What Retrocausal Explanations Look Like

Ken Wharton*
Department of Physics and Astronomy, San José State University, San José, CA 95192-0106

While it is generally known that retrocausal models can provide an account of Bell-inequality violations in terms of spacetime-local beables, new models can now explicitly show how this comes about. By analyzing a simple local-beable model that precisely recovers the quantum joint probabilities for measurements on a Bell state, general concerns about retrocausal models can be analyzed at a much deeper level than previously possible. (Including questions of locality, fine-tuning, free-settings, etc.) With this framework it is possible to assess whether various general concerns apply to this specific model, instead of mere straw-man alternatives. In this workshop, I am particularly interested to see whether surviving concerns are better classified as outstanding physics questions or philosophical objections.

I. INTRODUCTION

Bell's theorem has ruled out local past-common-cause explanations of observed Bell-inequality violations, but this has not stopped research into more general causal explanations of such phenomena, in terms of beables localized in spacetime. The options on the table include superluminal causal influences, retrocausal explanations, and a casual restriction on the measurement settings themselves. I have always found retrocausal explanations by far more compelling than the other options (including giving up on causal explanations entirely). But this opinion is an extreme minority among physicists, largely, I suspect, because most physicists do not know what a retrocausal explanation might look like in the first place (or have an incorrect straw-man view of what retrocausality might entail). This short paper aims to rectify this situation, and to address the points at which physics-based objections might be raised against such models.

To this end, I will begin by describing an explicit retrocausal model with well-defined, continuous, spacetime-local beables. This model is completely successful at reproducing the joint probabilities predicted by QM for a Bell state of two spin-1/2 particles in a singlet configuration, but is limited in that there is no obvious extension to cases of non-maximally entangled states. Nevertheless, since this model explicitly violates the Bell-inequalities, it serves as an excellent testbed for an analysis of retrocausal models in general. The model provides the ability to directly answer questions like: Is this model "local"? Is it "fine-tuned" to prevent non-local signaling? Does it violate the ability of the experimenters to freely choose measurement settings?

While I doubt that many will be swayed by this particular model, it may help one to sharpen the main points of concern, beyond a superficial and circular "I don't like retrocausal models because I don't like retrocausality". Indeed, I suspect that most of the sharpest objections will boil down to not physics arguments, but rather a philosophical distaste of concepts that are already present in other well-established physical theories. If so, logical consistency requires one to also reject these problematic concepts in all contexts. And if one is *not* willing to reject the block universe of general relativity, or to replace the perfect CPT-symmetry of quantum field theory with some deep time-asymmetry, the logical consequence may be that one is also forced to seriously consider retrocausal explanations of Bell-inequality violations. At minimum, by aligning itself with these well-established concepts, retrocausal explanations arguably comprise just as viable a research program as anything else on the table.

---

*kenneth.wharton@sjsu.edu

II. THE RETROCAUSAL MODEL

A. Schulman's One-Particle Model

The retrocausal model in question has been outlined in a recent paper [1], with further motivation to be found here [2]. The inspiration was a one-particle model proposed by Schulman [3], and understanding this one-particle (retrocausal) model is central to understanding the two-particle model. (Unlike standard QM, the two-particle model is essentially a trivial extension.)

Schulman's model applies to a single spin-1/2 particle which is subject to two consecutive measurements. The spin-vector $\boldsymbol{S}$ represents the actual state of the system at any given time (equivalent to a unit vector on the Bloch sphere); it points in the direction of the expectation value of the vector spin operator $<\boldsymbol{S}>$ for a pure state. Schulman's ansatz is that between any two measurements, this spin-vector is permitted undergo an anomalous rotation through a net angle α with a global probability proportional to

$$W(\alpha) = \frac{1}{\alpha^2 + \gamma^2}. \tag{1}$$

Here γ is a small free parameter in the model that cannot quite go to zero (without making $W(\alpha)$ ill-defined for anomaly-free histories). It is also important to note that these global probabilities are assigned to entire histories, not particular instants.

All spin measurements on the particle are imposed as boundary conditions. This is already standard for preparations; if one initially measures the spin in some particular direction $\boldsymbol{S_i}$, it is standard to assume that the prepared spin-vector is aligned with $\boldsymbol{S_i}$. But here there is no projection postulate upon measurement; if the second measurement finds the spin to be in some particular direction $\boldsymbol{S_f}$, then this direction is imposed as a *final boundary condition* on the spin-vector, just like the preparation. If $\boldsymbol{S_i} \neq \boldsymbol{S_f}$, and if there is no standard dynamical process which would take the former to the latter, then an anomalous rotation must necessarily occur, taking the spin-vector through an angle α (with a probability distribution proportional to $W(\alpha)$ over all allowable rotations).

The only asymmetry between past and future in Schulman's model comes in as a restriction on the experimenters themselves. If Alice is making the first measurement, and Bob is making the second measurement, both Alice and Bob can choose the angles of their spin-measurement settings to be anything they choose. (Their choice arises from outside the system of interest, effectively as freely-chosen external boundary conditions on the particle.) But only Alice can select the *actual* outcome (of the two possibilities, $\pm\hbar/2$), and send the corresponding particle to Bob. Bob has no control over his actual outcome, and can only read off the sign of the result.

Suppose Alice and Bob make consecutive spin measurements, each in some particular chosen direction, and their two chosen two angles differ by θ. Alice chooses her spin-vector output to be aligned with her setting angle; her outcome is always $+\hbar/2$. But Bob cannot make such a choice; the spin-vector will either be aligned ($+\hbar/2$) or anti-aligned ($-\hbar/2$). These two possibilities are therefore the two possible final boundary conditions on the particle. Therefore, the net anomalous rotation between measurements must be either $\theta(mod\, 2\pi)$ or $\pi+\theta(mod\, 2\pi)$, corresponding to Bob's two possible outcomes. But note that $W(\alpha)$ is not a cyclic function; $W(\theta)$ is different from $W(\theta + 2\pi)$. To generate probabilities, then, one must sum over all possible anomalous rotations that lead to the same result. Schulman finds this ratio of probabilities as

$$\frac{P(\theta)}{P(\pi+\theta)} = \frac{\sum_{n=-\infty}^{\infty} W(2n\pi+\theta)}{\sum_{n=-\infty}^{\infty} W(2n\pi+\pi+\theta)} = \frac{cos^2(\theta/2) + sin^2(\theta/2)tanh^2(\gamma/2)}{sin^2(\theta/2) + cos^2(\theta/2)tanh^2(\gamma/2)}. \tag{2}$$

In the limit that $\gamma \to 0$, this obviously reduces to the standard Born rule probabilities, and since γ is an arbitrary parameter, Schulman's ansatz can be made arbitrarily close to the Born rule.

Note that Schulman's one-particle model is effectively retrocausal. If Bob chooses the same measurement angle as Alice, there will never be any anomalous rotation. But if Bob chooses a different angle, then *before* he makes his choice, an anomalous rotation may occur. Because a different event in the past depends on Bob's future choice, this is a retrocausal model – as generally is any model with a choosable final boundary condition. The "bilking" argument (that Bob may find out the spin-vector and therefore learn about his future choice before he makes it) is naturally resolved by the link between measurements and boundary conditions; the only way to directly learn anything about the spin-vector is to physically make an intermediate measurement, which would change the boundaries of the experiment. Also note that this block-universe style of retrocausal model does not remove Bob's free choice of the setting; that is coming from outside the system of interest, as an effective external boundary.

B. Extension to a Bell State

Schulman's one-particle model does not assign conditional probabilities to instantaneous states, but rather joint probabilities to entire histories: it is the entirety of the anomalous rotation angle α that appears in Eqn (1). This feature allows a natural extension of the one-particle model to a two-particle system, where each particle has a local spin-vector. For two spin-1/2 particles, if the spin-vector of particle $P1$ undergoes an anomalous rotation of angle α, and the spin-vector of particle $P2$ undergoes an anomalous rotation of angle β, it is natural to define the unnormalized joint probability of these two events to be simply $W(\alpha)W(\beta)$.

Remarkably, this usual combination of joint probabilities is all that is needed to recover a retrocausal explanation of Bell-inequality violations, in terms of spacetime-local beables (the two spin-vectors). Specifically, we are interested in reproducing the Born Rule as applied to an entangled Bell state $\psi_{Bell} = (|00> - |11>)\sqrt{2}$, but without using such a non-local state to describe the system.

Instead, we can simply model the preparation of such an entangled system as a local constraint between two distinct and localized spin-vectors: the spin-vectors of $P1$ and $P2$ are initially constrained to point in opposite directions, but that direction is completely unspecified. This is *not* to say that each spin-vector has a uniform probability distribution over all directions; the probabilities are assigned to entire histories, not instantaneous states. Rather, the absolute direction of P1's spin-vector is unconstrained, meaning that all possible directions should be considered when constructing the possibility space on which the weights $W(\alpha)W(\beta)$ can be assigned. But with this caveat aside, this merely describes "classically entangled" particles: learning the initial spin-vector of $P1$ would precisely inform us about the initial spin-vector of $P2$.

After this classically-entangled pair of particles is produced, $P1$ is sent to Alice, and $P2$ is sent to Bob. Both Alice and Bob freely choose spin measurements to perform on the particles they receive. Using the above analysis, this is effectively Schulman's one-particle model applied twice; $P1$ undergoes an anomalous rotation of α to match one of the two allowed outcomes of Alice's setting, and $P2$ undergoes an anomalous rotation of β to match one of the two allowed outcomes of Bob's setting. What links the probabilities of the two outcomes is merely the classical entanglement imposed locally at the preparation.

To recover the Born rule, the free parameter γ must be very small, or else (2) would deviate from known probabilities [3]. For this reason, we can expect $W(0)$ to be overwhelmingly larger than $W(\alpha)$ for any non-zero α that can be distinguished by experimenters. The joint probability $W(\alpha)W(\beta)$, then, will be dominated by cases where either α or β is zero. (Given the initial constraint, they cannot both be zero unless Alice and Bob choose the same spin-axis to measure.) In other words, the small value of γ makes two separate anomalies very unlikely, so the distribution $W(\alpha)W(\beta)$ will naturally force all of the anomaly to occur on either $P1$ *or* $P2$, not a combination of both.

As in the single-particle case, the key parameter is the net angle θ between Alice's and Bob's measured spin directions. In fact, if $\alpha = 0$ and all of the anomalous rotation is in β, then this is essentially the same one-particle problem as before: Alice's setting chooses the axis that $P1$'s spin-vector is aligned with, and the preparation ensures that $P2$ is in the opposite direction. This vector must then anomalously rotate by either θ or $\theta + \pi$ to match Bob's setting, and the probabilities of such a rotation are the same as before. This also goes through if $\beta = 0$, by simply switching Alice and Bob.

The only complication here is that Alice can no longer choose her outcome; she's now in the same position as Bob, at another future boundary. This means that there are now *four* relevant joint probabilities to consider instead of two. Alice's measured spin-vector will either be aligned with her measurement setting (call this outcome $a = 0$) or anti-aligned ($a = 1$). Similarly, Bob's measured spin-vector will either be aligned with his measurement setting (call this outcome $b = 0$) or anti-aligned ($b = 1$). For the two outcome scenarios where $a \neq b$, the required anomalous rotation (on either α or β, but not both) is simply θ. For the two outcome scenarios where $a = b$, the required rotation is $\pi - \theta$. But since the anomaly is overwhelmingly likely to be on a single particle, the calculation already performed in (2) is also the answer here:

$$\frac{P(a \neq b)}{P(a = b)} = \frac{P(\theta)}{P(\pi + \theta)} = \frac{cos^2(\theta/2) + sin^2(\theta/2)tanh^2(\gamma/2)}{sin^2(\theta/2) + cos^2(\theta/2)tanh^2(\gamma/2)}. \tag{3}$$

In the $\gamma \to 0$ limit one can normalize these probabilities and find that the correlation $P(a \neq b) - P(a = b)$ becomes simply $cos^2(\theta)$, as would be expected from traditional QM.

One can drill down further and find the precise probability of (say) $P(a = 0, b = 1)$. The magnitude of the rotation anomaly needed for (a=0,b=1) is the same as for (a=1,b=0); in both cases a θ rotation is needed on one particle or the other. And since the probability weight $W(\theta)$ is only a function of θ, both of these sets of outcomes must have the same probability, or $cos^2(\theta/2)/2$. These probabilities not only violate the Bell inequality, they exactly match the known probabilities for measurements on a Bell state, in the limit that $\gamma \to 0$.

This result does not violate Bell's theorem because it is retrocausal. Specifically, Bell assumed that any hidden variable distribution could not depend on the future measurement settings. To see how this assumption is explictly violated here, notice that (because $\alpha = 0$ or $\beta = 0$) the original spin-vectors at preparation will always be aligned (or anti-aligned) with one of the measurement settings eventually chosen by Alice or Bob. One knows nothing about this "hidden variable" until after Alice and Bob make their setting decisions, and as above, it is immune to the bilking argument. Because the hidden variable of the initial spin-vector alignment is effectively caused by the eventual measurement choice, the premises behind Bell's theorem are explicitly violated; these so-called "retrocausal loopholes" are also available for other no-go theorems.

III. DISCUSSION

With this explicit model in hand, one can directly answer questions and criticisms that are sometimes directed at retrocausal models (or their imaginary straw-man versions).

Is this model local? For any given definition of locality, this model permits an answer, and for many definitions the answer is yes. All of the beables are spacetime-local, and any influence between Alice and Bob is explicable via spacetime-continuous beables. There is clearly no "action at a distance", if one defines locality in those terms.

Of course, if one ignores the underlying beables and merely asks questions about locality defined on the level of instantaneous probability distributions, one might get a different answer. After all, Alice's measurement settings are certainly a contributing causal factor to the probabilities at Bob's apparatus, and Alice and Bob need not be in each other's light-cones. But such definitions might also tend to ascribe non-locality to cases of "classical entanglement" where one does not pay attention to the hidden details (say, which shoe is in which shoebox) until they are revealed. Clearly it would be a double standard to pay attention to the hidden state of the shoes in this case, but not the hidden spin vector in the above model. And once the hidden spin vectors are considered, all of the local influences can be seen as confined to paths inside light-cones.

Another case for nonlocality can be levied against Schulman's ansatz in Eqn. (1), which assigns probabilities to global configurations/histories rather than to spacetime-local events. But any definition of locality that ruled out such probability distributions would be awkward; it would make most of classical statistical mechanics "nonlocal", and would seem to link the epistemic status of an observer (who might very well know only some global property) with the objective locality of an underlying system. Further analysis therefore requires some underlying explanation of how such epistemic distributions might arise. It may be relevant that the only known *derivation* of Eqn. (1) arises from a system comprised of spacetime-local beables (the Lagrangian density $\mathcal{L}$ and associated fields) and a spacetime-local constraint ($\mathcal{L} = 0$). [2]

Is this model fine-tuned to prevent signaling? This interesting question, raised by some general analysis in a recent paper by Wood and Spekkens [4], will be the subject of a detailed forthcoming analysis [5]. But in this case, it turns out the answer is no: varying the obvious free parameter γ deviates the probabilities away from the Born Rule, but does not allow signaling at any level, thanks in part to a key Alice-Bob exchange symmetry. We have found one violent change to the model which does permit signaling, but interestingly it also violates the uncertainty principle (and therefore the second law of thermodynamics [6]). Our tentative conclusion is that the second law also prevents any retro-signaling in generic retrocausal models, and the usual link between causation and signaling only applies to familiar thermodynamic situations. (Indeed, the second law is itself arguably subject to fine-tuning arguments.)

Isn't this just a fancy way to violate the free-settings/free-will assumption? Not in the slightest. There are indeed "superdeterministic" causal explanations of Bell-inequality violations where there is a common-past cause between the hidden variables of every entangled system and the hidden brain state of the agent(s) who will choose the measurement settings for that system. Any beable-level description of such a model would need to include 1) the distant past common cause C, 2) a causal link between C and the hidden variables, 3) a causal link between C and the hidden brain state(s), and 4) the causal link between the hidden brain states and the measurement settings. This beable-level model includes none of those items, and is therefore not a superdeterministic model.

In the model presented here, the choices of the experimenters are imposed as external boundaries on the two-particle system of interest. (Yes, they were final boundary constraints, but external to the spacetime-system of interest all the same.) Any settings are permissible; there is no restriction on what the boundaries might be. If we enlarged the system of interest to include the experimenters, one would need future boundary constraints on that larger system (not to mention a detailed model of the agents themselves).

One common point of confusion here is the claim that the hidden spin state vectors in the past constrain what the measurement settings are allowed to be in the future. Such a claim has several deep flaws. One is that it implicitly *assumes* forward-only causality, taking the view that the past always causes the future, thereby denying the model's very premise of an external future boundary constraint. Once this point is addressed by treating the past and the future on the same block-universe footing, an equivalent argument is that *no* consequential choice can be free, because the result of that choice in the future determines what your choice will be in the past. This is of course fatalist nonsense; just because the future is as real as the past does not mean that there are no causal relationships.

Any continued objections along these lines are therefore revealed to be philosophical objections against the block-universe concept in general, not retrocausal models in particular. And given that the block-universe is a well-established framework for general relativity, at the very least, casting aside retrocausal explanations on such grounds is not a step to be taken lightly. One can also complain that these models do not account for some deep time-asymmetry that many feel must be present in fundamental physics, but this is also a dangerous position to take given the importance of CPT-symmetry in quantum field theory. Indeed, if one accepts time-symmetry, quantum phenomena almost force one into considering retrocausal models. [7]

The largest remaining concern seems to be that models as described above need not be taken seriously because they are not complete; they have not yet been expanded to the point where they can explain the full gamut of results predicted by QM. The key problems with the above model are a missing generalization to cases of partial-entanglement and a missing account of how measurement settings (past or future) act as boundary constraints on the measured system. (Some proposed answers to the latter may be found in [2]).

But from my perspective, this incompleteness is all the *more* reason why retrocausal models deserve more attention; there is a lot of low hanging fruit in this field. After all, for those who take the position that looking for causal explanations of any reproducible correlations is "central to the scientific enterprise" [4], Bell's theorem pushes scientific inquiry into a corner with extremely few options. If one has respect for the spacetime-picture of general relativity (not to mention Lorentz invariance!), I feel that retrocausal explanations essentially stand alone as the best of all options. [8]

What retrocausal explanations amount to, then, is simply the time-symmetric block-universe picture of general relativity, where external boundary constraints on spacetime-subsystems can be chosen by external agents. (As Evans has put it, "retrocausality at no extra cost" [9].) This picture can easily be scaled up to the entire universe, where now the external constraints are merely cosmological boundary conditions (including, but not necessarily limited to, the Big Bang). There are details standing in the way of a full reformulation and explication of quantum theory, but no fundamental showstoppers – with the possible exception of our own instinctual biases.

---

[1] K. Wharton, "Quantum States as Ordinary Information", Information **5**, 190 (2014).
[2] K.B. Wharton, "Lagrangian-Only Quantum Theory", arXiv:1301.7012 (2013).

[3] L.S. Schulman, *Time's Arrows and Quantum Measurement* (Cambridge Univ. Press, Cambridge, 1997); L.S. Schulman, "Experimental Test of the 'Special State' Theory of Quantum Measurement", *Entropy* **2012**, *14*, 665-686.
[4] C.J. Wood and R.W. Spekkens, "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning", arXiv:1208.4119 (2012).
[5] D. Almada, K. Ch'ng, S. Kintner, B. Morrison, and K.B. Wharton, "Are Retrocausal Accounts of Entanglement Fine-Tuned?", manuscript in preparation.
[6] E. Hänggi and S. Wehner, "A violation of the uncertainty principle implies a violation of the second law of thermodynamics", *Nature Comm.*, **4**, 1670 (2013).
[7] H. Price and K. Wharton, "Dispelling the Quantum Spooks – a Clue that Einstein Missed?" arXiv:1307.7744 (2013).
[8] The only other local-beable option being conspiratorial superdeterminism stories that seem to require a working model of consciousness before any progress can be made.
[9] P.W. Evans, "Retrocausality at no extra cost", *Synthese* DOI: 10.1007/s11229-014-0605-0 (2014).