e_2 \in \sigma$, there is $e \in \sigma$ s.t. $e_1 \leq e$ and $e_2 \leq e$. Two points x, y are *compatible* iff there is a history to which they both belong. If two points x, y are incompatible, then they will belong to different histories. By the *prior choice principle*, for incompatible x, y there is a point e s.t. $e < x$ and $e < y$ at which histories containing x and y, resp., split from each other. This principle implies *historical connection*, i.e., all histories intersect ("in the distant past"). For histories σ and η, their set of *splitting points* $C_{\sigma,\eta}$ is defined to be the set of all maximal elements in their intersection $\sigma \cap \eta$.

The branching framework does not decide the question of the space-time metric. To keep things simple, in what follows I require that histories are Minkowski space-times.

2.2. Comparative Similarity of Histories

A key ingredient in the definition of the counterfactual conditional, to be given below, is the notion of comparative similarity of histories (or "worlds"). For example, in order to assess whether a counterfactual like "If kangaroos had no tails, they would topple over" is to be counted true or false, the recipe is to find, among the (possible) worlds in which kangaroos have no tails, the one that is most similar to the actual world. Then, the mentioned counterfactual is true iff in that world, kangaroos do topple over, false otherwise. The counterfactual is counted vacuously true if there simply is no situation in which the antecedent holds. (This verbal description is to be taken with a grain of salt; for the technical details, cf. (Lewis, 1973).)

On the basis of a branching structure $\langle W, \leq \rangle$, it is possible to introduce similarity orderings that have some of the technical properties needed for the definition of the counterfactual. For histories σ, η, γ, the ordering $\eta \sqsubset_\sigma \gamma$ is read "η is more similar to σ than γ." The intuition behind a branching notion of similarity is that the later two histories split, the more similar they are. On the basis of a branching structure, three similarity orderings that spell out this intuitive motivation can be defined:[2]

Definition 2.1. (Strong Version of Strict Comparative Similarity). η is more similar to σ than γ in the strong sense ($\eta \sqsubset_\sigma^S \gamma$) iff the set of splitting points between σ and γ, $C_{\sigma,\gamma}$, causally precedes the set of splitting points between σ and η, $C_{\sigma,\eta}$, i.e., iff $\forall x \in C_{\sigma,\gamma} \ \ \forall y \in C_{\sigma,\eta} \ \ x < y$.

Definition 2.2. (Mild Version of Strict Comparative Similarity). η is more similar to σ than γ in the mild sense ($\eta \sqsubset_\sigma^M \gamma$) iff $\forall x \in C_{\sigma,\gamma} \ \exists \ y \in C_{\sigma,\eta} \ \ x < y$.

Definition 2.3. (Weak Version of Strict Comparative Similarity). η is more similar to σ than γ in the weak sense ($\eta \sqsubset_\sigma^W \gamma$) iff $\forall x \in C_{\sigma,\gamma} \ \exists \ y \in C_{\sigma,\eta} \ x \leq y$ and for some $x' \in C_{\sigma,\gamma}$, $y' \in C_{\sigma,\eta} \ x' < y'$.

These definitions yield three partial orderings $\sqsubset$, "strictly more similar than." The definition of the counterfactual, to be given below, is based on an ordering $\sqsubseteq$, "at least as similar as," which is defined as

$$\eta \sqsubseteq_\sigma \gamma \quad \text{iff} \quad \text{not} \ \ \gamma \sqsubset_\sigma \eta. \tag{1}$$

[2] For a detailed discussion of these definitions, including proofs of the formal properties, cf. (Placek, 2000, p. 153) and Tomasz Placek's contribution to the IQSA V conference (this volume). Thanks to T.P. for allowing me to reproduce the definitions here.

In Lewis' original formulation, the relation $\sqsubseteq$ is required to be a weak ordering (preordering), meaning that it satisfies the following two conditions:

1. Connectedness: For all σ, η, γ, we have $\eta \sqsubseteq_\sigma \gamma$ or $\gamma \sqsubseteq_\sigma \eta$ (and possibly both).
2. Transitivity: For all $\sigma, \eta, \gamma, \delta$, if $\eta \sqsubseteq_\sigma \gamma$ and $\gamma \sqsubseteq_\sigma \delta$, then $\eta \sqsubseteq_\sigma \delta$.

Our orderings $\eta \sqsubseteq_\sigma^S \gamma$, $\eta \sqsubseteq_\sigma^M \gamma$ and $\eta \sqsubseteq_\sigma^W \gamma$ are connected, but not transitive (since the relation of being space-like separated is not transitive). Still, the given similarity orderings allow for the definition of a counterfactual connective that captures at least most of the important counterfactual inferences.[3] The three orderings are formally equally valuable, as they all have the same formal properties. However, the definitions, and thus the orderings, are obviously not equivalent.

A transitive similarity preordering could be defined, e.g., on the basis of a real-valued measure of distance between histories. We have not been able to come up with a good intuitive motivation for such a distance measure, but a first try might be to use the distance of splitting points from a given point of evaluation as a distance measure: Set the distance between σ and η, as viewed from Y, to be

$$D(\sigma, Y, \eta) = \sum_{Z \in C_{\sigma,\eta}} |Y - Z|, \tag{2}$$

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^4$.[4] Using this distance measure, we define

$$\eta \sqsubseteq_{\sigma,Y} \gamma \quad \text{iff} \quad D(\sigma, Y, \eta) \geq D(\sigma, Y, \gamma). \tag{3}$$

3. SEMANTICS

Based on the branching models, I now define modal operators and the counterfactual connective. Modal assertions are to be evaluated from the perspective of a certain space-time point Y and a certain history σ from Our World W.

3.1. Possibility and Necessity

There are two modal operators that are commonly employed: a strong operator, read "necessarily" ($\Box$), and a weak one, read "possibly" ($\Diamond$). Since the two modalities are inter-definable, I will only treat the weak modality and define "necessarily" to mean "not possibly not."

One has to distinguish a number of different concepts of modality in order to select the right one to use in the description of quantum correlation experiments.

[3] Cf. (Lewis, 1981) for a definition of the counterfactual that is based directly on a partial ordering $\sqsubset$.

[4] Note that D will thus be frame-dependent. This is awkward, but the (Lorentz-invariant) Minkowskian space-time distance cannot be used in the definition, as it is not necessarily positive, leading to counterintuitive results.

A statement $\phi(X)$ (stating something about a space-time point X) can be called "possible" at a space-time point Y in a history σ in at least the following five senses:

1. $\diamond_1$, logical possibility, meaning the absence of contradiction. This is the most liberal notion of possibility, counting everything as possible that is not contradictory. For our purposes, this is too wide.
2. $\diamond_2$, physical possibility, meaning absence of contradiction with established physical laws. This concept is also too broad for our purposes. Furthermore, it relies on a notion of physical laws that is not uncontroversial.
3. $\diamond_3$, possibility in the given model of Our World W. This is the first useful concept. It states that $\phi(X)$ is possible in W (in any history and at any point – this information is discarded) iff there is a history η such that $\phi(X)$ holds in η.
4. $\diamond_4$, possibility in the given universe based on accessibility: $\phi(X)$ is possible in σ at Y iff there is a history η that is accessible from σ (i.e., $\sigma \cap \eta \neq \emptyset$) in which $\phi(X)$ is true at Y. This concept will usually coincide with $\diamond_3$, since normally we assume historical connection. If historical connection does not hold, however, $\diamond_4$ is a different concept. Most importantly, $\diamond_4$ is the concept of the outer modalities based on the counterfactual to be defined below.
5. $\diamond_5$, possibility based on reality. This is the most important concept. According to $\diamond_5$, $\phi(X)$ is possible in a history σ and at a space-time point Y iff either (1) X is in the past or present of Y (i.e., outside the future light cone above Y), and in history σ, $\phi(X)$ is or was in fact true, or (2) X is in the causal future of Y, and in some branch of the universe above Y (i.e., in some future that is accessible from Y in history σ), $\phi(X)$ is true. This concept is called "possibility based on reality" because the real course of events determines what is possible: either something has become actual and is thus possible as well as necessary, or it is still open to occur, given what has occurred so far. (For the phrase "possibility based on reality," cf. also (Xu, 1997).) According to this notion, possibility and necessity coincide for the past, but differ for the future. This is exactly as it should be: the past is fixed, the future is open.

Since the concept $\diamond_5$ is the important one in our context, I use the simple "$\diamond$" for it. To give the formal definition explicitly: $\diamond\psi(X)$ is true at history σ and at point Y iff there is a history η in W such that σ and η agree for all points outside the future light cone above Y, and $\phi(X)$ holds at Y in η.

Example: At the source of a quantum correlation experiment (point Y = location of source), all outcomes $L\alpha+$ (left: setting α, outcome $+$), $L\beta-$ (setting β, outcome $-$), etc. are possible in this sense. After the setting α has been selected on the left (point Y above the selection event in an α-branch), $L\beta+$ is *no longer possible*.

3.2. Counterfactuals

In a situation where something is no longer possible, it may have been possible, and one can sometimes sensibly ask what *would be the case, had something else been the case*. For simplicity's sake, I deal with counterfactuals that have this asymmetric temporal structure only; they are the ones that figure importantly in reasoning about hidden variables and non-locality.

The basic form of the "would" counterfactual[5] is thus, "at Y in history σ, if $\phi(X)$ had been the case, $\psi(X')$ would be the case," symbolized as "$\phi(X) \Box\!\!\rightarrow \psi(X')$." In line with Lewis' definition (Lewis, 1973), I will use two auxiliary notions: the set W_σ of histories *accessible* from σ, defined as the set of histories that share some past region with σ,[6] and the weak ordering $\sqsubseteq_\sigma$ from the family $\sqsubseteq$ defined above in Section 2.2, where $\eta \sqsubseteq_\sigma \gamma$ means "history γ is no more similar to σ than η." Given these notions, the formal definition of the counterfactual reads as follows: The counterfactual statement "if $\phi(X)$ had been the case, $\psi(X')$ would be the case" is true in history σ at space-time point Y iff either (1) there is no history in W_σ in which $\phi(X)$ holds, or (2) there is a history η in W_σ in which $\phi(X)$ holds, and for all γ in W_σ, the following holds: If $\gamma \sqsubseteq_\sigma \eta$,[7] then in γ, the plain (material) conditional "if $\phi(x)$, then $\psi(X')$" holds: in γ, either $\phi(X)$ is false or $\psi(X')$ is true (or both).

3.3. Outer Modalities

From a counterfactual, one can retrieve a notion of conceivability, the so-called outer modalities (cf. Lewis, 1973, p. 22]). Relative to $\Box\!\!\rightarrow$, it is possible that $\phi(X)$ iff $\phi(X)$ is true in some history in W_σ. This is a far wider notion of possibility than the notion employed for possibility based on reality—indeed, it is the notion behind $\Diamond_4$, as already advertised above. This makes sense: the counterfactual by its very name is not tied to reality and thus needs to take into account more than what is still possible.

4. STAPP'S PURPORTED PROOF OF NON-LOCALITY

Henry Stapp has recently given a formal proof of non-locality from quantum mechanics alone, making heavy use of modal symbolism (Stapp, 1997). The overall structure of the proof is the following: Stapp starts with three premises

[5] The "might" counterfactual is definable from this in the usual fashion. As it is not important in quantum mechanical reasoning, I do not define it explicitly; cf. (Lewis, 1973).

[6] Assuming historical connection, this set will be the set of all histories in W; cf. the discussion about $\Diamond_4$ above.

[7] If the transitive similarity ordering based on the distance function D is used, '$\sqsubseteq_\sigma$' needs to be replaced by '$\sqsubseteq_{\sigma,Y}$'.

LOC1–LOC3 that, as he claims, express the concept of locality. Stapp then goes on to show that a contradiction can be derived from these three premises and some innocuous assumptions. If this were correct, it would follow that quantum mechanics has been shown to be nonlocal, without any recourse to reasoning about hidden variables, elements of reality or the like. (With Stapp, I agree that his innocuous assumptions are indeed innocuous.)

Here is a relevant portion of the proof (lines 15 and 16 are my addition to make explicit the formal contradiction that Stapp claims to have derived):

$$L\beta \wedge R\beta \wedge L\beta+ \;\strictif\; (R\alpha \;\Box\!\!\rightarrow L\beta \wedge R\alpha \wedge L\beta+) \qquad \text{LOC1} \quad S1$$

$$L\beta \wedge R\beta \wedge R\beta+ \;\strictif\; L\beta \wedge R\beta \wedge L\beta+ \qquad \text{QM} \quad S2$$

$$L\beta \wedge R\alpha \wedge L\beta+ \;\strictif\; L\beta \wedge R\alpha \wedge R\alpha- \qquad \text{QM} \quad S3$$

$$L\beta \wedge R\beta \wedge R\beta+ \;\strictif\; (R\alpha \;\Box\!\!\rightarrow L\beta \wedge R\alpha \wedge R\alpha-) \qquad \text{LOGIC, S1, S2, S3} \quad S4$$

$$L\beta \;\strictif\; (R\beta \wedge R\beta+ \;\strictif\; (R\alpha \;\Box\!\!\rightarrow R\alpha-)) \qquad \text{LOGIC?, fromS4} \quad S5$$

$$L\alpha \;\strictif\; (R\beta \wedge R\beta+ \;\strictif\; (R\alpha \;\Box\!\!\rightarrow R\alpha-)) \qquad \text{LOC2, S5} \quad S6$$

$$\ldots$$

$$\neg\Diamond(L\alpha \wedge R\beta) \qquad \text{LOGIC, S11, S14} \quad 15$$

$$\Diamond(L\alpha \wedge R\beta) \qquad \text{FREECHOICE} \quad 16$$

The proof is based on the properties of the Hardy state (Hardy, 1992). "$L\beta$" is to be read as "in the left wing of the experiment, setting β has been chosen," "$R\alpha+$" as "in the right wing, α was chosen, and the outcome was $+$," etc. Besides the counterfactual conditional "$\Box\!\!\rightarrow$" and the possibility operator "$\Diamond$," the proof employs the usual symbols "$\wedge$" for "and," "$\neg$" for "not," and the strict implication "$\strictif$," where "$\phi \strictif \psi$" is defined as "necessarily, if ϕ then ψ."

In its published form (Stapp, 1997), the proof contains formal errors resulting from an incautious mixing of strict and material conditionals.[8] These problems can however be circumvented.[9] For an assessment of the proof, it is conceptually most important to analyze the premises LOC1 and LOC2 that Stapp refers to in lines S1 and S6.

LOC1. Stapp's premise LOC1 "asserts that if under the condition that the choices were $L2$ [setting on the left] and $R2$ [on the right] the outcome in **L** at some earlier time were $L2+$, then if the (later) choice in **R** were to be $R1$, instead

[8] Cf., e.g., the inference from line S4 to S5—A very detailed and careful analysis of Stapp's proof is given in (Shimony and Stein, 2003), but their discussion is not based on an explicitly given formal semantics.

[9] At the Los Alamos preprint archive (`http://xxx.lanl.gov`) there is an ongoing discussion about the proof, with some new versions by Stapp, e.g., quant-ph/0010047.

of $R2$, but the free choice in **L** were to remain unchanged, then the outcome $L2+$ in **L** would likewise remain unchanged" (Stapp, 1997, p. 301). Counterfactual statements of this kind are true if the counterfactual is based on the weak notion of comparative similarity, but false otherwise: As there is a choice point common to both the actual and the counterfactual scenario, only definition 2.3. applies, being the only one of the three definitions that allows for ties (via $\leq$ instead of $<$ as in the other definitions). Thus, the validity of Stapp's first premise depends on a fine detail of the modal semantics.[10] As none of the three given notions of comparative similarity can be selected on purely formal grounds, a case for or against Stapp's premise LOC1 will have to based on some intuitive, at least non-formal, considerations. No case for or against LOC1 will be made in this paper.

LOC2. In the form given in the published proof, the rule LOC2 is inappropriate, again due to a mixing up of strict and material conditionals. It can however be repaired by changing the second conditional in lines S5 and S6 to a material one. Call the amended lines S5' and S6,' resp. The amended rule that allows one to infer S6' from S5' will be called LOC2.'

LOC2' is incorrect, as can be shown by a simple counter-model. Taking the branching model $\langle W, \leq \rangle$ to be a model of the Hardy experiment (Hardy, 1992) that Stapp himself uses in his argument, we have a case where the antecedent of the purported rule of inference, statement S5,' is true, while the consequent, statement S6,' is false. This one counterexample shows that, contrary to the intuitive motivation given by Stapp in his paper, LOC2' cannot be a valid rule of inference. The purported proof thus does not show that quantum mechanics of itself is non-local.

Even after Stapp's proof, all known valid arguments for the non-locality of quantum mechanics are based on assumptions about hidden variables of some sort. These arguments thus leave open the possibility that the assumptions made about hidden variables are fallacious, in which case quantum mechanics would not have been shown to be non-local.

ACKNOWLEDGMENTS

Thanks to the audience at the Cesenatico IQSA V conference, especially to Jeremy Butterfield, for helpful discussions. This paper is part of a larger project carried out with Tomasz Placek, whom I wish to thank for numerous stimulating comments and suggestions.[11]

[10] Thanks to Jeremy Butterfield for a discussion of this point.

[11] Since this paper was presented at IQSA V in 2001, the project has made some progress; of T. Placek and T. Müller, Counterfactuals and historical possibility, *Synthese*, forthcoming 2005.

REFERENCES

Belnap, N. D. (1992). Branching space-time. *Synthese* **92**, 385–434.

Hardy, L. (1992). Quantum-mechanics, local realistic theories and Lorentz-invariant realistic theories. *Physical Review Letters* **68**, 2981–2984.

Lewis, D. K. (1973). *Counterfactuals*, Harvard University Press, Cambridge, MA.

Lewis, D. K. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* **10**, 217–234.

Müller, T. and Placek, T. (2001). Against a minimalist reading of Bell's theorem: Lessons from Fine. *Synthese* **128**, 343–379.

Placek, T. (2000). *Is Nature Deterministic?*, Jagiellonian University Press, Cracow.

Placek, T. (2000). Stochastic outcomes in branching space-time. An analysis of Bell's theorem. *British Journal for the Philosophy of Science* **51**, 445–475.

Shimony, A. and Stein, H. (2003). On quantum non-locality, special relativity, and counterfactual reasoning. In Abhay Ashtekar, Robert S. Cohen, Don Howard, Juergen Renn, Sahotra Sakar and Abner Shimony, (eds.), *Revisiting the Foundations of Relativistic Physics. Festschrift in Honor of John Stachel*. Kluwer, Dordrecht, pp. 507–529.

Stapp, H. P. (1997). Nonlocal character of quantum theory. *American Journal of Physics* **65**, 300–304.

Xu, M. (1997). Causation in branching time (I): Transitions, events and causes. *Synthese* **112**, 137–192.