

Environment as a witness: Selective proliferation of information and emergence of objectivity in a quantum universe

Harold Ollivier,^{1,2} David Poulin,^{1,3,*} and Wojciech H. Zurek⁴

¹Perimeter Institute, 31 Caroline Street N, Waterloo, Ontario, Canada N2L 2Y5

²Projet Codes, INRIA, Boîte Postale 105, F-78153 Le Chesnay, France

³Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

⁴Theory Division, MS-B213, LANL, Los Alamos, New Mexico 87545, USA

(Received 26 April 2005; published 31 October 2005)

We study the role of the information deposited in the environment of an open quantum system in the course of the decoherence process. Redundant spreading of information—the fact that some observables of the system can be independently read off from many distinct fragments of the environment—is investigated as the key to effective objectivity, the essential ingredient of classical reality. This focus on the environment as a communication channel through which observers learn about physical systems underscores the importance of quantum Darwinism—selective proliferation of information about “the fittest states” chosen by the dynamics of decoherence at the expense of their superpositions—as redundancy imposes the existence of preferred observables. We demonstrate that the only observables that can leave multiple imprints in the environment are the familiar pointer observables singled out by environment-induced superselection (einselection) for their predictability. Many independent observers monitoring the environment will therefore agree on properties of the system as they can only learn about preferred observables. In this operational sense, the selective spreading of information leads to appearance of an objective classical reality from within the quantum substrate.

DOI: [10.1103/PhysRevA.72.042113](https://doi.org/10.1103/PhysRevA.72.042113)

PACS number(s): 03.65.Yz, 03.65.Ta, 03.67.–a

I. INTRODUCTION

Emergence of a classical reality within the quantum Universe has been the focus of discussions on the interpretation of quantum theory ever since its inception. Measurement—the process through which we learn about the world—has the power to transform fuzzy quantum states into solid classical facts. Understanding measurements has been therefore rightly regarded as the key to unlocking the mystery of the quantum-classical transition since the early days [1]. Bohr’s interpretation proposed in 1928 [2] introduced the classical domain “by hand,” with a demand that much of the Universe—including measuring devices—must be classical. This *Copenhagen interpretation* proved to be workable and durable, but is ultimately unsatisfying, because of the arbitrary split between the quantum and the classical. Thus, Copenhagen interpretation notwithstanding, attempts to explain the emergence of the classical, objective reality (including measurement outcomes) using only quantum theory were made ever since its structure became clear in the late 1920s.

Von Neumann [3] introduced a particularly influential model of the measurement process. In his approach—and in contrast to Bohr’s view—the apparatus \mathcal{A} is also quantum. “Bit-by-bit measurement” [4] is the simplest example: a two-dimensional system \mathcal{S} in pure state $\alpha|0\rangle + \beta|1\rangle$ interacts with a two-dimensional apparatus initially in state $|\mu_0^{\mathcal{A}}\rangle$. In course of the controlled-NOT (or “measurement”) gate the apparatus becomes—as one would now say—entangled with the system \mathcal{S} :

$$(\alpha|0\rangle + \beta|1\rangle) \otimes |\mu_0^{\mathcal{A}}\rangle \rightarrow \alpha|0\rangle \otimes |\mu_0^{\mathcal{A}}\rangle + \beta|1\rangle \otimes |\mu_1^{\mathcal{A}}\rangle. \quad (1)$$

This is premeasurement. It implies correlation of \mathcal{S} and \mathcal{A} , but does not yield a definite outcome.

The structure of Eq. (1) suggests a relative state interpretation of quantum theory [5]. However, to make contact with the familiar reality, one must point out the preferred relative states. Yet, in the bipartite setting of the premeasurement, such proposals are difficult to make without some *ad hoc* assumptions (e.g., about the special role of either memory states [5] or of the Schmidt basis [6]).

The presence of entanglement in the state obtained after the premeasurement implies a *basis ambiguity*—correlation of observables of \mathcal{A} with incompatible sets of pure states of the system which cannot be resolved without some modification of the model [4]. For example, the states $\{|\mu_0^{\mathcal{A}}\rangle, |\mu_1^{\mathcal{A}}\rangle\}$ of \mathcal{A} are in one-to-one correspondence with the states $\{|0\rangle, |1\rangle\}$ of \mathcal{S} , while the states $\{(|\mu_0^{\mathcal{A}}\rangle \pm |\mu_1^{\mathcal{A}}\rangle)/\sqrt{2}\}$ are in one-to-one correspondence with the states $\{(\alpha|0\rangle \pm \beta|1\rangle)/\sqrt{2}\}$ of \mathcal{S} . Thus, von Neumann’s model does not account for the existence of a fixed “menu” of possible measurement outcomes—an issue that must be addressed before the apparent selection of one position on this menu (i.e., the collapse of the wave packet) is contemplated.

Decoherence theory (see, e.g., [7–11] for reviews) added a new element that goes well beyond the von Neumann measurement model: in addition to \mathcal{S} and \mathcal{A} , decoherence recognizes the role of the environment \mathcal{E} that surrounds \mathcal{A} and interacts with the apparatus (or with any other object immersed in \mathcal{E}). The resulting “openness” of \mathcal{A} invalidates the egalitarian principle of superposition: while all states in the Hilbert space of the apparatus \mathcal{A} are “legal” quantum super-

*Present address: School of Physical Sciences, The University of Queensland, Australia.

positions, only some of them will retain their identity—will be stable in spite of the coupling to \mathcal{E} .

Returning to our example, the environment may interact with \mathcal{A} in such a way that an arbitrary superposition $\xi_0|\mu_0^A\rangle + \xi_1|\mu_1^A\rangle$ is transformed into a mixture $|\xi_0|^2|\mu_0^A\rangle\langle\mu_0^A| + |\xi_1|^2|\mu_1^A\rangle\langle\mu_1^A|$ after a very short time. Thus, only the two states $|\mu_0^A\rangle$ and $|\mu_1^A\rangle$ remain pure over time. Selection of such preferred set of states is known as environment-induced superselection, or *einselection*. The persistence of the correlation between \mathcal{A} and \mathcal{S} is the desired prerequisite of measurements, and only the stable *pointer basis* of \mathcal{A} selected by the interaction with \mathcal{E} fits the bill [4,9–12]. Indeed, after the decoherence time, the joint state of \mathcal{S} and \mathcal{A} , Eq. (1), becomes mixed:

$$|\alpha|^2|0\rangle\langle 0| \otimes |\mu_0^A\rangle\langle\mu_0^A| + |\beta|^2|1\rangle\langle 1| \otimes |\mu_1^A\rangle\langle\mu_1^A|. \quad (2)$$

As a consequence of decoherence, only classical correlations of \mathcal{A} with the system states $\{|0\rangle, |1\rangle\}$ persist.

Understanding the reason for the loss of validity of the quantum principle of superposition is a significant step in the understanding of the quantum-classical transition, but it does not go all the way in justifying objectivity: the einselected pointer states are still ultimately quantum. Thus, they remain sensitive to direct measurements—a purely quantum problem. An observer trying to find out about the system directly will generally disturb its state, unless he happens to make a nondemolition measurement [13,14] in the pointer basis. As a consequence, it is effectively impossible for an initially ignorant observer—someone who does not know what are the pointer states of the system—to find out the state of a physical system through a direct measurement without perturbing it: immediately after a direct measurement the state will be what the observer finds out it is, but not—in general—what it was before.

The situation becomes even more worrisome when one considers *many* initially ignorant observers attempting to find out about the system. As a consequence of the disturbance caused by a direct measurement on the system, the information gained by the first observer’s measurement can get invalidated by the second observer’s measurement, etc., unless they all happen to measure commuting observables—or more precisely, unless the measured observables share the system’s pointer states as eigenstates.

Quantum subjectivity is to be contrasted with the objectivity of classical physics, where many ignorant observers can—at least in principle—find out the state of the system without modifying it. This is because classical systems admit an underlying objective description (classical reality), and classical states can be found out by initially ignorant observers. This is generally not the case for quantum systems. Thus, objective information about quantum systems can be acquired directly by many only by a constrained set of preagreed measurements on \mathcal{S} (see, e.g., [15,16]).

Of course—as noted in past discussions of einselection [7,17,18]—there are good reasons for the observers to focus on the set of states singled out by decoherence: only such pointer states of \mathcal{S} (or their dynamically evolved descendants) continue to faithfully describe the system in spite of

its interaction with \mathcal{E} . All other states are affected by \mathcal{E} , making loss of predictability inevitable. Predictability is characteristic of the states of classical systems, and is thus a symptom of a classical reality. But, above all, prediction is the reason for measurements. One can therefore understand how observers with practical experience with the emergent classicality (imposed in our Universe by einselection) will be forced to choose the same pointer observables as they make their choice of what to measure: save for pointer observables, there is no other choice if measurements are to be useful for prediction. This may look like a preagreement, but it involves no consultation between observers: competing with the environment is simply not an option.

In effect, the environment acts as a superobserver, monitoring the same pointer observable over and over, with frequency and accuracy that cannot be matched by other (e.g., human) observers. They all have to measure observables that commute with the pointer observable. Last but not least, interactions available to observers are similar in structure (e.g., depend on distance, etc.) to the interactions responsible for the einselection [10,18]. So, predictive utility in the presence of decoherence and limited choice of the Hamiltonians available in our Universe motivate preagreement by constraining measurements to pointer observables. However, even if such preagreement imposed by einselection can help single out what observables can become objective, the actual role of the environment in what happens in practice far more dramatic and decisive: the environment is not just a superobserver; it becomes a witness. Observers use it to find out about systems of interest. Hence, they must be content with the information that can be extracted from its fragments (as, generally, they will never be able to intercept all of \mathcal{E}).

In its original formulation, decoherence theory treats the information transferred to \mathcal{E} as inaccessible. However, in the real world, this is typically *not* the case. Indeed, as was pointed out by one of us some time ago [17–19], the fact that we gain most of our information by intercepting a small fraction of the photon environment is significant for the emergence of effectively classical states from the quantum substrate. The purpose of this paper is to investigate the consequences of such an indirect information acquisition for the quantum-classical transition, and to explore the relation of this “environment as a witness” [10] point of view to the predictability of the pointer states as well as to other issues raised and partially explored in [20]. We shall demonstrate that the manner in which the information is stored in the environment is the reason for the inevitable consensus among many observers about the state of the effectively classical (but ultimately quantum) systems. In other words, the structure of information deposition in \mathcal{E} is responsible for the emergence of the objective classical reality from the quantum substrate.

We shall also begin to explore *quantum Darwinism*: the dynamics responsible for the proliferation of correlations that leads to the *survival of the fittest information*. This is a natural complement to the environment as a witness approach that is focused on how the data about \mathcal{S} can be extracted by interrogating \mathcal{E} . Quantum Darwinism allows the environment to act as a witness [10,20,21], adding a new dimension to the modern decoherence-based view of the emergence of the classical.

In the next section, we propose an operational notion of objectivity and discuss how we will use it to investigate the quantum-classical transition. Sections III and IV set up the notation and introduce tools of information theory required for the present study. Section V contains the core information-theoretic analysis of the manuscript. There, we establish a number of facts about the structure of the information stored in the environment, and study consequences of *redundant* imprinting of selected system observables in \mathcal{E} . These general properties are then illustrated in Sec. VI on a dynamical model used extensively in the study of decoherence. This allows us to extend the results of our analysis, and establish a direct connection between einselection and the emergence of an objective classical reality. Finally, we consider some open questions in Sec. VII and conclude in Sec. VIII with a summary.

II. OPERATIONAL DEFINITION OF OBJECTIVITY

Quantum Darwinism [10,20] aims to show that a consensus about the properties of a quantum system—the key symptom of classical reality—arises naturally and inevitably from within quantum theory when one recognizes the role of the environment as a broadcasting medium that acquires—in the process of monitoring the system of interest that leads to decoherence and einselection—multiple copies of the information about preferred properties of the system of interest.

We will set up a rigorous operational framework for the analysis of the emergence of objective classical reality of quantum systems, based on the following definition of objectivity.

Definition 1 (objective property). A property of a physical system is *objective* when it is

- (1) simultaneously accessible to many observers,
- (2) who are able to find out what it is without prior knowledge about the system of interest, and
- (3) who can arrive at a consensus about it without prior agreement.

This operational definition of what is objective is inspired by the notion of “element of physical reality” used by Einstein, Podolsky, and Rosen (EPR) in their famous paper [22] on entanglement: “If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity. [...] Regarded not as a necessary, but merely as a sufficient condition of reality, this criterion is in agreement with classical as well as quantum-mechanical ideas of reality.” Any property of the system satisfying this requirement would be considered an element of objective reality according to the definition we adopt.

The environment as a witness approach recognizes that in our everyday acquisition of data, the decision of what to measure on S is taken out of our hands by the environment, and rests with the dynamics responsible for decoherence—i.e., for the monitoring of the system by \mathcal{E} . Nevertheless, in the absence of any structure of \mathcal{E} , this would still be insufficient to explain the emergence of a consensus about the state of S . Measurements on the environment suffer from the same

basis ambiguity problems as direct measurements on the system: they can be performed in arbitrary bases. Moreover, they generally disturb the state of the environment, and hence, the correlations between S and \mathcal{E} . Unless all observers had agreed to measure the environment in the same basis, their subsequent measurements on the environment might not yield consensus about the system, and one would not be able to attribute “objective properties” to its state.

The solution to this puzzle becomes obvious after a close inspection of how we learn about systems in the real world. Not only do the independent observers gather information about S indirectly by measuring the environment, but different observers have access to *disjoint* fragments of \mathcal{E} . By definition, when the same information about S can be discovered from different fragments, it must have been imprinted in the environment *redundantly*. In addition, when many disjoint fragments of the environment contain information about the state of the system, its properties can be found out by different observers without the danger of invalidating each other’s conclusions. This is because observables acting on disjoint fragments of \mathcal{E} always commute with each other. Hence, the first two requirements for objectivity are satisfied.

Moreover, and this is a crucial result of our study, we will demonstrate that redundant imprinting in the environment selects a preferred set of system observables. In particular, for obvious reasons [23] the environment-promoted amplification of information required to arrive at a redundant imprinting cannot amount to cloning. Amplification comes at the price of singling out commuting system observables. As a consequence, even initially ignorant observers performing arbitrary measurements on their fragments of \mathcal{E} will find out only about this unique observable. Selectivity of amplification establishes that redundant imprinting in the environment is a sufficient requirement for the emergence of classical objective reality. Thus, the state of the system is *de facto* objective when its complete and redundant imprint can be found in \mathcal{E} .

Quantum Darwinism makes novel use of information theory by focusing on the communication capacity of the environment. This approach complements the conventional system-based treatments of decoherence. There, the environment is above all an information sink, a source of decoherence responsible for irreversible loss of information [4,8,9,24]. However, these two complementary approaches do agree in their conclusions: as we will show, the pointer observables singled out by einselection are the only ones that can leave a redundant imprint on the environment. In part, this can be understood as a consequence of the ability of the pointer states to persist while immersed in the environment. Moreover, in both approaches *predictability* of the preferred states of the system (i.e., either from initial conditions, or from many independent observations on \mathcal{E}) is the key criterion. This predictability is tied to the resilience that allows the information about the pointer observables to proliferate, very much in the spirit of the survival of the fittest, and corroborates conjectures about the role of quantum Darwinism in the emergence of objectivity we have described before [10,20,21].

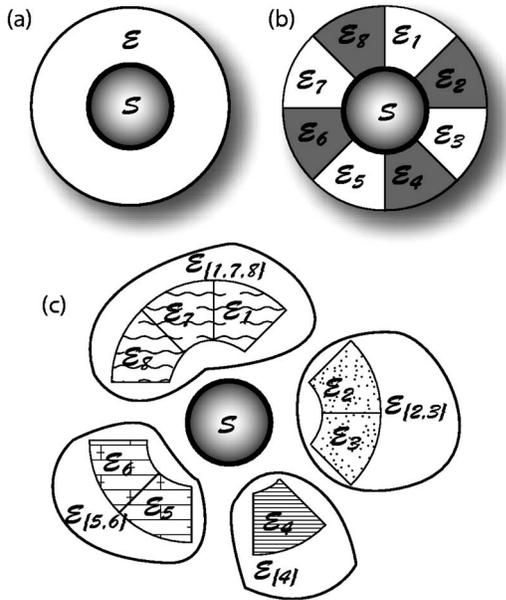


FIG. 1. The role and the structure of the environment in decoherence and in quantum Darwinism. (a) Decoherence treats the environment as a monolithic entity inaccessible to observers. The only role of \mathcal{E} is to be a sink of information about S . This is enough to understand einselection—the emergence of the preferred pointer observable in the system, with decoherence-free pointer eigenspaces. In our Universe (b) environments are typically *not* monolithic. Rather, they consist of subsystems, e.g., atoms or photons. From the point of view of decoherence, there is no difference between the paradigms illustrated by (a) and (b): The same environment will lead to the same evolution of the system (e.g., in quantum Brownian motion the same environment can be modeled either as a monolithic field or as a collection of harmonic oscillators). Quantum Darwinism (c) recognizes that observers acquire their information about S without interacting with it directly—from the imprints left by S on the fragments of \mathcal{E} . Such fragments are generally much smaller than \mathcal{E} . The central result of this paper is to show that the selected information about the system that is inscribed redundantly—in many copies—on the environment, and can be therefore found out independently from different fragments of \mathcal{E} by many observers, concerns the einselected pointer observable.

III. DEFINITIONS AND CONVENTIONS

In the setting we are considering, a system S with Hilbert space \mathcal{H}^S interacts with an environment \mathcal{E} with Hilbert space $\mathcal{H}^\mathcal{E}$. We denote the dimension of these state spaces by d^S and $d^\mathcal{E}$ respectively. Furthermore, we assume that the environment is composed of N environmental subsystems $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_N$ (see Fig. 1). That is, its Hilbert space has a *natural* tensor product structure $\mathcal{H}^\mathcal{E} = \otimes_{k=1}^N \mathcal{H}^{\mathcal{E}_k}$. This partition plays an important role in our analysis, as it suggests a natural definition of the independently accessible fragments of \mathcal{E} . We will comment on it at the end of Sec. IV.

The joint quantum state of S and \mathcal{E} is described by the density operator $\rho^{S\mathcal{E}}$ defined on $\mathcal{H}^S \otimes \mathcal{H}^\mathcal{E}$. The *reduced state* of the system is obtained by tracing out the environment $\rho^S = \text{Tr}_{\mathcal{E}}\{\rho^{S\mathcal{E}}\}$. It will often be useful to consider the joint state of the system and a *fragment* of the environment. Such

fragment—i.e., a subset of $\mathcal{E} = \{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_N\}$ —will be denoted by \mathcal{F} . The reduced state of S and \mathcal{F} is obtained by tracing out the complement of \mathcal{F} : $\rho^{S\mathcal{F}} = \text{Tr}_{\overline{\mathcal{F}}}\{\rho^{S\mathcal{E}}\}$, where $\overline{\mathcal{F}} = \mathcal{E} - \mathcal{F}$.

Following textbook quantum mechanics, we call an “observable of S ” (resp. \mathcal{E}) any Hermitian operator defined on \mathcal{H}^S (resp. $\mathcal{H}^\mathcal{E}$). By convention, we will use the first letters of the alphabet **A, B, C** to denote *system* observables while the last letters **X, Y, Z** will be reserved for *environment* observables. Hermitian operators can be written in their spectral decomposition, e.g.,

$$\mathbf{A} = \sum_j a_j A_j. \quad (3)$$

Adding to our convention, observables are denoted by bold capital letters, their eigenvalues by lowercase letters, and spectral projectors by capital letters. Only the spectral projectors are of interest to us as they completely characterize the measurement process, and the correlations between observables. We note that as we shall deal with Hermitian pointer states in many situations of interest (e.g., in underdamped harmonic oscillators) are beyond the scope of our study.

We will use the words “system” and “environment” in a very broad sense. Without loss of generality, we will suppose that \mathcal{H}^S is the part of the Hilbert space of S containing the degrees of freedom of interest. Even when the system is macroscopic, e.g., a baseball, we are typically only interested in a few of its degrees of freedom, e.g., center of mass, local densities, etc. Moreover, the degrees of freedom of S that do not couple to \mathcal{E} (directly or indirectly) play no role in our analysis. Hence d^S can remain reasonably small even for fairly large systems: d^S is really the number of relevant distinct physical configurations of S . This considerably simplifies the notation without compromising the rigor of our analysis.

Similarly, it is not necessary to incorporate all the rest of the Universe in \mathcal{E} . N really counts the number of environmental subsystems that may have been influenced by the system: only they may contain information about S . Hence, in many situations—such as a photon environment scattering off an object—the “size” of relevant \mathcal{E} can grow over time.

IV. INFORMATION

The approach to classicality outlined above is based on the existence of correlations between S and its environment that can be exploited by various observers to find out about the system. As both S and \mathcal{E} are quantum systems, quantum information theory may appear to be the right tool to study these correlations. This avenue has been considered in the past [10,19] in parallel with the approach we pursue here and in [20] and is currently also under investigation [25]. However, the emphasis here is on the observables, and information about them is easier to characterize through the relevant hypothetical measurements. As in [20], we focus on what can actually be found out about various observables of the system by monitoring observables in fragments of the environment.

The core question we ask is the following: How much does one learn about observable \mathbf{A} by measuring a different observable \mathbf{X} ? This question has an operational meaning. An observer may be considering measurement of the observable \mathbf{A} , but cannot predict its outcome with certainty. To reduce his ignorance, he can choose to measure a *different* observable \mathbf{X} . By doing so, he may decrease his uncertainty about the value of \mathbf{A} . The amount by which his uncertainty decreases is precisely the information gain we are going to study. It represents the average number of *bits* required to write down, in the most efficient way, the relevant data about \mathbf{A} acquired through the measurement of \mathbf{X} .

We will mostly be interested in the case where \mathbf{A} acts on the system and \mathbf{X} acts on a fragment of the environment—which automatically implies that $[\mathbf{A}, \mathbf{X}] = 0$. However, we will occasionally need to consider the correlations between the measurements carried successively on the same system. Hence, we present here the general case and return to the special case of commuting observables in the next section. Thus, \mathbf{A} and \mathbf{X} , with spectral projectors A_i and X_j , are arbitrary physical observables acting on an arbitrary system, in the state described by the density matrix ρ . In the following paragraph, there is no environment, just one system and two observables that may or may not commute.

The observer's uncertainty about the measurement outcome of \mathbf{A} is given by the corresponding *Shannon entropy*:

$$H(\mathbf{A}) = - \sum_i p(A_i) \ln p(A_i) \quad (4)$$

where the probability associated with the measurement outcome i —with the spectral projector A_i —is given by Born's rule $p(A_i) = \text{Tr}\{A_i \rho\}$. Entropy measures ignorance about the value of \mathbf{A} , the average number of bits missing to completely determine its value. When the measurement of observable \mathbf{X} is performed and outcome X_j is obtained, the state of the system is updated to

$$\rho \xrightarrow{X_j} \rho_{|X_j} = \frac{X_j \rho X_j}{p(X_j)} \quad (5)$$

according to the projection postulate of quantum theory [3]. This state update changes the probability assignment of the measurement outcomes of \mathbf{A} :

$$p(A_i|X_j) = \text{Tr}\{A_i \rho_{|X_j}\} = \frac{\text{Tr}\{A_i X_j \rho X_j\}}{p(X_j)}. \quad (6)$$

It is customary to call $p(A_i|X_j)$ “the conditional probability of A_i given X_j ” and similarly, $\rho_{|X_j}$ is “the conditional state of the system given X_j .”

Thus, when \mathbf{A} is measured subsequently to \mathbf{X} , the randomness of its outcome will be characterized by

$$H(\mathbf{A}|X_j) = - \sum_i p(A_i|X_j) \ln p(A_i|X_j). \quad (7)$$

The *conditional entropy* of \mathbf{A} given \mathbf{X} is the average of this quantity over the measurement outcomes of \mathbf{X} : $H(\mathbf{A}|\mathbf{X}) = \sum_j p(X_j) H(\mathbf{A}|X_j)$. The difference between the initial entropy of \mathbf{A} and its entropy posterior to the measurement of \mathbf{X} defines the *mutual information* we shall use throughout:

$$I(\mathbf{A}:\mathbf{X}) = H(\mathbf{A}) - H(\mathbf{A}|\mathbf{X}). \quad (8)$$

This is the average amount of information about \mathbf{A} obtained by measuring \mathbf{X} .

In quantum mechanics, it is possible that a certain measurement *decreases* one's ability to predict the outcome of a subsequent measurement, so mutual information is not necessarily positive. This is in fact the reason why direct measurements on the system cannot be used to arrive at a consensus about the state of the system. A direct measurement by one observer will invalidate the knowledge acquired by another when their measurements do not commute. However, this disturbance can be avoided when the measurements are carried out on different subsystems, since the observables commute automatically. The mutual information between such observables has extra properties that we shall now describe.

A. Correlations between system and environment

Let us now consider the case where \mathbf{A} acts on \mathcal{S} and \mathbf{X} on \mathcal{E} , or on a fragment \mathcal{F} of \mathcal{E} . As $[\mathbf{A}, \mathbf{X}] = 0$, the order in which the measurements are carried out does not change the joint probability distribution $p(A_i, X_j) = \text{Tr}\{A_i X_j \rho X_j\} = \text{Tr}\{X_j A_i \rho A_i\}$. It follows from Eqs. (5)–(8) that the mutual information defined above can be written in an explicitly symmetric form:

$$I(\mathbf{A}:\mathbf{X}) = \sum_{ij} p(A_i, X_j) \ln \frac{p(A_i, X_j)}{p(A_i)p(X_j)} \quad (9)$$

$$= H(\mathbf{A}) + H(\mathbf{X}) - H(\mathbf{A}, \mathbf{X}). \quad (10)$$

The amount of information about \mathbf{A} that is obtained by measuring \mathbf{X} is equal to the amount of information gained about \mathbf{X} by measuring \mathbf{A} , and is always positive. In this special case, there is a nice diagrammatic representation of the information theoretic quantities, shown on Fig. 2.

Of particular interest to us is the *maximum* amount of information about observable \mathbf{A} of \mathcal{S} that can be gained by measuring a fragment \mathcal{F} of the environment. This is defined as

$$\hat{I}_{\mathcal{F}}(\mathbf{A}) = \max_{\mathbf{X} \in \mathcal{M}_{\mathcal{F}}} I(\mathbf{A}:\mathbf{X}) \quad (11)$$

where $\mathcal{M}_{\mathcal{F}}$ is the set of measurements acting on \mathcal{F} only, i.e., the set of Hermitian operators that act trivially on $\mathcal{H}^{\bar{\mathcal{F}}}$ $= \otimes_{\mathcal{E}_k \in \bar{\mathcal{F}}} \mathcal{H}^{\mathcal{E}_k}$. The hat above is to emphasize that it is the maximum amount of information.

In particular, the maximum amount of information about the system observable \mathbf{A} that can be retrieved from the *entire* environment is denoted by $\hat{I}_{\mathcal{E}}(\mathbf{A})$. This quantity plays a crucial role in our analysis as only when $\hat{I}_{\mathcal{E}}(\mathbf{A}) \approx H(\mathbf{A})$ can we hope to find out about \mathbf{A} by probing the environment: the amount of information in the environment, $\hat{I}_{\mathcal{E}}(\mathbf{A})$, must be sufficiently large to compensate for the observer's initial ignorance, $H(\mathbf{A})$, about the value of \mathbf{A} .

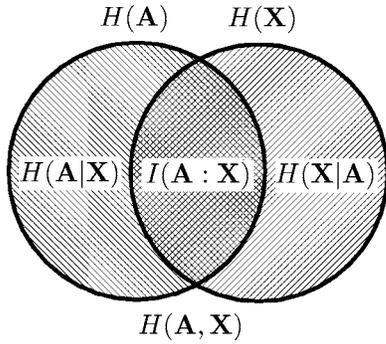


FIG. 2. Venn diagram for classical information. $H(\mathbf{A})$ is the entropy associated with a measurement of \mathbf{A} : it is the information needed to completely determine its outcome. $H(\mathbf{X})$ is the same quantity for the observable \mathbf{X} . $H(\mathbf{A}, \mathbf{X})$ is the *joint* entropy of \mathbf{A} and \mathbf{X} . $H(\mathbf{A}|\mathbf{X})$ is the average uncertainty about \mathbf{A} remaining after a measurement of \mathbf{X} . The information learned about \mathbf{A} by measuring \mathbf{X} is thus $I(\mathbf{A}:\mathbf{X})=H(\mathbf{A})-H(\mathbf{A}|\mathbf{X})$. The equivalent definition $I(\mathbf{A}:\mathbf{X})=H(\mathbf{A})+H(\mathbf{X})-H(\mathbf{A}, \mathbf{X})$ can also be understood from the diagram. These two definitions of the mutual information are equivalent [26] when probabilities can be consistently assigned to the outcomes of all the possible measurements of \mathbf{A} and \mathbf{X} both separately and jointly: this is ensured when \mathbf{A} and \mathbf{X} commute. They need not coincide otherwise.

B. Redundancy of information in the environment

When $\hat{I}_{\mathcal{E}}(\mathbf{A}) \approx H(\mathbf{A})$, the value of \mathbf{A} can be found out indirectly by probing the environment. However, as noted in Sec. II, for many observers to arrive at a consensus about the value of an observable \mathbf{A} , there must be many copies of this information in \mathcal{E} . As a consequence, independent observers will be able to perform measurements on disjoint subsets of the environment, without the risk of invalidating each other's observations.

Redundancy is therefore defined as the number of disjoint subsets of the environment containing almost all—all but a fraction δ —of the information about \mathbf{A} present in the entire environment. Formally, let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_R$ be R disjoint fragments of the environment, $\mathcal{F}_i \cap \mathcal{F}_j = \emptyset$ for $i \neq j$. Then,

$$R_{\delta}(\mathbf{A}) = \max_{\{\mathcal{F}_j\}} \{R \cdot \hat{I}_{\mathcal{F}_j}(\mathbf{A}) \geq (1 - \delta) \hat{I}_{\mathcal{E}}(\mathbf{A}), \forall j = 1, \dots, R\} \quad (12)$$

where the maximization is carried over all partitions of $\{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_N\}$ into disjoint subsets. Clearly, for any observable \mathbf{A} , $1 \leq R_{\delta}(\mathbf{A}) \leq N$. Redundancy simply counts the number of copies of the imprint of \mathbf{A} in \mathcal{E} , and hence the maximum number of observers that can independently find out about \mathbf{A} from \mathcal{E} .

C. Fragments of the environment and elementary subsystems

Before closing this section, we wish to emphasize the distinction we are making between fragments of \mathcal{E} and elementary subsystems, and comment on the role they play in using environment as a witness. The elementary subsystems of \mathcal{E} are defined through the natural tensor product structure

of $\mathcal{H}^{\mathcal{E}}$. We assume this structure to be given and fixed. In the case of a photon environment for instance, an elementary subsystem could consist of a single photon. A fragment of \mathcal{E} on the other hand is a collection of such elementary subsystems. For example, while no single photon can reveal the position of an object, a small collection of them, say 1000, may be enough to do so.

The optimization over *partitions* of the environment appearing in Eq. (12) is necessary to arrive at a proper mathematical definition of redundancy as there is no *a priori* preferred partition. This will allow us to derive very general consequences of redundancy, at the price of some technical (and perhaps also conceptual) complications. However, for the purposes of the emergence of a consensus among several observers—in essence an operational objective reality—this partition should reflect the distinct fragments of environment accessible to the different observers. While our results hold for any such partition of the environment into disjoint subsystems, Nature ultimately determines what part of \mathcal{E} is available to each observer.

The entire environment as a witness approach—and more precisely the very concept of redundancy—capitalizes on the fact that the environment has a tensor product structure $\mathcal{H}^{\mathcal{E}} = \otimes_{\mathcal{E}_k} \mathcal{H}^{\mathcal{E}_k}$. This raises the obvious question of who decides what are the elementary subsystems of \mathcal{E} . Our primary concern here is to provide a mechanism by which quantum systems can exhibit objective existence, the key symptom of the classical behavior. As we will demonstrate, this can be achieved given that the environment has a partition into subsystems, regardless of what these subsystems are. Hence, what really matters is that *various observers monitor different fragments of the environment*. As long as this requirement (which guarantees that the observables they measure commute) is satisfied, their definition of subsystems of \mathcal{E} need not coincide. Section V will present unavoidable consequences of this fact, without paying attention to the definition of the environmental subsystems. Thus, our most important conclusion—that redundancy implies selection of preferred observables—is independent of any particular choice of a tensor product in the environment.

However, different tensor product structure of the environment can *a priori* yield different redundantly imprinted observables since redundancy itself makes reference to the tensor product structure. There is no definite answer to what defines an elementary environmental subsystem, but some considerations point toward “locality” as a judicious guideline. For instance, particles are conventionally defined by the symmetries of the fundamental Hamiltonians of Nature, that are *local*. When we choose the particles of the standard model as the elementary subsystems, we are naturally led to local couplings between \mathcal{S} and \mathcal{E} . They will determine how the information is inscribed in \mathcal{E} . After all, there is no information without representation. Moreover, the information acquisition capacities of the observers are also ultimately limited by the fundamental Hamiltonians of Nature. The different observers occupy, and therefore monitor, different spatial regions. Therefore, the monitored fragments \mathcal{F}_i entering in the definition of redundancy—as well as the elementary subsystems \mathcal{E}_k composing them—should reflect these distinct spatial regions.

The fact that some division of the Universe into sub-systems is needed has been pointed out before. Indeed, the measurement problem disappears when the Universe cannot be divided into subsystems [10,11,17]. Therefore, assuming that such division exists in the discussion of the information-theoretic aspects of the origin of the classical does not seem to be a very costly assumption.

V. CONSEQUENCES OF REDUNDANCY

We now have all the necessary ingredients to study the consequences of the existence of redundant information about the system in the environment. Here, we derive several properties of the system's redundantly imprinted observables, as well as properties of the environmental observables revealing this information. While each of these results is interesting in its own right, our ultimate goal is to combine them and to show that the redundancy singles out a preferred set of commuting observables, the already familiar pointer observables. Throughout this section, we assume the existence of a *perfect* and *redundant* record of the system observables $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ in the environment, i.e., $\hat{I}_{\mathcal{E}}(\mathbf{A})=H(\mathbf{A})$ and $R_{\delta=0}(\mathbf{A}) \gg 1$, and similarly for $\mathbf{B}, \mathbf{C}, \dots$. The general case of imperfect imprints will be addressed in the next section. Let us begin by studying the consequences of the existence of a record about the value of \mathbf{A} in the environment.

Lemma 1. $\hat{I}_{\mathcal{F}}(\mathbf{A})=H(\mathbf{A})$, if and only if there exists an observable $\mathbf{X}' \in \mathcal{M}_{\mathcal{F}}$ for which $H(\mathbf{A}|\mathbf{X}')=0$ and $H(\mathbf{X}'|\mathbf{A})=0$. The measurement of \mathbf{X}' on a fragment of the environment reveals all the information of the system observable \mathbf{A} , and vice versa.

Proof. When observable \mathbf{A} is completely encoded in a fragment of the environment, $\hat{I}_{\mathcal{F}}(\mathbf{A})=H(\mathbf{A})$, there exists $\mathbf{X} \in \mathcal{M}_{\mathcal{F}}$ for which $I(\mathbf{A}:\mathbf{X})=H(\mathbf{A})$, which implies $H(\mathbf{A}|\mathbf{X})=0$. As noted in Sec. IV A, the mutual information between \mathbf{A} and \mathbf{X} is symmetric when these observables act on distinct systems, i.e., \mathcal{S} and \mathcal{F} . Therefore, measuring \mathbf{A} directly on the system provides an amount of information $I(\mathbf{A}:\mathbf{X})=H(\mathbf{A})$ about the value of \mathbf{X} , thereby decreasing its entropy to $H(\mathbf{X}|\mathbf{A})=H(\mathbf{X})-H(\mathbf{A})$. In general, this is not *all* the information about \mathbf{X} , as $H(\mathbf{X})$ may be larger than $H(\mathbf{A}):\mathbf{X}$ reveals all of the information about \mathbf{A} but the reverse is in general not true. However, by picking a suitable coarse graining \mathbf{X}' of \mathbf{X} , it is always possible to establish the duality $H(\mathbf{A}|\mathbf{X}')=0$ and $H(\mathbf{X}'|\mathbf{A})=0$.

This can be seen quite simply. The equality $H(\mathbf{A}|\mathbf{X})=0$ implies that given X_j, A_i is determined: each measurement outcome of \mathbf{X} points to a unique measurement outcome of \mathbf{A} . This defines a map $f: X_j \rightarrow A_i$. Such a map may not be one to one, so the conditional probability $p(X_j|A_i)$ of X_j given A_i is not necessarily deterministic. However, we can construct the *coarse-grained projectors*

$$X'_i = \sum_{j:f(X_j)=A_i} X_j \quad (13)$$

by regrouping the X_j in the preimage of A_i . The outcome of the associated measurements \mathbf{X}' are therefore in one-to-one

correspondence to the outcomes of \mathbf{A} , yielding the stated duality. The converse is trivial with $X=X'$. ■

An important corollary can be derived from Lemma 1 and the following observation: when the outcome of a projective measurement on a system is deterministic, the act of measuring does not modify the state of the system.

Corollary 1. Measurements of \mathbf{A} and \mathbf{X}' have exactly the same effect on the joint state of the system and the environment:

$$X'_j \rho^{SE} X'_j = A_j X'_j \rho^{SE} X'_j A_j = A_j \rho^{SE} A_j, \quad (14)$$

which implies $\rho_{\tilde{X}}^{SE} = \rho_{A_j}^{SE}$.

The contents of Lemma 1 and Corollary 1 formalize our intuitive understanding of the existence of a perfect record of the information about \mathbf{A} in \mathcal{E} : it allows perfect *emulation* of the direct measurement \mathbf{A} by the indirect measurement \mathbf{X}' . This emulation is not only perfect from the point of view of its information yield, it also has the same physical effect on the state of \mathcal{SE} . This can be regarded as an illustration of the fact that “information is physical” [27]: whenever the same information can be retrieved by two different means, the disturbance on the quantum state *can* be identical.¹ Here, the measurement of \mathbf{X} can in principle yield more information than the direct measurement \mathbf{A} itself, e.g., \mathbf{A} may be a very coarse-grained observable. This is why it is in general necessary to coarse-grain \mathbf{X} in order to obtain equivalent information, and therefore the same physical effect.

Lemma 2. When $\hat{I}_{\mathcal{F}}(\mathbf{A})=H(\mathbf{A})$, the observable \mathbf{A} commutes with the reduced density matrix of the system, $[\rho^S, \mathbf{A}]=0$.

Proof. Following Lemma 1 and Corollary 1, there exists $\mathbf{X}' \in \mathcal{E}$ for which $X'_j \rho^{SE} X'_j = A_j \rho^{SE} A_j$. The reduced state of the system can be therefore written as

$$\rho^S = \text{Tr}_{\mathcal{E}}\{\rho^{SE}\} \quad (15)$$

$$= \sum_j \text{Tr}_{\mathcal{E}}\{X'_j \rho^{SE} X'_j\} \quad (16)$$

$$= \sum_j \text{Tr}_{\mathcal{E}}\{A_j \rho^{SE} A_j\} = \sum_j A_j \rho^S A_j, \quad (17)$$

so ρ^S is block diagonal in the eigensubspaces of \mathbf{A} . ■

The relation between decoherence and the existence of an environmental record has been pointed out and investigated in the past [4,12,17–19,29–31]. The above lemma confirms that whenever the environment acquires a single copy of the information about \mathbf{A} , the state of \mathcal{S} decoheres with respect to the spectral projectors A_i of \mathbf{A} . This decoherence-induced diagonalization of the reduced state has additional important implications when the information about the system is encoded redundantly in the environment.

Corollary 2. Let \mathbf{A} be a system observable redundantly

¹In the theory of generalized measurements [28], one can in principle make a measurement that yields no information whatsoever, yet greatly disturb the system; hence the emphasized “can” in the above sentence.

imprinted in \mathcal{E} , and let \mathbf{X}' be the coarse-grained observable of the fragment \mathcal{F} of the environment that contains all the information about \mathbf{A} (as constructed in Lemma 1). Then the reduced state $\rho^{\mathcal{S}\mathcal{F}}$ commutes with both \mathbf{A} and \mathbf{X}' : the correlations between these observables are classical.

Proof. By assumption \mathbf{A} is redundantly imprinted in \mathcal{E} . Thus, its value can be inferred by independent measurements, say \mathbf{X} and \mathbf{Y} , acting on two distinct fragments of the environment, \mathcal{F} and $\bar{\mathcal{F}}=\mathcal{E}-\mathcal{F}$, respectively. Following Lemma 1, we can construct a coarse-grained observable \mathbf{X}' which retains essential correlations to \mathbf{A} and with the property that $H(\mathbf{X}'|\mathbf{A})=0$ —a measurement of \mathbf{A} on \mathcal{S} reveals all the information about \mathbf{X}' . As a consequence, the measurement \mathbf{Y} on $\bar{\mathcal{F}}$ that reveals all the information about \mathbf{A} must also reveal all the information about \mathbf{X}' . Hence, Corollary 2 follows from Lemma 2.

We will now characterize the class of system observables that can be redundantly imprinted in the environment. When \mathbf{A} and \mathbf{B} are redundantly imprinted in \mathcal{E} and redundancy is sufficiently high, their value can be inferred from disjoint fragments of the environment. This implies the following relation between any redundantly imprinted observables.

Lemma 3. Let \mathcal{F} be a fragment of the environment containing a copy of the information about \mathbf{A} . When the rest of the environment $\bar{\mathcal{F}}$ contains a copy of the information about \mathbf{B} , then \mathbf{A} and \mathbf{B} necessarily commute on the support of $\rho^{\mathcal{S}}$.

Proof. The assumptions of Lemma 3 imply the existence of two environmental observables \mathbf{X} and \mathbf{Y} acting on disjoint fragments of the environment \mathcal{F} and $\bar{\mathcal{F}}$ for which $H(\mathbf{A}|\mathbf{X})=0$ and $H(\mathbf{B}|\mathbf{Y})=0$. Following the procedure of Lemma 1, we can coarse-grain these two observables into \mathbf{X}' and \mathbf{Y}' that are in one-to-one correspondence with \mathbf{A} and \mathbf{B} , respectively. The observables \mathbf{X}' and \mathbf{Y}' obviously commute with each other as they act on disjoint fragments of \mathcal{E} , and also commute with the system observables \mathbf{A} and \mathbf{B} . Then, by Corollary 1, we have

$$\begin{aligned} A_i B_j \rho^{\mathcal{S}\mathcal{E}} B_j A_i &= A_i Y'_j \rho^{\mathcal{S}\mathcal{E}} Y'_j A_i \\ &= Y'_j A_i \rho^{\mathcal{S}\mathcal{E}} A_i Y'_j \\ &= Y'_j X'_i \rho^{\mathcal{S}\mathcal{E}} X'_i Y'_j \\ &= X'_i Y'_j \rho^{\mathcal{S}\mathcal{E}} Y'_j X'_i \\ &= X'_i B_j \rho^{\mathcal{S}\mathcal{E}} B_j X'_i \\ &= B_j X'_i \rho^{\mathcal{S}\mathcal{E}} X'_i B_j \\ &= B_j A_i \rho^{\mathcal{S}\mathcal{E}} A_i B_j, \end{aligned}$$

proving the lemma. \blacksquare

The above lemma states that it is impossible to acquire perfect information about two noncommuting observables by measuring two disjoint fragments of \mathcal{E} simultaneously. This is reminiscent of the Heisenberg indeterminacy principle. However, in spite of this similarity, these two results differ in the precise setting in which they hold. Heisenberg principle asserts that it is impossible to know simultaneously the values of two noncommuting observables of an otherwise isolated system. In our lemma, the system is already correlated

with its environment, and the observer is trying to find out about the value of two observables using these correlations. In other words, Heisenberg indeterminacy concerns information about the system before it interacted with \mathcal{E} , while Lemma 3 focuses on the information about \mathcal{S} that is present in \mathcal{E} after their interaction.

This difference illustrates the nature of our operational approach: we are describing the physical properties of an open quantum system, hence we focus on information about its present state, not on what it was prior to the interaction with \mathcal{E} . Decoherence happens, so we are dealing with it.²

We can derive a similar result for the commutation of environmental observables.

Lemma 4. Let \mathcal{F} be a fragment of the environment containing information about both \mathbf{A} and \mathbf{B} , two commuting system observables. Then, there exist two observables \mathbf{X}' and $\mathbf{Y}' \in \mathcal{M}_{\mathcal{F}}$ that commute on the support of $\rho^{\mathcal{S}\mathcal{E}}$ and reveal all the information about \mathbf{A} and \mathbf{B} , respectively.

Proof. The assumptions of Lemma 4 imply the existence of \mathbf{X} and $\mathbf{Y} \in \mathcal{M}_{\mathcal{F}}$ such that $I(\mathbf{A}:\mathbf{X})=H(\mathbf{A})$ and $I(\mathbf{B}:\mathbf{Y})=H(\mathbf{B})$. Following the procedure of Lemma 1, \mathbf{X} and \mathbf{Y} can be coarse grained to \mathbf{X}' and \mathbf{Y}' while retaining all the essential correlations to \mathbf{A} and \mathbf{B} . Since \mathbf{A} and \mathbf{B} commute, the proof of Lemma 3 can be applied here, yielding the desired result. \blacksquare

Given these properties of redundantly imprinted observables, we can now state a very important result, which essentially ensures the uniqueness of redundantly imprinted observables.

Theorem 1. Let \mathbf{B} and \mathbf{C} be two system observables redundantly imprinted in the environment, and let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_R$ be $R \geq 2$ disjoint fragments of \mathcal{E} , each containing a copy of the information about \mathbf{B} and \mathbf{C} . Then, there exists a refined system observable \mathbf{A} satisfying

- (1) $\hat{I}_{\mathcal{E}}(\mathbf{A})=H(\mathbf{A})$,
- (2) $R_0(\mathbf{A}) \geq R$,
- (3) $H(\mathbf{B}|\mathbf{A})=0$ and $H(\mathbf{C}|\mathbf{A})=0$.

Proof. The assumptions of the theorem imply the existence of $\mathbf{Y}^{\mathcal{F}_n}$ and $\mathbf{Z}^{\mathcal{F}_n} \in \mathcal{M}_{\mathcal{F}_n}$ for which $H(\mathbf{B}|\mathbf{Y}^{\mathcal{F}_n})=0$ and $H(\mathbf{C}|\mathbf{Z}^{\mathcal{F}_n})=0$ for all $n=1, \dots, R$. By Lemma 3, \mathbf{B} and \mathbf{C} must

²To fully grasp the distinction between these two settings requires a somewhat technical discussion. Assume that the system and environment are initially in the uncorrelated state $\rho^{\mathcal{S}\mathcal{E}}(0)=\rho^{\mathcal{S}}(0) \otimes \rho^{\mathcal{E}}(0)$. They interact for a time t , yielding the joint state $\rho^{\mathcal{S}\mathcal{E}}(t)=U^{\mathcal{S}\mathcal{E}}\rho^{\mathcal{S}\mathcal{E}}(0)U^{\mathcal{S}\mathcal{E}\dagger}$ where $U^{\mathcal{S}\mathcal{E}}=\exp\{-i\mathbf{H}^{\mathcal{S}\mathcal{E}}t\}$. When a measurement \mathbf{X} is carried on the environment and outcome j is observed, the conditional state of the system will be

$$\rho_{X_j}^{\mathcal{S}}(t) = \frac{\text{Tr}_{\mathcal{E}}\{X_j \rho^{\mathcal{S}\mathcal{E}}(t) X_j^\dagger\}}{p(X_j)}.$$

According to the generalized theory of measurement [28], this can be described as a positive operator valued measure (POVM) acting on the initial state of the system, $\rho^{\mathcal{S}}(0)$. Formally, there exists a set of operators A_{ij} acting on $\mathcal{H}^{\mathcal{S}}$ such that $\rho_{X_j}^{\mathcal{S}}(t)=\sum_i A_{ij} \rho^{\mathcal{S}}(0) A_{ij}^\dagger$. However, it does not correspond to any kind of measurement acting on $\rho^{\mathcal{S}}(t)$, the state of the system after it has interacted with \mathcal{E} (see, for example, [32] Eq. (100)). Our lemma applies also to this type of information gathering processes that cannot be described within the POVM formalism.

commute as they can be inferred from disjoint fragments of the environment. This fact, together with Lemma 4, implies that there exist *commuting* coarse-grained observables $\mathbf{Y}^{\mathcal{F}_n}$ and $\mathbf{Z}^{\mathcal{F}_n} \in \mathcal{M}_{\mathcal{F}_n}$ that reveal all of the information about \mathbf{B} and \mathbf{C} , respectively. The observables \mathbf{B} and \mathbf{C} can be merged into a more refined observable \mathbf{A} with spectral projectors given by the product of the spectral projectors of \mathbf{B} and \mathbf{C} :

$$A_{ij} = B_i C_j. \quad (18)$$

As B_i and C_j commute, the operators A_{ij} form a complete set of mutually orthogonal projectors. Similarly, the environment observables $\mathbf{Y}^{\mathcal{F}_n}$ and $\mathbf{Z}^{\mathcal{F}_n}$ can be merged into a more refined observable $\mathbf{X}^{\mathcal{F}_n}$ with spectral projectors $X_{ij}^{\mathcal{F}_n} = Y_i^{\mathcal{F}_n} Z_j^{\mathcal{F}_n}$. A measurement of $\mathbf{X}^{\mathcal{F}_n}$ on \mathcal{F}_n reveals all the information about \mathbf{A} . Indeed, following Corollary 1,

$$\begin{aligned} p(A_{i'j'} | X_{ij}^{\mathcal{F}_n}) &= \frac{\text{Tr}\{A_{i'j'} X_{ij}^{\mathcal{F}_n} \rho^{\mathcal{SE}} X_{ij}^{\mathcal{F}_n} A_{i'j'}\}}{\text{Tr}\{X_{ij}^{\mathcal{F}_n} \rho^{\mathcal{SE}} X_{ij}^{\mathcal{F}_n}\}} \\ &= \frac{\text{Tr}\{B_{i'} C_{j'} Y_i^{\mathcal{F}_n} Z_j^{\mathcal{F}_n} \rho^{\mathcal{SE}} Z_j^{\mathcal{F}_n} Y_i^{\mathcal{F}_n} C_{j'} B_{i'}\}}{\text{Tr}\{Y_i^{\mathcal{F}_n} Z_j^{\mathcal{F}_n} \rho^{\mathcal{SE}} Z_j^{\mathcal{F}_n} Y_i^{\mathcal{F}_n}\}} \\ &= \delta_{ii'} \delta_{jj'} \end{aligned}$$

which implies $H(\mathbf{A} | \mathbf{X}^{\mathcal{F}_n}) = 0$. This automatically implies $\hat{I}_{\mathcal{F}_n}(\mathbf{A}) = H(\mathbf{A})$ and, since the above holds for all $n=1, \dots, R$, we have $R_0(\mathbf{A}) \geq R$. Finally, note that $B_i = \sum_j A_{ij}$ and $C_j = \sum_i A_{ij}$; \mathbf{B} and \mathbf{C} are obtained by coarse-graining \mathbf{A} . Therefore, a measurement of \mathbf{A} reveals all the information about \mathbf{B} and \mathbf{C} , hence completing the proof. ■

The meaning of this theorem is that whenever more than one observable can be redundantly inferred from a fixed set of disjoint fragments of \mathcal{E} , they necessarily correspond to some coarse-grained version of a *maximally refined redundantly imprinted observable* \mathbf{A} . The *modus tollens* of this theorem is also very enlightening. Given a decomposition of \mathcal{E} into fragments and the associated maximally refined redundantly imprinted observable \mathbf{A} , the only observables \mathbf{B} that can be completely and redundantly inferred from these fragments of \mathcal{E} are those for which $I(\mathbf{B} : \mathbf{A}) = H(\mathbf{B})$. The value of \mathbf{B} must be entirely determined by a measurement of the maximally refined \mathbf{A} . This is obviously a sufficient condition: if a direct measurement of \mathbf{A} reveals all the information about \mathbf{B} and \mathbf{A} is redundantly imprinted in \mathcal{E} , then so is \mathbf{B} . Theorem 1 shows that this requirement is also necessary.

Note that the proof relies only on a redundancy greater than 2 and on a fixed decomposition of \mathcal{E} into disjoint fragments from which \mathbf{B} and \mathbf{C} are inferred. Relaxing this last assumption raises the possibility that different decompositions lead to incompatible maximally refined redundant observables. Figure 3 represents two decompositions of \mathcal{E} into fragments for which the corresponding maximally refined redundantly imprinted observables could not be compared through the above theorem. However, this possibility is ruled out when redundancy is high.

Theorem 2. Let \mathbf{B} and \mathbf{C} be two system observables both highly redundantly encoded in \mathcal{E} , such that $R_0(\mathbf{B}) \times R_0(\mathbf{C}) > N$. Then, $[\mathbf{B}, \mathbf{C}] = 0$ on the support of $\rho^{\mathcal{S}}$.

Proof. By the assumptions of the theorem, there exists a

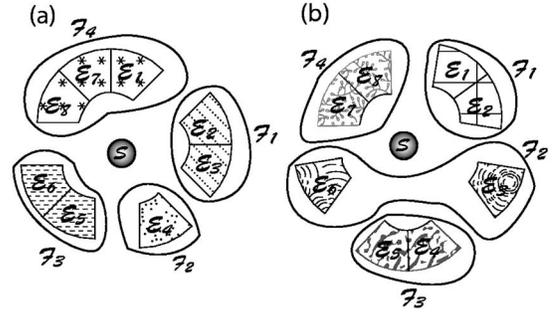


FIG. 3. Suppose \mathbf{B} and \mathbf{C} are redundantly imprinted in the environment, but are associated, respectively, with decompositions (a) and (b), represented by different arrangements of the subsystems of \mathcal{E} into fragments. Then, from Theorem 1 one cannot yet conclude that there exists a maximally refined observable \mathbf{A} from which both \mathbf{B} and \mathbf{C} can be inferred. This is because the two decompositions do not match. However, when the redundancy of \mathbf{B} and \mathbf{C} satisfies $R_0(\mathbf{B})R_0(\mathbf{C}) > N$, then Theorem 2 proves that the maximally refined observables obtained for each decomposition of \mathcal{E} into fragments must commute (i.e., they are compatible).

fragment \mathcal{F} of size at most $N/R_0(\mathbf{B})$ containing all information about \mathbf{B} . Since $R_0(\mathbf{C}) > N/R_0(\mathbf{B})$, there exists a fragment \mathcal{F}' disjoint from \mathcal{F} , i.e., $\mathcal{F} \cap \mathcal{F}' = \emptyset$, that contains all information about \mathbf{C} . By Lemma 3 we have $[\mathbf{B}, \mathbf{C}] = 0$. ■

It is intriguing to note that—in view of the above discussion—large redundancy implies $R > \sqrt{N}$. Theorems 1 and 2 prove our claim of Sec. II. Redundant spreading of information comes at the price of singling out a preferred set of commuting observables: those obtained by coarse-graining maximally refined redundantly observables.

So far, we have not considered any dynamical evolution of the system and environment: we have focused on the correlations present at a fixed time without paying attention to how they arise. We studied consequences of regarding environment as a witness (and interrogating it about the state of the system), but we haven't yet enquired about the dynamics that allowed the environment to acquire this information in the first place. Ultimately, it is the coupling Hamiltonian between \mathcal{S} and \mathcal{E} which establishes these correlations. The coupling is also responsible for the emergence of preferred system observables—the pointer observables—that are least affected by the openness of the system. Thus, as in the case of einselection of the pointer states, the connection between the emergence of an objective reality and the selection of preferred observables has to be ultimately traced to the dynamics responsible for the “environmental monitoring” of \mathcal{S} by \mathcal{E} —i.e., through dynamical considerations. When the correlations between maximally refined observables and fragments of the environment *persist in time*, Lemma 2 and Theorems 1 and 2 imply the following conclusion.

Corollary 3. Highly redundantly imprinted observables are the system's pointer observables.

In other words, under the idealized assumption of perfect records we have demonstrated that only the already familiar pointer observables can leave a redundant and robust imprint on \mathcal{E} . Corollary 3 can be understood as a consequence of the ability of the pointer states to persist while immersed in the

environment. This resilience allows the information about the pointer observables to proliferate, very much in the spirit of the Darwinian survival of the fittest. This, along with other facts, will be illustrated on a dynamical model in the next section.

VI. QUANTUM DARWINISM: DYNAMICAL EMERGENCE OF OBJECTIVITY

It should by now be clear that the presence of redundant information in the environment imposes severe physical constraints on the state of the system. Surprisingly, the derivation of these important facts required little physics—most of them followed from information-theoretic arguments, once again illustrating that information is physical. One could sum up our conclusions by noting that the environment can be a perfect witness—with multiple perfect records—but only for one set of commuting observables defined by the perfect maximally refined pointer observables. Such perfection will generally be only approximated in the real world. It is therefore important to investigate the situation where multiple, yet imperfect, copies arise through the dynamics. To this end we shall consider selective proliferation of information in a simple model that exhibits quantum Darwinism.

By focusing on the *dynamics* of \mathcal{SE} rather than on their *state* we will be able to extend our analysis. In particular, we will show that when the imprints of the objective observable are nearly perfect—i.e., $\hat{I}_{\mathcal{E}}(\mathbf{A}) \gg (1 - \vartheta)H(\mathbf{A})$ and $R_{\delta}(\mathbf{A}) > 1$ for $\vartheta \ll 1$ and $\delta \ll 1$ finite but small—the conclusions obtained earlier still hold. Above all, along the lines of Theorem 1, there exists a unique maximally refined observable whose information is the only one available in fragments of \mathcal{E} . Finally, we will be able to specify the optimal measurement to be performed on fragments of \mathcal{E} to learn about the system. This is a considerable improvement over the result of the last section where, in the absence of any model Hamiltonian, we could only demonstrate the existence of such a measurement.

A. Dynamics of quantum Darwinism and decoherence

The model we consider is a generalization of the simple early model of decoherence put forward in [12] and thoroughly investigated as a tractable, yet nontrivial, paradigm for decoherence: despite its simplicity, it captures the essence of einselection. Hence, our simple model serves as a special case that sets a conceptual framework for the study of more sophisticated models, such as a photon environment scattering on an object and carrying away potential visual data. As any specific model, it requires some fairly specific assumptions. We shall consider their role in the next section and discuss the extent to which relaxing some of these assumptions affects our conclusions.

A system \mathcal{S} is coupled to an environment $\mathcal{E} = \{\mathcal{E}_1, \dots, \mathcal{E}_N\}$ through the Hamiltonian:

$$\mathbf{H}^{\mathcal{SE}} = \sum_{\mathcal{E}_k \in \mathcal{E}} \mathbf{A} \otimes \mathbf{Z}^{\mathcal{E}_k}, \quad (19)$$

where \mathbf{A} and $\mathbf{Z}^{\mathcal{E}_k}$ are operators acting on \mathcal{S} and \mathcal{E}_k , respectively. The joint initial state of the system and the different

environmental subsystems is assumed to be a pure product state:

$$|\Phi^{\mathcal{SE}}(0)\rangle = |\phi^{\mathcal{S}}\rangle \otimes |\phi^{\mathcal{E}_1}\rangle \otimes |\phi^{\mathcal{E}_2}\rangle \otimes \dots \quad (20)$$

Hence, before the interaction, there is no correlation between the system and the environment, nor among the environmental subsystems.

A convenient way of writing $|\phi^{\mathcal{S}}\rangle$ in view of expressing the time evolution of the joint state of \mathcal{SE} is to decompose it in an eigenbasis $|i\rangle$ of \mathbf{A} ($\mathbf{A}|i\rangle = a_i|i\rangle$):

$$|\phi^{\mathcal{S}}\rangle = \sum_i \alpha_i |i\rangle. \quad (21)$$

Without loss of generality, we can assume that the vectors $|i\rangle$ with nonzero coefficients α_i in Eq. (21) are associated with distinct eigenvalues a_i of \mathbf{A} . This can be done by choosing appropriate bases for the degenerate eigenspaces of \mathbf{A} . Thus, after an interaction time t with the N subsystems of the environment, the joint state of \mathcal{SE} evolves into

$$|\Phi^{\mathcal{SE}}(t)\rangle = \sum_j \alpha_j |j\rangle \bigotimes_{\mathcal{E}_k \in \mathcal{E}} e^{-ita_j \mathbf{Z}^{\mathcal{E}_k}} |\phi^{\mathcal{E}_k}\rangle \quad (22)$$

$$= \sum_j \alpha_j |j\rangle \bigotimes_{\mathcal{E}_k \in \mathcal{E}} |\phi_j^{\mathcal{E}_k}\rangle, \quad (23)$$

where we have defined $|\phi_j^{\mathcal{E}_k}\rangle = e^{-ita_j \mathbf{Z}^{\mathcal{E}_k}} |\phi^{\mathcal{E}_k}\rangle$.

In the following sections, we will analyze the correlations between the system \mathcal{S} and the “observed” part of the environment \mathcal{F} . For example, in the case of a photon environment scattering on an object, \mathcal{F} is the set of photons that hit the observer’s retina, while $\bar{\mathcal{F}}$ represents those photons that have scattered on \mathcal{S} but that are not intercepted by the observer. Equation (23) allows us to compute the state of \mathcal{S} and \mathcal{F} :

$$\rho^{\mathcal{SF}} = \sum_{ij} \alpha_i \alpha_j^* |i\rangle\langle j| \otimes |\Phi_i^{\mathcal{F}}\rangle\langle \Phi_j^{\mathcal{F}}| \times \gamma_{ij}^{\bar{\mathcal{F}}}, \quad (24)$$

where

$$|\Phi_j^{\mathcal{F}}\rangle = \bigotimes_{\mathcal{E}_k \in \mathcal{F}} e^{-ita_j \mathbf{Z}^{\mathcal{E}_k}} |\phi^{\mathcal{E}_k}\rangle = \bigotimes_{\mathcal{E}_k \in \mathcal{F}} |\phi_j^{\mathcal{E}_k}\rangle, \quad (25)$$

and $\gamma_{ij}^{\bar{\mathcal{F}}} = \prod_{\mathcal{E}_k \in \bar{\mathcal{F}}} \langle \phi_j^{\mathcal{E}_k} | \phi_i^{\mathcal{E}_k} \rangle = \langle \Phi_j^{\bar{\mathcal{F}}} | \Phi_i^{\bar{\mathcal{F}}} \rangle$. Similarly, the reduced state of the system at time t reads

$$\rho^{\mathcal{S}} = \sum_{ij} \alpha_i \alpha_j^* |i\rangle\langle j| \times \gamma_{ij}^{\mathcal{E}}, \quad (26)$$

where $\gamma_{ij}^{\mathcal{E}} = \prod_{\mathcal{E}_k \in \mathcal{E}} \langle \phi_j^{\mathcal{E}_k} | e^{-it(a_i - a_j) \mathbf{Z}^{\mathcal{E}_k}} | \phi_i^{\mathcal{E}_k} \rangle$. Above, the γ ’s are called the *decoherence factors*.

Except in carefully controlled experiments—where the system can be very well isolated from its environment—the number of environmental subsystems interacting with \mathcal{S} is huge. In this case, the $\gamma_{ij}^{\mathcal{E}}$ ’s will be typically very small for $i \neq j$. Each environmental subsystem contributes a factor $\langle \phi_j^{\mathcal{E}_k} | e^{-it(a_i - a_j) \mathbf{Z}^{\mathcal{E}_k}} | \phi_i^{\mathcal{E}_k} \rangle$ to $\gamma_{ij}^{\mathcal{E}}$, so unless $|\phi^{\mathcal{E}_k}\rangle$ is an eigenstate of $\mathbf{Z}^{\mathcal{E}_k}$ —in which case \mathcal{E}_k is effectively decoupled from the system—this strictly decreases the decoherence factor.

Hence, $\gamma_{ij}^{\mathcal{E}}$ goes to δ_{ij} as N increases. To be more specific, when the initial states of the environmental subsystems are distributed uniformly at random, the decrease of $\gamma_{ij}^{\mathcal{E}}$ is exponential with N and typically Gaussian with time [33].

This reasoning also applies to the decoherence factors $\gamma_{ij}^{\bar{\mathcal{F}}}$: they tend to zero for $i \neq j$ as the number of *unobserved* environmental subsystems increases. Even though a considerable number of the environmental subsystems can be intercepted by the observer, an even larger fraction will usually escape his monitoring. Thus, the $\gamma_{ij}^{\bar{\mathcal{F}}}$'s will also typically be very small for $i \neq j$.

We have numerically studied a version of this model (the case of two dimensional system and environmental subsystems) in [20]. Our results indicate that for sufficiently large environment, several system observables can become nearly perfectly imprinted in the environment (see Fig. 1(a) of [20]). However, the results also clearly show that only those observables very “close” to the pointer observable can leave a *redundant* imprint in the environment (see Fig. 1(b) of [20]). The information about the other system observables can only be accessed by measuring the *entire* environment: their value cannot be learned from small fragments of \mathcal{E} . In what follows, we will derive these facts analytically.

B. Decoherence: The focus on the system

In order to contrast the environment as a witness approach to classicality as well as to illustrate significance of the dynamics that leads to quantum Darwinism, we will now review some results that were obtained using more conventional approaches. Thus, we will abandon for a moment the study of correlations between \mathcal{S} and parts of the environment to focus uniquely on the state of the system, as is done in the the standard studies of decoherence [8,9].

We see from Eq. (26) that the off-diagonal terms of the system's density matrix tend to be very small when expressed in the $|i\rangle$ basis. It is therefore natural to expect that these are the quasi-classical states of the system: the coupling to the environment suppresses quantum superpositions of the states $|i\rangle$. Of course, the exact instantaneous diagonal basis of $\rho^{\mathcal{S}}$ might differ from $|i\rangle$, especially when the coefficients α_j of the system's initial state Eq. (21) are almost equal. Hence, for someone focused on the instantaneous state of the system, it may not be clear why the basis $|i\rangle$ deserves any special attention in spite of the existence of small—yet nonzero—off-diagonal elements for $\rho^{\mathcal{S}}$. However, a simple analysis shows that the basis $|i\rangle$ is the only basis for which all off-diagonal terms tend to zero independently of the initial state of the system, and, thus, retain correlations with the rest of the Universe. This ability to retain correlations is behind the predictability of the pointer states—it is in fact their defining feature (see [4,12] for the initial formulation, and [10,11,34] for a recent re-assessment).

Persistence of correlations implies continued existence of the states [10,18]. Independence of the pointer states from the initial state of the system is obviously a very important property of the to-be-classical states: the quasiclassical domain must be independent of the initial state of the system—e.g., the set of the pointer states of the apparatus should not

depend on the state of the measured quantum system. When the system is initially prepared in one of the states $|i\rangle$, it will not be affected by the interaction with the environment: pointer states are stable. In more general cases (more complicated interactions, etc.), the above simple analysis cannot be carried out, and one usually relies on the predictability sieve [8,9,17,24] to find the pointer basis. Predictability sieve seeks most predictable initial states—states that produce the smallest amount of entropy over time while subject to interaction with the environment.

C. Perfect correlations

We now return to the study of the correlations between \mathcal{S} and fragments of the environment. Our goal is to characterize what kind and how information is stored in the environment when it is redundant. To be more precise, we will analyze the structure of information in a fragment \mathcal{F} of \mathcal{E} under the assumption that the rest of the environment, i.e., $\bar{\mathcal{F}} = \mathcal{E} - \mathcal{F}$, contains at least one additional copy of this information. Our demand that both \mathcal{F} and $\bar{\mathcal{F}}$ contain a copy of the information ensures a minimum redundancy of $R \geq 2$ which—although insufficient to ensure the emergence of a consensus about the properties of \mathcal{S} among many observers—is enough to rederive most of the results established in Sec. V.³ In addition, we will be able to specify optimal measurement strategies on \mathcal{F} , the fragment of \mathcal{E} accessible to the observer, for inferring any kind of information about \mathcal{S} .

To build our intuition, it is enlightening to first consider the limiting case $\gamma_{ij}^{\bar{\mathcal{F}}} = \gamma_{ij}^{\mathcal{F}} = \delta_{ij}$, where $\gamma_{ij}^{\mathcal{F}} = \langle \Phi_i^{\mathcal{F}} | \Phi_j^{\mathcal{F}} \rangle$. We return to the general case in Sec. VI D. Following the general argument of Sec. VI A, perfect correlation will be recovered when the environment is infinite. In our analysis, this condition reflects the physical assumption that both the observed and the unobserved fragments of \mathcal{E} are very large. It is also possible to obtain perfect correlations in much smaller systems when the dynamics is fine-tuned so that the resulting state is Greenberger-Horne-Zeilinger-like. In either case, the density matrix of \mathcal{S} and \mathcal{F} reads

$$\sigma^{\mathcal{S}\mathcal{F}} = \sum_i |\alpha_i|^2 |i\rangle\langle i| \otimes |\Psi_i^{\mathcal{F}}\rangle\langle\Psi_i^{\mathcal{F}}| \quad (27)$$

where the $|\Psi_i^{\mathcal{F}}\rangle$'s are mutually orthogonal. Note that ρ and the $|\Phi_i^{\mathcal{F}}\rangle$'s of Eqs. (24) and (25) have been replaced respectively by σ and $|\Psi_i^{\mathcal{F}}\rangle$'s to emphasize $\gamma_{ij}^{\mathcal{F}} = \delta_{ij}$.

Consider the observable $\mathbf{X} \in \mathcal{M}_{\mathcal{F}}$ that perfectly distinguishes between the orthogonal states $|\Psi_i^{\mathcal{F}}\rangle$:

$$\mathbf{X} = \sum_i x_i |\Psi_i^{\mathcal{F}}\rangle\langle\Psi_i^{\mathcal{F}}| = \sum_i x_i X_i. \quad (28)$$

Following Eq. (5), the state of the system after a measurement of \mathbf{X} on \mathcal{F} with outcome X_j is

³In this toy model, all the environmental subsystems couple to the system in the same way, so only the size of the fragments \mathcal{F} and $\bar{\mathcal{F}}$ matters. Thus, complications illustrated in Fig. 3 and analyzed in Theorem 2 do not occur, and redundancy $R \geq 2$ is indeed sufficient to establish uniqueness of the preferred observable.

$$\sigma_{|X_j}^S = \text{Tr}_{\mathcal{F}}\{\sigma_{|X_j}^{S\mathcal{F}}\} = \frac{\text{Tr}_{\mathcal{F}}\{X_j \sigma^{S\mathcal{F}} X_j\}}{p(X_j)} = |j\rangle\langle j|. \quad (29)$$

This is the same as the state of \mathcal{S} after a *direct* measurement of \mathbf{A} with outcome A_j , $\sigma_{|X_j}^S = \sigma_{|A_j}^S$. Thus, a subsequent measurement of \mathbf{A} on the system yields outcome A_i with certainty, so $H(\mathbf{A}|\mathbf{X})=0$. The same conclusion can be reached for a measurement acting on $\bar{\mathcal{F}}$: thus, \mathbf{A} is encoded redundantly in \mathcal{E} (there are at least two copies). As a consequence, all the results of Sec. V (with the exception of Theorem 2) hold. It is however instructive to derive them directly for our specific model, without appealing to the general lemmas.

As seen from Eq. (29), the indirect measurement \mathbf{X} perfectly emulates the direct measurement \mathbf{A} on the system. Consequently, all the information about the observable \mathbf{A} can be extracted by the measurement \mathbf{X} on the environment. In addition, we clearly see the no information–no disturbance principle at work: once the outcome of \mathbf{X} is known, measuring \mathbf{A} directly does not disturb the state of the system any further, as the measurement of \mathbf{X} “projects” the system in an eigenstate of \mathbf{A} (see Corollary 1). Averaging over the measurement outcomes of \mathbf{X} yields the reduced density matrix of \mathcal{S} , which is a mixture of the eigenstates of \mathbf{A} , so $[\rho^S, \mathbf{A}]=0$ as specified by Lemma 2.

Specifying the Hamiltonian responsible for the correlations between \mathcal{S} and its environment allows one to go beyond the purely information-theoretic analysis of Sec. V. Given the explicit form of the density matrix $\sigma^{S\mathcal{F}}$, we can address more interesting questions such as: what is the optimal measurement \mathbf{Y} on \mathcal{F} that yields the largest amount of information about a given system observable \mathbf{B} ? In fact, it turns out that \mathbf{X} is the optimal measurement on \mathcal{F} to find out the value of any system observable \mathbf{B} . Regardless of what we wish to learn about the system, the optimal strategy consists in measuring \mathbf{X} on \mathcal{F} :

$$I(\mathbf{B}:\mathbf{X}) \geq I(\mathbf{B}:\mathbf{Y}) \quad (30)$$

for all $\mathbf{B} \in \mathcal{M}_{\mathcal{S}}$ and $\mathbf{Y} \in \mathcal{M}_{\mathcal{F}}$. Similarly, when a nonoptimal measurement \mathbf{Y} is carried on \mathcal{F} , it is always primarily correlated with the pointer observable \mathbf{A} . In other words, when a measurement \mathbf{Y} is carried on a fragment of the environment, it always reveals more information about the pointer observable \mathbf{A} than any other system observable:

$$I(\mathbf{A}:\mathbf{Y}) \geq I(\mathbf{B}:\mathbf{Y}) \quad (31)$$

for all $\mathbf{B} \in \mathcal{M}_{\mathcal{S}}$ and $\mathbf{Y} \in \mathcal{M}_{\mathcal{F}}$.

Let us prove these two very important assertions. First, note that following Eqs. (8) and (10), the mutual information between any system observable \mathbf{B} and environmental observable \mathbf{Y} takes the following form:

$$I(\mathbf{B}:\mathbf{Y}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{Y}) \quad (32)$$

$$= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{B}). \quad (33)$$

When \mathbf{B} is a *fixed* system observable, maximizing $I(\mathbf{B}:\mathbf{Y})$ amounts to minimizing $H(\mathbf{B}|\mathbf{Y})$. To do this, consider the conditional state $\sigma_{|Y_i}^S$ of the system given a measurement out-

come Y_i on \mathcal{F} . Simple algebraic manipulations can be used to show that

$$\sigma_{|Y_i}^S = \frac{\text{Tr}_{\mathcal{F}}\{Y_i \sigma^{S\mathcal{F}} Y_i\}}{p(Y_i)} = \sum_j p(X_j|Y_i) \sigma_{|X_j}^S. \quad (34)$$

We see that $\sigma_{|Y_i}^S$ is a convex combination of the $\sigma_{|X_j}^S$. (This is no coincidence: it follows from Corollaries 1 and 3 and the existence of information about \mathbf{A} , and hence about \mathbf{X} , in the rest of the environment $\bar{\mathcal{F}}$.) Hence, the inequality of Eq. (30) follows from the convexity of entropy (cf. Proposition 4). Similarly, when \mathbf{Y} is a fixed measurement on \mathcal{F} , maximizing $I(\mathbf{B}:\mathbf{Y})$ amounts to minimizing $H(\mathbf{Y}|\mathbf{B})$. We can consider the state of \mathcal{F} following a measurement of \mathbf{B} on \mathcal{S}

$$\sigma_{|B_i}^{\mathcal{F}} = \frac{\text{Tr}_{\mathcal{S}}\{B_i \sigma^{S\mathcal{F}} B_i\}}{p(B_i)} = \sum_j p(A_j|B_i) \sigma_{|A_j}^{\mathcal{F}}, \quad (35)$$

a convex combination of the states of \mathcal{F} conditioned on a measurement of \mathbf{A} on \mathcal{S} . Again, the inequality Eq. (31) follows from convexity of entropy (cf. Proposition 4).

By combining Eq. (30) with the fact that measuring \mathbf{X} perfectly emulates the direct measurement \mathbf{A} on \mathcal{S} , we get the following equality:

$$\hat{I}_{\mathcal{F}}(\mathbf{B}) = I(\mathbf{B}:\mathbf{A}). \quad (36)$$

The information about \mathbf{B} accessible from the fragment \mathcal{F} is inherently limited by its correlation with the maximally refined observable \mathbf{A} . The only assumption required to arrive at this important equality is that correlations with both the observed and the unobserved parts of the environment impose $\gamma_{ij}^{\mathcal{F}} = \gamma_{ij}^{\bar{\mathcal{F}}} = \delta_{ij}$. This requirement is, according to Corollary 2, equivalent to saying that both \mathcal{F} and $\bar{\mathcal{F}}$ contain a perfect copy of the information about \mathbf{A} .

The consequences of Eq. (36) can be better appreciated once we recognize that it allows to evaluate the redundancy for any system observable straightforwardly. When the value of \mathbf{B} can be deduced from knowledge of the value of \mathbf{A} , then each fragment of \mathcal{E} containing information about \mathbf{A} will inevitably contain information about \mathbf{B} . Formally,

$$I(\mathbf{B}:\mathbf{A}) \geq (1 - \delta) \hat{I}_{\mathcal{E}}(\mathbf{B}) \Rightarrow R_{\delta}(\mathbf{B}) \geq R_{\delta=0}(\mathbf{A}). \quad (37)$$

On the other hand, when the value of \mathbf{B} *cannot* be deduced from knowledge of \mathbf{A} , \mathbf{B} cannot be redundant:

$$I(\mathbf{B}:\mathbf{A}) < (1 - \delta) \hat{I}_{\mathcal{E}}(\mathbf{B}) \Rightarrow R_{\delta}(\mathbf{B}) = 1. \quad (38)$$

This is because Eq. (36) holds as long as $\bar{\mathcal{F}}$ contains information about \mathbf{A} . To have $\hat{I}_{\mathcal{F}}(\mathbf{B}) \geq \hat{I}_{\mathcal{E}}(\mathbf{B})$, forces to take $\bar{\mathcal{F}}$ small enough so that it does not contain much information about \mathbf{A} . By virtue of Eq. (30), $\bar{\mathcal{F}}$ also contains little information about \mathbf{B} , implying $R_{\delta}(\mathbf{B})=1$.

Note that these results can also be derived with the help of the data processing inequality (see Proposition 5). In effect, the update rule of Eq. (5) shows that the sequence of measurements $\mathbf{B}, \mathbf{A}, \mathbf{X}, \mathbf{Y}$ forms a Markov chain (\mathbf{B} and \mathbf{Y} are arbitrary): the joint probability of the measurement results of $\mathbf{B}, \mathbf{A}, \mathbf{X}, \mathbf{Y}$ satisfies

$$p(Y_i, X_j, A_k, B_l) = p(Y_i|X_j)p(X_j|A_k)p(A_k|B_l)p(B_l), \quad (39)$$

yielding Eqs. (30) and (31) directly. This alternative derivation provides a very clear interpretation of our previous result. Trying to gather information about \mathbf{B} (instead of \mathbf{A}) with the indirect measurement \mathbf{Y} (instead of \mathbf{X}) can be viewed as the addition of noise over the perfect communication channel that allows the transmission of information about \mathbf{A} in the environment.

Equations (37) and (38) extend Theorem 1 to imperfect redundant imprints (i.e., finite δ). Only observables “close” to the maximally refined redundant observable \mathbf{A} —where closeness is measured with the help of mutual information—can leave a redundant (even imperfect) imprint in their environment. Moreover, it confirms the behavior found in the numerical study presented in [20].

D. Imperfect correlations

The previous section analyzed the consequences of perfect correlations between \mathcal{S} and \mathcal{F} in terms of optimal measurement strategies. However, perfect correlations are rarely found in Nature. Even for our simple model, perfect correlations can be assumed only in an asymptotic limit, as N tends toward infinity, or by a careful tuning of the interaction time and strength. Hence, it is important to understand what happens when the conditions are not perfect. Here, we show that nearly perfect correlations— $\gamma^{\mathcal{F}}$ and $\gamma^{\bar{\mathcal{F}}}$ sufficiently small—are enough to ensure the validity of the results established above, up to small correction terms.

The technique we use is inspired by perturbation theory. It relies on the construction of a “perfectly correlated” state $\sigma^{\mathcal{S}\mathcal{F}}$ of the form Eq. (27) which is close to the actual state $\rho^{\mathcal{S}\mathcal{F}}$ of Eq. (24) generated by the dynamics. Then, using various bounds on entropies, we will conclude that all the information-theoretic quantities extracted from the ideal $\sigma^{\mathcal{S}\mathcal{F}}$ are approximately equal to those extracted from the actual $\rho^{\mathcal{S}\mathcal{F}}$. We will also examine the regime of validity of this approximation.

Let us define

$$\sigma^{\mathcal{S}\mathcal{F}} = \sum_i |\alpha_i|^2 |i\rangle\langle i| \otimes |\Psi_i^{\mathcal{F}}\rangle\langle\Psi_i^{\mathcal{F}}|, \quad (40)$$

where the $|\Psi_i^{\mathcal{F}}\rangle$'s are obtained by applying the Gram-Schmidt orthonormalization procedure to the states $|\Phi_i^{\mathcal{F}}\rangle$ (see the Appendix for the details of this construction). With this definition, we will show that for any two observables \mathbf{B} and \mathbf{Y} , when $\gamma^{\mathcal{F}} = \max_{ij} |\gamma_{ij}^{\mathcal{F}}|$ and $\gamma^{\bar{\mathcal{F}}} = \max_{ij} |\gamma_{ij}^{\bar{\mathcal{F}}}|$ are small,

$$I_\rho(\mathbf{B}; \mathbf{Y}) \approx I_\sigma(\mathbf{B}; \mathbf{Y}). \quad (41)$$

Above, the subscripts ρ and σ refer to the state, either the actual $\rho^{\mathcal{S}\mathcal{F}}$ or the perfect $\sigma^{\mathcal{S}\mathcal{F}}$, used to derive the probabilities of the measurement outcomes, that in turn are used to quantify information (this shorthand notation will be used in the rest of this section). Therefore, when Eq. (41) holds, all the conclusions derived from the perfect correlation case remain approximately true.

Equation (41) is a consequence of simple inequalities that give an upper bound on the difference $|I_\rho(\mathbf{B}; \mathbf{Y}) - I_\sigma(\mathbf{B}; \mathbf{Y})|$.

First, Cauchy-Schwartz inequality (Proposition 1) gives $\text{Tr}[\rho^{\mathcal{S}\mathcal{F}} - \sigma^{\mathcal{S}\mathcal{F}}] \leq \sqrt{d^{\mathcal{S}}} \|\rho^{\mathcal{S}\mathcal{F}} - \sigma^{\mathcal{S}\mathcal{F}}\|_2$. Second, by definition of the trace distance between two density matrices (Definition 2), $|\sum_{ij} p_\rho(B_i, Y_j) - p_\sigma(B_i, Y_j)| \leq \text{Tr}[\rho^{\mathcal{S}\mathcal{F}} - \sigma^{\mathcal{S}\mathcal{F}}]$. Finally, by combining these results with Lemma 1 and Fanne's inequality (Proposition 3) applied separately to each of the three entropies involved in $I(\mathbf{B}, \mathbf{Y}) = H(\mathbf{B}) + H(\mathbf{Y}) - H(\mathbf{B}, \mathbf{Y})$, we find

$$|I_\rho(\mathbf{B}; \mathbf{Y}) - I_\sigma(\mathbf{B}; \mathbf{Y})| \leq -3f(\gamma^{\mathcal{F}}, \gamma^{\bar{\mathcal{F}}}) \ln[f(\gamma^{\mathcal{F}}, \gamma^{\bar{\mathcal{F}}})/d^{\mathcal{S}}] + O((\gamma^{\mathcal{F}})^{3/4} + \gamma^{\bar{\mathcal{F}}}), \quad (42)$$

with $f(\gamma^{\mathcal{F}}, \gamma^{\bar{\mathcal{F}}}) = \sqrt{d^{\mathcal{S}}[2(d^{\mathcal{S}} - 1)(\gamma^{\mathcal{F}})^2 + (\gamma^{\bar{\mathcal{F}}})^2]}$. Thus, this difference tends to zero when $\gamma^{\mathcal{F}}$ and $\gamma^{\bar{\mathcal{F}}}$ tend to zero.

To gain further insight into the regime of validity of Eq. (41), recall that $\gamma^{\mathcal{F}}$ and $\gamma^{\bar{\mathcal{F}}}$ typically decrease exponentially with the number of environmental subsystems that have effectively interacted with \mathcal{S} . Therefore, for equation Eq. (41) to hold within constant accuracy, it is sufficient that each fragment \mathcal{F} and $\bar{\mathcal{F}}$ contains a number of subsystems \mathcal{E}_k scaling as $\ln d^{\mathcal{S}}$ (or as a polynomial in $\ln d^{\mathcal{S}}$). As discussed at the end of Sec. III, $d^{\mathcal{S}}$ is the effective dimension of the probed degree of freedom of \mathcal{S} , i.e., the number of distinguishable outcomes of an hypothetical measurement \mathbf{B} . Clearly then, $\ln d^{\mathcal{S}}$ is generally much smaller than the size of the environment. Therefore, the conclusions drawn from the perfect correlation study remain essentially unchanged: \mathbf{A} is the observable of the system that leaves the strongest imprint in fragments of the environment; the optimal environmental measurement to learn about any system observable is the one that reveals information about \mathbf{A} ; and finally failing to interrogate \mathcal{F} about \mathbf{A} has the same effect as introducing noise in the measurement results. These conclusions hold to within an accuracy of roughly $\ln d^{\mathcal{S}}/N$, which is enough in most situations involving macroscopic systems.

VII. DISCUSSION: OPEN QUESTIONS AND CONNECTIONS

Environment as a witness as well as quantum Darwinism—the dynamics responsible for the redundant imprinting of certain states that leads to their objectivity—are based on the same dynamical paradigm that is used in the study of decoherence and einselection. The to-be-classical object of interest (the system \mathcal{S} or the apparatus \mathcal{A}) is immersed in the environment. What is now different—and this is a dramatic departure from the usual view of what matters—is the focus. Instead of analyzing either the state of \mathcal{S} *per se*, or even the fate of the correlations between \mathcal{S} and \mathcal{A} , attention shifts to the information available to the observer in *fragments* of \mathcal{E} .

This shift of focus has been precipitated by the realization [17–19] that observers only rarely (if ever) acquire information by direct interaction with systems of interest: rather, we use \mathcal{E} as a communication channel. The message—the state of the object of interest—is imprinted on \mathcal{E} in multiple copies. However, the *no-cloning theorem* implies that arbitrary unknown states cannot be “advertised” throughout the environment in this fashion. Hence, as one can expect (and as our

analysis indeed shows), selection of a preferred observable is a prerequisite for this redundant imprinting.

The basic intuition—the implication of the information-theoretic redundancy of the record for the emergence of the classical—was noted already some time ago [12,19,29]. It was refined into a more precise measure of redundancy—into R_S we employed here—only recently [20]. Using this δ redundancy we have demonstrated—either under very specific assumptions of [20], or in a still rather specific, but somewhat more general setting of this paper—that the familiar pointer observables are easiest to find indirectly, from the records imprinted on the fragments of the environment.

The aim of this section is to examine assumptions that went into our discussion and, by doing so, to explore the range of validity of our conclusions. The results obtained in Sec. V did not require any assumption about the physical model, but only apply to the case of perfect correlations. Thus, the model studied in Sec. VI is a convenient focus of attention. To establish uniqueness of a set of commuting preferred observables that is easiest to infer from \mathcal{E} we have assumed that:

- (1) the initial state is a pure product state;
- (2) the system has no self-Hamiltonian;
- (3) every environmental subsystem couples to the same system observable;
- (4) the environment has no Hamiltonian.

To find out which of these assumptions can be relaxed, and the extent to which they are responsible for the conclusions we have reached will eventually require investigation of other, more realistic, or at least “differently oversimplified” models. For standard decoherence, a similarly idealized model was put forward over two decades ago, but the investigation of various related and unrelated models of decoherence and einselection is still an ongoing activity, often yielding new insights. The same can be expected of quantum Darwinism. Different models may also require different mathematical tools. One such investigation—using von Neumann entropy and the symmetric Eq. (10) for mutual information—is currently under way [25] and points toward compatible conclusions.⁴ All we can offer here is a brief (and possibly premature) discussion of the role and importance of various assumptions, including these listed above.

Assumption 1 is probably the most unrealistic: in practice composite environments are rarely in a pure product state. Luckily, perfect purity *per se* is not essential to rely on the environment as a trustworthy witness. Having enough infor-

mation about \mathcal{E} will, however—and in contrast to decoherence and einselection—prove to be indispensable.

Indeed, it is often convenient to study decoherence assuming a very mixed state of the environment (i.e., thermal equilibrium or even a perfect mixture) when deriving master equations used to implement predictability sieve. Consequences of different initial mixtures of \mathcal{E} are, generally, somewhat different time dependences, and minor changes in the structure of the master equation (which often becomes less Markovian and less tractable when the environment is further away from a convincing thermal state), but the key qualitative conclusions that characterize decoherence and einselection—the localized nature of the preferred states and the fact that relaxation can be much slower than decoherence—seem to be usually unaffected.

By contrast, the degree of ignorance about the fragments of the environment plays an essential role in the study of the environment as a witness. The reader may be surprised by this. After all, the results of Secs. V and VI were derived by an information-theoretic analysis of a rather general scenario, so—given the assumptions—conclusions should be independent of the initial state of \mathcal{E} . In particular, redundancy—when present—remains a highly selective criterion. However, and this is the crux of the matter, the starting point of our investigation—the assumption that fragments of the environment contain information about the system—depends strongly on the initial state of \mathcal{E} .

When the environment is initially mixed, it will “know less” about the system. The *change* of the state of the environment after it has interacted with the system provides the evidence about the state of \mathcal{S} . But when either all of \mathcal{E} or some subsystems of it are totally mixed, they cannot be altered by the conditional dynamics that causes decoherence. And even partial mixing will make it more difficult for the environment to serve as a witness. To see why, let the initial state of the environment be an arbitrary nonentangled state:

$$\rho^{\mathcal{E}} = \sum_{\ell} p_{\ell} \bigotimes_{\mathcal{E}_k} |\phi^{\ell, \mathcal{E}_k}\rangle \langle \phi^{\ell, \mathcal{E}_k}|. \quad (43)$$

In particular, it could be a product of mixed states $\rho^{\mathcal{E}} = \rho^{\mathcal{E}_1} \otimes \rho^{\mathcal{E}_2} \otimes \cdots \otimes \rho^{\mathcal{E}_N}$, such as the thermal state of noninteracting environmental subsystems. By linearity of the Schrödinger equation, the joint state of \mathcal{S} and a fragment \mathcal{F} of \mathcal{E} after an interaction time t will be

$$\rho^{\mathcal{S}\mathcal{F}} = \sum_{\ell} p_{\ell} \sum_{ij} \alpha_i \alpha_j^* |i\rangle \langle j| \bigotimes_{\mathcal{E}_k} |\phi_i^{\ell, \mathcal{E}_k}\rangle \langle \phi_j^{\ell, \mathcal{E}_k}| \times \gamma_{ij}^{\ell, \bar{\mathcal{F}}} \quad (44)$$

where $|\phi_i^{\ell, \mathcal{E}_k}\rangle$ and $\gamma_{ij}^{\ell, \bar{\mathcal{F}}}$ are defined as in Sec. VI A.

In the previous sections, we have studied the kinds of correlations that arise given any of the components $\bigotimes_{\mathcal{E}_k} |\phi_i^{\ell, \mathcal{E}_k}\rangle$ of this mixture. It follows from the convexity of entropy (cf. Proposition 4) that the information contained in this mixture will be strictly less than the average information contained in the individual components of the mixture. As we have noted above, mixing the state of the environmental subsystems decreases their information storage capacity; when the state of the environment is perfectly mixed, it will simply be impossible to learn anything from it. Yet, the usual

⁴The two formulas for the mutual information, Eqs. (8) and (10), cease to coincide when the von Neumann entropy of the two subsystems and the relevant joint and conditional entropies are used on their left hand sides. The difference between the symmetric and asymmetric mutual information is the *discord* [10,19,35]. Discord is a measure of the *quantumness* of the correlation. It can be used to show that even separable correlated states of two systems need not be classical. Discord is suppressed by decoherence. So, when a small fraction of the environment is interrogated by the observer, its correlations with \mathcal{S} tend to be effectively classical. It follows that—in that case—estimates of redundancy based on either the mutual information defined by Eq. (8) or by Eq. (10) are essentially identical.

decoherence still leads to einselection—there will still be a pointer basis (defined, e.g., through the predictability sieve).

When the initial state of \mathcal{E} is not completely mixed, it is usually possible to extract from it some information about the system. In fact, given the assumption of redundancy, the qualitative conclusions of the previous section will still hold: the information about any system observable available in fragments of \mathcal{E} is intrinsically limited by the information about it obtained through a direct measurement of the maximally refined redundantly imprinted observable \mathbf{A} ; Eq. (36) becomes the inequality $\hat{I}_{\mathcal{F}}(\mathbf{B}) \leq I(\mathbf{B}:\mathbf{A})$. Therefore, we again anticipate the conclusion that only observables close to \mathbf{A} can get redundantly imprinted in \mathcal{E} .

Mixed initial states of \mathcal{E} will decrease channel capacity of the environment. Moreover, modes of the environment that can be imprinted with information about different aspects of \mathcal{S} will be in general mixed to a different degree. This may be reflected in the selective coarse-graining with which the observer perceives the system. This raises an interesting possibility: observables that are selected by the predictability sieve need not coincide with the (typically coarse-grained) observables that are easiest to find out about from \mathcal{E} . For example, this may happen when the modes that dominate decoherence (and hence einselection) are so mixed that they cannot serve as a useful communication channel, and the only modes that are sufficiently pure to be useful for this purpose monitor some other observable. An observer will then find out about that other coarse-grained observable, but with the resolution limited by indeterminacy: after all, pointer observable is still einselected by the environmental monitoring. Presence or absence of this effect is related to the assumption of the “typical environment fragment” we have noted in [20]. In other words, when the fragments of \mathcal{E} that the observer can use to extract information about \mathcal{S} contain or reveal data in a selective manner, environment acts as a witness with a selective memory (or a partial amnesia). Exactly how mixed can \mathcal{E} be to be still useful as a witness is of course the key question. We leave it as a subject for further research.

When the initial state of the *system* is mixed, but the environment is sufficiently pure and large so that significant redundant imprinting can arise from their interaction, environment can become a useful and trustworthy witness, and our approach should go through essentially unimpeded. Similarly, preexisting correlation with \mathcal{E} need not undermine our conclusions. In effect, it will typically mean that the environment was gathering evidence about the system in the past (although one can certainly imagine pathological preexisting correlations that get undone by the subsequent interaction).

To sum up, while the assumption of the initial product state has simplified our analysis, it can be relaxed to some extent without undermining our basic conclusions. What is non-negotiable is the demand that the final state allows one to discern evidence about \mathcal{S} in the fragments of \mathcal{E} . One can already anticipate that this demand will also put a constraint on the entanglement between the fragments of \mathcal{E} , as we shall note in the discussion of assumption 4 below.

Assumption 2 precludes non-trivial evolution of the sys-

tem. Relaxing it has been studied in the context of decoherence and einselection, and—as we have done above—we shall gain insight into its role in quantum Darwinism by recounting the implications of evolution for einselection, and analyzing their significance for the role of the environment as a witness.

In the studies of einselection the relative strength of the self-Hamiltonian \mathbf{H}^S and the interaction Hamiltonian $\mathbf{H}^{S\mathcal{E}}$ is an important ingredient that decides the course and effect of einselection. When the self-Hamiltonian is negligible or when \mathbf{H}^S and $\mathbf{H}^{S\mathcal{E}}$ commute, the environment monitors static states of the system—pointer states selected solely by the interaction [4,12]. In a sense, environment contains a record of a very uneventful history—things stay the same forever.

When the effect of \mathbf{H}^S is no longer negligible, these histories become more interesting. One may enter, for example, the “partial Zeno effect” regime where the evolution of the state of the system induced by the self-Hamiltonian is impeded (but not completely suppressed) by the environmental monitoring mediated by $\mathbf{H}^{S\mathcal{E}}$ (see, e.g., [8]). Models of such situations have been analyzed using rather different approach of quantum trajectories [36] that shares with quantum Darwinism the focus on the information gained indirectly from \mathcal{E} . There the environment is usually assumed to consist of photons, and the system is often a two-level atom. Pointer states still emerge and turn out to be the easiest states to infer from a fraction of such \mathcal{E} [37]. The basis ambiguity may reappear only if the whole \mathcal{E} could be captured. This ability to rely on a fraction of \mathcal{E} allows for the possibility of consensus between multiple observers [38].

As the coupling to the environment becomes weak the problem may become more tractable again, especially in two rather different limits. When the environment is slow, with the high-frequency cutoff that is small compared to the level spacing of the spectrum of \mathbf{H}^S , einselection enters its quantum limit [39]. Energy eigenstates are imprinted on the environment more or less regardless of the specific form of the coupling. Histories are again becoming rather uneventful, although for reasons quite different from these we have mentioned above in the negligible \mathbf{H}^S case.

By contrast, when the environment is broadband, with all the frequencies from very low to very high present, and the system is (at least approximately) linear and coupled to \mathcal{E} through its position, coherent states are the most predictable, and hence, pointer. Weak coupling does not preclude redundancy of the records. In this underdamped limit one can recover approximately reversible Liouvillian dynamics with localized states (i.e., approximately reversible classical trajectories that follow Newton’s laws emerge; see [9,10], and references therein). When the system is linear (e.g., harmonic oscillator) preferred pointer states turn out to be Gaussian [17,24], and one may anticipate that sequences of the most redundant “environmental records” will correlate with classical trajectories.

The quick summary of different possibilities above is meant to be suggestive rather than exhaustive, and is definitely not conclusive. What one would like to recover using environment as a witness is the notion of objective histories defined through sequences of time-ordered redundant records. Such objective histories could consist of sequences

of approximate instantaneous pointer states. They were conjectured to exist in the limit of very efficient and frequent broadband monitoring by the environment in [10]. Histories deduced from redundant records would be objective, and may be only approximate. It will be interesting to investigate their relation to consistent histories [40–44], especially when consistency is assured through the “strong decoherence” condition introduced in [30] (see also [31]) that invokes existence of a perfect environmental record.

Assumption 3 assures that there is a preferred observable of \mathcal{S} that is recorded by \mathcal{E} , and that it is singled out by the interaction Hamiltonian. This is a strong assumption, but often a reasonably realistic one: typical couplings between the systems (and, hence, between \mathcal{S} and \mathcal{E}) tend to depend on the distance (i.e., relative position) between them. As a consequence—as was already pointed out in the context of einselection [7,12]—pointer states tend to be localized in position.

One can of course imagine situations where this is not the case. For instance, there may be several physically distinct environments, each attempting to monitor a different observable of \mathcal{S} . One would expect that a dominant coupling would then impose its selection of the preferred observable, and the effect of the others would be perceived as noise.

This last remark leads to interesting questions about the correspondence between the approximate pointer observable that is selected by the predictability sieve as a result of the interaction of \mathcal{S} and \mathcal{E} , and the states of the system that are easiest to infer for an observer who has access to a single subsystem of the environment. Clearly, in full generality this case is not covered by our idealized assumptions earlier in the paper. For example, Gaussian pointer states are often selected by the predictability sieve, and yet they are not orthogonal and form an overcomplete basis, so some of the steps we have used in our proofs may not go through. Nevertheless, one may expect that—as was the case with decoherence—results that are exact for perfect pointer states (that are left untouched by decoherence) are still approximately valid for approximate pointer states (that are least touched by decoherence). Given the role Gaussian states play in the discussion of various models of decoherence, the extent to which Gaussian states coincide with the states that are easiest to infer is of obvious interest, and a likely area of future studies, but beyond the scope of this paper (see however [45]). Along the same line of thought, when the environment fragments that are the most useful to the observer interested in \mathcal{S} are not a typical sample of the environments responsible for decoherence, one may still guess that the states of \mathcal{S} that can be inferred from them will represent an appropriately coarse-grained version of the pointer states.

Assumption 4 assures that the environment does not evolve (and, therefore, does not complicate or even obliterate the record it has made of \mathcal{S}). More careful study of the consequences of relaxing this assumption can be found elsewhere [46]. Here we offer only a brief discussion of the basic possibilities.

One can distinguish three components of this requirement: fragments of the environment could evolve separately, they could interact with each other, or could interact with more distant second-order environments.

Let us first note that—in models where pointer observable commutes with the system-environment interaction Hamiltonian [e.g., $\mathbf{H}^{\mathcal{S}\mathcal{E}} = \sum_{\mathcal{E}_k \in \mathcal{E}} \mathbf{A} \otimes \mathbf{Z}^{\mathcal{E}_k}$, Eq. (19)] self-Hamiltonians of the individual fragments of \mathcal{E} obviously cannot change the fact that $[\mathbf{H}^{\mathcal{S}\mathcal{E}}, \mathbf{A}] = 0$, so the selection of the pointer basis is not affected by the dynamics of the environment. However, even in this simple case the presence of such self-Hamiltonians can influence the rate of acquisition of information by \mathcal{E} and, hence, redundancy [46]. We also note that in more realistic cases when the system has a self-Hamiltonian that does not commute with $\mathbf{H}^{\mathcal{S}\mathcal{E}}$ the pointer observable is approximate, and even its choice can be influenced by how the rate of information acquisition by \mathcal{E} compares with the dynamical time scales in \mathcal{S} .

The interaction between the subsystems of \mathcal{E} may be a significant complication. It will typically lead to entanglement between the fragments of \mathcal{E} , and could make their individual states mixed. Thus—as we have already pointed out in discussion of assumption 1—even though the environment as a whole may still contain redundant imprint of the state of the system, the relevant global observables could become effectively inaccessible to the observers who can sample \mathcal{E} only through local measurements on its fragments.

Last but not least, \mathcal{E} could be immersed in its own environment \mathcal{E}' . Effect of this second-order environment will depend on the nature of their interaction. It could of course erase the records of \mathcal{S} in \mathcal{E} , but it is also conceivable that \mathcal{E}' could monitor the record-keeping observables of the fragments of \mathcal{E} , which would increase the redundancy [10].

These are very interesting complications, very much worthy of further study. We note, however, that human observers tend to rely on photons for the vast majority of data about systems of interest. And photons, in effect, satisfy our assumption 4. This seems to us to be no accident.

It is clear from the above discussion that quantum Darwinism requires stronger assumptions than decoherence. This is not surprising: it is easier to use any system (including the environment) as a disposal, to get rid of unwanted information, than as a communication channel. Furthermore, existence of the pointer states is a necessary but not a *sufficient* condition for their redundant imprinting on \mathcal{E} . Nevertheless, even though quantum Darwinism relies on stronger assumptions than einselection implemented, e.g., through the predictability sieve, it is also clear from our discussion that these more restrictive assumptions are met in the real world often enough to employ environment as a reliable witness.

VIII. SUMMARY AND CONCLUSION

Observers use environment as a witness. This has obvious and profound implications for the interpretation of quantum theory. Nevertheless, it is not our intent here to go beyond the comments we have made throughout and to delve into a more extensive discussion of the implications of quantum Darwinism for the interpretation of quantum theory. In part, this is because some of the interpretational issues have been noted elsewhere, on the occasion of discussions of the role of redundancy in the *existential interpretation* (see [10,17,18,21], and references therein). The basic idea—that

the classical domain can arise from within a quantum Universe, and that it consists of states that can be found out without being destroyed, and, hence, in that operational sense exist independently of the observer—is obviously very much assisted by employing environment as a witness.

Quantum Darwinism uses the same model of the information transfer that was introduced to study decoherence and einselection, but asks a different question: instead of focusing—as does decoherence—on how the information is lost from the system, it analyzes how the information is gained and stored in the environment. In decoherence, the role of the environment is limited to hiding the underlying quantumness (view especially emphasized in discussions of quantum computation). Einselection recognizes that this information loss is selective, and, thus, that the environment has a capacity to single out preferred pointer states that are best at surviving immersion in \mathcal{E} . Moreover, they are chosen in the process reminiscent of a measurement— \mathcal{E} acts as an apparatus that (pre)measures the pointer observable of the system of interest.

This much was known. In particular, the monitoring role of \mathcal{E} was recognized early on [4]. Quantum Darwinism answers to the next logical question: since the environment interacts with the to-be-classical observable of the system of interest as would an apparatus, can it be used as an apparatus? To address it we have had to develop a measure of the objectivity of the records deposited in the environment. Redundancy (briefly considered [29], see also [12] in a similar context as a measure of amplification) was refined [20] and used to prove that—given reasonable assumptions—there is a unique set of commuting observables that are easiest to find out from fragments of \mathcal{E} , and that these redundantly recorded observables are indeed the familiar pointer observables.

As a consequence, observers monitoring fragments of the environment will be able to reach compatible conclusions about the value of redundantly imprinted observables of the system, i.e., about the *objective* properties of the system. In this operational sense, quantum Darwinism provides a satisfying explanation for the emergence of the objective classical world we perceive from the underlying quantum substrate.

We have illustrated quantum Darwinism on a dynamical model that is easy to analyze from an information-theoretic perspective, and yet is inspired by a photon environment scattering on an object: all the environmental subsystems couple to the object with the same Hamiltonian and do not interact with each other. This model enabled us to push our analysis further, as we could explicitly determine the optimal environmental measurements to learn about the properties of the system. It also demonstrated that quantum Darwinism tolerates reasonable departures from the “ideal” assumptions used to derive our main results in Sec. V. In this respect, it corroborates recent studies [10,20] and earlier conjectures [19,29] that noted the role of redundancy in assuring *resilience* of certain states—an essential feature of the classical domain. Hence, redundancy is a robust and selective criterion for determining objective properties of open quantum systems.

While redundantly imprinted observables are dynamically stable—as defined by einselection—the reverse is not neces-

sarily true: quantum Darwinism requires stronger assumptions, especially regarding structure, size, and the initial state of \mathcal{E} . For example, a single photon can be enough to decohere an object in superposition of two distinct positions. However, it takes a macroscopic number of them to redundantly broadcast the position of this object throughout the environment. While the entropy production of the system—a signature of decoherence—rapidly saturates, redundancy can then continue to grow with time. This illustrates how the increase of the redundancy of the record in the environment captures the fact that the system is still continuously under observation, and that information about the pointer observable is getting amplified up to the macroscopic level. In this sense, quantum Darwinism can be motivated using Bohr’s original ideas [2] about the role of *amplification* in the transition from quantum to classical.

ACKNOWLEDGMENTS

We would like to thank Robin Blume-Kohout for extensive and enjoyable discussions. This research was supported in part by NSA and ARDA. H.O. receives financial support from ACI Sécurité Informatique, and D.P. from Canada’s NSERC and Québec’s NATEQ.

APPENDIX

Definition 2 (trace distance). The trace distance between two density operators ρ, σ is defined by $D(\rho, \sigma) = \text{Tr}|\rho - \sigma|$. When ρ and σ commute, the trace distance is equal to the L_1 distance between the classical probability distributions defined by their eigenvalues. Alternatively, the trace distance is equal to $2 \max_P \text{Tr} P(\rho - \sigma)$ where the maximum is taken over positive operators $P \leq I$.

Definition 3 (Euclidean distance). The Euclidean distance between a pair of density operators ρ, σ is defined by $\|\rho - \sigma\|_2 = \sqrt{\text{Tr}\{(\rho - \sigma)^2\}}$.

Proposition 1 (Cauchy-Schwartz inequality). $D(\rho, \sigma) \leq \sqrt{d} \|\rho - \sigma\|_2$.

Proposition 2 Let $p_i = \text{Tr}\{A_i \rho\}$ and $q_i = \text{Tr}\{A_i \sigma\}$ be the probabilities of the outcomes of a measurement $\{A_i\}$ for states ρ and σ , respectively. Then, $D(p, q) \leq D(\rho, \sigma)$.

Proposition 3 (Fanne’s inequality). Let p_i and q_i with $i=1, 2, \dots, d$ be two classical probability distributions such that $D(p, q) \leq 1/e$. Then $|H(p) - H(q)| \leq D(p, q) \ln d + \eta(D(p, q))$, with $\eta(x) = -x \ln x$.

Proposition 4 (convexity of entropy). Let $P(X_i|Y_j)$ be a set of conditional probability distribution for the random variable X , and $P(Y_j)$ be the probability distribution for the random variable Y . Then

$$\sum_j \left[P(Y_j) \left(- \sum_i P(X_i|Y_j) \ln P(X_i|Y_j) \right) \right] \leq - \sum_i P(X_i) \ln P(X_i) \quad (\text{A1})$$

where $P(X_i) = \sum_j P(X_i|Y_j)P(Y_j)$.

Proposition 5 (data-processing inequality). Suppose $X \rightarrow Y \rightarrow Z$ is a Markov chain, then $H(X) \geq I(X:Y) \geq I(X:Z)$.

Lemma 5. For any matrix $\rho = \sum_{ij} \alpha_i \alpha_j^* |i\rangle\langle j| \otimes |\Phi_i^{\mathcal{F}}\rangle\langle\Phi_j^{\mathcal{F}}| \gamma_{ij}^{\bar{\mathcal{F}}}$, with $\alpha_i \neq 0$, and such that

- (1) $\langle i|j\rangle = \delta_{ij}$;
- (2) $|\Phi_i^{\mathcal{F}}\rangle$ are linearly independent;
- (3) $|\gamma_{ij}^{\bar{\mathcal{F}}}| \leq \gamma^{\bar{\mathcal{F}}}$ for $i \neq j$, and $\gamma_{ii}^{\bar{\mathcal{F}}} = 1$ for all i ;
- (4) $|\langle\Phi_i^{\mathcal{F}}|\Phi_j^{\mathcal{F}}\rangle| = |\gamma_{ij}^{\mathcal{F}}| \leq \gamma^{\mathcal{F}}$ for $i \neq j$,

there exists a matrix $\sigma = \sum_i |\alpha_i|^2 |i\rangle\langle i| \otimes |\Psi_i^{\mathcal{F}}\rangle\langle\Psi_i^{\mathcal{F}}|$, satisfying

- (1) $\langle\Psi_i^{\mathcal{F}}|\Psi_j^{\mathcal{F}}\rangle = \delta_{ij}$;
- (2) $\|\rho - \sigma\|_2 \leq \sqrt{2(d^{\mathcal{S}} - 1)} \gamma^{\mathcal{F}} + \gamma^{\bar{\mathcal{F}}} + O((\gamma^{\mathcal{F}})^{3/2} + (\gamma^{\bar{\mathcal{F}}})^2)$,

where $d^{\mathcal{S}}$ is the number of coefficients α_i .

Proof. The proof of this lemma relies on a careful analysis of Gram-Schmidt orthonormalisation procedure applied to the states $|\Phi_i^{\mathcal{F}}\rangle$.

We define the new set of orthogonal states $|\Psi_i^{\mathcal{F}}\rangle$ by the following equations:

$$|\Psi_i^{\mathcal{F}}\rangle = \frac{1}{N_i} \left(|\Phi_i^{\mathcal{F}}\rangle - \sum_{j < i} \langle\Psi_j^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle |\Psi_j^{\mathcal{F}}\rangle \right), \quad (\text{A2})$$

where N_i is a normalization factor.

More precisely, we prove by induction on i that the following properties hold:

$$|\langle\Phi_j^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle| \begin{cases} \leq 2\gamma^{\mathcal{F}} + O((\gamma^{\mathcal{F}})^2), & j < i, \\ \geq 1 - \frac{i-1}{2}(\gamma^{\mathcal{F}})^2 + O((\gamma^{\mathcal{F}})^3), & j = i, \\ \leq \gamma^{\mathcal{F}} + O((\gamma^{\mathcal{F}})^2), & j > i. \end{cases} \quad (\text{A3})$$

For $i=1$, the result is trivial, since $|\Psi_1^{\mathcal{F}}\rangle = |\Phi_1^{\mathcal{F}}\rangle$.

For $i > 1$, we assume that Eq. (A3) holds for $k < i$. By definition of N_i we have,

$$N_i^2 = 1 - \sum_{k < i} |\langle\Phi_i^{\mathcal{F}}|\Psi_k^{\mathcal{F}}\rangle|^2 \quad (\text{A4})$$

$$\geq 1 - (i-1)(\gamma^{\mathcal{F}})^2 + O((\gamma^{\mathcal{F}})^3). \quad (\text{A5})$$

Moreover, with Eq. (A2) we also have

$$|\langle\Phi_j^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle| = \frac{1}{N_i} \left| \langle\Phi_j^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle - \sum_{k < i} \langle\Psi_k^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle \langle\Phi_j^{\mathcal{F}}|\Psi_k^{\mathcal{F}}\rangle \right|.$$

(1) When $j < i$ we get

$$|\langle\Phi_j^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle| \leq \frac{1}{N_i} \left(|\langle\Phi_j^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle| + \left| \sum_{k < i, k \neq j} \langle\Psi_k^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle \langle\Phi_j^{\mathcal{F}}|\Psi_k^{\mathcal{F}}\rangle \right| + |\langle\Psi_j^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle \langle\Phi_j^{\mathcal{F}}|\Psi_j^{\mathcal{F}}\rangle| \right) \quad (\text{A6})$$

$$\leq 2\gamma^{\mathcal{F}} + O((\gamma^{\mathcal{F}})^2). \quad (\text{A7})$$

(2) For $i=j$,

$$|\langle\Phi_i^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle| = N_i \quad (\text{A8})$$

$$\geq 1 - \frac{i-1}{2}(\gamma^{\mathcal{F}})^2 + O((\gamma^{\mathcal{F}})^3). \quad (\text{A9})$$

(3) Finally, in the case $j > i$, the contribution at order $\gamma^{\mathcal{F}}$ comes from $\langle\Phi_j^{\mathcal{F}}|\Phi_i^{\mathcal{F}}\rangle$ which gives

$$|\langle\Phi_j^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle| \leq \gamma^{\mathcal{F}} + O((\gamma^{\mathcal{F}})^2). \quad (\text{A10})$$

With this result, we can easily calculate an upper bound on $\|\rho - \sigma\|_2$:

$$\begin{aligned} \text{Tr}(\rho - \sigma)^2 &= \text{Tr} \rho^2 + \text{Tr} \sigma^2 - 2 \text{Tr} \rho \sigma \\ &= 2 \sum_i |\alpha_i|^4 (1 - |\langle\Phi_i^{\mathcal{F}}|\Psi_i^{\mathcal{F}}\rangle|^2) + \sum_{i \neq j} |\alpha_i|^2 |\alpha_j|^2 |\gamma_{ij}^{\bar{\mathcal{F}}}|^2 \\ &\leq 2(d^{\mathcal{S}} - 1)(\gamma^{\mathcal{F}})^2 + (\gamma^{\bar{\mathcal{F}}})^2 + O((\gamma^{\mathcal{F}})^3). \end{aligned} \quad (\text{A11})$$

Therefore we have $\|\rho - \sigma\|_2 \leq \sqrt{2(d^{\mathcal{S}} - 1)(\gamma^{\mathcal{F}})^2 + (\gamma^{\bar{\mathcal{F}}})^2} + O((\gamma^{\mathcal{F}})^{3/2})$, concluding the proof. ■

- [1] *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. H. Zurek (Plenum Press, Princeton, NJ, 1983).
- [2] N. H. D. Bohr, *Nature (London)* **121**, 580 (1928), reprinted in [1].
- [3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1955).
- [4] W. H. Zurek, *Phys. Rev. D* **24**, 1516 (1981).
- [5] H. Everett III, *Rev. Mod. Phys.* **29**, 454 (1957).
- [6] H. D. Zeh, *Found. Phys.* **3**, 109 (1973); e-print quant-ph/0306151.
- [7] W. H. Zurek, *Phys. Today* **44**(10), 36 (1991).
- [8] E. Joos, H. D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I.-O. Stamatescu, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer-Verlag, Heidelberg, 2003).
- [9] J.-P. Paz and W. H. Zurek, in *Coherent Atomic Matter Waves*, edited by R. Kaiser, C. Westbrook, and F. David, *Proceedings*

of the Les Houches Summer School, LXXII, 1999 (Springer, Berlin, 2001), pp. 533–614.

- [10] W. H. Zurek, *Rev. Mod. Phys.* **75**, 715 (2003).
- [11] M. Schlosshauer, *Rev. Mod. Phys.* **76**, 1267 (2004).
- [12] W. H. Zurek, *Phys. Rev. D* **26**, 1862 (1982).
- [13] C. M. Caves, K. S. Thorne, W. P. Drever, V. D. Sandberg, and M. Zimmermann, *Rev. Mod. Phys.* **52**, 341 (1980).
- [14] V. B. Braginsky and F. Y. Khalili, *Rev. Mod. Phys.* **68**, 1 (1996).
- [15] P. Grangier, *Eur. J. Phys.* **23**, 331 (2002).
- [16] D. Poulin, *Phys. Rev. A* **71**, 022102 (2005).
- [17] W. H. Zurek, *Prog. Theor. Phys.* **89**, 281 (1993).
- [18] W. H. Zurek, *Philos. Trans. R. Soc. London, Ser. A* **356**, 1793 (1998).
- [19] W. H. Zurek, *Ann. Phys.* **9**, 855 (2000).
- [20] H. Ollivier, D. Poulin, and W. H. Zurek, *Phys. Rev. Lett.* **93**, 220401 (2004).

- [21] W. H. Zurek, e-print quant-ph/0308163.
- [22] A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777 (1935).
- [23] W. K. Wootters and W. H. Zurek, *Nature (London)* **299**, 802 (1982); D. Dieks, *Phys. Lett.* **92A**, 271 (1982).
- [24] W. H. Zurek, S. Habib, and J.-P. Paz, *Phys. Rev. Lett.* **70**, 1187 (1993).
- [25] R. Blume-Kohout and W. H. Zurek, e-print quant-ph/0408147; see also e-print quant-ph/0505031.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Series in Telecommunication (John Wiley and Sons, New York, 1991).
- [27] R. Landauer, *Phys. Today* **44**(5), 23 (1991).
- [28] K. Kraus, *States, Effects and Operations*, Fundamental Notions of Quantum Theory (Academic Press, Berlin, 1983).
- [29] W. H. Zurek, in *Quantum Optics and Experimental Gravitation and Measurement Theory*, edited by P. Meystre and M. O. Scully (Plenum, New York, 1983), pp. 87–116.
- [30] M. Gell-Mann and J. B. Hartle, in *Proceedings of the Fourth Drexel Symposium on Quantum Non-integrability*, edited by D.-H. Feng and B.-L. Hu, (International Press of Boston, Hong Kong, 1997).
- [31] J. J. Halliwell, *Phys. Rev. D* **60**, 105031 (1999).
- [32] C. Fuchs, e-print quant-ph/0205039.
- [33] W. H. Zurek, F. Cucchietti, and J.-P. Paz, e-print quant-ph/0312207.
- [34] W. H. Zurek, *Phys. Rev. A* **71**, 052105 (2005).
- [35] H. Ollivier and W. H. Zurek, *Phys. Rev. Lett.* **88**, 017901 (2002).
- [36] H. Carmichael, *An Open Systems Approach to Quantum Optics*, Lecture Notes in Physics Vol. 18 (Springer-Verlag, Berlin, 1993).
- [37] D. A. R. Dalvit, J. Dziarmaga, and W. H. Zurek, *Phys. Rev. Lett.* **86**, 373 (2001).
- [38] D. Dalvit, J. Dziarmaga, and W. H. Zurek, *Phys. Rev. A* **69**, 022109 (2004).
- [39] J.-P. Paz and W. H. Zurek, *Phys. Rev. Lett.* **82**, 5181 (1999).
- [40] R. B. Griffiths, *J. Stat. Phys.* **36**, 219 (1984).
- [41] M. Gell-Mann and J. B. Hartle, in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Reading, MA, 1990).
- [42] R. Omnès, *Rev. Mod. Phys.* **64**, 339 (1992).
- [43] R. Omnès, *Understanding Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1999).
- [44] R. B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, Cambridge, U.K., 2002).
- [45] D. A. R. Dalvit, J. Dziarmaga, and W. H. Zurek, e-print quant-ph/0509174.
- [46] R. Blume-Kohout and W. H. Zurek (unpublished).