Title: Multi-agent paradoxes beyond quantum theory

Speakers: Vilasini Venkatesh

Series: Quantum Foundations

Date: February 25, 2021 - 10:00 AM

URL: http://pirsa.org/21020033

Abstract: With ongoing efforts to observe quantum effects in larger and more complex systems, both for the purposes of quantum computing and fundamental tests of quantum gravity, it becomes important to study the consequences of extending quantum theory to the macroscopic domain. Frauchiger and Renner have shown that quantum theory, when applied to model the memories of reasoning agents, can lead to a conflict with certain principles of logical deduction. Is this incompatibility a peculiar feature of quantum theory, or can modelling reasoning agents using other physical theories also lead to such contradictions? What features of physical theories are responsible for such paradoxes?

Multi-agent paradoxes have been previously analysed only in quantum theory. To address the above questions, a framework for analysing multi-agent paradoxes in general physical theories is required. Here, we develop such a framework that can in particular be applied to generalized probabilistic theories (GPTs). We apply the framework to model how observersâ€™ memories may evolve in box world, a post-quantum GPT and using this, derive a stronger paradox that does not rely on post-selection. Our results reveal that reversible, unitary evolution of agentsâ€™ memories is not necessary for deriving multi-agent logical paradoxes, and suggest that certain forms of contextuality might be.

https://iopscience.iop.org/article/10.1088/1367-2630/ab4fc4

Multi-agent paradoxes beyond quantum theory

V. Vilasini[1,2] Nuriya Nurgalieva[2] Lídia del Rio[2]

[1]Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK

[2]Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland

Perimeter Institute

February 2021

V. Vilasini, N. Nurgalieva, L. del Rio. *New Journal of Physics, 21, 113028 (2019)*

Talk overview


Motivation


A quantum multi-agent paradox (Frauchiger-Renner)


Beyond quantum: Generalised reasoning, memories and measurements


Memory update in box-world


A post-quantum multi-agent paradox

MOTIVATION

Motivation

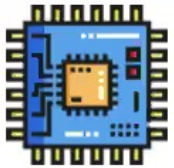
- When applying quantum theory to the real world, we often make a 'cut' between what is modelled as quantum or classical (e.g., observed vs observer, micro vs macroscopic).

"Quantum"

"Classical"

- But there is no clear quantum vs classical divide. In principle, quantum computers could play the role of agents/observers.



Quantum or classical?

- There are global efforts towards observing quantum effects in meso/macroscopic systems, both for quantum computing and fundamental tests of quantum gravity.

Important to study the implications of extending quantum theory to larger, more complex systems.

Specific questions and our contribution

- **Previous work:** Modelling the memories of reasoning agents as quantum systems can lead to a conflict with simple principles of logical reasoning. (Frauchiger and Renner Nat. Comm. 2018, Nurgalieva and del Rio EPCTS 2019)
- **Questions:** Which properties of quantum theory are responsible for such multi-agent logical paradoxes? Are they unique to quantum theory?
- **Potential applications:** Rules of logical reasoning that are applicable to these general scenarios. These could play a role in future quantum computers.

Need a general (theory-independent) framework for analysing multi-agent paradoxes.

- **Our contributions:** We propose such a framework, suggest a possible model of memory update in box-world (a post-quantum theory) and find a stronger paradox there.

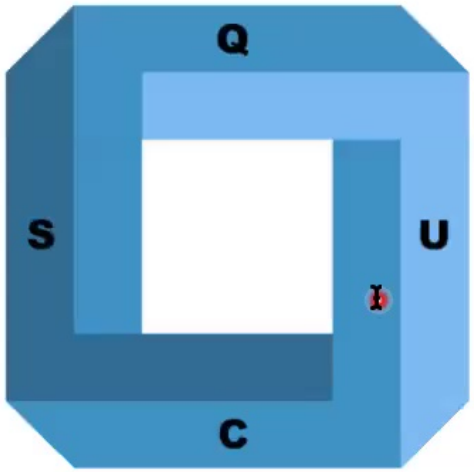
A QUANTUM MULTI-AGENT PARADOX

The Frauchiger-Renner no-go theorem

Theorem (Frauchiger and Renner 2018, Nurgalieva and del Rio 2019)

No physical theory $\mathbb{T}$ can simultaneously satisfy the four assumptions Q, U, C *and* S.

Q and U are assumptions relating to the (universal) validity of quantum theory,
C and S are assumptions relating to the validity of certain basic logical principles.



An impossible square: There is no physical theory that squares with the 4 assumptions!

Spoiler alert: We will see that contextuality, which also has the structure of "local consistency" and "global inconsistency" (Abramsky and Brandenburger, NJP 2011) is key for interpreting such results.

The assumptions

Q: Validity of the Born rule (deterministic case)

If the Born rule predicts, for a given state that a given measurement completed at time t yields the outcome $x = \xi$ with probability 1, then an agent who knows the state and measurement can be certain that $x = \xi$ at time t.

U: Quantum evolution of agents' memories

Suppose Alice measures a system S and stores the outcome in her memory A. Then another agent Bob can model the measurement process in Alice's lab as a reversible evolution given by a unitary U_{AS}.

E.g., If Alice measures the $|+\rangle_S$ state in the Z basis and stores the outcome in her memory (initialised to $|0\rangle_A$), then U_{AS} is simply the CNOT gate:

$$U_{AS} : \frac{1}{\sqrt{2}}(|0\rangle_S + |1\rangle_S) \otimes |0\rangle_A \mapsto \frac{1}{\sqrt{2}}(|0\rangle_S |0\rangle_A + |1\rangle_S |1\rangle_A)$$

C: Consistency between agent's perspectives

If a theory $\mathbb{T}$ satisfies C, then for any two agents Alice and Bob reasoning using the theory,

Alice is certain that Bob is certain that $x = \xi$ at time t

$\Downarrow$

Alice is certain that $x = \xi$ at time t.

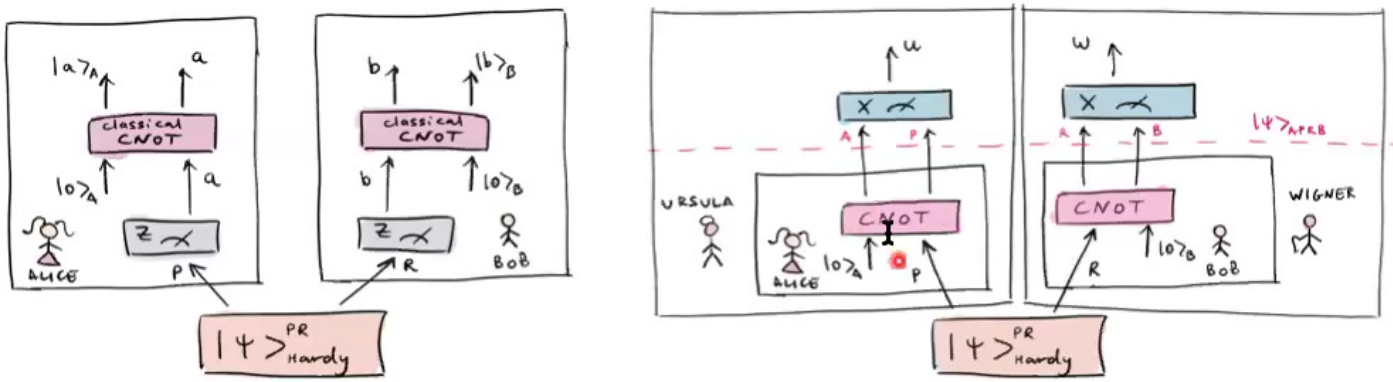
S: Measurement outcomes have a single value

No agent can conclude with certainty that $x = \xi$ at time t and $x \neq \xi$ at time t.

The paradox

Protocol:

- Alice and Bob share the bipartite Hardy state $|\psi\rangle^{PR}_{Hardy} = \frac{1}{\sqrt{3}}(|00\rangle_{PR} + |10\rangle_{PR} + |11\rangle_{PR})$.
- Alice and Bob measure the systems P and R in Z basis and store the corresponding outcome (a and b) in their memory A and B respectively.
- Ursula and Wigner measure the joint systems AP and RB in the "X basis" $\{|ok\rangle = (|00\rangle - |11\rangle)/\sqrt{2}, |fail\rangle = (|00\rangle + |11\rangle)/\sqrt{2}\}$ to obtain the outcomes u and w respectively.



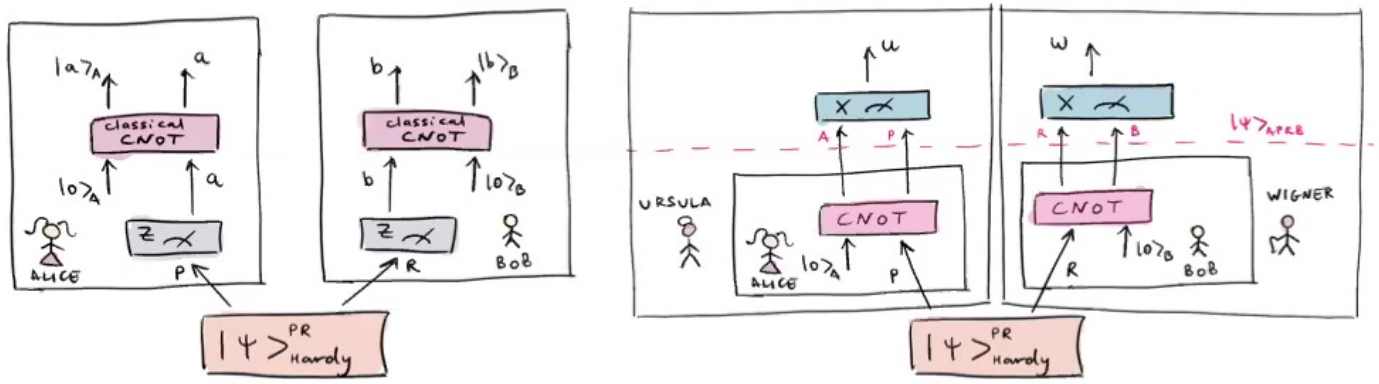
(a) Inside perspective

(b) Outside perspective

An entanglement-based version of the Frauchiger-Renner thought-experiment.

(Refs: Frauchiger and Renner 2018, Hardy 1992, Pusey 2018.)

The paradox



(a) Inside perspective

(b) Outside perspective

Reasoning:

- The state of the joint system $APRB$ after Alice and Bob measure, from the perspective of Ursula and Wigner would be (using U):

$$|\psi\rangle_{APRB} = \frac{1}{\sqrt{3}}(|0000\rangle + |1100\rangle + |1111\rangle)_{APRB}.$$

- The Born rule yields the probability $P(u = w = ok) = 1/12$ along with $P(b = 1|u = ok) = P(a = 1|b = 1) = P(w = fail|a = 1) = 1$. Therefore the parties can repeat protocol until $u = w = ok$ is obtained and then reason using Q and C to obtain the following paradoxical chain which contradicts S.

$$w = ok, u = ok \Rightarrow b = 1 \Rightarrow a = 1 \Rightarrow w = fail.$$

GENERALISED REASONING, MEMORIES AND MEASUREMENTS

Reasoning about knowledge

Theory-independent condition that tells us that rational agents can reason about each other's knowledge in the usual way, formalized using *epistemic modal logic*.

Definition (Reasoning agents)

An experimental setup with multiple agents $A_1, \dots A_N$ can be described by knowledge operators $K_1, \dots K_N$ and statements $\phi \in \Phi$, such that $K_i\phi$ denotes "agent A_i knows ϕ." It should allow agents to make deductions, that is

$$K_i[\phi \wedge (\phi \implies \psi)] \implies K_i\, \psi. \tag{1}$$

Furthermore, each experimental setup defines a trust relation between agents : we say that an agent A_i trusts another agent A_j (and denote it by $A_j \rightsquigarrow A_j$) iff for all statements ϕ, we have

$$K_i(K_j\, \phi) \implies K_i\, \phi. \tag{2}$$

- Equation (1) is known as the *distributive axiom* in modal logic.
- This notion of a trust relation (Eq. (2)) was introduced in Nurgalieva, del Rio 2019.

Reasoning about knowledge

An example:

- Suppose "Alice knows that Bob knows that Eve doesn't know the secret key k" and Alice trusts Bob to be a rational, reliable agent, then she can deduce that "I know that Eve doesn't know the key", and forget about the source of information (Bob) i.e.,

$$K_A(K_B\ \neg K_E\ k) \Longrightarrow\ K_A\ \neg K_E\ \ k.$$

- Alice can also make deductions of the type "since Eve does not know the secret key, and one would need to know the key in order to recover the encrypted message m, I conclude that Eve cannot know the secret message," i.e.,

$$K_A[(\neg K_E\ k) \wedge (\neg K_i\ k\ \Longrightarrow\ \neg K_i\ m,\ \forall\ i)]\ \Longrightarrow\ K_A \neg K_E\ m.$$

Agents as physical systems

A "physical system" is any object of a physical study and can be characterized, according to a theory $\mathbb{T}$, by a set of possible states $\mathscr{P}_S$ and a set of allowed operations, $\mathscr{O}_S$ on these states i.e., : $O_S : \mathscr{P}_S \mapsto \mathscr{P}_S$ for $O_S \in \mathscr{O}_S$.

Definition (Agents)

A physical setting may be associated with a set $\mathcal{A}$ of agents. An agent $A_i \in \mathcal{A}$ is described by a knowledge operator $K_i \in \mathcal{K}_{\mathcal{A}}$ and a physical system $M_i \in \mathcal{M}_{\mathcal{A}}$, which we call a "memory." Each agent may study other systems according to the theory $\mathbb{T}$. An agent's memory M_i records the results and the consequences of the studies conducted by A_i. The memory may be itself an object of a study by other agents.

Note: One agent $\neq$ one human. E.g., Alice before vs after tampering of her memory by Ursula would be considered different agents in the FR setting.

Physical theories as common knowledge

This condition incorporates the physical theory into the reasoning framework used by agents.

Definition (Common knowledge)

A physical theory shared by all agents $\{A_i\}_i$ in a given setting is a set $\mathbb{T}$ of statements that are common knowledge shared by all agents, i.e.

$$\phi \in \mathbb{T} \iff (\{K_i\}_i)^n \, \phi, \quad \forall \, n \in \mathbb{N},$$

where $(\{K_i\}_i)^n$ is the set of all possible sequences of n operators picked from $\{K_i\}_i$.

Measurement

This condition models measurement as perceived by the agent performing it.

Measurement

This condition models measurement as perceived by the agent performing it.

Definition (Measurements)

A measurement is a type of study that can be conducted by an agent A_i on a system S, the essential result of which is the obtained "outcome" $x \in \mathcal{X}_S$. If witnessed by an agent A_j, the measurement is characterized by a set of propositions $\{\phi_x\} \in \Phi$, where ϕ_x corresponds to the outcome x, satisfying:

- $K_j(K_i(\exists\, x \in \mathcal{X}_S : K_i\ \phi_x))$,
- $K_j\ K_i\ \phi_x \implies K_j\ K_i\ \neg(\phi_y), \quad \forall\ y \neq x.$

Example:

For a perfect Z measurement of a qubit, ϕ_0 may include statements like:
"the qubit is now in state $|0\rangle$; before the measurement it was not in state $|1\rangle$; if I measure it again in the same way, I will obtain outcome 0" and so on.

Memory update

This condition models measurement from the perspective of an outside agent who models the first agents' memory as a physical system.

Intuition:

In the quantum case, there was a unitary evolution of the system and memory

$$\left(\sum_{x=0}^{N-1} \alpha_x \ |x\rangle_{\text{system}}\right) \otimes |0\rangle_{\text{memory}} \to \sum_{x=0}^{N-1} \alpha_x \underbrace{|x\rangle_{\text{system}} \otimes |x\rangle_{\text{memory}}}_{=:\ |\tilde{x}\rangle_{SM}}.$$

- the set of states $\mathscr{P}_{SM} = \text{span}\{|x\rangle_{\text{system}} \otimes |x\rangle_{\text{memory}}\}_{x=0}^{N-1}$ of the system and memory post-measurement, is isomorphic to the set of states $\mathscr{P}_S = \text{span}\{|x\rangle_{\text{system}}\}_{x=0}^{N-1}$ of system alone, pre-measurement.

- That is, for every transformation ϵ_S that you could apply to the system before the measurement, there is a corresponding transformation ϵ_{SM} acting on the $\mathscr{P}_{SM}$ that is operationally identical.

Memory update

Definition (Information-preserving memory update)

Let $\mathscr{P}_S$ be a set of states of a system S that is being studied by an agent A_i with a memory M_i, and $\mathscr{P}_{SM_i}$ be a set of states of the joint system SM_i. If for a given initial state $Q^{in}_{M_i} \in \mathscr{P}_{M_i}$ of the memory, there exists a corresponding map $U^Q : \mathscr{P}_{SM_i} \to \mathscr{P}_{SM_i}$ $(\in \mathscr{O}_{SM_i})$ that satisfies the following conditions (1) and (2), then U^Q is called an information-preserving memory update.

1. *Local operations on S before the memory update can be simulated by joint operations on S and M_i after the update. That is, for all $P_S \in \mathscr{P}_S$, $O_S \in \mathscr{O}_S$, $A_j \in \mathcal{A}$, ϕ, there exists an operation $O_{SM_i} \in \mathscr{O}_{SM_i}$ such that*

$$K_j\ \phi[O_S(P_s)] \Rightarrow K_j\ \phi[O_{SM_i} \circ U^Q(P_S \parallel Q^{in}_{M_i})],$$

   where $\phi[\dots]$ are arbitrary statements that depend on the argument.

2. *The memory update does not factorize into local operations. That is, there exist no operations $O'_S \in \mathscr{O}_S$ and $O'_{M_i} \in \mathscr{O}_{M_i}$ such that*

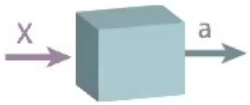
$$U^Q = O'_S \parallel O'_{M_i}$$

MEMORY UPDATE IN BOXWORLD

Generalised probabilistic theories and boxworld

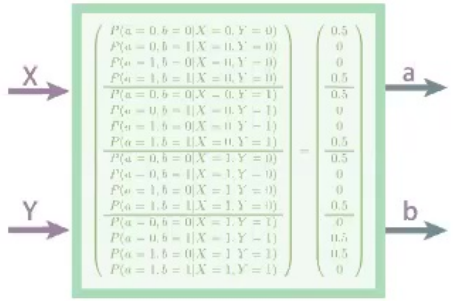
GPTs are a class of operational theories that include classical, quantum theory and boxworld.

Individual states: Characterised by probabilities of outcomes of a set of fiducial measurements.



$$\vec{P}_{gbit} = \begin{pmatrix} P(a=0|X=0) \\ P(a=1|X=0) \\ \hline P(a=0|X=1) \\ P(a=1|X=1) \end{pmatrix}$$

Composite states: $\vec{P}^{AB} = \sum_i r_i \, \vec{P}_i^A \otimes \vec{P}_i^B$ where r_i are real coefficients. E.g., PR box.

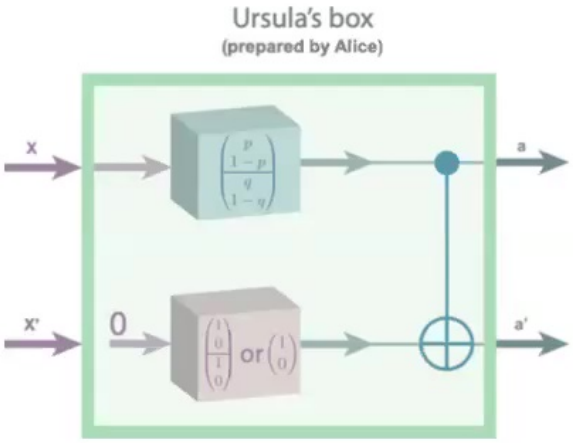


Measurements and transformations:

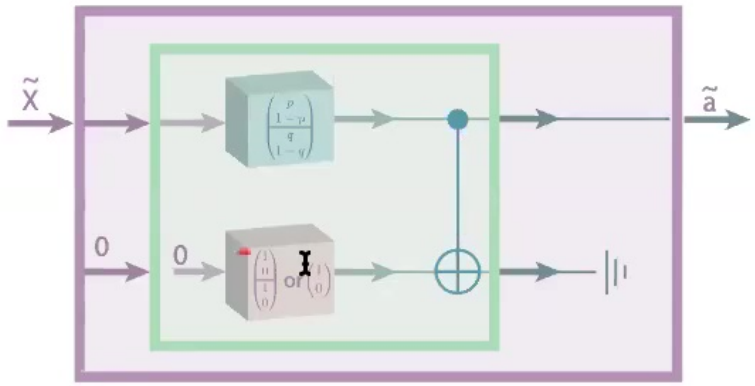
- Bipartite operations only comprise of classical input/output wirings, characterised in Barrett PRA 2007. No analogue of entangling operations such as Bell measurements.
- All reversible operations are trivial (only relabellings) as shown in Gross et. al. PRL 2010.

Possible memory update map

Theorem 1: Suppose an agent Alice measures a boxworld system S and stores the outcome in her memory M. Then, there exists an information preserving memory update transformation that describes the evolution of Alice's lab from the perspective of an outside agent, Ursula.



Outside perspective corresponds to a wiring, where the inside agent simply connected the output wires to form a CNOT gate.



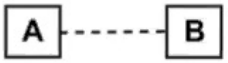
$\exists$ post-processing that Ursula can apply to "undo" the effect of Alice's measurement.

Note: Update maps that are not reversible in general can also satisfy our definition of an information preserving update.

A STRONGER POST-QUANTUM PARADOX

The protocol

(Credits: Tony Sudbery for the diagrammatic notation)



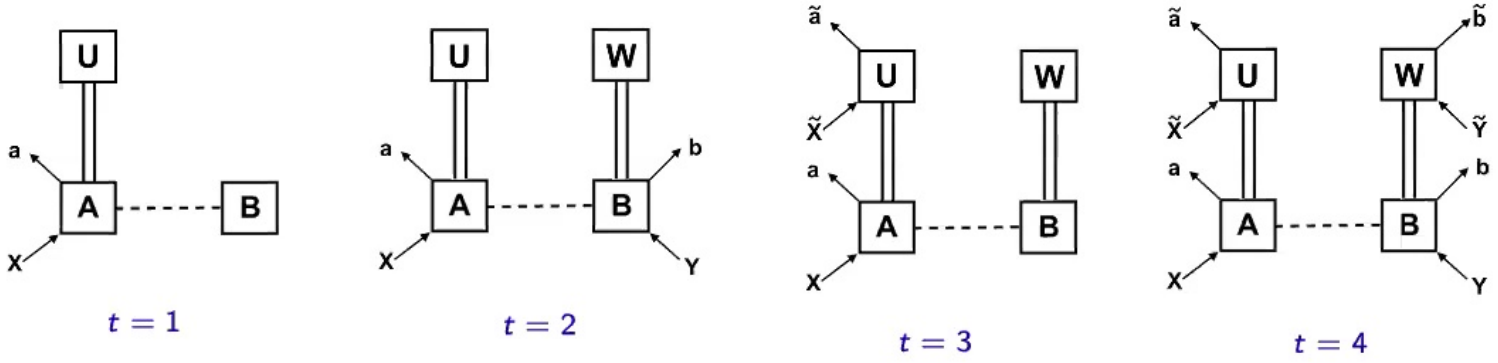
t=0 Alice and Bob share a PR box.

t=1 Alice measures setting X, and stores the outcome a in her memory A.

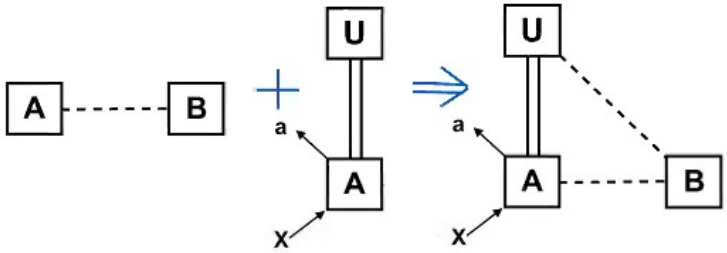
t=2 Bob measures setting Y, and stores the outcome b in his memory B.

t=3 Ursula measures Alice's lab, with setting $\tilde{X} = X \oplus 1$, obtaining outcome $\tilde{a}$.

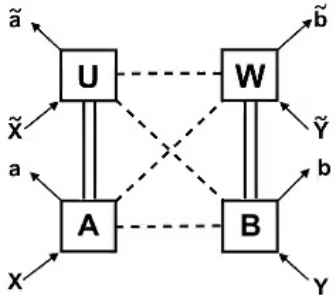
t=4 Wigner measures Bob's lab, with measurement setting $\tilde{Y} = Y \oplus 1$, obtaining outcome $\tilde{b}$.



Theorem 2: The memory update preserves the bipartite correlations i.e.,



The paradox



Trust relations:

$A_j \rightsquigarrow A_j$ (A_i trusts A_j) $\quad\Leftrightarrow\quad K_i K_j \phi \Rightarrow K_i \phi \ \forall \phi.$

$$A_{t=1,2} \leftrightsquigarrow B_{t=1,2}$$
$$B_{t=2,3} \leftrightsquigarrow U_{t=3}$$
$$U_{t=3,4} \leftrightsquigarrow W_{t=4}$$
$$W_{t=4} \leftrightsquigarrow A_{t=1}.$$

$$X.Y = a \oplus b$$
$$\tilde{X}.Y = \tilde{a} \oplus b$$
$$X.\tilde{Y} = a \oplus \tilde{b}$$
$$\tilde{X}.\tilde{Y} = \tilde{a} \oplus \tilde{b}$$
$$\tilde{X} = X \oplus 1$$
$$\tilde{Y} = Y \oplus 1$$

Taking $X = Y = 0$, $\tilde{X} = \tilde{Y} = 1$ and $\tilde{b} = 0$, we have:

$$K_W(\tilde{b} = 0 \implies \tilde{a} = 1),$$
$$K_W K_U(\tilde{a} = 1 \implies b = 1),$$
$$K_W K_U K_B(b = 1 \implies a = 1) \text{ and}$$
$$K_W K_U K_B K_A(a = 1 \implies \tilde{b} = 1).$$

Using trust relations, this gives: $K_W(\tilde{b} = 0 \implies \tilde{b} = 1)$.

A contradiction! (without post-selection)

Intuition: Contextuality of shared state gets elevated to the level of observed outcomes in a single experimental run, through the information preserving memory update.

A poetic summary

We talk, so we reason.
We reason about what we know,
We reason about what others know
And if we trust them, make their knowledge our own.

We learn, so we store.
We store our knowledge in a part of our memory.
We model that memory by a physical theory.
But if that theory is quantum, this leads to an inconsistency. ——— *FR 2018*

We learn, so we wonder.
Which theories lead to such inconsistensies,
Between reasoning agents and their memories modelled "physically"?
Proposing a framework, and in it a map for update of memories, ——— *"Results 1 and 2"*
An example we find in box world, a GPT, ——— *"Result 3"*
where using a PR box, agents find a stronger paradox.

We answer questions, so we question more.
What properties of theories lead to these paradoxes galore?
Seems not to be reversibility, but an information preserving quality, ——— *Conclusions*
And along with it contextuality,
Though, which forms of it, we are not yet sure
That, in future work we shall explore!

Open questions

- **Interpretations** of boxworld based on our generalised no-go theorem?
- **Relations to contextuality**
  - Contextuality hierarchy (Abramsky and Brandenburger NJP 2011), logical Bell ineqs (Abramsky and Hardy PRA 2012).
  - Logical pre-post selection paradoxes (Leifer and Pusey EPTCS 2015).
  - No-go theorems for objectivity of facts (Č. Brukner Entropy 2018, K. W. Bong et. al. Nat. Phys. 2020).
- **Notion of causality** in settings where agents are treated as physical systems? Framework lacking!
- More general **structure of logic** applicable to these settings?
- **Practical realisation** of FR type thought experiments using quantum computers? Implications for **measurement problem**?

THANK YOU!