Open Science Framework DOI:10.31219/osf.io/bsnh7

The rule of conditional probability is valid in quantum theory

P.G.L. Porta Mana
Kavli Institute, Trondheim <pgl@portamana.org>

13 July 2020; updated 15 July 2020

In a recent manuscript, Gelman & Yao (2020) claim that "the usual rules of conditional probability fail in the quantum realm" and purport to support that statement with an Young-slit example. Such a statement is false. In the present note I recall some literature in quantum theory that shows why it is false, sum up the incorrect reasoning underlying their example, and also correct some wrong or imprecise statements about the quantum physics in that example.

In fact, an opposite statement can be made, from two different perspectives, which will be discussed in the rest of this note:

- The example given in that manuscript confirms, rather than invalidate, the probability rules. The probability calculus shows that a particular relation between probabilities *cannot a priori* be assumed to be an equality or an inequality. In the quantum example it turns out to be an inequality, thus confirming what the probability calculus says.
- But actually the same inequality can be shown to appear in very non-quantum examples, such as drawing from an urn. Thus there is nothing peculiar to quantum theory in this matter.

Also, let me point out at the outset that the rule of conditional probability (and the other two rules, sum and negation) are in fact routinely used in quantum theory, with full validity, especially in problems of state "retrodiction" and measurement reconstruction (Jones 1991; Slater 1995; de Muynck 2002 chs 7, 8; Barnett et al. 2003; Ziman et al. 2006; D'Ariano et al. 2004; see Månsson et al. 2006 § 1 for many further references), for example to infer the state of a quantum laser given its output through different optical apparatus (Leonhardt 1997).

Similar incorrect claims with similar examples have appeared before in the quantum literature. Bernard O. Koopman (of the Pitman-Koopman theorem for sufficient statistics, Koopman 1936) discussed the falsity of such claims already in 1957. The Introduction in his work is very clear:

> Ever since the advent of modern quantum mechanics in the late 1920's, the idea has been prevalent that the classical laws of probability cease, in some sense, to be valid in the new theory. More or less explicit statements to this effect have been made in large number and by many of the most eminent workers in the new physics [...]. Some authors have even gone farther and stated that the formal structure of logic must be altered to conform to the terms of reference of quantum physics [...].
>
> Such a thesis is surprising, to say the least, to anyone holding more or less conventional views regarding the positions of logic, probability, and experimental science: many of us have been apt – perhaps too naively – to assume that experiments can lead to conclusions only when worked up by means of logic and probability, whose laws seem to be on a different level from those of physical science.
>
> The primary object of this presentation is to show that the thesis in question is entirely without validity and is the product of a confused view of the laws of probability.

A claim somewhat similar to Gelman & Yao's, with a similar supporting example, was made in a work by Brukner & Zeilinger (2001); although their focus was on an alleged inconsistency of some properties of the Shannon entropy in quantum theory.

The fallacy in their reasoning, which is similar to Gelman & Yao's, rests in the neglect of the experimental setup, leading either to an incorrect calculation of conditional probabilities, or to the incorrect claim that the probability calculus yields an equality, whereas it actually does not. The incorrect reasoning was discussed at length by Porta Mana (2004) through a step-by-step analysis and calculation. This work also showed, through simple examples (ibid. § IV), that the same incorrect statements can be obtained *with completely non-quantum systems*, such as drawing from an urn, if the setup is neglected. Let us start with such a counter-example.

A non-quantum counter-example Consider an urn with one *B*lue and one *R*ed ball. There are two possible drawing setups:

- D_{a} With replacement for blue, without replacement for red. That is, if blue is drawn, it is put back before the next draw (and the urn is shaken); if red is drawn, it is thrown away before the next draw.
- D_{b} With replacement for red, without replacement for blue.

These two setups are obviously mutually exclusive.

We can easily find the unconditional probability for blue at the second draw in the setup D_a:

$$\mathrm{P}(B_2 \mid D_a) = \tfrac{3}{4} \;. \tag{1}$$

Note that these can be intuitively found by simple enumeration, à la Boole, with a "possible worlds" diagram. Out of four possible worlds, half of which has blue at the first draw, and the other half has red, we see that three worlds have blue at the second draw.

The conditional probabilities for blue at the second draw, conditional on the first draw, are also easily found:

$$\mathrm{P}(B_2 \mid B_1 \land D_a) = \tfrac{1}{2} \qquad \mathrm{P}(B_2 \mid R_1 \land D_a) = 1 \;. \tag{2}$$

We find that

$$\mathrm{P}(B_2 \mid D_a) = \mathrm{P}(B_2 \mid B_1 \land D_a)\, \mathrm{P}(B_1 \mid D_a) + \mathrm{P}(B_2 \mid R_1 \land D_a)\, \mathrm{P}(R_1 \mid D_a) \;, \tag{3}$$

which is just the rule of conditional probability. It is in fact just the systematization and generalization of the intuitive "possible worlds" reasoning done above.

Now consider the setup D_b. We easily find

$$\mathrm{P}(B_2 \mid D_b) = \tfrac{1}{4} \;. \tag{4}$$

$$\mathrm{P}(B_2 \mid B_1 \land D_b) = 0 \qquad \mathrm{P}(B_2 \mid R_1 \land D_b) = \tfrac{1}{2} \;, \tag{5}$$

$$\mathrm{P}(B_2 \mid D_b) = \mathrm{P}(B_2 \mid B_1 \land D_b)\, \mathrm{P}(B_1 \mid D_b) + \mathrm{P}(B_2 \mid R_1 \land D_b)\, \mathrm{P}(R_1 \mid D_b) \;. \tag{6}$$

Now compare the unconditional probability for blue at the second draw in the setup D_a, with the probabilities for blue at the second draw conditional on the first in the setup D_b:

$$\mathrm{P}(B_2 \mid D_a) \neq \mathrm{P}(B_2 \mid B_1 \land D_b)\, \mathrm{P}(B_1 \mid D_b) + \mathrm{P}(B_2 \mid R_1 \land D_b)\, \mathrm{P}(R_1 \mid D_b) \;. \tag{7}$$

This inequality is not surprising – we are comparing different setugs. It is *not* an instance of the conditional-probability rule. In fact the probability calculus has nothing to say about the relation between the left side and right side, which are conditional on different statements or, if you like, pertain to two different sample spaces.

You can call the inequality above "interference" if you like; for further examples see Kirkpatrick (2003a; 2003b) and Porta Mana (2004 § IV).

Now consider another pair of drawing setups: setup D_c, with replacement for both colours; and setup D_d, without replacement for either colour. You can easily find that

$$\mathrm{P}(B_2 \mid D_c) = \mathrm{P}(B_2 \mid B_1 \land D_c)\,\mathrm{P}(B_1 \mid D_c) + \mathrm{P}(B_2 \mid R_1 \land D_c)\,\mathrm{P}(R_1 \mid D_c)\ , \tag{8}$$

$$\mathrm{P}(B_2 \mid D_d) = \mathrm{P}(B_2 \mid B_1 \land D_d)\,\mathrm{P}(B_1 \mid D_d) + \mathrm{P}(B_2 \mid R_1 \land D_d)\,\mathrm{P}(R_1 \mid D_d)\ , \tag{9}$$

$$\mathrm{P}(B_2 \mid D_c) \mathbin{\mathbf{=}} \mathrm{P}(B_2 \mid B_1 \land D_d)\,\mathrm{P}(B_1 \mid D_d) + \mathrm{P}(B_2 \mid R_1 \land D_d)\,\mathrm{P}(R_1 \mid D_d)\ . \tag{10}$$

The first two equalities are expressions of the conditional-probability rule. The last, however, is *not*, despite being an equality. It is simply a peculiar equality contingent on the two specific setups.

The probability calculus thus correctly handles situations such as (7) and (10).

Strictly speaking it is wrong to use the expression 'B_2' for all these setups, because 'B_2' in each setup denotes a different statement (or random variable) than in the others. Just like "it rains (on 2020-07-14T09:00+0200 in Trondheim)" is different from "it rains (on 2019-01-20T18:00+0200 in Rome)". I should have used different symbols. The explicit presence of '$D_{\dots}$', which represents given information, luckily avoided any ambiguities. But if in our formulae we omit the notation of the setup *and* we use the same notation for actually different statements or random variables, then we're in for trouble and for incorrect applications of the probability rules.

The inequality (7) is what Gelman & Yao (2020 p. 2) complain about, but in the context of a pair of quantum setups. I do not see how one can complain about it, or claim inconsistencies. It is obviously correct even from an intuitive analysis of the two setups, and the probability calculus, correctly, also leads to it. The probability calculus correctly leads also to the equality (10). As already said, given two mutually exclusive setups, the probability calculus a priori neither commits to an equality nor to an inequality.

I will now show that the simple example above is conceptually actually quite close to the quantum experiment mentioned by Gelman & Yao – and also correct some statements they make about that experiment.

The closeness is especially clear from the experimental and mathematical developments of quantum theory of the past 40 years (at the very least), as the literature cited below shows.

The quantum two-slit experiments The basic argument of Gelman & Yao is that, in a given setup of the quantum two-slit experiment, we have a specific probability distribution for the appearance of an emulsion or excitation on some point of the screen. We can call this a “screen detection”, but please keep in mind that in so doing we are adding an extra interpretation that modern quantum theory does not actually commit to – see references below. In a different experimental setup we have conditional probabilities for screen detection conditional on slit detection. Now, the probability in the first setup is not equal to the combination of the conditional probabilities of the second setup.

But this is exactly what happens in the urn example above, eq. (7). Also in this quantum case it is *not* a violation of the conditional probability rule – if anything it is a confirmation.

To see the analogy more clearly, let me present some additional facts from quantum theory.

The experimental setup without detectors at the slits, the two setups with one slit closed, and the setup with slit detectors are actually limit cases of a continuum of experimental setups (Wootters & Zurek 1979; Banaszek et al. 2013 for a recent review and further references see). In the general case, such a setup has “noisy” slit detectors, with a degree of noise $q \in [1/2, 1]$ that can be chosen in the setup. These noise degrees are of course mutually exclusive, so these setups are mutually exclusive.

The slit detector is called “noisy” in the following sense. Let us call y the detection position on the screen, X_1 the statement that detection occurs at slit #1, and X_2 at slit #2 (you can translate to random-variable notation if you prefer). We can prepare the electromagnetic field (quantum state S_1) in such a way that, in the setup with non-noisy slit detectors, detector #1 always fires; that is, statement X_1 has probability 1 and statement X_2 has probability 0 in this setup. We can also prepare a state S_2 such that the opposite holds: X_1 has zero probability, and X_2 unit probability.

If we prepare these states, but use the setup with noisy detectors of degree q – denote this setup by D_q – then we have probability q

that slit detector #1 fires for state S_1, and that #2 fires for state S_2 (the probabilities of the other slit are $1 - q$). That is, the statements X_1 and X_2 have probabilities q and $1 - q$. (We should actually not be using the same X symbol for all such setups, because the sample spaces are different; see the discussion for the urn example). The setup with non-noisy detectors is the limit case $q = 1$. In the limit, fully noisy case $q = 1/2$ there is basically no relation between the light states and the firing of the slit detectors; that is, we are always fully uncertain as to what detector would fire, no matter how the light state is prepared.

In each setup D_q (and given the light state S) we also have the conditional probability distribution $\mathrm{p}(y \mid D_q, S)$ for detection at y on the screen, and the conditional probability distributions $\mathrm{p}(y \mid X, D_q, S)$ for detection at y on the screen, conditional on detection X at the slits. We have

$$\mathrm{p}(y \mid D_q, S) = \\ \mathrm{p}(y \mid X_1, D_q, S)\, \mathrm{p}(X_1 \mid D_q, S) + \mathrm{p}(y \mid X_2, D_q, S)\, \mathrm{p}(X_2 \mid D_q, S)\,. \quad (11)$$

This is an instance of the conditional-probability rule. This equality also holds for long-run frequencies (see point (iii) below). Note that such frequencies are experimentally observed. I would like you to convince yourself, though, that the *equality* above (not the specific values of the frequencies) is not an experimental fact, since it rests on the very definition of conditional frequency.

The conditional and unconditional distributions above will of course be different depending on the setup D_q and the light state S.

There are a couple of interesting experimental facts about this collection of setups:

First, *both the conditional and unconditional* probability (or long-run frequency) distributions for the screen detection y generally have an oscillatory profile (depending on the specific light state, of course, which we assume fixed), typical of interference (Wootters & Zurek 1979; Banaszek et al. 2013; see also Chiao et al. 1995 for further variations). The oscillatory character is maximal for $q = 1/2$ and decreases as $q \to 1$; for $q = 1$ it is gone completely.

Second, the unconditional (frequency) distribution that we observe in the setup without slit detectors is experimentally equal to that observed in the setup $D_{1/2}$ with fully noisy slit detectors.

Third, one conditional distribution observed in the setup with one slit closed is experimentally equal to one in the setup D_1 with perfect slit detectors. Here we must be careful, because there is no slit detection in the second setup; rather, we speak of appearance or non-appearance at the screen, and in the latter case no conditional distribution is defined.

Note that the equalities in the last two cases should not be expected a priori, because the setups are physically different. Of course one can look for physical, "hidden variables" explanations of such equalities. Experimental quantum optics simply acknowledges the fact that two setups are equivalent for such detection purposes.

It should be noted how this range of cases in statistically analogous to the urn examples and setups previously discussed. In each setup, the rule of conditional probability holds (and in the quantum case we can have distributions, conditional and unconditional, with oscillatory profiles). Across different setups probability theory says that such a rule cannot be applied; and indeed we find inequalities across some setups and equalities across others, both in the quantum and non-quantum case, see eqs (7), (10).

It is also possible to consider situations where part of the setup, such as slit width or presence or absence of detectors, is unknown. This is similar, in the urn example above, to not knowing whether D_a or D_b applies. In this case one can make inferences by giving the probability for each setup, e.g. $\mathrm{P}(D_\mathrm{a})$, and applying the conditional-probability rule. The same rule is indeed applied in the analogous quantum situation (e.g. Ziman et al. 2006). Again, no violations of the probability rules in the quantum realm. ***

Further remarks and curiosities about quantum two-slits experiments I would like to mention a couple more experimental facts – which are, besides, statistically very interesting – to correct some statements by Gelman & Yao in relation to the two-slit experiment.

(**i**) It *does* matter whether many photons are sent at once, or one at a time (cf. Gelman & Yao 2020 § 2 point 1); as well as their wavelength, temporal spread, and so on. These details are part of the specification of the light state S mentioned above, and lead to different probabilities distributions of detection at the screen.

For example, in some setups and for some states we can have a detection probability density $p(y_1)$ for the first photon, and a *different* density for the second photon $p(y_2 \mid y_1)$, conditional on the detection of the first – both being different from the long-run joint density of detections $f(y)$. See e.g. the phenomena of higher-order coherence, bunching, anti-bunching, and so on[1].

The rules of the probability calculus also apply in such situations. We can infer, for example, the position of the first photon detection given the second from $p(y_1 \mid y_2) \propto p(y_2 \mid y_1)\, p(y_1)$.

(ii) The details about the light source and the setup are not "latent variables": they specify the quantum state of light and the measurement performed on it. They are like the initial and boundary conditions necessary to the specification of the behaviour of any physical system.

(iii) In view of point (i) above, it is important not to conflate the *probabilities* for single-photon detections and the *frequency* distribution of a long-run of such detections (Gelman & Yao 2020 § 2, seem to conflate the two).

I may add that the idea and parlance of "photons passing through slits" are used today only out of tradition; maybe a little poetically. The technical parlance, as routinely used in quantum-optics labs for example (Leonhardt 1997; Bachor & Ralph 2004), has a different underlying picture. The basic "system" in a quantum-optics experiment are not photons, but the modes of the field-configuration operator (note that this is not yet Quantum ElectroDynamics). "Photon numbers" denote the discrete outcomes of a specific energy-measurement operator; "photon states" denote specific states of the field operators. As another example, "entanglement" is strictly speaking not among photons, but among modes of the field operator (van Enk 2003). Several quantum physicists indeed oppose the idea and parlance of "photons", owing to the confusion it leads to. Lamb (of the Lamb shift, Lamb & Retherford 1947) wrote in 1995:

> the author does not like the use of the word "photon", which dates from 1926. In his view, there is no such thing as a photon. Only a comedy of errors

---

1 e.g. Mandel & Wolf 1965; Morgan & Mandel 1966; Paul 1982; Jacobson et al. 1995; and textbooks such as Loudon 2000; Mandel & Wolf 2008; Scully & Zubairy 2001; Bachor & Ralph 2004; Walls & Milburn 1994.

> and historical accidents led to its popularity among physicists and optical scientists.

Wald (1994) warns:

> standard treatments of quantum field theory in flat spacetime rely heavily on Poincaré symmetry (usually entering the analysis implicitly via plane-wave expansions) and interpret the theory primarily in terms of a notion of "particles". Neither Poincaré (or other) symmetry nor a useful notion of "particles" exists in a general, curved spacetime, so a number of the familiar tools and concepts of field theory must be "unlearned" in order to have a clear grasp of quantum field theory in curved spacetime. [p. ix] [. . .] the notion of "particles" plays no fundamental role either in the formulation or interpretation of the theory. [p. 2]

A summary of the modern formalism of quantum theory It may be useful to give a summary of how modern quantum theory works. A quantum system is defined by its sets of possible states and possible measurements. A state ρ is represented by an Hermitean, unit-trace matrix $\boldsymbol{\rho}$ (which satisfies additional constraints), called ' density matrix '. States traditionally represented by kets, $|\psi\rangle$, are just special cases of density matrices. A measurement setup M is represented by a set of Hermitean matrices $\{\boldsymbol{M}_r\}$ adding up to the identity matrix, of the same order as the density matrices, They are called ' positive-operator-valued measures '. Traditional von Neumann projection operators $\{|\phi_r\rangle\langle\phi_r|\}$ are just special cases. Each matrix $\boldsymbol{M}_r$ is associated with an outcome $\boldsymbol{r}$ of the measurement. These outcomes are mutually exclusive. An outcome can actually represent a combination of simpler outcomes, $\boldsymbol{r} \equiv (x, y, z, \dots)$, such as two intensities at two detectors.

The probability of observing outcome $\boldsymbol{r} \equiv (x, y, \dots)$ given the measurement setup M and the state S is encoded in the trace-product of the respective matrices:

$$\mathrm{p}(x, y, \dots \mid M \land S) \equiv \operatorname{tr}(\boldsymbol{M}_{x,y,\dots}\, \boldsymbol{\rho}) , \tag{12}$$

forming a probability distribution. Such probabilities come from repeated measurement experiences in the same experimental conditions – we could invoke de Finetti's theorem here, and some quantum physicists indeed do (Caves et al. 2002; van Enk & Fuchs 2002; Fuchs et al. 2004).

Once the probability distribution above is given, we can use the full-fledged probability calculus. We can for example sum (or integrate) over the detector outcomes $y, \dots$, obtaining the marginal probabilities for the detector outcomes x:

$$\mathrm{p}(x \mid M \land S) \equiv \mathrm{tr}\big[(\textstyle\sum_{y,\dots} \boldsymbol{M}_{x,y,\dots})\ \boldsymbol{\rho}\big]\ , \tag{13}$$

obviously represented by the matrices $\{\sum_{y,\dots} \boldsymbol{M}_{x,y,\dots}\}$.

obviously encoded by the matrix

Once the probabilities ab

First, in the experiments with only one slit open or both slits open, the outcome space is $\{\text{'no event'}\} \cup \mathbf{R}$, because either an emulsion is produced at some point on the screen, or none is produced.

After all, we do not expect that marginal probabilities from, say, a drawing-without-replacement urn setup should be obtainable from the joint distribution of a drawing-with-replacement setup, or of Pólya drawings. These setups are mutually exclusive. A random variable of one of them is not the same random variable of another. If you thought that you could consistently combine probabilities from such different urn-drawing setups and find that you actually cannot, well, too bad for you. The probability calculus, in fact, makes clear at the outset that the probabilities of these setups cannot generally be combined. It is thus somewhat funny that one ends up blaming the probability calculus, which makes a clear distinction, for one's neglect of that distinction.

Likewise, measurement setups in quantum theory – and in many classical-physics situations – are generally mutually exclusive.

This kind of incorrect claims about

I refer to the work just cited for the full analysis and counterexamples. Here I summarize the basic fallacy with a simpler counterexample.

The basic fallacy is the confusion of the probability conditional with a temporal ordering.

Take the conditional probability $\mathrm{P}(A \mid B)$, where the statement or event A refers to a time t_2, and B to a time t_1 that precedes t_2. This probability is related to the reverse conditional $\mathrm{P}(B \mid A)$ by

$$\mathrm{P}(A \mid B)\ \mathrm{P}(B) = \mathrm{P}(B \mid A)\ \mathrm{P}(A) = \mathrm{P}(A \land B)\ . \tag{14}$$

It goes without saying that *the statements or events A and B must be the same on both sides of each equation*. In particular, in the conditional $\mathrm{P}(B \mid A)$

the statement B still refers to the time t_1, and A to the time t_2, with t_1 preceding t_2. We can represent these times explicitly and rewrite (14) as

$$\mathrm{P}(A_2 \mid B_1)\,\mathrm{P}(B_1) = \mathrm{P}(B_1 \mid A_2)\,\mathrm{P}(A_2) = \mathrm{P}(A_2 \land B_1)\ . \tag{15}$$

The fallacy is to think that, in calculating $\mathrm{P}(B \mid A)$, we should now ensure that B refers to time t_2, and A to time t_1, *swapping the times*. But these would be *different* events, not the original events. In symbols, we would be calculating $\mathrm{P}(B_2 \mid A_1)$, which is different from $\mathrm{P}(B_1 \mid A_2)$, to which the conditional-probability rule (14) refers.

Probability theory has nothing to say, a priori, about the relation between $\mathrm{P}(B_1 \mid A_2)$ and $\mathrm{P}(B_2 \mid A_1)$. They are two logically different situations. A simple example can illustrate this point.

Consider an urn with *R*ed and *B*lue balls. The drawing scheme is with replacement for blue, and without replacement for red. That is, if a blue ball is drawn, it is put back in the urn (and the urn shaken) before the next draw. If a red ball is drawn, it is not put back before the next draw.

The urn initially has one blue and one red ball. The probabilities for the first draw are straightforward:

$$\mathrm{P}(B_1) = \tfrac{1}{2} \qquad \mathrm{P}(R_1) = \tfrac{1}{2}\ , \tag{16}$$

as are the conditional probabilities for the second draw, conditional on the first:

$$\mathrm{P}(B_2 \mid B_1) = \tfrac{1}{2} \qquad\qquad \mathrm{P}(R_2 \mid R_1) = 0 \tag{17a}$$

$$\mathrm{P}(R_2 \mid B_1) = \tfrac{1}{2} \qquad\qquad \mathrm{P}(B_2 \mid R_1) = 1\ . \tag{17b}$$

These probabilities can be visualized with the "equiprobable worlds" diagram of fig. 1. In half of the worlds the first draw yields red; in the other half, blue, according to (16). In all worlds with red first, the second must yield blue. In half of the worlds with blue first, the second yields blue; and in the other half, red; as for eqs (17).

Let us now calculate the conditional probabilities for the first draw conditional on the second (imagine the first draw was hidden from you, and upon seeing the second you're asked to guess what the first

Figure 1 "Equiprobable worlds" diagram

$$\begin{array}{lcccc} \text{1st draw} & \mathrm{R} & \mathrm{R} & \mathrm{B} & \mathrm{B} \\ & \downarrow & \downarrow & \downarrow & \downarrow \\ \text{2nd draw} & \mathrm{B} & \mathrm{B} & \mathrm{B} & \mathrm{R} \end{array}$$

was). Using the conditional-probability rule (15) in the form of Bayes's theorem,

$$\mathrm{P}(X_1 \mid Y_2) = \frac{\mathrm{P}(Y_2 \mid X_1)\,\mathrm{P}(X_1)}{\sum_X \mathrm{P}(Y_2 \mid X_1)\,\mathrm{P}(X_1)} , \tag{18}$$

we obtain

$$\mathrm{P}(B_1 \mid B_2) = \tfrac{1}{3} \qquad \mathrm{P}(R_1 \mid R_2) = 0 \tag{19a}$$

$$\mathrm{P}(R_1 \mid B_2) = \tfrac{2}{3} \qquad \mathrm{P}(B_1 \mid R_2) = 1 . \tag{19b}$$

These conditional probabilities are intuitively correct, as can be checked by simple enumeration of the possible cases in fig. 1. For example, among all three worlds that have blue at the second draw, two of them have red at the first; hence $\mathrm{P}(R_1 \mid B_2) = 2/3$. Also, if we have red at the second draw, then logically red cannot have been drawn at the first, otherwise there would not have been any red left; hence blue must have been drawn at the first: $\mathrm{P}(B_1 \mid R_2) = 1$.

The following two relations between the conditional probabilities (17b) and (19b) are relevant to our discussion:

$$\mathrm{P}(R_2 \mid B_1) \neq \mathrm{P}(R_1 \mid B_2) \qquad \mathrm{P}(B_2 \mid R_1) = \mathrm{P}(B_1 \mid R_2) . \tag{20}$$

As we see, the probability-calculus does not prescribe the equality nor the inequality between probabilities for events of similar kind but at different times (which therefore are *not* the same event). Any relations of this kind will depend on the specific situation. You can check, for example, that in a scheme of drawing with replacement for both blue and red, we would obtain two equalities in place of (20). In a scheme of Pólya draws for blue (blue is returned plus an additional blue) and replacement draws for red, instead, we would obtain two inequalities in place of (20).

Something analogous (see Porta Mana 2004 for an exact parallel see) happens in inferences about quantum measurements. The conditional probability for the result of measurement R made at time t_2, given the result of measurement B made at time t_1, is not necessarily the same as that for measurement R made at time t_1, given measurement B at time t_2. The time ordering of measurements is extremely important in quantum theory (as it is in the urn example above). In particular, if our goal is to actually *retrodict* the result of a previous measurement or state from the information gained in a subsequent measurement, we must be very careful not to confuse the conditional probabilities. The times of the two measurements cannot be swapped.

It should be noted that the conditional-probability rule (14) is in fact routinely used in quantum theory in problems of state retrodiction and measurement reconstruction (Jones 1991; Slater 1995; de Muynck 2002 chs 7, 8; Ziman et al. 2006; D'Ariano et al. 2004; see Månsson et al. 2006 § 1 for many further references), for example to infer the state of a quantum laser given its output through different optical apparatus (Leonhardt 1997).

Bibliography

("de *X*" is listed under D, "van *X*" under V, and so on, regardless of national conventions.)

Bachor, H.-A., Ralph, T. C. (2004): *A Guide to Experiments in Quantum Optics*, 2nd rev. and enl. ed. (Wiley-VCH, Weinheim). First publ. 1998.

Banaszek, K., Horodecki, P., Karpiński, M., Radzewicz, C. (2013): *Quantum mechanical which-way experiment with an internal degree of freedom*. Nat. Comm. **4**, 2594.

Barnett, S. M., Jeffers, J., Pegg, D. T. (2003): *Retrodictive quantum optics*. In: Bigelow, Eberly, Stroud, Walmsley (2003), 87–94.

Bigelow, N. P., Eberly, J. H., Stroud, C. R., Walmsley, I. A., eds. (2003): *Coherence and Quantum Optics VIII*. (Springer, New York).

Brukner, Č., Zeilinger, A. (2001): *Conceptual inadequacy of the Shannon information in quantum measurements*. Phys. Rev. A **63**, 022113. See also Porta Mana (2004).

Caves, C. M., Fuchs, C. A., Schack, R. (2002): *Unknown quantum states: the quantum de Finetti representation*. J. Math. Phys. $\mathbf{43}^9$, 4537–4559.

Chiao, R. Y., Kwia, P. G., Steinberg, A. M. (1995): *Quantum non-locality in two-photon experiments at Berkeley*. Quant. Semiclass. Opt. $\mathbf{7}^3$, 259–278.

D'Ariano, G. M., Maccone, L., Lo Presti, P. (2004): *Quantum calibration of measurement instrumentation*. Phys. Rev. Lett. **93**, 250407.

de Muynck, W. M. (2002): *Foundations of Quantum Mechanics, an Empiricist Approach*. (Kluwer, Dordrecht).

Fuchs, C. A., Schack, R., Scudo, P. F. (2004): *De Finetti representation theorem for quantum-process tomography*. Phys. Rev. A **69**, 062305.

Gelman, A., Yao, Y. (2020): *Holes in Bayesian statistics*. arXiv:2002.06467.

Jacobson, J., Björk, G., Chuang, I., Yamamoto, Y. (1995): *Photonic de Broglie waves*. Phys. Rev. Lett. **74**24, 4835–4838.

Jones, K. R. W. (1991): *Principles of quantum inference*. Ann. of Phys. **207**1, 140–170.

Kirkpatrick, K. A. (2003a): *"Quantal" behavior in classical probability*. Found. Phys. Lett. **16**3, 199–224. First publ. 2001.

— (2003b): *Classical three-box "paradox"*. J. Phys. A **36**17, 4891–4900. First publ. 2002. See also Ravon, Vaidman (2007) and Kirkpatrick (2007).

— (2007): *Reply to 'The three-box paradox revisited' by T. Ravon and L. Vaidman*. J. Phys. A **40**11, 2883–2890. See Kirkpatrick (2003b) and Ravon, Vaidman (2007).

Koopman, B. O. (1936): *On distributions admitting a sufficient statistic*. Trans. Am. Math. Soc. **39**3, 399–409.

— (1957): *Quantum theory and the foundations of probability*. In: MacColl (1957), 97–102.

Lamb Jr., W. E. (1995): *Anti-photon*. Appl. Phys. B **60**2–3, 77–84.

Lamb Jr., W. E., Retherford, R. C. (1947): *Fine structure of the hydrogen atom by a microwave method*. Phys. Rev. **72**3, 241–243.

Leonhardt, U. (1997): *Measuring the Quantum State of Light*. (Cambridge University Press, Cambridge).

Loudon, R. (2000): *The Quantum Theory of Light*, 3rd ed. (Oxford University Press, Oxford). First publ. 1973.

MacColl, L. A., ed. (1957): *Applied Probability*. (McGraw-Hill, New York).

Mandel, L., Wolf, E. (1965): *Coherence properties of optical fields*. Rev. Mod. Phys. **37**2, 231–287.

— (2008): *Optical coherence and quantum optics*, repr. with corrections. (Cambridge University Press, Cambridge). First publ. 1995.

Månsson, A., Porta Mana, P. G. L., Björk, G. (2006): *Numerical Bayesian state assignment for a three-level quantum system. I. Absolute-frequency data; constant and Gaussian-like priors*. arXiv:quant-ph/0612105.

Morgan, B. L., Mandel, L. (1966): *Measurement of photon bunching in a thermal light beam*. Phys. Rev. Lett. **16**22, 1012–1015.

Paul, H. (1982): *Photon antibunching*. Rev. Mod. Phys. **54**4, 1061–1102.

Porta Mana, P. G. L. (2004): *Consistency of the Shannon entropy in quantum experiments*. Phys. Rev. A **69**6, 062108. Rev. version at arXiv:quant-ph/0302049.

Ravon, T., Vaidman, L. (2007): *The three-box paradox revisited*. J. Phys. A **40**11, 2873–2882. See Kirkpatrick (2003b). Unfortunately the arguments of this work are marred by vagueness and contradictions. A reply is given in Kirkpatrick (2007).

Scully, M. O., Zubairy, M. S. (2001): *Quantum Optics*. (Cambridge University Press, Cambridge). First publ. 1997.

Slater, P. B. (1995): *Reformulation for arbitrary mixed states of Jones' Bayes estimation of pure states*. Physica A **214**4, 584–604.

van Enk, S. J. (2003): *Entanglement of electromagnetic fields*. Phys. Rev. A **67**, 022303.

van Enk, S. J., Fuchs, C. A. (2002): *Quantum state of an ideal propagating laser field*. Phys. Rev. Lett. **88**, 027902.

Wald, R. M. (1994): *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. (University of Chicago Press, Chicago).

Walls, D. F., Milburn, G. J. (1994): *Quantum Optics*. (Springer, Berlin).

Wootters, W. K., Zurek, W. H. (1979): *Complementarity in the double-slit experiment: Quantum nonseparability and a quantitative statement of Bohr's principle.* Phys. Rev. D **19**², 473–484.
Ziman, M., Plesch, M., Bužek, V. (2006): *Reconstruction of superoperators from incomplete measurements*. Found. Phys. **36**¹, 127–156. First publ. 2004.