Conditional Effects, Observables and Instruments

Stan Gudder

Department of Mathematics, University of Denver, Denver, Colorado, USA. E-mail: sgudder@du.edu

Editors: ***James F. Glazebrook & Danko D. Georgiev***

Article history: Submitted on February 2, 2024; Accepted on February 5, 2024; Published on February 6, 2024.

We begin with a study of operations and the effects they measure. We define the probability that an effect a occurs when the system is in a state ρ by $P_\rho(a) = \mathrm{Tr}(\rho a)$. If $P_\rho(a) \neq 0$ and $\mathcal{I}$ is an operation that measures a, we define the conditional probability of an effect b given a relative to $\mathcal{I}$ by $P_\rho(b \mid a) = \mathrm{Tr}[\mathcal{I}(\rho)b]/P_\rho(a)$. We characterize when Bayes' quantum second rule $P_\rho(b \mid a) = \frac{P_\rho(b)}{P_\rho(a)} P_\rho(a \mid b)$ holds. We then consider Lüders and Holevo operations. We next discuss instruments and the observables they measure. If A and B are observables and an instrument $\mathcal{I}$ measures A, we define the observable B conditioned on A relative to $\mathcal{I}$ and denote it by $(B \mid A)$. Using these concepts, we introduce Bayes' quantum first rule. We observe that this is the same as the classical Bayes' first rule, except it depends on the instrument used to measure A. We then extend this to Bayes' quantum first rule for expectations. We show that two observables B and C are jointly commuting if and only if there exists an atomic observable A such that $B = (B \mid A)$ and $C = (C \mid A)$. We next obtain a general uncertainty principle for conditioned observables. Finally, we discuss observable conditioned quantum entropies. The theory is illustrated with many examples.

Quanta 2024; 13: 1–10.

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY-3.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1 Effects and Operations

It is sometimes stated that all probabilities in quantum mechanics are conditional probabilities and there is some sense to this statement. Underlying most quantum experiments or observations, there are basic observables A_i and calculations are performed according to the outcomes obtained for A_i. For example, many quantum experiments consist of scattered particles and these involve the positions A_i of the various particles. The probabilities for another observable is thus conditioned by the outcomes of A_i.

According to complexity, there is a hierarchy of quantum measurements. The simplest are effects, the next are observables and finally we have instruments. Each of these types of measurements can be conditioned in a systematic way. They can even be conditioned among each other.

Let H be a finite-dimensional complex Hilbert space representing a quantum system. The set of linear operators on H is denoted by $\mathcal{L}(H)$ and the set of self-adjoint operators is denoted by $\mathcal{L}_S(H)$.

A *state* is a positive operator $\rho \in \mathcal{L}_S(H)$ with trace $\mathrm{Tr}(\rho) = 1$ and the set of states is denoted by $\mathcal{S}(H)$. States describe the conditions of the system and are employed to compute probabilities of measurement outcomes.

An operator a satisfying $0 \leq a \leq I$ is called an *effect*. An effect represents a two outcome *yes*-*no* experiment that either occurs or does not occur [1–5]. We represent the set of effects by $\mathcal{E}(H)$. If $a \in \mathcal{E}(H)$ occurs, then its *complement* $a' = I - a$ does not occur.

An *operation* is a completely positive linear map $\mathcal{I}\colon \mathcal{L}(H) \to \mathcal{L}(H)$ such that $\mathrm{Tr}[\mathcal{I}(\rho)] \leq \mathrm{Tr}(\rho)$ for all

$\rho \in \mathcal{S}(H)$ [1–5]. An operation that satisfies $\mathrm{Tr}[\mathcal{I}(\rho)] = \mathrm{Tr}(\rho)$ for all $\rho \in \mathcal{S}(H)$ is called a *channel* [3, 5, 6]. Any operation $\mathcal{I}$ has a *Kraus decomposition* $\mathcal{I}(A) = \sum_i K_i A K_i^*$ where $K_i \in \mathcal{L}(H)$ and $\sum_i K_i^* K_i \leq I$. We call K_i, $i = 1, 2, \ldots, n$, *Kraus operators* for $\mathcal{I}$ [4]. When $\mathcal{I}$ is a channel, we have $\sum_i K_i^* K_i = I$.

Corresponding to an operation $\mathcal{I}$ we have the *dual operation* [7–9] $\mathcal{I}^* \colon \mathcal{L}(H) \to \mathcal{L}(H)$ where $\mathcal{I}^*$ is linear and satisfies $\mathrm{Tr}[\mathcal{I}(\rho)A] = \mathrm{Tr}[\rho\mathcal{I}^*(A)]$ for all $\rho \in \mathcal{S}(H)$, $A \in \mathcal{L}(H)$. If $\mathcal{I}$ has Kraus decomposition $\mathcal{I}(A) = \sum K_i A K_i^*$, then $\mathcal{I}^*(A) = \sum K_i^* A K_i$ for all $A \in \mathcal{L}(H)$. If $\mathcal{I}$ is a channel then $\mathcal{I}^*(I) = I$. It is easy to check that if $\mathcal{I}$ is an operation, then $\mathcal{I}^* \colon \mathcal{E}(H) \to \mathcal{E}(H)$ and $\mathcal{I}^*(a) \leq a$ for all $a \in \mathcal{E}(H)$. We say that an operation $\mathcal{I}$ *measures an effect* a if $\mathrm{Tr}[\mathcal{I}(\rho)] = \mathrm{Tr}(\rho a)$ for all $\rho \in \mathcal{I}(H)$ [7, 8, 10]. We interpret $P_\rho(a) = \mathrm{Tr}(\rho a)$ as the probability that a occurs when the system is in state ρ. It follows that an operation measures a unique effect. However, as we shall see, there are many operations that measure an effect a. If $\mathcal{I}$ measures a, then

$$\mathrm{Tr}[\rho\mathcal{I}^*(I)] = \mathrm{Tr}[\mathcal{I}(\rho)] = \mathrm{Tr}(\rho a)$$

for every $\rho \in \mathcal{S}(H)$. Hence, $\mathcal{I}$ measures a if and only if $\mathcal{I}^*(I) = a$.

If $a, b \in \mathcal{E}(H)$ we write $a \perp b$ if $a + b \in \mathcal{E}(H)$. If $a, b \in \mathcal{E}(H)$ and $\mathcal{I}$ measures a, we define the *$\mathcal{I}$-sequential product of a then b* by $a[\mathcal{I}]b = \mathcal{I}^*(b)$. It is easy to check that $a[\mathcal{I}]b \leq a$, if $b \perp c$ then $a[\mathcal{I}](b + c) = a[\mathcal{I}]b + a[\mathcal{I}]c$ and $a[\mathcal{I}]I = a$ [7, 8]. An effect a is *sharp* if a is a projection and a is *atomic* if a is a one-dimensional projection.

Lemma 1. Let $\mathcal{I}$ be an operation that measures $a \in \mathcal{E}(H)$. (i) $\mathcal{I}^*(b) \leq a$ for all $b \in \mathcal{E}(H)$. (ii) If a is sharp, then $\mathcal{I}^*(b)a = a\mathcal{I}^*(b)$ for all $b \in \mathcal{E}(H)$. (iii) If a is atomic, then $\mathcal{I}^*(b) = \lambda a$ for some $\lambda \in [0, 1]$.

Proof. (i) Since

$$\mathcal{I}^*(b) + \mathcal{I}^*(b') = \mathcal{I}^*(b + b') = \mathcal{I}^*(I) = a$$

we conclude that $\mathcal{I}^*(b) \leq a$.

(ii) Since $\mathcal{I}^*(b) \leq a$, $\mathcal{I}^*(b)$ and a coexist [3]. Then a being sharp implies that $\mathcal{I}^*(b)a = a\mathcal{I}^*(b)$.

(iii) If a is atomic and $\mathcal{I}^*(b) \leq a$, we have that $\mathcal{I}^*(b) = \lambda a$ for some $\lambda \in [0, 1]$ [3]. □

If $P_\rho(a) \neq 0$ and $\mathcal{I}$ measures a, we define the *conditional probability of b given a relative to $\mathcal{I}$* by [9]

$$P_\rho(b \mid a) = \frac{\mathrm{Tr}[\mathcal{I}(\rho)b]}{P_\rho(a)}$$

We then have

$$\begin{aligned} P_\rho(b \mid a) &= \frac{\mathrm{Tr}[\mathcal{I}(\rho)b]}{\mathrm{Tr}[\mathcal{I}(\rho)]} = \frac{\mathrm{Tr}[\rho\mathcal{I}^*(b)]}{\mathrm{Tr}(\rho a)} \\ &= \frac{\mathrm{Tr}(\rho a[\mathcal{I}]b)}{\mathrm{Tr}(\rho a)} = \frac{P_\rho(a[\mathcal{I}]b)}{P_\rho(a)} \\ &= \frac{P_{\mathcal{I}(\rho)}(b)}{P_\rho(a)} \end{aligned}$$

We have that $b \mapsto P_\rho(b \mid a)$ is a probability distribution in the sense that $P_\rho(I \mid a) = 1$ and if $b_i \in \mathcal{E}(H)$ with $b_1 + b_2 + \cdots + b_n \leq I$, then

$$P_\rho\left(\sum_{i=1}^n b_i \mid a\right) = \sum_{i=1}^n P_\rho(b_i \mid a)$$

We also see that $\widetilde{\rho} = \mathcal{I}(\rho)/P_\rho(a)$ is a state called the *updated state* for $\mathcal{I}$ and we have

$$P_\rho(b \mid a) = \mathrm{Tr}(\widetilde{\rho}b) = P_{\widetilde{\rho}}(b)$$

Thus, to find $P_\rho(b \mid a)$ we first measure a using $\mathcal{I}$, update the state to $\widetilde{\rho}$ and then compute the probability of b using $\widetilde{\rho}$. If $\mathcal{I}$ and $\mathcal{J}$ are operations, we define the *sequential product of $\mathcal{I}$ then $\mathcal{J}$* as the operation given by $(\mathcal{I} \circ \mathcal{J})(\rho) = \mathcal{J}(\mathcal{I}(\rho))$ for all $\rho \in \mathcal{S}(H)$ [7, 8]. In a similar way we define $(\mathcal{I}^* \circ \mathcal{J}^*)(A) = \mathcal{J}^*(\mathcal{I}^*(A))$.

Theorem 2. Let $\mathcal{I}$ and $\mathcal{J}$ be operations. (i) $(\mathcal{I} \circ \mathcal{J})^* = (\mathcal{J}^* \circ \mathcal{I}^*)$. (ii) If $\mathcal{I}$ measures a and $\mathcal{J}$ measures b, then $\mathcal{I} \circ \mathcal{J}$ measures $a[\mathcal{I}]b$. (iii) If a is measured with $\mathcal{I}$, b with $\mathcal{J}$ and $a[\mathcal{I}]b$ with $\mathcal{I} \circ \mathcal{J}$, then

$$a[\mathcal{I}](b[\mathcal{J}]c) = (a[\mathcal{I}]b)[\mathcal{I} \circ \mathcal{J}]c$$

(iv) For all $\rho \in \mathcal{S}(H)$ we have

$$\mathrm{Tr}(\rho a)P_\rho(b[\mathcal{J}]c \mid a) = \mathrm{Tr}(\rho a[\mathcal{I}]b)P_\rho(c \mid a[\mathcal{I}]b)$$

Proof. (i) For all $\rho \in \mathcal{S}(H)$ $A \in \mathcal{L}(H)$ we obtain

$$\begin{aligned} \mathrm{Tr}[\rho(\mathcal{I} \circ \mathcal{J})^*(A)] &= \mathrm{Tr}[(\mathcal{I} \circ \mathcal{J})(\rho)A] = \mathrm{Tr}[\mathcal{J}(\mathcal{I}(\rho))A] \\ &= \mathrm{Tr}[\mathcal{I}(\rho)\mathcal{J}^*(A)] = \mathrm{Tr}[\rho\mathcal{I}^*(\mathcal{J}^*(A))] \\ &= \mathrm{Tr}[\rho(\mathcal{J}^* \circ \mathcal{I}^*)(A)] \end{aligned}$$

It follows that $(\mathcal{I} \circ \mathcal{J})^* = \mathcal{J}^* \circ \mathcal{I}^*$.

(ii) Since

$$\begin{aligned} \mathrm{Tr}[\mathcal{I} \circ \mathcal{J}(\rho)] &= \mathrm{Tr}[\mathcal{J}(\mathcal{I}(\rho))] = \mathrm{Tr}[\mathcal{I}(\rho)b] \\ &= \mathrm{Tr}[\rho\mathcal{I}^*(b)] = \mathrm{Tr}(\rho a[\mathcal{I}]b) \end{aligned}$$

it follows that $\mathcal{I} \circ \mathcal{J}$ measures $a[\mathcal{I}]b$.

(iii) Applying (i) gives

$$\begin{aligned} a[\mathcal{I}](b[\mathcal{J}]c) &= a[\mathcal{I}](\mathcal{J}^*(c)) = \mathcal{I}^*(\mathcal{J}^*(c)) = \mathcal{J}^* \circ \mathcal{I}^*(c) \\ &= (\mathcal{I} \circ \mathcal{J})^*(c) = (a[\mathcal{I}]b)[\mathcal{I} \circ \mathcal{J}]c \end{aligned}$$

(iv) This follows from

$$\begin{aligned}\mathrm{Tr}(\rho a)P_\rho\left(b\left[\mathcal{J}\right]c\mid a\right) &= \mathrm{Tr}(\rho a)\frac{\mathrm{Tr}\left(\mathcal{I}(\rho)b\left[\mathcal{J}\right]c\right)}{\mathrm{Tr}(\rho a)}\\ &= \mathrm{Tr}\left[\mathcal{I}(\rho)\mathcal{J}^*(c)\right]\\ &= \mathrm{Tr}\left[\mathcal{J}\left(\mathcal{I}(\rho)\right)c\right]\\ &= \mathrm{Tr}\left[(\mathcal{I}\circ\mathcal{J})(\rho)c\right]\\ &= \mathrm{Tr}\left[\rho(\mathcal{I}\circ\mathcal{J})^*\right]\\ &= P_\rho\left(a\left[\mathcal{I}\right]b\right)P_\rho\left(c\mid a\left[\mathcal{I}\right]b\right)\ \square\end{aligned}$$

Bayes' second rule says that

$$P_\rho(b\mid a)=\frac{P_\rho(b)}{P_\rho(a)}P_\rho(a\mid b) \tag{1}$$

The following lemma shows that this result does not always hold.

Lemma 3. The following statements are equivalent. (i) Equation (1) holds. (ii) Whenever $\mathcal{I}$ measures a and $\mathcal{J}$ measures b, then

$$\mathrm{Tr}\left(\rho a\left[\mathcal{I}\right]b\right)=\mathrm{Tr}\left(\rho b\left[\mathcal{J}\right]a\right)$$

(iii) Whenever $\mathcal{I}$ measures a and $\mathcal{J}$ measures b, then $\mathrm{Tr}\left[\mathcal{I}(\rho)b\right]=\mathrm{Tr}\left[\mathcal{J}(\rho)a\right]$.

Proof. (i)⇒(ii) If (i) holds, then

$$\begin{aligned}\mathrm{Tr}\left(\rho a\left[\mathcal{I}\right]b\right) &= \mathrm{Tr}\left[\rho\mathcal{I}^*(b)\right]=\mathrm{Tr}\left[\mathcal{I}(\rho)b\right]=P_\rho(a)P_\rho(b\mid a)\\ &= P_\rho(b)P_\rho(a\mid b)=\mathrm{Tr}\left[\mathcal{J}(\rho)a\right]\\ &= \mathrm{Tr}\left[\rho\mathcal{J}^*(a)\right]=\mathrm{Tr}\left(\rho b\left[\mathcal{J}\right]a\right)\end{aligned}$$

Hence, (ii) holds.

(ii)⇒(iii). If (ii) holds, then

$$\begin{aligned}\mathrm{Tr}\left[\mathcal{I}(\rho)b\right] &= \mathrm{Tr}\left[b\mathcal{I}^*(b)\right]=\mathrm{Tr}\left(\rho a\left[\mathcal{I}\right]b\right)=\mathrm{Tr}\left(\rho b\left[\mathcal{J}\right]a\right)\\ &= \mathrm{Tr}\left[\rho\mathcal{J}^*(a)\right]=\mathrm{Tr}\left[\mathcal{J}(\rho)a\right]\end{aligned}$$

Hence, (iii) holds.

(iii)⇒(i) If (iii) holds then

$$P_\rho(b\mid a)=\frac{\mathrm{Tr}\left[\mathcal{I}(\rho)b\right]}{P_\rho(a)}=\frac{\mathrm{Tr}\left[\mathcal{J}(\rho)a\right]}{P_\rho(a)}=\frac{P_\rho(b)\mathrm{Tr}(a\mid b)}{P_\rho(a)}$$

Hence, (i) holds. □

Corollary 4. If $\mathcal{I}$ measures a and $\mathcal{J}$ measures b, then the following statements are equivalent. (i) (1) holds for every $\rho\in\mathcal{S}(H)$. (ii) $a\left[\mathcal{I}\right]b=b\left[\mathcal{J}\right]a$. (iii) $\mathcal{I}^*(b)=\mathcal{J}^*(a)$.

Example 1. For $a\in\mathcal{E}(H)$ we define the Lüders operation $\mathcal{L}^{(a)}(\rho)=a^{\frac{1}{2}}\rho a^{\frac{1}{2}}$ [2, 11, 12]. Then

$$\mathrm{Tr}\left[\mathcal{L}^{(a)}(\rho)\right]=\mathrm{Tr}(a^{\frac{1}{2}}\rho a^{\frac{1}{2}})=\mathrm{Tr}(\rho a)$$

for all $\rho\in\mathcal{S}(H)$ so $\mathcal{L}^{(a)}$ measures a. Notice that $\mathcal{L}^{(a)*}=\mathcal{L}^{(a)}$ for all $a\in\mathcal{E}(H)$ and $a\left[\mathcal{L}^{(a)}\right]b=a^{\frac{1}{2}}ba^{\frac{1}{2}}$. We call $a\left[\mathcal{L}^{(a)}\right]b$ the *standard sequential product of a then b* [7, 13]. Relative to $\mathcal{L}^{(a)}$ we have for all $\rho\in\mathcal{S}(H)$, $b\in\mathcal{E}(H)$ that

$$P_\rho(a\mid b)=\frac{\mathrm{Tr}\left[\mathcal{L}^{(a)}(\rho)b\right]}{P_\rho(a)}=\frac{\mathrm{Tr}(a^{\frac{1}{2}}\rho a^{\frac{1}{2}}b)}{\mathrm{Tr}(\rho a)}=\frac{\mathrm{Tr}(\rho a^{\frac{1}{2}}ba^{\frac{1}{2}})}{\mathrm{Tr}\rho a)}$$

Applying Corollary 4 we have that Bayes' second rule holds relative to $\mathcal{L}^{(a)}$ and $\mathcal{L}^{(b)}$ for all $\rho\in\mathcal{S}(H)$ if and only if $a^{\frac{1}{2}}ba^{\frac{1}{2}}=b^{\frac{1}{2}}ab^{\frac{1}{2}}$. This is equivalent to $ab=ba$; that is, a and b commute [13]. Thus, (1) does not hold, in general. We also have from Theorem 2(iii) that

$$\begin{aligned}a\left[\mathcal{L}^a\right]\left(b\left[\mathcal{L}(b)\right]c\right) &= \left(a\left[\mathcal{L}^{(a)}b\right]\left[\mathcal{L}^{(a)}\circ\mathcal{L}^{(b)}\right]c\right)\\ &= a^{\frac{1}{2}}b^{\frac{1}{2}}cb^{\frac{1}{2}}a^{\frac{1}{2}}\end{aligned}$$

It follows from Theorem 2(ii) that $\mathcal{L}^{(a)}\circ\mathcal{L}^{(b)}$ measures $a\left[\mathcal{L}^{(a)}\right]b=a^{\frac{1}{2}}ba^{\frac{1}{2}}$. However,

$$\begin{aligned}(\mathcal{L}^{(a)}\circ\mathcal{L}^{(b)})(\rho) &= \mathcal{L}^{(b)}(\mathcal{L}^{(a)}(\rho))=\mathcal{L}^{(b)}(a^{\frac{1}{2}}\rho a^{\frac{1}{2}})\\ &= b^{\frac{1}{2}}a^{\frac{1}{2}}\rho a^{\frac{1}{2}}b^{\frac{1}{2}}\end{aligned}$$

and

$$\mathcal{L}^{(a^{\frac{1}{2}}ba^{\frac{1}{2}})}(\rho)=(a^{\frac{1}{2}}ba^{\frac{1}{2}})^{\frac{1}{2}}\rho(a^{\frac{1}{2}}ba^{\frac{1}{2}})^{\frac{1}{2}}$$

so $\mathcal{L}^{(a)}\circ\mathcal{L}^{(b)}\neq\mathcal{L}^{(a[\mathcal{L}^{(a)}]b)}$. We conclude that

$$a\left[\mathcal{L}^{(a)}\right]\left(b\left[\mathcal{L}^{(b)}\right]c\right)\neq\left(a\left[\mathcal{L}^{(a)}\right]b\right)\left[\mathcal{L}^{(a[\mathcal{L}^{(a)}]b)}\right]c$$

in general. □

Example 2. If $a\in\mathcal{E}(H)$, $\alpha\in\mathcal{S}(H)$, we define the *Holevo operation* [6, 14]

$$\mathcal{H}^{(a,\alpha)}(\rho)=\mathrm{Tr}(\rho a)\alpha$$

Then for every $\rho\in\mathcal{S}(H)$, $b\in\mathcal{E}(H)$ we obtain

$$\begin{aligned}\mathrm{Tr}\left[\rho\mathcal{H}^{(a,\alpha)*}(b)\right] &= \mathrm{Tr}\left[\mathcal{H}^{(a,\alpha)}(\rho)b\right]=\mathrm{Tr}\left[\mathrm{Tr}(\rho a)\alpha b\right]\\ &= \mathrm{Tr}(\rho a)\mathrm{Tr}(\alpha b)=\mathrm{Tr}\left[\rho\mathrm{Tr}(\alpha b)a\right]\end{aligned}$$

Hence,

$$\mathcal{H}^{(a,\alpha)*}(b)=a\left[\mathcal{H}^{(a,\alpha)}\right]b=\mathrm{Tr}(\alpha b)a$$

Since $\mathrm{Tr}\left[\mathcal{H}^{(a,\alpha)}(\rho)\right]=\mathrm{Tr}(\rho a)$ we see that $\mathcal{H}^{(a,\alpha)}$ measures a. This shows that for any $a\in\mathcal{E}(H)$, there are many operations that measure a. The conditional probability of b given a relative to $\mathcal{H}^{(a,\alpha)}$ becomes

$$P_\rho(b\mid a)=\frac{\mathrm{Tr}\left[\mathcal{H}^{(a,\alpha)}(\rho)b\right]}{P_\rho(a)}=\frac{\mathrm{Tr}(\rho a)\mathrm{Tr}(\alpha b)}{\mathrm{Tr}(\rho a)}=\mathrm{Tr}(\alpha b)$$

which curiously is independent of ρ and a. Applying Corollary 4 we have that Bayes' second rule holds for all $\rho \in \mathcal{S}(H)$ relative to $\mathcal{H}^{(a,\alpha)}$ and $\mathcal{H}^{(b,\beta)}$ if and only if

$$\mathrm{Tr}(\alpha b)a = \mathrm{Tr}(\beta a)b$$

If a and b are sharp this is equivalent to $a = b$ and $\mathrm{Tr}(\alpha a) = \mathrm{Tr}(\beta a)$. Moreover, Theorem 2(iii) becomes

$$\begin{aligned} a\left[\mathcal{H}^{(a,\alpha)}\right]\left(b\left[\mathcal{H}^{(b,\beta)}\right]c\right) &= \left(a\left[\mathcal{H}^{(a,\alpha)}\right]b\right)\left[\mathcal{H}^{(a,\alpha)}\circ\mathcal{H}^{(b,\beta)}\right]c \\ &= a\left[\mathcal{H}^{(a,\alpha)}\right]\left(\mathcal{H}^{(b,\beta)*}(c)\right) \\ &= a\left[\mathcal{H}^{(a,\alpha)}\right](\mathrm{Tr}(\beta c)b) \\ &= \mathrm{Tr}(\beta c)a\left[\mathcal{H}^{(a,\alpha)}\right]b \\ &= \mathrm{Tr}(\beta c)\mathcal{H}^{(a,\alpha)*}(b) \\ &= \mathrm{Tr}(\beta c)\mathrm{Tr}(\alpha b)a \end{aligned}$$

Unlike the Lüders operations, we have

$$\mathcal{H}^{(a,\alpha)}\circ\mathcal{H}^{(b,\beta)} = \mathcal{H}^{\left(a\left[\mathcal{H}^{(a,\alpha)}\right]b,\beta\right)}$$

Indeed,

$$\begin{aligned} \mathcal{H}^{(a,\alpha)}\circ\mathcal{H}^{(b,\beta)}(\rho) &= \mathcal{H}^{(b,\beta)}\left[\mathcal{H}^{(a,\alpha)}(\rho)\right] = \mathcal{H}^{(b,\beta)}(\mathrm{Tr}(\rho a)\alpha) \\ &= \mathrm{Tr}(\rho a)\mathcal{H}^{(b,\beta)}(\alpha) = \mathrm{Tr}(\rho a)\mathrm{Tr}(\alpha b)\beta \\ &= \mathrm{Tr}\left[\rho\mathrm{Tr}(\alpha b)a\right]\beta = \mathcal{H}^{(\mathrm{Tr}(\alpha b)a,\beta)}(\rho) \\ &= \mathcal{H}^{(\mathcal{H}(a,\alpha)^*(b),\beta)}(\rho) \\ &= \mathcal{H}^{\left(a\left[\mathcal{H}^{(a,\alpha)}\right]b,\beta\right)}(\rho) \end{aligned}$$

□

2 Observables and Instruments

A (finite) *observable* is a collection of effects $A = \{A_x : x \in \Omega_A\}$ on H satisfying $\sum_{x\in\Omega_A} A_x = I$ [1–3, 5]. We assume that the set Ω_A is finite and call Ω_A the *outcome space* for A. We think of A as an experiment or measurement and when the outcome x results, then we say that the effect A_x *occurs*. The condition $\sum_{x\in\Omega_A} A_x = I$ means that one of the outcomes occurs when a measurement of A is performed. If $\rho \in \mathcal{S}(H)$, then $P_\rho(A_x) = \mathrm{Tr}(\rho A_x)$ is the probability that the outcome x results and A_x occurs. We call $A(\Delta) = \sum\{A_x : x \in \Delta\}$, where $\Delta \subseteq \Omega_A$, a *positive operator-valued measure* (POVM). The *probability distribution* of A in the state ρ is the measure given by $\Phi_\rho^A(\Delta) = \sum_{x\in\Delta} P_\rho(x)$ for all $\Delta \in \Omega_A$ and we usually write

$$\Phi_\rho^A(x) = \Phi_\rho^A(\{x\}) = P_\rho(A_x)$$

A (finite) *instrument* is a finite collection of operations $\mathcal{I} = \{\mathcal{I}_x : x \in \Omega_\mathcal{I}\}$ such that $\overline{\mathcal{I}} = \sum_{x\in\Omega_\mathcal{I}} \mathcal{I}_x$ is a channel [1–3, 5, 15]. Then for all $\rho \in \mathcal{S}(H)$ and $\Delta \subseteq \Omega_\mathcal{I}$

$$\Phi_\rho^\mathcal{I}(\Delta) = \sum\{\mathrm{Tr}\left[\mathcal{I}_x(\rho)\right] : x \in \Delta\}$$

is a probability measure on $\Omega_\mathcal{I}$. We say that $\mathcal{I}$ *measures* an observable A if for all $\rho \in \mathcal{S}(H)$, we have $\mathrm{Tr}\left[\mathcal{I}_x(\rho)\right] = \mathrm{Tr}(\rho A_x)$ for every $x \in \Omega_A$. Clearly, $\mathcal{I}$ measures a unique observable and they both have the same probability distribution. As with operations and effects, an observable is measured by many instruments. If $\mathcal{I}$ is an instrument, its *dual instrument* $\mathcal{I}^* : \mathcal{L}(H) \to \mathcal{L}(H)$ satisfies [8, 10]

$$\mathrm{Tr}\left[\rho\mathcal{I}_x^*(A)\right] = \mathrm{Tr}\left[\mathcal{I}_x(\rho)A\right]$$

for all $A \in \mathcal{L}(H)$

$$\mathcal{I}_{\Omega_\mathcal{I}}^*(I) = \sum_{x\in\Omega_\mathcal{I}} \mathcal{I}_x^*(I) = I$$

It is easy to check that $\mathcal{I}_x^* : \mathcal{E}(H) \to \mathcal{E}(H)$ and $\mathcal{I}_x^*(I)$ is the observable measured by $\mathcal{I}$.

If $a \in \mathcal{E}(H)$ and A is an observable on H measured by the instrument $\mathcal{I}$, the effect *a conditioned by A* is the effect

$$(a \mid A) = \sum_{x\in\Omega_A} \mathcal{I}_x^*(a) = \mathcal{I}_{\Omega_A}^*(a) = \sum_{x\in\Omega_A} A_x\left[\mathcal{I}_x\right]a$$

It is clear that $a \mapsto (a \mid A)$ is a morphism in the sense that $(I \mid A) = I$ and if $a_i \in \mathcal{E}(H)$ with $\sum_{i=1}^n a_i \leq I$ then $\left(\sum_{i=1}^n a_i \mid A\right) = \sum_{i=1}^n (a_i \mid A)$. A *sub-observable* is a finite collection of effects $A = \{A_x : x \in \Omega_A\}$ on H satisfying $\sum_{x\in\Omega_A} A_x \leq I$ [8]. If A is a sub-observable, then A possesses a unique minimal extension to an observable by adjoining the effect $I - \sum_{x\in\Omega_A} A_x$ to A. If A is an observable and $a \in \mathcal{E}(H)$ is measured by an operation $\mathcal{I}$, then *A conditioned by a* is the sub-observable given by $(A \mid a)_x = a\left[\mathcal{I}\right]A_x$ [9]. Notice that we have $\sum_{x\in\Omega_A} (A \mid a)_x = a\left[\mathcal{I}\right]I = a$. If A and B are observables on H and $\mathcal{I}$ is an instrument that measures A, then *B conditioned on A relative to $\mathcal{I}$* is the observable [9]

$$(B \mid A)_y = \sum_{x\in\Omega_\mathcal{I}} \mathcal{I}_x^*(B_y) = \sum_{x\in\Omega_A} A_x\left[\mathcal{I}_x\right]B_y$$

If $\mathcal{I}$ and $\mathcal{J}$ are instruments on H we define the instrument *$\mathcal{J}$ conditioned by $\mathcal{I}$* as [9, 10]

$$(\mathcal{J} \mid \mathcal{I})_y(\rho) = \sum_{x\in\Omega_\mathcal{I}} \mathcal{J}_y(\mathcal{I}_x(\rho)) = \mathcal{J}_y\left[\overline{\mathcal{I}}(\rho)\right]$$

for all $\rho \in \mathcal{S}(H)$, $y \in \Omega_\mathcal{J}$. The next result corresponds to Theorem 2.

Theorem 5. Suppose $\mathcal{I}$ measures A and $\mathcal{J}$ measures B. (i) $(\mathcal{J} \mid \mathcal{I})$ measures $(B \mid A)$. (ii) For any observable C we have $((C \mid B) \mid A) = (C \mid (B \mid A))$.

Proof. (i) For every $\rho \in \mathcal{S}(H)$ we have

$$\begin{aligned}\mathrm{Tr}\left[(\mathcal{J} \mid \mathcal{I})_y(\rho)\right] &= \mathrm{Tr}\left[\sum_{x\in\Omega_{\mathcal{I}}} \mathcal{J}_y\left(\mathcal{I}_x(\rho)\right)\right]\\ &= \sum_{x\in\Omega_{\mathcal{I}}} \mathrm{Tr}\left[\mathcal{J}_y\left(\mathcal{I}_x(\rho)\right)\right]\\ &= \sum_{x\in\Omega_{\mathcal{I}}} \mathrm{Tr}\left[\mathcal{I}_x(\rho)B_y\right]\\ &= \sum_{x\in\Omega_{\mathcal{I}}} \mathrm{Tr}\left[\rho\mathcal{I}_x^*(B_y)\right]\\ &= \mathrm{Tr}\left[\rho\sum_{x\in\Omega_{\mathcal{I}}} \mathcal{I}_x^*(B_y)\right]\\ &= \mathrm{Tr}\left[\rho(B \mid A)_y\right]\end{aligned}$$

It follows that $(\mathcal{J} \mid \mathcal{I})$ measures $(B \mid A)$.

(ii) It follows from (i) that $(\mathcal{J} \mid \mathcal{I})$ measures $(B \mid A)$. Then for all $z \in \Omega_C$ we obtain

$$\begin{aligned}((C \mid B) \mid A)_z &= \overline{\mathcal{I}}^*(C \mid B)_z = \overline{\mathcal{I}}^*\left[\overline{\mathcal{J}}^*(C_z)\right]\\ &= \overline{\mathcal{J}}^* \circ \overline{\mathcal{I}}^*(C_z) = \overline{(\mathcal{I} \circ \mathcal{J})}^*(C_z)\\ &= \overline{(\mathcal{J} \mid \mathcal{I})}^*(C_z) = (C \mid (B \mid A))_z\end{aligned}$$

which gives the result. □

Theorem 6. If $\mathcal{I}$ measures A and $a \in \mathcal{E}(H)$, then for all $\rho \in \mathcal{S}(H)$ we have

$$\sum_{x\in\Omega_A} P_\rho(A_x)P_\rho(a \mid A_x) = P_\rho\left[(a \mid A)\right] = P_{\overline{\mathcal{I}}(\rho)}(a) \qquad (2)$$

Proof. We have that

$$\begin{aligned}\sum_{x\in\Omega_A} P_\rho(A_x)P_\rho(a \mid A_x) &= \sum_{x\in\Omega_A} \mathrm{Tr}(\rho A_x)\frac{\mathrm{Tr}\left[\rho\mathcal{I}_x^*(a)\right]}{\mathrm{Tr}(\rho A_x)}\\ &= \mathrm{Tr}\left[\rho\sum_{x\in\Omega_{\mathcal{I}}} \mathcal{I}_x^*(a)\right]\\ &= \mathrm{Tr}\left[\rho(a \mid A)\right] = \mathrm{Tr}\left[\rho\mathcal{I}_\Omega^*(a)\right]\\ &= \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)a\right] = P_{\overline{\mathcal{I}}(\rho)}(a)\end{aligned}$$

and the result follows. □

We call (2) *Bayes' quantum first rule*. This is the same as the classical Bayes' first rule except it depends on the instrument used to measure A. We then say that (2) is *context dependent* and that $\mathcal{I}$ is the *context* in which A is measured. In classical probability theory there is only one context available and no context dependence.

We say that a sub-observable A is *real-valued* if $\Omega_A \subseteq \mathbb{R}$ [16]. If A is real-valued and $\rho \in \mathcal{S}(H)$ the *ρ-average* (or *ρ-expectation*) of A is

$$E_\rho(A) = \sum_{x\in\Omega_A} xP_\rho(A_x) = \sum_{x\in\Omega_A} x\mathrm{Tr}(\rho A_x)$$

If A is real-valued, we define its *stochastic operator* [16] to be the self-adjoint operator $\widetilde{A} = \sum_{x\in\Omega_A} xA_x$. We then have

$$E_\rho(A) = \mathrm{Tr}\left(\rho\sum_{x\in\Omega_A} xA_x\right) = \mathrm{Tr}(\rho\widetilde{A})$$

which is the expectation of $\widetilde{A}$ in the state ρ. We also define the *conditional ρ-average*

$$E_\rho(A \mid a) = \sum_{x\in\Omega_A} xP_\rho(A_x \mid a) = \sum_{x\in\Omega_A} \frac{x\mathrm{Tr}\left[\rho\mathcal{I}^*(A_x)\right]}{\mathrm{Tr}(\rho a)}$$

where $\mathcal{I}$ measures a. The next result is called *Bayes' quantum first rule for expectations*.

Theorem 7. If $\mathcal{I}$ measures A and B is a real-valued observable, then

$$\sum_{x\in\Omega_A} P_\rho(A_x)E_\rho(B \mid A_x) = E_\rho\left[(B \mid A)\right] = E_{\overline{\mathcal{I}}(\rho)}(B)$$

Proof. For all $\rho \in \mathcal{S}(H)$, $x \in \Omega_A$ we have

$$E_\rho(B \mid A_x) = \sum_{y\in\Omega_B} \frac{y\mathrm{Tr}\left[\rho\mathcal{I}_x^*(B_y)\right]}{P_\rho(A_x)}$$

It follows that

$$\begin{aligned}\sum_{x\in\Omega_A} P_\rho(A_x)E_\rho(B \mid A_x) &= \sum_{x\in\Omega_A}\sum_{y\in\Omega_B} y\mathrm{Tr}\left[\rho\mathcal{I}_x^*(B_y)\right]\\ &= \sum_{y\in\Omega_B} y\mathrm{Tr}\left[\rho\sum_{x\in\Omega_A} \mathcal{I}_x^*(B_y)\right]\\ &= \sum_{y\in\Omega_B} y\mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)B_y\right]\\ &= E_{\overline{\mathcal{I}}(\rho)}(B)\\ &= \sum_{y\in\Omega_B} y\mathrm{Tr}\left[\rho(B \mid A)_y\right]\\ &= E_\rho\left[(B \mid A)\right]\end{aligned}$$

□

Example 3. Let A be the atomic observable

$$A = \{P_x \colon x \in \Omega_A\} = \{|\phi_x\rangle\langle\phi_x| \colon x \in \Omega_A\}$$

and let $\mathcal{I}$ be the instrument

$$\mathcal{I}_x(\rho) = P_x\rho P_x = \langle\phi_x, \rho\phi_x\rangle|\phi_x\rangle\langle\phi_x|$$

that measures A. Then

$$A_x\left[\mathcal{I}_x\right]a = \mathcal{I}_x^*(a) = \langle\phi_x, a\phi_x\rangle|\phi_x\rangle\langle\phi_x|$$

Moreover, if $B = \left\{B_y \colon y \in \Omega_B\right\}$ is an observable on H, then $(B \mid P_x)$ is the sub-observable $(B \mid P_x)_y = P_xB_yP_x$

and if $a \in \mathcal{E}(H)$, then $(a \mid A)$ is the effect $\mathcal{I}_\Omega^*(a)$. For all $\rho \in \mathcal{S}(H)$ we obtain

$$\begin{aligned} P_\rho(a \mid A) &= P_\rho\left[\mathcal{I}_\Omega^*(a)\right] = P_{\overline{\mathcal{I}}(\rho)}(a) \\ &= \operatorname{Tr}\left[\sum_{x\in\Omega_A} \langle \phi_x, \rho\phi_x\rangle P_x a\right] \\ &= \sum_{x\in\Omega_A} \langle \phi_x, \rho\phi_x\rangle\langle \phi_x, a\phi_x\rangle \end{aligned}$$

If B is a real-valued observable, we obtain

$$\begin{aligned} E_\rho(B \mid A) &= E_{\overline{\mathcal{I}}(\rho)}(B) = \operatorname{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right] \\ &= \sum_{x\in\Omega_A} \langle \phi_x, \rho\phi_x\rangle\left\langle \phi_x, \widetilde{B}\phi_x\right\rangle \end{aligned}$$

Bayes' quantum first rule gives

$$\begin{aligned} \sum_{x\in\Omega_A} P_\rho(P_x)P_\rho(a \mid P_x) &= P_\rho(a \mid A) \\ &= \sum_{x\in\Omega_A} \langle \phi_x, \rho\phi_x\rangle\langle \phi_x, a\phi_x\rangle \end{aligned}$$

and Bayes' quantum first rule for expectations gives

$$\begin{aligned} \sum_{x\in\Omega_A} P_\rho(P_x)E_\rho(B \mid P_x) &= E_\rho(B \mid A) \\ &= \sum_{x\in\Omega_A} \langle \phi_x\rho\phi_x\rangle\left\langle \phi_x\widetilde{B}\phi_x\right\rangle \end{aligned}$$ □

Example 4. If $A = \{A_x \colon x \in \Omega_A\}$ is an observable and $\alpha_x \in \mathcal{S}(H)$, $x \in \Omega_A$, we define the *Holevo instrument* $\mathcal{H}_x^{(A,\alpha)}(\rho) = \operatorname{Tr}(\rho A_x)\alpha_x$ [6, 14]. Then $\mathcal{H}^{(A,\alpha)}$ measures A because

$$\begin{aligned} \operatorname{Tr}\left[\mathcal{H}_x^{(A,\alpha)}(\rho)\right] &= \operatorname{Tr}\left[\operatorname{Tr}(\rho A_x)\alpha_x\right] = \operatorname{Tr}(\rho A_x)\operatorname{Tr}(\alpha_x) \\ &= \operatorname{Tr}(\rho A_x) \end{aligned}$$

Also, the dual of $\mathcal{H}^{(A,\alpha)}$ becomes

$$\mathcal{H}_x^{(A,\alpha)*}(a) = \operatorname{Tr}(\alpha_x a)A_x$$

and

$$\begin{aligned} (a \mid A) &= \mathcal{H}_{\Omega_A}^{(A,\alpha)*}(a) \\ &= \sum_{x\in\Omega_A} \mathcal{H}^{(A,\alpha)*}(a) \\ &= \sum_{x\in\Omega_A} \operatorname{Tr}(\alpha_x a)A_x \end{aligned}$$

Then Bayes' quantum first rule becomes

$$\sum_{x\in\Omega_A} P_\rho(A_x)P_\rho(a \mid A_x) = P_\rho(a \mid A) = \sum_{x\in\Omega_A} \operatorname{Tr}(\rho A_x)\operatorname{Tr}(\alpha_x a)$$

Moreover, if $B = \left\{B_y \colon y \in \Omega_B\right\}$ is a real-valued observable, then

$$\begin{aligned} E_\rho(B \mid A) &= \sum_{y\in\Omega_B} y\operatorname{Tr}\left[\overline{\mathcal{H}^{(A,\alpha)}}(\rho)B_y\right] \\ &= \sum_{y\in\Omega_B} y\operatorname{Tr}\left[\sum_{x\in\Omega_A} \mathcal{H}_x^{(A,\alpha)}(\rho)B_y\right] \\ &= \sum_{y\in\Omega_B} y\operatorname{Tr}\left[\sum_{x\in\Omega_A} \operatorname{Tr}(\rho A_x)\alpha_x B_y\right] \\ &= \sum_{x\in\Omega_A} \operatorname{Tr}(\rho A_x)\operatorname{Tr}(\alpha_x\widetilde{B}) \end{aligned}$$

Bayes' quantum first rule for expectations becomes

$$\sum_{x\in\Omega_A} P_\rho(A_x)E_\rho(B \mid A_x) = \sum_{x\in\Omega_A} \operatorname{Tr}(\rho A_x)\operatorname{Tr}(\alpha_x\widetilde{B})$$ □

Example 5. If $\mathcal{H}^{(A,\alpha)}$ and $\mathcal{H}^{(B,\beta)}$ are Holevo instruments, we show that

$$\mathcal{H}^{(A,\alpha)} \circ \mathcal{H}^{(B,\beta)} = \mathcal{H}^{(C,\beta)}$$

is the Holevo instrument with $C_{(x,y)} = \operatorname{Tr}(\alpha_y B_y)A_x$. Indeed

$$\begin{aligned} \left(\mathcal{H}^{(A,\alpha)} \circ \mathcal{H}^{(B,\beta)}\right)_{(x,y)}(\rho) &= \mathcal{H}_y^{(B,\beta)}(\mathcal{H}_x^{(A,\alpha)})(\rho) \\ &= \mathcal{H}_y^{(B,\beta)}\left[\operatorname{Tr}(\rho A_x)\alpha_x\right] \\ &= \operatorname{Tr}(\rho A_x)\operatorname{Tr}(\alpha_x B_y)\beta_y \\ &= \operatorname{Tr}\left[\rho\operatorname{Tr}(\alpha_x B_y)A_x\right]\beta_y \\ &= \operatorname{Tr}(\rho C_{(x,y)})\beta_y = \mathcal{H}_{(x,y)}^{(C,\beta)}(\rho) \end{aligned}$$

In contrast, if $\mathcal{L}^A$, $\mathcal{L}^B$ are Lüders instruments $\mathcal{L}_x^{(A)}(\rho) = A_x^{\frac{1}{2}}\rho A_x^{\frac{1}{2}}$, $\mathcal{L}_y^{(B)} = B_y^{\frac{1}{2}}\rho B_y^{\frac{1}{2}}$, we show that $\mathcal{L}^A \circ \mathcal{L}^B$ is not Lüders, in general. Indeed, suppose $\mathcal{L}^A \circ \mathcal{L}^B = \mathcal{L}^C$. We then obtain

$$\begin{aligned} (\mathcal{L}^A \circ \mathcal{L}^B)_{(x,y)}(\rho) &= \mathcal{L}_y^B(\mathcal{L}_x^A(\rho)) = B_y^{\frac{1}{2}}A_x^{\frac{1}{2}}\rho A_x^{\frac{1}{2}}B_y^{\frac{1}{2}} \\ &= C_{(x,y)}^{\frac{1}{2}}\rho C_{(x,y)}^{\frac{1}{2}} \end{aligned}$$

for all $\rho \in \mathcal{S}(H)$. Taking the trace of both sides gives $C_{(x,y)} = A_x^{\frac{1}{2}}B_y A_x^{\frac{1}{2}}$ and we conclude that

$$B_y^{\frac{1}{2}}A_x^{\frac{1}{2}}\rho A_x^{\frac{1}{2}}B_y^{\frac{1}{2}} = (A_x^{\frac{1}{2}}B_y A_x^{\frac{1}{2}})^{\frac{1}{2}}\rho(A_x^{\frac{1}{2}}B_y A_x^{\frac{1}{2}})^{\frac{1}{2}}$$

for all $\rho \in \mathcal{S}(H)$. Letting $\rho = I/n$ where $n = \dim H$ gives

$$B_y^{\frac{1}{2}}A_x B_y^{\frac{1}{2}} = A_x^{\frac{1}{2}}B_y A_x^{\frac{1}{2}}$$

This holds if and only if $A_x B_y = B_y A_x$, in which case $(\mathcal{L}^A \circ \mathcal{L}^B)_{(x,y)} = \mathcal{L}^{A_x B_y}$ for every $x \in \Omega_A$, $y \in \Omega_B$. In a similar way, if $a, b \in \mathcal{E}(H)$, then

$$\mathcal{H}^{(a,\alpha)} \circ \mathcal{H}^{(b,\beta)} = \mathcal{H}^{(\operatorname{Tr}(\alpha b)a,\beta)}$$

and $\mathcal{L}^a \circ \mathcal{L}^b$ is not Lüders unless $ab = ba$ in which case $\mathcal{L}^a \circ \mathcal{L}^b = \mathcal{L}^{ab}$. □

We say that an observable $A = \{A_x \colon x \in \Omega_A\}$ is *commuting* if $A_x A_y = A_y A_x$ for all $x, y \in \Omega_A$. Also, two observables B, C are *jointly commuting* if B and C are commuting and $B_x C_y = C_y B_x$ for all $x \in \Omega_B$, $y \in \Omega_C$.

Theorem 8. Two observables B, C are jointly commuting if and only if there exists an atomic observable A and observables B_1, C_1, such that $B = (B_1 \mid A)$, $C = (C_1 \mid A)$ relative to some instrument that measures A.

Proof. If $B = (B_1 \mid A)$, then $B_y = \sum_{x\in\Omega_A} \mathcal{I}_x^*(B_1 y)$ and by Lemma 1(iii) $\mathcal{I}_x^*(B_1 y) = \lambda_{x,y} A_x$ for $\lambda_{x,y} \in [0,1]$. Hence, $B_y = \sum_{x\in\Omega_A} \lambda_{x,y} A_x$. In a similar way, $C_z = \sum_{x\in\Omega_A} \mu_{x,z} A_x$ for $\mu_{x,z} \in [0,1]$. It follows that B and C are jointly commuting. Conversely, if B and C are jointly commuting, then all the effects in $\left\{B_y, C_z \colon y \in \Omega_B, z \in \Omega_C\right\}$ commute so they are simultaneously diagonalizable. Hence, there exists an atomic observable A such that $B_y = \sum_{x\in\Omega_A} \mathrm{Tr}(A_x B_y) A_x$ and $C_z = \sum_{x\in\Omega_A} \mathrm{Tr}(A_x C_z) A_x$ for all $y \in \Omega_B$, $z \in \Omega_C$. Using the Lüders instrument $\mathcal{L}_x^A \rho = A_x \rho A_x$ we have

$$(B \mid A)_y = \sum_{x\in\Omega_A} \mathcal{L}_x^{A^*} B_y = \sum_{x\in\Omega_A} A_x B_y A_x = \sum_{x\in\Omega_A} \mathrm{Tr}(A_x B_y) A_x = B_y$$

Similarly, $(C \mid A)_z = C_z$ so $B = (B \mid A)$ and $C = (C \mid A)$. □

Corollary 9. Observables B, C are jointly commuting if and only if there exists an atomic observable A such that $B = (B \mid A)$, $C = (C \mid A)$ relative to some instrument that measures A.

A similar proof gives the following.

Theorem 10. The following statements are equivalent. (i) An observable B is commuting. (ii) There exists an atomic observable A such that $B = (B \mid A)$. (iii) There exists an observable C and an atomic observable A such that $B = (C \mid A)$.

3 Uncertainty Principle and Entropy

Let B be a real-valued observable with stochastic operator $\widetilde{B} = \sum_{y\in\Omega_B} y B_y$. We have seen that $E_\rho(B) = \mathrm{Tr}(\rho\widetilde{B})$. Also, if A is an arbitrary observable and the instrument $\mathcal{I}$ measures A, then relative to $\mathcal{I}$ we have $E_\rho(B \mid A) = \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right]$. We call $E_\rho(B \mid A)$ the ρ-expectation of B in *context* A. If A, B, C are observables and B, C are real-valued, we define the *ρ-correlation of B and C in the context A* by [16]

$$\mathrm{Cor}(B, C \mid A) = \mathrm{Tr}\left[\rho(B \mid A)^\sim (C \mid A)^\sim\right] - E_\rho(B \mid A) E_\rho(C \mid A) = \mathrm{Tr}\left[\rho(B \mid A)^\sim (C \mid A)^\sim\right] - \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right] \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{C}\right]$$

Although $\mathrm{Cor}_\rho(B, C \mid A)$ need not be a real number, it is easy to check that

$$\overline{\mathrm{Cor}_\rho(B, C \mid A)} = \mathrm{Cor}_\rho(C, B \mid A)$$

We call $\Delta_\rho(B, C \mid A) = \mathrm{Re}\left[\mathrm{Cor}_\rho(B, C \mid A)\right]$ the *ρ-covariance of B and C in the context A* [16]. We define the *ρ-variance of B in the context of A* [16]

$$\Delta_\rho(B \mid A) = \mathrm{Cor}_\rho(B, B \mid A) = \Delta_\rho(B, B \mid A) = \mathrm{Tr}\left\{\rho\left[(B \mid A)^\sim\right]^2\right\} - \left\{\mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right]\right\}^2$$

Defining the *commutator* of $(B \mid A)^\sim$ with $(C \mid A)^\sim$ by

$$\left[(B \mid A)^\sim, (C \mid A)^\sim\right] = (B \mid A)^\sim (C \mid A)^\sim - (C \mid A)^\sim (B \mid A)^\sim$$

we obtain the *uncertainty principle* [16]:

$$\tfrac{1}{4} \left|\mathrm{Tr}\left(\rho\left[(B \mid A)^\sim, (C \mid A)^\sim\right]\right)\right|^2 + \left[\Delta_\rho(B, C \mid A)\right]^2 = \left|\mathrm{Cor}_\rho(B, C \mid A)\right|^2 \le \Delta_\rho(B \mid A) \Delta_\rho(C \mid A) \tag{3}$$

The variance $\Delta_\rho(B \mid A)$ gives the amount of uncertainty or lack of information about B provided by ρ relative to a first measurement of A. The less $\Delta_\rho(B \mid A)$ is, the more information ρ provides about B. Equation (3) gives a lower bound for the product of the uncertainties. Notice that (3) generalizes the usual uncertainty principle.

Example 6. Suppose A is sharp in which case $A_xA_{x'} = \delta_{xx'}$ for all $x, x' \in \Omega_A$. Let us measure A with the Lüders instrument $\mathcal{I}_x(\rho) = A_x\rho A_x$. We can now compute the various statistical quantities more completely. To simplify the notation we write $D_x = A_xDA_x$ for $D \in \mathcal{L}(H)$. We then have $\overline{\mathcal{I}}(\rho) = \sum_{x\in\Omega_A} \rho_x$ and

$$E_\rho(B \mid A) = \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right] = \sum_{x,y} y\mathrm{Tr}(\rho_x B_y) = \sum_x \mathrm{Tr}(\rho_x\widetilde{B})$$
$$(B \mid A)^\sim = \sum_y y(B \mid A)_y = \sum_x \widetilde{B}_x$$

We then obtain

$$\mathrm{Cor}_\rho(B, C \mid A) = \mathrm{Tr}\left(\rho\sum_x \widetilde{B}_x \sum_{x'} \widetilde{C}_{x'}\right) - \mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{B}\right]\mathrm{Tr}\left[\overline{\mathcal{I}}(\rho)\widetilde{C}\right] = \sum_x \mathrm{Tr}(\rho_x\widetilde{B}A_x\widetilde{C}) - \sum_{x,x'} \mathrm{Tr}(\rho_x\widetilde{B})\mathrm{Tr}(\rho_{x'}\widetilde{C})$$

$$\begin{aligned}\Delta_\rho(B, C \mid A) &= \mathrm{Re}\left[\mathrm{Cor}_\rho(B, C \mid A)\right] = \tfrac{1}{2}\left[\mathrm{Cor}_\rho(B, C \mid A) + \mathrm{Cor}_\rho(C, B \mid A)\right] \\ &= \tfrac{1}{2}\sum_x \mathrm{Tr}\left[\rho_x(\widetilde{B}A_x\widetilde{C} + \widetilde{C}A_x\widetilde{B})\right] - \sum_{x,x'} \mathrm{Tr}(\rho_x\widetilde{B})\mathrm{Tr}(\rho_{x'}\widetilde{C})\end{aligned}$$

$$\Delta_\rho(B \mid A) = \sum_x \mathrm{Tr}\left[\rho(\widetilde{B}_x)^2\right] - \left[\sum_x \mathrm{Tr}(\rho_x\widetilde{B})\right]^2$$
$$\Delta_\rho(C \mid A) = \sum_x \mathrm{Tr}\left[\rho(\widetilde{C}_x)^2\right] = \left[\sum_x \mathrm{Tr}(\rho_x\widetilde{C})\right]^2$$

Finally, the commutator term becomes

$$\mathrm{Tr}\left\{\rho\left[(B \mid A)^\sim, (C \mid A)^\sim\right]\right\} = \mathrm{Tr}\left(\rho\left[\sum_x \widetilde{B}_x, \sum_{x'} \widetilde{C}_{x'}\right]\right) = \mathrm{Tr}\left[\rho\sum_x(\widetilde{B}_x\widetilde{C}_x - \widetilde{C}_x\widetilde{B}_x)\right] = \sum_x\left[\rho_x(\widetilde{B}A_x\widetilde{C} - \widetilde{C}A_x\widetilde{B})\right]$$

Substituting these terms into (3) gives the uncertainty principle for this case. □

Example 7. Suppose A is measured by the Holevo instrument $\mathcal{H}_x^{(A,\alpha)}(\rho) = \mathrm{Tr}(\rho A_x)\alpha_x$. Then

$$\overline{\mathcal{H}^{(A,\alpha)}}(\rho) = \sum_{x\in\Omega_A} \mathrm{Tr}(\rho A_x)\alpha_x$$

and we saw in Example 5 that

$$E_\rho(B \mid A) = \sum_x \mathrm{Tr}(\rho A_x)\mathrm{Tr}(\alpha_x\widetilde{B})$$

Since $\mathcal{H}_x^{(A,\alpha)*}(B_y) = \mathrm{Tr}(\alpha_x B_y)A_x$ we obtain

$$(B \mid A)^\sim = \sum_y y(B \mid A) = \sum_y y\sum_x \mathcal{H}_x^{(A,\alpha)*}(B_y) = \sum_y y\sum_x \mathrm{Tr}(\alpha_x B_y)A_x = \sum_x \mathrm{Tr}(\alpha_x\widetilde{B})A_x$$

It follows that

$$\begin{aligned}\mathrm{Cor}_\rho(B, C \mid A) &= \mathrm{Tr}\left[\rho\sum_x \mathrm{Tr}(\alpha_x\widetilde{B})A_x\sum_{x'} \mathrm{Tr}(\alpha_{x'}\widetilde{C})A_{x'}\right] - \left[\sum_x \mathrm{Tr}(\rho A_x)\mathrm{Tr}(\alpha_x\widetilde{B})\right]\left[\sum_x \mathrm{Tr}(\rho A_x)\mathrm{Tr}(\alpha_x\widetilde{C})\right] \\ &= \sum_{x,x'} \mathrm{Tr}(\alpha\widetilde{B})\mathrm{Tr}(\alpha_{x'}\widetilde{C})\left[\mathrm{Tr}(\rho A_xA_{x'}) - \mathrm{Tr}(\rho A_x)\mathrm{Tr}(\rho A_{x'})\right]\end{aligned}$$

$$\Delta_\rho(B, C \mid A) = \mathrm{Re}\left[\mathrm{Cor}_\rho(B, C \mid A)\right] = \sum_{x,x'} \mathrm{Tr}(\alpha_x\widetilde{B})\mathrm{Tr}(\alpha_{x'}\widetilde{C})\left\{\tfrac{1}{2}\left[\mathrm{Tr}\left(\rho(A_xA_{x'} + A_{x'}A_x)\right)\right] - \mathrm{Tr}(\rho A_x)\mathrm{Tr}(A_{x'})\right\}$$

$$\Delta_\rho(B \mid A) = \sum_{x,x'} \mathrm{Tr}(\alpha_x \widetilde{B})\mathrm{Tr}(\alpha_{x'} \widetilde{B}) \left[\mathrm{Tr}(\rho A_x A_{x'}) - \mathrm{Tr}(\rho A_x)\mathrm{Tr}(\rho A_{x'})\right]$$

with a similar formula for $\Delta_\rho(C \mid A)$. Finally, the commutator term becomes

$$\mathrm{Tr}\left\{\rho\left[(B \mid A)^{\sim}, (C \mid A)^{\sim}\right]\right\} = \mathrm{Tr}\left\{\rho\left[\sum_x \mathrm{Tr}(\alpha_x \widetilde{B})A_x, \sum_{x'} \mathrm{Tr}(\alpha_{x'} \widetilde{C})A_{x'}\right]\right\} = \sum_{x,x'} \mathrm{Tr}(\alpha_x \widetilde{B})\mathrm{Tr}(\alpha_{x'} \widetilde{C})\mathrm{Tr}\left(\rho\left[A_x, A_{x'}\right]\right)$$

Substituting these terms into (3) gives the uncertainty principle for this case. □

The uncertainty $\Delta_\rho(A)$ measures the lack of information about A provided by the state ρ. In the dual picture, we have the lack of information $S_A(\rho)$ that a measurement of A provides about the state ρ and this is called entropy. We now briefly discuss conditional entropy. If $a \in \mathcal{E}(H)$, $\rho \in \mathcal{S}(H)$, we define the *ρ-entropy* of a by [17–20]

$$S_a(\rho) = -\mathrm{Tr}(\rho a) \ln\left[\frac{\mathrm{Tr}(\rho A)}{\mathrm{Tr}(a)}\right]$$

We interpret $S_a(\rho)$ as the amount of uncertainty that a measurement of a provides about ρ. The smaller $S_a(\rho)$ is, the more information a measurement of a gives about ρ. It follows that if $\mathcal{I}$ measures a, then

$$\begin{aligned} S_{a[\mathcal{I}]b}(\rho) &= -\mathrm{Tr}\left(\rho a\left[\mathcal{I}\right]b\right) \ln\left[\frac{\mathrm{Tr}\left(\rho a\left[\mathcal{I}\right]b\right)}{\mathrm{Tr}\left(a\left[\mathcal{I}\right]b\right)}\right] \\ &= -\mathrm{Tr}\left[\rho \mathcal{I}^*(b)\right] \ln\left[\frac{\mathrm{Tr}\left[\rho \mathcal{I}^*(b)\right]}{\mathrm{Tr}\left[\mathcal{I}^*(b)\right]}\right] \\ &= -\mathrm{Tr}\left[\mathcal{I}(\rho)b\right] \ln\left[\frac{\mathrm{Tr}\left[\mathcal{I}(\rho)b\right]}{\mathrm{Tr}\left[\mathcal{I}^*(b)\right]}\right] \end{aligned}$$

We define the *a-conditional ρ-entropy of b* as

$$\begin{aligned} S_{(b|a)}(\rho) &= S_b\left[\mathcal{I}(\rho)\right] \\ &= -\mathrm{Tr}\left[\mathcal{I}(\rho)b\right] \ln\left[\frac{\mathrm{Tr}\left[\mathcal{I}(\rho)b\right]}{\mathrm{Tr}(b)}\right] \end{aligned}$$

Notice that there is a close connection between these two entropies. Since *lnx* is an increasing function we have the following.

Lemma 11. $S_{a[\mathcal{I}]b}(\rho) \leq S_{(b|a)}(\rho)$ for every $\rho \in \mathcal{S}(H)$ if and only if $\mathrm{Tr}\left[\mathcal{I}^*(b)\right] \leq \mathrm{Tr}(b)$.

Example 8. If a is measured by the Lüders operation $\mathcal{L}^{(a)}(\rho) = a^{\frac{1}{2}}\rho a^{\frac{1}{2}}$, then

$$\mathrm{Tr}\left[\mathcal{I}^*(b)\right] = \mathrm{Tr}(a^{\frac{1}{2}}ba^{\frac{1}{2}}) = \mathrm{Tr}(ab) \leq \mathrm{Tr}(b)$$

so in this case we have $S_{a[\mathcal{I}]b}(\rho) \leq S_{(b|a)}(\rho)$ for all $\rho \in \mathcal{S}(H)$. □

Example 9. If a is measured by the Holevo operation $\mathcal{H}^{(a,\alpha)}(\rho) = \mathrm{Tr}(\rho a)\alpha$, then

$$\mathrm{Tr}\left[\mathcal{H}^{(a,\alpha)*}(b)\right] = \mathrm{Tr}\left[\mathrm{Tr}(\alpha b)a\right] = \mathrm{Tr}(\alpha b)\mathrm{Tr}(a)$$

Hence, $\mathrm{Tr}\left[\mathcal{H}^{(a,\alpha)*}(b)\right] \leq \mathrm{Tr}(b)$ if and only if $\mathrm{Tr}(\alpha b)\mathrm{Tr}(a) \leq \mathrm{Tr}(b)$. Depending on a, b, α this inequality sometimes holds and sometimes does not hold. We conclude that $S_{a[\mathcal{I}]b}$ and $S_{(a|b)}$ give different measures of information about ρ. □

If A is an observable, we define the *ρ-entropy* of A by [17, 19]

$$\begin{aligned} S_A(\rho) &= \sum_{x \in \Omega_a} S_{A_x}(\rho) \\ &= -\sum_{x \in \Omega_A} \mathrm{Tr}(\rho A_x) \ln\left[\frac{\mathrm{Tr}(\rho A_x)}{\mathrm{Tr}(A_x)}\right] \end{aligned}$$

If $\mathcal{I}$ measures A, we define the *A-conditional ρ-entropy* of the observable B by [17, 19]

$$\begin{aligned} S_{(B\|A)}(\rho) &= S_B\left[\overline{\mathcal{I}}(\rho)\right] \\ &= \sum_{y \in \Omega_B} S_{B_y}\left[\overline{\mathcal{I}}(\rho)\right] \end{aligned}$$

As with effects, this can be compared with

$$S_{(B|A)}(\rho) = \sum_{y \in \Omega_B} S_{(B|A)_y}(\rho)$$

and these are not related in general.

One of the advantages of $S_{(B\|A)}$ over $S_{(B|A)}$ is the following. If $\mathcal{I}$ measures A and $\mathcal{J}$ measures B we obtain

$$\begin{aligned} S_{((C\|B)\|A)}(\rho) &= S_{(C\|B)}\left[\overline{\mathcal{I}}(\rho)\right] \\ &= S_C\left[\overline{\mathcal{J}}\left(\overline{\mathcal{I}}(\rho)\right)\right] \\ &= S_{(C\|(B\|A))} \end{aligned}$$

but in general

$$S_{((C|B)|A)} \neq S_{(C|(B|A))}$$

We can continue this to obtain results concerning more than three observables.

References

[1] P. Busch, M. Grabowski, P. J. Lahti. Operational Quantum Physics. Vol. 31 of *Lecture Notes in Physics Monographs*. Springer, Berlin, 1995. doi:10.1007/978-3-540-49239-9.

[2] E. B. Davies, J. T. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics* 1970; **17**(3):239–260. doi:10.1007/bf01647093.

[3] T. Heinosaari, M. Ziman. The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement. Cambridge University Press, Cambridge, 2012. doi:10.1017/cbo9781139031103.

[4] K. Kraus. States, Effects, and Operations: Fundamental Notions of Quantum Theory. Vol. 190 of *Lecture Notes in Physics*. Springer, Berlin, 1983. doi:10.1007/3-540-12732-1.

[5] M. A. Nielsen, I. L. Chuang. Quantum Computation and Quantum Information. Cambridge University Press, Cambridge, 2010. doi:10.1017/cbo9780511976667.

[6] A. S. Holevo. Quantum coding theorems. *Russian Mathematical Surveys* 1998; **53**(6):1295–1331. arXiv:quant-ph/9809023. doi:10.1070/rm1998v053n06abeh000091.

[7] S. P. Gudder. Sequential products of quantum measurements 2021; arXiv:2108.07925.

[8] S. P. Gudder. Dual quantum instruments and subobservables 2023; arXiv:2302.12243.

[9] S. P. Gudder. Conditioned observables in quantum mechanics 2020; arXiv:2005.04775.

[10] S. P. Gudder. Combinations of quantum observables and instruments. *Journal of Physics A: Mathematical and Theoretical* 2021; **54**(36):364002. arXiv:2010.08025. doi:10.1088/1751-8121/ac1829.

[11] G. Lüders. Über die Zustandsänderung durch den Meßprozeß. *Annalen der Physik* 1951; **443**(5-8):322–328. doi:10.1002/andp.19504430510.

[12] G. Lüders. Concerning the state-change due to the measurement process. *Annalen der Physik* 2006; **15**(9):663–670. doi:10.1002/andp.200610207.

[13] S. P. Gudder, R. Greechie. Sequential products on effect algebras. *Reports on Mathematical Physics* 2002; **49**(1):87–111. doi:10.1016/S0034-4877(02)80007-6.

[14] A. S. Holevo. Probabilistic and Statistical Aspects of Quantum Theory. Scuola Normale Superiore Edizioni della Normale, Pisa, Italy, 2011. doi:10.1007/978-88-7642-378-9.

[15] S. P. Gudder. Quantum instruments and conditioned observables 2020; arXiv:2005.08117.

[16] S. P. Gudder. Real-valued observables and quantum uncertainty. *International Journal of Theoretical Physics* 2023; **62**(5):94. arXiv:2301.07185. doi:10.1007/s10773-023-05342-8.

[17] S. P. Gudder. Entropy of quantum measurements. *Entropy* 2022; **24**(11):1686. doi:10.3390/e24111686.

[18] M. Ohya, D. Petz. Quantum Entropy and Its Use. Texts and Monographs in Physics. Springer, Berlin, 2004.

[19] D. Šafránek, J. Thingna. Quantifying information extraction using generalized quantum measurements. *Physical Review A* 2023; **108**(3):032413. arXiv:2007.07246. doi:10.1103/PhysRevA.108.032413.

[20] A. Wehrl. General properties of entropy. *Reviews of Modern Physics* 1978; **50**(2):221–260. doi:10.1103/RevModPhys.50.221.