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 14 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 such statement with an example. Such a statement is false. I would like to recall some literature that shows why it is false, and to sum up the fallacy underlying their example.

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, 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 (Koopman 1936; Pitman 1936 of the Pitman-Koopman theorem for sufficient statistics,) discussed the falsity of such claims already in 1957. The introduction to his work (Koopman 1957) 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 similar to Gelman & Yao's, with a similar supporting example, was made in a work by Brukner & Zeilinger (2001), somewhat famous in the quantum community; although their focus was on an alleged inconsistency of some properties of the Shannon entropy in quantum theory.

The fallacy in the reasoning of Brukner & Zeilinger and Gelman & Yao rests in the neglect of the experimental setup, leading to an incorrect calculation of conditional probabilities. Such fallacy was exposed and 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 classical systems*, such as drawing from an urn, if the setup is neglected. Here is a simple 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.

The two setups are obviously mutually exclusive.

We can easily calculate what is the unconditional probability for blue at the second draw in the setup D_{a}:

$$\mathrm{P}(B_2 \mid D_{\mathrm{a}}) = \tfrac{3}{4} \,, \tag{1}$$

and the conditional probabilities for blue at the second draw, conditional on the first draw, in the setup D_{b}:

$$\mathrm{P}(B_2 \mid B_1 \land D_{\mathrm{b}}) = 0 \qquad \mathrm{P}(B_2 \mid R_1 \land D_{\mathrm{b}}) = \tfrac{1}{2} \,. \tag{2}$$

We find that

$$\mathrm{P}(B_2 \mid D_{\mathrm{a}}) \neq \mathrm{P}(B_2 \mid B_1 \land D_{\mathrm{b}})\, \mathrm{P}(B_1 \mid D_{\mathrm{b}}) + \mathrm{P}(B_2 \mid R_1 \land D_{\mathrm{b}})\, \mathrm{P}(R_1 \mid D_{\mathrm{b}}) \,. \tag{3}$$

This inequality is no surprising. And it does not contradict the rule of conditional probability, because that rule is supposed to be used within the same probability space. You can call the inequality above "interference" if you like; for further examples see Kirkpatrick (2003a; 2003b) and Porta Mana (2004 § IV).

Note that, had we considered a drawing setup D_{c} with replacement for both colours, and a setup D_{d} without replacement for either colour, we would have found

$$\mathrm{P}(B_2 \mid D_{\mathrm{c}}) = \mathrm{P}(B_2 \mid B_1 \land D_{\mathrm{d}})\, \mathrm{P}(B_1 \mid D_{\mathrm{d}}) + \mathrm{P}(B_2 \mid R_1 \land D_{\mathrm{d}})\, \mathrm{P}(R_1 \mid D_{\mathrm{d}})\ . \quad (4)$$

The equality above, however, is *not* and expression of the conditional-probability rule, because the probability spaces are different. It is simply a peculiar equality contingent on the two specific setups. The probability calculus handles correctly situations such as (3) or (4).

In fact, strictly speaking it is wrong to use the expression 'B_2' for all these setups, because 'B_2' in the one setup denotes a different statement (or random variable) than in another. 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 example above stresses the importance of the probability space – even in "non-quantum" situations. But it is not meant as a parallel of Gelman & Yao's (2020 § 2) example. In fact their example has important differences and their analysis of it is incorrect from the point of view of quantum theory. Here are the main points:

(i) It *does* matter whether many photons are sent at once, or one at a time, as well as their wavelength, temporal spread, and so on. These details lead to different probabilities distributions of detection at the screen[1]. More precisely, the spatio-temporal dependence of the optic-field operator (which define the photon state) must be specified. Note that these details are not "latent variables": they are the initial and boundary conditions that define the physical system; they correspond to the different drawing setups in the example above. The rules of probability apply seamlessly in each case.

[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.

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 to now knowing whether D_{a} or D_{b} above 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 can be applied in the analogous quantum situation (see e.g. Barnett et al. 2003). Again, no violations of the rules in the quantum realm.

(ii) Owing to the point above, it is important not to conflate the *probability* for a single-photon detection and the *frequency* distribution of a long-run of such detections, as Gelman & Yao instead do. For example, in some setups we can have a detection probability density $\mathrm{p}(y_1)$ for the first photon, and a *different* density for the second photon $\mathrm{p}(y_2 \mid y_1)$, 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 and bunching in the references above). The rules of the probability calculus also apply in such situations. For example, we can guess the position of the first photon detection given the second from $\mathrm{p}(y_1 \mid y_2) \propto \mathrm{p}(y_2 \mid y_1)\, \mathrm{p}(y_1)$.

(iii)

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

$$\mathrm{P}(A \mid B)\,\mathrm{P}(B) = \mathrm{P}(B \mid A)\,\mathrm{P}(A) = \mathrm{P}(A \land B)\;. \tag{5}$$

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 (5) as

$$\mathrm{P}(A_2 \mid B_1)\,\mathrm{P}(B_1) = \mathrm{P}(B_1 \mid A_2)\,\mathrm{P}(A_2) = \mathrm{P}(A_2 \land B_1)\;. \tag{6}$$

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 (5) 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:

$$\mathrm{P}(B_1) = \tfrac{1}{2} \qquad \mathrm{P}(R_1) = \tfrac{1}{2}\;, \tag{7}$$

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

$$\mathrm{P}(B_2 \mid B_1) = \tfrac{1}{2} \qquad\qquad \mathrm{P}(R_2 \mid R_1) = 0 \tag{8a}$$

$$\mathrm{P}(R_2 \mid B_1) = \tfrac{1}{2} \qquad\qquad \mathrm{P}(B_2 \mid R_1) = 1\;. \tag{8b}$$

Figure 1 "Equiprobable worlds" diagram

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

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 (7). 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 (8).

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 was). Using the conditional-probability rule (6) in the form of Bayes's theorem,

$$\mathrm{P}(X_1 \mid Y_2) = \frac{\mathrm{P}(Y_2 \mid X_1)\,\mathrm{P}(X_1)}{\sum_X \mathrm{P}(Y_2 \mid X_1)\,\mathrm{P}(X_1)}\;, \tag{9}$$

we obtain

$$\mathrm{P}(B_1 \mid B_2) = \tfrac{1}{3} \qquad\qquad \mathrm{P}(R_1 \mid R_2) = 0 \tag{10a}$$

$$\mathrm{P}(R_1 \mid B_2) = \tfrac{2}{3} \qquad\qquad \mathrm{P}(B_1 \mid R_2) = 1\;. \tag{10b}$$

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 (8b) and (10b) are relevant to our discussion:

$$\mathrm{P}(R_2 \mid B_1) \boldsymbol{\neq} \mathrm{P}(R_1 \mid B_2) \qquad \mathrm{P}(B_2 \mid R_1) \boldsymbol{=} \mathrm{P}(B_1 \mid R_2)\;. \tag{11}$$

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 (11). 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 (11).

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 (5) 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.

Barnett, S. M., Jeffers, J., Pegg, D. T. (2003): *Retrodictive quantum optics*. In: **bigelowetal2003**, 87–94.

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

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).

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.

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.

Pitman, E. J. G. (1936): *Sufficient statistics and intrinsic accuracy*. Math. Proc. Camb. Phil. Soc. **32**^{4}, 567–579.

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.

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

Ziman, M., Plesch, M., Bužek, V. (2006): *Reconstruction of superoperators from incomplete measurements*. Found. Phys. **36**^{1}, 127–156. First publ. 2004.