Realism without interphenomena: Reichenbach's cube, Sober's evidential realism, and quantum solipsism

Florian J. Boge

ABSTRACT
In 'Reichenbach's cubical universe and the problem of the external world', Elliott Sober attempts a refutation of solipsism à la Reichenbach. I here contrast Sober's line of argument with observations a hypothetical scientist makes on what is known as the 'Mermin contraption', and show that Sober's extension of Reichenbach's argument fails. However, there is a further line of defense against solipsism, endorsed by some interpreters of the quantum theory whose position is otherwise rather close to solipsism. As I shall argue, this line of defense remains viable, but we still fall short of a story in terms of what Reichenbach calls 'interphenomena' in the case of the Mermin contraption.

1. Introduction

In 'Reichenbach's cubical universe and the problem of the external world' (*Synthese* (2011) 181:3–21), Elliott Sober attempts a refutation of solipsism à la Reichenbach, albeit in epistemic rather than semantic, and Bayesian rather than frequentist terms. As is well known, Hans Reichenbach had attempted, in *Experience and Prediction*, to replace the verificationist account of meaning by a probabilist one. Whereas according to the former only verifiable sentences have meaning and two sentences with the same verification conditions are synonymous, in the latter 'verifiability' is replaced by the possibility of associating a probability and the same probability assignment under any possible observation implies synonymy (Reichenbach 1938, 54).

Using his probabilistic account, Reichenbach (1938, 133) then shows how to meaningfully distinguish between solipsism and realism—a philosophical opposition that Rudolf Carnap (2003, §5) had coined 'an idle linguistic dispute.' As a theory of meaning, however, Reichenbach's account fails:

> a sentence can be meaningful even if we have no clue as to how probable it is. Consider the sentence 'undetectable angels exist.' If this really were meaningless gibberish, we could not discuss its epistemic status or its logical relationships to other sentences. But we can. (Sober 2011, 5)

Similarly, the frequentist interpretation of probability has certainly gone out of fashion as a general account of probability in science, for various reasons, and Sober (2011, 4.–5) instead opts for a (normative) Bayesian account.

But Reichenbach here also aims at a *refutation of solipsism*, and Sober (2011) shows that it is possible to divorce that part of Reichenbach's arguments from his problematic appeals to semantics and frequencies. The heart of the argument is, in essence, a move from observed correlations to a plea for rejecting solipsism on account of the evidential force these correlations

Interdisciplinary Centre for Science and Technology Studies, Wuppertal University, Gaußstr. 20, 42119 Wuppertal, fjboge@uni-wuppertal.de. This is an Accepted Manuscript of an article accepted for publication in *International Studies in the Philosophy of Science*. The final version will be available at https://doi.org/10.1080/02698595.2021.1964335.

have in favor of a common cause hypothesis.

The intuitive appeal of Reichenbach's argument is captured by his story about the cubical universe and the genius 'Copernicus' who discovers a world outside of it. This is a philosophical thought experiment certainly reminiscent of classics like Plato's cave, 'a beautiful analogy for the problem of the external world.' (Sober 2011, 20)

In this story, mankind is confined to a cubical universe on whose outside walls shadows appear. The shadows displayed on two adjacent walls of the cube exhibit astonishing correlations, so that the story's hero, Copernicus, comes up with the following explanation: There are objects outside the cube, and two matching shadows are just images caused as common effects by some mechanism involving one single object. That's why they are so remarkably correlated!

Indeed, Copernicus is right. Outside the cube, there are birds flying around and a single bird's shadow is projected, by a cleverly contrived system of lights and mirrors, onto two adjacent walls. The appeal of the story, as an argument against solipsism, is that the cube represents an individual's being confined to her physical proportions and sensory capacities, but that an outside world-hypothesis *predicts* the correlations met with in experience, and so has higher evidential support: The correlations exhibited between, say, our auditory and visual experiences is predicted by the assumption that they are both simultaneously caused by an external reality, but not necessarily by one that traces them back to some opaque set of mental states alone. In Sober's words: 'It's the external world that is doing the work, stupid' (Sober 2011, 18).

2. Sober's evidential argument against solipsism

Actually, the story is more complicated than this. In fact, that goes for both, the Copernicus as well as the Reichenbach-Sober story. For in *Experience and Prediction*, Reichenbach does not make use of his famous *Principle of the Common Cause* (Reichenbach 1965) yet, a version of which may be stated as follows:

(PCC) If X and Y are correlated but logico-analytically independent,[1] then either X causes Y, Y causes X, or X and Y are joint effects of a common cause Z that renders X and Y conditionally probabilistically independent ('screens them off').[2]

In *Experience and Prediction*, Reichenbach rather tries to make the point that the relevant common cause hypothesis is favorable over a coincidence hypothesis for various reasons, without invoking PCC directly.

Following Sober (2011, 8), we can reconstruct the reasoning as follows. Let f denote the observed frequency with which an event of a particular type occurs, $F(x)$ the event that x takes on some generic property F at some point in time, and a, b distinct shadows on adjacent walls. The Cubists hence note the following observation:

O: $f(F(a) \wedge F(b)) > f(F(a))f(F(b))$

Call such a relation between frequencies a (positive) *empirical correlation*.[3] There are then (at least) two rival hypotheses to be considered:

CC: $F(a)$ and $F(b)$ are effects of a common cause.
$\bar{C}$: $F(a)$ and $F(b)$ are causally unconnected.

Sober's point now is that, given that CC *implies* a correlation between $F(a)$ and $F(b)$, it renders the observed regular coincidences more probable than the rival coincidence model

($\bar{C}$). Hence,

$$P(O|CC) > P(O|\bar{C}), \tag{L}$$

where P denotes probability, and we can count O as *evidence* for CC.[4] According to the Law of Likelihood (Royall 1997), we should thus favor CC over $\bar{C}$.

However, Sober (2011, §3 & §4) notes two subtleties here: First, it seems that Reichenbach is aiming rather to establish that CC is being made *more probable* by the observations, i.e., that

$$P(CC|O) > P(\bar{C}|O),$$

which does not follow from (L) unless substantial assumptions are being made about priors. And second, Reichenbach also has to address the following additional hypothesis:

CE: $F(a)$ and $F(b)$ are related to one another as cause and effect.

Some of the reasons Reichenbach offers for rejecting CE are rather incoherent (see Sober 2011, 10). But Reichenbach also

> has a footnote [...] that makes it clear that we are not here mistaking what his argument is. He says that the Common Cause hypothesis has a higher posterior probability than the Cause/Effect hypothesis because the former has the higher prior and in spite of the fact that the two hypotheses confer on the observed association the same probability (they have identical likelihoods). (Sober 2011, 10)

It is opaque to me what justifies the association of a higher prior to the bird hypothesis than to a direct causal relation between the shadows other than *petitio*. For given what the story is supposed to represent, assigning a higher prior to the bird story is tantamount to *stipulating* realism, rather than successfully arguing against solipsism. Reichenbach's argument here begs the question.

That does not leave the argument beyond repair: Sober (2011, 13 ff.) considers a possible re-rendering of the general line of argument by appeal to *interventionist* considerations (Woodward 2003). That repair, however, is complicated by yet another fact:

> Reichenbach's thought experiment about the cubical world includes the detail that the people in the cube cannot manipulate the shadow-images that appear on the cube's surfaces. [...] he stipulates that powerful force fields prevent them from doing so. This makes the puzzle harder to solve. Had Copernicus or his successors sent rockets to alter or remove one of the shadow-images in a pair (perhaps by smearing whitewash on it), these manipulations would have yielded evidence that favors [CC] over [CE]. (Sober 2011, 14)

The point Reichenbach is trying to make by stipulating this is presumably that our experiences cannot be altered *ad libitum*. However, that is only partially true, and the point Sober (2011) is trying to make here is exactly that we *are*, presumably, able to intervene on our sensory experiences in a suitable way.

For instance (Sober 2011, §7 & §8), closing my eyes and making the visual impression, V, of waves crashing on the beach disappear does not make my auditory impression, A, of the crashing-sound go away. Similarly, shutting my ears eliminates A but not V.

Now this could be the result of a 'ham fisted' manipulation, not a proper intervention: My shutting my eyes might simultaneously eliminate and (*via* a different causal path) *create* A anew. But that is an obviously conspiratorial story, which corresponds to what causal modlers call an 'unfaithful' causal model: It features a causal connection that does not show up in the statistics. This is not an attractive objection on the part of the solipsist.

Given this purported intervention, we may infer that both V and A are common effects of a single cause: Something in the outside world, such as my actually being at the beach. However, the solipsist still has an ace up her sleeve, for she can always maintain that the common cause is *internal* to experience.

Against this, Sober (2011, 17) objects that the following inequality 'is sanctioned by frequency data', where I_t denotes my intending to go the beach at some point in time t:

$$P(A_{t'>t} \wedge V_{t'>t}|I_t) > P(A_{t'>t}|I_t)P(V_{t'>t}|I_t). \tag{1}$$

Hence, my intentions do not screen off the auditory and visual experiences from one another. Sober then also considers the possibility of an intervening, or an additional common cause variable that represents a mental state which explains the correlation. But he comes down with the following verdict on the prospects of that:

> Suppose, for example, that right after I form the intention to go to the beach that I am rendered unconscious; the next thing I know, I am either experiencing wavy visual and auditory experiences, or I am experiencing neither. When I introspect, I find no further experiences that I can cite to explain this uncanny correlation of V and A. Realists will maintain that I not only have evidence for a common cause, but also have evidence that this common cause is not one of my mental states. (Sober 2011, 17)

Hence, because the the hypothesis CC_{ext} (with obvious meaning) renders the observed correlation expected, but CC_{int} does not, my finding a screener in the external but not in my purely mental world serves as *evidence* for the existence of an external world, and so against solipsism.

3. Spurious correlations: Short digression

A potential distraction is created by the objections raised by Sober (2001) against the PCC in detail. For why should Sober even care that the solipsist has a hard time pointing to a screener if the PCC does not generically hold?

To address this question step by step, first recall that Sober (1987; 2001) invokes the (empirical) correlation between British bread prices and Venetian sea levels as a potential counterexample to the PCC. As clarified in his 2001, in response to an objection by Spirtes, Glymour, and Scheines (2000), the example is not about the *changes* in these quantities, which do not exhibit a correlation, but about the fact that, in a surprising number of years, each of the two quantities has been simultaneously greater than average.[5]

It is common to refer to such correlations as *spurious*, for they appear to imply a causal connection which is quite likely not there. However, many spurious correlations can be explained away in terms of further causal structure, such as latent common cause-variables. Alternatively, such a correlation might persist only over a narrow time interval or in an otherwise select set of event-pairs, and so maybe is not really in need of explanation.

Call an empirical correlation *transient* if only present over a narrow time-interval, and *biased* if present only in an otherwise select sample. If neither of the two obtains, call it *robust*. I believe it is perfectly in line with the causal thinking underpinning the PCC to exclude transient and biased empirical correlations from application, in virtue of their being likely coincidental or artificial. Claiming that the sea level / bread price case is not robust would be tantamount to invoking the 'unrepresentative sample'-objection considered by Meek and Glymour (1994).[6] This option is rebutted by Sober (2001, 333) as follows:

> I claim that the sample is *not* unrepresentative. Go back and draw more and more measurements

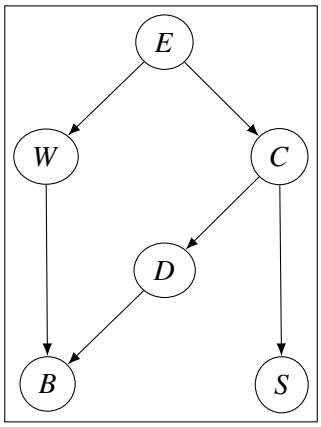

Figure 1.: Possible causal explanation of Sober's bread price / sea-level correlation

from the two centuries in question; the association of values will remain in place.

It seems that the years in which the association obtains has not been somehow consciously postselected, and the quote makes it clear that the correlation is presumably not transient.[7]

Let us assume that Sober is right. Then it is still unclear what the counterexample really establishes. For there could still be further, latent causal variables that serve to explain the correlation. For instance, consider the model depicted in Figure 1. Taking our cues from causal modlers (Spirtes, Glymour, and Scheines 2000; Pearl 2009) we retrieve the following causal independencies:

$$W \perp\!\!\!\perp \{C, D, S\}|E, \quad C \perp\!\!\!\perp W|E, \quad D \perp\!\!\!\perp \{E, W, S\}|C, \quad B \perp\!\!\!\perp \{E, C, S\}|\{D, W\}, \quad S \perp\!\!\!\perp \{D, E, W, B\}|C \tag{2}$$

The interpretation of this structure is as follows: First, Britain's (more precisely: the UK's) dominant sector, the service industry,[8] profits from *global economic growth* (E), since more people from around the world (tourists, business travelers...) can come into the UK to spend their money in, e.g., hotels and restaurants, and more foreign businesses can outsource tasks to UK service-businesses such as transport, storage, IT, finance, or scientific and technical services. Hence, there is a clear causal path from the global economy to *Britain's local wealth* (W).

Second, global economic growth is also clearly linked to *climate change* (C), which is known to be responsible for the *rising sea levels in Venice* (S), and elsewhere. Third, only if people are wealthy enough can they buy ever more expensive bread. So there is also a causal path from Britain's local wealth to the increasing *bread prices* (B). However, fourth, a known driving force behind increasing bread prices is the *dry weather in summer* (D),[9] which is itself caused by climate change. Hence there is also a direct causal link here.[10]

As we can see, the model has a sub-structure according to which Venice sea levels and British bread prices are both causally influenced by climate change. However, the model specifies additional causal structure that explains on what further factors the local bread price depends, and how these in turn relate to climate change themselves. So, while certainly not complete yet, this is not a naïve model that bluntly traces both sequences of events back to climate change without further ado.

Nevertheless, according to the Causal Markov Condition, which is generally recognized as a generalization of PCC to causal graphs, we can infer the following probabilistic independence

from the model:[11]

$$P(S \wedge B|C) = P(S|C)P(B|C) \tag{3}$$

I believe that this sensibly captures certain causal intuitions about this spurious correlation in a non-naïve way. The point, however, is not that this is necessarily the *right* causal model. Rather, the point is that this spurious correlation, like many others, can *possibly* be explained away by investigating deeper into latent causal variables.

It is a plausible dogma that any *robust* correlation that is not explained well by a hypothesis of type CE can be explained in similar ways: By a deeper analysis of the causal structure it is embedded into. Hence, that the PCC applies in all sufficiently detailed causal models that feature only robust correlations is not, or not clearly, ruled out by Sober's counterexample. However, where does that leave us regarding Sober's own assessment of the PCC and what is the relevance for the evidential argument against solipsism?

Sober (2001, 342) actually finds that:

> Although correlations need not have explanations of the sort described in [the PCC], such explanations often are plausible [...] *when they compete with separate cause explanations that predict that correlations do not exist*.

So in other words, the PCC might apply only in case a separate cause explanation that equally predicts the correlation is unavailable.

What is the separate cause explanation that predicts the correlation in the sea level / bread case? Sober (2001) does not spell this out in detail, but rather refers to 'the obvious separate cause explanation' (Sober 2001, 342). I believe that he must have in mind the explanation of rising sea levels by climate change and the explanation of bread prices by local economic factors and dry spells. However, that is not really a separate cause explanation, as we have just seen. If there was no common cause explanation, i.e., if bread prices could be traced back to purely local economic factors, I believe the association would be likely way more transient.

It is certainly *conceivable* that independent causal factors could create a robust empirical correlation that would defy the universality of PCC. For instance, this could happen if there were unconnected causes that happen to happen at the same time and set up a very similar causal etiology for two subsequent sequences of events in virtue of some similar mechanism.[12] However, something like this seems to be rarely the case, insofar as the (empirical) correlation is *robust*, and the discovery of two time sequences that exhibit remarkable correlations certainly invites for a search for deeper causal structure.

That said, the reason that Sober *does* seek for a screener in the case of V and A (wavy auditory and visual sense impressions) after having been unconscious (whatever that means for the solipsist) presumably is that there does not seem to be any sensible independent cause explanation that equally predicts the observed correlation.

On account of the above arguments, however, I believe that we should also be struck by the extraordinary *robustness* of empirical correlations regarding our own sense-experiences of different types: Confessedly, I have *never*, during my entire life span (which provides the entire 'solipsistic sample space'), had a sensation of type V without one of type A, unless intervening on one or the other in the ways indicated by Sober.[13] That holds regardless of whether I have fallen asleep during a bus ride or so.

4. Quantum solipsism

Let me here turn to a position that, so far as I know, is not really defended explicitly by anyone in the literature, but attested by critics to a number of mathematicians and mathematical physicists interested in the foundations of quantum theory (QT).[14] Call somebody a 'QBist' who takes herself to be restricted to her own, immediate sense experience as a source of evidence, much like the Cubists are restricted to events inside the cube, but in addition uses a particular mathematical calculus that corresponds to the formalism of QT (plus, maybe, some inputs from Bruno de Finetti's theory of probability). This formalism she uses to compute probabilities that she can then employ as an instruction manual for what to expect to experience in the future upon certain 'actions' (which, as far as she can tell, are just particular complex experiences). Such a QBist, if she also was to disregard any questions as to an external realty (outside the 'QBe') entirely, could by some standards be called a *quantum solipsist*.

This closeness of the writings of actual QBists, such as Fuchs, Mermin, and Schack (2014), to solipsism has been brought to the fore by Travis Norsen (2016). Here are some reasons:

> In QBism, a measurement is an action an agent takes to elicit an experience. The measurement outcome is the experience so elicited. The measurement outcome is thus *personal to the agent who takes the measurement action*. In this sense, quantum mechanics, like probability theory, is *a single user theory*.
>
> [...]
>
> According to QBism, quantum mechanics can be applied to any physical system. QBism treats all physical systems in the same way, including atoms, beam splitters, Stern-Gerlach magnets, preparation devices, measurement apparatuses, *all the way to living beings and other agents*. (Fuchs and Schack 2014, 3)

> In QBism, the only phenomenon accessible to Alice that she does not model with quantum mechanics is her own direct internal awareness of her *own private experience*. This (and *only this*) plays the role of the 'classical objects' of Landau and Lifshitz for Alice (and *only for Alice*). Her awareness of her past experience forms the basis for the beliefs on which her state assignments rest. And her probability assignments express her expectations for her future experience. (Fuchs, Mermin, and Schack 2014, 750)

However, like the Cubists, the QBists have a hero – 'Christopherus' – who declares to them that:

> We do [...] hold evidence for an independent world [...] external to ourselves precisely because we find ourselves getting unpredictable kicks (from the world) all the time. (Fuchs 2002, 11)

There is a superficial symmetry between Copernicus and Christopherus here, for both point to the explanation of patterns in experience – the correlated motion of the birds' shadows, and the repeated occurrence of 'random kicks', i.e., experiences one considers surprising or unexpected – by means of a common cause: (objects in) the external world.

However, in contrast to Copernicus, Christopherus does not really point to a correlation, but rather to a kind of (irregular, but repeated) *recalcitrance*. Moreover, Christopherus' explanation is *wholesale* rather than directed at individual objects (the birds), and so allows for 'weirder' realisms in which the relation between experience and reality is more complicated than that between the birds, the hidden mechanism, and the correlated shadows.[15]

Nevertheless, an important commonality remains, at least between Sober's *reappraisal* of Reichenbach and Fuchs's grounds for realism: In Sober's view, it is particularly the fact that my *intentions* do not always produce the desired, correlated visual and auditory experiences that provides evidence of an external world. Hence, his analysis also refers to features of recalcitrance. But the story again gets complicated. For consider the following setup (Fig. 2), a version of what Jon P. Jarrett (2009) coined a 'Mermin contraption' (after Mermin 1981).

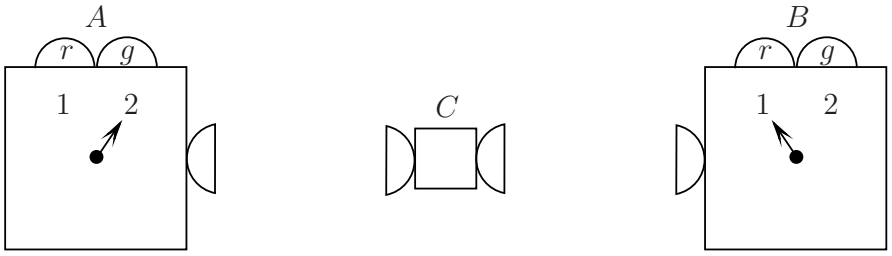

Figure 2.: Variation of the Mermin contraption

This contraption is an arrangement of three devices, A, B, and C, where A and B are placed on two diametrically opposite sides of C. A and B each have two settings, $1, 2$, and a red (r) and green (g) light on top of them. At certain points in time, lights flash on each of the two devices simultaneously, and the nob on each of A and B may or may not switch automatically just shortly before the lights flash, with no discernible pattern or correlation between both nobs.[16] Now imagine a scientist writing down a protocol as depicted in Tab. 1. After watching a very

A	B	A	B	$\cdots$
$1, r$	$1, r$	$1, g$	$2, r$	$\cdots$
$2, g$	$1, g$	$2, g$	$2, g$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 1.: Hypothetical protocol for observations on the Mermin contraption

long sequence of some thousands of flashes and writing down settings and lights flashed, she notices that the frequency of joint occurrences of r and g on A and B, respectively, is somewhere near $(2 + \sqrt{2})/8$ among the runs in which A was set to 1 and B was set to either 1 or 2, and equally when A was set to 2 and B to 1, but somewhere near $(2 - \sqrt{2})/8$ when both are set to 2. However, for each setting, the occurrences of either r or g on either A or B settle down around ½. Consequently, she notices an empirical correlation:

$$\begin{aligned} &f(A = g \wedge B = r|a \wedge b) \approx f(A = r \wedge B = g|a \wedge b) < f(A = g|a)f(B = r|b) \\ \approx &f(A = r|a)f(B = g|b) \qquad \text{for } a = b = 2, \end{aligned} \tag{4}$$

$$\begin{aligned} &f(A = g \wedge B = r|a \wedge b) \approx f(A = r \wedge B = g|a \wedge b) > f(A = g|a)f(B = r|b) \\ \approx &f(A = r|a)f(B = g|b) \qquad \text{else}, \end{aligned} \tag{5}$$

where a denotes A's setting and b denotes B's.

The scientist of course realizes immediately that if these visual experiences would be caused by some state of an external reality, λ, – maybe featuring, among other things, the states of two invisible particles being emitted from C – this would predict the observed correlations between the flashes. I.e., without crunching too many numbers, she can assume that a probabilistic model

$$P(A \wedge B|a \wedge b \wedge \lambda) = P(A|a \wedge \lambda)P(B|b \wedge \lambda) \tag{CC_λ}$$

would *predict* a correlation, and so $P((4) \wedge (5)|CC_\lambda) > P((4) \wedge (5)|\bar{C}_\lambda)$. Finding nothing inside her own mind that could otherwise be cited to predict the correlated visual impressions,

she concludes that this remarkable correlation between the flashing light bulbs should count as evidence of a mind-independent reality if anything does.

However, as is well known, our scientist soon runs into trouble. If she assumes that λ is causally unrelated to the setting (i.e., $P(\lambda|a \wedge b) = P(\lambda)$) and uses a numerical variable that maps g to $+1$ and r to -1 to compute expectation values, she easily derives the infamous CHSH inequality (Clauser et al. 1969),

$$\begin{aligned} &|\sum_{x,y} xyP(A = x \wedge B = y|1_A \wedge 1_B) + \sum_{x,y} xyP(A = x \wedge B = y|1_A \wedge 2_B) + \\ &+ \sum_{x,y} xyP(A = x \wedge B = y|2_A \wedge 1_B) - \sum_{x,y} xyP(A = x \wedge B = y|2_A \wedge 2_B)| \leq 2, \end{aligned} \tag{6}$$

with $x, y \in \{-1, +1\}$, and obvious meanings for the numbers conditioned on. But using the values approximated by her noted frequencies in place of the probabilities, she finds that this yields

$$2\sqrt{2} \leq 2, \tag{7}$$

which is clearly false.

What is our scientist to do? She could infer that the two flashes are related as cause and effect (CE_λ), rather than as effects of a common cause. However, that option is undermined by the observation that there is no conceivable intervention she can undertake to alter the statisics in A by manipulating B (Friederich 2015, 132), at bottom because the situation at both devices, A and B, is perfectly symmetric (Boge 2016).

She could also come up with weirder causal stories, such as the latent variable λ being influenced by, or influencing, the settings (an option sometimes called 'superdeterminism'), or maybe even that there are causal influences going back in time. However, all these options are by now known to be possible only on pains of stipulating causal relations that have no manifestation in the frequency data (Wood and Spekkens 2015),[17] and our scientist, intelligent as she is, will soon figure this out.

It seems as though the scientist can only explain the situation causally by appeal to some unfaithfulness, or causal-statistical 'fine tuning', just as involved in the ham-fisted intervention of Sect. 2. Clearly, embracing a causal account in which causation is 'naturally ham-fisted' just to be able to point to a screener amounts to embracing a conspiracy theory in as much is as this was the case for the solipsist with respect to her defense against interventionism. This is not an attractive option for our realist scientist.

Given the considerations of Sect. 3, the realist could now insist that there need not be a screener, λ, there at all. For the correlation could be coincidental, *qua* being caused individually by causal mechanisms inside A and B respectively that set up similar etiologies. However, this latter option was already contested in the sea-level bread case, and it just seems ludicrous to claim in the present example—maybe at best like the mystical kind of explanation debated under the header 'synchronicity' or 'parallelism' by Wolfgang Pauli and Carl Gustav Jung (e.g. Miller 2010). I do not believe any realist would seriously want to suggest as much.

Moreover, repeating the experiment over and over, the scientist will find these correlations to be highly non-transient; and relativization to runs in which the joint settings take on particular values seems like a fully legitimate 'post selection', not the creation of a somehow biased sample. The correlation, in other words, is extremely robust.

The major difference to the sea-level / bread price case is that we there do not (or at least not obviously) run into *contradictions with experience* when looking into the possibility of latent

causal variables that might explain the correlation. But in the case of the Mermin contraption, we surely do.

Now this is as devastating for the realist *QBist* as it is for any realist. For QBists argue that their reference to the experiences of individual agents *especially* restores the compatibility between QT and relativity: Imagine the Mermin contraption with Bob sitting at B and Alice at A, where the two are separated by a spacelike distance, i.e., so far away that nothing that is slower than, or equally fast as, light can propagate between them during a single run of the experiment. Then any causal explanation would have to violate the relativistic speed limit on top of the causal problems already indicated.[18] However:

> Quantum mechanics, in the QBist interpretation, cannot assign correlations, spooky or otherwise, to space-like separated events, since they cannot be experienced by any single agent. Quantum mechanics is thus *explicitly* local in the QBist interpretation. (Fuchs, Mermin, and Schack 2014, 750–1)

In particular, when Alice and Bob meet and compare their lists, they will elicit experiences in one another that exhibit the pattern described earlier. This is the moment at which the correlated 'measurement outcomes' are created, for the interaction between both is a measurement for the QBist. Yet this hardly makes the list-like experiences in these 'measurements' *understandable*:

> The [QBist] story explains why both [Bob] and Alice *expect*, with degree of belief one, to find anticorrelated spins, but [...] the [QBist] story does not explain why the measured spins *are in fact* anticorrelated. (Earman 2019, 418)

Unless subject to a kind of 'inverse McGruk effect', in which believing is seeing, it seems as though the realist QBist, Alice, should consider her interaction with *Bob* as the cause of the observed correlations between table entries: Intervening on her visual impressions of Bob by refusing to meet up with him will certainly also make the list of results correlated to hers disappear. However, holding Bob causally responsible seems ludicrously incompatible with any other conceptions Alice will certainly entertain about him, based on her past experience. Unless, of course, Alice takes the whole situation to be a hoax concocted, and never resolved, by Bob and the team of scientists setting up the contraption—something akin to another conspiracy theory.

Our would-be quantum solipsist appears to be gaining an edge here: She can simply compute the desired expectations using the (rotation invariant) singlet state $2^{-1/2}(|\uparrow\rangle\,|\downarrow\rangle - |\downarrow\rangle\,|\uparrow\rangle)$ on spin operators representing observables with misalignments by angles $\pi/4$ and $(3\pi)/4$ (which arise from choosing settings on A and B respectively), to find her predictions match the empirical observation quite exactly. Furthermore, she can use the contradiction between the causal prediction and the empirical evidence as grounds for rejecting the *causal closure* of experience.

Returning to the single-scientist Mermin contraption, the situation now turns out to be related to my wavy visual and auditory impressions after my falling asleep on the bus as follows: Against the background of there only being 'unfaithful' cause-effect explanations, my inability to find a screener within experience and the ability to find one without in the beach case was counted as evidence that 'It's the external world that is doing the work, stupid'! However, against the background that there are *no* faithful causal models that respect screening off in the Mermin case, my inability to find a screener *at all* without contradicting observations now seems to undermine that evidence.

Hence, the solipsist is at liberty to point out that sometimes searching for a screener – even when faced with robust empirical correlations that cannot be explained as coincidentally arising from independent causes – leads to contradictions with frequency data. Accordingly, she is not required to take the correlations between V and A after what she would describe as

'falling asleep' as evidence of the existence of an external world. Experience simply has these causal gaps, such as my sitting in the bus and suddenly lying on the beach, maybe next to what I should describe as 'my friends'—who, on the realist account, surely must have dragged me there.

5. Conclusion

Where does all that leave us on Christopherus' and Copernicus' claims to the existence of an external world outside the cube (or 'QBe')? The resemblance between Copernicus' situation and popular level accounts of quantum correlations, as rehearsed in the previous section, is remarkable, and it is surprising that this has so far been scarcely addressed.[19]

It is telling, moreover, that Sober (2011, 9) likens the cube to Wesley Salmon's (1984, 162) example of two student essays resembling each other because they are both the result of plagiarism from a common source: Matching results in student exams or like situations have a history of being used to explicate what quantum correlations *cannot* be like (Schrödinger 1935, 845–6; Maudlin 2011, 13 ff.). Hence, it is equally surprising that the potential connection between arguments about realism and experimental violations of inequalities such as the CHSH-one is pretty unpopular among philosophers.

As the penultimate section shows, Copernicus's reasons for embracing a hypothesis of the form CC are undermined by the fact that he soon runs into contradictions if the correlation is like that between the lights on the Mermin contraption. There is no cube-external common cause Copernicus can cite without either abandoning screening off or embracing a fine-tuned conspiracy theory. However, that does not at all address *Christopherus'* path from recalcitrance to reality, so what are the prospects of that?

First note that actually any QBist is a solipsist of the *epistemological* or at least *methodological* variety: Treating one's own conscious experiences as the sole basis for knowledge and everything else as constructed from that is, by definition, a methodological stance to which the label 'solipsism' properly attaches.

> [T]he choice of an autopsychological basis amounts merely to an application of the form and method of solipsism, but not to an acknowledgment of its central thesis, [so] we may describe our position as *methodological solipsism*. (Carnap 2003, 102)

However, notice that thus only a QBist confining her *ontological commitments* to her own experiences would be a 'proper' solipsist, one of the 'bad kind'.

Given that she is a 'harmless' kind of solipsist, the QBist can, at some level of construction, indeed point to recalcitrance as evidence that there is something outside of her. Recalcitrance has also been used by proper metaphysicians as evidence for, say, the modal force of the laws of nature (Hüttemann 2014), or even for the independence of fundamental interactions (Bartels 2019). So why not use it at a lower level of metaphysical theorizing?

This leaves us in the following position regarding the 'unscreenable' correlations exhibited by the Mermin contraption: If there is an external world (EW) that is not fully controllable by willful interventions, then, in general, the occurrence of surprises and uncontrollable events (OS) is to be expected. This is not so on the complement hypothesis ($E\bar{W}$). The surprising correlations exhibited by the Mermin contraption, which are so radically out of our control (remember: no possibility to intervene locally on the remote outcome which is strongly correlated but, like the local one, apparently random), hence serve as proper evidence *for* the existence of an external world – even if not for particular objects or events in that world *causing* these correlated experiences.

I submit that even this argument will never convince the proper, 'bad', metaphysical solipsist

though. For she can always maintain that, in addition to causal gaps, experience simply has a willful and a non-willful part, and that's that. Against Reichenbach, I hence agree with Carnap to the extent that the dispute between realism and solipsism (or even intersubjective idealism), though maybe not purely verbal, cannot be settled by reference to any possible experience. These are just rival metaphysical systems.

Furthermore, consider Reichenbach's own approach to QT, in which he distinguishes *phenomena*, i.e., something that 'can be 'directly' verified by such devices as the Geiger counter, a photographic film, a Wilson cloud chamber, etc.', from *interphenomena*, i.e., something 'introduced by inferential chains of a much more complicated sort [and] constructed in the form of an *interpolation* within the world of phenomena' (Reichenbach 1944, 20–1). Reichenbach then goes on to discuss the usual wave-particle story as providing two possible, complementary interpolations in the double-slit experiment. However, given the problems induced by the assumption of a screener λ, it seems that in the case of the Mermin contraption, we are short of *any* interpolating story.

The bottom line is that the combination of an empiricist position in the quantum foundations, such as QBism, with standard principles of causation (no fine-tuning, screening off for robust empirical correlations without direct causation) does *not* lead to metaphysical *solipsism*. But we could still take it to lead to a rejection of full-blown *scientific* realism: If we have no representational means for telling an interpolating causal story about the robust correlation exhibited by the Mermin contraption – if we cannot always postulate Reichenbachian interphenomena – then maybe this indicates that scientific theorizing ultimately does not equip us with the right semantic tools for referring properly to that external world whose existence we take to be evidenced by our recalcitrant experiences.

Acknowledgements Some research for this paper was done during my employment with the research unit *The Epistemology of the Large Hadron Collider*, funded by the German Research Foundation (DFG; grant FOR 2063). I thank Alex Gebharter for some help with causal graphs, and an anonymous referee for helpful comments.

References

Bartels, Andreas. 2019. "Explaining the modal force of natural laws." *European Journal for Philosophy of Science* 9 (1): 6.

Boge, Florian. 2016. "Book Review: "Simon Friederich: Interpreting Quantum Theory: A Therapeutic Approach"." *Erkenntnis* 82 (2): 443–449.

Carnap, R. 2003. *The Logical Structure of the World And Pseudoproblems in Philosophy*. Open Court.

Cavalcanti, Eric G. 2018. "Classical causal models for bell and kochen-specker inequality violations require fine-tuning." *Physical Review X* 8 (2): 021018.

Clauser, John F., Michael A. Horne, Abner Shimony, and Richard A. Holt. 1969. "Proposed Experiment to test Local Hidden-Variable Theories." *Physical Review Letters* 23 (15): 880–884.

Earman, John. 2019. "Quantum Bayesianism Assessed." *The Monist* 102 (4): 403–423.

Forster, Malcolm, and Elliott Sober. 2004. "Why Likelihood." In *The Nature of Scientific Evidence*, edited by Mark L. Taper and Subhash R. Lele, 153–190. University of Chicago Press.

Friederich, S. 2015. *Interpreting Quantum Theory: A Therapeutic Approach*. Basingstoke: Palgrave Macmillan.

Fuchs, Christopher A. 2002. "The anti-vaxjo interpretation of quantum mechanics." *arXiv preprint quant-ph/0204146* .

Fuchs, Christopher A, N David Mermin, and Rüdiger Schack. 2014. "An introduction to QBism with an application to the locality of quantum mechanics." *American Journal of Physics* 82 (8): 749–754.

Fuchs, Christopher A, and Rüdiger Schack. 2014. "QBism and the Greeks: why a quantum state does not represent an element of physical reality." *Physica Scripta* 90 (1): 015104.

Hitchcock, Christopher, and Miklós Rédei. 2020. "Reichenbach's Common Cause Principle." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, spring 2020 ed. Metaphysics Research Lab, Stanford University.

Hüttemann, Andreas. 2014. "Scientific practice and necessary connections." *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia* 29 (1): 29–39.

Ismael, Jenann. 2020. "What Entangelment Might be Telling us: Space, Quantum Mechanics, and Bohm's Fish Tank." In *The Foundation of Reality: Fundamentality, Space, and Time*, edited by David Glick, George Darby, and Anna Marmodoro. Oxford University Press.

Jarrett, Jon P. 2009. "On the separability of physical systems." In *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle*, 105–124. Springer.

Maudlin, Tim. 2011. *Quantum Non-Locality and Relativity. Metaphysical Intimations of Modern Physics*. 3rd ed. Malden, Oxford: Wiley-Blackwell.

Meek, Christopher, and Clark Glymour. 1994. "Conditioning and intervening." *The British journal for the philosophy of science* 45 (4): 1001–1021.

Mermin, N David. 1981. "Quantum mysteries for anyone." *The Journal of Philosophy* 78 (7): 397–408.

Miller, A.I. 2010. *137: Jung, Pauli, and the Pursuit of a Scientific Obsession*. W. W. Norton.

Näger, Paul M. 2016. ""The causal problem of entanglement"." *Synthese* 193 (4): 1127–1155.

Norsen, T. 2016. "Quantum Solipsism and Nonlocality." In *Quantum Nonlocality and Reality: 50 Years of Bell's Theorem*, edited by Mary Bell and Shan Gao, 204–237. Cambridge: Cambridge University Press.

Pearl, JC, and EG Cavalcanti. 2019. "Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios." *arXiv preprint arXiv:1909.05434* .

Pearl, Judea. 2009. *Causality. Models, Reasoning, and Inference*. 2nd ed. Cambridge, New York: Cambridge University Press.

Putnam, Hilary. 1981. *Reason, Truth and History*. Cambrdige, New York: Cambridge University Press.

Reichenbach, Hans. 1938. *Experience and prediction: An analysis of the foundations and the structure of knowledge*. University of Chicago press.

Reichenbach, Hans. 1944. *Philosophic Foundations of Quantum Mechanics*. Mineloa, New York: Dover Publications, Inc.

Reichenbach, Hans. 1965. *The Direction of Time*. Mineloa, New York: Dover Publications, Inc. Edited by Maria Reichenbach.

Royall, R. M. 1997. *Statistical Evidence: A Likelihood Paradigm*. Chapman and Hall.

Salmon, W.C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Schrödinger, Erwin. 1935. "Die gegenwärtige Situation in der Quantenmechanik." *Naturwissenschaften* 23 (50): 844–849.

Sober, Elliott. 1987. "The principle of the common cause." In *Probability and causality: Essays in Honor of Wesley Salmon*, edited by J. Fetzer, 211–228. Springer.

Sober, Elliott. 2001. "Venetian sea levels, British bread prices, and the principle of the common cause." *The British Journal for the Philosophy of Science* 52 (2): 331–346.

Sober, Elliott. 2011. "Reichenbach's cubical universe and the problem of the external world." *Synthese* 181 (1): 3–21.

Spirtes, P., C.N. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*. Cambridge, Massachusetts: MIT Press.

Wood, Christopher J, and Robert W Spekkens. 2015. ""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.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford, New York: Oxford University Press.

Wroński, Leszek. 2014. *Reichenbach's Paradise. Constructing the Realm of Probabilistic Common 'Causes'*. Warsaw, Berlin: De Gruyter Open.

Yule, G Udny. 1926. "Why do we sometimes get nonsense-correlations between Time-Series? A study

in sampling and the nature of time-series." *Journal of the royal statistical society* 89 (1): 1–63.

Notes

1. By 'logico-analytic independence', I mean that the co-occurrence of X and Y is neither fixed by a meaning postulate, such as the co-occurrence of being a bachelor and being unmarried, nor by logical implication, such as the co-occurrence of heads with heads-or-tails. Neither of these would be in need of explanation other than by pointing out the logical or conceptual relations between them.

2. I deliberately leave it open whether X and Y are event types or tokens, for reasons raised in Fn. 7.

3. Sober uses 'correlation' to exclusively refer to a theoretical property associated with probabilities, and 'association' to refer to the observed regular coincidence of two events of distinct types. I find 'empirical correlation' the more natural terminology, and I will allow myself to sometimes suppress the 'empirical' when the context allows this.

4. I will here not question the notion that likelihood provides a sensible measure for evidence in this way. A detailed defense is offered, for instance, by Forster and Sober (2004).

5. As a referee has rightly pointed out to me, however, this already makes the example considerably less forceful.

6. The debate actually reaches back further, to Yule (1926).

7. It is unclear to me whether this is equally true of the other counterexample discussed by Sober (2001, 335). Sober considers the correlation, over a lifetime, between the heights of specimens from two species that are homo*plaseous* rather than homo*logous*. That is, the two species have a common trait that 'evolved independently in the two lineages' and is not 'inherited unmodified from a common ancestor' (Sober 2001, 335). To give Sober's own example: robins and sparrows have inherited their wings from a common ancestor (they are homologous) whereas robins and bats have not (they are homoplaseous). However, finding specimens from two species which have extremely similar height development requires a contrived selection of individuals. If a fair sample is used, I suspect that the empirical correlation is rather nullified. On the other hand, if expressed as a relation on the type level, such as 'the average relative height of specimen of species x is almost exactly the same as that of specimen of species y after such and such a fraction of their life span', there is a perfectly acceptable type-level common cause: similar selective environmental pressures. Sober's (2001, 339) claim that 'the principle of the common cause[...] is about common *token* causes' I believe to be incorrect as it stands (see, for instance, the discussion of Reichenbach's theater troupe in Hitchcock and Rédei 2020).

8. See https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/fivefactsabouttheukservicesector/2016-09-29.

9. See, e.g., https://www.bbc.com/news/science-environment-53921121

10. Wroński (2014, 28) actually identifies very similar causal variables, but never extends his considerations to asking whether melting sub-polar glaciers and deteriorating crop levels are causally related, and so does not offer a common cause explanation.

11. In particular, the last independence implies that $P(S \wedge D \wedge E \wedge W \wedge B|C) = P(S|C)P(E \wedge D \wedge W \wedge B|C)$. Marginalizing both sides for D, E and W, we have the desired result. It also follows directly from the 'decomposition axiom' assumed in some causal modeling frameworks (Pearl 2009, 11).

12. I believe this is what Sober is getting at with his homoplasy-example, but my reservations expressed in Fn. 7 remain.

13. I do not count, say, impressions of waves on a muted TV here; these are too clearly distinct from 'original Vs'.

14. I have in mind, of course, especially Christopher Fuchs, N. David Mermin, and Rüdiger Schack.

15. For instance, consider Hilary Putnam's (1981, 63) reading of Kant in this connection, according to which 'any judgment about external or internal objects (physical things or mental entities) says that the noumenal world as a whole is such that this is the description that a rational being (one with our rational nature) given the information available to a being with our sense organs (a being with our sensible nature) would construct. [...] *There is not even a one-to-one correspondence between things-for-us and things in themselves*.'

16. We can imagine this to be brought about by further devices hooked up to A and B respectively.

17. Note that this is also true of proposals beloved by many philosophers, such as Bohmian approaches (cf. Wood and Spekkens 2015, 20).

18. Paul M. Näger (2016) is one of the few to distinguish the causal aspects of this type of scenario explicitly from its implications for the relativity theories. Note that by now, these impossibility results based on purely causal considerations have been generalized also to the impossibility of classical causal models of *contextuality* (Cavalcanti 2018; Pearl and Cavalcanti 2019).

19. A notable exception is Jennan Ismael (2020). However, she does not really address the problem that *any* common cause explanation

of quantum correlations involves fine-tuning (Näger 2016; Wood and Spekkens 2015).