Referee's report on #311-2 "Defending the Time-Symmetrized Quantum Theory"

I wish I could recommend this paper for publication. The idea of the time-symmetrized interpretation of quantum theory, indeed of any version which uses two state vectors rather than one, is very intriguing, and it would be welcome to have any novel interpretation to discuss. There are some obscurities in the two-state theory, and an article which would clarify those obscurities would be useful. Unfortunately, this paper does not serve to make the situation clearer. Rather, the author seems to refuse to grant the gist of the objections being made, and never clearly addresses them. The net result is more confusion rather than less.

Let us make the problem clear. At the heart of the theory is a formula. The formula takes as input three items: a state vector determined by a measurement result at an earlier time t1, a state vector determined by the result of a measurement at a later time t2, and a projection operator associated with an intermediate time, t, between t1 and t2. The projection operator is associated with a particular outcome of sort of measurement. From these three items, the theory calculates a number, which it calls a probability. The question which must be addressed is: what is this number supposed to represent? Is it a) the probability that a measurement of the right sort, which actually took place a t, had the given result b) the probability that the measurement, which did not take place a t, would have given that result had it been performed or c) none of the above?

The difference between a and b is clear, but to make it absolutely patent, consider an example. A friend of mine is coming to visit me. He calls at t1 to tell me he is on the way. I know that he will either take the train or the bus. There are two cars on the train, and my friend always chooses at random between them. I hear over the radio at t that there has been a train crash, that only 10 of the passengers were left uninjured, and that 9 of these were sitting in the last car of the train. At t2 my friend appears uninjured at my door. Now we must distinguish two questions.

Question 1) If my friend was on the train, what is the probability that he was sitting in the last car? Answer: .9.
Question 2) My friend turns out not to have taken the train. But what is the probability that had he taken the train, he would have been sitting in the last car? Answer: .5.
Observation: The fact that my friend is unhurt at t2 is critical for calculating the answer to Question 1. It is entirely irrelevant for calculating the result of Question 2. If my friend actually took the bus, his health at t2 has no bearing on the question of whether he would have sat in the last car had he taken the train. So asking for the likelihood of P given Q, and for the likelihood of P if Q had been the case (when it was not) are two quite different things, yielding

different answers which depend on different data.

The ABL rule which lies at the hear of this interpretation is the correct rule for answering questions like question 1. As such, it is uncontroversial: indeed, it is exactly the rule one gets from standard quantum theory together with Bayes' rule. Because of this latter fact, the rule itself gives us no insight into an interpretation, as it is part of
the standard interpretation. On the other hand, the ABL rule seems just incorrect as an answer to questions like 2. Using it there would be just like thinking that my friend's health, given that he did not take the train, somehow makes it more likely (or even certain!) that he would not have been hurt if he had taken the train. And that is just wrong. Standard quantum theory, though, does provide answers to questions like 2, giving probabilities for results of counterfactual measurements. These probabilities depend only on the one usual state function.

The critics which this paper aims to respond to try repeatedly to make this simple distinction, and to point out that the ABL rule does not apply to counterfactual situations. Instead of responding
ing directly
to these objections, the author claims not to understand the meaning of them, in particular claiming that one cannot make sense of ascribing probabilities to measurements which did not occur. The author repeats several makes it more likely (or even certain!) that he would not have been hurt if he had taken the train. And that is just wrong. Standard quantum theory, though, does provide answers to questions like 2, giving probabilities for results of counterfactual measurements. These probabilities depend only on the one usual state function.

The critics which this paper aims to respond to try repeatedly to make this simple distinction, and to point out that the ABL rule does not apply to counterfactual situations. Instead of responding directly to these objections, the author claims not to understand the meaning of them, in particular claiming that one cannot make sense of ascribing probabilities to measurements which did not occur. The author repeats several time the dictum that "unperformed experiments have no results". But the issue is not what results an unperformed experiment had, but the probability for various results had the experiment been performed. Standard quantum theory provides these, as the author is aware. The author's repeated claims to be unable to assign any meaning to these counterfactuals is extremely unuseful here, and conveys the sense that no real communication is going on.

What, then, is the author's own position. It is, unfortunately, hard to say. One could say, of course, that the rule is meant only to apply to situations where an intermediate measurement was made, and one is calculating the probability of various outcome based both on the outcomes of antecedent and subsequent measurements. Then the rule is correct, uncontroversial, and, as yet, metaphysically uninteresting. If this were the position, then the author ought to simply reject talk of counterfactuals about what would have hapntrary to fact"? There are alsment been made. But the author continues to want to talk about counterfactuals. It is very hard to know quite what is meant though.

For example, on the simple question of whether the relevant probabilities are for results of experiments which were actually performed or results of experiment had they been performed, no clear answer is forthcoming. The insistence on talking about counterfactuals suggests the latter, but then one is mysteriously told that "I challenge the interpretation according to which counterfactual statements are necessarily about events which do not happen" (p. 16). This is simply unintelligible: what does the author intend by "counterfactual" if not "subjective conditional whose antecedent is contrary to fact"? There are also the mysterious phrases "except for the measurement at the time t if performed' (p. 12) and "except, may be, a measurement at time t" (p. 18), which suggest that one is somehow unsure whether or nor a measurement occurred at t, and hence unsure whether one is talking about the likelihood that an outcome actually occurred, or that it would have occurred had a measurement been made. One might be so unsure (as I might be unsure if my friend took the train or not), but one then needs subjective probabilities for the occurrence and non-occurrence of the measurement to get any numbers out at all. Given the subjective probabilities, one would get numbers in the usual way.

The author suggests a way of evaluating counterfactuals which holds the results at both t1 and t2 fixed. If this is a true counterfactual, then this simply disagrees with how we do evaluate them, just as my friend's health having taken the bus is not held fixed when considering what would have happened to him had he taken the train. If we made measurements at t1 and t2 and no measurement at t and got certain results, then we can, of course ask: if we had made the same measurements at t1 and t2 and as well a measurement at t, and if the results at t1 and t2 had been the same, then what are the probabilities of various outcomes at t? And standard quantum theory will give an answer using the ABL rule. But this answer is not what we would be looking for if we ask, in normal circumstances, the probabilities for various results had I made the measurement. Similarly, if my friend did not take the train, I can ask for the probabilities that he would have been in the last car had he taken the train and arrived uninjured. But this is not the probability that he would have been in the car had he taken the train full stop. The rather nonstandard counterfactual can be made sense of, and maybe this is what the author intends. But since standard quantum theory, using the standard rules and Bayes's theorem will use the ABL rule here, so invoking the rule does not point us to an alternative interpretation.

As this question about interpreting the ABL rule is the center of the paper, and as the author's position is obscure, I could not recommend the paper. There are also some other annoyances. There are some technical errors: formula 1 on p. 2 is incorrect (it should not be squared) as is formula 5 on p. 10 (the sin should be raised to the fourth power).
The quotation on p.8 draws a distinction between the orthodox and standard interpretations of quantum theory, but it is impossible to tell what the orthodox interpretation is. The author thinks the

quotation is important, but then uses "orthodox or standard" on the page in a way to suggest that they are interchangeable, and both mean what the earlier passage means by "standard".
The phrase "all possible measurements" on p. 9 is ambiguous over just the main issue (is this merely epistemic or ontological possibility).
The calculation on p. 14, meant to show the compatibility of the ABL formula with standard quantum theory, is unintelligible from the point of view of the ABL interpretation. The calculation does show that standard quantum, together with the probability calculus (equation 6) yield the ABL rule for a certain situation. But this only works by using standard quantum theory (not the ABL rule) for prob(1f) and prob (2f). If the project is to systematically replace the standard probability calculation with ABL calculations, this fails.

The author consistently uses the phrase that the measurement at time t2 "find(s) the system at t2 in the state |psi2>" (e.g. p. 3). The terminology "finds the state" is tendentious. In the standard theory, the system is typically not in |psi2> until after the measurement, and so is not "found to be" in that state, but comes to be in that state. This, of course, makes the propagation of |psi2> backwards in time from t2 look physically meaningless. The author may well disagree with the stanrdard interpretation, but the question is not decided simply by adopting one set of terminology. And the terminology obscures some major issues, e.g. are there collapses in the two-state formalism, and if so, when do they occur?

Reply to the comments of the referee

I am encouraged by the referee statement that "it would be welcome to have any novel interpretation (which uses two state vectors rather than one) to discuss."

I admit that I "refuse to grant the gist of the objections [to the ABL rule (which uses two-state vector)] being made", but I cannot agree that I "never clearly address them". Of course "clearly" is a very subjective term, but I hope that my reply will clarify my view to the referee.

The ABL rule gives probabilities for the results of measurements performed at a time t, $t_1 < t < t_2$, given the results of measurements at t_1 and at t_2. The question is: can these probabilities be useful in cases in which the measurement at time t have not occur or might have not occur.

The referee asks a somewhat different question: Is the meaning of the probability given by the ABL rule a) the probability for the results of the measurements which were actually performed, "b) the probability that the measurement, which did not take place at t, would have given that result had it been performed or c) none of the above?"

There is a full consensus about (a), the question is: can we give meaning to the ABL rule beyond this case? The meaning of (b), however, as stated is very obscure. It is not clear under which condition the probability is being considered. According to the basic formulation of the ABL rule the condition must be that the results of the measurements at t_1 and at t_2 are given. But the referee makes it clear in his analysis of a simple example that the condition he considers is the result of the measurement at t_1 only. Naturally, the time-symmetrized ABL formula does not give the probability for the time-asymmetrical case (b).

According to the referee, the distinction between (a) and (b) (and the inapplicability of the ABL rule to the case (b)) is an essential part of the criticism to which my paper tries to respond. The referee complains that I do not respond directly to this issue. This is not so. The case (b) of the referee is "the interpretation (b)" of Section 4 of my paper. I do discuss it in detail there and reach the same conclusion as the referee that the ABL formula is not appropriate in this case. This, however, does not prove anything, since the case is explicitly asymmetric in time. The question is can we have a time-symmetrized theory for problems posed in a time-symmetric way as are all the problems I and my colleagues investigated using the ABL rule?

The referee writes:

> "instead of responding directly to these objections, the author claims not to understand the meaning of them, in particular claiming that one cannot make sense of ascribing probabilities to measurements which did not occur."

It seems, however, that he agrees with me on this last point. Indeed, he continues:

> "But the issue is not what results an unperformed experiment had, but the probability for various results had the experiment been performed."

Thus, he rejects, as I do, interpretation (a) of Section 4 of my paper. The referee understands that interpretation (a) is meaningless and this makes my discussion of (a) unnecessary for him. I also cannot comprehend (a) (I state it clearly in the paper) and the only reason I consider it in detail is because the writings of some of the critics (as quoted in the paper) strongly suggest that this is the interpretation they adopt. In particular, only this interpretation seems to justify the calculations in "the proof of inconsistency" of the critics, the key point of the whole argument.

The referee complains that it is hard to say what is the author's own position. I state it clearly in the paper, that it is interpretation (c) of Section 4. From reading p.3 of the referee report I get the feeling that the referee understood the essential part of it, but he was not ready to accept the meaning of counterfactuals which is different from the usual usage as in his example. He showed that he has the same prejudice of time-asymmetric counterfactuals which lead to the discussed criticism of the ABL rule.

Another prejudice of the referee which I suspect is common to some of the critics is that counterfactuals must necessarily be contrary to the fact. The referee writes:

> 'This is simply unintelligible: what does the author intend by "counterfactual" if not "subjective conditional whose antecedent is contrary to the fact"?'

To answer, let me quote the main "authority" in the field:

> "Counterfactuals with true antecedents–counterfactuals which are not counterfactual–are not automatically false, nor they lack truth value. This stipulation does not seem to me at all artificial..."(Lewis, D. (1973), *Counterfactuals*, p.3).

The most common situation for the application of the ABL rule is that we know the results at t_1 and t_2, but we do not know which measurement out of set of incompatible measurements was performed at time t. In discussing possible outcomes of these measurements there is a "counterfactual" element since it is not possible to perform all of them on the same system. It will be very strange in this case to discuss only the results of measurements which were not performed. Note, however, an interesting application of the ABL rule for "true" counterfactuals, see my footnote 10.

The referee understood that I suggest "a way of evaluating counterfactuals which holds the results at both t_1 and t_2 fixed." He continues not to accept it:

> "If this a true counterfactual, then this simply disagrees with how we do evaluate them..."

And again:

> "But this answer is not what we would be looking for if we ask, in normal circumstances, the probability..."

Finally the referee admits:

> "The rather nonstandard counterfactual can be made sense of, and maybe this is what the author intend."

Yes, this is what I have in mind. The circumstances in which I, Aharonov, and others working in the field apply the ABL rule are not of the kind (b) in the referee's example. We consider possible measurements on pre- and post-selected quantum systems.

The referee, however, continues:

> "But since standard quantum theory, using the standard rules and Bayes's theorem will use the ABL rule here, so invoking the rule does not point us to an alternative interpretation.
>
> As this question of interpreting the ABL rule is the center of the paper, and as the author's position is obscure, I could not recommend the paper."

Here I have difficulty in understanding the argument of the referee. He understood my position. He granted that "The rather nonstandard counterfactual can be made sense of", but then he says that "it does not point us to alternative interpretation." After all, the critics were supposed to show that there is no counterfactual interpretation of the ABL rule whatsoever.

Maybe the following paragraph can explain the thoughts of the referee:

> "The calculation on p.14, meant to show the compatibility of the ABL formula with standard quantum theory, is unintelligible from the point of view of the ABL interpretation. The calculation does show that standard quantum, together with probability calculus (Eq. 6) yield the ABL rule for a certain situation. But this only works by using standard quantum theory (not the ABL rule) for prob(1f) and prob (2f). If the project is systematically replace the standard probability calculation with ABL calculations, this fails."

Yes, now I can see why I do not understand the referee: the reason is that I do not know what he means by "the ABL interpretation", "the ABL rule", and "ABL calculations". I have been working with Aharonov now for 15 years and I have used the ABL rule in more than 20 papers and the only way to make ABL calculations I know is the one I have used in calculations on p. 14. According to the referee this way is "unintelligible from the point of view of the ABL interpretation". I asked Prof. Aharonov for help, but he also said that this "unintelligible" way (the way which is consistent with standard quantum theory) is the one he always used and it is the only one he knows.

I thank the referee for spotting two mis-prints (just for the record: these are not errors: the equations were used correctly later in the paper). I will make these and a few other corrections (if the paper will be accepted) before publication. In particular, I will change the phrasing so that the reader will not get an impression that I want to discuss a distinction between orthodox and standard interpretation. This semantic-historical issue is not relevant for my paper. But, I would prefer not to change my usage of "the system was found in the state $|\psi_i\rangle$". This is a standard way to say that the result of a measurement of some variable A was the nondegenerate eigenvalue a_i corresponding to the state $|\psi_i\rangle$ without specifying irrelevant A and a_i. I also do not think that I should enter in the paper into a discussion of the collapse in the two-state formalism beyond the remark made at the end of the paper that the formalism fits well with the many-worlds interpretation (in which there is no collapse).

Comments on "Time-Symmetrized Counterfactuals in Quantum Theory" #311

B

The author has extensively revised this paper in an effort to address difficulties identified by previous reviewers. Nonetheless I am afraid that I have to recommend against publication, if for no other reason that most of the presentation is extremely obscure. Section 2, which attempts to outline the time-symmetrized formalism will not communicate the ideas to readers who are not already thoroughly familiar with the ideas. Section 3, which attempts a novel account of counterfactuals, leaves the ideas unclear, so that I was not able to see in later sections with any confidence how they were supposed to apply. I simply was not able to follow the argument in section 4 - neither the presentation of the alleged inconsistency arguments, nor the proposed reply. I was not able to understand section 6 at all. In section 7, proposing various conceptions of "elements of reality", I fail to understand what is meant by inferring "with certainty" from data at one time to an alleged element of reality at a *prior* time. The relevant bit of the EPR reality condition is that, if we chose, we can always examine one of the inferred elements of reality. But we can never travel backwards in time to examine a retrodicted "element of reality"; and I see no indirect way of examining such posits. Section 8 appears to be brief exposition of failure of separability effects already discussed by others; and contrary to what the authors suggests, I fails to see how time symmetric analyses helps us understand any of them more clearly. I found section 9, on weak measurements, completely obscure.

Given the over all obscurity of the presentation, I cannot be quite sure whether or not the author has a correct argument defending time symmetric analyses against the objections of Sharp and Shanks and others. However, I am quite doubtful.

Referee C

The revisions in this paper do not adequately correct those failings I noted in my previous report on an earlier version, and so once again I recommend against its publication in *Philosophy of Science*. The authors have not succeeded in restating their response to Sharp's and Shanks's objection more clearly and persuasively in a briefer and more carefully focussed discussion note. Instead, this version of the paper contains significantly expanded treatments of other topics only tangentially related to this central objection. The response to the objection itself in section 4 remains essentially unchanged, and the comments on this response in the concluding section 9 do not significantly clarify the response.

This is particularly unfortunate, since after reading the authors' "Reply to the referee whose report starts with 'I recommend...'" it is at last clear to me that they *do* have a reply to Sharp's and Shanks's objection that effectively addresses that objection (unlike the reply they actually give, which misses the point of the objection)! Since several objections similar to that of Sharp and Shanks have now been published, it seems highly desirable for the present authors to present such a reply in a paper worthy of publication in a journal like *Philosophy of Science*. At the risk of appearing to sketch such a paper for them, I shall try to make these points clear.

Consider the following situation, which we suppose to be actual. An ensemble of spin 1/2 systems with spin up in the a direction is prepared at t_i by performing an ideal measurement of spin in the a direction on an ensemble of spin 1/2 systems, and selecting only those that give the result "up". Later, at t_f, a measurement of spin in the c direction is performed on each member of this ensemble. Each member of the ensemble evolves freely between t_i and t_f (in particular, no measurement of spin in a b direction coplanar with a and c is carried out on any of them). Now consider the following question:

> 'If an ideal measurement of spin in the b direction had been carried out on each member of the ensemble at a time t intermediate between t_i and t_f, what would have been the probability of getting the result "up"?'

Answers to this question take the form of counterfactual statements--subjunctive conditionals whose antecedent is false (since, as we supposed, no measurement of spin in a b direction coplanar with a and c is actually carried out between t_i and t_f). On the usual treatment of counterfactuals, the question is to be answered by considering worlds similar to the actual world up to (just before) t, but subsequently diverging from the actual world because of the performance of an ideal measurement of spin in the b direction on each member of the ensemble at a time t. Now quantum mechanics specifies a probability

$$\text{Prob}_{QM}(\text{up}) = \cos^2(\theta_{ab}/2)$$

which is usually thought of as the objective chance for an "up" result of the b-spin measurement in any such world. Moreover, that objective chance is assumed to be the same whether (as in the actual world) a measurement of spin in the c direction is performed on each member of this ensemble at t_f, or (unlike the actual world) no such measurement is performed. Indeed, if this were not the case, one could influence the objective chance at t by performing a **later** measurement at t_f -- a clear case of backward causation. Standard quantum mechanics also implies that this objective chance is the same in any of these worlds in which a measurement of spin in the c direction is performed on each member of this ensemble at t_f, **whatever** the outcomes of all these spin measurements are in that world.

Sharp's and Shanks's objection is now that time-symmetrized quantum theory specifies a probability **different** from $\mathrm{Prob}_{QM}(\mathrm{up})$ in its answer to our question: i.e. it implies the answer

'If an ideal measurement of spin in the b direction had been carried out on each member (TS) of the ensemble at a time t intermediate between t_i and t_f, the probability of getting the result "up" would have been $\mathrm{Prob}_{TS}(\mathrm{up})$.' [where $\mathrm{Prob}_{TS}(\mathrm{up}) \neq \mathrm{Prob}_{QM}(\mathrm{up})$]

In section 4 of the paper under review the authors argue

(1) that Sharp and Shanks must be considering an **actual** situation in which an ideal measurement of b spin is performed, since otherwise discussion of the probability of the result of the measurement at t is meaningless ("Unperformed measurements have no results"), and

(2) that time-symmetrized quantum theory gives the same probability as standard quantum mechanics for such an actual situation.

(2) is correct, but (1) is incorrect and its justification is spurious. It is because they incorrectly maintain (1) that the authors of the paper under review fail to address Sharp's and Shanks's actual objection, which I restated above. I suggest that proponents of time-symmetrized quantum theory **should** respond to this objection as follows.

Sharp and Shanks are right that time-symmetrized quantum theory can be applied to a merely counterfactual intervening measurement so as to give the answer (TS) to our counterfactual question. Moreover, they are right to claim that $\mathrm{Prob}_{TS}(\mathrm{up}) \neq \mathrm{Prob}_{QM}(\mathrm{up})$. But this does not show that time-symmetrized quantum theory conflicts with ordinary quantum mechanics, because $\mathrm{Prob}_{TS}(\mathrm{up})$ and $\mathrm{Prob}_{QM}(\mathrm{up})$ are the values of two distinct probabilities. From the perspective of ordinary quantum mechanics, while $\mathrm{Prob}_{TS}(\mathrm{up})$ gives the wrong value for the **objective chance** of the "up" result in the b spin measurement, it does give the correct value of the **epistemic** probability of the "up" result in the b spin measurement for the counterfactual situation in which the "up" result of the measurements of spin in the c direction performed on each member of the ensemble at t_f is known to occur with frequency $\cos^2(\theta_{ac}/2)$ -- the quantum mechanically predicted probability for the situation in which c spin is measured at t_f in the absence of any intervening measurement of b spin at t. From the point of view of time-symmetrized quantum theory, this epistemic probability must be treated as basic, and not derived from any underlying objective quantum mechanical chance. After all, objective chance is itself a time-asymmetric concept that has no place in a thoroughly time-symmetrized theory! But time-symmetrized quantum theory can still reproduce the predictions of ordinary quantum theory for what proponents of the latter usually think of as chances: the argument for this (which now becomes a conciliatory point) is essentially that given in section 4 of the paper.

If proponents of time-symmetrized quantum theory take this line (as I think they should), then they do have an answer to Sharp's and Shanks's objection. But then they will face more interesting challenges of a different kind. By reading all probabilities as epistemic, they make it unclear how these arise from an underlying indeterminism. More importantly, on this reading of the time-symmetrized counterfactuals it seems quite misleading to associate these with "elements of reality" in the way the authors proceed to do in later sections of their paper. And their suggested application to the "quantum puzzles" then threatens to make these seem more rather than less puzzling. Such more interesting challenges will have to be faced later. For now, the important thing is for the authors to **briefly** and carefully formulate their response to Sharp and Shanks-type arguments in a way that proponents of those arguments can understand and to which they will feel the need to respond.

APPENDIX WHICH WAS NOT SENT

"…. I also add an appendix in which I present a curious gedanken situation which is a variation on the theme of the example of the referee. I believe that my example serves better for demonstrating counterfactual interpretation of the ABL rule. However, the example is not directly connected to the main issue–the validity of inconsistency proofs–and I prefer not to have this example be the topic of the future discussion about publication of my paper.

Appendix: gedanken story

I have a special friend. He is a quantum invisible man. He can be in a superposition of different macroscopic states and under normal circumstances he does not leave any trace on the environment. He can come to me taking a bus or taking one of the two train cars. He has chosen to make the journey to me in a superposition

$$\frac{1}{\sqrt{3}}(|car1\rangle + |car2\rangle + |bus\rangle).$$

In order to meet him I made a special measurement and found him coming in the state

$$\frac{1}{\sqrt{3}}(|car1\rangle + |car2\rangle - |bus\rangle).$$

At the intermediate time I heard the news that a team with special equipment looked for invisible quantum men in one of the train cars. Now the ABL formalism will help me to make a warning: "My friend, I know for sure that they saw you!" I do not know in which car the team made the measurement, but the ABL formula tells me that if the measurement was made in car 1, they saw my friend in car 1, and if, instead, they looked in car 2, they saw him for sure in car 2.

(A discussion of this example with one particle in three boxes instead of mysterious invisible quantum man in the train and the bus can be found in Aharonov and Vaidman (1991) and Vaidman (1996)."