IOP Publishing New J. Phys. 25 (2023) 023018 https://doi.org/10.1088/1367-2630/acb3ef

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
2 August 2022

REVISED
16 January 2023

ACCEPTED FOR PUBLICATION
17 January 2023

PUBLISHED
14 February 2023

PAPER

Toys can't play: physical agents in Spekkens' theory

Ladina Hausmann*, Nuriya Nurgalieva* and Lídia del Rio

Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland
* Authors to whom any correspondence should be addressed.

E-mail: hladina@phys.ethz.ch and nuriya@phys.ethz.ch

Keywords: epistemic theories, logical paradoxes, physical agents, Spekkens' toy theory

Abstract

Information is physical (Landauer 1961 *IBM J. Res. Dev.* **5** 183–91), and for a physical theory to be universal, it should model observers as physical systems, with concrete memories where they store the information acquired through experiments and reasoning. Here we address these issues in Spekkens' toy theory (Spekkens 2005 *Phys. Rev.* A **71** 052108), a non-contextual epistemically restricted model that partially mimics the behaviour of quantum mechanics. We propose a way to model physical implementations of agents, memories, measurements, conditional actions and information processing. We find that the actions of toy agents are severely limited: although there are non-orthogonal states in the theory, there is no way for physical agents to consciously prepare them. Their memories are also constrained: agents cannot forget in which of two arbitrary states a system is. Finally, we formalize the process of making inferences about other agents' experiments and model multi-agent experiments like Wigner's friend. Unlike quantum theory (Nurgalieva and del Rio Lidia 2019 *Electron. Proc. Theor. Comput. Sci.* **287** 267–97; Fraser *et al* 2020 Fitch's knowability axioms are incompatible with quantum theory arXiv:2009.00321; Frauchiger and Renner 2018 *Nat. Commun.* **9** 3711; Nurgalieva and Renner 2021 *Contemp. Phys.* **61** 1–24; Brukner 2018 *Entropy* **20** 350) or box world (Vilasini *et al* 2019 *New J. Phys.* **21** 113028), in toy theory there are no inconsistencies when physical agents reason about each other's knowledge.

> There were realities the human mind was never meant to withstand, pressures it was never meant to survive. Knowledge is like the sea. Go too deep, and the crushing weight of it could kill you.
>
> Seanan McGuire, *Laughter at the Academy*

> You… are… a… toy!!! You aren't the real Buzz Lightyear, you're an… aw, you're an action figure! You are a child's… plaything!
>
> Toy Story

Contents

1. Introduction and summary of the toy theory

1.1. Physical information

Physical theories that aim to describe the world at macroscopic scales should ideally be able to physically model observers, the experiments they perform, and their reasoning within the theory. For instance, the information acquired by agents must be stored in some physical form, in systems that we call memories. These can encompass biological brains but also classical and quantum computer memories, or even just measurement devices. Processing that information is ultimately a physical process, and manipulations of an agent's memory do not simply result abstract epistemic changes of their knowledge, but also in concrete physical changes. For example, quantum measurement schemes [1] show us that the process of acquiring information about a system entangles the system with our physical memory, and Landauer's principle [2, 3] tells us that forgetting that information is to shuffle those correlations to the environment, at a thermodynamic cost.

1.2. Abstract logic

Another feature that physical theories should satisfy is to allow for the information contained in the memories to be operated according to simple reasoning principles. Ideally we would like inferences such as 'if Alice knows that a is true, and she knows that a implies b, then she knows that b is true' to hold independently of the physical origins of a and b[1]. In other words, an abstract system of *epistemic logic* should ideally be applicable to any physical setting within the theory [4, 5].

1.3. Tension between abstract logic and physical information

It was shown that for some theories where both of of these requirements are satisfied—where agents reason about each other's knowledge and are themselves modelled as physical memories within the scope of the model—experience inconsistencies. For example, this is the case for the quantum theory and generalized probability theories (in particular, so-called box world), where agents applying standard logic to reason about physical experiments can come to contradictory conclusions [6, 7]. Our ultimate goal is to understand which classes of theories exhibit similar incompatibility between multi-agent logic and physics. In previous work [7], we discuss the elements needed to define a reasoning agent for an arbitrary physical theory (including physical descriptions of agents' memories and measurements, subjective state update rules, and how these interface with an abstract logic system). In this work we analyse these concepts specifically in Spekkens' toy theory.

1.4. Spekkens' toy theory subsection

Spekkens' toy theory is an *epistemically-restricted theory* [8–15]. Such theories distinguish two types of states: ontic states, which encode the physical state of a system, and epistemic states—the states of knowledge that an observer has about the system. The theory imposes restrictions on agents' knowledge (in Spekkens' case, an agent can only have access to half of the total information about the ontic state). The evolution of an ontic state is governed by the dynamics of underlying ontic theory, whereas the epistemic state is on top subject to particular rules, for example for updating after a measurement. For a detailed description of the formalism of the toy theory, we refer the reader to our review [16].

1.5. Contribution and structure

In section 1.6, we provide a minimal summary of the essence of Spekkens' toy theory. In section 2, we discuss the physical implementations of measurements in quantum theory and their analogues in the toy theory, and analyze restrictions on agents' choices in conditional preparation scenarios in the toy theory in section 3. In section 4, we consider the process of forgetting, and show that not all expressions of ignorance are valid in the toy theory. We discuss how agents can make inferences in section 5, and consider the consequences of our results for thought experiments like the Frauchiger–Renner setup [17], formulated for the case of the toy theory. We summarize our conclusions and outline future directions in section 6. To keep the main manuscript light, we present only intuitive examples in low dimensions; our general results hold for arbitrary dimensions, and are formalized in the appendix.

[1] For example, Alice knows that (a) it is raining, and (b) whenever it rains outside, Bob carries an umbrella with him. She concludes that Bob indeed is walking with an umbrella today. Here Alice is able to combine her knowledge of Bob's reasoning (b)) with her own observation (a) to infer Bob's behaviour. We will examine quantum examples later in the manuscript.

1.6. Minimal summary of the Spekkens' toy theory

1.6.1. Knowledge balance principle

Spekkens' toy theory is an **epistemically-restricted theory** [8]. Such theories distinguish two types of states: **ontic states**, which encode the physical state of a system, and **epistemic states**—the states of knowledge that an observer has about the system. In the toy theory, epistemic states are constrained by the *knowledge balance principle*, inspired by the Heisenberg uncertainty principle:

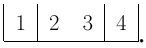
> 'If one has maximal knowledge, then for every system, at every time, the amount of knowledge one possesses about the ontic state of the system at that time must equal the amount of knowledge one lacks' [18].

In practice, suppose that a system can be in one of $2d$ possible ontic states, that is, specifying the value of $\log_2(2d) = 1 + \log_2 d$ degrees of freedom is sufficient to identify each ontic state. Then an observer is only allowed to access at most $\log_2 d$ bits of information, and we say that the system *has dimension d*; this and other constraints will become clearer with the examples ahead.

1.6.2. System dimensions

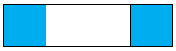
Our results apply to the general case where a toy system can be decomposed into N subsystems of dimension d each. The continuous limit, analogous to a particle moving in space, corresponds to $d \to \infty$. For example, the elementary systems analogous to qubits (toy bits) have $d = 2$. In the following we stick to intuitive visualizations for small discrete dimensions (one or two toy bits); we review the general case in appendix A.

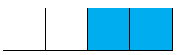
1.6.3. Notation for toy bits

We can represent up to two discrete toy systems via simple grid diagrams. For higher dimensions, grid diagrams are not as useful; the interested reader may find a review of the necessary mathematical notation in the appendix. A single system with two degrees of freedom ($d = 2$) has four different ontic states (labelled for example 1,2,3,4). To see this, note that answering two binary questions would be sufficient to identify the ontic state of the system, for example 'is the ontic state odd?' and 'is the ontic state smaller than 3?'. Each ontic state is represented by one of four boxes, 

. To represent an observer's knowledge about the ontic state—that is the epistemic state from their perspective—we colour in some of the boxes. For example, $\{1,2\} =$ 

represents 'the observer knows that the system's ontic state is either $o = 1$ or $o = 2$.' For simplicity we usually omit the number labels. One can draw an analogy between toy states and quantum states:

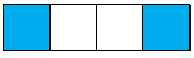
$|0\rangle \sim$  $= \{1,2\}$, $|+\rangle \sim$  $= \{1,3\}$, $|+i\rangle \sim$  $= \{1,4\}$,

$|1\rangle \sim$  $= \{3,4\}$, $|-\rangle \sim$  $= \{2,4\}$, $|-i\rangle \sim$  $= \{2,3\}$.

These are the only pure states at $d = 2$: they are states of maximal allowed information, according to the knowledge balance principle, for which the observer knows half of the degrees of freedom (for example  represents 'the ontic state is odd'). The quantum analogy carries through to a stabilizer formulation of the toy theory [9]. The fully mixed state (a state of maximal ignorance) is represented as the mixture:

$\{1,2,3,4\} =$  $=$  $\vee$  $\sim |0\rangle\langle 0| + |1\rangle\langle 1|$

$=$  $\vee$  $\sim |+\rangle\langle +| + |-\rangle\langle -|$

$=$  $\vee$  $\sim |i\rangle\langle i| + |-i\rangle\langle -i|.$

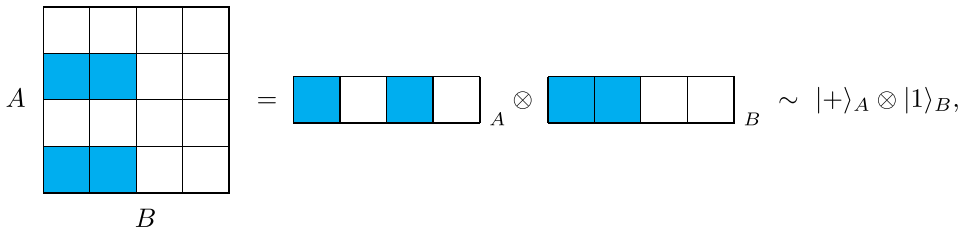
Unlike quantum theory, this is the only physical mixed state at $d = 1$, that is the only mixed state that can emerge as a marginal of a globally pure state (like the entangled state which we will consider in a moment). The epistemic restriction implies that for a system composed of N elementary systems, an agent is only allowed to have access to exactly $0 \leqslant J \leqslant N$ bits of information (corresponding to an epistemic state spanning

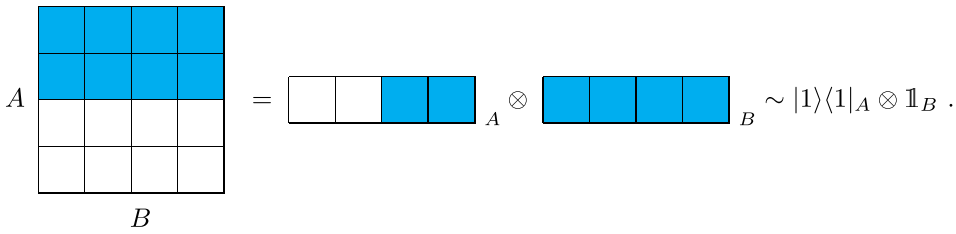
2^{2N-J} ontic states). For example, for $N = 1$ the valid epistemic states can span either 2 or 4 ontic states. The mixture:

is not a valid epistemic state [18].[2]

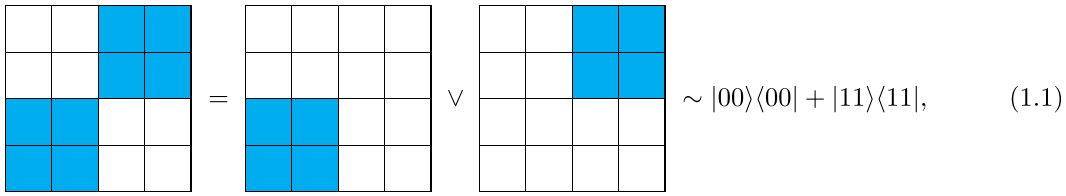
1.6.4. Composing systems

When we consider two toy systems, we represent the ontic states with a 4×4 grid, where the rows determine the possible ontic states of system A and the columns those of system B, for example:



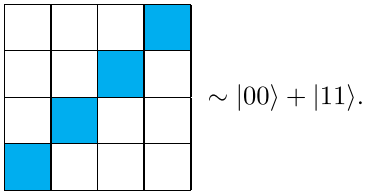


We omit the subsystem labels when they are clear from context. There are also global toy states which are not product states, for example the classically correlated state:



(1.1)

which is a mixed state, and the pure entangled state:

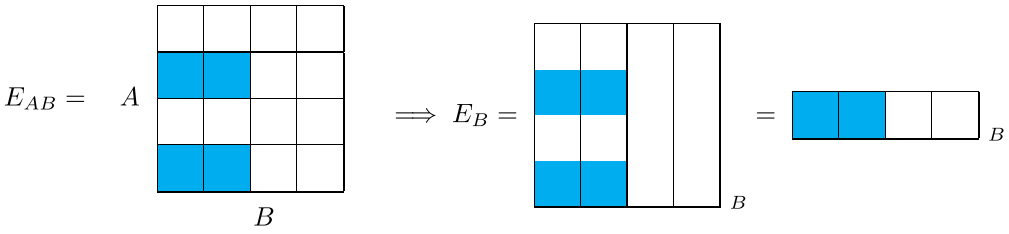


Analogously to a Bell state in quantum theory, this latter state is used for toy teleportation and dense coding protocols [18]. Like their quantum analogues, entangled toy states are globally pure states with mixed marginals. In this example, the observer has maximal information about the correlations between A and B, and maximal ignorance about the reduced state of individual subsystems.

1.6.5. Reduced states

For discrete toy systems, taking the reduced state over one system corresponds to projecting the grid diagram into one axis. For example,

[2] At the time of writing (late 2022), Spekkens and colleagues are investigating how to articulate a layer of (Bayesian) probabilistic knowledge on top of these epistemic 'physical' states. If successful, this would allow for Bayesian states whose measurement statistics are identical to the illegal 'physical' epistemic state  , at least for local measurements. We leave it as future work to study the consequences and stability of that approach..

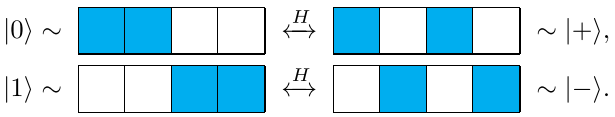


More formally, if the global epistemic state is E_{AB}, the reduced (or marginal) state on system B is $E_B = \{o_B : (o_A, o_B) \in E_{AB}\}$. In the above example, $E_{AB} = \{(1,1),(1,2),(3,1),(3,2)\}$ so $E_B = \{1,2\}$.

1.6.6. Notation for transformations

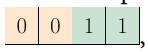
For discrete dimensions, the allowed toy transformations are permutations of ontic states, constrained to mapping valid epistemic states to valid epistemic states (in all subsystems). For example consider the transformation that permutes the second and third ontic states,

This permutation acts analogously to the Hadamard gate in quantum theory,



We will see other examples (like a CNOT gate) further ahead; in particular we will see that a controlled Hadamard is not a valid operation, and we will explore the implications of this for agents' free choice.

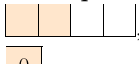
1.6.7. Notation for measurements

In the toy theory, a measurement is a partition of the ontic state space into valid epistemic states, called the measurement basis. The outcome of the measurement is determined by the position of the ontic state. The observer can then update their knowledge; in the toy theory the measurement update rule leads to an ontic disturbance, which we will later see emerges from the physical implementation of a measurement. The consequence for the updated epistemic state is a bit cumbersome to explain, but will become clear from examples right ahead. In short, the updated description should guarantee that if the observer repeats the measurement they obtain the same outcome, and be compatible with their overall knowledge [18]. For example, consider the measurement: $\mathcal{M}_Z =$ 

, where the numbered partitions correspond to the epistemic states of the measurement basis. This is analogous to a Pauli-Z measurement of a single qubit. Suppose that Bob measures a toy bit in state  in this basis, and obtains outcome 

. This allows him to deduce that the ontic state previous to the measurement was 0; however

is not a valid epistemic state. Rather, it is a **pre- and post-selected state**, a concept that has analogues in quantum theory. The smallest epistemic state compatible with his knowledge of the pre- and post-selection and with the requirement for repeated outcomes is . To see this, first note that there are four epistemic states compatible with Bob's knowledge,

The requirement for repeated outcomes translates to 'the post-measurement description must be a subset of  , so that if I apply the same measurement again, I am guaranteed to obtain the same outcome 

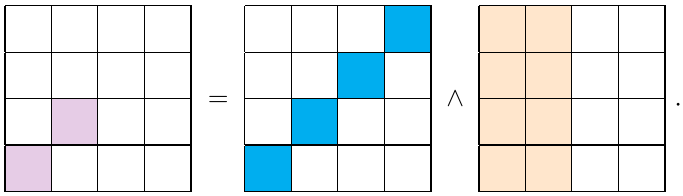
.' Of the four candidates, only the first epistemic state satisfies the condition,

This is analogous to the quantum measurement update rule for projective outcomes: if Bob measured $|+\rangle$ in the Z basis and obtained outcome 0, he would henceforth describe the state as $|0\rangle$. Now suppose that Bob is

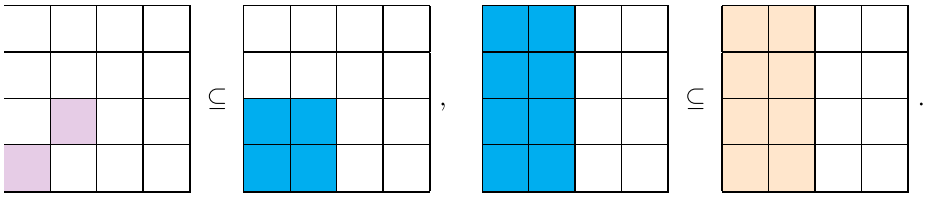
making the same measurement, but now on his half of the entangled toy Bell state (1.1). The global view of Bob's local measurement is:

$$\mathcal{I}\otimes\mathcal{M}_Z = \begin{array}{|c|c|c|c|}\hline 0&0&1&1\\\hline 0&0&1&1\\\hline 0&0&1&1\\\hline 0&0&1&1\\\hline\end{array}$$

If Bob obtains outcome $\boxed{0}$ he can deduce that the global ontic state before the measurement was one of the two coloured squares in the bottom-left corner,



This is the pre- and post-selected state that describes his knowledge of the ontic state of *AB* just before the measurement. There are two valid epistemic states that are compatible with this knowledge and live in the measurement partition $\boxed{0}$,



Of the two candidates, Bob picks the leftmost (the smallest state), which is the description that makes the most use of his knowledge about the state before the measurement (picking the state on the right would be discarding what he knew of the correlations between the two systems). The quantum analogy is when Bob makes a local Z measurement on a Bell state $\propto |00\rangle + |11\rangle$, and upon seeing outcome 0, updates his global description of the state to $|00\rangle$.

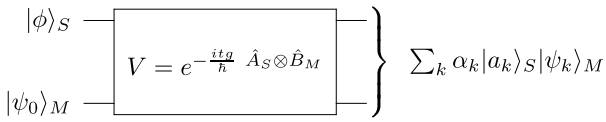
2. Learning: physical implementations of measurements

2.1. Minimal settings

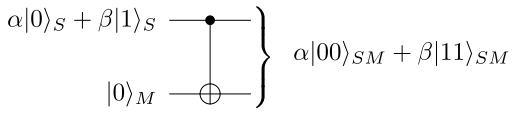
In theories where agents can improve their knowledge about the state of the system by posing questions or performing measurements on the system, they have to update their epistemic state as a consequence. Hence, if we model agents as physical systems, the bare minimum of their description has to include the degree of freedom of the memory entry corresponding to the system being inquired by the agent. We call the process of revising the memory entry after a measurement a *memory update*. In this section, we consider how one can model the process of learning—measuring a system and registering the outcome in a memory—in the toy theory. It is in these settings that we find the most dramatic differences between quantum and toy theory. At every step, we start by reviewing the quantum process, and then try to find an analogous in the toy theory.

2.2. Quantum measurements as entangling operations

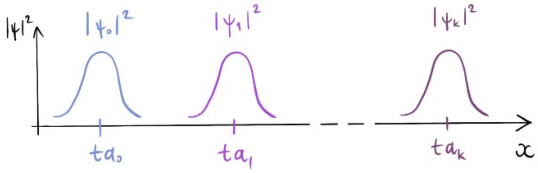
In quantum mechanics, measurements are entangling processes between the system of interest and a measurement device. For example, in the Stern–Gerlach experiment, the internal spin of a particle is coupled



(a) **Von Neumann measurement scheme [9].** An observable $\hat{A}_S = \sum_k a_k|a_k\rangle\langle a_k|$ is measured by coupling the system of interest S to a pointer M through the Hamiltonian $\hat{H} = g\ \hat{A}_S \otimes \hat{B}_M$. If the original state of the system is $|\phi\rangle_S = \sum_k \alpha_k|a_k\rangle_S$, then the final state is $\sum_k \alpha_k|a_k\rangle_S \otimes |\psi_k\rangle_M$, where $|\psi_k\rangle_M = e^{\frac{-itga_k}{\hbar}\hat{B}_M}|\psi_0\rangle_M$.



(b) **Discrete quantum measurement.** A qubit S is measured in computational basis: it is coupled to another qubit M, acting as a memory system, by a CNOT gate. The operation coherently copies the state of S to M.



(c) **Continuous quantum measurement.** If M is a continuous system and $\hat{B}_M = \hat{P}_M$ is the momentum operator and then each $|\psi_k\rangle_M$ has the original wave function of the pointer shifted by a_k, $\psi_k(x) = \psi_0(x - ta_k)$. For strong measurements, these peaks don't overlap.

Figure 1. Quantum measurements as physical evolutions. In quantum theory, we can implement measurements as unitary processes, which entangle the state of the system measured to the measurement device in the basis of the observable. Here are two familiar examples. Zooming out to the perspective of an external agent who sees the experimenter as just another quantum system, we can include the experimenter's memory and lab as part of the measurement device, and use the same unitary view to model the whole measurement process.

to its position degree of freedom through the application of a magnetic field; it is the position of the particle that acts as a pointer or measurement device when it hits a screen, as the arrival position is correlated with the internal spin.

2.3. Quantum von Neumann measurement scheme

More concretely, we can measure a discrete observable $\hat{A}_S = \sum_k a_k|a_k\rangle\langle a_k|_S$ on a system S by coupling it to a measurement device (or pointer) M, through a Hamiltonian of the form $H = g\,\hat{A}_S \otimes \hat{B}_M$, where $\hat{B}_M$ is a suitable observable on M and g a tunable constant [1]. Letting the two systems evolve for some time t corresponds to the reversible evolution modelled by the unitary $V = e^{\frac{-it}{\hbar}H}$. If S is initially in an arbitrary state $|\phi\rangle_S = \sum_k \alpha_k|a_k\rangle_S$ (with unknown coefficients $\alpha_k = \langle a_k|\phi\rangle_S$), and we prepare the measurement device (or 'pointer') in state $|\psi_0\rangle_M$, then after time t the global state becomes:

$$e^{\frac{-it}{\hbar}H}(|\phi\rangle_S \otimes |\psi_0\rangle_M) = e^{\frac{-igt}{\hbar}\ \hat{A}_S\otimes\hat{B}_M} \sum_k \alpha_k|a_k\rangle_S \otimes |\psi_0\rangle_M = \sum_k \alpha_k|a_k\rangle_S \otimes \underbrace{e^{\frac{-itga_k}{\hbar}\hat{B}_M}|\psi_0\rangle_M}_{|\psi_k\rangle_M}.$$

This simple unitary evolution can be modified to account for noise, finite-size effects, coarse and continuous observables and other corrections, in order to cover realistic implementations of quantum measurements. A relevant remark for later is that ultimately, the observer's memory is itself a quantum system that becomes entangled with the system measured—at least from the perspective of an external agent.

2.4. Quantum examples

Two examples for continuous and discrete measurements are summarized in figure 1. The simplest case is when both the system to measured and the pointer are single qubits (figure 1(b)). An entangling CNOT gate[3] between S and M implements a strong measurement of S in the Z basis:

$$(\alpha|0\rangle_S + \beta|1\rangle_S) \otimes |0\rangle_M \quad \xrightarrow{V = \text{CNOT}} \quad \alpha|0\rangle_S|0\rangle_M + \beta|1\rangle_S|1\rangle_M \tag{2.1}$$

A familiar example in continuous systems is the position measurement of a particle, which entangles the particle and pointer in the position basis. This is achieved for example by setting $\hat{A}_S = \hat{X}_S$ and $\hat{B}_M = \hat{P}_M$, and initializing the pointer to a well-localized state (like a Gaussian wave[4]). The measurement process results in the physical evolution:

[3] An interaction Hamiltonian that implements this gate in a suitable amount of time (e.g. $t = \pi$) is for example $H_{\text{int}} = \frac{1}{4}Z'_S \otimes X'_M$, where Z'_S and X'_M are shifted Z and X operators, $Z'_S = Z_S - \mathbb{1}_S$, $X'_M = X_M - 3\mathbb{1}_M$.

[4] This measurement can be made sharper or weaker by tuning the parameters of the initial wave function of the pointer and the interaction time.

$$\left(\int dx\, \phi(x)\, |x\rangle_S\right) \otimes \left(\int dx'\, \psi_0(x')\, |x'\rangle_M\right) \quad \xrightarrow{V} \quad \int dx \int dx'\, \phi(x)\, \psi_0(x' - tgx)\, |x\rangle_S |x'\rangle_M.$$

2.4.1. Emergence of the post-measurement state

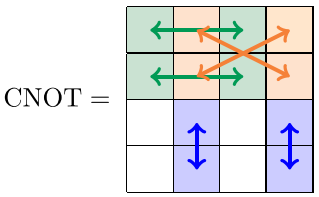
In quantum theory, the post-measurement state of a system S emerges from the physical picture by moving the Heisenberg cut one level up and thinking of a projective measurement on the agent's memory M after the physical entangling operation between S and M. For example, the (unnormalized) post-measurement state when obtaining outcome 1 on a Z measurement of a qubit in state $|\psi\rangle_S$ can be expressed as:

$$\mathrm{Tr}_M \left[\underbrace{\mathbb{1}_S \otimes |1\rangle\langle 1|_M}_{\text{measuring memory}} \underbrace{\mathrm{CNOT}\ (\ |\psi\rangle\langle\psi|_S \otimes |0\rangle\langle 0|_M)\ \mathrm{CNOT}}_{\text{entangling } S \text{ and memory}} \right] = \underbrace{|1\rangle\langle 1|_S}_{\text{measuring } S}\ |\psi\rangle\langle\psi|_S.$$

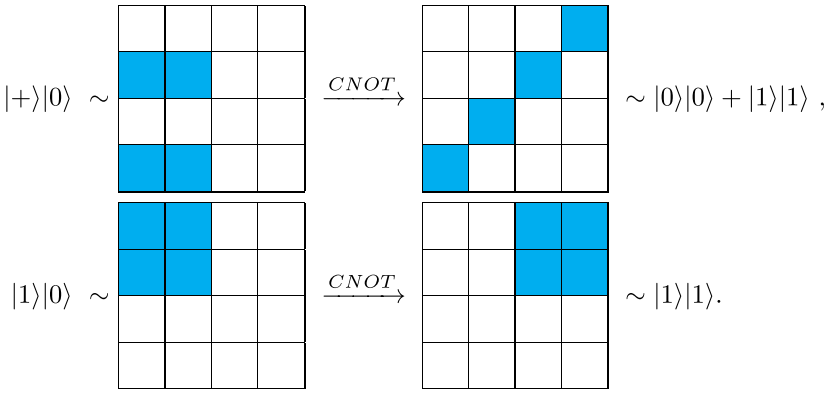
For details on the general case we refer to appendix C.1.

2.5. Measurements in the toy theory: discrete example

We can now try to bring over the same concept of physical measurement to the toy theory. Let us consider two separate toy bit systems: the first one is the system we measure, the second is the memory where the outcome is registered. We can proceed analogously to the quantum case, as there is an allowed transformation corresponding to the CNOT gate in the original toy theory [8]. In the notation of grid diagrams it permutes the ontic states as:



Analogously to the quantum CNOT gate, the toy CNOT transformation correlates two toy-bits, for example:



Once again, the post-measurement state of the system emerges from conditioning on a projective measurement on the memory, and then finding the reduced state.

2.6. Measurements in the toy theory: general case

For the general case of arbitrary continuous or discrete dimensions, we can also find global transformations on the system and pointer that implement any valid measurements (theorem B.6). The transformation is essentially analogous to the von Neumann measurement procedure, correlating the two toy subsystems in a way that reflects the properties of the observable, and recovers the post-measurement state of the system measured when we condition on the outcome. There are a few extra constraints (for example, on how the dimensions of system and pointer should match), but the main qualitative difference to quantum measurements comes from the fact that toy observables are already restricted at the abstract level, as we saw in the introduction. See appendix B.2 for examples in continuous dimensions (like a position measurement) and detailed derivations of the formal results.

3. Acting: restrictions on agents' choices

3.1. Choices start with measurements

If we model agents' actions as physical processes, we see that they can always be decomposed as a measurement followed by a conditional transformation. The process of deciding which of a series of actions to take is ultimately a measurement—of one's memory, of a randomness generated, or any relevant external systems. We look at the weather to decide what to wear, consult our agenda to decide on appointments, and even when making random decisions we can model our source of randomness as an explicit physical system. Taking this to the extreme, when we try to implement a statement like 'a system S can be prepared in one of many states $\{\psi_k\}_k$', whoever chooses the state k does it by measuring another system, obtain an outcome k, and then apply a physical transformation on their own memory and S that prepares the state. We will see that in Spekkens' toy theory, physical agents are dramatically restricted in this action. Indeed, only agents outside the toy theory can prepare non-orthogonal states. In other words, toys cannot play.

3.2. Quantum conditional preparation scenarios

First consider a simple quantum measure-and-prepare scenario: Alice measures a state $|\phi\rangle_R = \sum_k \alpha_k |a_k\rangle_R$ and, depending on her observed outcome k, prepares another system S in state $|\eta_k\rangle_S$. This procedure can be described from the outside as a global two-step unitary process (figure 2). First, Alice's measurement is modelled by a unitary V that couples R to Alice's memory A; This is followed by another unitary U, which implements the conditional state preparation, correlating S with A. In the case of a strong measurement ($\langle\psi_k|\psi_\ell\rangle_A = \delta_{k\ell}$), the overall transformation acts as:

$$\left(\sum_k \alpha_k |a_k\rangle_R\right) \otimes |0\rangle_A \otimes |0\rangle_S \xrightarrow{V} \left(\sum_k \alpha_k |a_k\rangle_R \otimes |\psi_k\rangle_A\right) \otimes |0\rangle_S \xrightarrow{U} \sum_k \alpha_k |a_k\rangle_R \otimes |\psi_k\rangle_A \otimes |\eta_k\rangle_S.$$

After a strong quantum measurement ($\langle\psi_k|\psi_\ell\rangle_A = \delta_{k\ell}$), Alice is not restricted in the states $\{|\eta_k\rangle_S\}_S$ that she can conditionally prepare. In particular, she can conditionally prepare non-orthogonal states: for example, she can prepare $|0\rangle_S$ if she observes outcome 0, and $|+\rangle_S$ if the outcome is 1:

$$(\alpha|0\rangle_R + \beta|1\rangle_R)|0\rangle_A|0\rangle_S \xrightarrow{V} (\alpha|0\rangle_R|0\rangle_A + \beta|1\rangle_R|1\rangle_A)|0\rangle_S \xrightarrow{U} \alpha|0\rangle_R|0\rangle_A|0\rangle_S + \beta|1\rangle_R|1\rangle_A|+\rangle_S.$$

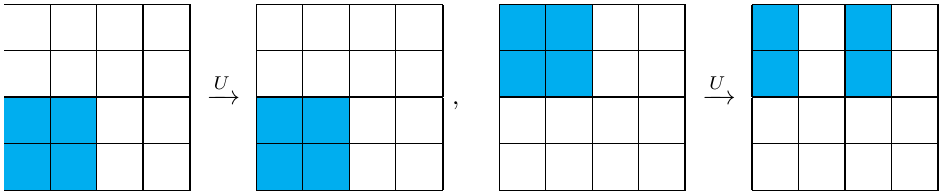
In contrast, the toy theory imposes unexpected constraints, and this kind of conditional preparation is not allowed.

3.3. Forbidden conditional preparations in the toy theory

Conditional preparations of non-orthogonal states are not allowed in the toy theory. The problem is not in Alice's measurement (which we have seen are very similar to quantum measurements), but in the second step, U, where Alice performs the conditional preparation of non-orthogonal states. This transformation would have to act on the joint state of A and S analogously to:

$$\begin{aligned} U: \quad & |0\rangle_A|0\rangle_S \to |0\rangle_A|0\rangle_S, \\ & |1\rangle_A|0\rangle_S \to |1\rangle_A|+\rangle_S, \end{aligned}$$

which in the toy theory would look like:



There are no allowed transformations that implement this action in the toy theory, even when we consider transformations on a larger system, which are irreversible at this scale (corollary 1). The generalization of this example is a dramatic restriction on agents' actions; see theorem B.8 for the formal version of this result. This theorem applies to arbitrary system dimension and arbitrary amount of systems, that can later be traced out.

Theorem 3.1 (Restrictions on conditional action of agents in the Spekkens' toy theory). *In Spekkens toy theory, if an agent measures a system R, obtaining outcome k, and prepares a second system S in one of several*

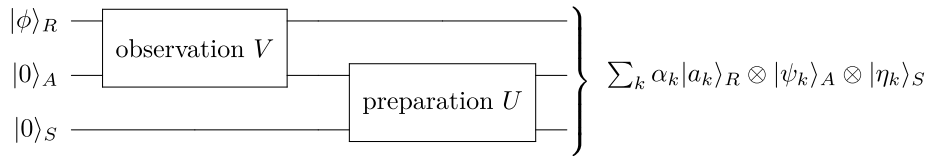


Figure 2. Conditional preparation. Alice first observes the system R, writing down the outcome to her memory, and then initializes the state of the system S depending on her outcome.

states $\{\psi_k\}_k$ depending on the outcome k, then any two of these states (ψ_k, ψ_ℓ) must be either identical or orthogonal. The number of identical states of each type is the same.

4. Forgetting: valid expressions of ignorance

4.1. Forgetting with explicit quantum memories

The physical process of forgetting information originally stored in a memory can be modelled through an interaction between the memory and its environment; this may lead to the loss of correlation between the current memory content and the system it referred to. For example, suppose that you write your credit card number on a notepad—the notepad is correlated with your credit card. If later the notepad is smudged, erased or burned (all physical interactions with an environment), it will no longer be perfectly correlated with the credit card. The same can be said of information stored in a (quantum or classical) hard drive which is subject to noise and decoherence originating from the interaction with its environment, as expressed by the data processing inequality. In a simple example (figure 3), suppose that the agent measures a bipartite system $S = S_1 \otimes S_2$, storing the outcome in their bipartite memory $M = M_1 \otimes M_2$, through a standard von Neumann scheme which from an outside perspective is modelled like a coherent copy operation, entangling S and M,

$$\left(\sum_{k,j} \alpha_{kj} \, |\phi_k\rangle_{S_1} |\psi_j\rangle_{S_2}\right) \otimes |0\rangle_{M_1} |0\rangle_{M_2} \longrightarrow \sum_{k,j} \alpha_{kj} |\phi_k\rangle_{S_1} |\psi_j\rangle_{S_2} \otimes |k\rangle_{M_1} |j\rangle_{M_2} =: |\Psi\rangle_{SM}.$$

Now let the memory interact with an environment system E. An example is complete thermalization of the second register, in which the environment is initially in a thermal state τ_E and the interaction swaps the state of the second register M_2 with the environment,

$$U_{ME} : \mathbb{1}_{M_1} \otimes \mathrm{SWAP}_{M_2 E}$$

After the global state evolves under $\mathbb{1}_S \otimes U_{ME}$, the final (mixed) state of S and the memory is

$$\begin{aligned} \rho_{SM} &= \mathrm{Tr}_E([\mathbb{1}_S \otimes U_{ME}] \, [|\Psi\rangle\langle\Psi|_{SM} \otimes \tau_E] \, [\mathbb{1}_S \otimes U_{ME}^\dagger]) \\ &= \sum_{k,k',j} \alpha_{kj} \, \alpha^*_{k'j} \, |\phi_k\rangle\langle\phi_{k'}|_{S_1} \otimes |\psi_j\rangle\langle\psi_j|_{S_2} \otimes |k\rangle\langle k'|_{M_1} \otimes \tau_{M_2}, \end{aligned}$$

where all the information about S_2 is lost to the environment, as can be seen from the mutual information

$$I(S_2 : M)_\rho = 0, \quad I(S_1 : M)_\rho = I(S_1 : M)_\Psi.$$

4.2. Abstract uncertainty in quantum theory

In addition to this physical process of forgetting, in quantum theory mixed states $\rho_S = \sum_i p_i \rho_i$ can be used to describe a state of knowledge where an agent has abstract uncertainty about which of the states $\{\rho_i\}_i$ describes S, and their best Bayesian guess is described by a probability distribution $\{p_i\}_i$. In quantum theory the probabilities and states in these abstract mixtures can be arbitrary, and one can always find a physical forgetting process on an explicit memory that connects the physical and abstract representations. In the toy theory (as it stands at the time of writing) descriptions of uncertainty are severely limited, perhaps because there are no natural physical sources for this uncertainty in the toy world.

4.3. Forgetting with an implicit toy memory

In the toy theory, not all states can be mixed in a way such that the resulting state is allowed by the epirestricted picture. For example, while it is possible to mix the states:
we are not allowed to mix states as:

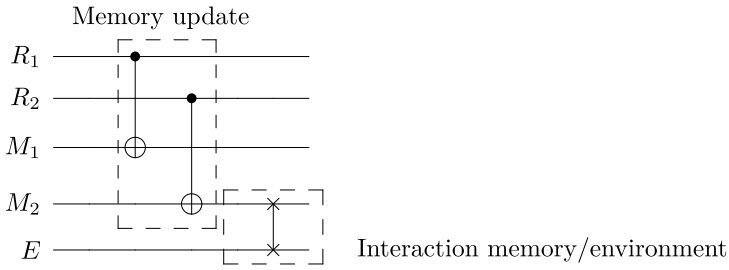
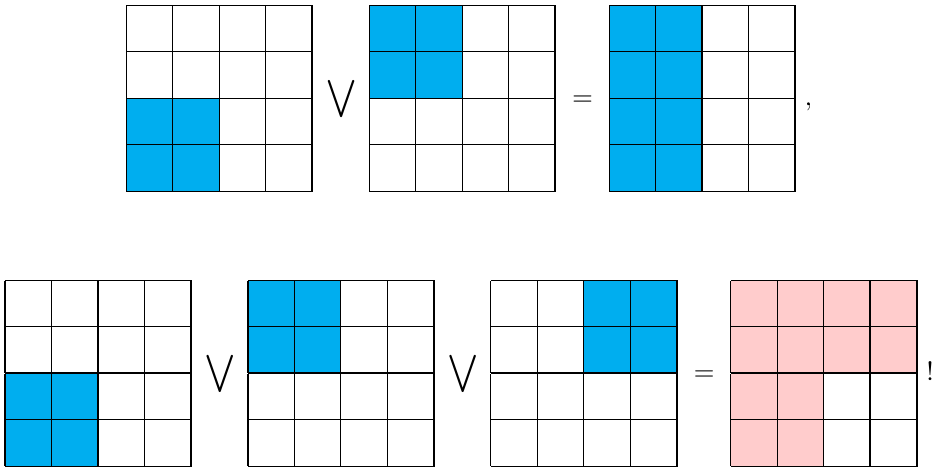


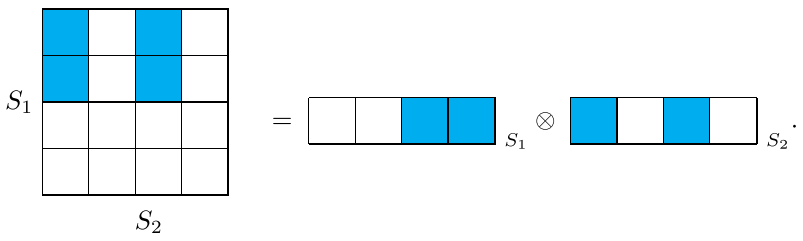
Figure 3. Losing information as a quantum evolution. One can imagine a process where the information stored in an agent's memory is partially lost to the environment. Here, the agent first performs a memory update, writing down the outcomes of measurements on systems S_1 and $S_@$ in their memory systems M_1 and M_2. Due to an interaction with the environment E (for example a complete thermalization represented here as a SWAP gate), the information contained in memory qubit M_2 is exchanged with the environment, so that correlations between the memory and S_2 are lost to the agent through this process.



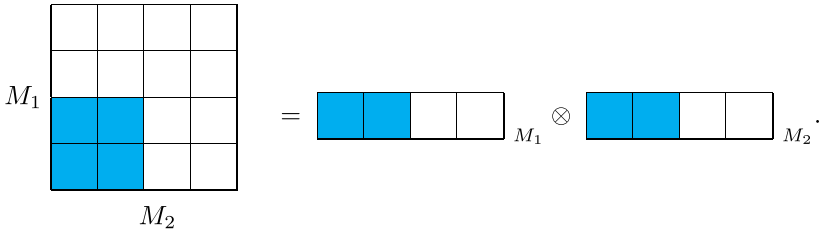
This means that for certain sets of states we are not allowed to forget which states we had initially. Moreover, even if we are able to 'forget', we only forget each state with an equal probability: the epistemic states always constitute uniform probability distributions over the corresponding set of ontic states. One can argue that only uniform distributions over the states we choose to forget (when we can) are physical, as it is not clear how one would assign non-uniform priors to the probabilities of forgetting for different states. For example, we cannot model a setting where we forget that the system is in the state $\{1,2\}$ with probability $\frac{1}{3}$, and in the state $\{3,4\}$ with probability $\frac{2}{3}$—we are only allowed to forget both states with an equal probability of $\frac{1}{2}$.

4.4. Forgetting as a physical process with explicit toy memories

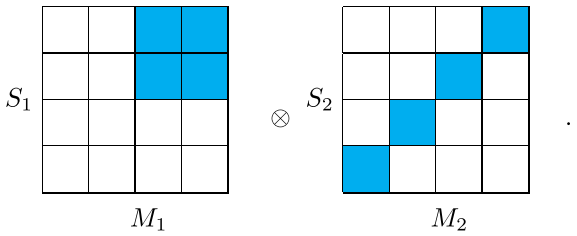
We can now see the equivalent of the quantum example where we explored the physical process of forgetting. Suppose that S starts in state:



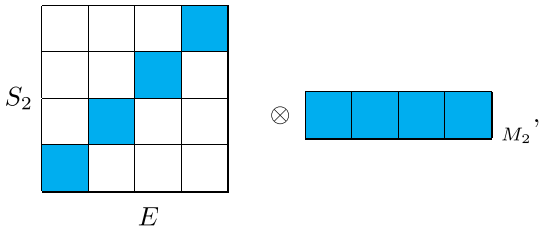
and the agent has a memory $M = M_1 \otimes M_2$ originally in state



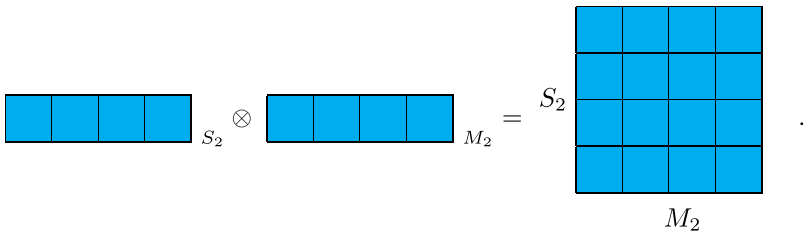
The agent measures the two toy bits of S in the Z basis, storing the outcome of the first measurement in M_1 and the second in M_2. From the outside we can model this measurement as an entangling operation $\mathrm{CNOT}_{S_1M_1} \otimes \mathrm{CNOT}_{S_2M_2}$ resulting in the global state:



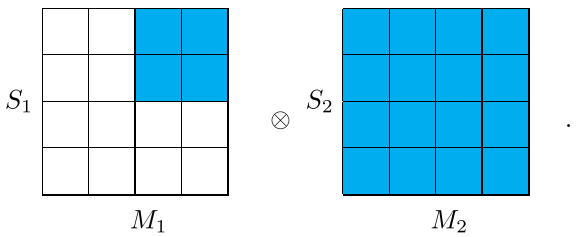
Now we simulate the decoherence process in memory M_2, through swapping with an environment in a fully mixed state, ρ_E. In the quantum analogy, this corresponds to full thermalization of M_2 with an environment at infinite temperature or with a degenerate Hamiltonian. The grid diagram of the joint state of $S_2 \otimes M_2 \otimes E$ would span three dimensions, but a convenient visualization is:



from which it is easier to see that the reduced state of $S_2 \otimes M_2$ is:



This gives us the joint state of S and M:



In this example, measuring their memory M could give the agent some information about S_1 but not about S_2 — they have forgotten about S_2. This physical picture helps us understand why forgetting in arbitrary ways is not allowed in the toy theory: to model partial ways of forgetting, we can vary the interaction V_{ME}

between memory and the environment, and the initial state of the environment. Because both transformations and epistemic states are restricted in the toy theory, the final states of $S \otimes M$ are also restricted in form. The information an agent still remembers about S after interaction with an environment can be again modelled through the reduced states of S conditioned on a measurement on the memory, and these states are, as we have seen, restricted.

4.5. Interpretation

Another intuition for why forgetting always results in a fully mixed state lies in how we understand knowledge in the toy theory [19]. The principle of knowledge balance imposes that we either possess information about the individual states of the systems, or information about correlations between them. After a physical measurement, the memory and the measured system are perfectly correlated, so from the outside we have no information about the reduced state of the system; erasing the memory effectively erases the information about correlations, leaving us maximally ignorant about the state of the measured system.

5. Reasoning: making inferences about other agents' experiments

In this section, we formalize the conditions under which agents can reason about measurement outcomes—their own and each other's. First, we look at what it means to get a certain outcome or predict it with certainty. Then, we apply this result to a particular subset of statements agents can make, namely, inferential statements. Finally, we demonstrate how these rules are applied, using examples of Bell scenario, Wigner's friend, and Frauchiger–Renner thought experiment in the toy theory, and discuss the differences from their quantum counterparts.

5.1. Deterministic predictions

In the following thought experiments, agents are able to reason about each other's outcomes—for deterministic statements. Let us formalize what it means to measure something with certainty or to predict something with certainty in the toy theory. In appendix B.3 we prove lemma B.10 which gives two conditions for an outcome of a measurement to happen with certainty. The first condition certifies that an outcome happens with certainty, while the second condition ensures that this outcome is the desired one.

5.2. Conditions for making inferences

In thought experiments we often make inferences of the type 'A $= 1 \implies$ B $= 1$', which predict other agent's measurement outcome based on our observation. How do we formalize the certainty of such an inference in the toy theory? The statement 'A $= 1 \implies$ B $= 1$' corresponds to the conditional probability $P(B = 1|A = 1) = 1$. Lemma B.11 gives two conditions for when we can make valid inferences. Intuitively, the first condition ensures that there is an outcome of the measurement of observable B that can be inferred if observable $A = 1$ is known. The second condition ensures that the outcome that can be inferred is indeed the outcome $B = 1$.

5.3. Example: a Bell scenario

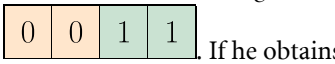
5.3.1. Quantum Bell setting

In quantum theory, if Alice and Bob share a Bell state and measure their individual qubits in the computational basis (corresponding to the observables Z_A and Z_B), they can make inferences about each other's outcomes (figure 4). For example, if the shared state is:

$$\frac{1}{\sqrt{2}}\left(|00\rangle_{AB} + |11\rangle_{AB}\right),$$

and Bob obtains outcome $b = 0$, he can update his description of the global post-measurement state to $|00\rangle_{AB}$ and infer with certainty that Alice must obtain outcome $a = 0$; analogously for $b = 1$.

5.3.2. Toy Bell scenario

This scenario can be reenacted in the toy theory, with similar results: Bob can make a deterministic prediction about Alice's outcome (figure 5). Alice and Bob share the entangled state analogous to a Bell state, Bob measures his system in the toy-Z basis, $M_Z =$ 

. If he obtains outcome $B = 0 =$, he can update his description of the shared state to

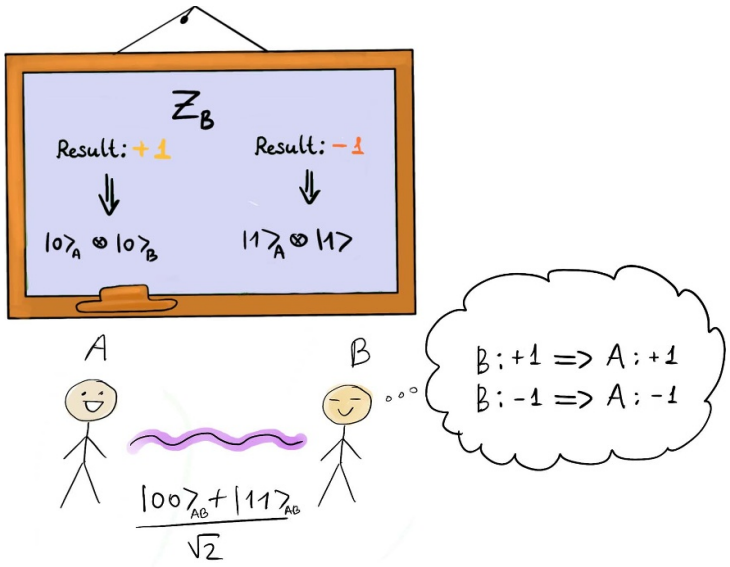


Figure 4. Reasoning in a quantum Bell experiment. Alice and Bob share a Bell state, and each perform a Z basis measurement on their qubit. Given his outcome, Bob reasons about Alice's outcome and can predict it with certainty.

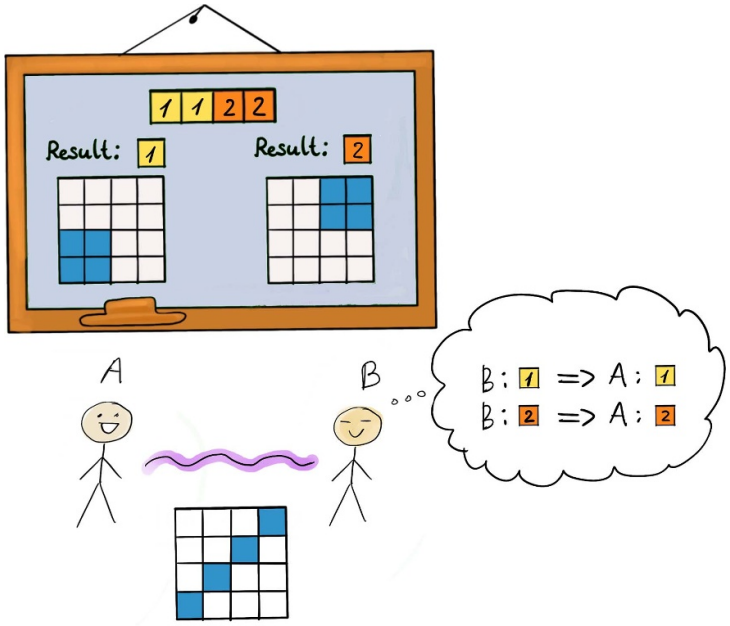
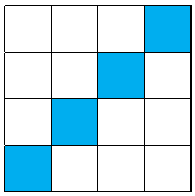
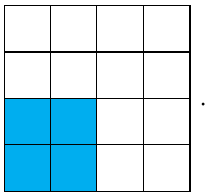


Figure 5. Reasoning in a toy Bell experiment. Analogously to the quantum Bell scenario, Alice and Bob share an entangled bipartite system and perform the toy version of local Z measurements. Based on his outcome, Bob can predict Alice's outcome with certainty.



From this description, Bob can infer that if Alice now measures her system in the same basis, she will obtain outcome $A = 0 =$  with certainty, that is, '$B = 0 \implies A = 0$' Analogously, he can conclude that '$B = 1 \implies A = 1$'. A formal proof can be found in appendix B.3.



5.4. Example of meta measurements: Wigner's friend

5.4.1. Wigner's quantum friend

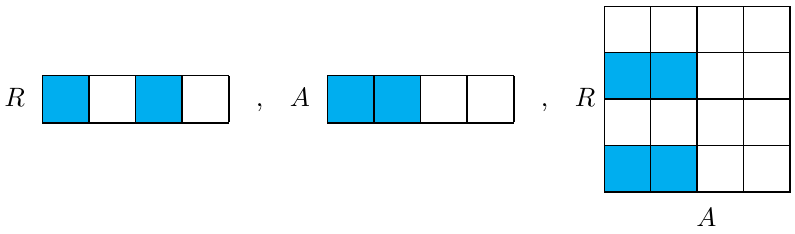
Wigner's friend experiment was first proposed by Wigner [20]. The setting involves a quantum system R and an observer A (Alice) performing a measurement on this system in a closed laboratory, as well as an outside observer Wigner. For Alice in the lab, the outcome of the experiment is recorded in the device she is using to measure the system R, for example, as a position of a pointer (or an entry in her memory). However, Wigner does not have any information about Alice getting a particular outcome, and describes the evolution of the closed lab as a unitary (reversible) process, and assigns an entangled state to R and A. In the language of quantum mechanics, if we assume that the pointer is initially in the state $|0\rangle_A$, and the state of the measured system is $\frac{1}{\sqrt{2}}|0\rangle_R + \frac{1}{\sqrt{2}}|1\rangle_R$ this process corresponds to:

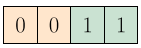
$$\left(\frac{1}{\sqrt{2}}|0\rangle_R + \frac{1}{\sqrt{2}}|1\rangle_R\right)|0\rangle_A \to \frac{1}{\sqrt{2}}|0\rangle_R|0\rangle_A + \frac{1}{\sqrt{2}}|1\rangle_R|1\rangle_A. \tag{5.1}$$

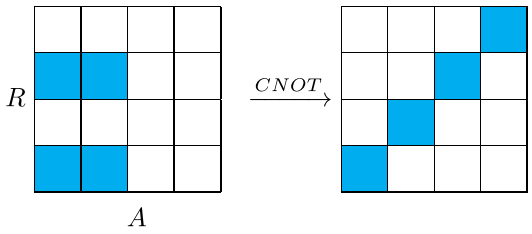
Alice and Wigner turn out to have descriptions of the same setting which are vastly different from each other. We will not discuss numerous conceptual implications of the original thought experiment here—a review can be found in [21]. However, we would still like to see how we can model this setting in the toy theory (here we will do so in the original epirestricted picture).

5.4.2. Wigner's toy friend

In the toy theory, we consider again two subsystems: the measured system R and Alice's memory register A. The individual states and the joint state of the systems R and A can be pictured as:



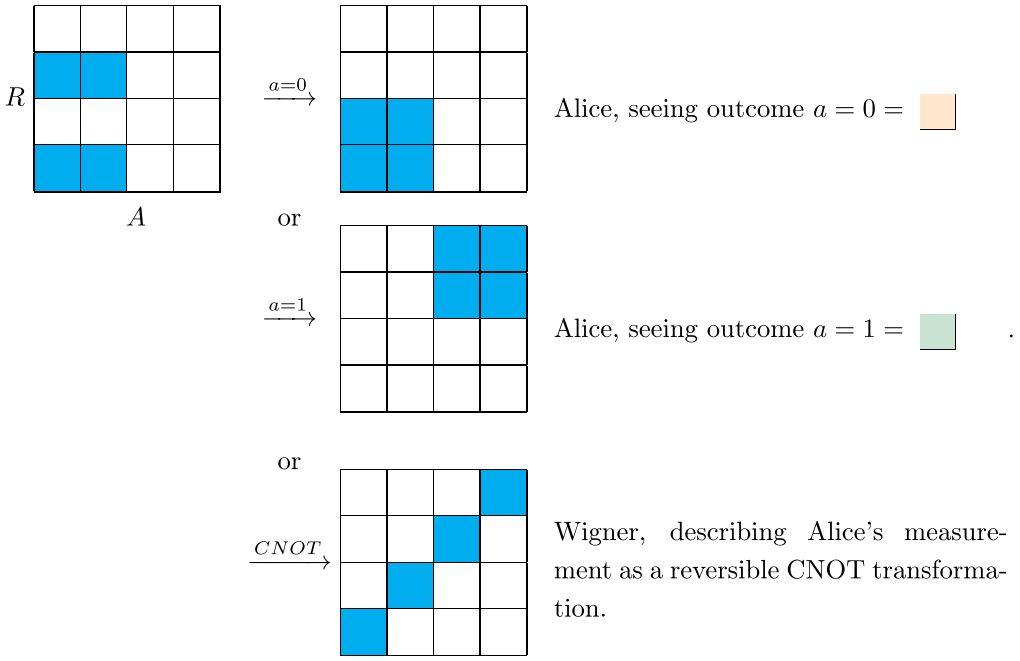
Alice measures R in the toy-Z basis. She describes her measurement as $M_Z =$  and sees a definite outcome a. Wigner sees Alice's measurement as a reversible transformation (the toy CNOT), and updates his description of the joint state of R and A as:



Alice, on the other hand, has the subjective experience of seeing one outcome a and write it to her memory. We can describe the knowledge update from the perspective of the different agents as:

5.4.3. Interpretation

In the framework of the toy theory, the difference between Alice's and Wigner's descriptions has a straightforward interpretation. Thanks to the knowledge balance principle, in the state of the maximal knowledge an agent can have maximal information either about an individual system or about how these



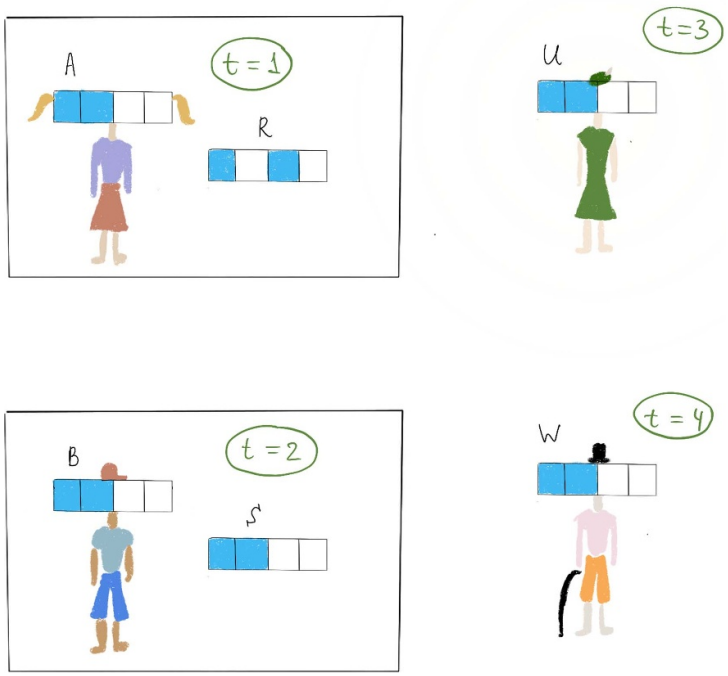


Figure 6. The Frauchiger–Renner-like setup of agents in the toy model. The setting includes four agents *A*, *B*, *U* and *W*, memories of which are all modelled as two-level systems in the toy model, and two additional two-level systems *R* and *S* in Alice's and Bob's labs respectively. In the toy version of the FR setup, we are not concerned with the course of the protocol four agents follow. The question we ask is whether there exists a joint state of systems *A*, *B*, *R* and *S*, and measurements performed on the systems *R*, *S*, *RA* and *SB*, such that the agents come to the same contradicting chain of statements as in the original quantum setting [6]. A summary of the quantum experiment can be found in appendix C.2.

individual systems are correlated. Hence, Alice's and Wigner's epistemic states do not contradict each other, and simply represent two different ways an agent can view a composite system (two states of knowledge about the same ontic state). In [19], it is shown that the Wigner and Alice's views discrepancy can be reproduced (and interpreted!) in any realistic toy model. Alice's collapsed state can simply be understood as more coarse-grained compared to Wigner's; the correlations of her state are beyond her description level. She can still get away with it, though, as the correlations are not used in later dynamics—which is not the case for the next scenario we present.

5.5. Example of multi-agent paradoxes: the Frauchiger–Renner scenario

5.5.1. Quantum multi-agent paradox

In the Frauchiger–Renner setting [17], four quantum agents (Alice, Bob, Ursula and Wigner) perform a series of measurements and reasoning steps, reaching a logical contradiction,

$$(w = \text{ok} \wedge u = \text{ok}) \implies b = 1 \implies a = 1 \implies w = \text{fail}, \tag{5.2}$$

that is, when both Ursula and Wigner obtain outcomes 'ok' in their measurements, they can reason based on their observation that that Bob predicted that Alice predicted that Wigner would obtain a different outcome, 'fail', with certainty. The experimental protocol consists of individual steps that we covered so in this manuscript: two qubits R and S are initially prepared in an entangled Hardy state [22]; Alice measures R and Bob measures S; then Ursula measures Alice's lab (including R and Alice's memory A), and finally Wigner measures Bob's lab (including S and Bob's memory B). The contradiction is found for a specific choice of initial state and measurement bases, which are described in appendix C.2.[5] For a pedagogical discussion of the original paradox and the assumptions behind it, we refer to our previous work [23]; for discussions of broader implications for abstract logic and interpretations of quantum theory see for example [5, 21, 24, 25]. The paradox has also been shown to arise in other physical theories, namely box world [7].

5.5.2. Toy multi-agent scenario

Our question is whether a similar multi-agent logical paradox can be found in Spekkens' toy theory; we will see that it cannot, partially because of the restrictions in the individual operations like conditional state preparation, and partly because this is an explicitly non-contextual epistemic theory. Since the toy theory does not allow Alice to perform non-orthognal conditional state preparation of S, we follow the version of the experiment where all relevant correlations are encoded in the initial state of RS. The global system of Alice and Bob's labs is composed of four subsystems: two systems R and S measured by Alice and Bob, and Alice's and Bob's memory registries A and B (figure 6). The order of measurements follows the original experiment, and the systems can be of an arbitrary dimension k. We show that in the toy theory, there is no choice of initial state and measurements by the four agents that can lead to a logical contradiction in this experimental scenario. A formal description of the setting and proof can be found in appendix B.4.

Theorem 5.1. *There exists no valid epistemic state that can be used to reach a logical contradiction in the Frauchiger–Renner setting in the toy theory.*

6. Discussion

6.1. Learning, reasoning and forgetting as physical processes in the toy theory

In this work, we found a way to model the physical evolution of systems and agents' memories that implement in Spekkens' toy theory the abstract proccess of measurement and forgetting information, analogously to how quantum measurements and information loss are modelled as explicit quantum evolutions. We found conditions on experimental settings that guarantee that agents can reason with certainty about each other's experiments, including settings where agents can measure each other.

6.2. Restrictions on free choice of agents

One can interpret the impossibility of an arbitrary conditional preparation in the toy theory as a limitation on the free choice of agents. An experimenter cannot decide to prepare an arbitrary valid epistemic state depending on her observations—she is constrained to an orthogonal set of states. In addition, agents cannot set arbitrary probability distributions as inputs for future experiments (because such distributions would be encoded in physical systems like biased coins). Note that agents can still perform deterministic operations that entangle other systems; they just cannot make a decision about which entangling operation to apply that does not result in uniformly-distributed orthogonal states on those systems.

6.3. In the toy theory, limited knowledge is. . . limited

While in classical and quantum theories we can always lose information, in Spekkens' toy theory it is impossible to model many natural expressions of limited knowledge, like not knowing which of two non-orthogonal states a system is in; conditional state preparation of arbitrary states is also forbidden. This shows us that even aspects of logic and information theory that we take for granted and consider

[5] In the original formulation of the experiment, Alice measures R and based on the outcome she makes a conditional preparation of S in one of two non-orthogonal states, which she sends to Bob. The two formulations (conditional preparation or initially entangled state) are equivalent for the logical analysis of the experiment, as they express the same correlations.

independent of the physical theory (like having probabilistic knowledge) are indeed dramatically physical.

6.4. Foils of the toy theory

To understand to which extent these peculiarities are an artifact of the knowledge balance principle, one could consider possible relaxations of the principle[6] — for example, imposing it for pure states but allowing all probabilistic mixtures of pure states. One must proceed with caution, as any such relaxation may have unintended consequences for the stability of the theory: for example, we considered the relaxation in the original formulation where we only require valid marginals, but not that validity is preserved under subsystem measurement, and tried to come up with a different measurement update that does not require this. However, we found that it is not possible to define such a measurement update and, additionally, found mixed epistemic states that that cannot be written as the mixture of pure states. We leave the investigation of other relaxations of the theory to future work.

6.5. Forgetting in other epistemic theories

Epistemic models provide insight into how different epistemic restrictions influence the set of transformations and measurements an agent can perform, and how their memory can be modelled. We have already mentioned epistemically restricted Liouville mechanics [26], and here we have taken a look at a particular epistemic limitation of 'knowledge balance principle'. However, one can imagine that the restrictions above can be weakened or modified. One possible direction of the future research then would be modelling memories of agents and their reasoning for an arbitrary relation between information contained in the epistemic and ontic states of the system, and see in which cases conclusions made in quantum mechanics are reproduced. This could lead to a better understanding of which types of epistemology can be admitted by quantum mechanics, and what are essential properties of such epistemic theories. We leave the investigation of how the striking limitations in the process of forgetting information plays out for other (epistemically restricted) theories as future work.

6.6. No multi-agent logical paradox in the Frauchiger–Renner setting

We proved that in settings analogous to the Frauchiger–Renner experiment there is no assignment of states and measurements that can lead to a logical paradox in the toy theory. We do not claim that our model of agents is exhaustive; in our analysis, we only capture one degree of freedom of the agent which corresponds to the memory register for the outcome of their measurement, and for this particular model (albeit a minimally reasonable one) the paradox does not come to be. We conjecture that there is no model that can lead to a paradox, in arbitrary multi-agent settings, because the theory is non-contextual; we will investigate this in future work.

6.7. Relation to contextuality

Multi-agent logical paradoxes involve chains (or possibly more general structures) of statements that cannot be simultaneously true in a consistent manner. Failures of noncontextuality can often be expressed in terms of the inability to consistently assign definite outcome values to a set of measurements [8, 27]. Examples of paradoxical chains of reasoning in quantum theory [17] and box world [7] — two contextual theories—and the intuition of the impossibility of finding such a chain in Spekkens' toy model, as shown here, suggests the following conjecture: logical multi-agent paradoxes are proofs of contextuality, and all contextual physical theories can model multi-agent logical paradoxes. The connection of contextuality to logical contradictions has already been to some extent explored in existing research. For example, it can be shown that the patterns of reasoning which are used in finding a contradiction in the Liar cycles [28] are similar to the reasoning we make use of in FR-type arguments, and in [29] the connection is established between such logical cycles and contextuality. Additionally, in [30], it has been shown that every proof of a logical pre-post selection paradox is a proof of contextuality. The question of how proofs of contextuality relate to proofs of multi-agent logical paradoxes will be formally addressed in future work.

6.8. Weak and noisy measurements

In [31], the analysis of weak measurements and weak values is applied to the epistemically-restricted theory of Liuoville classical mechanics. Weak values in that case coincide with the ones obtained for gaussian quantum mechanics; no anomalous weak values are observed, as the theory is non-contextual. It would also be interesting to apply our analysis of physical measurements to try to implement noisy and weak measurements in generalized Spekkens' theory; we leave this as an open project.

[6] 'I think I know what to do. We're gonna have to break a few rules, but if it works, it will help everybody.' — Toy story.

6.9. Wigner's other friends

The thought experiment analyzed in this paper has many similarities to the thought experiments proposed by Brukner [32] and Bong *et al* [33], which also build on the original Wigner's friend scenario. However, the conclusions drawn from the latter two differ from the original FR experiment and its toy analogue discussed in this paper. While Brukner's and Cavalcanti's results provide a strengthening of the Bell's theorem, considering the FR thought experiment in various theories is an exploration of what it means to be a user of the theory and also be described within the said theory, and what operational restrictions such a user might have. As the toy theory does not exhibit non-local features, and its correlations do not violate Bell's inequalities, it is not suitable for formulating Brukner's and Cavalcanti's results. We leave the interesting question of identifying fundamental connections between the assumptions used in all of these scenarios as future work.

6.10. Open questions and generalizations

We did not use the explicit form of the epistemic restriction in our proof of non-existence of the paradox in the toy theory; the assumptions we made are not unique to classical complementary. In principle, we could have defined a different criterion for joint knowability of observables. To preserve the general structure of the theory this new criterion needs to be subject to some conditions: for example, we require jointly knowable variables to have a linear structure; this follows from the fact that if two variables are known, any linear combination can be calculated. The formalization and exploration of different epistemic restrictions, as well as the investigation of other more general settings in which agents could reason about each other, is left for future work.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank Matthew F Pusey, David Schmid, Yilè Ying and Rob Spekkens for pointed discussions. NN and LdR acknowledge support from the Swiss National Science Foundation through SNSF Project No. 200020_165843 and through the National Centre of Competence in Research *Quantum Science and Technology*(QSIT). LdR further acknowledges support from the FQXi large grant *Consciousness in the Physical World*. LdR is grateful for the hospitality of Perimeter Institute where part of this work was carried out. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities.

Author contributions

This manuscript is the second part of LH's semester project as a masters student (the first part was the review of the toy theory [16]). All authors contributed equally to the ideas and techniques developed here. LH wrote the first draft of the manuscript, and it was revised by all authors.

Appendix A. Toy theory formalism in arbitrary dimensions

The complete review of the toy theory formalism, including its original, stabilizer and arbitrary formulations, can be found in [16].

A.1. Formalism for arbitrary dimensions

A.1.1. Is this formalism necessary to understand the results of this paper?

To tackle the more general case of toy systems of arbitrary dimensions, we must review heavier formalism [10, 11]. Our main results are expressed in this language, but we add intuitive descriptions that convey the main message and do not require learning the formalism.

A.1.2. Epistemic states

In the continuous case, we represent toy systems through observables q_i and p_i for each of the n subsystems, analogous to position and momentum: for example, a toy particle moving in 3D would have $n = 3$. For arbitrary dimensions, we represent a valid epistemic state $(\mathbf{V}, \mathbf{v})$ by the known observables $\mathbf{V} = \langle \mathbf{f}_1, \ldots, \mathbf{f}_k \rangle$ and the valuation $\mathbf{v}$ of those observables. For example, if all we know about a 1D continuous system is that

$\mathbf{f}_1 = 2q_1 - p_1 = 5$, then $\mathbf{V} = \left\langle \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\rangle$ and we can have for example $\mathbf{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, as $(2,-1)\begin{pmatrix} 3 \\ 1 \end{pmatrix} = 2 \times 3 - 1 \times 1 = 5$. The set of ontic states compatible with this epistemic state are all those that share the valuation $\mathbf{v}$ for the observables in $\mathbf{V}$; this includes $o_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, but also for instance $o_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, as $(2,-1)o_2 = 5$, $o_3 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$, and so on.

A.1.3. Epistemic restriction for continuous systems

The complete epistemic restriction in then given by the principle of classical complementarity:

> "The valid epistemic states are those wherein an agent knows the values of a set of quadrature variables that commute relative to the Poisson bracket, and is maximally ignorant otherwise." [10]

Here, "maximal ignorance" means that there is a uniform probability over all other values of variables. It can be shown that this complementarity principle requires observables to be linear, i.e. quadrature observables.

We represent quadrature observables as a vector $\mathbf{f} \in \mathbb{Z}_d^{2n}/\mathbb{R}^{2n}$ if we consider n systems of dimension d or n continuous systems. Then the Poisson bracket is defined for both the continuous and discrete case, where the sum has to be understood in mod d in the discrete case:

$$[f,g] = \sum_{i=1}^{n} f_{2i-1}g_{2i} - f_{2i}g_{2i-1} \tag{A.1}$$

where f_j denotes the jth entry of $\mathbf{f}$ [7,8] [10].

A.1.4. Composing continuous systems

To continue our example suppose that we bring in a second 1D system, and we know the local observable corresponding to the position of this system, for instance $\mathbf{f}_2 = q_2 = 10$. The global epistemic state is then specified by:

$$(\mathbf{V},\mathbf{v}') = (\langle \mathbf{f}_1, \mathbf{f}_2 \rangle, \mathbf{v}') = \left(\left\langle \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} 3 \\ 1 \\ 10 \\ 0 \end{pmatrix} \right),$$

which is a product state. On the other hand, if instead of q_2 we knew a global property, like that the positions of the two systems were perfectly correlated, $q_1 = q_2$, we could represent this through a new observable $\mathbf{f}_3 = q_1 - q_2 = 0$, and so our global epistemic state would be:

$$(\mathbf{V}',\mathbf{v}'') = (\langle \mathbf{f}_3 \rangle, \mathbf{v}'') = \left(\left\langle \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} 3 \\ 1 \\ 3 \\ 0 \end{pmatrix} \right).$$

A.1.5. Reversible transformations

Valid reversible transformations are *sympletic transformations*, represented by a pair $(U,\mathbf{a})$, where $\mathbf{a}$ is an ontic state[9] and U is a sympletic matrix. Sympletic matrices are those that satisfy $U^T J U = J$, where

$$\mathbf{J} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ -1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & -1 & \\ 0 & 0 & 1 & 0 & \\ \vdots & & & & \ddots \end{pmatrix}, \tag{A.2}$$

[7] The Poisson bracket can also be understood as the symplectic inner product of $\mathbf{f}$ and $\mathbf{g}$.

[8] Because of this definition we call the $2i-1$th entry the position of the system i and the $2i$th entry the momentum of system i.

[9] Throughout this paper $\mathbf{a} = 0$ unless otherwise stated.

and is used to write the symplectic inner product[10]. The transformation $(U,\mathbf{a})$ transforms an each ontic state from $\mathbf{o}$ to $U(\mathbf{o}+\mathbf{a})$. An epistemic state $(V,\mathbf{v})$ transforms under such a symplectic transformation as:

$$(V,\mathbf{v}) \rightarrow (\ (U^T)^{-1}V,\ U(\mathbf{v}+\mathbf{a})\ ). \tag{A.3}$$

The reason the transformation implements $(U^T)^{-1}$ on the vector space of known variables is that the transformation transforms an ontic state $\mathbf{o}$ to $U\mathbf{o}$, so if we know that an ontic state was compatible with the known variables before it must also be compatible afterwards.

For example consider the state:

$$\left(\mathbf{V}=\left\langle\begin{pmatrix}2\\-1\end{pmatrix}\right\rangle,\mathbf{v}=\begin{pmatrix}3\\1\end{pmatrix}\right),$$

and the transformation that swaps position and momentum:

$$U=\begin{pmatrix}0&1\\1&0\end{pmatrix}.$$

This transforms the above state to:

$$\left(\mathbf{V}'=\left\langle\begin{pmatrix}-1\\2\end{pmatrix}\right\rangle,\mathbf{v}'=\begin{pmatrix}1\\3\end{pmatrix}\right).$$

The ontic state $o_2=\begin{pmatrix}2\\-1\end{pmatrix}$ which was compatible with the knowledge before the transformation is still valid after the transformation as $\begin{pmatrix}-1\\2\end{pmatrix}^T(Uo_2)=5$.

A.1.6. Measurements

In the continuous case measurement consists of a vector space V_π of observables that which can have outcomes $\mathbf{v}_\pi\in\Omega$, all possible (inequivalent) outcomes also result in a partition of the ontic state space, like in the discrete case. The probability to get outcome $\mathbf{v}_\pi$ if the system is in an ontic state $\mathbf{m}$ is given by [10]:

$$\xi(\mathbf{v}_\pi|\mathbf{m})=\delta_{V_\pi^\perp+\mathbf{v}_\pi}(\mathbf{m}). \tag{A.4}$$

Intuitively, this means that we can only obtain measurement outcomes that are compatible with the ontic state of the system. We can denote a measurement V_π and its outcome $\mathbf{v}_\pi$ by the pair $(V_\pi,\mathbf{v}_\pi)$. With the conditional probability distribution $\xi(\mathbf{v}_\pi|\mathbf{m})$ we can calculate the probability for a measurement outcome given the epistemic state $(V,\mathbf{v})$ [10],

$$P(\mathbf{v}_\pi|(V,\mathbf{v}))=\sum_{\mathbf{m}\in\Omega}\xi(\mathbf{v}_\pi|\mathbf{m})\ \mu_{(V,\mathbf{v})}(\mathbf{m}), \tag{A.5}$$

where $\mu_{(V,\mathbf{v})}(\mathbf{m})$ is the equal probability distribution over all ontic states compatible with the knowledge of the observables in V having valuations $\mathbf{v}$.

A.1.7. Example of a measurement

For example consider the state:

$$\left(\mathbf{V}=\left\langle\begin{pmatrix}2\\0\end{pmatrix}\right\rangle,\mathbf{v}=\begin{pmatrix}3\\1\end{pmatrix}\right),$$

and we want to measure the position:

$$\left\langle\begin{pmatrix}1\\0\end{pmatrix}\right\rangle.$$

[10] For two observables $\mathbf{f}$ and $\mathbf{g}$, their symplectic inner product corresponds to the Poisson bracket $[f,g]=\sum_{i=1}^{n}f_{2i-1}g_{2i}-f_{2i}g_{2i-1}$, with f_j denoting the jth entry of $\mathbf{f}$. It can then be rewritten making use of the matrix $\mathbf{J}$: $[f,g]=\mathbf{f}^T\,\mathbf{J}\,\mathbf{g}$.

The epistemic state is such that we know the position to be $2 \cdot 3 = 6$, so the only compatible measurement outcome is:

$$\mathbf{v}_\pi = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

A.1.8. Post-measurement state

The post-measurement state of the system (given the information about the outcome) is as follows.

Theorem A.1 (Measurement update rule [16]). *When an epistemic state $(V, \mathbf{v})$ is subjected to a measurement V_π, and outcome $\mathbf{v}_\pi$ is obtained, the epistemic state is updated to $(V', \mathbf{v}')$, where*

$$V' = V_\pi \oplus V_{commute}, \tag{A.6}$$

$$\mathbf{v}' \in (V_\pi^\perp + \mathbf{v}_\pi) \cap (V_{commute}^\perp + \mathbf{v}), \tag{A.7}$$

$$V_{commute} = \{\mathbf{f} \in V : \ [\mathbf{f}, \mathbf{f}_\pi] = 0, \ \forall \, \mathbf{f}_\pi \in V_\pi\} \subseteq V. \tag{A.8}$$

Appendix B. Formal results and proofs

B.1. Linear algebra lemmas

Here we list the results from linear algebra we used in the refinement of the generalization of Spekkens' toy theory.

Lemma B.1 ([34]). *Let $W \subset \Omega$ be a subvector space or submodule and $\mathbf{w} \in \Omega$. Then for any $\mathbf{a} \in W + \mathbf{w}$*

$$W + \mathbf{w} = W + \mathbf{a} \tag{B.1}$$

Proof. Let $\mathbf{b} \in W + \mathbf{w}$ then

$$\mathbf{b} = \mathbf{w}_1 + \mathbf{w}, \tag{B.2}$$

for some $\mathbf{w}_1 \in W$. As $\mathbf{a} \in W + \mathbf{w}$ we know that

$$\begin{aligned} \mathbf{a} &= \mathbf{w}_2 + \mathbf{w} \\ \iff \mathbf{w} &= \mathbf{a} - \mathbf{w}_2 \end{aligned} \tag{B.3}$$

for some $\mathbf{w}_2 \in W$. Plugging the expression for $\mathbf{w}$ into the expression for $\mathbf{b}$ we find:

$$\mathbf{b} = \mathbf{w}_1 - \mathbf{w}_2 + \mathbf{a} \in W + \mathbf{a} \tag{B.4}$$

Therefore, $W + \mathbf{w} \subset W + \mathbf{a}$.

Let $\mathbf{c} \in W + \mathbf{a}$ then we can write

$$\mathbf{c} = \mathbf{w}_3 + \mathbf{a} = \mathbf{w}_3 + \mathbf{w}_2 + \mathbf{w} \in W + \mathbf{w}, \tag{B.5}$$

where we used the expression for $\mathbf{a}$ from above. From this Equation we can conclude that $W + \mathbf{a} \subset W + \mathbf{w}$. □

Lemma B.2 ([34]). *Let $W, V \subset \Omega$ be two subvector spaces or submodules and $\mathbf{v}, \mathbf{w} \in \Omega$. Then if $(W + \mathbf{w}) \cap (V + \mathbf{v}) \neq \emptyset$, it holds that:*

$$(W + \mathbf{w}) \cap (V + \mathbf{v}) = (W \cap V) + \mathbf{u}, \tag{B.6}$$

with $\mathbf{u} \in (W + \mathbf{w}) \cap (V + \mathbf{v})$.

Proof. If $(W+\mathbf{w})\cap(V+\mathbf{v})\neq\emptyset$ then there exists a $\mathbf{u}\in(W+\mathbf{w})\cap(V+\mathbf{v})$. Lemma B.1 allows us to write:

$$\begin{aligned} W+\mathbf{w}&=W+\mathbf{u}\\ V+\mathbf{v}&=V+\mathbf{u}. \end{aligned} \tag{B.7}$$

This means each element in $V+\mathbf{v}$ is of the form $\mathbf{u}_1=\mathbf{v}_1+\mathbf{u}$ for $\mathbf{v}_1\in V$, and each element in $W+\mathbf{w}$ is of the form $\mathbf{u}_2=\mathbf{w}_1+\mathbf{u}$ for $\mathbf{w}_1\in W$. Therefore $\mathbf{u}_1$ is in $W+\mathbf{w}$ if and only if $\mathbf{v}_1\in W$ and $\mathbf{u}_2$ is in $V+\mathbf{v}$ if and only if $\mathbf{w}_1\in V$. Therefore we can conclude that $(W+\mathbf{w})\cap(V+\mathbf{v})=(V\cap W)+\mathbf{u}$.

□

Lemma B.3 ([11]). *Let $V,W\subset\Omega$ be two subvector spaces or submodules then it holds that:*

$$(V\oplus W)^{\perp}=V^{\perp}\cap W^{\perp} \tag{B.8}$$

Proof. Let $\mathbf{a}\in(V\oplus W)^{\perp}$ and, therefore, for all vectors $\mathbf{u}\in(V\oplus W)$ it holds that $\mathbf{a}^T\mathbf{u}=0$. In particular, this holds for $\mathbf{a}\in V^{\perp}$ and $\mathbf{a}\in W^{\perp}$ as V and W are subsets of $(V\oplus W)$. Therefore, we can conclude that $(V\oplus W)^{\perp}\subset V^{\perp}\cap W^{\perp}$.

Let $\mathbf{b}\in V^{\perp}\cap W^{\perp}$ and let $\mathbf{u}\in(V\oplus W)$ be arbitrary. Then we find that

$$\mathbf{u}^T\mathbf{b}=\mathbf{u}_w^T\mathbf{b}+\mathbf{u}_v^T\mathbf{b}=0 \tag{B.9}$$

for $\mathbf{u}_w\in W$ and $\mathbf{u}_v\in V$ such that $\mathbf{u}_w+\mathbf{u}_v=\mathbf{u}$. Therefore, $V^{\perp}\cap W^{\perp}\subset(V\oplus W)^{\perp}$ □

Lemma B.4 ([35, 36]). *Let $V\subset\Omega$ be a subset or submodule. Then it holds that $(V^{\perp})^{\perp}=V$.*

Proof. The proof for this in the case for general d can be found in [35]. In the case for general vector spaces (d prime or the continuous case) the proof can be found in [36]. □

B.2. Measurement as a physical process

Here you can find the proofs of statements used to formulate rules for measurement process.

B.2.1. Example: measuring position with a continuous 1D pointer

We consider the case where both the measured system and the pointer are continuous 1D systems, characterized by the observables q_S,p_S,q_M and p_M. Note that we cannot start the pointer in the analogous of a Gaussian state, as each toy observable can only be either fully known or completely unknown. We start instead with a pointer well-localized in position space, with $q_M=x_0$. Suppose that we want to measure the position of the first system, S. If S starts in a state of well-defined position q, then we expect to end up in a 'classically corelated' toy state analogous to $|q\rangle_S|x_0+q\rangle_M$. If on the other hand S starts with well-defined momentum p and undefined position, we would expect the final global state to be somehow analogous to a superposition $\sum_q\alpha(q,x_0,p)|q\rangle_S|x_0+q\rangle_P$. Theorem B.6 shows that in the toy theory there exists a transformation that produces a final state that is analogous to the quantum case.

Theorem B.5 (Coherent copy of position). *Let the epistemic state of the memory system be initialized as $q_{memory}=0$, and let the initial epistemic state of the measured system be $(V_{information}=\langle\mathbf{v}_1\rangle,\mathbf{v})$ with $\mathbf{v}_1,\mathbf{v}\in\mathbb{Z}_d^2$ or $\mathbb{R}^2$ so that the initial epistemic state of the composite system is:*

$$\left(\left\langle\begin{pmatrix}v_1^1\\v_1^2\\0\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\\0\end{pmatrix}\right\rangle,\begin{pmatrix}v^1\\v^2\\0\\0\end{pmatrix}\right). \tag{B.10}$$

Then the transformation:

$$S=\begin{pmatrix}1&0&0&0\\0&1&0&-1\\1&0&1&0\\0&0&0&1\end{pmatrix} \tag{B.11}$$

correlates the position of the memory system and the information system. In prime and continuous dimensions tracing out the memory system results in a mixture of all possible measurement outcomes of the position of the information system.

Proof. We apply the transformation to the initial state:

$$S: \left(\left\langle \begin{pmatrix} v_1^1 \\ v_1^2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} v^1 \\ v^2 \\ 0 \\ 0 \end{pmatrix} \right) \to \left(\left\langle \begin{pmatrix} v_1^1 \\ v_1^2 \\ 0 \\ v_1^2 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} v^1 \\ v^2 \\ v^1 \\ 0 \end{pmatrix} \right). \tag{B.12}$$

This transformation ensures that the position of the memory system is always equal to the position of the information system.

Furthermore, if we trace out the memory system, that is taking the marginal of the probability distribution of the ontic state over the memory system. The probability distribution before marginalisation is the uniform distribution over $V^\perp + \mathbf{v} = \mathrm{span}((0,-1,0,1)^T, (v_1^\perp, v_2^\perp, v_1^\perp, 0)^T) + (v^1, v^2, v^1, 0)^T$ where $(v_1^\perp, v_2^\perp)$ is the vector spanning $V^\perp_{information}$. After marginalisation this results in a uniform probability distribution over $V^\perp + \mathbf{v}$ with the last two entries removed $V'^\perp + \mathbf{v}' = \mathrm{span}((0,1)^T, (v_1^\perp, v_2^\perp)^T) + (v^1, v^2)$. Therefore, if $v_1^2 \neq 0$ the traced out state is the maximally mixed state and therefore an equal mixture of all positions of the information system. In non prime dimensions, even if $v_1^2 \neq 0$, the state does not need to be the maximally mixed state. On the other hand if $v_1^2 = 0$, the state is the state where the information system has definite position v^1. □

B.2.2. Examples

To obtain some intuition, let us look at two examples. Recall that this sympletic transformation acts on an initial state as $(V, \mathbf{v}) \to ((S^T)^{-1}V, S\mathbf{v})$. Consider the initial state analogous to $|\phi\rangle_S|x_0\rangle_M$, that is the product state between an arbitrary state of S and a well-defined memory position $q_M = x_0$. This state is transformed as:

$$\left(\left\langle \begin{pmatrix} v_1 \\ v_2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} w_1 \\ w_2 \\ x_0 \\ 0 \end{pmatrix} \right) \longrightarrow \left(\left\langle \begin{pmatrix} v_1 \\ v_2 \\ 0 \\ -v_2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} w_1 \\ w_2 \\ x_0 + w_1 \\ 0 \end{pmatrix} \right). \tag{B.13}$$

In particular, after a quick simplification we can see that it transforms the initial state analogous to $|q\rangle_S|x_0\rangle_M$ as:

$$\left(\left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} q \\ 0 \\ x_0 \\ 0 \end{pmatrix} \right) \longrightarrow \left(\left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} q \\ 0 \\ q + x_0 \\ 0 \end{pmatrix} \right), \tag{B.14}$$

that is, a state analogous to $|q\rangle_S|x_0 + q\rangle_M$, as desired (it satisfies $q_S = q$ and $q_M = q + x_0$). On the other hand, U transforms the state analogous to $|p\rangle_S|x_0\rangle_M$ as:

$$\left(\left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} 0 \\ p \\ x_0 \\ 0 \end{pmatrix} \right) \longrightarrow \left(\left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} 0 \\ p \\ x_0 \\ 0 \end{pmatrix} \right), \tag{B.15}$$

that is, a state such that $p_S - p_M = p$ and $q_S + q_M = x_0$. Similarly to the quantum case, the reduced states of just S or just M are fully mixed.

Theorem B.6 (Coherent copy of an arbitrary observable). *Let the state of the information system be* $(W, \mathbf{w})$*, where W such that it defines a state on an arbitrary amount of systems. Let* $\mathbf{f}$ *be an observable, such that* $\mathbf{f} = ((S^M)^T)^{-1}\mathbf{q}_1$ *where* $\mathbf{q}_1$ *the position of the first system of the information system and* S^M *a symplectic transformation. Furthermore, let* $\mathbf{v} = (T^T_{mem})^{-1}\mathbf{q}$ *be an observable with T a symplectic transformation and* $\mathbf{q}$ *the position of the memory system. Let the memory system be initially in the state where the value of* $\mathbf{v}$ *is 0, such that the total initial state is* $(V_{tot}, \mathbf{v}_{tot}) = \left(\langle(\mathbf{v}, 0, 0, \dots)^T, (0, 0, \mathbf{w}_1)^T, \dots (0, 0, \mathbf{w_k})^T\rangle, (0, 0, \mathbf{w})^T\right)$*, where* $\mathbf{w}_1, \dots, \mathbf{w_k}$ *span W.*
Then the transformation

$$(T_{mem} \otimes \mathbb{1}_{mes})(\mathbb{1}_{mem} \otimes S^M_{mes})(S \otimes \mathbb{1}_{mes_2,..,mes_N})(\mathbb{1}_{mem} \otimes (S^M_{mes})^{-1})(T^{-1}_{mem} \otimes \mathbb{1}_S)), \tag{B.16}$$

correlates the value of $\mathbf{v}$ *of the memory system and with the value of* $\mathbf{f}$ *of the information system. In prime and continuous dimensions tracing out the memory system results in a mixture of all possible measurement outcomes of* $\mathbf{f}$ *of the information system.*

Proof. The first part of the transformation $(\mathbb{1}_{mem} \otimes (S^M_{mes})^{-1})(T^{-1}_{mem} \otimes \mathbb{1}_S))$ performs a basis change, such that the observable $\mathbf{f}$ of the information system in the old basis is transformed to $\mathbf{q}_1$ in the new basis and the observable $\mathbf{v}$ of the memory system is transformed to the position of the memory system. Therefore, the case where we copy $\mathbf{q}_1$ into the position of the memory system is related by a basis change to the case where we copy $\mathbf{f}$ into the value of $\mathbf{v}$ of the memory system. By an analogous calculation as in theorem B.5 the transformation $(S \otimes \mathbb{1}_{mes_2,..,mes_N})$ correlates the position of the memory system with $\mathbf{q}_1$. This means that after the transformation with $(S \otimes \mathbb{1}_{mes_2,..,mes_N})$ the state $(V'_{tot}, \mathbf{v}')$ is such that V'_{tot} contains a vector of the form $(1, 0, -1, 0, 0, 0, 0, \dots, 0)^T \in V'_{tot}$ and the valuation of $(1, 0, -1, 0, 0, 0, 0, \dots, 0)^T$ is zero. The last part of the transformation $(\mathbb{1}_{mem} \otimes S^M_{mes})(T_{mem} \otimes \mathbb{1}_S))$ transforms $(1, 0, -1, 0, 0, 0, 0, \dots, 0)^T \to (\mathbf{v}, -\mathbf{f})$ and the valuation of the new vector is still zero as the transformation is symplectic. Therefore, the above transformation correlates the value of $\mathbf{f}$ on the information system with the value of $\mathbf{v}$ on the memory system.

As the transformation $(\mathbb{1}_{mem} \otimes (S^M_{mes})^{-1})(T^{-1}_{mem} \otimes \mathbb{1}_S))$ acts only locally on the information system and the memory system, we can first transform back the joint memory and system state, then trace out the memory system and finally only transform the information system back and get the same result as if we would have just traced out the memory system. To determine the marginal we must consider the probability distribution over the ontic states induced by the epistemic state. This probability distribution is just the uniform distribution over the ontic states in $U^\perp + \mathbf{u}$, where $(U, \mathbf{u})$ is the state after the measurement update. The vector space $(S^T_M \otimes T^T_{mem} U)^\perp$ is given by $V^\perp_M \oplus \langle(0, 1, 0, -1, \dots, 0)^T\rangle$ where $V^\perp_M$ is spanned by the vectors $(v_i^{(1)}, 0, \mathbf{v}_i)$ where $\mathbf{v}_i$ are the vectors spanning $(S^T_M W)^\perp$ and $v_i^{(k)}$ is the kth entry of $\mathbf{v}_i$. Taking the marginal of the probability distribution over the memory system gives the uniform probability distribution over $(S^T_M W)^\perp \oplus \langle(0, 1, 0, \dots, 0)^T\rangle + S^{-1}_M \mathbf{w}$, as marginalisation results in removing the the entries in the vectors of the systems we marginalized over. Therefore, we can conclude the state of the marginal system is $(M = (S^T_M W) \cap \langle(0, 1, \dots, 0)^T\rangle^\perp, S^{-1}_M \mathbf{w})$. Thus all vectors in M are of the form $(1, 0, \dots)^T$ or $(0, 0, \dots)^T$. This means that $M \oplus \langle(1, 0, 0, \dots, 0)^T\rangle$ is a set of commuting observables and $(1, 0, 0, \dots, 0)^T$ is linearly independent of M if and only if $(0, 1, 0, \dots, 0)^T$ was not already contained in $(S^T_M W)$. In this case, for each value $q \in Z_d$ or $\mathbb{R}$ there exists a valuation vector $\mathbf{v}_q$ such that $(1, 0, 0, \dots, 0)^T$ has the valuation q and the valuation of M is constant. This means that M is a mixture of states with all possible values of q_1 except if $\mathbf{q}_1$ was already known. After transforming the marginalized system back with S_M, the results we found for $\mathbf{q}_1$ before the transformation with S_M hold now for $\mathbf{f}$. □

Lemma B.7 (Existence of symplectic transformation). *Consider the same setting as in theorem B.6. In continuous and prime dimensions for any observable* $\mathbf{f}$ *on the information system and any observable* $\boldsymbol{v}$ *on the memory system there exist symplectic transformations such that*

$$\mathbf{f} = ((S_M)^T)^{-1}\mathbf{q}_1 \tag{B.17}$$

$$\mathbf{v} = (T^T)^{-1}\mathbf{q}. \tag{B.18}$$

Proof. Both transformations exist if we can find for any amount of systems a symplectic transformation where the first column of the transformation is given by an arbitrary vector. Let us call this vector $\mathbf{w} = (w_1, \dots, w_{2n})^T$. The set of vectors:

$$C = \left\{(w_1, \dots, w_{2n})^T, (0, 0, w_3, w_4, \dots, w_{2n})^T, \dots, (0, \dots, 0, w_{2n-1}, w_{2n})^T\right\} \tag{B.19}$$

all commute with each other. Furthermore, the first vector listed in this set $(w_1, \dots, w_{2n})^T$ has symplectic inner product 1 with the vector $(-1/w_2, 0, 0, \dots, 0)$ if $w_2 \neq 0$ (otherwise choose $(0, 1/w_1, 0, \dots, 0)$ and commutes

with the rest of the set C. For the second vector the same holds for $(1/w_2, 0, -1/w_4, 0, \ldots, 0)$ if $w_4 \neq 0$ (otherwise choose $(1/w_2, 0, 0, 1/w_3, \ldots, 0)$ or $(0, -1/w_1, 0, 1/w_3, \ldots, 0)$ if $w_2 = 0$). In a similar way for all vectors $\mathbf{u}$ in the set C such a vector $\mathbf{u}'$ can be constructed such that it has symplectic inner product is 1 with $\mathbf{u}$ and commutes with all other vectors in C. A transformation is sympletic if its columns are such that, the first two columns have symplectic inner product 1 with each other but commute with the rest, the same for the third and fourth column and so on for the following pairs of columns. Therefore, the transformation where the odd columns are the vectors from the set C, starting with the vector $\mathbf{w}$ and the even rows are the vectors we constructed from the previous odd column in the manner as described above is by construction symplectic and has $\mathbf{w}$ as its first column. □

Theorem B.8 (Restrictions on conditional transformations). *Let $(V_S^i, \mathbf{v}_s^i)$ be the state of a system and $(V_T^i, \mathbf{v}_T^i)$ the initial state of the system to be prepared in a conditional preparation scenario. Additionally, let S_{ST} be a symplectic transformation. Then, for $\mathbf{v}_{s,1}^i, \ldots, \mathbf{v}_{s,k}^i$ generating a partition of the ontic state space, i.e. $((V_S^i)^\perp + \mathbf{v}_{s,1}^i) \cup \cdots \cup ((V_S^i)^\perp + \mathbf{v}_{s,k}^i) = \Omega$ the marginals of the target system the final state $(V_T^f, \mathbf{v}_T^f)$ must be either identical or orthogonal. The number of identical states is equal for each separate type of identical states.*

Proof. After the transformation the state of the system is:

$$\left(\left(S_{ST}^{-1}\right)^T \left(V_S^i \oplus V_T^i\right), S_{ST}\left(\mathbf{v}_s^i \oplus \mathbf{v}_T^i\right)\right). \tag{B.20}$$

This state is allowed as V_S^i only has support on the first 2 amount of systems in S entries of the system and V_T^i on the rest. The marginal of the target system after the transformation is:

$$\left(\left(R_T \left(S_{ST}^{-1}\right)^T \left(V_S^i \oplus V_T^i\right)\right), \Pi_T S_{ST} \mathbf{v}_s^i \oplus \mathbf{v}_T^i\right) \tag{B.21}$$

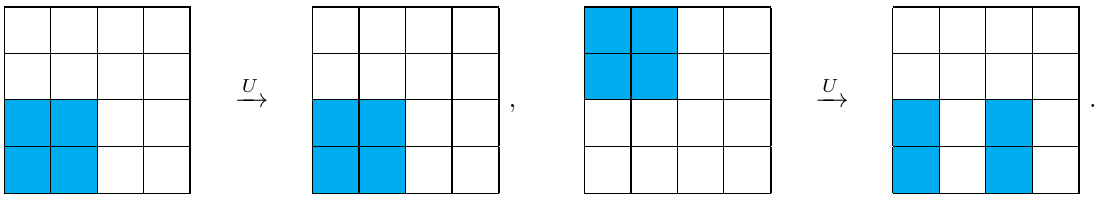
where Π_T the projection on the last 2 amount of systems in T entries of the system and R_T removes the vectors that have support in S. As the transformation cannot depend on $\mathbf{v}_S^i$, $\mathbf{V}_T^f$ is independent of $\mathbf{v}_S^i$. Therefore, changing $\mathbf{v}_S^i$ can only change the valuation $\mathbf{v}_s^f$ of the final state. Furthermore, changing the valuation either result in orthogonal or identical states.

If the transformation is irreversible, then we can always see the transformation as a reversible transformation on a larger system and subsequently tracing out some systems. Taking the trace effectively removes vectors from V_T^f. Therefore it can make states that were orthogonal identical, but it cannot produce non orthogonal or identical states, as also in this case the transformation cannot depend on $\mathbf{v}_S^i$.

This establishes that the different marginalised states on the target system only differ in their valuations which depend only on $\mathbf{v}_s^i$. By equation (B.21) this dependence is a linear map. Let us call this map C. So, two target valuations $\mathbf{v}_{T,1}^f, \mathbf{v}_{T,2}^f$ are identical if and only if $\mathbf{v}_{T,1}^f - \mathbf{v}_{T,2}^f \in Ker(C)$. Therefore, the sets of identical marginal target states are given by $Ker(C) + \mathbf{v}_T^f$ with $\mathbf{v}_T^f$ some valuation for a target state. Therefore, each set of identical target marginal states has the same number of elements.

Note that all transformations in Spekkens' toy theory are given by a symplectic matrix S and a shift vector $\mathbf{a} \in \Omega$, the shift vector only changes the value of the different known observables by a constant, independent of the vector $\mathbf{v}_s^i$. Therefore, adding shifts to transformations also does not help to obtain non-orthogonal states. □

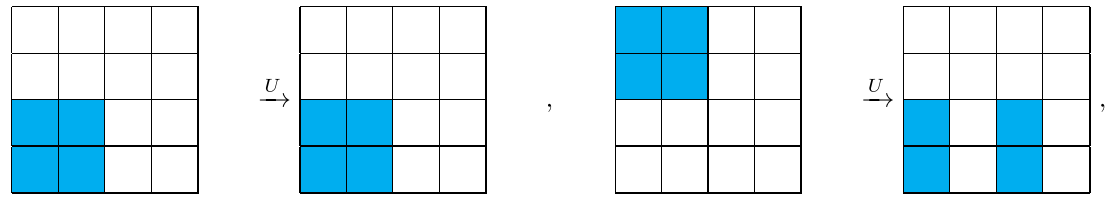
Corollary 1 (Restrictions on conditional transformations: example). *In Spekkens' toy theory, there are no reversible transformations U that implement the action:*



Proof. This is a direct application of theorem B.8, which is the formal version of theorem 3.1. It also has a direct and simple proof in the stabilizer version of the toy theory; we present that proof here for pedagogical purposes for the readers familiar with the stabilizer formalism.

In the stabilizer formalization of the toy theory [9], for appropriate dimensions, each valid epistemic state can be isomorphically identified with a set $S = \langle g_1 g_2, \ldots \rangle$ of commuting or anti-commuting Pauli operators forming a valid stabilizer. Each allowed transformation on toy states corresponds to a permutation Π on the

subset of stabilizers. This permutation can be represented by a (unitary or anti-unitary) permutation matrix V_Π that acts on each stabilizer $g \in S$ as $V_\Pi g V_\Pi^T$. In this case, we have:



$|00\rangle \sim \langle Z_A, Z_S\rangle$ $\langle Z_A, Z_S\rangle$ $|10\rangle \sim \langle -Z_A, Z_S\rangle$ $|0+\rangle \sim \langle Z_A, X_S\rangle$.

This transformation Π would have to map the stabilized states as:

$$\begin{aligned}\Pi: \quad &\mathrm{span}\{\mathcal{Z}_A, \mathcal{Z}_S\} \to \mathrm{span}\{\mathcal{Z}_A, \mathcal{Z}_S\},\\ &\mathrm{span}\{-\mathcal{Z}_A, \mathcal{Z}_S\} \to \mathrm{span}\{\mathcal{Z}_A, \mathcal{X}_S\}.\end{aligned}$$

This means that permutation Π would depend on the sign of $\mathcal{Z}_A$. This dependence should be explicit in the corresponding permutation matrix V_Π (which needs to be either unitary or anti-unitary [9]). However, permutation matrices cannot depend on the sign of stabilizers [9], due to the linear nature of quantum theory. To see this, note that the permutation matrix would have to act on each stabilizer $g \in \{\mathcal{Z}_A \otimes \mathbb{1}_S, -\mathcal{Z}_A \otimes \mathbb{1}_S, \mathbb{1}_A \otimes \mathcal{Z}_S\}$ as $V_\Pi\, g\, V_\Pi^\dagger$, so if the transformation on the first state behaves as:

$$\Pi: \quad \langle \mathcal{Z}_A, \mathcal{Z}_S\rangle \to \langle V_\Pi(\mathcal{Z}_A \otimes \mathbb{1}_S)V_\Pi^\dagger,\ V_\Pi(\mathbb{1}_A \otimes \mathcal{Z}_S)V_\Pi^\dagger\rangle = \mathrm{span}\{\mathcal{Z}_A, \mathcal{Z}_S\},$$

then it must act on the second state as:

$$\begin{aligned}\Pi: \quad \mathrm{span}\{-\mathcal{Z}_A, \mathcal{Z}_S\} \to \quad &\mathrm{span}\{V_\Pi(-\mathcal{Z}_A \otimes \mathbb{1}_S)V_\Pi^\dagger,\ V_\Pi(\mathbb{1}_A \otimes \mathcal{Z}_S)V_\Pi^\dagger\}\\ =\,&\mathrm{span}\{-V_\Pi(\mathcal{Z}_A \otimes \mathbb{1}_S)V_\Pi^\dagger,\ V_\Pi(\mathbb{1}_A \otimes \mathcal{Z}_S)V_\Pi^\dagger\}\\ =\,&\mathrm{span}\{-\mathcal{Z}_A, \mathcal{Z}_S\} \neq \mathrm{span}\{\mathcal{Z}_A, \mathcal{X}_S\}.\end{aligned}$$

□

B.3. Predictions with certainty

Here you can find the proofs of statements used to formulate rules for agents making predictions with certainty. The linear algebraic statements used in the proofs are formally justified in appendix B.1.

Lemma B.10. *Given an epistemic state $(V, \mathbf{v})$ and a measurement V_π, an outcome $\mathbf{v}_\pi$ will be measured with certainty if and only if:*

$$V^\perp + \mathbf{v} \subset V_\pi^\perp + \mathbf{v}_\pi. \tag{B.22}$$

This condition is fulfilled if and only if the following two properties hold:

(a) $V_\pi \subset V$,
(b) $(V^\perp + \mathbf{v}) \cap (V_\pi^\perp + \mathbf{v}_\pi) \neq \emptyset$

Proof. The probability of the measurement outcome $\mathbf{v}_\pi$ given that the system is in an epistemic state $(V, \mathbf{v})$ is given by the condition in [16]:

$$P(\mathbf{v}_\pi|(V,\mathbf{v})) = \sum_{\mathbf{m}\in\Omega} \xi(\mathbf{v}_\pi|\mathbf{m})\mu_{(V,\mathbf{v})}(\mathbf{m}) = \frac{1}{N_V}\sum_{\mathbf{m}\in\Omega} \delta_{V_\pi^\perp + \mathbf{v}_\pi}(\mathbf{m})\delta_{V^\perp+\mathbf{v}}(\mathbf{m}). \tag{B.23}$$

We are able to predict with certainty that the measurement outcome is $\mathbf{v}_\pi$ if:

$$(\delta_{V^\perp+\mathbf{v}}(\mathbf{m}) = 1) \implies (\delta_{V_\pi^\perp+\mathbf{v}_\pi}(\mathbf{m}) = 1). \tag{B.24}$$

By the definition of $\delta_{V^\perp+\mathbf{v}}$ and $\delta_{V_\pi^\perp+\mathbf{v}_\pi}$ this condition is equivalent to [16]:

$$V^\perp + \mathbf{v} \subset V_\pi^\perp + \mathbf{v}_\pi. \tag{B.25}$$

One might wonder why the symmetry is broken between $(V, \mathbf{v})$ and $(V_\pi, \mathbf{v}_\pi)$. The reason here is that $\delta_{V^\perp + \mathbf{v}}$ is normalized, while $\delta_{V_\pi^\perp + \mathbf{v}_\pi}$ is not. The above condition can be further simplified with the following lemma.

Let both properties be fulfilled. Then there exists a vector $\mathbf{w} \in (V^\perp + \mathbf{v}) \cap (V_\pi^\perp + \mathbf{v}_\pi)$ and it holds that $V_\pi \subset V$. From the latter, it follows that $V_\pi^\perp \supset V^\perp$. This can be seen in the following was: if $\mathbf{w} \in V^\perp$ then for any $\mathbf{u} \in V$ it holds that $\mathbf{u}^T\mathbf{w} = 0$ and as $V_\pi \subset V$ this holds in particular for all $\mathbf{u} \in V_\pi$. Thus, $\mathbf{w} \in V_\pi^\perp$. With lemma B.1 we can then conclude:

$$V_\pi^\perp + \mathbf{v}_\pi = V_\pi^\perp + \mathbf{w} \tag{B.26}$$

$$V^\perp + \mathbf{v} = V^\perp + \mathbf{w}. \tag{B.27}$$

Thus, it holds that:

$$V^\perp + \mathbf{v} \subset V_\pi^\perp + \mathbf{v}_\pi \neq \emptyset. \tag{B.28}$$

Let

$$V^\perp + \mathbf{v} \subset V_\pi^\perp + \mathbf{v}_\pi. \tag{B.29}$$

be fulfilled. Then the condition 2 holds, and $V^\perp \subset V_\pi^\perp$. Similarly as above, it holds that $(V^\perp)^\perp \subset (V_\pi^\perp)^\perp$. Since $V \subset \Omega$ is a subset or a submodule, it is true that $(V^\perp)^\perp = V$ (lemma B.4, which implies the condition 1. □

Lemma B.11. *Let $(V_A, v_{A=1})$ be a measurement on subsystem A and $(V_B, v_{B=1})$ a measurement on subsystem B. The statement 'A = 1 $\implies$ B = 1' can be made if and only if following conditions are fulfilled:*

(a) $V_B \subset (V_{commute,\,A} \oplus V_A)$,
(b) $(V_{commute,\,A}^\perp + \mathbf{v}) \cap (V_A^\perp + \mathbf{v}_{A=1}) \cap (V_B^\perp + \mathbf{v}_{B=1}) \neq \emptyset$.

Proof. The updated state if we measure $A = 1$ is $((V_{\text{commute, A}} \oplus V_a), \mathbf{v}')$ where $\mathbf{v}' \in (V_A + \mathbf{v}_{A=1}) \cap (V_{\text{commute, A}} + \mathbf{v})$. Thus, lemma B.10 says that the statement 'A = 1 $\implies$ B = 1' can be made if and only if

(a) $V_B \subset (V_{\text{commute, A}} \oplus V_A)$
(b) $((V_{\text{commute, A}} \oplus V_A)^\perp + \mathbf{v}') \cap (V_B^\perp + \mathbf{v}_{B=1}) \neq \emptyset$

The second condition is equivalent to

$$(V_{\text{commute, A}}^\perp + \mathbf{v}') \cap (V_A^\perp + \mathbf{v}') \cap (V_B^\perp + \mathbf{v}_{B=1}) \neq \emptyset. \tag{B.30}$$

Based on lemma B.1 we can conclude that the second condition is equivalent to:

$$(V_{\text{commute, A}}^\perp + \mathbf{v}) \cap (V_A^\perp + \mathbf{v}_{A=1}) \cap (V_B^\perp + \mathbf{v}_{B=1}) \neq \emptyset. \tag{B.31}$$

□

B.3.1. Toy Bell scenario

The example of the Bell scenario in the generalized formalism goes as follows. Alice and Bob's shared entangled state can be expressed as:

$$(V, \mathbf{v}) = \left(\left\langle \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right). \tag{B.32}$$

Bob's Z measurement corresponds to observable $\left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle$. If he obtains outcome p, he updates his description of the shared state to

$$\left(\left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle, \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \right) = \left(\left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle, \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \right). \tag{B.33}$$

Now, if Alice measures her system in $\left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle$, she will get the outcome p with certainty. Thus, if Alice and Bob share the state $(V, \mathbf{v})$, the conclusion '$B = p \implies A = p$' can be made with certainty. We can also check that the conditions of lemma B.11 are fulfilled:

(a) $$\left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \subset \left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle \oplus \left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle,$$

(b) $$\begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \in \left(\left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle^{\perp} + \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \right) \cap \left(\left\langle \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle^{\perp} + \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \right) \cap \left(\left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle^{\perp} + \begin{pmatrix} 0 \\ p \\ 0 \\ p \end{pmatrix} \right).$$

B.4. No Frauchiger–Renner paradox in the toy theory

B.4.1. Experimental setting

We can write an arbitrary global ontic state of Alice and Bob's labs before the measurements start as:

$$\mathbf{o} = \begin{pmatrix} \mathbf{o}_R \\ \mathbf{o}_A \\ \mathbf{o}_S \\ \mathbf{o}_B \end{pmatrix}, \tag{B.34}$$

with $\mathbf{o}_R \in \mathbb{Z}_d^{n_R}$ or $\mathbb{R}^{n_R}$, $\mathbf{o}_A \in \mathbb{Z}_d^{n_A}$ or $\mathbb{R}^{n_A}$, $\mathbf{o}_S \in \mathbb{Z}_d^{n_S}$ or $\mathbb{R}^{n_S}$, and $\mathbf{o}_B \in \mathbb{Z}_d^{n_B}$ or $\mathbb{R}^{n_B}$. We allow for arbitrary measurements by the different agents, fixing only the range of systems measured and the order of measurements:

Alice at t = 1 performs the following measurement

$$V_A = \left\langle \begin{pmatrix} \mathbf{v}_{A,1} \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \mathbf{v}_{A,k_A} \\ 0 \end{pmatrix} \right\rangle \tag{B.35}$$

where $\mathbf{v}_A \in \mathbb{Z}_d^{n_R}$ or $\mathbb{R}^{n_R}$. Alice says she got the outcome '1' if she got the result $\mathbf{v}_{A=1}$.

Bob at t = 2 performs the following measurement

$$V_B = \left\langle \begin{pmatrix} 0 \\ \mathbf{v}_{B,1} \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \mathbf{v}_{B,k_B} \\ 0 \end{pmatrix} \right\rangle, \tag{B.36}$$

where $\mathbf{v}_B \in \mathbb{Z}_d^{n_S}$ or $\mathbb{R}^{n_S}$. Bob says he got the outcome 1 if he got the result $\mathbf{v}_{B=1}$.

Ursula at t = 3 performs the following measurement:

$$V_U = \left\langle \begin{pmatrix} \mathbf{v}_{U,1} \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \mathbf{v}_{U,k_U} \\ 0 \end{pmatrix} \right\rangle, \tag{B.37}$$

where $\mathbf{v}_{U,1}, \mathbf{v}_{U,2} \in \mathbb{Z}_d^{n_A+n_R}$ or $\mathbb{R}^{n_A+n_R}$. Ursula says she measured 'ok' if she got the outcome $\mathbf{v}_{U,ok}$ and 'fail' if she got the outcome $\mathbf{v}_{U,fail}$. These two vectors need to correspond to distinct outcomes, i.e. $\mathbf{v}_{U,ok} - \mathbf{v}_{U,fail} \notin V_U^{\perp}$.

Wigner at t = 4 performs the following measurement:

$$V_W = \left\langle \begin{pmatrix} 0 \\ \mathbf{v}_{W,1} \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \mathbf{v}_{W,k_W} \end{pmatrix} \right\rangle, \tag{B.38}$$

where $\mathbf{v}_{W,1}, \mathbf{v}_{W,2} \in \mathbb{Z}_d^{n_B+n_S}$ or $\mathbb{R}^{n_W+n_s}$. Wigner says he measured 'ok' if he got the outcome $\mathbf{v}_{W,ok}$ and 'fail' if he got the outcome $\mathbf{v}_{W,fail}$. These two vectors need to correspond to distinct outcomes, i.e. $\mathbf{v}_{W,ok} - \mathbf{v}_{W,fail} \notin V_W^{\perp}$.

The linear algebraic statements used in the proof are formally justified in appendix B.1.

Theorem 5.1. *There exists no valid epistemic state that can be used to reach a logical contradiction in the Frauchiger–Renner setting in the toy theory.*

Proof. We separate the proof in four parts. First, we find necessary conditions on the epistemic state such that predictions can be made with certainty. Second, we find conditions such that the probability for Wigner and Ursula both get 'ok' is non zero. Finally, we show that two measurement outcomes which would lead to a contradiction must be the same and, thus, do not lead to a contradiction. □

B.4.2. Epistemic state that allows predictions with certainty

We follow the chain of reasoning and determine what V needs to fulfill to allow for predictions with certainty:

'U = ok ⟹ B = 1' The first condition for predictions with certainty requires that $V_B \subset V_{\text{commute},U} \oplus V_U$ where $V_{\text{commute},U}$ denotes subset of V that commutes with all vectors in V_U. This means for all vectors $\mathbf{v}_B \in V_B$ these exist vectors $\mathbf{v} \in V_{\text{commute},U}, \mathbf{v}_U \in V_U$ such that $\mathbf{v}_B = \mathbf{v} + \mathbf{v}_U$.

'B = 1 ⟹ A = 1' The first condition for predictions with certainty requires that $V_A \subset V_{\text{commute},B} \oplus V_B$ where $V_{\text{commute},B}$ denotes subset of V that commutes with all vectors in V_B. This means for all vectors $\mathbf{v}_A \in V_A$ these exist vectors $\mathbf{v} \in V_{\text{commute},B}, \mathbf{v}_B \in V_B$ such that $\mathbf{v}_A = \mathbf{v} + \mathbf{v}_B$.

'A = 1 ⟹ W = fail' The first condition for predictions with certainty requires that $V_W \subset V_{\text{commute},A} \oplus V_A$ where $V_{\text{commute},A}$ denotes subset of V that commutes with all vectors in V_A. This means for all vectors $\mathbf{v}_W \in V_W$ these exist vectors $\mathbf{v} \in V_{\text{commute},A}, \mathbf{v}_A \in V_A$ such that $\mathbf{v}_W = \mathbf{v} + \mathbf{v}_A$.

In total, this means for each $\mathbf{v}_W \in V_W$ there exist vectors $\mathbf{v}_1 \in V_{\text{commute},A}$,$\mathbf{v}_2 \in V_{\text{commute},B}$,$\mathbf{v}_3 \in V_{\text{commute},U}$, $\mathbf{v}_A \in V_A$,$\mathbf{v}_B \in V_B$, and $\mathbf{v}_U \in V_U$ such that

$$\begin{aligned} \mathbf{v}_W &= \mathbf{v}_1 + \mathbf{v}_A \\ &= \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_B \\ &= \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_U. \end{aligned} \tag{B.39}$$

B.4.3. $P(ok, ok) \neq zero$

Even if the above chain of reasoning holds, the paradox only occurs if, additionally, the probability for Ursula and Wigner to both get 'ok' $P(ok, ok)$ is not zero. As they perform a joint measurement their measurement is given by $V_U \oplus V_W$. Because Ursula and Wigner measure different subsystems we can choose equivalent measurement outcome vectors such that $\mathbf{v}_{U,ok} \in V_W^{\perp}$ and $\mathbf{v}_{W,ok} \in V_U^{\perp}$. Thus, the measurement outcome $W = ok, U = ok$ has has a shift vector $\mathbf{v}_{U,ok} + \mathbf{v}_{W,ok}$. For $P(ok, ok) \neq 0$ to hold the state $(V, \mathbf{v})$ has to fulfill

$$((V_U \oplus V_W)^{\perp} + \mathbf{v}_{U,ok} + \mathbf{v}_{W,ok}) \cap (V^{\perp} + \mathbf{v}) \neq \emptyset. \tag{B.40}$$

Lemma B.2 allows us to conclude that this condition is equivalent to:

$$\mathbf{v}_{U,ok} + \mathbf{v}_{W,ok} - \mathbf{v} \in (V_U \oplus V_W)^\perp \oplus V^\perp. \tag{B.41}$$

Due to $(V^\perp)^\perp = V$ for a submodule or subset V (lemma B.4), we can rewrite the expression $(V_U \oplus V_W)^\perp \oplus V^\perp$:

$$\begin{aligned}(V_U \oplus V_W)^\perp \oplus V^\perp &= \left(\left((V_U \oplus V_W)^\perp \oplus V^\perp\right)^\perp\right)^\perp \\ &= \left(\left((V_U \oplus V_W)^\perp\right)^\perp \cap \left(V^\perp\right)^\perp\right)^\perp \\ &= ((V_U \oplus V_W) \cap V)^\perp.\end{aligned} \tag{B.42}$$

Thus, equation (B.39) is equivalent to:

$$\mathbf{v}_{U,ok} + \mathbf{v}_{W,ok} - \mathbf{v} \in ((V_U \oplus V_W) \cap V)^\perp. \tag{B.43}$$

B.4.4. Getting the correct outcomes with certainty

We want to ensure that in the reasoning chain we cannot only conclude with certainty but we can also make the desired conclusions. For example, the previous conditions ensure that when Bob measures he gets an outcome with certainty (which can be calculated from the epistemic state), but the condition we consider here ensures that this outcome is $B = 1$:

'U = ok $\Longrightarrow$ B = 1' We apply the second condition for predictions with certainty. Thus, it has to hold that

$$(V_B^\perp + \mathbf{v}_{B=1}) \cap (V_{\text{commute},U}^\perp + \mathbf{v}) \cap (V_U^\perp + \mathbf{v}_{U,ok}) \neq \emptyset. \tag{B.44}$$

The two measurements are defined on two different subsystems. Therefore, we can choose, without loss of generality, $\mathbf{v}_{B=1} \in V_U^\perp$ and $\mathbf{v}_{U,ok} \in V_B^\perp$. Additionally, it does not matter which intersection is calculated first. Therefore, we can first calculate the intersection $(V_B^\perp + \mathbf{v}_{B=1}) \cap (V_U^\perp + \mathbf{v}_{U,ok})$ using lemmas B.1 and B.3

$$\begin{aligned}&(V_B^\perp + \mathbf{v}_{B=1}) \cap (V_U^\perp + \mathbf{v}_{U,ok}) \\ &\quad = (V_B^\perp + \mathbf{v}_{U,ok} + \mathbf{v}_{B=1}) \cap (V_U^\perp + \mathbf{v}_{U,ok} + \mathbf{v}_{B=1}) \\ &\quad = (V_B^\perp \cap V_U^\perp) + \mathbf{v}_{B=1} + \mathbf{v}_{U,ok} \\ &\quad = (V_B \oplus V_U)^\perp + \mathbf{v}_{B=1} + \mathbf{v}_{U,ok}.\end{aligned} \tag{B.45}$$

Plugging this result into equation (B.44) we find the condition

$$((V_B \oplus V_U)^\perp + \mathbf{v}_{B=1} + \mathbf{v}_{U,ok}) \cap (V_{\text{commute},U}^\perp + \mathbf{v}) \neq \emptyset. \tag{B.46}$$

Due to lemma B.1 this condition is equivalent to the condition:

$$\begin{aligned}\mathbf{v}_{B=1} + \mathbf{v}_{U,ok} - \mathbf{v} &\in (V_B \oplus V_U)^\perp \oplus V_{\text{commute},U}^\perp \\ &= ((V_B \oplus V_U) \cap V_{\text{commute},U})^\perp.\end{aligned} \tag{B.47}$$

'B = 1 $\Longrightarrow$ A = 1' With the same reasoning as above we find

$$\mathbf{v}_{A=1} + \mathbf{v}_{B=1} - \mathbf{v} \in ((V_B \oplus V_A) \cap V_{\text{commute},B})^\perp. \tag{B.48}$$

'A = 1 $\Longrightarrow$ W = fail' With the same reasoning as above we find

$$\mathbf{v}_{A=1} + \mathbf{v}_{W,fail} - \mathbf{v} \in ((V_A \oplus V_W) \cap V_{\text{commute},A})^\perp. \tag{B.49}$$

In total, the epistemic state $(V, \mathbf{v})$ and the measurements of Alice, Bob, Ursula, and Wigner need to fulfil:

(a) $V_B \subset V_{\text{commute},U} \oplus V_U$
(b) $V_A \subset V_{\text{commute},B} \oplus V_B$
(c) $V_W \subset V_{\text{commute},A} \oplus V_A$

(d) $\mathbf{v}_{U,ok} + \mathbf{v}_{W,ok} - \mathbf{v} \in ((V_U \oplus V_W) \cap V)^{\perp}$
(e) $\mathbf{v}_{B=1} + \mathbf{v}_{U,ok} - \mathbf{v} \in ((V_B \oplus V_U) \cap V_{\text{commute},U})^{\perp}$
(f) $\mathbf{v}_{A=1} + \mathbf{v}_{B=1} - \mathbf{v} \in ((V_B \oplus V_A) \cap V_{\text{commute},B})^{\perp}$
(g) $\mathbf{v}_{A=1} + \mathbf{v}_{W,fail} - \mathbf{v} \in ((V_A \oplus V_W) \cap V_{\text{commute},A})^{\perp}$

Let us assume we have found a state $(V, \mathbf{v})$ and measurements such that the above chain of reasoning holds, and $P(ok, ok) \neq 0$. Such a state would lead to a paradox. In the following, we show that such a state and measurements cannot exist.

For all $\mathbf{v}_W$ there exist $\mathbf{v}_A, \mathbf{v}_B$, and $\mathbf{v}_U$ be as in equation (B.39). Without loss of generality, we can choose equivalent measurement outcomes such that $\mathbf{v}_{B=1} \in V_U^{\perp}$ and $\mathbf{v}_{U,ok} \in V_B^{\perp}$. Then it holds that $\mathbf{v}_U - \mathbf{v}_B \in V_{\text{commute},U}$ and $\mathbf{v}_U - \mathbf{v}_B \in V_B \oplus V_U$. Thus, equation (B.47) implies that:

$$-\mathbf{v}_B^T \mathbf{v}_{B=1} + \mathbf{v}_U^T \mathbf{v}_{U,ok} + \mathbf{v}_B^T \mathbf{v} - \mathbf{v}_U^T \mathbf{v} = 0. \tag{B.50}$$

With the same argument we can find the following conditions:

$$-\mathbf{v}_B^T \mathbf{v}_{B=1} + \mathbf{v}_A^T \mathbf{v}_{A=1} + \mathbf{v}_B^T \mathbf{v} - \mathbf{v}_A^T \mathbf{v} = 0, \tag{B.51}$$

$$-\mathbf{v}_W^T \mathbf{v}_{W,fail} + \mathbf{v}_A^T \mathbf{v}_{A=1} - \mathbf{v}_A^T \mathbf{v} + \mathbf{v}_W^T \mathbf{v} = 0. \tag{B.52}$$

Subtracting equation (B.51) from equation (B.52) find the condition:

$$\mathbf{v}_W^T \mathbf{v}_{W,fail} - \mathbf{v}_W^T \mathbf{v} + \mathbf{v}_B^T \mathbf{v}_{B=1} - \mathbf{v}_B^T \mathbf{v} = 0. \tag{B.53}$$

Adding up equations (B.53) and (B.50) we find the condition

$$-\mathbf{v}_W^T \mathbf{v}_{W,fail} + \mathbf{v}_W^T \mathbf{v} + \mathbf{v}_U^T \mathbf{v}_{U,ok} - \mathbf{v}_U^T \mathbf{v} = 0. \tag{B.54}$$

If $P(ok, ok) \neq 0$ it holds that for all $\mathbf{w} \in (V_U \oplus V_W) \cap V$

$$\mathbf{w}^T (\mathbf{v}_{U,ok} + \mathbf{v}_{W,ok} - \mathbf{v}) = 0. \tag{B.55}$$

In particular, it holds that $\mathbf{w} = \mathbf{v}_U - \mathbf{v}_W \in (V_U \oplus V_W) \cap V$. Thus we can conclude that the following condition holds

$$-\mathbf{v}_W^T \mathbf{v}_{W,ok} + \mathbf{v}_U^T \mathbf{v}_{U,ok} - \mathbf{v}_U^T \mathbf{v} + \mathbf{v}_W^T \mathbf{v} = 0. \tag{B.56}$$

We can subtract equation (B.54) from equation (B.56) and find

$$\mathbf{v}_W^T (\mathbf{v}_{W,ok} - \mathbf{v}_{W,fail}) = 0. \tag{B.57}$$

Because $\mathbf{v}_U, \mathbf{v}_A, \mathbf{v}_B$ as in equation (B.39) can be found for all $\mathbf{v}_W$ it holds that $\mathbf{v}_{W,ok} - \mathbf{v}_{W,fail} \in V_W^{\perp}$. Thus, $\mathbf{v}_{W,ok} - \mathbf{v}_{W,fail}$ correspond to the same measurement outcome, as they have same valuation for any $\mathbf{v}_W \in V_W$. In summary, if there is a state and measurements such that the paradoxical chain of reasoning would hold, we have shown that then the two measurement outcomes that would lead to a paradox have to be the same outcome. Therefore, no such paradoxical chain of reasoning is possible. □

Appendix C. Review of quantum processes and experiments

C.1. Quantum measurements as physical processes

The usual way to describe the measurement process in quantum theory is to start with von Neumann measurements [1], which project the system into one of the eigenstates of the observable – such a measurement can be characterized by a set of projectors. However, von Neumann measurements only represent a certain class of measurements, as they contain all information about the observable. Generally, we are also interested in measurements which extract information only partially – while they reduce the uncertainty about the observable, they do not remove it completely. These generalized measurements are known as POVMs (positive operator-valued measurements)[11]. Operationally, POVMs can be implemented by introducing another quantum system, an ancilla (which can play the part of a pointer or a memory), performing a joint unitary on both systems, and then subjecting the ancilla to a von Neumann measurement. For example, the simplest way to measure a qubit in its computational basis with projectors $\{|0\rangle\langle 0|_S, |1\rangle\langle 1|_S\}$ is to perform a joint CNOT gate on the system and the memory (figure 1(b)).

Now let us consider the memory update for the case of the continuous system. For a system $\mathcal{H}_S$, we define the orthonormal basis $\{|x\rangle_S \mid x \in \mathbb{R}\}$ such that the states of the basis fulfil:

$$\hat{x}|x\rangle_S = x|x\rangle_S. \tag{C.1}$$

Let us also introduce a memory system $\mathcal{H}_M$, isomorphic to the Hilbert space $\mathcal{H}_S$. Our aim is to describe an operation which would coherently copy the state of the system S to the system M. We define the $CNOT_X$ gate as the transformation:

$$CNOT_X : |x_1\rangle_S \otimes |x_2\rangle_M \to |x_1\rangle_S \otimes |x_2 + x_1\rangle_M. \tag{C.2}$$

We can call $CNOT_X$ a *X-memory update*, as it is written w.r.t. the position basis. If we choose $x_2 = 0$, then the position of the first system S (the one being measured) is copied into the memory system M. After performing the memory update on the state $|x_1\rangle_S \otimes |0\rangle_M$, the state evolves to $|x_1\rangle_S|x_1\rangle_M$. Physically, this can be implemented as the action of the Hamiltonian $H_{SM} = \hat{X}_S \otimes \hat{P}_M$ for time $t = 1$, which couples two systems in the following way (we assume $\hbar = 1$):

$$\begin{aligned} U_{SM}(t)\left(\sum_k \alpha_k |x_k\rangle_S \otimes |x\rangle_M\right) &= \exp\left(-\frac{it}{\hbar}(X_S \otimes P_M)\right)\left(\sum_k \alpha_k |x_k\rangle_S \otimes |x\rangle_M\right) \\ &= \sum_k \alpha_k |x_k\rangle_S \otimes \exp\left(-ix_k P_M\right)|x\rangle_M \\ &= \sum_k \alpha_k |x_k\rangle_S \otimes |x + x_k\rangle_M. \end{aligned}$$

In principle, we can substitute the observable $\hat{X}_S$ on the system S with any other physical observable $\hat{A}_S = \sum_k a_k |a_k\rangle\langle a_k|_S$. In that case, the Hamiltonian takes the form $\hat{H}_{SM} = \hat{A}_S \otimes \hat{P}_M$, and the memory update $CNOT_X$ acts as:

$$CNOT_X : \sum \alpha_k |a_k\rangle \otimes |x\rangle_M \to \sum_k \alpha_k |a_k\rangle \otimes |x + a_k\rangle.$$

Suppose that the pointer M has the initial position wave function $\psi_0(x)$, for instance a Gaussian wave. The interaction Hamiltonian reads $\hat{H}_{SM} = g\hat{A}_S \otimes \hat{P}_M$ (where we also add the factor of g for quantifying the strength of the interaction), which leads to:

$$|\psi_k\rangle_M = e^{-itga_k\hat{P}_M}|\psi_0\rangle_M = e^{-itga_k\hat{P}_M}\int_{-\infty}^{+\infty} dx\, \psi_0(x)\, |x\rangle_M = \int_{-\infty}^{+\infty} dx\, \psi_0(x - t\, g\, a_k)\, |x\rangle_M.$$

The final global state is therefore:

$$\sum_k \alpha_k |a_k\rangle_S \otimes \int_{-\infty}^{+\infty} dx\, \psi_0(x - t\, g a_k)\, |x\rangle_M,$$

[11] They can be characterized by a generalising the set of projectors above: suppose that we pick $\{\Pi_i\}$ – a set of m operators with the only restriction $\sum_i \Pi_i^\dagger \Pi_i = \mathbb{1}$, where $m \leqslant n$.

where each outcome is correlated with a shift in the position of the pointer, and the weight of each peak corresponds to the probability of observing that outcome[12] (figure 1(c)). Let us just look at two examples that will be useful later to compare to the toy theory. For simplicity, suppose that we tune the interaction Hamiltonian and interaction time such that $tg = 1$, that the system S measured is continuous, and that we are measuring a continuous observable, $\hat{A}_S = \int_{-\infty}^{+\infty} dk\, a_k |a_k\rangle\langle a_k|_S$. In particular, let us see what happens when we measure the position ($\hat{A}_S = \hat{X}_S = \int_{-\infty}^{+\infty} dx\, x\, |x\rangle\langle x|_S$) of a system S that's initially in:

(a) A position eigenstate, $|x'\rangle_S$: the final state becomes

$$|x'\rangle_S \otimes |\psi_{x'}\rangle, \qquad \psi_{x'}(x) = \psi_0(x - x'). \tag{C.3}$$

If the initial pointer state was also a position eigenstate ($\psi_0(x) = \delta(x - x_0), |\psi_0\rangle_M = |x_0\rangle_M$), we obtain

$$|x'\rangle_S \otimes |x_0 + x'\rangle. \tag{C.4}$$

(b) A momentum eigenstate $|p'\rangle_S = (2\pi\hbar)^{-1/2} \int_{-\infty}^{+\infty} dx'\, e^{ip'x'/\hbar} |x'\rangle_S$: the final state is entangled:

$$\frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} dx'\, e^{ip'x'/\hbar} |x'\rangle_S |\psi_{x'}\rangle_M = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dx\, e^{ip'x'/\hbar}\, \psi_0(x - x')\, |x'\rangle_S |x\rangle_M. \tag{C.5}$$

Now we consider the case when the pointer is initialized in a position eigenstate $|x_0\rangle_M$ (that is, $\psi_0(x) = \delta(x - x_0)$). In this case, the global state is simply:

$$\frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{+\infty} dx'\, e^{ip'x'/\hbar} |x'\rangle_S |x_0 + x'\rangle_M = \int_{-\infty}^{+\infty} dp e^{-i(p'-p)x_0} |p\rangle_S |p' - p\rangle_M. \tag{C.6}$$

C.2. Frauchiger–Renner thought experiment

In this appendix, we present the technical derivation of the Frauchiger–Renner paradox in quantum theory [6]. Here, we assume that the reader has basic knowledge of the postulates and notation of quantum theory. This derivation is adapted from [23] without conditional state preparation.

C.2.1. Systems and initial state

We have two qubits R and S and two participants Alice and Bob, whose memory registers A and B are also modelled as one qubit each. There are two external agents Ursula and Wigner, whose quantum memories do not need to be explicitly modelled at this stage. The initial state of R, S, A, B is a Hardy state of R and S [22], and erased memories:

$$|\psi_0\rangle_{RSAB} = \frac{1}{\sqrt{3}} (|0\rangle_R |0\rangle_S + |1\rangle_R |0\rangle_S + |1\rangle_R |0\rangle_S) \otimes |0\rangle_A |0\rangle_B. \tag{C.7}$$

We will describe the protocol and the evolution of the state of $RSAB$ from the perspective of Ursula and Wigner, who put Alice and Bob's labs below the Heisenberg cut in a neo-Copenhagen interpretation (that is, they model Alice and Bob's measurements as reversible physical evolutions of quantum systems).

$t = 1$ Alice measures R in the Z basis and stores the result in her memory. From Wigner's perspective, she is now entangled with R:

$$|\psi_1\rangle_{RASB} = \frac{1}{\sqrt{3}} (|0\rangle_R |0\rangle_A |0\rangle_S + |1\rangle_R |1\rangle_A |0\rangle_S + |1\rangle_R |1\rangle_A |0\rangle_S) \otimes |0\rangle_B. \tag{C.8}$$

Notation: we changed the order of the subsystems because this will be easier later.

$t = 2$ Bob measures S in the Z basis and stores the result in his memory. The global state is now a Hardy state between the two labs,

$$|\psi_2\rangle_{RASB} = \frac{1}{\sqrt{3}} (|0\rangle_R |0\rangle_A |0\rangle_S |0\rangle_B + |1\rangle_R |1\rangle_A |0\rangle_S |0\rangle_B + |1\rangle_R |1\rangle_A |0\rangle_S |1\rangle_B). \tag{C.9}$$

[12] This measurement can be made sharper or weaker by tuning the parameters of the initial wave function of the pointer and the interaction time.

$t=3$ Ursula measures Alice's lab (RA) in the Bell basis; in particular we care about outcomes with non-zero probability, which correspond to the eigenstates:

$$|\text{ok}\rangle_{RA} = \frac{|0\rangle_R|0\rangle_A - |1\rangle_R|1\rangle_A}{\sqrt{2}}, \tag{C.10}$$

$$|\text{fail}\rangle_{RA} = \frac{|0\rangle_R|0\rangle_A + |1\rangle_R|1\rangle_A}{\sqrt{2}}. \tag{C.11}$$

$t=4$ Wigner measures Bob's lab (SB) in the Bell basis; again we label the two eigenstates with finite probability as:

$$|\text{ok}\rangle_{SB} = \frac{|0\rangle_S|0\rangle_B - |1\rangle_S|1\rangle_B}{\sqrt{2}}, \tag{C.12}$$

$$|\text{fail}\rangle_{SB} = \frac{|0\rangle_S|0\rangle_B + |1\rangle_S|1\rangle_B}{\sqrt{2}}. \tag{C.13}$$

C.2.2. Reasoning and analysis

The possibility of both Ursula and Wigner getting the outcome 'ok' is non-zero:

$$P[u=w=ok] = |(\langle \text{ok}|_{RA}\langle \text{ok}|_{SB})|\psi_2\rangle_{RASB}|^2 = \frac{1}{12}.$$

From now on, we post-select on this event. At time $t=3$, Ursula reasons about the outcome that Bob observed at $t=2$. Since we can regroup the global state before her measurement as:

$$|\psi_2\rangle_{RASB} = \sqrt{\frac{2}{3}}|\text{fail}\rangle_{RA}|0\rangle_S|0\rangle_B + \frac{1}{\sqrt{3}}|1\rangle_R|1\rangle_A|1\rangle_S|1\rangle_B. \tag{C.14}$$

Ursula concludes that the only possibility with non-zero overlap with her observation of $|\text{ok}\rangle_{RA}$ is that Bob measured $|1\rangle_S$. We can write this inference as '$u=\text{ok} \implies b=1$'. She can further reason about what Bob, at time $t=2$ thought about Alice's outcome at time $t=1$. Whenever Bob observes $|1\rangle_S$, he can use the same form of $|\psi_2\rangle_{RASB}$ to conclude that Alice must have measured $|1\rangle_R$. We can write this as '$b=1 \implies a=1$'. Finally, we can think about Alice's deduction about Wigner's outcome. Using the rewriting of the global state:

$$|\psi_2\rangle_{RASB} = \frac{1}{\sqrt{3}}|0\rangle_R|0\rangle_A|0\rangle_S|0\rangle_B + \sqrt{\frac{2}{3}}|1\rangle_R|1\rangle_A|\text{fail}\rangle_{SB}, \tag{C.15}$$

we see that Alice reasons that, whenever she finds R in state $|1\rangle_R$, then Wigner will obtain outcome 'fail' when he measures Bob's lab. That is, '$a=1 \implies w=\text{fail}$'. Thus, chaining together the statements (the same reasoning that allowed the reader to solve the three hats problem), we reach an apparent contradiction:

$$w=u=\text{ok} \implies b=1 \implies a=1 \implies w=\text{fail}.$$

That is, when the experiments stops with $u=w=\text{ok}$, the agents can make *deterministic* statements about each other's reasoning and measurement results, concluding that Alice had predicted $w=\text{fail}$, hence arriving to a logical contradiction.

ORCID iDs

Ladina Hausmann https://orcid.org/0000-0001-6827-7194
Nuriya Nurgalieva https://orcid.org/0000-0003-2443-2169
Lídia del Rio https://orcid.org/0000-0002-2445-2701

References

[1] John V N 1955 *Mathematical Foundations of Quantum Mechanics* (Princeton, NJ: Princeton University Press)
[2] Landauer R 1961 Irreversibility and heat generation in the computing process *IBM J. Res. Dev.* **5** 183–91
[3] Bennett C H and Shor P W 1998 Quantum information theory *IEEE Trans. Inf. Theory* **44** 2724–42
[4] Nurgalieva N and del Rio Lidia 2019 Inadequacy of modal logic in quantum settings *Electron. Proc. Theor. Comput. Sci.* **287** 267–97
[5] Fraser P, Nurgalieva N and Lídia del R 2020 Fitch's knowability axioms are incompatible with quantum theory (arXiv:2009.00321)
[6] Frauchiger D and Renner R 2018 Quantum theory cannot consistently describe the use of itself *Nat. Commun.* **9** 3711
[7] Vilasini V, Nurgalieva N and Lídia del R 2019 Multi-agent paradoxes beyond quantum theory *New J. Phys.* **21** 113028
[8] Spekkens R W 2005 Contextuality for preparations, transformations and unsharp measurements *Phys. Rev.* A **71** 052108
[9] Pusey M F 2012 Stabilizer notation for spekkens' toy theory *Found. Phys.* **42** 688–708
[10] G Chiribella and R W Spekkens (eds) 2016 Quasi-quantization: classical statistical theories with an epistemic restriction *Quantum Theory: Informational Foundations and Foils* (Dordrecht: Springer)
[11] Catani L and Browne D E 2017 Spekkens' toy model in all dimensions and its relationship with stabiliser quantum mechanics *New J. Phys.* **19** 073035
[12] Coecke B and Edwards B 2011 Spekkens's toy theory as a category of processes (arXiv:1108.1978)
[13] Coecke B, Edwards B and Spekkens R W 2011 Phase groups and the origin of non-locality for qubits *Electron. Not. Theor. Comput. Sci.* **270** 15–36
[14] Backens M and Duman A N 2016 A complete graphical calculus for spekkens' toy bit theory *Found. Phys.* **46** 70
[15] Comfort C and Kissinger A 2021 A graphical calculus for Lagrangian relations
[16] Hausmann L, Nurgalieva N and Lídia del R 2021 A consolidating review of Spekkens' toy theory (arXiv:2105.03277)
[17] Frauchiger D and Renner R 2018 Quantum theory cannot consistently describe the use of itself *Nat. Commun.* **9** 3711
[18] Spekkens R W 2007 Evidence for the epistemic view of quantum states: a toy theory *Phys. Rev.* A **75** 032110
[19] Lostaglio M and Bowles J 2021 The original wigner's friend paradox within a realist toy model *Proc. R. Soc.* A **477**
[20] Wigner E P 1961 Remarks on the mind-body question *The SCientist SPeculates* ed E J Good (London: Heinemann) pp 284–302Reprinted Wigner E P 1995 *Philosophical Reflections and Syntheses* (Berlin: Springer) pp 247–60
[21] Nurgalieva N and Renner R 2021 Testing quantum theory with thought experiments *Contemp. Phys.* **61** 1–24
[22] Hardy L 1993 Nonlocality for two particles without inequalities for almost all entangled states *Phys. Rev. Lett.* **71** 1665–8
[23] Nurgalieva N and Lídia del R 2019 Inadequacy of modal logic in quantum settings *EPCTS* **287** 267–97
[24] Boge F J 2019 Quantum information versus epistemic logic: an analysis of the Frauchiger–Renner theorem *Found. Phys.* **49** 1143–65
[25] Haddara M and Cavalcanti E G 2022 A possibilistic no-go theorem on the wigner's friend paradox (arXiv:2205.12223)
[26] Bartlett S D, Rudolph T and Spekkens R W 2012 Reconstruction of gaussian quantum mechanics from liouville mechanics with an epistemic restriction *Phys. Rev.* A **86** 012103
[27] Kochen S and Specker E P 1975 Logical structures arising in quantum theory *The Logico-Algebraic Approach to Quantum Mechanics* (Dordrecht: Springer) pp 263–76
[28] Cook R T 2004 Patterns of paradox *J. Symb. Logic* **69** 767–74
[29] Abramsky S, Barbosa R S, Kishida K, Lal R and Mansfield S 2015 Contextuality, Cohomology and Paradox *24th EACSL Annual Conf. on Computer Science Logic (CSL 2015) (Leibniz Int. Proc. in Informatics (LIPIcs))* vol 41, ed S Kreutzer (Dagstuhl: Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik) pp 211–28
[30] Pusey M F and Leifer M S 2015 Logical pre- and post-selection paradoxes are proofs of contextuality
[31] Karanjai A, Cavalcanti E G, Bartlett S D and Rudolph T 2015 Weak values in a classical theory with an epistemic restriction *New J. Phys.* **17** 073015
[32] Brukner Č 2018 A no-go theorem for observer-independent facts *Entropy* **20** 350
[33] Bong K-W, Utreras-Alarcón A, Ghafari F, Liang Y-C, Tischler N, Cavalcanti E G, Pryde G J and Wiseman H M 2020 A strong no-go theorem on the Wigner's friend paradox *Nat. Phys.* **16** 1199–205
[34] Gallier J 2011 Basics of affine geometry *Texts in Applied Mathematics* (New York: Springer) pp 7–63
[35] Wilding D, Johnson M and Kambites M 2013 Exact rings and semirings *J. Algebra* **388** 324–37
[36] Lam T Y 2004 *Introduction to Quadratic Forms Over Fields* (Providence, RI: American Mathematical Society)