CrossMark

Entropic dynamics: a hybrid-contextual model of quantum mechanics

Kevin Vanslette [iD]

Abstract   The Bell–Kochen–Specker (BKS) theorem and the more recent ψ-epistemic *no-go* theorems of Quantum Mechanics (QM) are discussed in the context of entropic dynamics. In doing so we find that the BKS theorem allows for, a perhaps overlooked, hybrid-contextual model of QM in which one set of commuting observables (position in this case) is non-contextual and all other observables are contextual. Entropic Dynamics is in a unique position as compared to other foundational theories of QM because it derives QM using standard techniques in Bayesian probability theory. In this formalism, position is the preferred basis from which inferences about other contextual operators are made. This leads to the interpretation that Entropic Dynamics is a hybrid-contextual model of QM, which we show to be consistent with the BKS theorem and QM.

1 Introduction

Quantum Mechanics is an odd mix of the predictable and unpredictable. On one hand, it is hugely successful in its ability to predict the set of eigenvalues, expectation values, and operators for a particle-system of interest. On the other hand, each measurement holds some amount of unpredictability, quantified by a probability distribution, except for a few trivial cases. This unpredictable nature leaves a space for the many interpretations of QM to coexist inharmoniously within the community—a community, no doubt, easily bothered by disharmony of any-type.

The community reduces and organizes this disharmony by ruling out interpretations and foundational theories of QM that disagree with the predictable findings of QM. This is done by first making a few reasonable assumptions a theory of QM may obey, and then by showing these assumptions lead to contradictions in the formalism, construct a *no-go* theorem. This is the basis of the Bell inequalities [4], the BKS theorem [5,23,25], as well as the findings of Pusey–Barret–Rudolph (PBR) [28] (reviewed in [24]) on the epistemic interpretation of the wavefunction. Post-analysis, the final results are sometimes tabulated in 2 by 2 tables—for example:

K. Vanslette (✉)
University at Albany (SUNY), Albany, NY 12222, USA
e-mail: kvanslette@albany.edu; kvanslette77@gmail.com

	ψ-ontic	ψ-epistemic
contextual	A	B
noncontextual	C	D

,

and are followed by statements like, "theories of type "C" or "D" are ruled out by the BKS theorem and "B" is ruled out by PBR". A reader may be inclined to conclude that QM must be a theory of type "A" (potential interpretation of Bohmian mechanics). The 2 by 2 is by nature an over simplification; it fails to span the entire set of plausible theories, and consequently interpretations, of QM. This is due the fact that *no-go* theorems are proofs by contradiction, and only theories which strictly adhere to their (shown invalid) assumptions are ruled out.

In particular, this paper will show entropic dynamics (ED) to be a theory of QM which lies on the line between theories "B" and "D", while not being ruled out by any of the aforementioned *no-go* theorems. We classify ED as a hybrid-contextual theory of QM because positions are treated noncontextually while all other observables are treated contextually—the main result of this paper. In the same way, other theories and interpretations of QM may slip between the cracks of these *no-go* theorems, which reopens the perhaps presumed closed universe of discourse for a few edge theories and interpretations of QM.

ED will be reviewed briefly as it pertains to the *no-go* theorems of interest. New insights into how these *no-go* theorems are handled within ED will be presented and some critiques will be given. The sense in which a theory can be hybrid-contextual and still obey QM will be discussed.

2 Entropic dynamics

The axioms of a foundational Physics program must in large originate from outside of Physics or else one runs the risk of "using Physics to derive Physics". In this sense, ED is a foundational framework for Physics [8,9]. Starting from the laws (or axioms) of probability theory, probability updating, and inference methods [8,9,12,18–20], ED is able to reformulate parts of Physics as tools for inference [8,9,17]. It should be noted that ED is not an "be all end all" for Physics—rather it generates models for inference that happen to be consistent with Physics. Along a similar line of thought, the "discoveries" in ED are the inferential constraints and pertinent information required to obtain Physics from probability theory rather than the physics equations themselves. Current and future research in ED involves formulating other well known laws of Physics, attempting to refine and strengthen methods in ED [7], and using ED to address conceptual or paradoxical issues in Physics—the later being the central focus of this article.

Here we are interested in the constraints and assumptions required to derive QM from the first principles of inference and probability updating.[1] We will briefly review the aspects of ED pertaining to the *no-go* theorems of interest. The full derivation of QM from ED may be found in [9].

2.1 From ED to QM

The first step in any inferences problem is to state the universe of discourse, the set of possible outcomes or microstates, one would like to infer on the basis of incomplete information. To derive quantum mechanics (in flat space) the universe of discourse spanned by N particles are their positions in a flat Euclidean space **X** (metric δ_{ab}). Our knowledge of the positions of particles is characterized by a probability density $\rho(x)$ where x is a coordinate in a $3N$ dimensional configuration space of particle coordinates x_n^a, where $a = 1, 2, 3$ denotes the ath spacial axis of the nth particle's position. When convenient we use a super-index notation $x^A \equiv x_n^a$, where $A = (n, a)$, and the Einstein summation convention. From the onset, particles have *definite* yet *unknown* positions and are treated as

---

[1] Other fields of Physics (Statistical Mechanics, QFT, ect.) have different relevant informational constraints in ED due to differences in phenomenology [8,17].

the "physical" or "ontological" quantities we are interested in inferring. QM will be derived using this universe of discourse—namely that particle positions are noncontextual.

Now that the microstates have been specified, we are inclined to ask how the position of these particles change. In particular, if the particles are located at $x = \{x^A\}$, we wish to know how probable it is for $x \to x'$, that is, we seek a transition probability of the form $P(x'|x)$ to quantify this uncertainty. Not knowing anything about how particles change position gives trivial dynamics, so we make the following assumptions: (1) particles move along continuous trajectories and (2) particles may have a tendency to be correlated and drift. These assumptions are represented by expectation value constraints on $P(x'|x)$. The first assumption is saturated by making large $\Delta x_n^a = x_n^{'a} - x_n^a$ improbable. This is done by imposing $P(x'|x)$ have small variances, κ_n, in particle coordinates,

$$\langle \Delta x_n^a \Delta x_n^b \rangle \delta_{ab} = \kappa_n, \qquad (n = 1, \ldots, N) \tag{1}$$

where motion is continuous in the limit $\kappa_n \to 0$. The second assumption is imposed by one additional expectation value constraint,

$$\langle \Delta x^A \rangle \partial_A \phi \equiv \sum_{n,a} \langle \Delta x_n^a \rangle \frac{\partial \phi(x)}{\partial x_n^a} = \kappa', \tag{2}$$

where κ' is another small constant and $\frac{\partial \phi(x)}{\partial x_n^a}$ is the "drift" gradient that guides probability flow and fixes the average displacement through the constant κ'.

There are many probability distributions $P(x'|x)$ that satisfy the above expectation value constraints. We, therefore, use the method of maximum entropy [8,18–20] to rank the candidate distributions and select the "least biased" distribution by maximizing the relative entropy $S[p(x), q(x)]$. Given a prior state of knowledge $q(x)$ and some expectation values one knows $p(x)$ ought to obey, the method of maximum entropy updates the prior distribution $q(x)$ to a new (posterior) state of knowledge $q(x) \to p(x)$. This method of inference is consistent with Bayesian probability updating [15,16] and, in-fact, generalizes Bayesian inference. The method of maximum entropy is, therefore, a nature tool for performing inference in ED.

In the present case, we are interested in updating a prior transition probability distribution $Q(x'|x)$ to a posterior transition probability distribution $P(x'|x)$ using the method of maximum entropy. The prior transition distribution $Q(x'|x)$ is a very broad normalizable Gaussian distribution to encode that, given nothing is known about particle motion (equations (1) and (2) are yet to be imposed), particles may jump anywhere with near to equal probability— there is no reason to believe otherwise. We will maximize the relevant relative entropy,

$$S[P(x'|x), Q(x'|x)] = -\int \mathrm{d}x' \, P(x'|x) \log \frac{P(x'|x)}{Q(x'|x)}, \tag{3}$$

with respect to the expectation value constraints, (1) and (2), via the Lagrange multiplier method. Letting $\{\alpha_n\}$ be the particle specific Lagrange multipliers that impose the N constraints from (1) and letting α' be the Lagrange multiplier which imposes (2), maximizing the entropy ($S = S[P(x'|x), Q(x'|x)]$) with respect to these constraints (and normalization) is setting the variation,

$$\delta \left(S + \sum_n \alpha_n (\langle \Delta x_n^a \Delta x_n^b \rangle \delta_{ab} - \kappa_n) + \alpha'(\langle \Delta x^A \rangle \partial_A \phi - \kappa') \right)$$

equal to zero. Varying $P(x'|x)$ above gives,

$$\int \left(-\log \left(\frac{P(x'|x)}{Q(x'|x)} \right) - 1 + \sum_n \alpha_n \Delta x_n^a \Delta x_n^b \delta_{ab} + \alpha' \Delta x^A \partial_A \phi \right) \delta P(x'|x) \, \mathrm{d}x' = 0,$$

which is stationary for arbitrary variations of $P(x'|x)$ when,

$$P(x'|x) = \frac{1}{Z} \exp \left[-\frac{1}{2} \sum_A \alpha_n (\Delta x^A - \langle \Delta x^A \rangle)^2 \right], \tag{4}$$

after completing the square, where $\langle \Delta x^A \rangle = \frac{\alpha'}{\alpha_n} \delta^{ab} \frac{\partial \phi}{\partial x_n^b}$. Because $Q(x'|x)$ is nearly constant over regions of interest, it has been absorbed into the normalization constant Z. In principle, at anytime during the calculation the expectation values on the RHS of (1) and (2) may be calculated by taking their corresponding expectation values over $P(x'|x)$.

There are explicit and in-depth arguments for the notion of an instant in ED [8,9] that we need not explore here; however, in summary, t is introduced as a convenient book keeping parameter to label changes in the probability distribution. In particular, $\alpha_n \propto \frac{1}{\Delta t}$ such that equal amounts of time are measured by equal fluctuations of position. Because α_n is a particle specific Lagrange multiplier, we expect it to have particle specific parameters, i.e $\alpha_n = \frac{m_n}{\eta \Delta t}$, where we see at the end of the calculation that m_n is the particle specific mass and η is a constant that fixes units and is related to $\hbar$. After these arguments, the transition probability reads,

$$P(x'|x, \Delta t) = \frac{1}{Z} \exp \left[-\sum_A \left(\frac{m_n}{2\eta \Delta t} \left(\Delta x^A - \langle \Delta x^A \rangle \right)^2 \right) \right], \tag{5}$$

such that the state of knowledge of the positions of particles at a later time t' is given by marginalizing over the previous position coordinates,

$$\rho(x'|t') = \int P(x'|x, \Delta t) \rho(x|t) \, dx, \tag{6}$$

where $\rho(x|t) \equiv \rho(x) \equiv \rho$. Equation (6) is the integral form of the Fokker–Planck (diffusion) equation and may be recast as the differential Fokker–Planck equation (an explication may be found in [8] which involves introducing a test function and integrating by parts),

$$\partial_t \rho = -\partial_A \rho \left(m^{AB} \partial_B \eta \phi - \eta \log \rho^{1/2} \right) = -\partial_A \left(\rho v^A \right), \tag{7}$$

where the current "velocity" v^A of the probability flow in configuration space is,

$$v^A = m^{AB} \partial_B \Phi \quad \text{where} \quad m^{AA'} = m_n^{-1} \delta^{aa'} \delta_{nn'}, \tag{8}$$

$$\text{and} \quad \Phi = \eta \phi - \eta \log \rho^{1/2}, \tag{9}$$

is a function defined in terms of previously defined variables. As the current velocity v^A is a gradient of Φ from (8), the current velocity separates into two parts: the drift velocity $b^A = m^{AB} \partial_B \eta \phi$ and the osmotic velocity $u^A = m^{AB} \partial_B(-\eta \log \rho^{1/2})$ such that the current velocity is $v^A = b^A + u^A$. In this sense, Φ is something like a "current potential" for the current velocity v^A that tells us how ρ is going to change in time by (7). For the purpose of transparency in this review, Φ eventually becomes the phase of the wavefunction in QM. At this point, Φ's only time dependence is through ρ, but it is important to evaluate what we have been able to derive using ED so far.

ED has managed to show that the the Fokker–Planck equation (7) may be interpreted as a mechanism of *probability updating* due to the maximum entropy considerations, arguments, and equations proposed up to this point. The "current potential" Φ, as argued above, is thus a mechanism or function that guides probability updating, and in this sense, is purely epistemic. To derive QM, we need an additional mechanism for updating the probability of the positions of the particles. We impose $\phi(x) \to \phi(x, t)$ (in (2)) to be a dynamical variable such that Φ has further functionality in its ability to update ρ.

First note that nothing prevents us from recasting (7) as the functional derivative $\partial_t \rho = \frac{\delta \tilde{H}}{\delta \Phi}$, where

$$\tilde{H}[\rho, \Phi] = \int dx \left[\frac{1}{2} \rho m^{AB} \partial_A \Phi \partial_B \Phi + F[\rho] \right] \tag{10}$$

satisfies (7), and has integration functional constant of ρ, named $F[\rho]$. Later $\tilde{H}$ plays the role of a Hamiltonian. At this point the dynamics of ϕ, and consequently Φ, are unknown and we need a natural way to tie down the functional form of the time dependence in ϕ—we suppose that the dynamics of Φ are set such that changes in Φ leads to changes in ρ such that $\frac{d\tilde{H}}{dt} = 0$. This assumption, and its motivation, is the subject of some of the current

research in ED (private communication A. Caticha, S. Ipek). At this point, we claim that "Physics assumptions" like $\frac{d\tilde{H}}{dt} = 0$ are potentially unavoidable constraints if the updating process of ρ is eventually going to describe physics.

Letting the evolution of ρ and Φ obey $\frac{d\tilde{H}}{dt} = 0$ means ρ and Φ are dynamically coupled (Hamilton–Jacobi) equations:

$$\partial_t \rho = \frac{\delta \tilde{H}}{\delta \Phi} \quad \text{and} \quad \partial_t \Phi = -\frac{\delta \tilde{H}}{\delta \rho}. \tag{11}$$

The remaining arguments in ED involve the direct specification of the integration constant $F[\rho]$, which is not pertinent to the remaining content of this article. I suggest the interested reader follow the remaining arguments in [9] that lead to a more complete specification of $\tilde{H}$, which is

$$\tilde{H}[\rho, \Phi] = \int dx \left[\frac{1}{2} \rho m^{AB} \partial_A \Phi \partial_B \Phi + \rho V + \xi m^{AB} \frac{1}{\rho} \partial_A \rho \partial_B \rho \right],$$

where $V(x)$ is a particle potential and $\xi m^{AB} \frac{1}{\rho} \partial_A \rho \partial_B \rho$ is attributed to information geometry. Following [9], nothing prevents combining the solutions ρ and Φ of (11) into a single complex function $\Psi \sim \rho^{1/2} \exp(i\Phi/\hbar)$. After some massaging [9], ED reproduces the *linear* Schrödinger equation (SE),

$$i\hbar \frac{\partial \Psi}{\partial t} = -\sum_n \frac{\hbar^2}{2m_n} \nabla_n^2 \Psi + V\Psi, \tag{12}$$

as an application of inference.

At this point, the standard Hilbert space formalism may be adopted to represent the epistemic state $\Psi(x)$ as a vector,

$$|\Psi\rangle = \int dx \, \Psi(x) |x\rangle \quad \text{with} \quad \Psi(x) = \langle x | \Psi \rangle. \tag{13}$$

The expression of $|\Psi\rangle$ in another basis is regarded as a potentially convenient way of expressing position space wavefunctions.

A more general SE equation, which includes the presence of external nonzero electromagnetic vector potentials $\mathbf{A}$, may also be generated within ED [9]. This is done by applying additional expectation value constraints $\triangle x^a A_a(x_n) = \kappa_n''$ when maximizing (3), similar in form to the drift potential from Eq. (2). Spin is generated by positing the existence of a "spin frame" field $\mathbf{s}(x)$ (motivated through geometric algebra), which again constrains the expected drift, and the Pauli equation for a single particle is found using ED (private communication A. Caticha). The drift potential ϕ, electromagnetic vector potential, and spin frame fields are introduced as epistemic parameters that update the position space distributions of quantum mechanical particles—they are epistemic as they only provide information about how epistemic probability distributions change. The correct spin statistics for identical multi-particle states has yet to be generated from ED, so at this point we impose a symmetrization postulate *ex post facto*, which of course is no better or worse than the standard quantum mechanical formalism.

2.2 Measurement in ED

A natural question is, "If position is the only definite quantity, how are other operators in quantum mechanics measured?". This question was originally addressed in [21] and more recently is addressed in [30] as well as how the notions of von Neuman, weak measurements, and Weak Values [1,13,14] fit into ED. This question is only addressed to the extent "operators" and "measurement" are defined within the entropic dynamics framework. Measurement is a two step process: the state of a system is first updated via a unitary and Schrödinger evolution for

the purpose of *detection*, which is a Bayesian update made due to the presence of data. As ED is an application of inference and probability updating, measurement is simply tackled by applying the appropriate rules of inference. Because position is the only beable in ED, a state vector $|\Psi\rangle$ which is expanded in the basis $|a\rangle$, is a potentially convenient stand-in for position space wavefunctions,

$$|\Psi\rangle = \sum_a c_a |a\rangle = \int \Psi(x)|x\rangle \, dx. \tag{14}$$

An operator such as $\hat{A}$ is an epistemic object, which does not require an ontic existence within the ED framework; however, their values may still be inferred. Operators, Weak Values, and eigenvalues are, therefore, a subset of a type of quantity we call the *inferables* of the theory [30]. As it is the position of particles that formulate the ontic objects in ED, we make inferences about inferables by detecting positions. In most cases, the detecting of position (or presence) of a particle in a detector is itself done by observing (and inferring) the result of a macroscopic amplification within the detector that has been generated by positional presence [21,30].

To accomplish this, [21] introduces the concept of a unitary measurement device $\hat{U}_A |a_i, t\rangle = |x_i, t'\rangle$, which maps states $|a\rangle$ (the state we wish to infer) at time t to a position x_a at a later time t' (presumably on a screen). This allows for the inference of $|a\rangle$ (the eigenvectors of an operator $\hat{A}$) by making detections of x_a at a later time. An example of such a unitary measurement device is a periodic crystal lattice or prism which diffracts "momentum" states to position states. We have,

$$|\Psi', t'\rangle = U_A |\Psi, t\rangle = \sum_a |x_a, t'\rangle\langle a, t|\Psi, t\rangle = \sum_a c_a |x_a, t'\rangle, \tag{15}$$

such that $p(x_a|t') = p(a|t) = |c_a|^2$, that is, the particle may be detected at x_a with probability $p(x_a|t')$ at a later time *as if* it were earlier in the state $|a\rangle$. Inferences can then be made about the operator $\hat{A} = \sum_a \lambda_a |a\rangle\langle a|$ where λ_a are *arbitrary* scalars. The actual detection of the location of a particle in a single experiment is facilitated with another detection device such as a photo-plate, CCD camera, or bubble chamber. In such instances, the probability of x given a detection D is given by,

$$P'(x) = q(x|D) = \frac{q(x, D)}{q(D)} = \frac{p(x)q(D|x)}{q(D)}, \tag{16}$$

where $q(D|x)$ is the likelihood function which accounts for the accuracy of the measurement device. In the present case, after a detection at D,

$$P'(x_a|t') = q(x_a|D, t') = \frac{p(x_a|t')q(D|x_a, t')}{q(D)}, \tag{17}$$

effectively "collapses" the wavefunction, which is to say the final state of the system is known with certainty for sharply peaked likelihood functions in ED [30]. Similarly, if we want to infer the spin of a particle, we preform a von Neumann or a weak measurement to entangle the position and spin of the particle (via unitary evolution as one would with a Stern-Gerlach device) such that by detecting position, we may infer the spin in a similar fashion.

2.3 Remark

Quantum Mechanics has been derived as a peculiar application of epistemic probability updating when the ontic elements of interest are the positions of particles. No further interpretation of Quantum Mechanics in ED is needed. The wavefunction is found to be a useful epistemic quantity for calculating probability distributions, which represent the state of knowledge of a system. Other quantum mechanical objects, like operators and Hilbert spaces, play a supporting role.

Concepts in ED are naturally communicated in the the language of probability. The language generated by the Copenhagen interpretation of quantum mechanics clashes somewhat with the language of probability; for instance, the notion of an "observable" makes little sense when nontraditional Hermitian operators are considered, i.e. does

one ever really claim to "observe" $\hat{p}^n$, $\hat{\rho}$, or one of its eigenvalues? In truth these quantities are inferred through the measurement process no differently than the average energy of a statistical system is inferred by measuring the height of mercury on a thermometer. What was observed in this scenario is the height of the mercury and what is inferred is the temperature and the energy of the system. The word "observable" really looses meaning when one considers "measuring" the Weak Value of an operator $A_W = \frac{\langle \Psi' | \hat{A} | \Psi \rangle}{\langle \Psi' | \Psi \rangle}$, which in general is a complex numbers that may lie outside the eigenvalue spectrum of the operator $\hat{A}$ [1,13,14]. In ED, Weak Values are simply categorized as potentially interesting inferables of the theory as they are inferred from pointer variable (positions in ED) detections [30]. The language used in contextuality proofs does not naturally coincide with the language of probability and inference, which is touched upon later.

3 ψ-epistemic?

In the previous section, we claimed that ψ is an epistemic object which represents our current knowledge of the system in question. This immediately runs into conflict with the ψ-epistemic *no-go* theorem from [28]; however, there is no issue. An excellent review of the ψ-epistemic/ontic dichotomy is presented in [24] and ED would be categorized as a realist (or partial realist) ψ-epistemic model. The first assumption in [28] is (paraphrased) 1) is that a ψ-epistemic state has physical values upon which inferences may be made. ED agrees with this assumption whole heartedly, and the variables which are "physical" in ED are alone the definite yet unknown positions of particles. The second assumption (verbatim) 2) is that "systems which are prepared independently (a) have independent physical states (b)".

The second assumption requires further investigation: first (a)—what does it mean for a system to be prepared independently and second (b)—what does it mean for a system to have independent physical states? The definition of independence in (a) seems to be saturated by the definition of independence in probability theory, namely that if two systems are prepared independently then their joint probability distribution is factorisable into independent probability distributions, $p(x_1, x_2) = p_1(x_1) p_2(x_2)$ and, therefore, there are no correlations between x_1 and x_2 at that time (however evolution may later induce correlations). The quantum mechanical analog is that these two states are non-entangled product states. The definition of independent physical states in (b) is rather unclear from the outset but later is defined quantitatively by,

$$D(\mu_0, \mu_1) = \frac{1}{2} \int |\mu_0(\lambda) - \mu_1(\lambda)| d\lambda, \tag{18}$$

or equivalently in our notation,

$$D[p_1, p_2] = \frac{1}{2} \int |p_1(x_1) - p_2(x_2)| \delta(x_1 - x_2) dx_1 dx_2, \tag{19}$$

such that if $D = 1$ then p_1 and p_2 are completely disjoint, and thus occupy (in their words) "independent physical states". It is easy to see now how the definition of independence in (a) differs from the definition of independence in (b), in-fact the (a) and (b) definitions clash in assumption 2) for any independent joint probability $p(x_1, x_2) = p_1(x_1) p_2(x_2)$ (a) which are not entirely disjoint (i.e. "physically" independent (b))—that is, if $p_1(x_1)$ and $p_2(x_2)$ overlap in $\mathcal{X}$. This regularly occurs in noninteracting multiparticle states in Quantum and Statistical Mechanics (can be obtained by marginalizing over the momentum states of a phase-space probability distribution $\rho(x, p)$). In ED, the "physicality" of particle positions is independent of the state of knowledge at hand. This is because probability is not a measure of physicality (or onticity) but rather as a degree of belief or plausibility [8,12,20] that the proposition "the particle is located at x" is true. As the leading assumptions of what entails a ψ-epistemic state differ, the ψ-epistemic *no-go* theorem does not apply, which is admitted as a possible exemption to their *no-go* theorem in the conclusion of [28]. We are, therefore, justified in treating ψ epistemically.

4 Hidden variables, realism, and non-locality

The subject of hidden variables, realism, and non-locality in ED has been touched upon in [8,21] and it will be further explored here. In Bell's landmark paper [4], he found a contradiction between QM and hidden variable theories which claimed local realism. It was accomplished by considering a hidden variable λ, which if known, would give the outcome of an experiment (an eigenvalue of an operator) with certainty $a_0 = A(\lambda = \lambda_0)$. By integrating over the probability of a hidden variable,

$$\langle A(\lambda)B(\lambda)\rangle = \int p(\lambda)A(\lambda)B(\lambda)\,\mathrm{d}\lambda, \tag{20}$$

he showed that such expectation values do not always agree with the expectations values of QM, for general $p(\lambda)$.

In ED there is no such hidden variable. The particle dynamics are non-deterministic as can be seen by the Brownian like paths particles take due to the form of the transition probability $P(x'|x)$ in (5), or after energy conservation, that the particles are undergoing a non-dissipative diffusion. Even if the initial conditions of particle positions are known exactly (or with near perfect precision), $\rho(x) \sim \delta(x - x_0)$, Eq. (6) is inevitably nondeterministic for time steps Δt. Because the Brownian paths of particles are non-differentiable, other equi-temporal quantities (e.g. momentum or energy) are simultaneously indefinite, which is another argument against position being a hidden variable. The process which is deterministic in ED is the evolution of the probability distribution as it follows the HJ-like equations from (11) given the appropriate constraints, boundary, and initial conditions are known. The drift potential $\phi(x) \sim \Phi$ updates the probability distribution of particle locations rather than guiding each particle at every point, again seen by inspecting (11) and (5). The solution is to realize that $\Phi(x)$ is an epistemic parameter that is coupled to $\rho(x)$ for each x in a complicated way through the HJ-like Eq. (11). The nonlocal nature of probability as a means for quantifying knowledge (of the future, past, or present) accounts for the nonlocal behavior of QM in ED. As any collapse is an epistemic change in the system, each observer assigns distributions that coincide with their current state of knowledge of the system (i.e. Alice and Bob may preform partial traces and the like).

5 BKS type theorems

The BKS theorem sheds light on the incompatibility of hidden variable theories and quantum mechanics [5,23]. Years later Mermin demonstrated what is considered to be the simplest expression of what is usually an algebra and geometry intensive BKS theorem [25]. BKS proofs have been generalized to the N-qubit Pauli group [31] and [27] gives a BKS proof using continuous position and momentum observables. In [31], they give a simple algorithm to convert observable based BKS proofs to a large number of projector based BKS proofs, so we will focus on the simpler observable based proofs.

The class of hidden variable theories the BKS theorem excludes have the following reasonable conditions: The value of an operator is *definite* yet *unknown* such that we may assign it a preexisting value (its eigenvalue) called its valuation [3,25,31]. The valuation of an operator $\hat{A}$ at any time is then,

$$v(\hat{A}) = \langle a|\hat{A}|a\rangle = a. \tag{21}$$

It is also assumed that functional relationships between operators $f(\hat{A}, \hat{B}, \hat{C}, \ldots) = 0$ hold throughout the valuation process,

$$v(f(\hat{A}, \hat{B}, \hat{C}, \ldots)) = f(v(\hat{A}), v(\hat{B}), v(\hat{C}), \ldots) = 0. \tag{22}$$

It should be noted that the considered operators must commute $v(\hat{A}\hat{B}) = v(\hat{A})v(\hat{B}) = v(\hat{B}\hat{A})$ when taking valuations for (22) to hold. Mermin demonstrates the contradiction of equations (21) and (22) with quantum mechanics by considering what is now know as the Peres–Mermin square:

ZI	IX	ZX
IZ	XI	XZ
ZZ	XX	YY

.

Each table entry is an observable from the 2-qubit Pauli group consisting of a joint eigenbasis consisting of 4 eigenvectors. As a notational convenience we will omit tensor products when there is no room for confusion and let $X = \sigma_x$ such that an arbitrary table entry $XZIY$ represents $\sigma_x^{(1)} \otimes \sigma_z^{(2)} \otimes I^{(3)} \otimes \sigma_y^{(4)}$, following the notational structure in [31]. The product of the operators along a given row or column is the rank 4 identity $I(4) = II$ (in this notation) with the exception of the last row, which is $-II$. Consider the valuation of the standard matrix product of the elements of the first row,

$$v(ZI \cdot IX \cdot ZX) = v(II) = 1. \tag{23}$$

Supposing (22) is true then

$$v(ZI \cdot IX \cdot ZX) = v(ZI)v(IX)v(ZX) = 1. \tag{24}$$

The valuation of the ijth element A^{ij} in the table is $v(A^{ij}) = \pm 1$ and, therefore, (22) imposes a constraint on the individual valuations $v(ZI)v(IX)v(ZX) = 1$, which is only satisfied if either 0 or 2 of the valuations are -1. This cuts the universe of discourse from $2^3 = 8$ possibilities down to 4. Let $A^{i\odot}$ be the product of the operators in the ith row and $A^{\odot j}$ the product of the operators in the jth column such that above $A^{1\odot} = ZI \cdot IX \cdot ZX$ is the standard matrix product between the listed operators. Mermin showed his square indeed leads to a contradiction when considering the product of the row and column valuations,

$$\prod_i v(A^{i\odot})v(A^{\odot i}) = v(II)^5 v(-II) = -1, \tag{25}$$

whereas applying (22) to each row and column, $v(A^{i\odot}) = \prod_j v(A^{ij})$, gives,

$$\prod_i v(A^{i\odot})v(A^{\odot i}) \rightarrow \prod_i \prod_j v(A^{ij})^2 = 1, \tag{26}$$

which is a contradiction. This is due to the fact that not all of the elements in Mermin's square commute and, therefore, all observables cannot be assigned definite eigenvalues. Quantum mechanical formalism and experiment agrees with (25) and not with (26), and thus (22) must be thrown out. Bell makes a point that it may be overconstraining for the valuation to produce identical values when different sets of commuting observables are being considered, just to refute it by noting that a space-like separated observer could change which set of commuting observables he/she wishes to measure mid-flight. A hidden variable theory would then have to explain this nonlocal change in the valuation meaning that the BKS theory only refutes local hidden variables theories.

5.1 Interpreting the contradiction: contextuality

The standard interpretation of the contradiction by Bell, Kochen, Specker, Mermin and others is that quantum mechanical observables are *contextual*, meaning that the operator's "aspect", "character", or "value" depend on the remaining set of commuting observables under which it is considered. Any observable which does not depend on the remaining set of commuting observables in this way is called *noncontextual*, which, for example, are the individual observables $v(A^{ij})$ from the Mermin square and (26).

In more recent years the interpretation of the BKS theorem, which in principle would rule out all local hidden variable theories obeying (21) and (22), has been under scrutiny, in essence, for having a more restrictive interpretation than the theorem merits. The work by [10,22,26] opens a loophole due to the impracticality of infinite measurement precision, and thus the BKS theorem is "nullified" in their language. Appleby (and others) find the "nullified" critique to be too harsh of a criticism [2]. De Ronde [29] points out that epistemic and ontic contextualty are consistently being scrambled into a omelet when perhaps the yoke and egg whites should be cooked separately. He defines ontic contextuality as the formal algebraic inconsistency of the operator and valuation formalism of Quantum Mechanics within the BKS theorem—having nothing to do with measurement. The epistemic counterpart is more aligned with the principles of Bohr in that Quantum Mechanics involves an interaction between system

and measurement apparatus whose outcomes are inevitably communicated in classical terms—the context is given by the measurement device. The difference is subtle but, as noted, ontic contextuality is defined to be independent of the differing interpretations of quantum mechanics where epistemic contextuality need not be. Our treatment of contextuality does separate in this fashion; however, de Ronde's usage of the word "ontic" refers to the quantum formalism, whereas our usage only refers to ontic particle positions in ED.

5.2 Critiques on representing onticity with valuations in QM

As we know, the assumptions (21) and (22) lead to contradictions. Inevitably (21) and (22) will be illogical on several levels, some of which are discussed below. The main critique we present is, how do we know that the valuation of an observable $v(\hat{A})$ accurately represents the notion of *definite, preexisting values* of an operator, that would be obtained if a measurement is carried out? The alleged strength of the BKS theorem is that analysis has been done independent of the particular state $|\Psi\rangle$, and thus it should hold for all $|\Psi\rangle$ in general. This is troubling for a number of reasons, the first being that a particular $|\Psi\rangle$ may not have components along every eigenvector of an operator $\hat{A}$, in which case a zero probability event could be assigned a definite existence, and one would never know because $|\Psi\rangle$, which all of the observables in question pertain to, has not been specified. This issue here is an interplay between the ontic and epistemic contextuality given by de Ronde, because only sensible valuations may be given if the state of the system is known—in general the density matrix $\hat{\rho}$.

If the valuation process is to be applicable to arbitrary "observables" independent of the state at hand, then one runs into another logical inconsistency when attempting to apply valuations to a density matrix, $\hat{\rho}$, because it represents the probabalistic state of a system. It makes little sense to have different sets of commuting observables $\{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3 \ldots\}$ which are required to span to the same Hilbert space as the state in question $|\Psi\rangle$ (or $\hat{\rho}$). Furthermore, the valuation of a density matrix $\hat{\rho} = \sum_i p_i |i\rangle\langle i|$ gives one of its eigenvalues, p_i, which are probabilities themselves and are never *directly* observed, but are usually inferred from the frequency of a large number of independent trials. One cannot possibly claim that a system is ontically expressing a definite preexisting probability value p_i. Probability by its nature is a measure of the indeterminance of a state $|i\rangle$ rather than a value (physically) carried by the state $|i\rangle$—which is as epistemic as it gets! If Alice knowingly prepares one system and Bob does not know which system Alice has prepared, then it is clear that p_i's cannot have a definite existence because both Alice and Bob disagree about said values over the same single "ontic" system of interest. Furthermore, when the a measurement is made to determine the state, the probability value updates (the eigenvalue changes) and in this sense the assignment of an eigenvalue $\hat{\rho}$ through valuation represents nothing physical about the state of the system's *definite, preexisting values* that would in principle be obtained if a measurement was carried out. In this sense, the eigenvalues of operators in general do not represent *definite, preexisting* (noncontextual) *values* of an operator.

Due to these critiques, and that in ED one may infer eigenvalues from position detections, it is difficult to know what precisely a valuation procedure represents meta-physically (linguistically), besides the simple choice of a matrix element. As discussed, the valuation of an operator may not always represent an ontic value of an observable and, therefore, we suggest relaxing this notion and replacing it by the more general statement, "The valuation of an operator (or set of operators) represents a quantity that in principle may be inferred", or in the language of [30], "The valuation of an observable is an *inferable* of the theory".

5.3 Hybrid-contextual theories

It should be noted that in Entropic Dynamics, the idea of valuation is very unnatural. An inference based theory allows us to state, quantify, and represent how much we do *not* know about the state of a system through a probability distribution, upon which we use the rules of inference and probability updating to determine what we do. Recall in Sect. 2.2 that the measurement procedure in ED allows for the inference of $\hat{A} = \sum_a \lambda_a |a\rangle\langle a|$ where λ_a are arbitrary scalars by making detections of position at a later time. Eigenvalues themselves are an afterthought of the inference process that are epistemically inferable parameters by the changes they make to (position) probability distributions.

Strictly speaking, the BKS theorem discards realist theories in which all of the considered operators are treated ontically through their valuation. This leaves open the possibility for a hybrid-contextual theory in which only a subset of commuting observables are *definite yet unknown*, or noncontextual, while other variables (or sets of commuting observables) are contextual. To date the only theory of Quantum Mechanics known to the author that seems to fit this description precisely is entropic dynamics [9].

The only operators required to undergo valuation in ED are the $3N$-particle position coordinates with their corresponding $3N$ operators $\hat{X}^{(n)}$. In the language of valuations, we would have,

$$v_x(\hat{x}_i^{(n)}) \equiv \langle x_i^{(n)}|\hat{x}_i^{(n)}|x_i^{(n)}\rangle = x_i^{(n)}, \tag{27}$$

for a particular coordinate $x_i^{(n)}$. Position operators trivially obey (22),

$$v_x(f(\hat{x}_i^{(n)}, \hat{x}_j^{(m)}, \ldots)) = f(v(\hat{x}_i^{(n)}), v(\hat{x}_j^{(m)}), \ldots) = 0, \tag{28}$$

for any function f, because all position operators mutually commute. No parity contradiction in the sense of [25,31] can be reached because all of the operators requiring valuation mutually commute. The BKS proofs are proofs by contradiction. This means a set of counter examples has been found which rule out the general applicability of assigning definite yet unknown values to all operators all the time; however, as seen above, there are instances in which there is no contraction and the assignment of definite yet unknown values in this instance is consequently *not* ruled out.

Operators other than position, A^{ij}, need not be noncontextual in ED as they are considered to be epistemic in nature. In this case, one should not claim A^{ij}, one of its eigenvalue a^{ij}, or a state $|a^{ij}\rangle = \int dx\, \psi_{a^{ij}}(x)|x\rangle$, to have a definite existence outside of characterizing our knowledge of the definite yet unknown positions of particles x. That being said, when one can expand A^{ij} in the position basis, we find that the A^{ij} are naturally contextual—although in principle this is unwarranted in ED as no valuation is required.

As it is the position that is definite, the valuation of the operator A^{ij}, before measurement (where $|x\rangle = |x_1\rangle \otimes |x_2\rangle \cdots \otimes |x_N\rangle$ for N particles), is one of the diagonal matrix elements in the x basis,

$$v_x(A^{ij}) \rightarrow \langle x_0|A^{ij}|x_0\rangle = \left\langle x_0\left|\sum_n |a_n^{ij}\rangle a_n^{ij}\right.\right\rangle a_n^{ij} x_0 = \sum_n a_n^{ij}|\langle x_0|a_n^{ij}\rangle|^2 \neq a^{ij}, \tag{29}$$

where in this case it is supposed that the definite yet unknown value of x is x_0. This is obviously not one of the eigenvalues or "observables" of A^{ij}, but in ED A^{ij} is an inferable and so is $v_x(A^{ij})$. The position space valuation $v_x(A^{ij})$ is some real number which in principle may be assigned to any position coordinate. In general, parity type proofs of the BKS theorem require A^{ij} to be simultaneously part of an even number of sets of commuting observables [31]. This means an operator A^{ij} is simultaneously diagonalized in (at-least two) different basis,

$$A^{ij} = \sum_n |a_n^{i\odot}\rangle a_n^{ij} \langle a_n^{i\odot}| = \sum_n |a_n^{\odot j}\rangle a_n^{ij} \langle a_n^{\odot j}|, \tag{30}$$

where, for example in the Peres–Mermin square, $|a_n^{i\odot}\rangle$ refers to the eigenvectors of the commuting set of variables from the ith row and $|a_n^{\odot j}\rangle$ refers to the eigenvectors of the commuting set of variables from the jth column. The largest number of distinct sets of eigenvectors is equal to the number of sets of commuting observables in the BKS proof. Using this notation we may denote the product of the operators in a commuting set by,

$$A^{i\odot} = \sum_n |a_n^{i\odot}\rangle a_n^{i1} a_n^{i2} \cdots a_n^{iN} \langle a_n^{i\odot}| = \sum_n |a_n^{i\odot}\rangle a_n^{i\odot} \langle a_n^{i\odot}|, \tag{31}$$

where N is the number of operators in the commuting set of observables. In general, the application of (22) to the position valuations of $\{A^{ij}\}$ will not hold,

$$v_x(A^{i\odot}) \rightarrow \prod_j v_x(A^{ij}), \tag{32}$$

because it would require,

$$\sum_n |\langle x_0|a_n^{i\odot}\rangle|^2 a_n^{i\odot} \rightarrow \prod_j \left(\sum_n a_n^{ij} |\langle x_0|a_n^{i\odot}\rangle|^2\right)_j, \tag{33}$$

which is potentially equal, but in the vast majority of cases is not. The lack of equality can be seen if one considers three commuting momentum observables $\hat{p}_1 \otimes \hat{1}_2$, $\hat{1}_1 \otimes \hat{p}_2$, and $\hat{p}_1 \otimes \hat{p}_2$ with $\{|a_n^{i\odot}\rangle\} = \{|p_1, p_2\rangle\}$—the LHS diverges while the RHS is zero because it involves products of odd integrals. This poses no issue in ED because $A^{i\odot}$ or the individual A^{ij} need only exist epistemically, so their valuations (matrix elements) need not agree—the product of matrix elements need not be the matrix element of the product so imposing equality is nonsensical. Furthermore (if above was not enough), contextuality is preserved among non-position observables (for noncontextual position) as can be seen when (22) is applied to the product of all of the commuting sets of observables,

$$\prod_i v_x(A^{i\odot}) v(A^{\odot i}) \rightarrow \prod_i \prod_j v_x(A^{ij})^2 \geq 0, \tag{34}$$

for situations when the LHS is less than zero or it is simply not equal to the RHS. This calculation shows that definite (noncontextual) positions before measurement do not imply definite (noncontextual) A^{ij} and, therefore, we are justified in treating the operators A^{ij} contextually—which means we should not apply valuations to them, or if we do, we should not expect (22) to hold. This suggests that we may not expect equations like (22) to hold true in general because, if valuations interpreted as *inferables* (Sect. 5.2), then expecting something like $v(A^2) = v(A)^2$ to hold true is potentially analogous to expecting expectation values like $\langle A^2 \rangle = \langle A \rangle^2$ to hold true, which of-course is not true in general. Spin in ED is not required to be noncontextual so valuation is not required. The current form of ED would potentially be ruled out if A^{ij} were noncontextual under position valuations in general—but this is not the case as the matrix elements are simply epistemic inferables.

Because all position operators always mutually commute with one another and, therefore, are all simultaneously diagonalizable in the same set of position eigenvectors (i.e. $|x\rangle = |x^{i\odot}\rangle = |x^{\odot j}\rangle = |x^{\odot\odot}\rangle$), they may be treated noncontextually together. If an operator is a product of contextual and noncontextual operators, it remains contextual because applying position space valuations on the noncontextual operators leaves the contextual operators contextual. This can be seen by applying position space valuations to the continuous operators defined in [27] ($|\langle S_{QM} \rangle| = 6$). As noted in the critiques, the valuation of an operator may not always express the definite yet unknown values of an observable—it may be best to relax this notion such that the valuation of an operator represents a quantity that in principle may be inferred, an inferable, in general.

A question of interest is, how, if everything is to be measured or inferred using a (non-contextual) position basis (Sect. 2.2 and [21,30]), is the contextual nature of a set of contextual operators $\hat{A}^{ij} \in \mathcal{A}$ non-contradictory? This question is especially tricky because it mixes the epistemic and ontic palates of contextuality in the sense of [29], who, as well as [6], quote Mermin , "the whole point of an experimental test of BKS [theorem] misses the point.". That being said, the contexuality of the operators A^{ij} is simply expressed through the lack of commutativity between sets of commuting observables, we do not need to do position valuations $\{A^{ij}\}$ to make inferences about the states. ED perhaps sheds some light onto Mermin's statement about the lack of an experimental test.

Suppose Alice prepares a two particle system and sends it to Bob who has a compound unitary measurement device (15) for each set of commuting observables (each row and column) of the Mermin square (for simplicity), but really this is applicable to any construction of sets of commuting observables. Because Bob can only measure one row or column for a given pair of particles sent from Alice, him choosing a measurement device from the ith row or column means he has chosen and applied the unitary measurement device $\hat{U}_{A^{i\odot}}$ to the incoming state and mapped it to position coordinates for detection and inference. That is, the physical application of $\hat{U}_{A^{i\odot}}$ picks, P^i, the ith set of commuting observables,

$$P^i * \begin{array}{|c|c|c|} \hline A^{11} & A^{12} & A^{13} \\ \hline A^{21} & A^{22} & A^{23} \\ \hline A^{31} & A^{32} & A^{33} \\ \hline \end{array} \longrightarrow \hat{U}_{A^{i\odot}} \begin{array}{|c|c|c|} \hline A^{i1} & A^{i2} & A^{i3} \\ \hline \end{array}$$

$$\longrightarrow \begin{array}{|c|c|c|} \hline x^{i1} & x^{i2} & x^{i3} \\ \hline \end{array},$$

and at a later time one may apply valuation(s) to the associated positions operators if one wishes, because the operator is position based (at that later time) $A^{ij} \rightarrow \sum_n |x_n^{ij}\rangle a_n^{ij} \langle x_n^{ij}|$. The notion of detectors picking sets of commuting observables is mentioned in [25], but here the process is specified to show how contextuality is preserved. The positions may be detected and the associated commuting set of A^{ij} may be inferred.

This process resembles the epistemic notion of contextuality presented in [29]. The operators A^{ij} are treated contextually—the position space valuation may be applied after the set of commuting observables has been chosen by the unitary measurement device. As only one set of commuting observables may be picked at a time by Bob, the quantum mechanical expectation values match that as read by Bob (and are, therefore, in the form of (25)). Alice, being in the dark, does not know which row Bob will pick and is free to assign a probability Bob picks the ith row or column, and after learning the chosen row or column may she update her probability accordingly.

6 Discussion

The most natural inferential tool in ED is probability. The critiques given in Sect. 5.2 are further motivation for the use of probability to make rational inferences, while the interpretation of valuation functions, which inevitably lead to contradictions in the BKS theorem, is not generally applicable. There, reason was given for the need of a more general interpretation of the valuation of an operator, which was stated, "The valuation of an operator (or set of operators) represents a quantity that in principle may be inferred". Because the probability of a state is only defined in terms of its set of commuting observables, and because there is no way to generate a unique joint probability distributions among non-commuting observables [11], a rational discussion on the potential simultaneous onticity between non-commuting observables is not possible—luckily ED formulates QM by assuming the position of particles to be the only ontic variables.

ED is in a unique position among foundational theories of QM because QM was derived by applying standard probability techniques to a system of particles with ontic positions. This naturally classifies the ontic and epistemic elements of QM and provides a clean cut interpretation of QM such that physical and conceptual problems in QM may be handled rigorously (as it has in [21,30] and other recent articles). The hope is that a full treatment of spin in ED will provide better notions of the symmetrization postulate and the Pauli exclusion principle, but this problem has yet to be tackled in full. At this point more work can be done to make sure that ED is able to reproduce all of the known results of QM and to hopefully shine some light on the many interpretational paradoxes, *no-go* theorems, and problems that surround QM. The end goal of ED is to show inferential origins of physical laws. This generates new interpretations and directions for old ideas, and hopefully, ED will generate some new physics as well.

This paper shows the sense in which a foundational QM theory may be hybrid-contextual, i.e. one set of commuting operators is noncontextual (ontic) while all others are contextual, and still obey the BKS theorem. In ED this occurs because contextual operators are not required to have definite ontological existence outside of their characterization of the state of knowledge of the noncontextual operators. A loose guide for a theory to be hybrid-contextual, and also agree with QM, is that its ontic variables are treated noncontextually while its epistemic variables are treated contextually. The values of interest associated with contextual operators (energy, momentum, and spin) are inferred by measurement of noncontextual observables (position in ED).

References

1. Aharonov, Y., Albert, D.Z., Vaidman, L.: How the result of a measurement of a component of the spin of a spin- 2 particle can turn out to be 100. Phys. Rev. Lett. **60**, 1351–1354 (1988)
2. Appleby, D.M.: The Bell–Kochen–Specker theorem. arXiv:quant-ph/0308114v1 (2003)
3. Ballentine, L.E.: Quantum Mechanics a Modern Development, 2nd edn. World Scientific Publishing Co. Pte. Ltd., Singapore (2014)
4. Bell, J.: On the einstein podolsky rosen paradox. Physics **1**(3), 195–200 (1964)
5. Bell, J.: On the problem of hidden variables in quantum mechanics. Rev. Mod. Phys. **38**, 447–452 (1966)
6. Cabello, A.: Experimentally testable state-independent quantum contextuality. Nat. Phys. **80**(3), 475–478 (1998)
7. Carrara, N., Caticha, A.: Quantum phases in entropic dynamics. In: MaxEnt 2017, the 37th international workshop on Bayesian inference and maximum entropy methods in science and engineering (2017)
8. Caticha, A.: Entropic Inference and the Foundations of Physics. (monograph commissioned by the 11th Brazilian Meeting on Bayesian Statistics—EBEB-2012) (2012). http://www.albany.edu/physics/ACaticha-EIFP-book.pdf
9. Caticha, A.: Entropic dynamics. Entropy **17**, 6110–6128 (2015)
10. Clifton, R., Kent, A.: Simulating quantum mechanics by non-contextual hidden variables. Proc. R. Soc. Lond. A **546**, 2101 (2000)
11. Cohen, L.: Frontiers of Nonequilibrium Statistical Physics, Series B: Physics, vol. 135, pp. 97–117. Springer, New York (1986)
12. Cox, R.: Probability, frequency and reasonable expectation. Am. J. Phys. **14**, 1–13 (1946)
13. Dressel, J., Malik, M., Miatto, F., Jordan, A., Boyd, R.: Colloquium: understanding quantum weak values: Basics and applications. Rev. Mod. Phys. **86**, 307–316 (2014)
14. Duck, I.M., Stevenson, P.M., Sudarshan, E.C.G.: The sense in which a 'weak measurement' of a spin 1/2 particle's spin component yields a value 100. Phys. Rev. D **40**, 2112–2117 (1989)
15. Giffin, A.: Maximum entropy: The universal method for inference. Ph.D. thesis, University at Albany (SUNY) (2009)
16. Giffin, A., Caticha, A.: Updating probabilities with data and moments. In: AIP conference proceedings, p. 74 (2007)
17. Ipek, S., Caticha, A.: Entropic quantization of scalar fields. In: MaxEnt 2014, the 34th international workshop on Bayesian inference and maximum entropy methods in science and engineering. AIP Publishing, Waterloo (2014)
18. Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**, 620 (1957)
19. Jaynes, E.: Information theory and statistical mechanics II. Phys. Rev. **108**, 171 (1957)
20. Jaynes, E.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge (2003)
21. Johnson, D., Caticha, A.: Entropic dynamics and the quantum measurement problem. In: Bayesian inference and maximum entropy methods in science and engineering: 31st international workshop. AIP Publishing, Waterloo, Canada (2011)
22. Kent, A.: Noncontextual hidden variables and physical measurements. Phys. Rev. Lett. **83**, 3755 (1999)
23. Kochen, S., Specker, E.: The problem of hidden variables in quantum mechanics. J. Math. Mech. **17**, 59–87 (1967)
24. Leifer, M.: Is the quantum state real? an extended review of ψ-ontology theorems. Quanta **3**, 67–155 (2014)
25. Mermin, N.: Hidden variables and the two theorems of John Bell. Rev. Mod. Phys. **65**(3), 803 (1993)
26. Meyer, D.A.: A simplified basis for the Bell–Kochen–Specker theorems. Phys. Rev. Lett. **83**(19), 3751 (1999)
27. Plastino, A., Cabello, A.: State-independent quantum contextuality for continuous variables. Phys. Rev. A **82**, 022114 (2010)
28. Pusey, M., Barret, J., Rudolph, T.: On the reality of the quantum state. Nat. Phys. **8**(3), 475–478 (2012)
29. de Ronde, C.: Unscrambling the quantum omelette of epistemic and ontic contextuality (part 1): preexistent properties or measurement outcomes?. **14** (2016). arXiv:1606.03967v3
30. Vanslette, K., Caticha, A.: Quantum measurement and weak values in entropic dynamics. In: MaxEnt 2016, the 36th international workshop on Bayesian inference and maximum entropy methods in science and engineering. AIP Publishing, Waterloo (2016)
31. Waegell, M., Aravind, P.K.: Proofs of the Kochen-Specker theorem based on the n-qubit pauli group. Phys. Rev. A **88**, 012102 (2013)