Yes Ghosts, No Unicorns: Quantum Modeling and Causality in Physics and Beyond

Kathryn Schaffer and Gabriela Barreto Lemos

Abstract Entanglement is often considered a signature of "true quantumness." But what counts as "true quantum entanglement?" Historically, physicists have relied on statistical tests—Bell tests—as a quantum-classical decider: entanglement that shows violations of Bell inequalities is taken to show non-classical correlation. But, meanwhile, claims of Bell-inequality violations with classical systems have proliferated, in physics and beyond. The situation is confusing. This chapter takes some steps toward clarity. Drawing from examples in physics, we urge caution in cross-disciplinary modeling comparisons and illustrate the kind of explanatory causal reasoning that underlies Bell tests. We then highlight the recent application of Causal Analysis to Bell tests to emphasize the role of "unicorn-like" fine-tuning. Finally, we discuss recent work in classical optics that shows that Bell inequalities need to be re-derived and interpreted with assumptions appropriate to the measurement scenario. While we do personally believe that quantum physics exhibits a type of spookiness (a quantum-physics-specific "ghost"), the more important point of this chapter is to argue that Bell inequalities are not portable: their bounds need to be re-derived and interpreted appropriately for each case.

Keywords Quantum modeling · Bell test · Causality · Entanglement · Contextuality · fine-tuning

1 Introduction

This chapter is part of an ongoing conversation between its authors exploring two beliefs we share: (a) that laboratory experiments in quantum physics reveal at

K. Schaffer (✉)
School of the Art Institute of Chicago, Chicago, IL, USA
e-mail: kschaf2@artic.edu

G. B. Lemos
Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

T. Veloz et al. (eds.), *Trends and Challenges in Cognitive Modeling*, STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health,
https://doi.org/10.1007/978-3-031-41862-4_9

least one distinct *spookiness* that remains unexplained and (b) that this distinct spookiness, because of how it uniquely arises in quantum physics, is not present in other disciplinary contexts that use similar mathematical frameworks for quantitative modeling. To defend claim (a), we will discuss a class of experiments in quantum physics referred to as "Bell tests," the interpretation of which remains hotly contentious within physics. These experiments involve quantum entanglement and famously exemplify what Einstein called "spooky action at a distance." The "yes ghosts" in the title of this chapter reflects, metaphorically, our commitment that Bell tests in quantum physics do reveal something distinctly spooky. Work recasting Bell tests in the language of Causal Analysis (Wood and Spekkens 2015; Cavalcanti 2018; Pearl and Cavalcanti 2021), as well as recent work analyzing Bell tests with classical light in the broader framework of Contextuality tests (Markiewicz et al. 2019), shows that the spookiness is not merely about the claimed "action at a distance" in entanglement scenarios. It is also more fundamental to the structure of causality in quantum physics. If fine-tuned, conspiratorial, "unicorn-like" causal mechanisms are forbidden, then Bell tests in quantum physics pose a puzzling contradiction.[1]

For this chapter, which is addressed to an interdisciplinary audience, our primary goal is to argue for claim (b). A consensus has not yet emerged about what is going on in Bell test experiments in physics, and not all physicists agree with us that there is something spooky involved. Nevertheless, the point of contention does not arise in other domains that employ similar mathematical modeling. We will argue this first by discussing features of mathematical modeling practices, emphasizing that the relevant causal mechanisms in a modeling situation depend on the specific measurements used to produce the data, and second by discussing classical-optics scenarios from physics.

Quantum-inspired modeling practices now appear in fields as diverse as cognition, economics, and language modeling (see Pothos and Busemeyer 2022; Lee 2020; Surov et al. 2021 and the references therein for some examples). Entanglement models are often central in such efforts. As a contribution to this growing discourse, we explain some reasons why importing the mathematics of quantum physics—the mathematics of entanglement and Bell inequalities, in particular—does not necessarily mean importing its interpretational problems.

2 Disciplinary and Interdisciplinary Considerations

The interdisciplinary context of this chapter invites a few prefatory comments. We underscore that we are focused on mathematical modeling practices and their causal assumptions, not on stances toward philosophy or worldview. It is part of the

[1] In the workshop preceding this volume, one of us (KS) used the phrase "no unicorns" to say that quantum physics does not provide magical ways around ordinary physics. For this chapter we use the unicorn metaphor for something more technical. The connecting thread between both uses is respect for the explanatory success of experimental physics.

culture around quantum physics to call it strange, but such judgments are a matter of opinion. It is a subjective question whether the wave-particle duality, observer-dependence, and uncertainty in quantum physics seem more or less intuitive than the atomism, determinism, and mechanism associated with classical physics. The word “spooky” may be subjective in the same way, but we are using it to refer to something specific and technical, not a ghost-filled worldview.

As “quantum modeling” gains popularity in non-physics fields, the philosophical implications of modeling choices sometimes take on an importance that is unfamiliar to physicists (Schaffer and Barreto Lemos 2021). Thus, for example, we have heard quantum modeling characterized as a revolutionary approach meant to unseat classical paradigms. But a physicist’s library will have textbooks that separately discuss classical mechanics, thermodynamics, quantum physics, classical electromagnetism, relativity, and so on. Data analysis for an experiment can sample from across the bookshelf without needing to commit to just one, or needing all philosophical conflicts to be resolved. In research physics, we are not making an either-or selection of a classical or quantum approach when we model data. In pursuit of explanations that make sense, classical modeling provides appropriate and self-consistent explanations for some measurements (or parts thereof), and quantum modeling does so for others.

Within quantum physics (the subfield of physics that focuses on testing core predictions of quantum theory), there are a few specific experimental results, e.g., Bell tests with quantum entangled particles (Giustina et al. 2015; Hensen et al. 2015; Shalm et al. 2015), that exhibit challenges to the expectation of self-consistency and making-sense. The subject of this chapter is those cases that define the discipline-specific, as-yet-unresolved quantum spookiness. This is an important point to make because Bell-like scenarios have also been explored in classical physics (Borges et al. 2010; Qian et al. 2015; Goldin et al. 2010; Frustaglia et al. 2016; Li et al. 2018) contexts as well as non-physics contexts. The mystique of quantum strangeness, and the desire for proof of “true quantumness,” can cloud discernment of important context-dependent differences in all of these cases.

We argue that Bell tests in quantum physics are distinct. Modeling is not a worldview choice, but it is not just the choice of a set of equations, either. To be explanatory, a modeling practice must embed measurement-specific assumptions about allowable causal processes. Context-dependent and disciplinary differences in causal assumptions are central to sense-making. Bell tests are one example of this truth, but in general, causal mechanisms are important differences in cross-disciplinary modeling comparisons.

3 What Counts as Quantum Modeling?

The phrases “quantum model” and “quantum modeling” circulate in non-physics contexts, but without a universal consensus on their scope of meaning. In this section we share some perspectives from physics.

We work with an operational definition of mathematical modeling as a research practice that associates mathematical structures with analogous relationships among measurable quantities. This definition emphasizes both that mathematical modeling involves a form of metaphor (tracing structural analogies between mathematics and a real-world phenomenon) and that we should think of modeling as a verb, an activity embedded in a disciplinary research context that produces peer-reviewed, published results. The criteria for success vary. In some modeling contexts, identifying a structural similarity between an equation and a set of measurements may be enough to constitute success. In others, the modeling effort is not considered successful unless it yields predictions for novel measurements, answers "why" questions, meets goodness-of-fit standards, or otherwise fits into an explanatory sense-making framework. In other words, the fact that a mathematical metaphor exists does not necessarily determine what it means, nor whether it is any good.

The word "model" is a can of worms. In practice, it often refers to some equation(s) used in a modeling effort. But even to draw structural correspondences—to make a mathematical metaphor—we need context-dependent specifications of how symbols on a page relate to measurements in the world. The word "quantum" is also a can of worms. Equations have no allegiance to any discipline. A certain formalism may be historically associated with quantum physics, but the label "quantum" is a convention that carries no intrinsic meaning nor well-defined scope of application. Is it only a quantum model if the equations are applied in physics? Does only a single version of the formalism count, or does the term apply to a class of related probabilistic models? Is every part of the formalism equally "quantum," even what is shared with models in classical physics (e.g., sinusoidal waves)?

There is a body of knowledge associated with the century-plus success of quantum physics. This knowledge, though it was generated through active modeling, now has a relatively static core: equations in textbooks that have not changed for many decades, a well-defined set of corresponding experiments that are now primarily demonstrative, not actively scrutinized. Most applications of this static body of knowledge do not function as critical tests. That is to say, most of the ways physicists use quantum physics knowledge are not aimed toward falsifying it or expanding it. The research discipline known as quantum physics *is* principally aimed toward those sorts of tests. Can experiments interrogate the textbook formalism in new contexts? What can we reveal through mathematical reformulations? How do we test questions in quantum interpretation? With questions like these, the discipline of quantum physics, over time, expands beyond what is already in the textbooks. In the ambiguity of unresolved open questions, boundary-defining vocabulary questions ("what counts as quantum") may be permanently premature. There is a well-defined canon of established facts from the past, but we do not know what will enter that canon in the future.

There are also lessons to draw from quantum-related modeling practices in physics, but beyond the subdiscipline of quantum physics. Consider a SQUID (Superconducting QUantum Interference Device). A SQUID is a macroscopic-scale device that leverages low-temperature material properties to realize quantum tunneling and interference for magnetic flux detection. To explain *why* a SQUID enables single-quantum flux sensitivity, textbook quantum formalism is involved. However researchers who use SQUID circuits (such as astrophysicists using them in a detector system) can treat them as "black box" circuit components. Even though a SQUID is based on quantum tunneling and interference, papers that model SQUID-based detector systems do not need to discuss quantum physics. No obvious "quantum modeling" is necessary (see, for example, Montgomery et al. 2020).[2]

A contrasting example is research applying quantum formalism to phenomena in classical optics, e.g., (Stoler 1981; Klyshko 1988; Spreeuw 1998; Simon and Agarwal 2000). In this case, physicists use mathematics sourced from quantum textbooks, but the experimental systems are causally governed by classical electrodynamics. This has spurred a vigorous debate in physics about vocabulary (Karimi and Boyd 2015). If a classical optics scenario can be described through the same mathematics as quantum entanglement, is it appropriate to label such a phenomenon "classical entanglement"? Or is that a contradiction in terms, because entanglement means something special to the quantum physics context? Many physicists would say that "classical entanglement" has nothing to do with quantum physics, but SQUIDs do. Such a judgment is not about the how, but about the why. Modeling in physics requires more than a structural metaphor because it engages causal explanations.

These examples also show that size scale does not determine whether quantum physics might be relevant. SQUIDs and the circuits that use them are macroscopic, and so are classical optics systems. As such, these examples can help to inform encounters with "quantum modeling" in non-physics disciplines that are also dealing with macroscopic physical systems. For example, consider quantum modeling of phenomena associated with brains or cognition. How might quantum models apply? In multiple ways, that need not relate to one another. On the scale of neurons or smaller, plausibly some brain components could play a role analogous to SQUIDs or other quantum-based circuit components, as macroscopic physical systems whose "how" is linked to a quantum-physics "why." Once their behavior is understood, modeling those components in context is likely to be similar to the SQUID case: the intrinsically quantum processes can be treated as a black box within a whole-system model.

It is also plausible that some structures in quantum formalism could make good metaphors for whole-brain phenomena, macroscopic human behavior, and some data sets involving language and cognition, e.g., (Pothos and Busemeyer 2022). In such cases, quantum formalism can apply to the "how," without any connection to a quantum-physics "why," as with classical entanglement in optics.

Both possibilities could be simultaneously true (quantum processes mattering in neurons as well as quantum formalism applying to whole-brain phenomena) with

[2] There are more mundane examples of "essentially quantum devices," such as transistors. Arguably quantum processes—e.g., those that enable chemistry—are ubiquitous. Not all quantum effects are equally possible in everyday conditions, though. We discuss SQUIDs as an example of a quantum device because they require special conditions. This is also likely to be true for, e.g., entanglement-based devices.

absolutely no link nor meaning across modeling contexts. This relates to a general point about modeling and mereology (the study of part-whole relationships): the kinds of structures we can model mathematically do not generically translate across scale, from part to whole or vice versa. Neither do the causal reasons for them.

To summarize, both “quantum” and “model” are slippery terms. Modeling, as a practice, involves mathematical metaphors for empirical relationships. Such metaphors may be portable from one context to another, but they do not translate trivially across mereological scale. Nor does it mean anything if a similar model works in two different contexts or across scales. Finally, modeling practices vary in their aims. In physics, modeling normally seeks to answer “why” questions beyond the “how” of a phenomenon. This does not mean resolving all of the philosophy. What it means is that modeling practices in physics aim to explain phenomena in terms of situation-specific causal mechanisms.

4 Sense-Making Is More than Metaphor

More matters in mathematical modeling than the structure of the equation(s) used to fit the data.

To explore this, consider a linear model. The conventional formula for a straight line is $y = mx + b$, where parameter m sets the slope of the line, and b sets the value of y when $x = 0$. Such a metaphor has many applications. Variables x and y can represent displacement in physical space, such that the line describes a path. With y relating to space and x to time, the model can describe linear motion. The model also works in contexts that make no explicit reference to space or time, e.g., x could represent the number of identical items in a shopping cart and y their total cost. We could even get creative and characterize mood in proportion to sunny weather.

While these scenarios share a structural similarity, there are context-dependent dissimilarities too. Variables can be continuous or discrete, bounded or unbounded, or have other constraints. In the cost-items case, no values for x or y should be negative; in the path-through-space case, they could be. What is important about the line, as a model, also differs in each case. Explanatory modeling involves evaluating alternatives; the line is thus conceptually embedded in a mathematical space of options that is case-specific. A cost-items scenario might allow sharp discontinuities (e.g., bulk discounts). Such discontinuities would be impossible in linear motion, but a friction term might appear. Expectations of monotonicity, continuity, and single-valuedness are situation-specific.

In research practices that use mathematical modeling for explanatory reasoning, most of the specificity relates to the data. Measurements require apparatuses. Data sets can have mistakes. Stochasticity matters. We cannot make sense of data without detailed knowledge of how all of this works. Such knowledge establishes assumptions about possible, impossible, plausible, and implausible causal mechanisms for features that might appear in a data set.

It is never part of a student lab report in a physics class to allow that unicorns may have secretly manipulated the outcome of a measurement for their pleasure. Explore this by imagining a kinematics lab in university physics. Each station in the room is equipped with an air track that allows approximately frictionless one-dimensional motion of an accompanying "car." When the student flips a switch, a spring releases the car, allowing it to drift across the track at approximately constant velocity. The same switch restarts a clock. Each track has a set of movable sensors that trigger when the car passes, recording elapsed time. The location of each sensor is measured by eye, using a ruler. A single experimental iteration, or run, involves a student placing the sensors, releasing the car, and collecting elapsed time and corresponding distance measurements for the car's motion.

A professor tells students that they can work individually or collaborate to take data from ten runs of the experiment, varying sensor positions. They are to tabulate and graph the data and then perform a linear fit to estimate the average velocity of the car and its standard deviation. The professor leaves them under the supervision of a Teacher Assistant (TA), returning to grade the papers later.

What criteria will she use to grade the papers? Well, the first thing she checks is whether the students performed the linear fit properly. Indeed, the TA must have helped: all students have correct mathematics applying a linear fit and estimating average velocities. Do they all get good grades? No. Clearly, more was going on, since many of the graphs look quite different from one another. Some of the fits are terrible. Does she award grades based on the apparent goodness-of-fit? Also no. This is experimental physics. To evaluate the lab reports she has to look at the data and use knowledge of the ways it might have been caused.

First she considers some of the papers with visibly poor fits. A few students apparently had equipment problems or made mistakes, which she guesses by noticing some unphysical data patterns. With one, the graph suggests a significant friction effect. Given the context, this is plausible. She marks off points, highlighting the issues students should have noticed and attempted to explain.

In one lab report, the data has such a large scatter that position and time values look barely correlated. How could this happen, given the apparatus? It suggests a serious data-taking problem. But then, the professor notices that the same data is shared by five students. While the reported data is inconsistent with the expected behavior of one air track, it is perfectly consistent with each student in the group performing two runs on a separate copy of the apparatus. The air track setups vary enough that each produces a different velocity and time-offset (m and b in the formula for a line, respectively). The combined data table does not test the behavior of one air track; it tests the behavior of a collection.

Papers that show good fits also deserve attention. Just because the graph looks reasonable does not mean it is without error. She checks for things like unphysical data values. In one paper, the fit is *too* good. The times from trial to trial are all the same, and the deviations in positions are small. This student gets a zero; the results were clearly fabricated.

The professor then comes across a paper with a unicorn drawn next to the graph. She checks the name on the paper. It was submitted by the TA, her graduate student, probably as a joke. So, what is the catch? The fit looks reasonable. She looks closely at the data table. There is an oddity that is so subtle that she almost misses it. Entries with an even time value (measured in milliseconds) tend to have position measurements that deviate high. Entries with odd time values tend to have position measurements that deviate low.

As a provocation, this is successful. It brings up subtle questions about statistics, causality, and "naturalness." A way to generate the pattern would be to record times in a run and then go back and read positions with an extra rule: add a small deviation if the time is even, and subtract the same amount if it is odd. There is more to speculate about. In what other ways could such a correlation be achieved without manipulating the data? Could it be achieved with a modified (vibrating?) apparatus? Could the correlation occur as a statistical fluke? What principles guide the interpretation here, and are they the same principles used to detect the paper that was clearly faked? Could the other student papers that "make sense" not also have similar hidden effects? But if we allowed for that, could we even hold physics class?

Experts across physics, statistics, and philosophy could debate these questions at length, but experimentalists need to cut the debate short to get anything done. It is obvious that the correlation in the unicorn paper is suspicious; evenness is arbitrary in a measurement of time. Likewise, we expect a certain arbitrariness with respect to exact choices of sensor placement. Shifting a given sensor a little (modifying its associated time and distance measurements slightly) should not matter to the substance of what we observe. If the extra correlation in the unicorn paper were physically real, it would thus be inconsistent with known causal mechanisms for linear motion, assessed with clocks and rulers.

Causality is subtle, and causal reasoning in science is not straightforward to formalize. In statistics, there is an approach called Causal Analysis that describes how the relationship between causal factors in a scenario relates to statistical correlations in the observed data (Pearl 2009). In experimental research, usually such knowledge is implicit. We assume that persistent correlations between random variables have two possible reasons (Reichenbach's Principle): one variable causes the other and/or both variables share common causes. The existence of a correlation under-determines the possible reasons why, but it implies that reasons should exist. Meanwhile, the absence of a correlation, if that too persists, shows independence.

This reasoning is part of paper-grading. An extra friction term is a plausible cause for certain extra correlations in linear motion. Variability among a set of five devices is one explanation for failing to see some expected correlations. Even if these judgments reflect general statistical principles, they are also hyper-specific in practice. The professor needs to know not just about how linear motion works, but about the devices in the room, about how clocks work, and how students work. A dishonest student is a plausible causal mechanism for a data set with unexpectedly low scatter; a grad student joke is a plausible explanation for the correlation in the unicorn paper.

In the language of Causal Analysis, fine-tuning refers to a case where the presence or the absence of correlations in a data set depends on specific values of parameters. It is a unicorn-like specialness where we expect the universe to be indifferent. The faked correlation in the unicorn paper is a good joke to play on your graduate advisor, because if it were real, it would be special in that way. Fine-tuning is not always associated with deliberate fakery, though. The results of the collaborative student group show how fine-tuning can happen when the causal model for the phenomenon is not faithful to the measurement process. The lack of an observed correlation in that data set is suspicious-looking to the professor. Given the assumed causal mechanisms associated with a single air track, the data look impossible—unicorn-esque. But in both of these examples, using situation-specific knowledge to identify extra causal factors (intentional manipulation in one case and the extra apparatus variability in the other) resolves any actual mystery.

The overall point of this section is that models (equations) are "just metaphors." By design, they are economical in their expression of structure and thus highly portable from one domain to another. Explanatory sense-making with real data, on the other hand, involves causal mechanisms. The details are anything but portable. The grading judgments (explanations) in our example showcase this. They might not even apply to a linear motion lab exercise done differently down the hall. They certainly would not apply to linear modeling in economics.

Experimental sense-making is more than applying a mathematical metaphor. It is hyper-specific. If we reject fine-tuned explanations (no unicorns), then persistent unexplained correlations are, perhaps, spooky.

5 Essential Quantumness? Entanglement and Bell Inequality Violations

Entanglement is frequently described as a (even the) quintessentially quantum phenomenon. In popular press, like many news reports surrounding the 2022 Nobel Prize in Physics, entanglement is associated with Einstein's famous phrase "spooky action at a distance" or with the claim that physics has officially rejected "local realism" once and for all.

Given the discussion in Sect. 3, skepticism is generally warranted in any conversation that attempts to identify essential quantumness. This is true with entanglement. Some people (ourselves included) believe there is something spooky in some entanglement experiments. But even in quantum physics textbooks, many examples of entanglement do not exhibit the spooky effect. Thus quantum entanglement alone is not enough to challenge philosophical ideas like "realism." The relationship to ideas about "locality" and "non-locality" is also complicated; some of the systems that exhibit the spooky kind of quantum entanglement have spatially separated parts, but some do not. Moreover, even among experts, "locality" and "non-locality" have a range of meanings (Cavalcanti and Wiseman 2012; Harrigan and Spekkens 2010; Wiseman 2006; Brown and Timpson 2014).

Meanwhile, the formalism that defines quantum entanglement is just as portable as the formula for a line. It codifies a type of non-separability that may be a perfectly reasonable mathematical metaphor for systems in many modeling contexts, not just quantum physics labs. It is certainly used elsewhere in physics, with classical optics systems as a notorious example (Collins and Popescu 2002; Aiello et al. 2015; McLaren et al. 2015). The linear modeling examples from earlier in this chapter are a prompt to approach cross-comparisons between these cases with caution: some modeled "how" structures may be similar, while important data-specific details and "why" explanations may differ.

Thus, the ability to model a phenomenon as entanglement is an insufficient marker that the phenomenon is "essentially quantum." But why seek such a marker in the first place? At present, there does seem to be a practical reason: there appear to be computational advantages associated with quantum algorithms.[3] Technology craves good quantum-classical deciders. Bell tests, as tests for evidence of the "spooky" effects in entangled systems, have historically been treated as useful in this way. Just as entanglement is more nuanced than the popular press might suggest, Bell tests do not function as a one-size-fits-all test for quantumness. There is more nuance because Bell tests are about the "why."

A Bell test in quantum physics is a sense-making test assessing correlations observed in data. The assumed measurement scenario for a Bell test is generic. The test assesses correlations observed between the outcomes of two or more detectors, each with two or more possible settings that determine exactly what is measured. Given situation-specific assumptions about plausible causal mechanisms and measurement outcomes, it is typically possible to derive both lower and upper bounds on possible correlations in Bell tests (Popescu and Rohrlich 1998). Such a derivation results in a "Bell inequality," characterizing those bounds. Data observed to violate the bounds can be interpreted as a challenge to the situation-specific causal assumptions.

Bell inequalities are not portable. It is impossible to derive and interpret such an inequality without articulating experiment-specific assumptions. The fictional linear motion lab from earlier in this chapter helps explain why. If an even-odd time correlation persisted in real linear motion data, we might devise a degree-of-spookiness statistical test to characterize what we saw as especially strange in that situation. This would be like a Bell test. The same test would not apply generically to another linear modeling context, e.g., for a cost-items model. A product could easily be cheaper in packs of two, violating the assumptions implicit in the linear-motion case.

The point is that correlations only register as spooky if there are situation-specific reasons to think the universe should not work that way. There have been a number of experiments in quantum physics, known as "loophole free" Bell tests (Giustina et al. 2015; Hensen et al. 2015; Shalm et al. 2015), that have repeatedly exhibited

[3] A review of quantum information theory is beyond the scope of this chapter, but some instructive related examples come from quantum game theory, e.g., the Mermin-Peres Magic Square (MPMS) game. Recent experimental results (Xu et al. 2022) have confirmed that the use of quantum entanglement in the MPMS game enables a winning strategy that exceeds what is possible with classical resources.

correlations that exceed what can be explained by classical causal mechanisms. The spookiness is assessed by applying situation-specific bounds on the expected degree of correlation based on classical statistical reasoning. Observed data violates these bounds.

It is subtle. Quantum physics is not supposed to be classical; quantum formalism explicitly predicts correlations that conflict with classical expectations. So is it really strange that quantum correlations match those predictions? Maybe not. Or maybe the spooky thing in Bell tests is not particular to Bell tests, or even entanglement, but is more about indeterminacy, or randomness, or (pick your favorite quantum oddity). But maybe what is going on here poses a deep challenge to our understanding of the nature of causality, indicating that there is something about quantum causality that remains to be understood (Pienaar 2017, 2020; Cavalcanti and Lal 2014), potentially a true essential quantumness, which relates to "why" not just "how." Do we have consensus yet about any of this in physics? Not yet.

Ultimately, technological development may force the issue. If you want to capitalize on quantum effects for computation, it would be cheaper to reproduce those same effects with classical systems when possible. Does physics serve The Man, or fundamental knowledge? Either way, physicists are working hard to find classical ways to simulate—and thus, possibly, explain (erase)—what currently seems spooky in quantum Bell tests.

6 Bell's Theorem: Two Ways

The most famous and canonical Bell test, articulated by John Bell himself (Bell 1964), involved mathematically derived correlation bounds—the Bell Inequalities—for the probabilities of two or more parties' measurement outcomes, conditional on measurement settings that are chosen independently for each party. A Bell inequality does not stand alone; it gains its physical interpretation through *Bell's Theorem*. The original version of that theorem states that statistics that arise from any "locally causal hidden variable model" for the scenario must satisfy the Bell inequality. Hidden variables are extra variables affecting the measurements but not included in the nominal model. In the fictional example earlier in this chapter, the data set made by the collaborative student group was subject to hidden variables: variations in air-track behaviors from one apparatus to another.

In the experiments addressed by Bell's Theorem, the parties that perform measurements are distantly separated. A locally causal hidden variable model is one that obeys the physics-based constraint that faster-than-light-speed causal influences between their locations are ruled out. When an experiment designed to realize the assumptions in Bell's Theorem shows measurement statistics that violate the appropriate Bell inequality, such a result cannot be explained by locally causal hidden variable models. This is the conclusion sometimes interpreted as conflicting with "local realism."[4]

However, Bell's Theorem has been reformulated in multiple ways, and not all of them have the same philosophical connotations. Bell inequalities can be understood, for example, as a special case of a broader class of contextuality inequalities (Bell 1966; Kochen and Specker 1967; Kernaghan and Peres 1995; Cabello and García-Alcaine 1996). A relatively recent reformulation has also been performed using Causal Analysis, which is a method of inferring possible causal structures from the nature of the correlations that arise in an experiment. Studying the Bell scenario through this framework recasts the interpretive stakes in a useful way.

A Causal version of Bell's Theorem was proposed by Wood and Spekkens in a seminal paper (Wood and Spekkens 2015). It enables a reformulation of Bell's Theorem based on three assumptions about the experimental setup: (i) There are no direct causes between measurement outcomes on one party's side and measurement outcomes on another party's side. So, according to Reichenbach's principle, any correlations that arise between measurement outcomes obtained by different parties *must* be due to a common cause. (ii) No-signaling: A choice of measurement setting on one party's side does not affect the outcome of another party's experiment and vice versa. (iii) Measurement setting independence: The choices of measurement settings by each party are made independently of each other.

When an experiment designed to meet these conditions violates a Bell inequality, it is interpreted to contradict one or more of the assumptions used to build the theorem. Those assumptions include the underlying assumptions of Causal Analysis, e.g., Reichenbach's principle, and the assumption of no fine-tuning. No fine-tuning, as we explained in Sect. 4, is the assumption that the observed statistics are typical for the given causal structure and do not depend on special values of any parameters. It is the "no unicorns" rule. In Bell-like tests, it means that any observed statistical independence (including the no-signaling condition) happens because the causal structure of the phenomenon implies such an independence, not because it was engineered, or the result of some special "accident."

Therefore, if the experiment ensures no-signaling and no direct causes between the outcomes on either side, we can conclude that there is no classical causal model that explains Bell-inequality-violating correlations without fine-tuning. This is why, from a causal modeling perspective, the (loophole-free) quantum violations of Bell inequalities register as distinctly spooky. Given causal models and hard-to-give-up causal assumptions, the observed correlations lead to a logical contradiction. Something is happening that does not fit into the framework of a classical causal model unless we allow unicorn-like interventions. And because no-fine-tuning is

[4] Local realism is generally defined as two assumptions taken *together*: (i) Realism: "If, without disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. The element of reality represents the predetermined value for the physical quantity." (Einstein et al. 1935) (ii) Locality: physical influences between spatially separated systems cannot propagate faster than the speed of light.

such an important concept in sense-making in other experimental contexts, that is not an easy thing to accept.

7 Classical Bell-Like Tests

Insofar as a Bell-inequality-violating entangled system may be a useful resource for quantum communication, it may in the future be treated as a special kind of black box, like the SQUIDs discussed earlier. And like with a SQUID, the physical conditions required to operate such a quantum black box would likely be expensive (e.g., requiring low temperatures and a high degree of isolation from the environment). Thus, if we view Bell-like correlations as a resource to be exploited, there is a motivation to understand whether cheaper, classically based black boxes could allow some or all of the same communication and cryptography tricks. This is one reason that optics researchers study which quantum computation effects can be simulated with classical light and which cannot (van Enk and Fuchs 2002). In particular, experiments ask: can classical light that exhibits non-separability ("classical entanglement") produce results that violate Bell inequalities (Borges et al. 2010; Qian et al. 2015; Li et al. 2018)?

The causal Bell's Theorem is helpful for carefully framing the analysis of such experiments. It replaces the notion of space-like separation with the weaker notion of no-signaling. The condition of no-signaling is naturally satisfied by any pair of non-interacting degrees of freedom, such as the polarization and spatial modes of (paraxial) light, even if they exist in the same beam and hence are not space-like separated. Thus, the Causal Bell's Theorem might be applicable to classical light, even if the original Bell's Theorem does not apply. To derive the inequalities, it is necessary that certain *non-disturbance* relations hold among the relevant variables, of which *no-signaling* is a special case. The idea to use an interferometric setup to violate a Bell-like inequality with classical light was originally proposed by Suppes et al. (1996). Shortly thereafter, Spreeuw (1998) introduced the concept of classical non-separability between two or more degrees of freedom of the same light beam. Numerous authors subsequently analyzed this concept (Aiello et al. 2015; Pereira et al. 2014; Qian and Eberly 2011; Van Enk 2003), including its application to experimental violations of Bell-like inequalities (Borges et al. 2010; Goldin et al. 2010; Frustaglia et al. 2016; Li et al. 2018; Qian et al. 2015).

Apparent violations of Bell inequalities with classical optics systems have been repeatedly demonstrated. Does this show that there is nothing especially spooky about quantum Bell tests? Or, the reverse that the same spookiness can be created in classical systems? We would argue, given the evidence to date, no. It does not show either of these things. The traditional Bell inequality (and its specific bounds on classically induced correlations) is not easily portable. When Bell inequalities are re-derived with situation-specific measurement assumptions, the inequality bounds may change, which changes the interpretation of measured correlations. A recent work by Markiewicz et al. (2019) makes this argument for the classical light case.

The authors point out that the type of measurement involved in classical light Bell tests differs significantly from the type of measurement in quantum Bell tests. The appropriate Bell's inequality for the classical light context would need to be an inequality between measured *field intensities* and not probabilities of "clicks" of detectors. A re-derived version of the Bell inequality for this case shows different bounds. Thus the observed correlations in the classical case are not "spooky." They are consistent with classical causal expectations and apparently have little direct bearing on interpreting the quantum Bell test case.

In other words, the exact nature of the measurement matters to the derivation of the Bell test logic. Sense-making is specific. There may be some resemblance between classical non-separability experiments and quantum entanglement experiments, but a careful analysis of the causal assumptions embedded in the measurement scenario differs. Characterizing what would be "spooky" in each case depends on situation-specific expectations about causality.

8 Application to Interdisciplinary Contexts

Paper titles across multiple disciplines advertise Bell-inequality violations in novel systems as key results that demonstrate "non-classical behavior" or purportedly demonstrate entanglement. Examples include analysis of concept combinations and word associations, e.g., (Beltran and Geriente 2019; Aerts et al. 2021). There is also a long history of toy-model examples, e.g., involving rubber bands (Sassoli de Bianchi 2013) or socks (Aerts and de Bianchi 2019). What do Bell-inequality violations in these cases mean?

Well first, it helps to remember the general cautions on modeling and metaphor from the beginning of this chapter, especially the how vs. why distinction. If the work references a "Bell inequality," probably some form of non-separability is involved, in metaphorical correspondence to entanglement. Likely the work describes a measurement scenario with some similarity to the ones used in quantum Bell experiments. However, as we have argued, for a Bell inequality to be used as a meaningful sense-making test, it must be derived appropriately for the types of causal mechanisms that the research intends to test. Parts of the Bell argument may function as portable metaphors, but the exact derived bounds, and the interpretation of measured correlations in a given case, are not portable pieces of argumentation. They relate to "why" questions that are measurement-specific.

Based on what we have learned from physics, it is possible to create apparent Bell-inequality violations in systems governed by classical physics and classical causality. This does not explain quantum Bell test correlations, nor does it show that the same thing is happening in a classical system. Historically, violation of Bell inequalities may have been considered as a type of special test for quantumness, but recent work, like that referenced in this chapter, shows that the meaning of a Bell-inequality violation can only be rigorously interpreted within a full-blown Bell test.

We have to derive appropriate statistical bounds articulating the "why" explanations we wish to test for these measurements.

The Causal Bell Theorem provides a useful tool for thinking about Bell tests, by recasting the situation in terms of unicorn-like fine-tuning. This reasoning is particularly applicable to some of the toy model examples we have seen, e.g., (Aerts and de Bianchi 2019). A scenario that explicitly defines a way to produce Bell-violating correlations in a measurement is like a graduate student explicitly inventing a way to introduce an even-odd time correlation in motion lab data. The extra correlation would be "spooky" only if we imagine that it could happen without the intentional, extra, fine-tuned causation. In other words, a professor "grading" the results would only find the data suspicious if she expected a different causal process than the one actually involved. Thus, toy-model examples that articulate a causal process to violate Bell inequalities are different from a fully wrought Bell test. The same comments could easily apply for other purported Bell-inequality violations. As physicists, we are less sure how to approach sense-making with data sets of word or concept associations. But we can ask: are there well-defined expectations about the statistical behavior of such data sets, given classical causal reasoning? A "yes" response to that question, and an experiment-specific Bell-inequality derivation, would seem necessary in order to interpret any purported violation.

As of the writing of this chapter, we think there is still something special in quantum physics Bell tests that does not happen in classical optics, in word associations, rubber bands, or other contexts. We subjectively judge the problem to be "spooky," but the important point is that the spookiness is not trivially imitated in systems that share some modeled structures. It has to do with our expectations about the causal processes involved.

9 In Conclusion: Prove Us Wrong

We began this chapter with two claims: (a) that there is something spooky in specific Bell test results in quantum physics and (b) that this spookiness is not present in other modeling contexts.

For us, claim (a) is amenable to modification as research in quantum foundations progresses. In the space of technical debates about quantum physics, we could plausibly be convinced that nothing is particularly spooky about loophole-free Bell test results, e.g., if we are convinced to adopt particular approaches to quantum interpretation (some of which view Bell tests as unremarkable). We also could have our views shifted by novel results in the cross-comparison of classical and quantum optics systems, as just one example. A number of the results we reference in this chapter are relatively recent. Thus the conclusions we draw from them may evolve as more experiments are performed and the discussion matures. Quantum Bell tests have seemed strange for some decades now, but they may not seem strange in the future, depending on what we learn next.

Claim (b) is the one we stand behind firmly for this chapter. It is based primarily on general points about modeling; that modeling employs a type of metaphor. Cross-disciplinary "how" similarities may exist between two disciplinary domains in which the causal "why" radically differs. Since the Bell test spookiness is about causal sense-making, it does not translate in a simple way from one measurement scenario to another. This is clearly demonstrated within physics, considering the differences in interpretation needed to evaluate Bell-like scenarios involving classical light.

Still, claim (b) can also be confronted and potentially falsified. If a Bell-like test situation in a non-quantum physics context is analyzed in such a way as to show persistent correlations that cannot be explained with the causal mechanisms that make sense in that case, it may also be (like Bell tests in quantum physics) spooky in an important way. To make the argument, though, most likely the relevant inequalities or important bounds need to be re-derived in a customized, situation-specific manner.

A methodological prescription for attempting such an effort (to prove that a Bell-like test in a novel context is similarly "spooky") might begin with formally articulating a causal model and assessing consistency within that framework. Informally, think about unicorns. Can the data be explained causally, with no fine-tuning? If so, it is different from what is happening in quantum Bell tests. If not, then possibly you will indeed join us in questioning causality in your measurement context, spooked by your own specific Bell test ghost.

Acknowledgments GBL acknowledges financial support from the Brazilian agencies CNPq and FAPERJ (JCNE E-26/201.438/2021). We thank reviewers Fernando de Melo and Jason Gallicchio for helpful comments that substantially improved the final draft.

References

Diederik Aerts, Jonito Aerts Arguëlles, Lester Beltran, Suzette Geriente, and Sandro Sozzo. Entanglement in cognition violating Bell inequalities beyond Cirel'son's bound. *arXiv preprint arXiv:2102.03847*, 2021.

Diederik Aerts and Massimiliano Sassoli de Bianchi. When Bertlmann wears no socks. common causes induced by measurements as an explanation for quantum correlations. *arXiv preprint arXiv:1912.07596*, 2019.

Andrea Aiello, Falk Töppel, Christoph Marquardt, Elisabeth Giacobino, and Gerd Leuchs. Quantum-like nonseparable structures in optical beams. *New Journal of Physics*, 17(4):043024, 2015.

John S Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

John S Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern physics*, 38(3):447, 1966.

Lester Beltran and Suzette Geriente. Quantum entanglement in corpuses of documents. *Foundations of Science*, 24:227–246, 2019.

C. V. S. Borges, M. Hor-Meyll, J. A. O. Huguenin, and A. Z. Khoury. Bell-like inequality for the spin-orbit separability of a laser beam. *Phys. Rev. A*, 82:033833, Sep 2010.

Harvey R Brown and Christopher G Timpson. Bell on Bell's theorem: The changing face of nonlocality. *arXiv preprint arXiv:1501.03521*, 2014.

Adán Cabello and Guillermo García-Alcaine. Bell-Kochen-Specker theorem for any finite dimension. *Journal of Physics A: Mathematical and General*, 29(5):1025, 1996.

Eric G Cavalcanti. Classical causal models for Bell and Kochen-Specker inequality violations require fine-tuning. *Physical Review X*, 8(2):021018, 2018.

Eric G Cavalcanti and Raymond Lal. On modifications of Reichenbach's principle of common cause in light of Bell's theorem. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424018, 2014.

Eric G Cavalcanti and Howard M Wiseman. Bell nonlocality, signal locality and unpredictability (or what Bohr could have told Einstein at Solvay had he known about Bell experiments). *Foundations of Physics*, 42:1329–1338, 2012.

Daniel Collins and Sandu Popescu. Classical analog of entanglement. *Physical Review A*, 65(3):032321, 2002.

Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

S. J. van Enk and C. A. Fuchs. Quantum State of a Propagating Laser Field. *Quantum Information and Computation*, 2:151–165, 2002.

Diego Frustaglia, José P Baltanás, María C Velázquez-Ahumada, Armando Fernández-Prieto, Aintzane Lujambio, Vicente Losada, Manuel J Freire, and Adán Cabello. Classical physics and the bounds of quantum correlations. *Physical review letters*, 116(25):250404, 2016.

Marissa Giustina, Marijn AM Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, et al. Significant-loophole-free test of Bell's theorem with entangled photons. *Physical review letters*, 115(25):250401, 2015.

Matias A Goldin, Diego Francisco, and Silvia Ledesma. Simulating Bell inequality violations with classical optics encoded qubits. *JOSA B*, 27(4):779–786, 2010.

Nicholas Harrigan and Robert W Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40:125–157, 2010.

Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575):682–686, 2015.

Ebrahim Karimi and Robert W Boyd. Classical entanglement? *Science*, 350(6265):1172–1173, 2015.

Michael Kernaghan and Asher Peres. Kochen-Specker theorem for eight-dimensional space. *Physics Letters A*, 198(1):1–5, 1995.

DN Klyshko. A simple method of preparing pure states of an optical field, of implementing the Einstein–Podolsky–Rosen experiment, and of demonstrating the complementarity principle. *Soviet Physics Uspekhi*, 31(1):74, 1988.

Simon Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17(1):59–87, 1967.

Raymond ST Lee. *Quantum finance*. Springer, 2020.

Tao Li, Xiong Zhang, Qiang Zeng, Bo Wang, and Xiangdong Zhang. Experimental simulation of monogamy relation between contextuality and nonlocality in classical light. *Opt. Express*, 26(9):11959–11975, Apr 2018.

Marcin Markiewicz, Dagomir Kaszlikowski, Paweł Kurzyński, and Antoni Wójcik. From contextuality of a single photon to realism of an electromagnetic wave. *npj Quantum Informationvolume*, 5:5, 2019.

Melanie McLaren, Thomas Konrad, and Andrew Forbes. Measuring the nonseparability of vector vortex beams. *Physical Review A*, 92(2):023833, 2015.

Joshua Montgomery, Adam J Anderson, Jessica S Avva, Amy N Bender, Matt A Dobbs, Daniel Dutcher, Tucker Elleflot, Allen Foster, John C Groh, William L Holzapfel, et al. Performance and characterization of the SPT-3G digital frequency multiplexed readout system using an

improved noise and crosstalk model. In *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy X*, volume 11453, pages 167–188. SPIE, 2020.
JC Pearl and EG Cavalcanti. Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios. *Quantum*, 5:518, 2021.
Judea Pearl. *Causality*. Cambridge university press, 2009.
L. J. Pereira, A. Z. Khoury, and K. Dechoum. Quantum and classical separability of spin-orbit laser modes. *Phys. Rev. A*, 90:053842, Nov 2014.
Jacques Pienaar. Causality in the quantum world. *Physics*, 10:86, 2017.
Jacques Pienaar. Quantum causal models via quantum Bayesianism. *Physical Review A*, 101(1):012104, 2020.
Sandu Popescu and Daniel Rohrlich. Causality and nonlocality as axioms for quantum mechanics. In *Causality and Locality in Modern Physics: Proceedings of a Symposium in honour of Jean-Pierre Vigier*, pages 383–389. Springer, 1998.
Emmanuel M Pothos and Jerome R Busemeyer. Quantum cognition. *Annual review of psychology*, 73:749–778, 2022.
Xiao-Feng Qian and J. H. Eberly. Entanglement and classical polarization states. *Opt. Lett.*, 36(20):4110–4112, Oct 2011.
Xiao-Feng Qian, Bethany Little, John C Howell, and JH Eberly. Shifting the quantum-classical boundary: theory and experiment for statistically classical optical fields. *Optica*, 2(7):611–615, 2015.
Massimiliano Sassoli de Bianchi. Using simple elastic bands to explain quantum mechanics: a conceptual review of two of Aerts' machine-models. *Open Physics*, 11(2):147–161, 2013.
Kathryn Schaffer and Gabriela Barreto Lemos. Obliterating thingness: an introduction to the "what" and the "so what" of quantum physics. *Foundations of Science*, 26:7–26, 2021.
Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical review letters*, 115(25):250402, 2015.
R Simon and GS Agarwal. Wigner representation of Laguerre–Gaussian beams. *Optics letters*, 25(18):1313–1315, 2000.
Robert JC Spreeuw. A classical analogy of entanglement. *Foundations of physics*, 28(3):361–374, 1998.
David Stoler. Operator methods in physical optics. *JOSA*, 71(3):334–341, 1981.
P. Suppes, J. Acacio de Barros, and A.S. Sant'Anna. A proposed experiment showing that classical fields can violate Bell's inequalities. *ArXiv:9606019*, 1996.
Ilya A Surov, E Semenenko, AV Platonov, IA Bessmertny, F Galofaro, Zeno Toffano, A Yu Khrennikov, and AP Alodjants. Quantum semantics of text perception. *Scientific Reports*, 11(1):1–13, 2021.
SJ Van Enk. Entanglement of electromagnetic fields. *Physical Review A*, 67(2):022303, 2003.
Howard M Wiseman. From Einstein's theorem to Bell's theorem: a history of quantum non-locality. *Contemporary Physics*, 47(2):79–88, 2006.
Christopher J Wood and Robert W Spekkens. The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning. *New Journal of Physics*, 17(3):033002, 2015.
Jia-Min Xu, Yi-Zheng Zhen, Yu-Xiang Yang, Zi-Mo Cheng, Zhi-Cheng Ren, Kai Chen, Xi-Lin Wang, and Hui-Tian Wang. Experimental demonstration of quantum pseudotelepathy. *Physical Review Letters*, 129(5):050402, 2022.