Do qubits dream of entangled sheep? Quantum measurement without classical output

Noah Lupu-Gladstein\(^1\), Aharon Brodutch\(^1,2,3\), Hugo Ferretti\(^4\), Weng-Kian Tham\(^5\), Arthur Ou Teen Pang\(^6\), Kent Bonsma-Fisher\(^1,4\), and Aephraim M Steinberg\(^1,4\)

\(^1\) Department of Physics and Center for Quantum Information and Quantum Control, University of Toronto, 60 St George St, Toronto, Ontario M5S 1A7, Canada

\(^2\) IonQ, Inc, College Park, MD, United States of America

\(^3\) National Research Council of Canada, 100 Sussex Dr, Ottawa, Ontario K1A 0R6, Canada

\(^4\) Canadian Institute for Advanced Research, Toronto, Ontario M5G 1M1, Canada

* Authors to whom any correspondence should be addressed.

E-mail: nlupugla@physics.utoronto.ca, brodutch@ionq.co, kent.bonsma-fisher@nrc-cnrc.gc.ca and steinberg@physics.utoronto.ca

Keywords: quantum measurement, quantum foundations, quantum networks

Abstract

Quantum mechanics is usually formulated with an implicit assumption that agents who can observe and interact with the world are external to it and have a classical memory. However, there is no accepted way to define the quantum–classical cut and no \textit{a priori} reason to rule out fully quantum agents with coherent quantum memories. In this work, we introduce an entirely quantum notion of measurement, called a \textit{sensation}, to account for quantum agents that experience the world through quantum sensors. Sensations eschew probabilities and instead describe a deterministic flow of quantum information. We quantify the information gain and disturbance of a sensation using concepts from quantum information theory and find that sensations always disturb at least as much as they inform. Viewing measurements as sensations could lead to a new understanding of quantum theory in general and to new results in the context of quantum networks.

1. Introduction

It is decisive to recognize that, however far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms.

\textit{Niels Bohr [1]}

Measurement has been central to quantum mechanics since its inception. Historically, measurements were formalized with the notion of observables: Hermitian operators \([2, 3]\) that represent quantities like position, momentum, and spin. This formulation of measurement has since been generalized several times \([4–7]\), but none of these advances have questioned the basic notion that a measurement yields a classical outcome that can be copied. To account for classical outcomes, quantum mechanical models of measurement invoke an external, classical observer.

As quantum technology improves, we draw closer to an age where quantum machines make complex decisions using quantum computations that run on quantum data gathered with quantum sensors. These systems would be quantum agents in the sense that they autonomously sense and act upon the quantum world \([8, 9]\). When a quantum agent measures its surroundings, is it really natural to view the results of that measurement as merely classical? Should the agent be viewed as external to the quantum world, or an active participant within it (see figure 1)?

Our aim is to model quantum measurements as a quantum agent might: free of classical outcomes and external observers. A fully quantum paradigm will facilitate the development of new experiments, measurement techniques, communication protocols, and computational algorithms that do not fit easily in a classical framework. A complete description of quantum agents would not just model their observations, but
also their decisions and actions. In this work, we focus solely on observation and leave the matter of decisions and actions for further study. We assume that a quantum agent has some ability to decide when and how to observe a given system. The mechanics behind this decision process are beyond the scope of this work, seeing as modeling how a classical agent makes decisions is already a difficult problem. We take for granted that a quantum agent can distinguish between its external environment and ‘itself’, leaving the question of how a quantum agent even develops a sense of ‘self’ for future research. Although this work does not paint a complete picture of quantum agency, its ideas have already inspired a forthcoming experiment [10].

Agents, both classical and quantum, experience the world through sensors. We define a sensor as a system with a memory that can interact with another system and store information from that system in its memory. We compare two types of relationships between an observer and a sensor:

- Classical—The sensor is separate from the observer and the sensor’s memory decoheres faster than the observer can process it
- Quantum—The sensor is a part of the observer and the sensor’s memory remains coherent throughout being processed by the observer

For a given observer, we classify a sensor as classical or quantum accordingly. A quantum sensor can take in incoherent information like a classical sensor, but it can also take in and store quantum information about a quantum system.

As humans, we are naturally equipped only with classical sensors and the standard notion of measurement (see figure 2) reflects that. Traditionally, a measurement involves a system and an observable property of that system. The measurement is completed when an observer obtains a result, i.e. a classical record which can be copied and shared. Von Neumann developed a pragmatic quantum measurement model that is satisfactory for predicting our classical experience, yet invokes an observing agent external to the system [2]. There have been many attempts to adapt von Neumann’s model and keep the observer inside the system (notably collapse theories and the approaches following Everett and Bohm [11–13]), but all cling to the central role of observables.

We adopt a radical view of measurement: a measurement is any interaction between two systems in which the final state of one system depends on the initial state of the other. To avoid confusion with the historically laden term ‘measurement’, we call such an interaction a sensation. The result of a sensation is a quantum state, not a classical label. Completing a sensation does not require an external agent.

Any traditional measurement can be cast as a sensation followed by a transition to a classical value (the sensation is the so-called pre-measurement [14]), but there are sensations that transcend traditional measurements. For example, consider a sensor that swaps a state in its memory with a state in the environment. This interaction does not result in a distinct set of classical outcomes, so it cannot be associated

![Figure 1. Agency with internal (quantum) agents compared to external (classical) agents.](image-url)
Figure 2. Quantum measurements with external agents. (a) The traditional approach to measurement characterised by the POVM elements \( \{ E_k \} \). The probability for an outcome \( a_k \) is 
\[
p(a_k) = \text{tr}(E_k \rho) \]
and the corresponding minimally disturbing state transformation is 
\[
|\psi\rangle \rightarrow \frac{\sqrt{E_k}}{\sqrt{\langle \psi | E_k | \psi \rangle}} |\psi\rangle
\]
(b) The von Neumann scheme for a measurement of a non-degenerate observable \( A \) provides a more detailed description than the textbook approach. It includes a quantum measurement device \( M \) and an amplification process whereby the information is copied onto multiple registers. The external observer reads out the state of some of these registers and records a classical result \( a_k \). The cut between the external agent and the other subsystems has no observable consequences.

with an observable or positive operator valued measure (POVM). It is nevertheless a sensation because the final state of the sensor depends on the initial state of the environment.

We explore how an agent that uses sensations might view the world. We do not attempt to describe the inner workings of a quantum agent, nor attempt to accommodate the way humans with classical memories observe the world through events and probabilities. Instead, our aim is to describe information flow in a deterministic theory with as few assumptions as possible. In this sense our work is also very different from approaches that put the observer at the center (e.g. [15, 16]) which tend to begin with probabilities as a primitive. Our results give a consistent way of understanding information in a deterministic (collapse-free) interpretation of quantum mechanics as a complete theory in the sense that all physical objects (including observers) and all dynamical processes (including sensation) are described by the same rules.

Our work takes a world-view which is similar to the Everettian (many-worlds, relative-state) interpretation, i.e. starting from section 3 we will consider the sensor as part of the quantum world and assume a unitary theory. However, our work is not an interpretation of quantum mechanics. It is instead the beginning of a framework for studying quantum agents.

We begin in the next section with a description of the traditional—classical sensor — approach to measurement. In section 3 we define sensations with quantum sensors. We introduce the result channel as the extension of the POVM into the deterministic quantum regime, and give two examples of sensations: the von Neumann sensation, and the swap sensation (section 3.1.2). We then show how such swap sensations can be implemented in practice (section 3.2). In section 4 we develop a method to quantify information gain, and disturbance. Using these tools, we compare the von Neumann and swap sensations. Implications of quantum agency are discussed in section 5 where we provide examples involving multiple agents sharing information, emphasising one of the limitations of the swap sensation, and mention the relation to quantum computing (section 5.2). Our main conclusion (section 6) is that a complete quantum treatment of agency which is free from anthropocentric pre-conceptions can lead to new results. It is conceivable that by ignoring the possibility of quantum agents with quantum sensors we are in danger of making an oversight similar to the one made by the founders of modern computing and information theory (many of whom had in-depth knowledge of quantum mechanics) who missed or ignored the possibility of quantum information processing.

2. External observers: observables and POVMs

John von Neumann first described the measurement of an observable \( A \) in terms of three separate subsystems:

1. A system to be measured with an associated Hilbert space \( \mathcal{H}_S \) on which the observable \( A \) is an Hermitian operator
2. A sensor with Hilbert space $\mathcal{H}_M$ on which there are two canonically conjugate operators: the pointer operator $P_M$ and its canonical conjugate $Q_M$.

3. An external agent with no mathematical representation.

The measurement process according to von Neumann’s scheme can be broken into two stages: interaction and readout. In the interaction stage a Hamiltonian of the type

$$H_i \propto A_S \otimes Q_M$$

(1)

is switched on to couple the system and sensor. The result can then be amplified (see figure 2(b)). At the readout stage, an external agent records the state of the sensor. Prior to the mathematical derivation, von Neumann (invoking Bohr [2, footnote 207]) argues that the agent being external is not unique to the quantum regime, and that the precise cut between the agent and the sensor is arbitrary even in the classical case. The aim of the derivation, which would later be called the von Neumann measurement scheme, was to regain the classical intuition that measurements are independent of the specific choice of the cut between the sensor and the agent. This motility of the cut is demonstrated by showing that the agent and the sensor can be treated as a single composite system which can be cut into subsystems arbitrarily without modifying the outcome probabilities or the state update rules.

To make a comparison between sensations and traditional measurements, we list a number of features that arise from the von Neumann scheme:

- Result—After the interaction, the sensor’s memory is in a state that generally depends on the initial system.
- Broadcastability—The result can be copied and broadcast to others.
- Constrained back-action—A second sensor interacting identically with the same system yields the same result up to statistical uncertainty.
- Motility of the cut—There is no accepted scientific theory that separates the sensor from the external observer.

The first feature is essential in the definition of a measurement [13]. The next two points play a key role in communication since they allow multiple agents to communicate the results of their measurements by making copies and broadcasting them. Such communication engenders objectivity in the sense that different agents can agree on some specific property by making individual measurements of the same system and comparing their results [13, 17–19]. Historically, these or some subset of these features are taken as the defining properties of measurements (e.g. [20]). Our aim in this work is to explore the consequences of keeping only the requirement that measurements have results that depend on the initial system.

### 3. The quantum observer

The features of von Neumann measurements match our classical experience, but do they also capture everything a quantum observer can experience? To start answering this question, we define sensations, the more general class of interactions enabled by access to a quantum sensor.

#### 3.1. Defining sensation

As before, the system $S$ to be observed is associated with the Hilbert space $\mathcal{H}_S$. The sensor is a part of an observer $O$ and interacts with the system. In contrast to traditional measurement, we consider the observer itself a quantum system. We associate the memory in the observer’s sensor with the Hilbert space $\mathcal{H}_O$. The sensation is a procedure which includes a unitary operation $U$ on $S$, $O$, and possibly other subsystems. We call these other subsystems the environment $E$, with Hilbert space $\mathcal{H}_E$, because it is conceptually separate from the measured system and the observer. Despite the terminology, the environment need not be noisy and could simply represent an additional sensor (see section 3.1.1).

For simplicity, we assume that free evolution can be ignored on the time scales of the sensation$^5$. To keep the notation simple, we will also use no subscripts when describing an operator on the joint $SEO$ Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_E \otimes \mathcal{H}_O$. Furthermore, we will use the term channel to refer to completely-positive trace-preserving maps. We denote such channels $\mathcal{C} : \mathcal{X} \to \mathcal{Y}$ when they take states on $\mathcal{H}_X$ to states on $\mathcal{H}_Y$.

A sensation is driven by an interaction between the system, observer, and environment. The interaction is defined by a unitary $U$ and the initial composite state $\chi_{EO}$ of the environment $E$ and the observer $O$.

---

$^5$ Note that in practice the free evolution can lead to significant corrections, see for example section 3.2.1 below.
An interaction $\rho$ identity, similarly the result channel plays the central role in the following definition of a measurement. One would say that a channel is a measurement if the associated POVM elements are not all proportional to the identity, i.e., the interaction is a quantum channel $M(U, \chi_{EO})$:

$$M(U, \chi_{EO})[\rho_S] = \text{tr}_E[U \rho_S \otimes \chi_{EO} U^\dagger].$$  

(See figure 3(a)). To simplify our notation we will henceforth drop the arguments in $M(U, \chi_{EO})$.

For any interaction $M : S \to SO$, we also define a result channel $\mathcal{R}_{\mathcal{M}} : S \to O$ that maps the state $\rho_S$ onto a state in $\mathcal{H}_O$ called the result,

$$\mathcal{R}_{\mathcal{M}}[\rho_S] = \text{tr}_S[M[\rho_S]].$$  

This result channel can be seen as the analogue of the POVM, but whereas the POVM maps system states to probabilities, the result channel $\mathcal{R}_{\mathcal{M}}$ maps system states to observer states. Similarly, we introduce the back-action

$$\mathcal{B}_{\mathcal{M}}[\rho_S] = \text{tr}_O[M[\rho_S]]$$

which is a channel that describes how the interaction modifies $S$. This channel is the standard channel associated with measurement back-action (as defined in [6] for example). In a theory with external agents we would say that a channel is a measurement if the associated POVM elements are not all proportional to the identity, similarly the result channel plays the central role in the following definition of a measurement.

**Definition 1 (Sensation, result).** An interaction $M : S \to SO$ with an associated result channel $\mathcal{R}_{\mathcal{M}} : S \to O$ is a sensation (of $S$ by $O$) if and only if there are two system states $\rho_S, \sigma_S$ such that $M[\rho_S] \neq M[\sigma_S]$. The result is the quantum state $\mathcal{R}_{\mathcal{M}}[\rho_S]$.

Note that since the channel is linear, it is sufficient to consider pure states. That is, a channel $\mathcal{M} : S \to SO$ with an associated result channel $\mathcal{R}_{\mathcal{M}} : S \to O$ is a sensation when there are two pure states $\rho_S, \sigma_S$ such that $\mathcal{R}_{\mathcal{M}}[\rho_S] \neq \mathcal{R}_{\mathcal{M}}[\sigma_S]$.

The asymmetry in our definition of sensations as channels from $S$ to $SO$ (as opposed to $SO$ to $S$) may seem like a ‘bug’, but it is actually a crucial feature. It provides a formal lever to distinguish the observer from the observed. If two quantum agents interact, they will each experience a different sensation because the two agents will define themselves as the observer and the other as the system. A coupling between two agents $A$ and $B$ establishes an objective channel from $AB$ to $AB$, but the resulting sensation is subjective in that it depends on whether $A$ is seen as the observer or the system.

We will soon explore concrete examples of specific sensations, but first we will give two examples of interactions that are not sensations. An interaction driven by any unitary of the form $U_{SEO} = U_{SE} \otimes U_O$, is not a sensation. Second, consider an interaction of the form $U_{SEO} = e^{i\chi_{EO}} \neq 1$. The interaction is not a sensation in the special case where $\chi_{EO}$ happens to commute with $Y_{EO}$, although it is a sensation otherwise.

*Figure 3.* Quantum measurements with a quantum agent: A measurement interaction $\mathcal{M}$ (blue box) fills an observer’s memory $O$ with information about a system $S$ in the presence of an environment $\mathcal{E}$. The interaction is defined in terms of two objects: a unitary $U$ (which couples the system $S$, environment $\mathcal{E}$ and the observer’s memory $O$), and the initial $\mathcal{E}O$ state $\chi_{EO}$. The interaction takes an initial system state $\rho_S$, to a final $SO$ state $M[\rho_S]$. The result $\mathcal{R}_{\mathcal{M}}[\rho_S]$ is the reduced state encoded in the observer’s memory. The interaction induces the result channel $\mathcal{R}_{\mathcal{M}}$ (dashed red on bottom right) from system states to observer memory states and the back-action $\mathcal{B}_{\mathcal{M}}$ (dashed red on bottom left), a channel from system states to disturbed system states. The entire process is deterministic. The interaction is considered a sensation (according to definition 1) as long as $\mathcal{R}_{\mathcal{M}}[\rho_S]$ is not a constant function of $\rho_S$, i.e., $\mathcal{M}$ is a sensation of $S$ whenever the result depends on the state of $S$. 

Mathematically, the interaction is a quantum channel $M(U, \chi_{EO}) : S \to SO$ on an arbitrary system state $\rho_S$:

$$M(U, \chi_{EO})[\rho_S] = \text{tr}_E[U \rho_S \otimes \chi_{EO} U^\dagger].$$

(See figure 3(a)). To simplify our notation we will henceforth drop the arguments in $M(U, \chi_{EO})$.
3.1.1. Example 1: the von Neumann measurement

In the von Neumann measurement, the environment is a classical sensor. To measure an observable $A = \sum_k a_k A_k$ with eigenspace projectors $A_k$, we start with an initial product state written as $|\psi\rangle |0\rangle |0\rangle$, where $|\psi\rangle$ is an arbitrary $S$ state and $|0\rangle |0\rangle$ is a fixed initial $EO$ state. We then generate an interaction $W_{SE}$ between the environment and the system and follow with an interaction $V_{EO}$ between the system and the observer, so $U = V_{EO} W_{SE}$. These interactions and the initial $EO$ states are chosen so that

$$|\psi\rangle |0\rangle |0\rangle \xrightarrow{W_{SE}} \sum_k [A_k |\psi\rangle] |a_k\rangle |0\rangle \xrightarrow{V_{EO}} \sum_k [A_k |\psi\rangle] |a_k\rangle |k\rangle$$

where the environment states $|a_k\rangle$ are orthogonal to each other, as are the memory states $|k\rangle$ (we assume that the dimensions of $H_S$ and $H_O$ are large enough). Each of these interactions can be generated by a Hamiltonian in the form of equation (1).

The result of the interaction is the final memory state $\sum_k \langle k | A_k |\psi\rangle |k\rangle$. The observer’s information about the system is captured by the correlations between $S$ and $O$. The result depends on the initial system state in the sense that each of the orthogonal sensor states ‘points’ at a corresponding system state with weights determined by the initial system state.

3.1.2. Example 2: the SWAP sensation

We now go to the extreme situation where the agent learns everything about the quantum state of the system by ‘swapping’ the latter, lock, stock, and barrel with her quantum sensor’s memory register. We refer to this operation as the swap sensation. For simplicity, we choose $O$ to be the same dimension as $S$ so $U$ can be the standard SWAP gate $\sum_{k} A_k |\psi\rangle |0\rangle |0\rangle \xrightarrow{V_{EO}} A_k |\psi\rangle |k\rangle |k\rangle$ (see figure 4(b)) which puts the state $|\psi\rangle$ in the observer’s sensor. Following this measurement, the observer’s sensor contains everything about the state of $S$ before the interaction, at the cost of a very strong back-action. The sensation’s result cannot be shared across multiple sensors and there is no system observable associated with it. The agent senses the state $|\psi\rangle$, rather than a specific property of $|\psi\rangle$.

The back-action of a swap sensation is so strong, one may wonder whether it is even appropriate to call $S$ the system once its contents have been entirely swapped with a part of the observer. In practice, $S$ is often a small piece of a much larger system. For example, if an observer uses a swap sensation to store a photon in a quantum memory (explored further in section 3.2.1), she only senses one small portion of the entire electromagnetic field. Whatever state was swapped out from her sensor’s memory would start to participate in the dynamics of electromagnetism, so there is a physical sense in which $S$ is still the system, even after its contents have been completely disturbed.

3.1.3. Example 3: the decohered swap

A more realistic scenario involves a memory which is open to the environment. Analysis in this case depends on the precise dynamics of the coupling to the environment. One possibility that allows us to recapture the...
classically intuitive broadcasting feature is dephasing of $O$ in some preferred basis $\{ |k\rangle \}$, through interaction with the environment (see figure 4(c)). The full transformation for the $SO$ subsystems would then be

$$\rho_S \otimes \tau_O \rightarrow \tau_S \otimes \rho_O \rightarrow \tau_S \otimes \sum_k \langle k | \rho | k \rangle \langle k | \rho | k \rangle_O.$$  

(6)

The observer’s state is then similar to the one in the von Neumann scheme above (equation (5)) with equivalence in the case $A_k = |k\rangle \langle k|$, however there are no longer any correlations with $S$. Correlations with $E$ would be similar to those of equation (5) and in principle $E$ could include many copies of the result. As we will show below, the swap-and-decohere process is similar to the behaviour of a photon detector. The main feature we want to highlight for now is that once decoherence kicks in, the sensation can be associated with an observable like a traditional measurement. The main features of all three sensations discussed above are presented in table 1 using terminology which we develop further in section 4.

3.2. Physical models for a swap sensation

The swap sensation is an extreme example of a new capability granted by quantum sensors with ideal quantum memories. We now show that this sensation is not only physically feasible, but that the dephased swap is also fairly commonplace.

3.2.1. Photon detectors

Photoelectric devices that measure light intensity are among the most common ways to sense the world. These devices typically work by swapping the state of the incoming light field with the state of photoelectrons, which then interact with each other and rapidly decohere (see appendix B). This detection process is essentially the dephased swap measurement described above in section 3.1.2. The fact that one of the most ubiquitous ways of observing the world does not follow the von Neumann scheme and moreover acts like a dephased SWAP operation (with all information in the field being lost) is significant. In particular any assertion (see [13] for example) that good observations should be repeatable by multiple observers must be assessed with this in mind. While photodetection with rapid decoherence does not fit within the narrow model of von Neumann measurements, it is nevertheless a classical measurement in the sense that its outcomes can be modelled as classical labels rather than coherent quantum states. Only the number (as opposed to the relative phase) of photoelectrons detected is meaningful because the environment washes out any definite phase relation between photoelectron number states.

3.2.2. Optical quantum memories

If the environment’s action on a photodetector can be sufficiently tamed, it becomes more than a mere measurement. It becomes a quantum memory for an optical mode. Such a memory is not beholden to any particular optical observable. Its dynamics are described by sensations where the optical mode is the system to be measured, and the memory mode represents the sensor. It has been demonstrated that photon storage in an off-resonant Raman quantum memory is equivalent to a beamsplitter interaction between a flying photon mode ($a^\dagger$) and a stationary spin-wave excitation ($b^\dagger$) [21,22] (see figure 5). The ‘reflectivity’ in this beamsplitter interaction is given by the storage efficiency of the memory which can, in theory, approach unity [23]. This storage process acts like a SWAP between the bosonic optical and spin-wave memory modes. Any superposition of photon number states in the optical mode becomes the same superposition of spin-wave excitations in the spin-wave mode and vice versa.

Perfect or near-perfect quantum memories are yet to be demonstrated, but it is reasonable to expect that such devices will exist in the not-too-distant future, most likely with some type of error correction mechanisms to increase coherence times. Some of these memories, in particular those directly connected to communication channels and sensors, will likely be optical memories with a mechanism that resembles a SWAP operation with an incoming light mode. A sufficiently advanced quantum computer might interact with the world mostly through swap sensations (see section 5.2).

4. Information theory for quantum observers

The information gained from a measurement is usually quantified through some function of the probability distribution associated with the possible measurement results, viewed as classical labels. This approach cannot be applied as is to our deterministic framework that defines results as quantum states. In this section we present an intuitive way to think about information gain and disturbance that can be applied in the quantum-agent scenario. We start by introducing maximally informative and maximally disturbing sensations in section 4.1 and point out that their non-informative and non-disturbing counterparts are not sensations. We then (section 4.2) introduce an additional agent $A$ who is initially entangled with $S$ (see
4.1. Maximal information gain and disturbance

A sensation $\mathcal{M}$ is maximally informative whenever the observer’s final memory state and the details of $\mathcal{M}$ are sufficient to produce at least one instance of a state identical to any arbitrary initial system state.

**Definition 2 (Maximally informative sensation).** A sensation $\mathcal{M} : S \rightarrow SO$ with an associated result channel $\mathcal{M}_1^{-1} : O \rightarrow S$ such that the composition of channels $\mathcal{M}_1^{-1} \circ \mathcal{M} = 1$ is the identity channel.

It is tempting to demand that a maximally informative sensation enables the observer to produce not just one, but any number of states identical to a given input. While that requirement might be appropriate for a classical observer, quantum observers must abide by the no-cloning theorem [24, 25]. A related, but distinct notion in standard quantum mechanics is an informationally complete measurement [26]. These measurements eventually allow an observer to produce many copies of a measured state, but only after measuring many identical instances of that state.

The opposite limit of a maximally informative sensation is an interaction in which the observer’s final state is independent of the initial system state. Such a non-informative interaction is in fact not a sensation under definition 1.

**Definition 3 (Non-informative interaction).** An interaction $\mathcal{M} : S \rightarrow SO$ with an associated result channel $\mathcal{M}_1^{-1} : O \rightarrow S$ such that $\mathcal{M}_1^{-1} \circ \mathcal{M} = 1$ is non-informative when $\mathcal{M}_1^{-1} [\sigma_S] = \mathcal{M}_1^{-1} [\sigma_S]$ for any pair of states $\rho_S, \sigma_S$ in $S$.

A sensation is maximally disturbing if the final system state is independent of the initial system state.

**Definition 4 (Maximally disturbing sensation).** A sensation $\mathcal{M} : S \rightarrow SO$ with an associated back-action $\mathcal{M}_1 : S \rightarrow S$ such that the composition of channels $\mathcal{M}_1 \circ \mathcal{M} = 1$ is maximally disturbing when $\mathcal{M}_1 [\rho_S] = \mathcal{M}_1 [\sigma_S]$ for any pair of states $\rho_S, \sigma_S$ in $S$.

By contrast, a non-disturbing interaction is one that allows the system to perfectly recover its initial state from its final state.

**Definition 5 (Non-disturbing interaction).** An interaction $\mathcal{M} : S \rightarrow SO$ with an associated back-action $\mathcal{M}_1 : S \rightarrow S$ is called non-disturbing when there exists a quantum channel $\mathcal{M}_1^{-1} : S \rightarrow S$ such that the composition of channels $\mathcal{M}_1^{-1} \circ \mathcal{M} = 1$.

4.2. Quantifying information gain and disturbance

We seek an information gain function $G(\mathcal{M})$ and a disturbance function $D(\mathcal{M})$ to quantitatively describe the results and back-action of an interaction $\mathcal{M}$. These functions must be compatible with the definitions given above for maximally informative, non-informative, maximally disturbing, and non-disturbing. Furthermore, we demand that no local operation on $O$ after the interaction increases $G(\mathcal{M})$. Similarly, no local operation on $S$ after the interaction may decrease $D(\mathcal{M})$. These requirements are summarized below.

- $G(\mathcal{M}) = 0$ when $\mathcal{M}$ is non-informative and strictly positive otherwise.
- $G(\mathcal{M})$ maximizes for a given system $S$ when $\mathcal{M}$ is maximally informative.
G(\(\mathcal{R}_{\mathcal{B}G}\)) is non-increasing under local operations on \(\mathcal{O}\): if \(\mathcal{M}' = \mathcal{C}_\mathcal{O} \circ \mathcal{M}\) and \(\mathcal{C}_\mathcal{O}\) is a local channel on \(\mathcal{O}\), then \(G(\mathcal{M}') \leq G(\mathcal{M})\).

\(D(\mathcal{B}_{\mathcal{M}}) = 0\) when \(\mathcal{M}\) is non-disturbing and strictly positive otherwise.

\(D(\mathcal{B}_{\mathcal{M}})\) maximizes for a given system \(S\) when \(\mathcal{M}\) is maximally disturbing.

\(D(\mathcal{B}_{\mathcal{M}})\) is non-decreasing under local operations on \(S\): if \(\mathcal{M}' = \mathcal{C}_\mathcal{O} \circ \mathcal{M}\) and \(\mathcal{C}_\mathcal{O}\) is a local channel on \(\mathcal{O}\), then \(D(\mathcal{M}') \geq D(\mathcal{M})\).

To construct functions \(G(\mathcal{R}_{\mathcal{B}G})\) and \(D(\mathcal{B}_{\mathcal{M}})\) that respect the above properties, we will consider a second agent \(A\) who is maximally entangled with \(S\) before the measurement interaction. Maximally entangling \(S\) with \(A\) reflects the assumption that the observer has no prior knowledge about the system state \(\rho_S^G\). For concreteness, we choose the entangled state to be \(\Phi^+_{AS} = \frac{1}{\sqrt{2}} \sum_{k=1}^d |k\rangle_A |k\rangle_S\) where \(d = \dim(H_S) = \dim(H_A)\) and \(\{|k\rangle\}_{k=1}^d\) is a complete, orthonormal basis. We will quantify the correlations between two subsystems \(X\) and \(Y\) via mutual information, defined as

\[ I_{X:Y}(\tau_{XY}) = S(\tau_X) + S(\tau_Y) - S(\tau_{XY}) \quad (7) \]

where \(S(\tau) = -\text{tr}[\tau \ln \tau]\) is the von Neumann entropy of a state \(\tau\). Mutual information enjoys some properties \([7, 27, 28]\) which are specifically useful for our purposes:

\[ I_{X:Y}(\tau_{XY}) \geq 0 \] with equality if and only if \(\tau_{XY}\) is a product state, i.e. \(\tau_{XY} = \tau_X \otimes \tau_Y\).

The mutual information reaches its maximal value \(2 \log d\) when \(\tau_{XY}\) is a maximally entangled state.

Mutual information is non-increasing under local operations.

We can now quantify both the information gain and back-action by looking at the change in mutual information after the measurement interaction. We define information gain \(G(\mathcal{R}_{\mathcal{B}G})\) as \(I_{A:O}\) after the interaction \(\mathcal{M}\)

\[ G(\mathcal{R}_{\mathcal{B}G}) \equiv I_{A:O}((1_A \otimes \mathcal{R}_{\mathcal{B}G}) [\Phi^+ \langle \Phi^+ |_{AS}]) \quad (8) \]

and the disturbance \(D(\mathcal{B}_{\mathcal{M}})\) as the drop in \(I_{A:S}\) after \(\mathcal{M}\) (see figure 6)

\[ D(\mathcal{B}_{\mathcal{M}}) \equiv I_{A:S}([\Phi^+ \langle \Phi^+ |_{AS}) - I_{A:S}((1_A \otimes \mathcal{B}_{\mathcal{M}}) [\Phi^+ \langle \Phi^+ |_{AS})]. \quad (9) \]

To see that these quantities make sense, we examine the qualitative behaviour of \(I_{A:O}\) and \(I_{A:S}\). First we note that once we fix \(\dim(H_S) = \dim(H_A) = d\), the initial correlations are \(I_{A:S}(\Phi^+ \langle \Phi^+ |_{AS}) = 2 \log d\). Since \(O\) is initially uncorrelated with \(A\) and \(S\), we also have \(I_{A:S,O}(\Phi^+ \langle \Phi^+ |_{AS} \otimes \chi_O) = 2 \log d\). The measurement interaction \(\mathcal{M}\) is a local operation on \(SO\), so \(I_{A:S,O}(\mathcal{M}[\Phi^+ \langle \Phi^+ |_{AS})] \leq 2 \log d\). Ignoring \(O\) to obtain the back-action \(\mathcal{B}_{\mathcal{M}}\) is a local operation, so \(I_{A:S}(\mathcal{B}_{\mathcal{M}}[\Phi^+ \langle \Phi^+ |_{AS})] \leq 2 \log d\). Ignoring \(S\) to obtain the result channel \(\mathcal{R}_{\mathcal{B}G}\) is also local, so \(I_{A:O}(\mathcal{R}_{\mathcal{B}G}[\Phi^+ \langle \Phi^+ |_{AS})] \leq 2 \log d\). As a result,

\[ 0 \leq G(\mathcal{R}_{\mathcal{B}G}), D(\mathcal{B}_{\mathcal{M}}) \leq 2 \log d. \quad (10) \]

4.3. Examples

Let us consider the simple case where each subsystem is a qubit. Both \(E\) and \(O\) are initially in the state \(|0\rangle\) and the maximally entangled \(AS\) state is \(\Phi^+ = \frac{1}{\sqrt{2}}(|00⟩ + |11⟩)\). In the von Neumann scheme, the final \(ASEO\) state is \(\frac{1}{\sqrt{2}}[|0000⟩ + |1111⟩]\), which gives \(G(\mathcal{R}_{\mathcal{B}G}) = D(\mathcal{B}_{\mathcal{M}}) = \log 2\).

For a swap sensation, the final \(ASO\) state is \(\frac{1}{\sqrt{2}}[|00⟩ + |11⟩]\), so the reduced \(AO\) state is maximally entangled and \(G(\mathcal{R}_{\mathcal{B}G}) = 2 \log 2\) (this can also be seen by noting that \(\mathcal{R}_{\mathcal{B}G}\) is the \(S \rightarrow O\) identity). The channel \(\mathcal{B}_{\mathcal{M}}\) takes all system states to \(|0⟩_{S}\) so recovery is impossible and the disturbance is maximal: \(D(\mathcal{B}_{\mathcal{M}}) = 2 \log 2\).

The decohered swap sensation yields \(\frac{1}{\sqrt{2}}[|00⟩ + |11⟩]\) as the final \(ASO\) state. The final \(AO\) state is the classically correlated state \(\frac{1}{\sqrt{2}}[|00⟩ |00⟩ + |11⟩ |11⟩]\), which has \(G(\mathcal{R}_{\mathcal{B}G}) = \log 2\). The final \(AS\) state is uncorrelated \(\frac{1}{\sqrt{2}}[|0⟩ |0⟩ + |1⟩ |1⟩] \otimes |0⟩\) and has \(D(\mathcal{B}_{\mathcal{M}}) = 2 \log 2\).

It is possible to modify the approach by encoding any prior knowledge the observer has in the initial \(AS\) state so that \(ASEO\) would be the purification of the ensemble of initial states, but the details of such modifications are beyond the scope of this work.

For the sake of simplicity, we consider only finite dimensional Hilbert spaces.
Figure 6. Information gain and disturbance: (a) A second agent $A$ is initially entangled with $S$ so that they are maximally correlated, $I_{A:S} = 2 \log d$. The observer $O$ is initially uncorrelated with $A$. (b) The information gain is defined as the change in mutual information between $A$ and $O$ after the sensation (equation (8)). (c) The disturbance is the change in mutual information between $A$ and $S$ (equation (9)).

Table 1. Comparing three types of sensations: the von Neumann scheme, the swap, and the swap with decoherence (see sections 3.1.1 and 3.1.2 for details). Motility of the cut refers to the possibility of identifying different cuts between the observer and the environment. The result is broadcastable when it can be copied and shared with other observers, and the measurement is repeatable if a second measurement of the same type will produce the same outcome. Reversibility refers to the resources required for reversing the operation so that the system will be restored to its original state (the resources are either access to $SO$ or access the entire $SEO$). Information gain and disturbance (Defined in section 4.2) are given in terms of two reference points defined in section 4.1. For a system of dimension $d$, the Heisenberg limit is $\log d$ and maximal is $2 \log d$. The swap sensation obtains maximal information at the cost of maximal disturbance. The decohered swap performs as well or worse than the von Neumann scheme on all accounts and is essentially equivalent to a destructive von Neumann measurements like photodetection.

|                      | von Neumann | SWAP | SWAP with decoherence |
|----------------------|-------------|------|-----------------------|
| Motility of the cut  | Yes         | No   | Yes                   |
| Broadcastable        | Yes         | No   | Yes                   |
| Repeatable           | Yes         | No   | No                    |
| Reversibility        | Requires environment | Yes | Requires environment |
| Information gain $G(\mathcal{M})$ | Heisenberg | Maximal | Heisenberg |
| Disturbance $D(\mathcal{M})$ | Heisenberg | Maximal | Maximal |

These results generalize naturally to $d$ dimensional systems. The von Neumann scheme has $G(\mathcal{M}_{\text{vN}}) = D(\mathcal{M}_{\text{vN}}) = \log d$, which we call the Heisenberg limit. The swap sensation has $G(\mathcal{M}_{\text{SWAP}}) = D(\mathcal{M}_{\text{SWAP}}) = 2 \log d$ and the decohered swap sensation has $G(\mathcal{M}_{\text{SWAP with decoherence}}) = \log d$ and $D(\mathcal{M}_{\text{SWAP with decoherence}}) = 2 \log d$. These results are summarised in table 1.

The fact that the information gain of a swap sensation is $2 \log d$ warrants an explanation. Intuitively, a swap should provide at least some information, even if it is not obvious how much. It should provide at least as much information as any von Neumann sensation, because once the system state falls into the memory of the agent’s sensor, she is free to post-process it with any von Neumann sensation she sees fit. Yet a naive analysis of the swap sensation suggests it provides no information. `To wit, the swap sensation maps any system state $|\psi\rangle$ to a separable state $|0\psi\rangle$. The system and observer are completely uncorrelated, no matter the initial system state. Any attempt to quantify information gain based on correlations of observer and system observables will fail to match the intuition that a swap offers non-zero information. It is a triumph of the sensation formalism that it succeeds in assigning the swap non-zero information. In fact, it assigns the swap the highest possible value of $2 \log d$ for a given system dimension $d$. 

10
4.4. Information-disturbance relation

The examples above suggest a relationship between the information gain $G(\mathcal{R})$ and disturbance $D(\mathcal{B})$ of a measurement interaction $\mathcal{M}$. In the first two examples (von Neumann and swap) we had $G(\mathcal{R}) = D(\mathcal{B})$ and in the last example (decohered swap) we had $D(\mathcal{B}) > G(\mathcal{R})$. In all three examples, the observer disturbed at least as much information as she gained. In the following theorem, we prove that $0 \leq D(\mathcal{B}) - G(\mathcal{R})$. Furthermore, the examples suggest that when the sensation yields a pure state, the disturbance equals the information gain. We show that this is true in general by proving that the difference between the disturbance and information gain is bounded from above by twice the joint entropy of the $ASO$ system.

**Theorem 1.** Consider a measurement interaction $\mathcal{M} : S \rightarrow SO$ of a system $S$ with a Hilbert space of finite dimension $d$. Let $|\Phi^+\rangle_{AS}$ be the state of that system maximally entangled with an ancilla $A$. Let $\rho_{ASO} = (1_A \otimes \mathcal{M}) [|\Phi^+\rangle \langle \Phi^+|_{AS}]$ be the state after the interaction. Then the information gain $G(\mathcal{R}) = I_{A,O}(\rho_{AO})$ and disturbance $D(\mathcal{B}) = 2 \log d - I_{A,S}(\rho_{AS})$ satisfy

$$0 \leq D(\mathcal{B}) - G(\mathcal{R}) \leq 2S(\rho_{ASO}).$$  (11)

**Proof.** We start by proving $0 \leq D(\mathcal{B}) - G(\mathcal{R})$. Our proof is based on the strong subadditivity of quantum information, which can be expressed as the following inequality valid for any tripartite state: $\tau_{XYZ}$: [7, equation 11.114]

$$I_{X,Y}(\tau_{XY}) + I_{X,Z}(\tau_{XZ}) \leq 2S(\tau_X).$$  (12)

Applying this inequality to $\rho_{ASO}$ shows that

$$I_{A,O}(\rho_{AO}) \leq 2S_A(\rho_A) - I_{A,S}(\rho_{AS}).$$  (13)

The interaction $\mathcal{M}$ does not act on the ancilla, so the $A$ subsystem remains in its initial state, which was half of a maximally entangled state. Consequently, $S(\rho_A) = \log d$. Substituting this value, along with the definitions for disturbance and information gain reveals $G(\mathcal{R}) \leq D(\mathcal{B})$.

Next we prove $D(\mathcal{B}) - G(\mathcal{R}) \leq 2S(\rho_{ASO})$. The proof relies on a corollary of strong subadditivity known as the Araki–Lieb triangle inequality [29], which states that any bipartite state $\tau_{XY}$ obeys the following inequality:

$$|S(\tau_X) - S(\tau_Y)| \leq S(\tau_{XY}).$$  (14)

We also reuse the fact that $S(\rho_A) = \log d$.

$$D(\mathcal{B}) - G(\mathcal{R}) = 2 \log d - I_{A,S}(\rho_{AS}) - I_{A,O}(\rho_{AO})$$  (15)

$$= 2S(\rho_A) - S(\rho_{AS}) - S(\rho_{AO})$$  (16)

$$= S(\rho_A) - S(\rho_{AO}) + S(\rho_{AO}) - S(\rho_{AS})$$  (17)

$$\leq 2S(\rho_{ASO})$$  (18)

The last line follows from applying the triangle inequality to obtain $S(\rho_{AS}) = S(\rho_O) \leq S(\rho_{ASO})$ and $S(\rho_{AO}) - S(\rho_{PS}) \leq S(\rho_{ASO})$.

\[\square\]

The upper bound in equation (11) shows that a sufficient condition for a sensation to inform as much as it disturbs is for it to map pure state to pure states, and thus output an $ASO$ state with zero entropy. However, this sufficient condition is not a necessary condition. The von Neumann measurement considered in section 4.3 produced the $ASO$ state $\frac{1}{\sqrt{2}}[|0000\rangle + |1111\rangle]$. Although the reduced $ASO$ state is mixed, the sensation has equal disturbance and information gain. A complete classification of the necessary and sufficient conditions for a sensation to inform as much as it disturbs would essentially give a recipe for ideal, minimally disturbing measurements, but it is beyond the scope of this work.

5. Agency in the quantum world

In the previous sections we developed a framework to study measurements involving agents equipped with quantum sensors. Within this framework we extended the classically motivated concepts of observation and uncertainty into the quantum regime without trying to force a connection to our own experience. While our experience of the world seems to be classical at least in the sense that our observations can be described using
POVMs, there seems to be no law of nature that can rule out the more general type of observation described in section 3 above (it should however be emphasised that a wide body of work has been devoted to finding such laws; e.g. the quantum Darwinism [17, 18, 30] and spontaneous collapse [31] programs). If we believe quantum theory applies at all levels, then quantum agents can exist in principle, for example in the form of sufficiently advanced fault tolerant quantum computers with peripheral quantum sensors (see figure 8). In this section we discuss quantum agency through well known results developed in the context of traditional measurements and explore possible directions for research in quantum computing.

5.1. Simultaneous observations

We now move to a scenario involving two observers $O_A$ and $O_B$ who try to sense the same system simultaneously (see figure 7). In the case of a von Neumann sensation of the same observable, this is a well studied scenario [30]. The two sensations commute and the outcomes are the same as those arising from the scenario where the observers perform the sensation one after the other, or a situation where the two observers monitor the same sensor. In all of these cases the observers do not influence each other’s sensation and all can reach the Heisenberg limit and moreover have correlated results.

What happens if two observers attempt to sense the same system simultaneously, both using the same swap sensation? The SWAP operator $S_{SO}$ that swaps system states with memory states in $O_A$’s sensor does not commute with the otherwise identical SWAP operator $S_{SO_B}$ for $O_B$. As a result, the two observers’ sensations will influence each other. We model the simultaneous swap using the interaction Hamiltonian $H_{i2} = S_{SO_A} + S_{SO_B}$. The system and observers are initially in the product state $|\psi\rangle_S|0\rangle_{O_A}|0\rangle_{O_B}$. The Hamiltonian $H_{i2}$ is applied for a time $t$, leading to the state $e^{-itH_{i2}}|\psi\rangle_S|0\rangle_{O_A}|0\rangle_{O_B}$. At every moment in time, we associate a sensation with each observer, $O_A$ and $O_B$, denoted $M_A$ and $M_B$ respectively. Both of these sensations refer to the same dynamical process, save for an exchange in the notion of observer and environment. Since the situation is symmetric, we will only study it from observer $O_A$’s perspective.

We explicitly calculate the time-evolved state in $CS$. From equation (C5), we compute the information gain and disturbance and plot it for $d = 2$ in figure 7(c). For comparison with this ‘double swap’ scenario, we plot the information gain and disturbance for a ‘single swap’ sensation in which only $O_A$ observes the system and the interaction Hamiltonian is $H_{i1} = S_{SO_A}$. Initially, each observer gains nearly as much information from the double swap as the single swap. By the time each observer gains 0.5 bits of information (half the Heisenberg Limit), the presence of the other observer is felt and the double swap information begins to lag behind the single swap. While the single swap reaches the maximal information gain and disturbance of 2 bits at time $t = \pi/2$, the double swap caps out sooner at $t = \pi/3$ with an information gain of 0.91 bits and a disturbance of 1.68 bits. If each observer had instead performed the same von Neumann measurement on the system, they would have each gained 1 bit of information, slightly more than the double swap. They also would have only disturbed the system by 1 bit, significantly less than the double swap.

Figure 7. Simultaneous observations using a swap interaction (a) Two observers $O_A$ and $O_B$ simultaneously couple to the same system using the swap Hamiltonian. (b) A circuit diagram for the sensation with an ancilla $A$ added to account for disturbance and information gain. (c) Informational quantities in bits as a function of time. The solid blue curve denotes the information gained by each observer individually during the double swap. Dashed purple shows the disturbance to the original system. It also equals the information gained by both observers when they cooperate and act as a single, joint observer. Olive dots denote the single swap information gain and disturbance, which describes what would have happened if only one observer attempted to sense the system. Green dashed dots show the Heisenberg limit for reference, which corresponds to the information gain and disturbance of an ideal von Neumann measurement. At the cost of extra disturbance compared to a von Neumann measurement, swap observers can nearly achieve the Heisenberg limit without cooperation and significantly exceed it with cooperation.
The double swap seems worse in terms of both information gain and disturbance than a von Neumann measurement, but there is more to the story. So far, we have only considered how much information $O_A$ and $O_B$ gain individually. However, if they cooperate we can treat them as a single joint observer and ask how much information $O = O_A O_B$ gains. For this joint observer, the interaction with the system is unitary, so their information gain equals their disturbance and maximizes to 1.68 bits. This information gain significantly exceeds the 1 bit of information that one, two, or indeed any number of cooperative observers using the same von Neumann measurement can extract. Even though each von Neumann observer individually gains 1 bit from their measurement, it is all the same, redundant bit. At the cost of extra disturbance, two observers using a swap sensation can gain almost as much information individually as a von Neumann measurement while gaining significantly more information as a single, cooperative unit.

5.2. Quantum computers as agents

Work on mechanized observers who are part of the quantum system goes back at least as far as Everett [13] who imagined quantum automata observing the system in a generic way, but focused his attention on von Neumann type measurements and a classical experience where results are classical labels. Later, Albert [32, 33] showed that a quantum automaton with access to its own memory registers could perform measurements whose (classically interpreted) results seem paradoxical. These works inspired much of the early theoretical work on quantum computing, in particular Deutsch’s pioneering work on universal quantum Turing machines [34]. Most subsequent research however, regards quantum computers as devices to be used by classical agents.

The usual quantum information processing paradigm involves classical inputs and outputs. In a sampling problem for example, the input is a classical description of some quantum circuit ending with a sequence of binary measurements, and the output is a string of bits that represents a sample from the probability distribution for the measurement outcomes. This picture is certainly realistic, but it can be modified in a quantum internet scenario where quantum computers are directly connected to quantum sensors that feed a quantum state as the input for a computation. These computers could interact with other quantum computers through quantum interconnects, so that they only share quantum information. If perfect quantum communication networks were available, we would expect most of the communication between quantum computers to involve no classical results (see figure 8). Quantum states generated by one computer could then be used to make decisions about which circuit to encode on other computers. While, most circuits are realized presently with classical control parameters, quantum controls have been used to run superpositions of circuits [36–38]. Quantum computers communicating in this way would have entangled circuits and classical results would only appear when these computers interact with us. Such communication would allow many small quantum computing nodes to be joined together into a single distributed quantum computer.

Work on quantum computing theory tends to be based on the assumption that quantum theory can be applied as-is to arbitrarily large systems, yet the vast majority of work takes an anthropocentric view: classical input, classical output. Notably this approach even extends to problems in the QMA computational complexity class [39] which involve a proof encoded in a quantum state that is sent to a verifier who outputs

---

8 Note that the possibility of having quantum computers as observers is sometimes mentioned in the context of quantum foundations [35].

9 and perhaps never if certain science fiction scenarios turn out to be correct.
a classical 'accept' or 'reject' with an appropriate probability. Exceptions include delegated quantum computing [40, 41] and compression protocols [42–44] which are usually framed as intermediate stages in a larger classical-in classical-out task. To the best of our knowledge, none of these have been studied as computational complexity classes. For example, one might study a type of decision problem where the language is defined in Hilbert space and the verifier's output is a quantum state.

There are a number of difficulties in extending ideas in computational complexity to the fully quantum regime. One technical issue involves dealing with imperfections and handling non-orthogonal output states. Classically it is customary to expect that the right result would appear with a high probability (say 2/3) and that the results can be distinguished from each other. In the quantum case, one could try to deal with imperfections by asking that the real output state be close to some ideal output state. The metric used for the distance should make sense in the context of the problem. The usual distance measures used in quantum mechanics are interpreted in terms of probabilities for measurement outcomes, however as we showed here (section 4.1), these can still make sense in a theory without probabilities. The same issue appears when dealing with non-orthogonal output states. There seems to be no inherent reason to claim that orthogonal states are more distinguishable to a quantum observer than non-orthogonal ones. More precisely: without the Born rule, there is no obvious notion of distinguishability which can be applied to orthogonal states. It might however be reasonable to expect that since unitary operators preserve the inner product, it has a special meaning. In fact this is probably unavoidable once we assign meaning to reduced density operators, or start thinking about broadcasting information to multiple observers [20].

6. Conclusions

The orthodox approach to measurement in quantum mechanics is based on an assumption that the observer is external to the quantum system and can observe the quantum world through observables. This approach works incredibly well to describe all known quantum phenomena and is unlikely to fail as long as observers (such as ourselves) have classical memories. Quantum memories do not seem to be a naturally occurring phenomenon, but much of the effort in quantum information processing is aimed at increasing coherence times and it is increasingly likely that engineered quantum memories will become commonplace in the not-too-distant future.

If we are to believe that the current technological trajectory will continue, we should entertain the idea of attributing agency to sufficiently advanced quantum computers. As we have shown in this work, such quantum agents would have access to a broader range of actions than those available to us. In particular, they would use sensations rather than conform to the narrow definition of measurement which has been adopted in quantum theory. We have explored these ideas and presented the swap sensation (section 3.1.2 and figure 4(b)) as an extreme case of quantum observation. Our ideas have already inspired an upcoming experiment [10] which empirically measures the information flow of several sensations.

We began with a generalized definition of measurement (definition 1) whose deterministic result is a quantum state recorded in the quantum memory of an observer's sensor. To handle this more general formalism (which includes the von Neumann scheme as a special case) we modified the basic mathematical tools of quantum measurement theory and replaced the POVM with the result channel $\mathcal{R}_{\text{mem}}$, which takes system states to states of the memory in the observer's sensor (see figure 3).

Using these tools, we suggested a method for quantifying information gain and disturbance (section 4 and figure 6). These definitions led to expected results for von Neumann measurements while faithfully describing the swap sensation's exotic ability to provide complete information at the cost of maximal disturbance. We showed that the disturbance of a sensation always equals or exceeds the information gained, which places a fundamental constraint on the information processing power of quantum agents. We provided a sufficient condition for a sensation to be ideal in the sense that it informs as much as it disturbs, but left open the task of identifying conditions that are both sufficient and necessary. We used these quantities to analyze a thought experiment in which two agents both attempt to use a swap sensation on the same system at the same time. The thought experiment revealed that cooperation amongst quantum agents is essential for making the most of certain sensations.

Our results are the first steps in studying fully quantum agents, but there is still much work to be done. We have only studied how quantum agents observe their surroundings, but the processes by which quantum agents might decide and act using those observations is still ripe for exploration. While much work is being done on studying new mechanisms for decoherence-free interactions, there is very little work on protocols, algorithms, and computational paradigms that involve purely quantum agents. While our framework formally separates the observer from the observed, the physical mechanism by which a quantum agent gains a sense of self is still a mystery.
We hope that our work encourages others to consider scenarios that involve multiple quantum agents who can interact with each other in a purely coherent, quantum mechanical way, for example a world-wide network of intelligent quantum computers (see figure 8). As we work towards building sophisticated quantum machines and a quantum internet, we should give careful consideration to how these machines would think and interact with the world around them. Learning from past oversights, we should not ignore the possibility of information being truly quantum mechanical.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank J E Sipe for useful discussions. This work was funded by Grant Number FQXi-RFP-1819 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation. All authors were also supported by by NSERC and the Fetzer Franklin Fund of the John E Fetzer Memorial Trust. A M S is a fellow of CIFAR.

Appendix A. Traditional quantum measurements

A.1. Observables, outcomes and probabilities

Measurements in quantum mechanics have traditionally been associated with ‘observables’ which are mathematically represented as Hermitian operators. Each observable \( A \) has unique set of eigenstates (eigenspaces if it is degenerate) and eigenvalues so that it can be written as \( A = \sum a_k a_k^\dagger \) where \( a_j \neq a_k \) unless \( j = k \) and \( \{ a_k \} \) are projectors onto orthogonal subspaces. The (real) eigenvalues \( \{ a_k \} \) are usually used to label the possible measurement results, while the projectors \( \{ a_k \} \) are used for calculating the probabilities for each result. For a system initially in the state \( \rho \), the probability that a measurement of \( A \) will yield the result \( a_k \) is given by the Born rule \( p(a_k) = \text{tr}(a_k \rho) \)

This formalism can be extended by noting that the set of orthogonal projectors \( \{ a_k \} \) can be replaced by a set of positive operators \( \{ E_k \} \) (not necessarily orthogonal) with \( \sum E_k = I \). Such a set of positive operators is called a POVM and its elements can be plugged into the Born rule to produce a probability distribution

\[ p(a_k) = \text{tr}(E_k \rho) \tag{A1} \]

for a set of possible measurement results labeled \( \{ a_k \} \). For simplicity we will refer to equation (A1) as the Born rule.

In many cases we are also interested in the back-action of the measurement on the measured system. The POVM does not provide sufficient information to predict a unique outgoing system state, although it can be used to identify a minimally disturbing measurement (as defined for example in [6]). For a minimally disturbing measurement, a result \( a_k \) with an associated POVM element \( E_k \) implies the transformation \( \rho \rightarrow \rho E_k / \text{tr}(E_k \rho) \) on the measured system (see figure 2(a)). In the special case of a measurement of an observable \( A = \sum a_k A_k \), the minimally disturbing measurement is called a von Neumann measurement and the update rule is \( \rho \rightarrow (a_k \rho a_k^\dagger) / \text{tr}(a_k \rho) \). When \( A_k \) is a rank-1 projector this rule has the simple form \( \rho \rightarrow A_k \) i.e. the state projects onto the eigenstate associated with \( a_k \). The disturbance is considered minimal since the update rule leaves eigenstates of \( A \) undisturbed.

The above description of a measurement (Born rule and state update rule) is in most cases sufficient for making predictions about the outcomes of experiments where the precise details of the measurement procedure and the observer can be ignored. It does however imply a hard cut between the observer and the measured system.

A.2. The von Neumann scheme

Von Neumann’s approach can be used to model the measurement process associated with any POVM by treating a measurement apparatus as a quantum mechanical system initially in a state \( \Upsilon \). The measurement begins with some interaction \( U_{SM} \) so that \( \rho \otimes \Upsilon \rightarrow U_{SM} (\rho \otimes \Upsilon) U_{SM}^\dagger \) after which the measurement result is encoded in the state of \( M \). To read out the result, an observer would need to measure an observable \( \Pi_M = \sum a_k \Pi_k \) on \( M \), where \( \Pi_k \) are orthogonal projectors and the labels \( a_k \) are distinct. The probability for a result \( a_k \) would then be \( P(a_k | \rho_S) = \text{tr} [(\mathbb{1} \otimes \Pi_k) U_{SM} (\rho \otimes \Upsilon) U_{SM}^\dagger] \). This equation can be written in the form of Born’s rule (A1) by identifying the POVM element \( E_k = \text{tr}_M [U_{SM}^\dagger (\mathbb{1} \otimes \Pi_k) U_{SM} (\mathbb{1} \otimes \Upsilon)] \) so that
$P(a_k|\rho_S) = \text{tr}_M[E_k \rho_S]$. The procedure also gives the post-measurement state of the observed system given a result $a_k$ as

$$\text{tr}_M[\epsilon(\rho_S) U_{SM}[\rho_S \otimes \chi^f_{SM}]]$$ [45]. The term von Neumann scheme is often used for the special case where this procedure is applied to the inner workings of a von Neumann measurement. For this measurement $U_{SM}$ is generated by a Hamiltonian in the form of equation (1) (see figure 2 above and section 3.1.1 below for more details).

The scheme above is more detailed than the Born and state update rules, but it is not complete since it involves an observation of the measurement device (via $\Pi_M$) with no details on how this measurement is constructed. It then begs the question ‘how is the measurement device observed?’ to which we could give the same answer ad infinitum. At this point von Neumann invoked the external observer (which he previously justified). It is however possible to treat the result quantum mechanically—i.e. as a state $|a_k\rangle$ rather than a classical label ‘$a_k$’ - so that no external observers are required. This fully coherent approach comes with interpretational issues, but it can be argued that these are no more problematic than the alternative. As we will show in section 3 the possibility of encoding the measurement result in a quantum state allows a more general definition of measurement where some observers can perceive the world in a way which cannot be modeled in the language of POVMs.

**A.3. Side remark: locality and the born rule**

Quantum mechanics with external (classical) observers is fairly well defined at least from an operational perspective where certain systems (e.g. humans, cameras, etc) are postulated to be classical and external. In such an approach one generally accepts the Born rule and a certain version of the collapse postulate as part of the quantum-to-classical transition. This in turn allows us to use some tools which we usually take for granted, especially those that have been adopted from statistical mechanics and information theory. One which is of particular significance here is the reduced state. Consider a system which is composed of two distinct subsystems: $\mathcal{S}$ and $\mathcal{M}$ such that its description is a state $|\psi\rangle_{SM}$ in the tensor product Hilbert space $\mathcal{H}_{SM} = \mathcal{H}_S \otimes \mathcal{H}_M$. The reduced state of the system is $\rho_S = \text{tr}_M(|\psi\rangle_{SM} \langle \psi|)$, where $\text{tr}_M$ is the partial trace over $\mathcal{H}_M$. This reduced state contains all the information necessary for calculating probabilities for the outcomes of measurements via the Born rule (A1). From the operational perspective of the external observer, this reduced state contains all there is to know about the system. The notion of a reduced state can also be extended to dynamics. Consider an initial product state $\rho_{SM} \otimes \tau_M$ on $\mathcal{H}_{SM} = \mathcal{H}_S \otimes \mathcal{H}_M$ and some Schrödinger evolution described by the unitary $U_{SM}$. The local (reduced) dynamics is described by the completely positive trace preserving map $\mathcal{E}(\rho_{SM}) = \text{tr}_M[U_{SM} \rho_{SM} \otimes \tau_M U_{SM}^\dagger]$ (also called the quantum channel).

Quantum theory with quantum observers provides a more difficult situation than a theory with external observers, and requires a more careful treatment of fundamental postulates and their interpretation. Our approach here is to take the ‘standard’ Schrödinger-picture quantum theory (without collapse) at face value, as is often done in many-worlds interpretations [10, 47, 48]. In particular, we assume that the reduced density operators are valid complete descriptions of the local states, and that reduced dynamics provide a complete description of the local dynamics. We note that this is not necessarily the situation in hidden variable theories such as those following de Broglie and Bohm’s pilot wave [11], and that even in the many worlds interpretation the justification for assigning meaning to reduced density operators requires a complicated argument and non-trivial assumptions [49].

**Appendix B. Photodetection**

The usual treatment of photodetection [50, 51] begins with an atomic system interacting with an electric field through an interaction Hamiltonian of the form $H_i = \vec{d} \cdot \vec{E}$, where $\vec{d}$ is the dipole moment of the atom and $\vec{E}$ is the electric field. This Hamiltonian has the von Neumann form (1) with the atom acting as the measurement apparatus and the field $\vec{E}$ as the measured operator, and so we might be tempted to say that this is a von Neumann measurement of the field. However, free evolution cannot be neglected at optical frequencies, and the free evolution term in the Hamiltonian does not commute with the interaction term. The dynamics are usually treated by moving to a rotating frame and making the rotating wave approximation to get rid of terms that oscillate rapidly (see [51, section 4.3 ]).

We treat the detector as a two-level atom and assume that the atoms are initially in the ground state and that the excited state is an ion and a free photoelectron. Each absorbed photon leads to the release of a photoelectron (see [51, section 5.2] for a detailed derivation) which induces an irreversible amplification sequence whose details depend on the specifics of the detector. Ideally the detector would perform a perfect

10 see Deutsch and Hayden [46] for a different (Heisenberg based) approach to many worlds.
SWAP followed by dephasing so that the incoming photon number would be perfectly correlated with the amplified output signal; however, practical issues mean that photon number sensitivity is usually far from perfect.

**Appendix C. Simultaneous swap**

We calculate the state evolution and information exchange of the simultaneous swap sensation described in section 5.1. There are two relevant (unnormalized) eigenstates of $H_{SW}$:

$$|e_2\rangle = |\psi 00\rangle + |0\psi 0\rangle + |00\psi\rangle,$$

$$|e_{-1}\rangle = 2|\psi 00\rangle - |0\psi 0\rangle - |00\psi\rangle. \quad (C1)$$

They have respective eigenvalues 2 and $-1$. These allow us to expand the initial state $|\psi 00\rangle$ in the eigenbasis of $H_{SW}$.

$$|\psi 00\rangle = \frac{|e_2\rangle + |e_{-1}\rangle}{3}. \quad (C2)$$

The evolved state is

$$e^{-iH_{SW}}|\psi 00\rangle = \frac{e^{-2it} |e_2\rangle + e^{it} |e_{-1}\rangle}{3}$$

$$= \frac{e^{-2it} + 2e^{it}}{3} |\psi 00\rangle + \frac{e^{-2it} - e^{it}}{3} \left[ |0\psi 0\rangle + |00\psi\rangle \right]. \quad (C3)$$

When $t$ is an integer multiple of $\frac{\pi}{2}$, the state evolves back to its original uncorrelated state times a phase factor. For these $t$ values, the interaction is not a sensation by definition 1.

To calculate information gain and disturbance for the $d = 2$ case, we add a third agent $A$ initially entangled with $S$, so that the initial $ASO_AO_B$ state is

$$|\Psi\rangle = \frac{|0000\rangle + |1100\rangle}{\sqrt{2}}. \quad (C4)$$

The state evolves to

$$e^{-iH_{AS}}|\Psi\rangle = \frac{1}{\sqrt{2}} \left[ e^{-2it} |0000\rangle + \frac{e^{-2it} + 2e^{it}}{3} |1100\rangle + \frac{e^{-2it} - e^{it}}{3} \left[ |1010\rangle + |1001\rangle \right] \right]. \quad (C5)$$

The information gain is the mutual information of the reduced $AO_A$ state,

$$\rho_{AO_A} = \frac{1}{2} \left[ |00\rangle + \frac{1 - e^{3it}}{3} |11\rangle \right] \left[ |00\rangle + \frac{1 - e^{-3it}}{3} |11\rangle \right] + \frac{7 + 2 \cos(3t)}{18} |10\rangle \langle 10|. \quad (C6)$$

The disturbance is 2 minus the mutual information of the reduced $AS$ state,

$$\rho_{AS} = \frac{1}{2} \left[ |00\rangle + \frac{1 + 2e^{3it}}{3} |11\rangle \right] \left[ |00\rangle + \frac{1 + 2e^{-3it}}{3} |11\rangle \right] + \frac{2 - 2 \cos(3t)}{9} |10\rangle \langle 10|. \quad (C7)$$

**ORCID iD**

Kent Bonsma-Fisher @ https://orcid.org/0000-0002-5581-8892

**References**

[1] Bohr N 1983 Discussions with einstein on epistemological problems in atomic physics Quantum Theory and Measurement ed J A Wheeler and W H Zurek (Princeton University Press)

[2] von Neumann J and Beyer R T (Translator) and Wheeler N A (eds) 2018 Mathematical Foundations of Quantum Mechanics (Princeton University Press)

[3] van Hove L 1958 Von Neumann’s contributions to quantum theory Bull. Am. Math. Soc. 64 95–99

[4] Kirkpatrick K A 2006 Translation of Luders’ “Über die Zustandsanderung durch den Messprozess” G. Luders, Ann. Phys. (Leipzig) 8 322–328 (1951) Ann. Phys., Lpz. 15 663–70

[5] Brian Davies E 1976 Quantum Theory of Open Systems (Academic)

[6] Wiseman H M and Milburn G J 2009 Quantum Measurement and Control (Cambridge University Press)

[7] Nielsen M and Chuang I 2010 Quantum Computation and Quantum Information (Cambridge University Press)

[8] Dunjko V, Taylor J M and Briegel H J 2016 Quantum-enhanced machine learning Phys. Rev. Lett. 117 130501
