Multi-Observer Quantum Mechanics from Classical Bayesian Inference

Kevin Player*

April 26, 2023

Solipsism may be logically consistent with present Quantum Mechanics, Monism in the sense of Materialism is not.

Eugene Wigner

Abstract

We present several quintessential quantum ideas and shed them in a classical light. We argue how quantum information theory can be understood as a generalization of classical information theory in a nonstandard way (without density matrices). In fact, we show how classical information theory can be embedded in quantum information theory using zero phase wavefunctions. We employ this embedding to motivate a new multi-observer extension of quantum mechanics. Finally, we outline an experiment to test the existence of our multi-observer theory.

1 Classical Information Theory

Lets start with a 3-bit example, where each bit is realized by coins which are either heads (H) or tails (T). With these 3 coins in hand, we conduct some thought experiments. During this section we outline an epistemic treatment of classical information where we represent our classical knowledge of the 3 coin ensemble. The ideas that we encounter all extend without loss of generality to n-bits or any finite set.

1.1 Tribits and Notation

Let us describe the situation once we throw the 3 coins. We move freely between binary bits (0 and 1) and coin states (H and T) mapping between them $H \leftrightarrow 0$ and $T \leftrightarrow 1$. Let

$$x_i \in \{H, T\}$$

be the outcome for the throws $i = 1, 2, 3$ and let $x = (x_1, x_2, x_3)$ denote the resulting sequence or tribit. There are 8 tribits and we will refer to them with the italic integers *0* through *7* in binary order. For example the tribit *6* has any of the following forms

- 110
- TTH

*kjplaye@gmail.com

- $(1, 1, 0)$
- (T, T, H)
- tribit *6*

Let C be the set of 8 possible tribits $\{HHH, HHT, \cdots, TTT\}$.

1.2 Rules for Constructing Knowledge Statements

We want to describe knowledge statements[1] about the coins. For any set A, let $\mathcal{K}(A)$ be the set of knowledge statements about A. Our goal will be to construct and understand $\mathcal{K}(C)$ by repeatedly applying the following simple rules.

Postulate 1.1. *For any tribit $x \in C$ we have the knowledge statement $b_x \in \mathcal{K}(C)$, that says we know x with certainty or probability 1. We call this a basic statement.*

There are 8 tribits, so there are 8 basic statements, $b_0, \cdots, b_7$.

Postulate 1.2. *A formal superposition can be made for any statements $s_1 \in \mathcal{K}(C)$ and $s_2 \in \mathcal{K}(C)$ and $\alpha \in [0, 1]$*

$$super_\alpha(s_1, s_2) = \alpha s_1 + (1 - \alpha)s_2. \tag{1}$$

This formal combination means to take statement s_1 with probability α and statement s_2 with probability $(1 - \alpha)$; This is a new statement in $\mathcal{K}(C)$.

1.3 Distributions on the Ensemble

For any set A let $\mathcal{P}(A)$ be the set of probability distributions on A. Then $\mathcal{P}(C)$ is the set of probability distributions on tribits. For a fixed $\eta \in \mathcal{P}(C)$ we let p_i be the probability of the tribit i in η (for instance p_6 is the probability of seeing TTH.) Then the eight p_i completely characterize η as the vector $(p_0, \cdots, p_7) \in \mathbb{R}^8$. Call this 8 dimensional ambient space $\mathcal{V}(C)$ (and in general $\mathcal{V}(A) = \mathbb{R}^A$ for set A).

Lemma 1.1. $\mathcal{K}(C) = \mathcal{P}(C)$

Proof. Let $\eta \in \mathcal{K}(C)$. The rules 1.1 and 1.2, that form η, output probability distributions at every step. So $\eta \in \mathcal{P}(C)$.

Next let $\eta = (p_i) \in \mathcal{P}(C)$. We inductively extend the distribution

$$\eta_k = \gamma_k(p_0, \cdots, p_{k-1}, 0, \cdots)$$

to

$$\eta_{k+1} = \gamma_{k+1}(p_0, \cdots, p_k, 0, \cdots)$$

using basic b_k and

$$\eta_{k+1} = super_\alpha(\eta_k, b_k)$$

for appropriate choices of normalizations γ_k, γ_{k+1} and α. So we have a sequence of rules that form η and hence $\eta \in \mathcal{K}(C)$. $\square$

Corollary 1.1.1. For any finite set A,

$$\mathcal{K}(A) = \left\{ (p_a)_{a \in A} \in \mathbb{R}^A \middle| \sum_{a \in A} p_a = 1 \right\} \subseteq \mathcal{V}(A)$$

where $\mathcal{K}(A)$ is affine codimension 1 in $\mathcal{V}(A)$ and $\mathcal{V}(A)$ is the vectorspace $\mathbb{R}^A$.

[1] Whenever we say "statement" in this paper, we mean knowledge statement, or Bayesian belief statement.

Proof. The right hand side is the definition of $\mathcal{P}(A)$ and the claim follows using a generalization of the proof of Lemma 1.1. $\square$

A general knowledge statement $\eta \in \mathcal{K}(C)$ combines the basic statements $b_0, \cdots, b_7$ using coefficients $p_0, \cdots, p_7$,

$$p_0, \cdots, p_7 \in \mathbb{R}^+,$$

so that $\eta = \sum p_i b_i$. To be a statistical distribution we also need it to be normalized with $\sum p_i = 1$. Note that now (1) is more than just a formal relation since it actually holds linearly inside of $\mathcal{V}(C)$.

Let $\mathcal{S}$ be the category of nonempty sets. Let $\mathcal{W}$ be the category of real vectorspaces. Then

Lemma 1.2. $\mathcal{V}$ is a functor from $\mathcal{S}$ to $\mathcal{W}$ that takes injections and surjections to surjections and injections respectively.

Proof. Let $f : A \hookrightarrow B$ be an injection. Define $\mathcal{V}(f) : \mathcal{V}(B) \twoheadrightarrow \mathcal{V}(A)$ to be an orthogonal projection which is a surjection

$$b \mapsto \begin{cases} b & \text{if } b \in A \\ 0 & \text{if } b \notin A \end{cases}$$

Now let $g : A \twoheadrightarrow B$ be a surjection. Define $\mathcal{V}(g) : \mathcal{V}(B) \hookrightarrow \mathcal{V}(A)$ to send each $b \in B$ to the sum over the fiber of b. That is we send

$$b \mapsto \sum_{a \in g^{-1}(b)} a.$$

Each of these fibers is disjoint so the map is injective.

Consider a general map $h : A \to B$. We write this map as a surjection onto the image followed by an injection $A \twoheadrightarrow Im(h) \hookrightarrow B$. We then send h to

$$\mathcal{V}(A) \hookleftarrow \mathcal{V}(Im(h)) \twoheadleftarrow \mathcal{V}(B).$$

Finally, we examine the composition of $r : A \to B$ and $s : B \to D$. For any $d \in D$ we have three possibilities

1. $d \notin Im(s)$
2. $d \in Im(s)$, but $s^{-1}(d) \cap Im(r) = \emptyset$
3. $d \in Im(s \circ r)$

Consider condition #1. In this case we have $\mathcal{V}(s)$ sending d to zero. For condition #2, $\mathcal{V}(s)$ sends d to a fiber of s, but then all of $s^{-1}(d)$ gets sent to zero by $\mathcal{V}(r)$. For condition #3, $\mathcal{V}(s)$ sends d to $s^{-1}(d)$ and then that gets sent by $\mathcal{V}(r)$ to the sum over the nonempty $r^{-1}(s^{-1}(d))$. These cases and behaviors are the same as directly considering $\mathcal{V}(s \circ r)$ where the cases #1 and #2 have d getting sent to zero and case #3 exhibits the same behavior since the r fibers of the fibers of s are just the fibers of $s \circ r$.

This demonstrates the composition property of the functor $\mathcal{V}(s \circ r) = \mathcal{V}(r) \circ \mathcal{V}(s)$. $\square$

Each coin corresponds to a 2 dimensional space $\mathcal{V}(\{H, T\})$ which we will call V_1. The 8 dimensional space $V_1^{\otimes 3}$ has a basis with vectors like $T \otimes H \otimes T$. Then $\mathcal{V}(C)$, is identified with the tensor power $V_1^{\otimes 3}$ by sending

$$(x_1, x_2, x_3) \mapsto x_1 \otimes x_2 \otimes x_3.$$

Lemma 1.3. $\mathcal{V}(A \times B) = \mathcal{V}(A) \otimes \mathcal{V}(B)$ and the n-bit knowledge space $\mathcal{V}(\mathbb{Z}_2^n)$ is $V_1^{\otimes n} \cong \mathbb{R}^{2^n}$. $\mathcal{A} \times \mathcal{B} \twoheadrightarrow \mathcal{A}$ identifies $\mathcal{V}(A)$ with $\mathcal{V}(A) \otimes \mathbb{1}_B$. The same is true for B and $\mathbb{1}_A$.

Proof. We map the $|A| \cdot |B|$ basis elements $(a, b) \mapsto a \otimes b$. The linear extension of the map is an isomorphism. For n-bits we repeatedly apply the rule. The sum over the fiber of $a \in A$ is just a tensored with the sum over $b \in B$ which we call $\mathbb{1}_B$. $\square$

1.4 Classical Bayesian Projections

We don't always get to measure the exact tribit. For instance, we might only get to know that the first coin is T and that the 2nd and 3rd coins are the same. This corresponds to a subset of C, which we will call S

$$\begin{aligned} S &= \{s \in C | x_1 = T, x_2 = x_3\} \\ &= \{4, 7\} \qquad \text{(THH, TTT)} \end{aligned}$$

We will use S as an example in the remaining text.

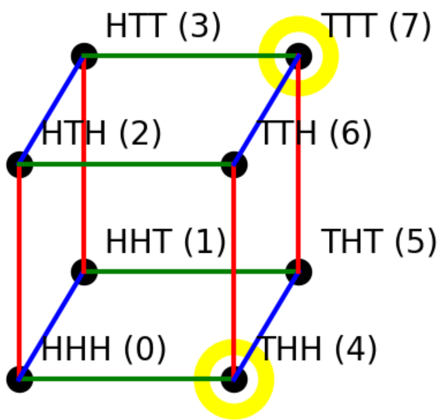

Figure 1: Coin sequences with S circled in yellow.

Someone could make a measurement that tells them that the state is in S and nothing else. This would be a specific case of what we will call a projection measurement ρ that updates our knowledge. The statement becomes zero exactly on p_i where $i \notin S$ and renormalizes p_i on the remaining $i \in S$. More specifically,

$$\rho : (p_0, \cdots, p_7) \mapsto \left(0, 0, 0, 0, \frac{p_4}{p_4 + p_7}, 0, 0, \frac{p_7}{p_4 + p_7}\right) \tag{2}$$

which motivates the following

Definition 1.1 (Bayesian Projection)**.** For any finite set A with $B \subseteq A$ we consider a measurement of B. It takes p_a to 0 when $a \notin B$ and $\frac{p_a}{\sum_{b \in B} p_b}$ when $a \in B$. We call this a Bayesian projection since the ambient spaces follow a projection $\mathcal{V}(B \hookrightarrow A)$.

1.5 Classical Bayesian Inference

A more general type of measurement is a probabilistic measurement. Someone could learn that there is a 95% chance that the state is in S. In full generality we will call such a probabilistic observation $\mathcal{O}$. We can figure out how to update our knowledge statement from p_i, to $\hat{p_i}$, using a relative[2] version of Bayes's rule

$$\frac{\hat{p_i}}{\hat{p_j}} = \underbrace{\frac{P(i|\mathcal{O})}{P(j|\mathcal{O})}}_{\text{Posterior}} = \underbrace{\frac{P(\mathcal{O}|i)}{P(\mathcal{O}|j)}}_{\text{Bayes Factor}} \quad \underbrace{\frac{p_i}{p_j}}_{\text{Prior}}$$

Pulling out the Bayes factor we find that we just multiply by the likelihood and re-normalize

$$\hat{p_k} = P(\mathcal{O}|k) \;\; p_k. \tag{3}$$

A special case are projections $P(\mathcal{O}|k) \in \{0, v\}$, for some fixed value v, like the S projection case in section 1.4. These are the Bayesian projections see Definition 1.1.

[2]In the ratio, the contribution of $P(\mathcal{O})$ is canceled out and accounted for during normalization. $P(\mathcal{O})$ only holds significance prior to its observation and it doesn't require consideration during the update process.

Definition 1.2 (Bayesian Inference)**.** An update of the form (3) is called a Bayesian Inference.

1.6 Classical Correlation

We consider the uniform distribution on S.

$$\frac{1}{2}(T \otimes H \otimes H) + \frac{1}{2}(T \otimes T \otimes T)$$

We know the first coin is T, which we promptly throw away.

We have HH or TT with equal probabilities for the second and third coins. If we give the second coin to Alice and the third coin to Bob then we have a classical correlation. If Bob finds that the third coin is T then we know that Alice will also find that the second coin is T (and similarly for F). The reduction is purely epistemic, it is the separate knowledge of Alice and Bob that is changing.

Definition 1.3 (Correlation)**.** Two bits are said to be correlated when our knowledge is written as

$$\frac{1}{2}(H \otimes H) + \frac{1}{2}(T \otimes T)$$

2 Quantum Information Theory

2.1 Classical to Quantum Embedding

General quantum information[3] about 8-qubits can be expressed as a wave function.

$$q_0, \cdots, q_7 \in \mathbb{C}$$

with $\sum |q_i|^2 = 1$. The Born rule is ostensibly a map $q_i \mapsto |q_i|^2 = p_i$ to classical probability. Here we can instrument all of the classical information theoretic constructs by restricting the phase[4] of q_i to be zero. We can map backward $p_i \mapsto \sqrt{p_i} = q_i$; which commutes with the Born rule (subject to normalization):

$$\text{(Quantum)} \qquad \mathbb{C}^n \rightleftarrows (\mathbb{R}_{\geq 0})^n \qquad \text{(Classical)}$$

So quantum information theory needs to generalize the classical. In particular we will see that quantum measurement and entanglement restrict to classical measurement and classical correlation for zero phase.

2.2 Quantum Bayesian Projection

We illustrate zero phase Bayesian projection with an example.

Within the embedding we define a projection operator π which projects onto the space generated by the subset S, from the previous subsection

$$\pi = \sum_{i \in S} |i\rangle \langle i| \,.$$

Then let A be any operator with an eigenspace, with eigenvalue λ, equal to the image of π. An observation of λ would correspond to an application of π, which is a Bayesian projection as in Definition 1.1.

[3] We are purposely leaving out density matrices and POVMs and just dealing with pure states.

[4] We consider negative values as 180 degrees out of phase, so zero phase means non-negative real.

Lemma 2.1. The set inclusion $B \subseteq A$ yields a Bayesian Projection as in Definition 1.1. Then a zero phase wavefunction instruments the classical measurement.

Proof. Let $q_i = \sqrt{p_i}$ be a zero phase wavefunction corresponding to $(p_i) \in \mathcal{K}$. Let

$$\lambda_i = \begin{cases} 1 \text{ if } i \in B \\ 2 \text{ if } i \notin B \end{cases}$$

and define an operator

$$D = \sum_{i \in A} \lambda_i \left|i\right\rangle \left\langle i\right|$$

If we make a measurement of 1, in D, then the wave function collapses by projecting onto the eigenspace of 1. The wave function is then renormalized. This projection on the B-space and normalization of q_i exactly matches the classical Bayesian projection and normalization of p_i. □

2.3 Quantum Bayesian Inference

Lemma 2.2. Let $P(\mathcal{O}|i)$ define a Bayesian inference as in Definition 1.2. Then a zero phase wavefunction instruments the classical measurement.

Proof. Consider, in the fashion of [1], a more general measurement with matrix $M_{\mathcal{O}}$ which is diagonal with entries

$$M_{\mathcal{O}}(i,i) = \sqrt{P(\mathcal{O}|i)}$$

This measurement is identical to the classical Bayesian inference for zero phase wavefunctions. □

2.4 Quantum Entanglement

Lemma 2.3. A classical correlation knowledge statement becomes an entangled zero phase wavefunction.

Proof. The wavefunction in Definition 1.3 maps to

$$\frac{1}{\sqrt{2}}(H \otimes H) + \frac{1}{\sqrt{2}}(T \otimes T)$$

which is a entangled Bell state. □

2.5 Entropy

Von Neumann entropy is left out of this note since it is not a generalization of classical entropy in the manner presented here. This is because the zero phase classical wavefunctions are pure states which all have zero Von Neumann entropy.

3 Overview of the Generalization

The quantum mechanical concepts wave function collapse, entanglements and ensembles are direct generalizations of the classical information theoretic concepts of Bayesian inference, correlation and statistical ensembles respectively. This is not to say that they are not strange, but to say that they need to be generalizations of the classical concepts. Wave function collapsing should generalize Bayesian inference and classical measurement. Entanglement should be a generalization of classical correlation. This is enough to motivate new ways of doing quantum mechanics, which we will encounter in the next sections.

3.1 Multi-Observer Quantum Mechanics

We focus on quantum wave function collapse as a generalization of classical Bayesian inference. We can implement the classical Bayesian inference within zero phase quantum mechanics as outlined above. So the classical Bayesian theory has a direct tie in; that observation and measurement occur in tandem with a change in knowledge[5].

In the classical theory, knowledge is local to the observer[6], multiple observers each have their own knowledge. For instance, Wigner's friend and Wigner each have their own classical knowledge. It would seem then that Wigner's friend and Wigner must have different zero phase wavefunctions as well, if they are to be generalizing classical knowledge. This is our prediction for multi-observer quantum physics, that every observer has a local wavefunction.

3.2 Prediction and Experiment

We predict that multi-observer quantum mechanics will be required for a proper generalization of classical knowledge. We propose an experiment where we inject an "observer" into a quantum eraser[7]. The injected observer will have to be a small apparatus that is

- able to record a measurement ψ of the eraser particle path.
- able to forget the measurement ψ.
- able to demonstrate that a measurement was made with a record ρ.

Let ϕ be the lab technician's wavefunction. In ϕ the apparatus is entangled with the subject particles in the eraser experiment. We need to find out if there is another wavefunction in play that is not just part of ϕ. The key here is that the apparatus is able to demonstrate that it "knew" something, or in other words that another wavefunction ψ existed, using ρ.

The apparatus can not keep ψ for proper erasure, but must "forget" it. Observed erasure, via an interference pattern, proves that ψ is not a part of ϕ. Finally, there should be a record ρ in the apparatus that it did at one point in time record a measurement ψ. Receipt of the report ρ and eraser interference pattern together show that ψ and ϕ were necessarily part of a two-observer system.

4 Acknowledgments

Thanks to Erik Ferragut and Dan Justice for useful discussions.

References

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2011.

[2] Kevin Player, "The Unruh effect as holographic thrust," 2023.

[5] This change in knowledge is tangible, it always occurs as a transfer of matter/energy from the environment to the observer [2].

[6] Note that current quantum theory is a theory of one observer, usually the experimenter in a lab. An immediate subject is quantum key distribution(QKD), which requires at least three observers, Alice, Bob, and Eve; the security of QKD depends on multi-observer quantum ontology.

[7] Here we probe the boundaries of what constitutes an observer and a measurement. We claim that all versions of observer and measurement will be detectable in this experiment whenever it can be done.