Learning from Theories

Michael Dascal*
University of Maryland,
College Park, MD

There is a fundamental, but often overlooked assumption in the practice of science. This is a simple belief that predictive accuracy is a good measure of representational accuracy. The more a theory's predictions are verified, the more we're inclined to believe that the theory describes what's genuinely 'out there' in the world. Unfortunately, this assumption isn't as straightforward as it appears.

In 1930, Gödel famously proved that a sufficiently strong theory can't prove every true sentence in its language. This result has a surprisingly consequence, though. It forces us to distinguish between sentences that are true 'merely within' a theory – true only about the theory itself – from those that are true about the theory's "subject matter", to borrow Boolos's term [2]. But once we acknowledge this distinction, we find that lacking *pre-theoretic* knowledge about a theory's subject matter can make it difficult, if not impossible to learn anything at all from how a theory represents its subject matter.

I'll explain how this follows here, first by looking to Gödel's theorem[1] in the context of arithmetic and then by extending the result to classical and relativistic physics, and finally to quantum mechanics. As I do this, I'll often refer to the *structure* of theories and subject matters. This isn't meant to carry any foundational or metaphysical weight. Rather, I believe it's simply a natural way to think about formal systems in the relevant manner. If it helps, you might think of it as an abstraction of objects and relations that are relevant to a theory or a domain of discourse.

I. GÖDEL'S INCOMPLETENESS AND WHAT A SENTENCE IS ABOUT

My first task is to explain how Gödel's theorem forces us to distinguish between a sentence's being true only within a theory from its being true of the theory's subject matter. This is easiest to see in the case of arithmetic because we have a very clear concept of its intended subject matter, the natural numbers ($\mathbb{N}$). This allows us to keep clear about what Gödel's theorem does and doesn't imply of arithmetic theories and of $\mathbb{N}$. We can then consider the implications of this distinction in other contexts.

There are many theories of arithmetic, each very different from the next. Just to mention a few, Peano arithmetic (A_{Pn}) contains the successor function, recursively defined addition and multiplication, and a second-order induction axiom (which can be changed to a first-order axiom schema). Robinson arithmetic (A_{R}) includes the successor function, recursive addition and multiplication, but not induction. Lastly, Presberger arithmetic (A_{Pb}) includes the successor function, addition, and a first-order induction axiom schema, but not multiplication. There are many other theories, of course, but I mention these because A_{R} and A_{Pb} are subsets of A_{Pn}, but both are complete.

Importantly, each of these is meant to represent the structure described by the Dedekind-

* mdascal@umd.edu

[1] Note that here and in what follows, I am referring to Gödel's first incompleteness theorem.

Peano axioms.[2] It may seem an odd point to raise, but I take it that this means that they all aim to represent *the* natural numbers. That is, there aren't different sets of natural numbers, $\mathbb{N}_{\mathsf{A_{Pn}}}$, $\mathbb{N}_{\mathsf{A_R}}$, etc., but rather a single $\mathbb{N}$. I also take it that $\mathbb{N}$ exhibits a unique structure and that any theory that can represent this structure can represent $\mathbb{N}$.[3]

Gödel's theorem states that given an axiomatic theory T that can represent $\mathsf{A_{Pn}}$ in a language L, there exist Gödel sentences – true sentences in L that T can't prove. This immediately raises a question though: given that there are true but unprovable sentences in Peano arithmetic, does this mean that there are 'true but unprovable' sentences (or predications, perhaps) about the natural numbers? In other words, does Gödel's theorem teach us about $\mathbb{N}$?

I believe the answer here is "no", but that we can only answer this *because* we have such a clear concept of the structure of $\mathbb{N}$. One way to see this is to consider two cases: Take G to be a Gödel sentence in $\mathsf{A_{Pb}}$. Either G is in the language of at least one of $\mathsf{A_R}$ or $\mathsf{A_{Pb}}$, or it isn't expressible in either. In both cases we see that the unprovability of G in $\mathsf{A_{Pn}}$ doesn't tell us anything about $\mathbb{N}$.

In the first case, if G is about $\mathbb{N}$ then it must be that its unprovability in $\mathsf{A_{Pn}}$ is the result of some additional structure afforded by the richer theory. That is, the fact that G is expressible, (and thus provable) in $\mathsf{A_R}$ or $\mathsf{A_{Pb}}$ means that the proof must somehow 'come apart' with the additional structure of $\mathsf{A_{Pn}}$. (Note that if G is about $\mathbb{N}$ then it can't be true in $\mathsf{A_{Pn}}$ and false in $\mathsf{A_R}$ or $\mathsf{A_{Pb}}$.)

It's very strange to think about a stronger theory 'losing' a proof that's available in a weaker one, but ignoring this difficulty, the fact that G is provable in some other theory means that its unprovability in $\mathsf{A_{Pn}}$ can't teach us anything about $\mathbb{N}$ – it's merely a fact about $\mathsf{A_{Pn}}$. This might be clearer if we consider a more intuitive scenario: Imagine some sentence P could be proven in some theory $\mathsf{T_1}$ but not in a weaker theory $\mathsf{T_2}$. We certainly wouldn't think that the unprovability of P in $\mathsf{T_2}$ would teach us anything about the subject matter - it would just be the result of the $\mathsf{T_2}$'s weakness. The case here is exactly the same, except that G is somehow provable in the weaker theory rather than the stronger one.

In the case where G isn't expressible in the other theories, we can't address what its unprovability might imply without first considering this difference in expressibility. More specifically, we must ask whether this difference implies anything about whether G is about $\mathbb{N}$. To do so, let's simplify things a bit. Take an arbitrary sentence P that's expressible in $\mathsf{A_{Pn}}$ but not $\mathsf{A_{Pb}}$. We might ask whether it's possible that P expresses something about $\mathbb{N}$ that can't be captured in $\mathsf{A_{Pb}}$. If P is simple – perhaps "7 is prime" – this might seem obviously true. However, if we consider that "prime" is only definable with recursive multiplication then we see that the property can't be captured with the structure of $\mathbb{N}$ alone. "7 is prime" is not a statement in the language of the natural numbers – it is only a statement within a theory that contains the right structural enrichment of $\mathbb{N}$.

Of course, this sentence is true *in virtue* of something about $\mathbb{N}$. We can even find a sentence that is truth-functionally equivalent to P in $\mathsf{A_{Pb}}$ (or a theory that closely resembles $\mathsf{A_{Pb}}$). For example, we could extend $\mathsf{A_{Pb}}$ into a theory $\mathsf{A'_{Pb}}$ by adding *non-recursive* multiplication in the form of axiomatic definitions of functions "$\cdot x$" for each number x. "7 is prime" would then be truth-functionally equivalent to the $\mathsf{A'_{Pb}}$ sentence, "there are no numbers x

[2] These are the axioms formalized by Peano (but attributed to Dedekind) to describe the natural numbers. Peano arithemetic then adds recursive addition, recursive multiplication and induction to these axioms. (See [18, p.90, ft.1], for example.)

[3] This is not to claim anything about the existence of $\mathbb{N}$ (or its structure) independent of the theories that represent it. Note, though, that when I talk about physics I do take the physical world to have a structure independent of our theories.

and y less than 7 such that $x \cdot y = 7$ or $y \cdot x = 7$." But this doesn't quite capture what it is for a number to be *prime*. Note that the $\mathsf{A}'_{\mathsf{Pb}}$ sentences built like this to be truth-functionally equivalent to "7 is prime" and "13 is prime" describe very different properties of 7 and 13. The property described in the first, but not the property described in the second, would be true of the number 77, for example. What we call "prime" would require an infinitely long definition in $\mathsf{A}'_{\mathsf{Pb}}$.

If my claim that "7 is prime" is only true within A_{Pn} seems strange then consider a slightly more obvious case – for example, the sentence that multiplication is commutative. Here it's a little clearer why the sentence is about the structural enrichment afforded by A_{Pn} and not the structure of $\mathbb{N}$. To assert that multiplication is commutative is to say something about recursively defined multiplication, and we know that this operation isn't required to capture $\mathbb{N}$.

My point here is that if G is not expressible in A_{R} or A_{Pb} then if it were truly about $\mathbb{N}$ – if it weren't true merely within A_{Pn} – then there would exist a sentence in the language of $\mathbb{N}$ that would capture the content of G perfectly and completely. But if so, there would necessarily be an equivalent (not merely truth-functionally equivalent) sentence in all theories that represent $\mathbb{N}$. In turn, this means that A_{R} and A_{Pb} would contain a sentence equivalent to G, and moreover, that this sentence would be provable in these theories (because they're complete). Putting this all together then, if G isn't expressible in the simpler theories then it's either true merely within A_{Pn} or, as in the first case, its unprovability doesn't teach us anything about $\mathbb{N}$.

In both cases then we find that the unprovability of the Gödel sentence tells us nothing about the natural numbers. This isn't a problem of course, and indeed it isn't very surprising. We have a very clear concept of $\mathbb{N}$, and the Dedekind-Peano axioms describe it perfectly (albeit not uniquely!), so we can always clearly see the differences between the structure of $\mathbb{N}$ and the structures of the theories about it. However, taking a step back we realize that provability of Gödel's theorem *itself* demonstrates something we may not have expected, which is that that our theories may contain provable sentences which have nothing to do with their subject matter. That is, Gödel's proof, *qua* proof of a sentence that describes the existence of a Gödel sentence, proves a sentence that is not about the theory's subject matter, but rather about the theory itself.

One way to think of all this is that a theory's axioms all stand on equal footing. There's no differentiation between those that capture the subject matter's structure and those that build on top of it. (Moreover, there's no guarantee that such a delineation exists in the axioms.) This means that we can't expect a theory to 'indicate' which of its theorems are about its subject matter and which aren't. This may not be worrisome in the case of arithmetic, but the consequences of this are troublesome when our knowledge of the subject matter is at all obscured, as is the case in physics.

II. CLASSICAL AND RELATIVISTIC PHYSICS

I don't want to enter into a discussion of scientific realism here, but many take it as the goal of physics to represent the objects and properties that exist in physical reality. As Einstein says, "the concepts of physics refer to a real external world, i.e. ideas are posited of things that claim a 'real existence' independent of the perceiving subject." [9] (as translated by [13, p.190]). Unlike the case of arithmetic though, when it comes to physics our pre-theoretic knowledge of the subject matter is at least somewhat obscured. In both classical and relativistic physics we have direct perceptual access to some of the subject matter – we want to model the behaviors of systems that we experience, from tables and chairs to gases and galaxies. However, there are also (1) cases where our theories add structure to that of the subject matter, and (2) 'blind spots' in the structure of the subject matter that become apparent when we consider some of the implications of our theories.

Cases of the prior include emergent macroscopic properties of microscopic ensembles. For example, temperature, which we identify with average molecular kinematic energy, can't play a fundamental role in how systems evolve. This doesn't mean that temperature isn't real or doesn't have explanatory power, of course. Rather, the addition of temperature to the fundamental structure of the universe is analogous to enriching the Dedekind-Peano axioms with recursive addition. It expands our ability to predict and predicate over the subject matter. Underlyingly, however, we take the microphysics of a system to give rise to its thermodynamics just as summations are only true in virtue of the successor function.

These types of cases usually mark known distinctions between the structure of the subject matter and the structure added by our physical theories. This is not so in the case of 'blind spots', which are somewhat trickier to point to just because the structure of the physical universe is somewhat obscured. Some *might* be identifiable through 'redundancies' that appear in our physics. For example, Newton's laws depend on the first and second time derivatives of position, and we speak of physical systems as if they genuinely have instantaneous velocities and accelerations within them somehow. However, it isn't clear that these properties are truly 'out there' in these systems. The laws may be correct and may accurately predict physical evolutions and interactions, but this doesn't necessitate the metaphysical existence of these properties in the structure of the world. Differentiation with respect to time could be to Newtonian physics as recursion is to Peano arithmetic – a method of enriching the underlying structure of a subject matter.

Given that there is at least this 'blurry' line between the structure of the subject matter and the theoretical enrichment afforded by our physics, we can now consider the results of what we've seen above. First, note that the applicability of Gödel's theorem is somewhat easily satisfied. If a physical theory is a mathematical theory combined with physical laws that constrain the evolution of some domain of objects, then all we need is for the mathematical theory be at least as strong as Peano arithmetic (which is trivially true in the case of physics), and that the laws be axiomatizable, which I take as granted here.[4] Gödel's theorem then tells us that there exists true sentences in our physical theories that they can't prove.

Knowing a theory contains a Gödel sentence is one thing but identifying it can be a different thing entirely. There is an alternative that will prove useful here, and this is to find 'Gödel-like' sentences – sentences which *can* be true despite being certainly unprovable in our physical theories. There are two examples I wish to look at, both inspired by Barrow's discussion on the subject [1]:

The first derives from Norton's dome [14], a theoretical dome designed so that if a sphere were to roll up its side with the right initial velocity it would come to rest at its apex in a finite amount of time. Norton describes the dome to highlight a difficulty with determinism in Newtonian mechanics: Given that Newtonian laws are time-symmetric, if a sphere were to be placed at the top of the dome it would be consistent with the theory that it wait for some indeterminable period of time before spontaneously rolling down its side.

[4] See [1, p.10] for a discussion of the applicability of the theorem to physics. Note that I don't mean to trivialize Hilbert's sixth problem! I take independent axiomatizations of classical, relativistic and quantum mechanics as sufficient here.

Now, say such a dome were built and sphere placed at its apex. Ignoring microfluctuations, etc., if the sphere doesn't sit at this point indefinitely then we know there would be a true sentence that is unprovable in Newtonian physics – the one that describes the time the sphere begins to move. And now we can ask, does the fact that physics can't establish the truth of this sentence represent a fact about reality, or merely a fact about the theory? More fundamentally, does the fact that Newton's laws are provably time-symmetric represent a property of the physical world or merely a property of the theory we use to describe it? Without some independent, pre-theoretical knowledge of the structure of the physical universe, it isn't at all clear how we might even begin to answer this question. That is, we can't know whether statements about time-symmetry represent anything about the universe at all. It may be that these represent only properties of our model.

An analogy may be helpful here: Recall that we need recursive multiplication to define 'prime'. We saw that this doesn't mean that a weaker theory can't express some truth-functionally equivalent sentence to one of the form "n is prime". Similarly, it may be that the structure of the universe describes each system as exhibiting time-symmetric evolution, but that it does not contain the structure required to describe *time-symmetry*. (Note, for example, that such a property seems to be second-order. It's a constraint on the constraints on physical evolution.)

The second example of a Gödel-like sentence follows from the hole argument. In developing general relativity Einstein noticed that with general covariance came an under-determination of how the metric field in an empty region of spacetime would 'sit' on the spacetime manifold [19]. This has led to a renewed debate between spacetime substantivalism and relationalism.[5] It may be obvious how this relates to our discussion: given that we are unsure of the true ontology of spacetime, it isn't clear how we should interpret this under-determination. More in line with how I've been speaking, it's possible that there is a true sentence that (uniquely) describes how the metric field attaches to the manifold in such a region, but such a sentence is certainly unprovable in our physics. Again, we're left unsure about whether this under-determination represents merely a feature of the theory or whether it reflects a true freedom in the physical universe.

I'm not the first to raise this point in regard to the hole argument. Curiel [7] argues something very similar:

> just because the mathematical apparatus of a theory appears to admit particular mathematical manipulations does not *eo ipso* mean that those manipulations admit of physically significant interpretation.... The mathematical formalism by itself cannot tell us what manipulations it admits have physical significance; one must determine what one is allowed to do with it – 'allowed' in the sense that what one does respects the way that the formalism actually represents physical systems. [7, p.452].

Without pre-theoretic knowledge of the structure of spacetime – knowledge of what is "allowed", as Curiel says – we don't know what to make of this mathematical freedom.

Now my point here isn't simply about Norton's dome or the hole argument. The point is that our theories contain provable sentences which might not reflect anything about the structure of the physical universe. The fact that we don't have clear access to this structure

[5] The original substantivalist/relationalist debate provides further examples of a Gödel-like sentences, but it seems better to look to relativity theory given its success.

precludes us from confidently differentiating between whether certain sentences are true merely within the theories or true of their subject matters.

Admittedly, just as in the case of arithmetic, these cases might not cause much concern. Our knowledge of the subject matter of classical and relativistic physics may be slightly obscured, but at the end of the day we have a good idea of what to make of most of what these theories say: We know how tables and chairs behave, even if we can't give them robust microscopic descriptions. We know how a gas disperses in a room even if we can't trace its microscopic evolution. We even know exactly how two observers will disagree about what they observe even though we can't know whether or not there exists a privileged inertial frame. Things become much worse when we consider quantum mechanics, however.

III. LEARNING FROM QUANTUM MECHANICS

There is a very interesting and highly relevant fact about the development of quantum mechanics. In 1925 and 1926, two independent formulations were developed to describe our empirical results, Heisenberg's matrix mechanics [12] and Schrödinger's wave mechanics [15]. Of course, Schrödinger immediately demonstrated that the two were equivalent [16], but nonetheless we have two theories that offer very different representations of quantum systems. (And I haven't even considered Feynman's path integral formulation!)

The existence of these competing pictures reflects something quite obvious: When it comes to quantum mechanics, we barely have any pre-theoretic knowledge of the theory's subject matter. We may have concepts of the particles and interactions the theory seems to represent, but these conceptual particles can fluctuate between being wave-like and particle-like as required by the problem or situation in front of us. Similarly, we move back and forth between using matrices and wave mechanics knowing full well that the result of a particular calculation won't depend on which we use. Although this mental flexibility has proven immensely productive in our ability to predict measurement outcomes, it also demonstrates that these concepts bear no resemblance to the objects that we aim to represent. It is (by a very strong assumption I think!) impossible for a quantum system's wave/particle nature to *actually* fluctuate in the way our concepts of them can, and it is certainly not the case that this nature changes *because* of how we calculate predictions about it. Of course, it may be that the true state of a quantum system is entirely unlike the concepts we use to represent them, but this only further demonstrates our ignorance of the true structure of our theory's subject matter.

This observation isn't new, of course. Only nine years after Schrödinger completed his wave mechanics did he use his famous cat thought experiment to demonstrate that naïve quantum state realism contradicts our experience and experimental data [17]. And as we know well, this discrepancy between the quantum state and our observations is at the root of the measurement problem. In fact, there is a sense in which the measurement problem is even worse than it appears at first. We might think that the problem is merely about how we can describe the evolution of a quantum superposition into a single measurement outcome given that the theory only permits reversible dynamics. However, if we consider a measurement of position, for example, we find that not only is there this 'dynamical divergence' between the quantum mechanical evolution and our observations, but there is also a kinematic one. Our classical concept of a position is perfectly precise – Euclidean positions are zero-dimensional points in space. However, position eigenstates, *qua* delta functions in position space, are not physically permitted in quantum mechanics. (They are not normalizable, and hence unphysical). This means that our conceptual picture of the structure of these objects is even further away from their true structure than we might have thought.

Given that the subject matter of quantum mechanics is so obscured, what should we say about whether we learn from quantum theory? One path forward might be to again look for Gödel or Gödel-like sentences to see where the theory and subject matter might come apart. And one possible case becomes immediately apparent if we consider sentences that describe single measurement outcomes. Say an x-spin measurement on a $|z+\rangle$ electron results in $|x+\rangle$. The sentence describing this result is certainly unprovable in quantum mechanics. But this case is importantly different from the ones we saw in the discussion of classical and relativistic physics. There it wasn't clear whether a sentence, for example about the time-symmetry of physical laws, was true outside of our theory. Here we're outright denying our theory from the start by positing something the it explicitly precludes, so using such a case to show that we can't learn from quantum mechanics would be entirely question-begging.

In fact, this is the fundamental problem with trying to learn from quantum mechanics. No matter which 'language' we pick to describe the theory (i.e. matrices, wave functions, path integrals or even fields) there will always be a significant discrepancy between what the theory describes and what we observe. In the case of arithmetic, and to some degree in the case of classical physics, we were able to consider how the content of a Gödel or Gödel-like sentence might relate to the intended subject matter. In the case of quantum mechanics, not only do we have no direct access to the structure of the subject matter, but we *know* that our concepts of the subject matter *cannot* represent its actual structure. (And this is worsened by the fact that there isn't even a unique picture drawn by the theory given its different formulations.) As such, it begins to look like we have very good reason to worry about taking the theory to reflect the structure of its subject matter.

One thing I should make explicit here is that Gödel's theorem has merely offered a reference or 'proof of concept' in what I've described, and this is because its proof is based entirely on the formal structure of a theory (independent of its subject matter). As such, it offers a good general direction for finding cases where it can be difficult to know whether or not a sentence is true merely within the theory (assuming the subject matter is at all obscured). But there is no reason to think that it's a unique source for these cases. Just as Gödel's proof demonstrates the existence of Gödel sentences, the provability of the theorem demonstrates the existence of sentences merely within the theory. And it isn't difficult to see that if these exist, then without clear pre-theoretical knowledge of the structure of the subject matter *there is no way to determine whether a given sentence in the theory reflects anything about its subject matter at all.* What this means is that projects that attempt to retrieve our classical observations from quantum state ontologies, for example, seem to be fundamentally misguided. They begin by assuming that the theory provides insight into the structure of the quantum world, and we've seen that this is certainly not something we can take for granted.

I highlight this detail both for clarity and because there has recently been some discussion in the quantum foundations literature that significantly mirrors the claims I've made here, albeit from an entirely different direction. In 2016, a manuscript by Frauchiger and Renner [10] claimed to prove that no single world theory of quantum mechanics could consistently maintain the universality of unitary evolution. (This was later revised into [11].) In response to this, Bub offered his own interpretation of their result, which parallels the opinions of Schrödinger in 1935. As Bub explains [3–6], Frauchiger and Renner prove that any (non-

Bohmian) theory which maintains the universality of unitarity and the existence of single measurement outcomes must take quantum mechanics to be a *probabilistic* theory, not a *representational* one. In other words, if we wish to accept our experience and empirical data (of single measurement outcomes) without 'breaking' quantum mechanics or adding new measurement dynamics to it, then we must give up any attempt to learn metaphysical or ontological lessons from the theory. Although we may use the theory to make highly accurate predictions, our assumption that this indicates some sort of successful representation of the 'true nature' of the world is unjustified – and can lead to unwanted conclusions, be they inconsistencies, branching universes, or perhaps extreme solipsism. As a result, we're forced to approach any attempt to learn from the theory's representation of its subject matter with a strong skepticism.

IV. CONCLUSION

We've seen from the relatively simple case of arithmetic that Gödel's theorem inadvertently proves a weakness of a sufficiently strong formal theory – it will be able to prove sentences that have nothing to do with the theory's subject matter. This may not be worrisome when we know the subject matter independently of the theory, as in the case of $\mathbb{N}$, but when we turn to physics this result has some deep implications. We find here that our ability to learn about the structure of the physical universe can be at best hindered, as in the case of classical and relativistic physics, and at worst blocked completely, as may be the case in quantum mechanics.

Admittedly, this skepticism is unsatisfying. It doesn't seem that this is what physics is *for* – it feels like it's meant to *represent and inform us* about the structure of the physical world, not merely to predict our observations of it. In the context of quantum physics alone this seems to describe part of the progress that's been made at least since Dirac's prediction of the anti-electron in 1931 [8]. However, it's a mistake to confuse accurate prediction with accurate representation. The Standard Model provides us with predictions of observations in particular experimental circumstances, and what we have seen here is that this is not the same as providing a list of fundamental particles that necessarily exist and have properties in the way that tables and chairs do. Of course, this is not to call any of the theory's predictions into question. Were they inaccurate then we would know our model to be incorrect and would be forced to reevaluate it entirely. Rather, what is left unknown is how we might begin to learn about the parts of reality that we don't *already* know about, from nomological possibility to spacetime ontology to the fundamental nature of quantum systems.

[1] Barrow, J. D. Gödel and Physics. In *Kurt Gödel and the Foundations of Mathematics: Horizons of Truth*, M. Baaz, Ed. Cambridge University Press, 2011, pp. 255–276.

[2] Boolos, G. *Logic, logic, and logic.* Harvard University Press, 1998.

[3] Bub, J. Why Bohr was (Mostly) Right. *arXiv preprint: arXiv:1711.01604* (2017).

[4] Bub, J. What is really there in the quantum world? In *Philosophers Look at Quantum Mechanics*, A. Cordero, Ed. Springer, 2019.

[5] Bub, J. 'Two Dogmas' Redux. In *Quantum, Probability, Logic. Jerusalem Studies in Philosophy and History of Science.*, M. Hemmo and O. Shenker, Eds. Springer, Cham, 2020.

[6] Bub, J. In Defense of a "Single-World" Interpretation of Quantum Mechanics. *Studies in the History and Philosophy of Modern Physics* (forthcoming). Preprint: arXiv:1804.03267.

[7] Curiel, E. On the existence of spacetime structure. *The British Journal for the Philosophy of Science 69*, 2 (2018), 447–483.

[8] Dirac, P. A. M. Quantised singularities in the quantum field. *Proceedings of the Royal Society of London, A 133*, 821 (1931), 60–72.

[9] Einstein, A. Quanten-mechanik und wirklichkeit. (Quantum Mechanics and Reality). *Dialectica 2*, 3-4 (1948), 320–324. Partially translated by Don Howard in [13].

[10] Frauchiger, D., and Renner, R. Single-world interpretations of quantum theory cannot be self-consistent. *arXiv preprint: arXiv:1712.07207* (2016).

[11] Frauchiger, D., and Renner, R. Quantum theory cannot consistently describe the use of itself. *Nature Communications 9* (2018).

[12] Heisenberg, W. Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift fur Physik 33* (1925), 879–893.

[13] Howard, D. Einstein on locality and separability. *Studies in History and Philosophy of Science Part A 16*, 3 (1985), 171–201.

[14] Norton, J. D. Causation as folk science. *Philosophers' Imprint 3*, 4 (2003), 1–22.

[15] Schrödinger, E. Quantisierung als eigenwertproblem i–iv. *Annalen der Physik 79, 79, 80, 81*, 4, 6, 13, 18 (1926), 361–376, 489–527, 437–490, 109–139.

[16] Schrödinger, E. Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinem. (On the relationship between Heisenberg-Born-Jordan quantum mechanics and my own). *Annalen der Physik 384*, 8 (1926), 734–756.

[17] Schrödinger, E. Die gegenwartige Situation in der Quantenmechanik. (The present situation in quantum mechanics). *Die Naturwissenschaften 23* (1935), 807–812. Translated as Trimmer J. The Present Situation in Quantum Mechanics: A Translation of Schrödinger's "Cat Paradox" Paper. *Proceedings of the American Philosophical Society 124*, 5 (1980), 323-338.

[18] Smith, P. *An introduction to Gödel's theorems*, 2nd ed. Cambridge University Press, 2013.

[19] Stachel, J. The hole argument and some physical and philosophical implications. *Living Reviews in Relativity 17*, 1 (2014).