Quantum AIXI: Universal Intelligence via Quantum Information

Elija Perrier[1][0000−0002−6052−6798]

Centre for Quantum Software & Information, UTS and Gradient Institute, Sydney
elija.perrier@gmail.com

Abstract. AIXI is a widely studied model of artificial general intelligence (AGI) based upon principles of induction and reinforcement learning. However, AIXI is fundamentally classical in nature - as are the environments in which it is modelled. Given the universe is quantum mechanical in nature and the exponential overhead required to simulate quantum mechanical systems classically, the question arises as to whether there are quantum mechanical analogues of AIXI which are theoretically consistent or practically feasible as models of universal intelligence. To address this question, we extend the framework to quantum information and present Quantum AIXI (QAIXI). We introduce a model of quantum agent/environment interaction based upon quantum and classical registers and channels, showing how quantum AIXI agents may take both classical and quantum actions. We formulate the key components of AIXI in quantum information terms, extending previous research on quantum Kolmogorov complexity and a QAIXI value function. We discuss conditions and limitations upon quantum Solomonoff induction and show how contextuality fundamentally affects QAIXI models.

Keywords: Quantum · AIXI · Universal Intelligence · Complexity.

1 Introduction

A cornerstone in the theoretical landscape of AGI is the AIXI model [33, 34], which provides a mathematically rigorous and complete framework for an optimal Bayesian reinforcement learning agent. AIXI's optimality is rooted in Solomonoff's theory of universal induction [50] and classical computability theory. However, AIXI and its underlying assumptions are fundamentally classical. Given that our universe is governed by quantum mechanics, a fundamental question arises: what constitutes universal intelligence in a quantum mechanical world? This paper addresses this question by developing the mathematical foundations for Quantum AIXI (QAIXI), an AGI model operating within the principles of quantum mechanics. Using the formal theory of quantum information processing, we extend prior work on quantum intelligence [15, 44] and situate QAIXI as an agent anchored in both quantum and classical registers interacting with the environment via quantum and classical channels. Using this formulation, we contribute the following:

(a) Describing the QAIXI agent's interaction loop using density operators and quantum measurement theory along with specifying quantum Kolmogorov complexity (K_Q) [10, 53] in terms of quantum information channels and its role in a universal quantum prior.
(b) Utilising the theory of quantum-to-quantum (QTQ) and quantum-to-classical channels [56] to describe the space of possible quantum environments.
(c) Formulating a quantum equivalent of the universal Bayesian mixture over quantum environments while adapting a quantum Bellman equation and the QAIXI policy based on quantum formalism.

(d) Describing circumstances in which the Kochen-Specker theorem [38] renders any QAIXI history fundamentally contextual.

As we discuss below, QAIXI is impractical to implement. Our contribution therefore aims to contribute to debate over the theoretical models of optimal intelligence that account for the fundamentally quantum nature of the universe, particularly limitations stemming from the inherent properties of quantum state spaces, quantum computation, and quantum measurement.

1.1 Related Work

AIXI [33–35] is among the leading AGI proposals at the centre of debates over AGI, as both a model of universal intelligence and a proposal against which other proposals are assessed. It takes Solomonoff's universal induction, folds it into sequential decision theory, and produces what is arguably the cleanest statement of an unbounded optimal agent. Subsequent symbolic, connectionist, and hybrid AGI proposals typically position themselves by (i) trying to approximate AIXI in practice (e.g. AIXItl, MC-AIXI, information-geometric variants) or (ii) critiquing AIXI's reliance on classical assumptions and incomputable quantities. Logicist programmes [9], emergentist lines [51], distributed cognitive architectures [26,27], and self-organising approaches [41] all share AIXI's basic agent–environment model, albeit while extending, varying or removing its component Bayesian machinery. AIXI has been subject to critical review over the last two decades on a number of grounds: (1) physical unrealisability or super-Turing requirements [39,55]; (2) the Cartesian boundary between software and hardware [7,8]; (3) inadequate treatment of resource constraints, counterfactual reasoning or multi-agent reflection [23,49]; and (4) challenges to the Kolmogorov-based prior itself [52]. Even so, AIXI remains a cornerstone of classical AGI (CAGI) theories. The advent of quantum information technologies (QIT) [2,43] and interest in their computational capabilities and quantum forms of AGI (QAGI) poses a natural question for AIXI models: if AIXI purports to be a universal agent in a *classical* world, what replaces it in a *quantum* universe? Most quantum-AI research to date is narrowly cast, focusing on specific technical features such as quantum decision theory, quantum machine learning [11, 16, 21, 45, 48], quantum reinforcement learning [20, 36, 42], and hybrid variational schemes [46, 47]. These propose access to quantum algorithms providing prospective quantum advantage (e.g. amplitude-estimation–based policy evaluation [40]), albeit with ongoing debate about their practical reach [3].

Attempts to synthesise AIXI with quantum mechanics (and quantum information processing) have been relatively limited. Work on Solomonoff induction [50] and quantum computation (e.g. Deutsch's quantum Turing machines [18]) have considered component features of the AIXI program, including substituting out specific classical components for quantum analogues (e.g. replacing classical with quantum Grover search [15]). Yet quantum mechanics operationally - and ontologically - carries profound differences from classical ontology (and computation). Quantum systems may subsist in superposition states, be entangled and lack definitive identity until measurement interaction. Moreover, any integration of quantum mechanics and AIXI must also reckon with seminal results in quantum foundations: contextuality, non-locality, and quantum measurement — all absent from classical AIXI. Work on quantum causality [14,57], algorithmic thermodynamics [22], and many-worlds decision theory [54] offer contributions on foundational quantum mechanical issues facing quantum AIXI-style agents whose quantum operations disturb the very environment they probe. At present, however, there is neither a consensus quantum prior (the role played by $2^{-K(\nu)}$ in AIXI) nor a tractable complexity class in which QAIXI could be approximated.

2   Classical AIXI

The AIXI agent [33] is a theoretical model for a universally optimal reinforcement learning agent. It interacts with an unknown environment μ in cycles. In each cycle t, the agent chooses an action $a_t \in \mathcal{A}$ and receives a percept $e_t = (o_t, r_t) \in \mathcal{O} \times [0, 1]$, consisting of an observation o_t and a reward r_t. The history is $h_{<t} = a_1 e_1 \ldots a_{t-1} e_{t-1}$. AIXI's policy π^{AIXI} aims to maximise future expected rewards:

$$a_t := \pi^{AIXI}(h_{<t}) = \arg\max_{a_t} \sum_{e_t} \cdots \sum_{e_m} \left(\sum_{k=t}^{m} \gamma^{k-t} r_k \right) \sum_{\nu \in \mathcal{M}_U} w_\nu \nu(e_t, \ldots, e_m | h_{<t} a_t \ldots a_m), \quad (1)$$

where m is the lifetime of the agent, $\gamma \in [0, 1)$ is a discount factor, and the inner sum is over all environments ν in a universal class $\mathcal{M}_U$ (e.g., all chronological semi-computable environments $\mathcal{M}_{sol}$), weighted by $w_\nu = 2^{-K(\nu)}$. $K(\nu)$ is the Kolmogorov complexity of the classical Turing Machine (CTM) describing environment ν. The sum over environments constitutes Solomonoff's universal prior ξ_U:

$$\xi_U(e_{1:m} || a_{1:m}) := \sum_{\nu \in \mathcal{M}_{sol}} 2^{-K(\nu)} \nu(e_{1:m} || a_{1:m}). \quad (2)$$

AIXI is Pareto-optimal and self-optimising in the sense of [33]. However, it is incomputable due to the use of $K(\cdot)$ and the sum over all CTMs. Its ontological assumptions are classical: deterministic or classically stochastic environments, objective histories, and classical information theory.

3   Quantum Computational Foundations

3.1   Quantum information

To construct QAIXI, we must replace classical computational notions with their quantum counterparts. However, this is not a straightforward isomorphic mapping. The unique properties of quantum mechanics problematise conventional assumptions underlying classical AIXI, such as agent identity, definitiveness of state and the separability and distinguishability of an agent from its environment. To formulate QAIXI in a way that accommodates and caters for these ontological differences, we turn to the language and formalism of quantum information theory [56]. In QIP, *systems* (agents) and *environments* are described in terms of *registers* X (e.g. bits) comprising information drawn from a classical alphabet Σ. Registers may be in either classical or quantum states. We define a Hilbert space $\mathcal{X} = \mathbb{C}^{|\Sigma|}$ with computational basis $\{|s\rangle\}_{s \in \Sigma}$. A *quantum state* of a register X (associated with space $\mathcal{X}$) is a density operator $\rho \in \mathcal{D}(\mathcal{X})$, i.e., a positive semi-definite operator with $\text{Tr}(\rho) = 1$. Classical states are then ρ which are diagonal in the basis of $\mathcal{X}$. Interactions with (and changes to) states occur via *channels* which are superoperators. They define how quantum and classical states interact. Classical-to-classical registers (CTC) preserve classical states, classical-to-quantum (CTQ) channels encode classical information in quantum states, quantum-to-classical (QTC) channels extract classical information from quantum states, decohering them in the process; quantum-to-quantum (QTQ) channels form coherent (e.g. unitary) transformations between quantum registers. In this framing, both agents and environments are registers which may be CAGI (classical state sets) or QAGI (quantum state sets). They may interact in ways that are coherent (quantum-preserving) or classical (via CTC or QTC maps).

3.2   Quantum AIXI

The components of QAIXI can then be understood as follows. Let the QAIXI agent be associated with a quantum register A representing the internal degrees of freedom, while its environment is represented by a quantum register E. Their corresponding (finite–dimensional) complex Hilbert spaces are denoted by $\mathcal{H}_A$ and $\mathcal{H}_E$. At interaction step t the agent's private state is a density operator $\rho_A^{(t)} \in \mathcal{D}(\mathcal{H}_A)$ that may encode its complete history, its current belief state, or—in the ideal case—an explicit representation of the universal quantum mixture Ξ_Q. The environment is simultaneously described by $\rho_E^{(t)} \in \mathcal{D}(\mathcal{H}_E)$. Taken together the composite system occupies the joint state $\rho_{AE}^{(t)} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_E)$. Note that $\rho_{AE}^{(t)}$ can be entangled, so that $\rho_{AE}^{(t)} \neq \rho_A^{(t)} \otimes \rho_E^{(t)}$. A history is therefore an operator-valued stochastic process rather than a sequence of point events.

Actions. At each cycle the agent chooses an *action* a_t, formally a completely–positive, trace–preserving (CPTP) map acting on the environment register (or on a designated subsystem of the joint register). The two canonical cases are as follows.

1. First, when the agent performs a coherent control operation, a_t is realised by a unitary channel $\Phi_{U_{a_t}} : X \mapsto U_{a_t} X U_{a_t}^\dagger$, applied either on $\mathcal{H}_E$ alone or on $\mathcal{H}_A \otimes \mathcal{H}_E$. Such a map preserves superposition and entanglement and therefore belongs to the quantum-to-quantum (QTQ) class of channels.
2. Second, the agent may decide to interrogate the environment by means of a *quantum instrument* $\mathcal{I}_{a_t} = \{\mathcal{E}_k^{a_t}\}_{k \in \Gamma_{\text{obs}}}$, where the outcome alphabet Γ_{obs} is classical. Each branch map $\mathcal{E}_k^{a_t} : \mathcal{L}(\mathcal{H}_{E'}) \to \mathcal{L}(\mathcal{H}_{E'})$ is completely positive and trace non–increasing, satisfies $\sum_k \mathcal{E}_k^{a_t} = \mathcal{E}^{a_t}$ (their sum is a CPTP map) and $\mathcal{E}_k^{a_t \dagger}(\mathbb{I}) = \mathbb{I}$, and typically takes the form $\mathcal{E}_k^{a_t}(X) = M_k^{a_t} X M_k^{a_t \dagger}$ for a POVM $\{M_k^{a_t}\}$. Because every branch deposits a classical record k while decohering the measured subsystem, $\mathcal{I}_{a_t}$ is a paradigmatic quantum-to-classical (QTC) channel.

Formally the action a_t can be represented via $\Phi_{a_t} : \mathcal{L}(\mathcal{H}_{AE}) \longrightarrow \mathcal{L}(\mathcal{H}_{AE})$ which may be indexed by the form of $\mathcal{L}$.

Percepts and rewards. What the QAIXI agent perceives is determined entirely by the instrument it applies. A measurement outcome $k \in \Gamma_{\text{obs}}$ becomes the observation component o_t of the percept $e_t = (o_t, r_t)$. The outcome is drawn with probability

$$\Pr\left(o_t = k \mid a_t, \rho_{AE}^{(t-1)}\right) = \text{Tr}\left[\mathcal{E}_k^{a_t}\left(\text{Tr}_A \rho_{AE}^{(t-1)}\right)\right].$$

The reward $r_t \in \Gamma_{\text{rew}}$ (with $\Gamma_{\text{rew}} \subseteq \mathbb{R}$ in the standard reinforcement-learning setting) is computed by a classical post-processing function that may depend on both o_t and the agent's prior internal state $\rho_A^{(t-1)}$. The resulting data o_t and r_t are stored in designated registers - but it is at this stage *classical* data. So it can be stored in a classical register via a CTC map, or encoded in a quantum register via a CTQ map.

Interaction loop. The QAIXI cycle of interaction with the environment is as follows. Given the pre-interaction state $\rho_{AE}^{(t-1)}$, the agent selects a_t. If a_t is a unitary, the composite state updates coherently to $\rho_{AE}^{(t)} = \Phi_{U_{a_t}}\left(\rho_{AE}^{(t-1)}\right)$. If instead a_t is an instrument, an outcome k is observed with

the above probability, the environment collapses to

$$\rho_{E'}^{(t)} = \mathrm{Tr}_A \rho_{AE}^{(t)} = \frac{\mathcal{E}_k^{a_t}\left(\mathrm{Tr}_A\,\rho_{AE}^{(t-1)}\right)}{\mathrm{Pr}(k)}, \tag{3}$$

The global post-measurement state is $\rho_{AE}^{(t)} = \rho_A^{(t-1)} \otimes \rho_{E'}^{(t)}$ (assuming no entanglement - see Appendix A for the general case). Finally, the agent applies an internal CPTP map—its update rule—to obtain $\rho_A^{(t)}$ from $\rho_A^{(t-1)}$ in the light of (a_t, o_t, r_t). This update may be trivial (identity) if $\rho_A^{(t)}$ is purely classical or it may itself be a non-trivial quantum channel when the agent maintains coherent beliefs. The agent's internal memory is refreshed by a chosen CPTP rule $\mathsf{U}_{\mathrm{int}} : \rho_A^{(t-1)} \mapsto \rho_A^{(t)}$ that may itself depend on (a_t, k, r_t). This operational definition fixes the semantics of *actions*, *percepts*, and *rewards* in the quantum setting, providing the concrete substrate on which the universal mixture ξ_Q and the QAIXI value functional are built.

3.3 Quantum Kolmogorov complexity

QAIXI relies upon quantum (rather than classical) Kolmogorov complexity. Let $\mathcal{Q}_{\mathrm{sol}}$ be the set of all chronological, semi–computable *quantum* environments, each such environment $Q \in \mathcal{Q}_{\mathrm{sol}}$ being represented by a Quantum Turing Machine that outputs an instrument sequence. An *environment* Q is a CPTP map acting on a register $\mathcal{H}_E$. Throughout we use the Choi-Jamiolkowski isomorphism [31] to identify Q with the purified vector $|Q\rangle := (\mathbb{1} \otimes Q)|\Phi^+\rangle \in \mathcal{H}_E^{\otimes 2}$, where $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i\,i\rangle$ for $d = \dim \mathcal{H}_E$. All trace-distance bounds on $|Q\rangle$ translate to diamond-norm bounds on Q. Its quantum Kolmogorov complexity is:

$$K_Q(Q) \;:=\; \min_{p \in \{0,1\}^\star}\left\{|p| \;\Big|\; \big\||U_{\mathrm{univ}}\,|p\rangle|0\rangle - |Q\rangle \otimes |\mathrm{aux}\rangle\big\| \le \varepsilon\right\}, \tag{4}$$

for a universal QTM U_{univ} and fixed $\varepsilon < 1$ with Hilbert-Schmidt norm $||\cdot||$. Equation (4) therefore measures program length needed to approximate the channel Q. The $|\mathrm{aux}\rangle$ term is an ancilla whose dimension is at most polynomial in $|p|$. Here we have assumed that for any two universal QTMs U_1, U_2 there exists a constant c_{U_1,U_2} such that for every environment Q:

$$\left|K_Q^{U_1}(Q) - K_Q^{U_2}(Q)\right| \;\le\; c_{U_1,U_2}. \tag{5}$$

The quantum Solomonoff mixture is the semi–density operator:

$$\Xi_Q\big(a_{1:m}\big) \;:=\; \sum_{Q \in \mathcal{Q}_{\mathrm{sol}}} 2^{-K_Q(Q)}\,\rho_E^Q(a_{1:m}), \tag{6}$$

where $\rho_E^Q(a_{1:m})$ is the environment state generated by Q under the action sequence $a_{1:m}$. Projecting Ξ_Q onto any classical POVM recovers the probability mixture $\xi_Q(e_{1:m}\|a_{1:m})$ that generalises (2). Whenever every admissible environment Q outputs commuting observables and the QAIXI agent restricts itself to instruments diagonal in that basis, each $\rho_E^Q(a_{1:m})$ becomes a diagonal density operator encoding an ordinary probability measure. In that limit $K_Q(Q)$ coincides (up to an additive constant) with the classical Kolmogorov complexity $K(\nu)$ of the induced CTM ν, and equations (6)–(9) reduce exactly to the classical Solomonoff prior ξ_U and the AIXI policy (1). Hence the quantum formulation strictly generalises the classical one.

3.4   QAIXI value functional

We also define the quantum analogue of the Bellman equation. For a policy π and environment Q define the discounted return

$$V_Q^\pi\big(\rho_{AE}^{(t-1)}\big) = \mathbb{E}_{\pi,Q}\Big[\sum_{k=t}^m \gamma^{k-t}r_k \ \Big| \ \rho_{AE}^{(t-1)}\Big]. \tag{7}$$

Because each branch map $E_k^{a_t}$ both yields o_t and updates $\rho_E^{(t)}(k)$, the expectation $E_{\pi,Q}\big[\sum_{k=t}^m \gamma^{k-t}r_k\big]$ is taken over the (non-Markovian) instrument-conditioned trajectory measure, not over an i.i.d. sequence. It terminates at finite horizon m. Averaging over the universal prior we obtain:

$$V_{\Xi_Q}^\pi\big(\rho_{AE}^{(t-1)}\big) = \sum_Q 2^{-K_Q(Q)}V_Q^\pi\big(\rho_{AE}^{(t-1)}\big). \tag{8}$$

The *QAIXI policy* maximises this functional at every step:

$$a_t := \pi^{\mathrm{QAIXI}}\big(\rho_{AE}^{(t-1)}\big) = \underset{a_t \in \mathsf{Act}}{\arg\max}\, \mathbb{E}_{k\sim p_t(\cdot|a_t)}\Big[r_t(k) + \gamma V_{\Xi_Q}^{\mathrm{QAIXI}}\big(\rho_{AE}^{(t)}(k)\big)\Big]. \tag{9}$$

where $p_t(k \mid a_t) = \mathrm{Tr}\big[\mathcal{E}_k^{a_t}\big(\,\Xi_Q^{(t)}(a_{1:t-1})\big)\big]$. Here $\rho_{AE}^{(t)}(k)$ is the post-measurement state with probabilities as $p_t(k \mid a_t)$. The fix-point equation implicit in ((8)–9) is the *quantum Bellman equation.* The act of observation fundamentally alters ρ_E due to quantum back-action. Note that as we are dealing with quantum trajectories, the classical notion of histories $h_{<t}$ must now account for the sequence of quantum operations and classical outcomes. Moreover, because a universal QTM may not halt on superposed inputs and is itself not computable. The substitution $K \rightsquigarrow K_Q$ leaves the mixture (6) just as incomputable as its classical counterpart and introduces an additional obstacle: evaluating $\rho_E^Q(a_{1:m})$ is in general computationally hard even for $m = 1$. However, preparing an arbitrary n-qubit quantum state $|\psi\rangle$ can require a quantum circuit of size exponential in n [37]. This implies that $K_Q(|\psi\rangle)$ can be exponentially larger (in qubit count) than n for complex states, unlike classical $K(x)$ which is at most $\ell(x) + c$ (only if x is the program itself, not an arbitrary binary string). The upper bound is still $O(2^n)$ bits. Practical approximations therefore hinge on identifying structured subclasses of $\mathcal{Q}_{\mathrm{sol}}$ (e.g. stabiliser processes or tensor-network environments) for which both K_Q and the Born probabilities are efficiently approximable.

4   Quantum Solomonoff Induction (QSI)

Classical Solomonoff induction assigns to every finite data string a universal a-priori probability obtained by summing, with complexity–based weights, over all computable environments that could have produced that string. In the quantum setting the data arriving at an agent are the classical outcomes of *instrument branches* executed on a quantum environment. The Quantum Solomonoff Induction formalism therefore replaces probability measures by *semi-density operators* and Kolmogorov codes by *quantum program states*.

Environment class and universal mixture. Let $\mathcal{Q}_{\mathrm{sol}}$ denote the set of all chronological, semi-computable quantum environments. Each $Q \in \mathcal{Q}_{\mathrm{sol}}$ is specified by a program $p(Q) \in \{0,1\}^*$ for a fixed universal QTM U_{univ}. Running $p(Q)$ produces, step by step, a sequence of CPTP maps that act on the environment register and a *measurement specification* for the classical transcript. Here Q plays the role of ν in the classical case. Its quantum Kolmogorov complexity is

$K_Q(Q) := \min\{|p(Q)| : U_{\text{univ}}(p(Q)) \simeq Q\}$. For an action sequence $a_{1:m}$ we write $\rho_E^Q(a_{1:m}) \in \mathcal{D}(\mathcal{H}_E)$ for the (generally mixed) state that Q prepares on the environment register immediately before the observation at cycle m. The *universal semi-density operator* conditional on the agent's actions is then:

$$\Xi_Q(a_{1:m}) := \sum_{Q \in \mathcal{Q}_{\text{sol}}} 2^{-K_Q(Q)} \rho_E^Q(a_{1:m}), \qquad 0 < \text{Tr}\big[\Xi_Q(a_{1:m})\big] \leq 1. \tag{10}$$

Projecting Ξ_Q onto the POVM that implements the agent's instrument at cycle m yields the scalar $\xi_Q(e_{1:m}\|a_{1:m}) := \text{Tr}\big[M_{e_{1:m}} \Xi_Q(a_{1:m})\big]$, which reduces to the classical Solomonoff prior ξ_U when all $M_{e_{1:m}}$ commute.

Bayesian updates. Given a history $h_{<t} = (a_{1:t-1}, e_{1:t-1})$, the posterior semi-density operator is obtained by the update followed by a renormalisation:

$$\Xi_Q^{(t)}(a_{1:t-1}) := \frac{\mathcal{M}_{e_{t-1}}\big(\Xi_Q^{(t-1)}(a_{1:t-2})\big)}{\text{Tr}[\mathcal{M}_{e_{t-1}}\big(\Xi_Q^{(t-1)}(a_{1:t-2})\big)]}, \qquad \Xi_Q^{(0)} := \Xi_Q. \tag{11}$$

Here $\mathcal{M}_{e_{t-1}}$ is the CP map of the realised branch of the agent's instrument at step $t-1$. The distribution for the next observation is:

$$\xi_Q(\,\cdot\,\|h_{<t}, a_t) = k \mapsto \text{Tr}\big[\mathcal{E}_k^{a_t}(\Xi_Q^{(t)}(a_{1:t-1}))\big]. \tag{12}$$

4.1   Convergence theorem

Analogous to classical Solomonoff induction, under certain conditions we might expect QSI to exhibit convergence properties. The convergence properties of QSI are complex and remain an open question. To sketch out the issues, we consider the following model (see Appendix D for more detail). Write $D(\rho\|\sigma) = \text{Tr}[\rho(\ln\rho - \ln\sigma)]$ for the *Umegaki relative entropy*. Fix a true quantum environment $Q^\star \in \mathcal{Q}_{\text{sol}}$ and let $\rho_E^\star(a_{1:m})$ be its state sequence. We assume:

(C1) Ergodicity. There is a $\delta > 0$ such that for every admissible action policy the time-averaged state satisfies: $\liminf_{m\to\infty} \frac{1}{m}\sum_{k=1}^m \big\|\rho_E^\star(a_{1:k}) - \rho_E^\star(a_{1:k-1})\big\|_1 \leq \delta$.

(C2) Informational completeness. Each cycle's instrument has a POVM refinement whose classical Fisher information matrix is full-rank up to error $\epsilon > 0$.

(C3) Complexity gap finite. $K_Q(Q^\star) < \infty$ and $g := \sum_{Q \neq Q^\star} 2^{-\big(K_Q(Q) - K_Q(Q^\star)\big)} < \infty$.

Theorem 1 (QSI convergence). *Under* (C1)–(C3) *the posterior density operator satisfies*

$$\mathbb{E}_{Q^\star}\big[D\big(\rho_E^\star(a_{1:t}) \,\big\|\, \Xi_Q^{(t)}(a_{1:t})\big)\big] \leq \frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t}. \tag{13}$$

Consequently, by the quantum Pinsker inequality [17, 32] (total variation distance given by KL-divergence),

$$\mathbb{E}_{Q^\star}\big[\tfrac{1}{2}\big\|\rho_E^\star(a_{1:t}) - \Xi_Q^{(t)}(a_{1:t})\big\|_1\big] = \mathcal{O}\big(t^{-1/2}\big). \tag{14}$$

Sketch of proof. Proving QSI convergence is an open question. One potential avenue is as follows. Define likelihood operators $\Lambda_t^Q := \mathcal{M}_{e_t} \circ \cdots \circ \mathcal{M}_{e_1}\big(\rho_E^Q(a_{1:t})\big)$. Monotonicity of quantum relative

entropy (the data–processing inequality; see [56]) implies that applying any branch map $\mathcal{M}_{e_k}$ can only *decrease* divergence:

$$D\big(\rho_E^\star \,\|\, \Xi_Q\big) \;-\; D\big(\Lambda_k^\star \,\|\, \Lambda_k^\Xi\big) \;\text{ is } \mathbb{E}_{Q^\star}\text{-martingale.} \tag{15}$$

Condition (C2) guarantees that each step's expected drop is non-negative. Summing these expected drops from $k = 1$ up to $k = t$ therefore shows that the *average* divergence after t cycles is bounded by the initial one,

$$\mathbb{E}_{Q^\star}\Big[D\big(\rho_E^\star \,\|\, \Xi_Q\big)\Big] \;\le\; \frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t}.$$

Finally, the quantum Pinsker inequality [17] turns this $1/t$ bound on relative entropy into an $\mathcal{O}(t^{-1/2})$ bound on trace distance, completing the convergence claim. The replacement of full informational completeness by ϵ-completeness introduces an extra $O(\epsilon)$ term in the Pinsker bound, which vanishes as $\epsilon \to 0$. Equation (11) is executed by the internal belief-revision QTQ channel $\mathsf{U}_{\mathrm{int}}$. The posterior $\Xi_Q^{(t)}$ supplies the conditional expectation required for the quantum Bellman equation (9). QSI is the inductive basis of QAIXI concentrating weight on environments that remain compatible with classical measurements. In the commuting limit the theorem reduces to the classical Solomonoff convergence bound [33]. A formal proof would need to justify martingale convergence for operator-valued universal mixtures under quantum measurement dynamics (App. D).

4.2   QSI Limitations

Aside from the computational complexity considerations of QSI (which we postpone for future work), several other significant challenges arise with QSI:

1. *Specifying $\mathcal{Q}_{\mathrm{sol}}$ and $K_Q(\cdot)$.* A universal class of *computable quantum environments* must be fixed, e.g. the set of all chronological, semi–computable QTMs that output instrument sequences. The quantum description-length $K_Q(Q)$ is then the length of the shortest qubit-string that drives a fixed universal QTM to approximate the environment Q within trace-distance ε. Proving universality and invariance up to an additive constant—standard for classical Kolmogorov complexity—is technically subtler in the quantum setting because programs may themselves be in superposition. Moreover, if an environment is weakly entangled, this may ultimately mean a classical AIXI can effectively approximate QAIXI's value function.
2. *Measurement back-action.* Born probabilities $\Pr_Q(e_{1:m}\|a_{1:m})$ are computed from the iterated CPTP maps that model the interaction loop (Sec. 3.2). Each measurement branch both yields the classical outcome o_t and *updates* the post-measurement state $\rho_E^{(t)} \mapsto \rho_E^{(t+1)}(o_t)$. Because measurement alters ρ_E, the quantum back-action makes the expectation in the Bellman equation dependent on the sequence of quantum operations and classical outcomes in a complex, non-Markovian way, so the likelihood of future data depends on the entire history $h_{<t}a_t$.
3. *Non-locality and contextuality.* When Q prepares entangled states, the joint distribution $\Pr_Q(e_{1:m}\|a_{1:m})$ can violate Bell inequalities [6] and exhibit contextual dependence on the full instrument sequence. QSI must therefore aggregate over models that are *not* representable by any classical hidden-variable process, which complicates identifiability and convergence analyses.
4. *Incomputability and resource sensitivity of the prior.* Like its classical counterpart, K_Q is uncomputable. Moreover, the weight $2^{-K_Q(Q)}$ does not penalise the physical resources needed either to prepare the initial state of Q or to simulate its dynamics—tasks that can be exponential-time or generally infeasible. For physically realistic prediction one may replace K_Q by a resource-aware complexity measure that incorporates, for example, circuit depth or Trotter step count.

5. *Quantum error correction (QEC).* Implementing QAIXI on real, noisy hardware requires quantum error correction. Encoding each logical qubit in hundreds-to-thousands of physical qubits both (i) inflates the effective description-length, shifting the universal prior toward far simpler environments, and (ii) introduces continual syndrome-measurement back-action whose residual logical noise is only bounded below a code-specific threshold—together, these effects raise the resource bar and weaken the assumptions (ergodicity and informational completeness) behind potential convergence guarantees.

Moreover, there are specific challenges to QAIXI and QSI raised by quantum foundational issues.

Bell non-locality. In any QAIXI environment that distributes entangled subsystems to space-like separated agent components the joint percept distribution can violate the CHSH inequality. No classical hidden-variable environment ν can reproduce those statistics. A classical ξ_U assigns zero mass to each Bell-violating local hypothesis, so only the quantum mixture ξ_Q has non-trivial mass to contribute. When the agent exploits Bell-type correlations for decision-making, its percepts are no longer conditionally-independent given the entire history, breaking the martingale structure assumed in Theorem 1. See Appendix F.

Kochen–Specker contextuality. If the true environment prepares a KS set of projectors, then there exists *no* history-independent map $v : \mathcal{P}(\mathcal{H}_E) \to \{0,1\}$ assigning pre-existing outcomes to every measurement the agent may perform. QSI must therefore sum over hypotheses whose outcome statistics depend on the entire future instrument sequence up to horizon m. Bayesian updates must track non-commuting observables (see Appendix E.1)..

Theorem 2. *Let $\{P_1, \dots, P_n\} \subset \mathcal{P}(\mathcal{H})$ be a Kochen–Specker uncolourable set whose projectors sum to the identity operator on $\mathcal{H}$, i.e. there exists* no *map $v : \mathcal{P}(\mathcal{H}) \to \{0,1\}$ satisfying $\sum_{i=1}^{k} v(\Pi_i) = 1$ for every projector decomposition $\sum_{i=1}^{k} \Pi_i = \mathbb{1}$ that contains only elements of the set. Let $\mathcal{I}$ be any instrument whose Kraus operators are polynomials in the P_j and whose action is confined to the support $\mathcal{H}$. Then no quantum Turing machine Q can output, for every action sequence $a_{1:m}$ an agent might perform, a commuting family of projectors $\{Q_{a_{1:m}}(e_{1:m})\}$ such that*

$$\forall a_{1:m}, \ \forall e_{1:m} \in \Gamma^m : \quad \Pr_{\text{env}}(e_{1:m} \| a_{1:m}) \ = \ \text{Tr}\big[Q_{a_{1:m}}(e_{1:m})\, \rho_E^\star\big], \tag{16}$$

with $\rho_E^\star$ the true environment state prepared before cycle 1.

Proof. Fix *one* action string $a_{1:m}^\dagger$ that instructs the agent to measure, at the final cycle, *every* projector in the KS-set. Because the QTM outputs commuting projectors $\{Q_{a_{1:m}^\dagger}(e_{1:m})\}$, for each j there is a classical random variable $X_j := v_Q(P_j) := e_{1:m}(P_j) \in \{0,1\}$, namely the indicator that the outcome sequence has eigenvalue 1 for P_j. Equation (16) asserts that these random variables reproduce the Born probabilities of the true state, hence the map $v_Q : \ P_j \longmapsto X_j(\omega)$ is *non-contextual*: it assigns 0 or 1 to P_j *without reference to the context in which P_j is measured.* By construction $\sum_{j \in \mathcal{C}} P_j = \mathbb{I}$ for every context $\mathcal{C} \subset \{1, \dots, n\}$, whence $\sum_{j \in \mathcal{C}} v_Q(P_j) = 1$ holds almost surely. Thus v_Q is a *global* transformation of the projector into $\{0,1\}$, contradicting KS uncolourability. Therefore a perfect, non-disturbing predictor Q cannot exist.

Contextuality problematises the straightforward analogy with classical history in AIXI albeit subject to the extent to which environments actually do manifest uncolourable equivalents. Because

a universal classical history $h_{<t}$ cannot determine simultaneously the outcomes of all future instruments, any Bayesian update rule that conditions the QSI posterior on $h_{<t}$ *alone* is necessarily incomplete; the posterior must be refined by the *entire future instrument schedule*. Consequently the martingale proof of Theorem 1 requires an adapted filtration that records measurement contexts.

No-cloning. Evaluating the quantum likelihood $\Pr_Q(e_{1:m}\|a_{1:m})$ requires *fresh* copies of the pre-measurement state $\rho_E^Q(a_{1:m})$, but the no-cloning theorem forbids their duplication from a single run. Hence each Bayesian update consumes the very evidence it needs for validation; the sample complexity of learning scales with the number of distinct instruments explored, whereas in the classical case the same trajectory can be replayed arbitrarily often. Two consequences of the above are that: (a) convergence proofs must be *context-sensitive*; a single universal mixture cannot assign fixed probabilities to all quantum experiments simultaneously; and (b) even under ideal identifiability the learning rate is limited by state-preparation resources giving rise to sample complexity consequences for any QAIXI agent. See Appendix F.2.

5   Conclusion and Open Questions

We have laid out an analysis of and quantum information processing-based mathematical framework for Quantum AIXI, a universal intelligent agent designed to operate within a quantum mechanical universe. We have shown, using our channel and register-based model of agent/environment interaction in the quantum setting, how quantum analogues of universal intelligence components can be constructed. However there are significant limitations to QAIXI. Firstly, it is nowhere near being a practical proposal: QEC requires millions of physical qubits for even modest logical qubit counts; maintaining extended coherence for QAIXI decision cycles remains far beyond existing technology. Theoretically, QAIXI relies on quantum Kolmogorov complexity for its universal prior and quantum Turing machines to model environments. Key challenges include the incomputability of quantum Kolmogorov complexity, the exponential resources often required for quantum state preparation and simulation of quantum dynamics (when performed classically), complexity of calculating probabilities. The quantum Bellman equation involves expectations over quantum measurements and evolutions that are contextual. Moreover, exactly when QAIXI would provide a quantum advantage is an open question (see App. B). As interest in both computational agency and quantum information processing grows, understanding the capabilities and constraints of integrating quantum technologies into AGI systems will become of increasing focus. Further research directions include:

1. *Computable Approximations*: Can meaningful, computable (or efficiently quantum-computable) approximations to QAIXI be developed? What restrictions on the class of quantum environments $\mathcal{M}_{Q,sol}$ or the definition of K_Q would make this possible?
2. *Convergence Rates and Bounds*: Rigorously establishing convergence theorems for QSI and deriving tight bounds on convergence rates, taking into account quantum information-theoretic limits (e.g., Holevo information, quantum data processing inequalities), in light of learnability and tomography [1] e.g. How does quantum measurement back-action affect learning speed?
3. *Quantum Resources*: How do quantum resources like entanglement, if available to the agent for its internal processing or for interacting with the environment, affect the performance or complexity of QAIXI?
4. *Impact of Quantum Interpretations*: While this paper focused on the standard (Copenhagen-like) measurement formalism, how would adopting alternative interpretations (e.g., Everettian Many-Worlds, Bohmian Mechanics [12], QBism [25]) alter the definition of QAIXI's optimality, its policy, or its perceived complexity? For example, in an Everettian setting, does QAIXI optimise expected reward across trajectories or branches [19, 54]?

References

1. Aaronson, S.: The learnability of quantum states. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **463**(2088), 3089–3114 (2007)
2. Aaronson, S.: Quantum computing since Democritus. Cambridge University Press (2013)
3. Aaronson, S.: Quantum machine learning algorithms: Read the fine print. Nature Physics p. 5 (2014)
4. Aaronson, S.: Shadow tomography of quantum states. In: Proceedings of the 50th annual ACM SIGACT symposium on theory of computing. pp. 325–338 (2018)
5. Aaronson, S., Arkhipov, A.: The computational complexity of linear optics. In: Proceedings of the forty-third annual ACM symposium on Theory of computing. pp. 333–342 (2011)
6. Bell, J.: Speakable and Unspeakable in Quantum Mechanics. Cambridge University Press, Cambridge, 2nd edn. (2004)
7. Bennett, M.T.: The optimal choice of hypothesis is the weakest, not the shortest. In: Artificial General Intelligence. Springer Nature (2023)
8. Bennett, M.T.: Technical appendices (2024). `https://doi.org/10.5281/zenodo.7641741`, `https://github.com/ViscousLemming/Technical-Appendices`
9. Bennett, M.T., Maruyama, Y.: The artificial scientist: Logicist, emergentist, and universalist approaches to artificial general intelligence. In: Artificial General Intelligence. Springer (2022)
10. Berthiaume, A., Van Dam, W., Laplante, S.: Quantum kolmogorov complexity. Journal of Computer and System Sciences **63**(2), 201–221 (2001)
11. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**(7671), 195–202 (2017)
12. Bohm, D.: A suggested interpretation of the quantum theory in terms of "hidden" variables. i and ii. Physical Review **85**(2), 166–193 (1952)
13. Bostanci, J., Watrous, J.: Quantum game theory and the complexity of approximating quantum nash equilibria. Quantum **6**, 882 (2022)
14. Brukner, Č.: Quantum causality. Nature Physics **10**(4), 259–263 (2014)
15. Catt, E., Hutter, M.: A gentle introduction to quantum computing algorithms with applications to universal prediction (2020), `https://arxiv.org/abs/2005.03137`
16. Cerezo, M., Verdon, G., Huang, H.Y., Cincio, L., Coles, P.J.: Challenges and opportunities in quantum machine learning. Nature Computational Science (2022). `https://doi.org/10.1038/s43588-022-00311-3`
17. Csiszár, I., Körner, J.: Information theory: coding theorems for discrete memoryless systems. Cambridge University Press (2011)
18. Deutsch, D.: Quantum theory, the church-turing principle and the universal quantum computer. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **400**(1818), 97–117 (1985)
19. Deutsch, D.: Quantum theory of probability and decisions. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **455**(1988), 3129–3137 (1999)
20. Dong, D., Chen, C., Chen, H., Tarn, T.J.: Quantum reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **38**(5), 1207–1220 (2008)
21. Dunjko, V., Briegel, H.J.: Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Reports on Progress in Physics **81**(7), 074001 (2018)
22. Ebtekar, A., Hutter, M.: Foundations of algorithmic thermodynamics. Physical Review E **111**(1), 014118 (2025)
23. Fallenstein, B., Soares, N., Taylor, J.: Reflective variants of solomonoff induction and aixi. In: International Conference on Artificial General Intelligence. pp. 60–69. Springer (2015)
24. Fang, K., Fawzi, O., Renner, R., Sutter, D.: Chain rule for the quantum relative entropy. Physical review letters **124**(10), 100501 (2020)
25. Fuchs, C.A., Mermin, N.D., Schack, R.: An introduction to qbism with an application to the locality of quantum mechanics. American Journal of Physics **82**(8), 749–754 (2014)

26. Goertzel, B.: The general theory of general intelligence: A pragmatic patternist perspective. Tech. rep., Singularity Net (2021)
27. Goertzel, B., et al.: Opencog hyperon: A framework for agi at the human level and beyond. Tech. rep., OpenCog (2023)
28. Grünwald, P.D., Vitányi, P., et al.: Algorithmic information theory. Handbook of the Philosophy of Information pp. 281–320 (2008)
29. Guo, H., Zhang, J., Koehler, G.J.: A survey of quantum games. Decision Support Systems **46**(1), 318–332 (2008)
30. Gutoski, G., Watrous, J.: Toward a general theory of quantum games. In: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing. pp. 565–574 (2007)
31. Haapasalo, E.: The choi–jamiołkowski isomorphism and covariant quantum channels. Quantum Studies: Mathematics and Foundations **8**(3), 351–373 (2021)
32. Hirota, O.: Application of quantum pinsker inequality to quantum communications. arXiv preprint arXiv:2005.04553 (2020)
33. Hutter, M.: Universal artificial intelligence: Sequential decisions based on algorithmic probability. Springer Science & Business Media (2004)
34. Hutter, M.: Universal Algorithmic Intelligence: A Mathematical Top→Down Approach, pp. 227–290. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
35. Hutter, M.: Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability. Springer, Heidelberg (2010)
36. Jerbi, S., Gyurik, C., Marshall, S., Briegel, H., Dunjko, V.: Parametrized quantum policies for reinforcement learning. Advances in Neural Information Processing Systems **34**, 28362–28375 (2021). `https://doi.org/10.5281/zenodo.5833370`
37. Knill, E.: Approximation by quantum circuits. arXiv preprint quant-ph/9508006 (1995)
38. Kochen, S., Specker, E.P.: The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics **17**(1), 59–87 (1967)
39. Leike, J., Hutter, M.: Bad universal priors and notions of optimality. COLT (2015)
40. Liu, Y., Arunachalam, S., Temme, K.: A rigorous and robust quantum speed-up in supervised machine learning. Nature Physics pp. 1–5 (2021). `https://doi.org/10.1038/s41567-021-01287-z`
41. McMillen, P., Levin, M.: Collective intelligence: A unifying concept for integrating biology across scales and substrates. Communications Biology **7**(1), 378 (Mar 2024)
42. Meyer, N., Ufrecht, C., Periyasamy, M., Scherer, D.D., Plinge, A., Mutschler, C.: A survey on quantum reinforcement learning. arXiv preprint arXiv:2211.03464 (2022)
43. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, 10th anniversary edn. (2010)
44. Özkural, E.: Ultimate intelligence part ii: Physical measure and complexity of intelligence. arXiv preprint arXiv:1504.03303 (2015)
45. Perrier, E.: Quantum Geometric Machine Learning. arXiv preprint arXiv:2409.04955 (2024)
46. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J., Killoran, N.: Evaluating analytic gradients on quantum hardware. Physical Review A **99**(3) (mar 2019). `https://doi.org/10.1103/physreva.99.032331`, `https://doi.org/10.1103%2Fphysreva.99.032331`
47. Schuld, M., Petruccione, F.: Introduction, p. 1–19. Quantum Science and Technology, Springer International Publishing (2018). `https://doi.org/10.1007/978-3-319-96424-9_1`, `https://doi.org/10.1007/978-3-319-96424-9_1`
48. Schuld, M., Petruccione, F.: Machine Learning with Quantum Computers. Springer (2021)
49. Soares, N., Fallenstein, B.: Two attempts to formalize counterpossible reasoning in deterministic settings. In: Artificial General Intelligence: 8th International Conference, AGI 2015, AGI 2015, Berlin, Germany, July 22-25, 2015, Proceedings 8. pp. 156–165. Springer (2015)
50. Solomonoff, R.J.: A Formal Theory of Inductive Inference. Part I & II. Information and Control **7**(1–2), 1–22, 224–254 (1964). `https://doi.org/10.1016/S0019-9958(64)90223-2`

51. Solé, R., Moses, M., Forrest, S.: Liquid brains, solid brains. Philosophical Transactions of the Royal Society B: Biological Sciences **374**(1774), 20190040 (2019)
52. Thórisson, K.R., Bieger, J., Thorarensen, T., Sigurðardóttir, J.S., Steunebrink, B.R.: Why artificial intelligence needs a task theory: and what it might look like. In: Artificial General Intelligence: 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016, Proceedings 9. pp. 118–128. Springer (2016)
53. Vitányi, P.M.: Quantum kolmogorov complexity based on classical descriptions. IEEE Transactions on Information Theory **47**(6), 2464–2479 (2002)
54. Wallace, D.: The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. Oxford University Press (2012)
55. Wang, P., Hammer, P.: Assumptions of decision-making models in agi. In: Artificial General Intelligence: 8th International Conference, AGI 2015, AGI 2015, Berlin, Germany, July 22-25, 2015, Proceedings 8. pp. 197–207. Springer (2015)
56. Watrous, J.: The Theory of Quantum Information. Cambridge University Press (2018)
57. Wiseman, H., Cavalcanti, E.: Causarum investigatio and the two Bell's theorems. Half a Century of Bell's Theorem, Quantum [Un]Speakables II (2017)

A   Entanglement in the Interaction Loop

In Section 3.2, the post-measurement joint state was written under the separability assumption $\rho_{AE}^{(t-1)} = \rho_A^{(t-1)} \otimes \rho_E^{(t-1)}$, leading to $\rho_{AE}^{(t)} = \rho_A^{(t-1)} \otimes \rho_{E'}^{(t)}$. When the agent and environment registers are initially entangled this factorisation is instead as follows. The outcome probability is given by:

$$\Pr(o_t = k \,|\, a_t, \rho_{AE}^{(t-1)}) \;=\; \mathrm{Tr}\Big[\big(\mathrm{id}_A \otimes \mathcal{E}_k^{a_t}\big)\rho_{AE}^{(t-1)}\Big]$$

with conditional post-measurement state:

$$\rho_{AE}^{(t)} \;=\; \frac{\big(\mathrm{id}_A \otimes \mathcal{E}_k^{a_t}\big)\big(\rho_{AE}^{(t-1)}\big)}{\Pr(k)}$$

and reduced environmental state:

$$\rho_{E'}^{(t)} = \mathrm{Tr}_A\,\rho_{AE}^{(t)} = \frac{\mathcal{E}_k^{a_t}\big(\mathrm{Tr}_A\,\rho_{AE}^{(t-1)}\big)}{\Pr(k)}.$$

Only when $\rho_{AE}^{(t-1)}$ is separable (or when the instrument completely decoheres the measured subsystem) does the joint state factorise into $\rho_A^{(t-1)} \otimes \rho_{E'}^{(t)}$. The subsequent internal QAIXI update $\mathsf{U}_{\mathrm{int}} : \rho_A^{(t-1)} \mapsto \rho_A^{(t)}$ now acts on $\rho_A^{(t)} = \mathrm{Tr}_E\,\rho_{AE}^{(t)}$, which already incorporates any entanglement-induced disturbance. This general form of interaction loop is valid for all pre-measurement states, entangled or not, and serves as the reference implementation for any rigorous analysis of QAIXI in the fully quantum regime.

B   Quantum Advantage for QAIXI

A fundamental question for any quantum extension of classical AIXI is: under what circumstances does quantum processing provide genuine computational advantages? We examine this question below in the case of specific environment classes where classical simulation becomes intractable.

B.1   Boson Sampling Environments

An common reference for assessments of quantum computational advantage over classical protocols is that of boson sampling [5], a non-universal photonic quantum protocol that injects indistinguishable single photons into a fixed linear interferometer and samples output patterns whose probabilities are governed by matrix permanents which is considered exponentially hard for classical computers. Let $\mathcal{B}_n$ denote the class of quantum environments that embed boson sampling processes as follows. Each environment $Q \in \mathcal{B}_n$ maintains an n-mode optical interferometer described by a random unitary matrix $U \in \mathrm{U}(n)$ drawn from the Haar measure. On each cycle, Q accepts a classical action a_t specifying input photon configuration. It then evolves the state via U, and outputs measurement outcomes from a fixed detection scheme. Suppose there exists a family $\{Q_n\} \subset \mathcal{B}_n$ such that (a) generating one percept from Q_n requires $\mathrm{poly}(n)$ quantum operations; and (b) computing the output distribution to total variation distance $< 2^{-n}$ is #P-hard. In such a case, any classical AIXI with total running time fails to approximate the optimal value unless the polynomial hierarchy collapses.

This can be understood by considering how AIXI's value function requires accurate estimation of environment transition probabilities. For $Q_n \in \mathcal{B}_n$, this reduces to sampling from the boson sampling distribution. The original boson sampling conjecture involves classical simulation which, to be within the required accuracy, would yield an algorithm for computing permanents of Gaussian random matrices—a problem conjectured to be #P-complete which would collapse the hierarchy, contradicting standard complexity-theoretic assumptions. In terms of value functions, let $\hat{V}^{\text{classical}}$ denote the value function computed by any polynomial-time classical agent. If $|\hat{V}^{\text{classical}} - V^*_{Q_n}| < 1/\text{poly}(n)$, then the agent must distinguish between boson sampling distributions and uniform distributions over the same support—a task equivalent to permanent estimation within exponential precision. In this way, boson sampling provides an in principle litmus test for quantum advantage, albeit with limited applicability to specific tasks AIXI or QAIXI would undertake. The boson sampling example illustrates how QAIXI's computational advantage over classical AIXI is contingent on environments involving problems which are classically hard.

B.2 Structured Environment Classes

The example above raises the obvious question of when do quantum environments admit efficient classical approximation? We draw upon results in quantum information to illustrate when this may be true in the case of weakly entangled systems. In quantum formalism, an environment Q is called a *Matrix Product Environment* (MPE) with bond dimension χ if its state evolution can be represented as:

$$\rho_E^{(t)} = \text{Tr}_{\text{aux}} \left[\bigotimes_{i=1}^{n} A_i^{[t]}(\chi) \right] \tag{17}$$

where each $A_i^{[t]}(\chi)$ is a $\chi \times \chi$ matrix depending on the history up to time t. The question then becomes how classically learnable is an MPE? Let $\mathcal{E}_\chi$ denote the class of all MPEs with bond dimension $\chi = O(1)$. Then there exists a classical polynomial-time agent that ε-approximates QAIXI's value function on any $Q \in \mathcal{E}_\chi$ with sample complexity $O(\text{poly}(n, \varepsilon^{-1}))$. The bond dimension of an MPE is the size of tensors used to represent the MPE and is a figure that relates to the complexity of the representation and the extent of its entanglement. The key observation is that MPEs with constant bond dimension admit efficient tomography using classical shadow tomography [4] which can, in certain cases, reduce the observables required to calculate the expectation values of exponentially many observables of an unknown quantum state to only logarithmic number of randomised measurements of that state.

C Information-Theoretic Limits on QAIXI

Information-theoretic limits upon quantum information processing and measurement [56] that are well-studied in QIP also have consequences for any QAIXI proposal. For example, the no-cloning theorem discussed above imposes fundamental limits on QAIXI's learning efficiency that have no classical analog. Since QAIXI observes each environment state only once per trajectory, convergence rates for learning the current environment state precisely are limited. For quantum environments with parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, the quantum Fisher information matrix provides fundamental

bounds on parameter estimation. For example, for a parametric quantum state $\rho(\boldsymbol{\theta})$, the quantum Fisher information matrix has elements:

$$[\mathcal{I}_Q(\boldsymbol{\theta})]_{jk} = \frac{1}{2}\mathrm{Tr}\left[\rho(\boldsymbol{\theta})\{\partial_j L(\boldsymbol{\theta}), \partial_k L(\boldsymbol{\theta})\}\right] \tag{18}$$

where $L(\boldsymbol{\theta})$ is the symmetric logarithmic derivative. Moreover, the quantum Cramér-Rao bound [56] (a bound on optimal precision for estimating parameters from quantum measurements) means that any QAIXI agent estimating environment parameters $\boldsymbol{\theta}$ from T measurement outcomes satisfies:

$$\mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \geq \frac{1}{T}\mathcal{I}_Q(\boldsymbol{\theta})^{-1} \tag{19}$$

This bound is generally tighter than the classical Fisher information bound, reflecting the additional constraints imposed by quantum measurement. QAIXI's potential advantage would likely emerge in environments where this structure aligns with quantum computational advantages.

D   QSI Detail and Limitations

Here we discuss conditions on QSI (Theorem 1) in particular some of the challenges facing proving convergence.

D.1   Convergence Conditions

Prior to measurement, each state is given a weight proportional to its description length. Recall from Theorem 1 that under assumptions *(C1)–(C3)*, the posterior semi-density operator $\Xi_Q^{(t)}$ (Eq. (11)) satisfies Eq. 13:

$$\mathbb{E}_{Q^\star}\left[D\left(\rho_E^\star(a_{1:t}) \,\|\, \Xi_Q^{(t)}(a_{1:t})\right)\right] = \mathrm{Tr}\left[\rho_E^\star(a_{1:t})\left(\ln \rho_E^\star(a_{1:t}) - \ln \Xi_Q^{(t)}(a_{1:t})\right)\right] \leq \frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t}. \tag{20}$$

To show this, recall that each density operator ρ_E^Q is associated with a quantum state Q where Q^* denotes the true environmental state. $\Xi_Q^{(t)}$ describes the agent's best assessment as to the true state of the environment ρ_E^* at time t (we leave the time dependence of ρ understood). The QAIXI belief state is a weighting using such descriptions over each possible state density operator:

$$\Xi_Q = \omega_{Q^*}\rho_E^* + \sum_{Q \neq Q^*} \omega_Q \rho_E^Q.$$

The ω_Q terms reflect the descriptions of state Q. The shortest description of Q is given by $K_Q(Q)$ where simpler descriptions get more weight via $\omega_Q = 2^{-K_Q(Q)}$. In classical minimum description length (MDL) formalism, the code length of a data sequence is a combination of the model complexity (description length) and likelihood of data under the chosen model (a dilution term). The $D\left(\rho_E^\star(a_{1:t}) \,\|\, \Xi_Q^{(t)}(a_{1:t})\right)$ term denotes the relative entropy and is reflective of the additional measurements required to minimise the estimate between Ξ and ρ_E^*. To see this, define:

$$Z = \omega_{Q^*} + \sum_{Q \neq Q^*} \omega_Q = \omega_{Q^*}(1+g) \tag{21}$$

where:

$$g = \sum_{Q \neq Q^*} \frac{\omega_Q}{\omega_{Q^*}} = \sum_{Q \neq Q^*} 2^{-(K_Q(Q) - K_Q(Q^*))} \tag{22}$$

Ξ_Q is not yet normalised and ρ_E remains a semi-density operator as $\mathrm{Tr}\rho_E \leq 1$ i.e. as with the classical case Z is sub-normalised e.g. $Z = \sum_Q \omega_Q \leq 1$ for prefix-free codes with binary word-lengths (see [28]). To apply the martingale and chain-rule identities we require density (not semi-density) operators. To do so, we note the correct model's share of the total description mass (the effective prior mass of the true environment inside the normalised mixture) is:

$$\frac{\omega_{Q^*}}{Z} = \frac{1}{1+g} = c_{max}$$
$$Z = \omega_{Q^*}(1+g)$$

The normalised prior at time $t = 0$ is then given by:

$$\Xi_Q = \frac{\Xi_Q^{(0)}}{Z} = \frac{\rho_E^*}{(1+g)} \tag{23}$$

The form of the bound D_0 depends on our choice of weighting c for ρ_E^* where $0 < c \leq c_{max}$. Choosing $c = c_{max}$ gives:

$$\Xi_Q^{(0)} = c\rho_E^* = \frac{1}{1+g}\rho_E^* \tag{24}$$

with relative entropy given by:

$$\begin{aligned}
D_0 &= \mathrm{Tr}[\rho_E(\ln \rho_E^* - \ln \Xi_Q)] \\
&\leq \mathrm{Tr}[\rho_E(\ln \rho_E^* - (\ln \rho_E^* - \ln(1+g)))] \\
&= \ln(1+g)
\end{aligned}$$

assuming $\mathrm{Tr}(\rho_E^Q) = 1$ for all states Q. This form includes only the dilution term. In the event of perfect information (so $g = 0$) it reduces to the complexity $K_Q(Q^*) \ln 2$. To assume the more familiar MDL form (which includes both complexity and dilution terms) we can scale c (while remaining within the bounds above) e.g. $c = \omega_{Q^*}^2/Z = \omega_{Q^*}/(1+g)$. Doing so gives:

$$D_0 = \mathrm{Tr}[\rho_E(\ln \rho_E^* - \ln \Xi_Q)] \tag{25}$$
$$\leq \mathrm{Tr}[\rho_E(\ln \rho_E^* - (\ln \omega_{Q^*} - \ln(1+g) + \ln \rho_E^*)] \tag{26}$$
$$= \underbrace{K_Q(Q^*) \ln 2}_{\text{model complexity}} + \underbrace{\ln(1+g)}_{\text{dilution}} \tag{27}$$

At each cycle s (action and observation), the agent's state update is a CPTP map $\mathcal{M}_s : \rho \mapsto \rho(a_{1:s})$ such that:

$$\Xi_Q^{(s)} = \frac{\mathcal{M}_s(\Xi_Q^{(s-1)})}{\mathrm{Tr}\mathcal{M}_s(\Xi_Q^{(s-1)})} \quad \text{and} \quad \rho_E^\star(a_{1:s}) = \frac{\mathcal{M}_s(\rho_E^\star(a_{1:s-1}))}{\mathrm{Tr}\mathcal{M}_s(\rho_E^\star(a_{1:s-1}))}. \tag{28}$$

Using the chain rule for quantum relative entropy [24] we obtain:

$$D_{s-1} = \mathop{\mathbb{E}}_{Q^\star}\left[D\big(\rho_E^\star(a_{1:s-1})\|\Xi_Q^{(s-1)}\big)\right] = \mathop{\mathbb{E}}_{Q^\star}\sum_{k=1}^{s-1} D\big(\rho_E^\star(a_k\,|\,a_{1:k-1})\,\big\|\,\Xi_Q^{(k)}(a_k\,|\,a_{1:k-1})\big). \qquad (29)$$

Applying (29) with $s = t+1$ and noting that each conditional divergence is non-negative gives

$$\mathop{\mathbb{E}}_{Q^\star} D\big(\rho_E^\star(a_{1:t})\,\|\,\Xi_Q^{(t)}(a_{1:t})\big) \;\leq\; D_0. \qquad (3)$$

Combining 27 and (3) and divide by t:

$$\mathbb{E}_{Q^\star} D\big(\rho_E^\star(a_{1:t})\,\|\,\Xi_Q^{(t)}(a_{1:t})\big) \;\leq\; \frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t} \qquad (30)$$

to obtain Eq. 13. In our sketch above, the transition from the relative entropy convergence to a convergence rate for distinguishability is achieved using the quantum Pinsker inequality, quantifying how the QAIXI agent's belief $\Xi_Q^{(t)}(a_{1:t})$ converges to the true state $\rho_E^\star(a_{1:t})$. In an information-theoretic sense, $\frac{1}{2}\|\rho - \sigma\|_1$ represents the maximum probability with which one can distinguish between two quantum states ρ and σ using an optimal measurement. The quantum Pinsker inequality (also known as the Csiszár–Kullback–Pinsker inequality for quantum states) provides that for any two density operators ρ and σ:

$$\tfrac{1}{2}\|\rho - \sigma\|_1 \;\leq\; \sqrt{\tfrac{1}{2}\,D(\rho\|\sigma)}.$$

We use the property that for non-negative random variables, if $A \leq B$, then $\mathbb{E}[A] \leq \mathbb{E}[B]$ and Jensen's inequality such that $\mathbb{E}[\sqrt{X}] \leq \sqrt{\mathbb{E}[X]}$ such that:

$$\mathbb{E}_{Q^\star}\big[\tfrac{1}{2}\|\rho - \sigma\|_1\big] \;\leq\; \sqrt{\tfrac{1}{2}\,\mathbb{E}_{Q^\star}\big[D(\rho\|\sigma)\big]}. \qquad (31)$$

Substituting into equation 31 with $\rho = \rho_E^\star(a_{1:t})$ and $\sigma = \Xi_Q^{(t)}(a_{1:t})$:

$$
\begin{aligned}
\mathbb{E}_{Q^\star}\big[\tfrac{1}{2}\big\|\rho_E^\star(a_{1:t}) - \Xi_Q^{(t)}(a_{1:t})\big\|_1\big] &\leq \sqrt{\tfrac{1}{2}\,\mathbb{E}_{Q^\star}\Big[D\big(\rho_E^\star(a_{1:t})\,\big\|\,\Xi_Q^{(t)}(a_{1:t})\big)\Big]} \\
&\leq \sqrt{\tfrac{1}{2}\cdot\frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t}} \\
&= \sqrt{\frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{2t}} \\
&= \left(\sqrt{\frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{2}}\right) t^{-1/2}.
\end{aligned}
$$

Hence :

$$\mathbb{E}_{Q^\star}\big[\tfrac{1}{2}\big\|\rho_E^\star(a_{1:t}) - \Xi_Q^{(t)}(a_{1:t})\big\|_1\big] = \mathcal{O}\big(t^{-1/2}\big). \qquad (32)$$

The significance of Eq. (32) is that the $\mathcal{O}(t^{-1})$ convergence rate for the expected relative entropy implies an $\mathcal{O}(t^{-1/2})$ convergence rate for the expected trace distance. Thus any measurement strategy can separate the agent's belief from the true state only with probability vanishing like $t^{-1/2}$.

This means that as the agent accumulates more data (as t increases), its belief state $\Xi_Q^{(t)}(a_{1:t})$ not only becomes a better approximation of the true state $\rho_E^\star(a_{1:t})$ in terms of information-theoretic divergence, but also becomes increasingly *indistinguishable* from the true state via any physical measurement. The maximum probability of successfully distinguishing the agent's belief from the true state diminishes as $t^{-1/2}$. This transition from a $1/t$ rate for a quadratic-like loss (relative entropy can be seen as a generalised squared error in the space of probability distributions or density matrices) to a $t^{-1/2}$ rate for an L_1-like distance (trace distance is the quantum analogue of total variation distance) is a common theme in statistical inference. It mirrors, for instance, aspects of the classical Bernstein–von Mises theorem where posterior distributions concentrate around the true parameter value. Importantly, this $\mathcal{O}(t^{-1/2})$ rate for distinguishability is derived without needing any assumptions beyond those already made for establishing the relative entropy bound (i.e., conditions (C1)–(C3) that underpin Theorem 1).

D.2   Conditions on QSI

Establishing the convergence of QSI is challenging because of the conditions that would need to subsist:

1. *(C1) Ergodicity:* There is a $\delta > 0$ such that for every admissible action policy the time-averaged (with respect to each admissible instrument schedule) state of the true environment $Q^\star$ satisfies

$$\liminf_{m \to \infty} \tfrac{1}{m} \sum_{k=1}^m \left\| \rho_E^\star(a_{1:k}) - \rho_E^\star(a_{1:k-1}) \right\|_1 \le \delta.$$

   This condition implies that the environment's state does not change too erratically on average, ensuring that past observations retain some relevance for predicting future states.

2. *(C2) Informational Completeness:* Each cycle's instrument (measurement) has a POVM refinement whose classical Fisher information matrix is full-rank up to error $\epsilon > 0$. This means the agent's measurements are sufficiently informative to distinguish different quantum states, which is crucial for learning.

3. *(C3) Complexity Gap:* The true environment $Q^\star$ has finite quantum Kolmogorov complexity, $K_Q(Q^\star) < \infty$, and the sum of relative complexities of other environments:

$$g := \sum_{Q \neq Q^\star} 2^{-\left(K_Q(Q) - K_Q(Q^\star) \right)}$$

   is finite. This ensures $Q^\star$ is not infinitely complex relative to the universal QTM and has a non-vanishing initial weight in the QSI mixture, and that the collective weight of alternative hypotheses is manageable.

D.3   Comparison with classical SI

In our sketch above, we have adapted the typical classical approach for convergence of classical SI.

1. Likelihood Operators and Initial Divergence Bound: The QSI mixture $\Xi_Q(a_{1:m})$ (Eq. (10)) is a semi-density operator representing the agent's belief about the environment state before the m-th observation, given actions $a_{1:m}$. The posterior $\Xi_Q^{(t)}$ is obtained via updates based on observed

outcomes (Eq. (11)). A first step, analogous to classical Solomonoff induction, is to bound the initial information deficit of the QSI mixture with respect to the true environment $Q^\star$. This initial divergence, $D_0 = D(\rho_E^\star \| \Xi_Q^{(0)})$, where $\Xi_Q^{(0)}$ is the initial QSI prior, can be bounded by the quantum Kolmogorov complexity of the true environment:

$$D_0 \leq K_Q(Q^\star) \ln 2 + \ln(1 + g). \tag{33}$$

This bound reflects that the cost of encoding the true environment within the universal mixture is related to its complexity $K_Q(Q^\star)$, and the term $\ln(1 + g)$ accounts for the mass of alternative hypotheses. Establishing this bound rigorously for semi-density operators and Umegaki relative entropy is central to a full proof.

2. Monotonicity of Quantum Relative Entropy and Martingale Argument: The core of the convergence argument hinges on the data-processing inequality for quantum relative entropy. This principle states that for any quantum operation (a completely-positive trace-preserving map, CPTP map), such as a measurement branch $\mathcal{M}_{e_k}$ occurring in the Bayesian update, the relative entropy between any two states cannot increase: $D(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)) \leq D(\rho \| \sigma)$. Recall the likelihood operators $\Lambda_t^Q :=$ $\mathcal{M}_{e_t} \circ \cdots \circ \mathcal{M}_{e_1}(\rho_E^Q(a_{1:t}))$ where $\rho_E^Q(a_{1:t})$ is interpreted as the initial state on which the sequence of CPTP maps $\mathcal{M}_{e_1}, \ldots, \mathcal{M}_{e_t}$ act.) Eq. (15) states that $D(\rho_E^\star \| \Xi_Q) - D(\Lambda_k^\star \| \Lambda_k^\Xi)$ is $\mathbb{E}_{Q^\star}$-martingale. Specifically, if $D_k = D(\rho_E^{\star(k)} \| \Xi_Q^{(k)})$ is the divergence between the true state and the posterior after k observation-update cycles, then $\mathbb{E}_{Q^\star}[D_k | \text{history } h_{<k}] \leq D_{k-1}$. The divergence, on average, tends to decrease or stay the same. The decline in divergence at step k, $\Delta_k = D_{k-1} - \mathbb{E}_{e_k \sim Q^\star}[D_k | h_{<k}]$, is therefore non-negative. As a result, the assumption $C2$ is needed to ensure that if the posterior $\Xi_Q^{(k-1)}$ is not perfectly aligned with the true state $\rho_E^{\star(k-1)}$. Measurement provides additional information that leads to a decrease in uncertainty (and greater weighting of the QAIXI belief Ξ towards the true state (g declines while ω_{Q^*} increases).

3. Average Divergence: By summing these non-negative expected drops over t interaction cycles, we get:

$$\sum_{k=1}^{t} \mathbb{E}_{Q^\star, \text{hist}}[\Delta_k] = \mathbb{E}_{Q^\star, \text{hist}}[D_0 - D_t] \leq D_0,$$

where D_0 is the initial divergence bounded as in Eq. (33), and $D_t = D(\rho_E^\star(a_{1:t}) \| \Xi_Q^{(t)}(a_{1:t}))$ is the divergence at step t. This implies that the sum of positive one-step learning gains (in terms of divergence reduction) is bounded by the initial total ignorance D_0. The result that:

$$\mathbb{E}_{Q^\star}[D_t] \leq \frac{D_0}{t} = \frac{K_Q(Q^\star) \ln 2 + \ln(1 + g)}{t}.$$

shows that the expected divergence at time t itself must decrease on average This establishes that the expected relative entropy between the true environment state and the QSI posterior converges to zero at a rate of $1/t$.

4. Quantum Pinsker Inequality for Trace Distance Convergence: Finally, the quantum Pinsker inequality provides a link between the trace distance $\|\rho - \sigma\|_1$ (a measure of distinguishability) and the relative entropy $D(\rho \| \sigma)$:

$$\frac{1}{2} \|\rho - \sigma\|_1^2 \leq D(\rho \| \sigma).$$

Applying this to the result from step 3:

$$\mathbb{E}_{Q^\star}\left[\frac{1}{2}\left\|\rho_E^\star(a_{1:t}) - \Xi_Q^{(t)}(a_{1:t})\right\|_1^2\right] \leq \mathbb{E}_{Q^\star}[D_t] \leq \frac{K_Q(Q^\star)\ln 2 + \ln(1+g)}{t}.$$

Using Jensen's inequality (or a linear version of Pinsker's inequality), this implies convergence in trace distance:

$$\mathbb{E}_{Q^\star}\left[\frac{1}{2}\left\|\rho_E^\star(a_{1:t}) - \Xi_Q^{(t)}(a_{1:t})\right\|_1\right] = \mathcal{O}(t^{-1/2}).$$

This shows that the QSI posterior state $\Xi_Q^{(t)}(a_{1:t})$ converges to the true environment state $\rho_E^\star(a_{1:t})$ in trace distance. The effect of ϵ-completeness in (C2), rather than perfect informational completeness, would typically introduce an additional error term in the bounds, which vanishes as $\epsilon \to 0$.

D.4   Limitations

The proof sketch above, while following a known pattern, relies on several steps that require careful and rigorous justification in the quantum setting:

1. *Martingale Theory for Quantum Processes.* Classical martingale convergence theorems are well-established for sequences of real-valued random variables with respect to a filtration. Extending these to sequences of quantum states (density operators or, as here, semi-density operators like Ξ_Q) that evolve via quantum operations (measurements and unitary evolutions) is a significant technical hurdle. The notion of a filtration $\mathcal{F}_t$ must properly capture the information gained from sequences of quantum measurements, which is complicated by measurement back-action, the non-commutative nature of observables, and potential contextuality.

2. *Relative Entropy with Semi-Density Operators.* The Umegaki relative entropy $D(\rho\|\sigma)$ is typically defined for density operators ρ and σ (where $\mathrm{Tr}\rho = \mathrm{Tr}\sigma = 1$). We have set out how this would be adjusted (normalised) in order to render it a full density operator. An alternative approach given the QSI mixture Ξ_Q is a semi-density operator ($\mathrm{Tr}\Xi_Q \leq 1$) is for the derivation - and properties (like monotonicity under CPTP maps), and the Pinsker inequality - to be established or appropriately adapted for semi-density operators.

3. *Measurement Back-Action.* Each measurement $\mathcal{M}_{e_k}$ provides a classical outcome e_k and transforms the quantum state. The sequences of true states $\rho_E^\star(a_{1:k})$ and posterior beliefs $\Xi_Q^{(k)}(a_{1:k})$ are conditioned on the full history of actions $a_{1:k}$ and prior outcomes $e_{1:k-1}$. The calculation of the expected drop in divergence requires careful averaging over outcomes e_k generated by the true environment $Q^\star$ acting on its state $\rho_E^\star(a_{1:k-1})$. The ergodicity condition (C1) is important in ensuring that the learning process is meaningful over time and does not get stuck due to highly non-stationary state sequences.

4. *Contextuality and Non-locality.* As we discuss, if the true environment $Q^\star$ exhibits Kochen-Specker contextuality or Bell non-locality, the classical intuition that a history uniquely determines future outcome probabilities (independent of measurement choices for other compatible observables) breaks down. The requirement for an adapted, context-aware filtration complicates the structure of the conditional expectations inherent in any martingale argument e.g. standard assumptions of conditional independence of percepts, often used in classical proofs, are violated. However, the practical implications of these foundational issues may be more or less significant depending on context (see the Appendix section below for an extended discussion).

5. *Informational Completeness.* Condition (C2) asserts that measurements are sufficiently informative (Fisher information matrix is full-rank up to ϵ), the precise manner in which this guarantees a sufficient trend towards convergence at each step needs would require further analysis. For example, how (and whether) the error ϵ propagates into the final convergence bounds and rates is would be open question.

6. *No-Cloning and Sample Complexity.* The no cloning theorem fundamentally limits the agent's ability to learn. A QAIXI agent cannot make multiple measurements on identical copies of a past unperturbed quantum state to refine its likelihood estimates for different actions from that state as measurement consumes the state. This underscores the online and single-shot nature of quantum learning from a trajectory. It also has consequences for the effective sample complexity of learning undertaken by any QAIXI and is a significant difference from classical scenarios where data can often be re-analysed. Note that if i.i.d. copies at negligible cost, the no-cloning impact may be minimal.

While the QSI convergence theorem sketch outlines a potentially plausible route to proving that a QAIXI-like agent can learn its quantum environment, a full proof faces unresolved theoretical and technical challenges. These primarily arise from the foundational differences between classical and quantum information processing, particularly concerning measurement, state representation, and contextuality. Addressing these issues remains a significant open research area in the theoretical foundations of quantum artificial general intelligence.

E   Foundational Implications for QAIXI

The Kochen-Specker (KS) theorem [38] is a cornerstone of quantum foundations, reflecting central differences between quantum and classical descriptions of reality. It states that for quantum systems of dimension three or higher, it is impossible to assign definite, pre-existing values to all possible measurements (observables) in a way that is independent of the context of measurement (i.e., the order of measurement operations and which other compatible observables are measured alongside it). Here we elaborate in more detail on the implications of KS contextuality for the QAIXI framework, particularly concerning Theorem 2.

E.1   Contextuality and QAIXI's Universal Mixture

Classical AIXI operates under the assumption that an environment ν has definite properties that can be observed. The history $h_{<t}$ is a sequence of such definite observations. However, if the true quantum environment $Q^\star$ leads to states whose measurements exhibit contextuality, then QAIXI's understanding of Q^* will differ from the classical case. Recall Theorem 2 provides that no quantum Turing machine Q (representing a predictive model for the environment) can output a commuting family of projectors $\{Q_{a_{1:m}}(e_{1:m})\}$ that perfectly predicts outcomes for a KS-uncolourable set of measurements (Eq. (16)). The consequence for QAIXI's belief state, $\Xi_Q(a_{1:m})$ (defined in Eq. (10)), is significant. QAIXI's belief is a mixture that sums over all semi-computable quantum environments $Q \in \mathcal{Q}_{\mathrm{sol}}$. If $Q^\star$ is KS-contextual, then any individual Q within the sum that attempts to model $Q^\star$ using classical, non-contextual hidden variables (or by outputting commuting projectors that aim to assign definite values independent of measurement choices) will fail to accurately reproduce the statistics of $Q^\star$. Thus, for Ξ_Q to converge to a description of $Q^\star$, it must implicitly give dominant weight to models Q that are themselves contextual or whose predictive mechanisms do not rely on non-contextual value assignments.

E.2   Contextuality Consequences for Learning

In classical AIXI, the history $h_{<t}$ is a sequence of action-percept pairs, where percepts are assumed to be objective records of definite environmental properties via measurement. For QAIXI situated within a contextual quantum environment, the situation is more nuanced. The outcome o_t of a measurement action a_t may not be interpretable as revealing a pre-existing property P_j of the system. Instead, o_t is realised only upon measurement. Intuitively, its probability (and even its meaning) can depend on the complete set of compatible observables measured in action a_t. To see this, let $Q^\star$ be an environment that prepares a qutrit ($d = 3$) system. Suppose QAIXI can choose actions a_t that correspond to measuring sets of compatible projectors from a KS set (e.g., a set of projectors for a qutrit as in the Peres-Mermin square, or simpler KS sets). Let $P_1, P_2, \ldots, P_n$ be such a KS set.

- If QAIXI performs action a_1 measuring the context $\mathcal{C}_1 = \{P_i, P_j, P_k\}$ (where $P_i + P_j + P_k = \mathbb{I}$), it obtains outcomes (o_i, o_j, o_k).
- If it later (or in a counterfactual scenario) performs action a_2 measuring context $\mathcal{C}_2 = \{P_i, P_l, P_m\}$ (where $P_i + P_l + P_m = \mathbb{I}$), it obtains outcomes (o'_i, o'_l, o'_m).

The KS theorem implies that QAIXI cannot learn a universal value assignment $v(P_x) \in \{0, 1\}$ such that $o_i = v(P_i)$ and $o'_i = v(P_i)$ consistently across all contexts for all projectors in the KS set, while

also satisfying $\sum_{P_x \in \mathcal{C}} v(P_x) = 1$ for all contexts $\mathcal{C}$. This means that the Bayesian update rule for QAIXI (Eq. (11)), $\Xi_Q^{(t)}(a_{1:t-1}) := \frac{\mathcal{M}_{e_{t-1}}\left(\Xi_Q^{(t-1)}(a_{1:t-2})\right)}{...}$, must process the information e_{t-1} (which includes o_{t-1}) in a way that acknowledges its contextual nature. The meaning of o_{t-1} for updating beliefs about $Q^\star$ is tied to the full instrument $\mathcal{M}_{e_{t-1}}$ and potentially the set of all measurements that constituted action a_{t-1}. This is why the posterior must be refined by the entire future instrument schedule (or at least the current one) if one aims for precise predictions in contextual scenarios. A simple classical history string $h_{<t}$ is insufficient.

E.3   Convergence Implications

The conditions for the convergence of QSI as sketched in Theorem 1, relies on a martingale argument. Martingales are defined with respect to a filtration, which represents the information accumulated over time. If measurement outcomes are contextual, the structure of this filtration becomes more complex. The information gained from an outcome o_t is not just about the specific observable measured but also about the context. This is why the martingale assumptions of Theorem 1 require an adapted filtration that records measurement contexts - because standard conditional independence assumptions, often implicit in simpler martingale proofs, may not hold if the probability of future outcomes depends on the context of past measurements in non-trivial ways. The ergodicity (C1) and informational completeness (C2) conditions would also need to be interpreted in light of contextuality. In essence, KS contextuality reflects the differences between classical and quantum ontology - and classical and quantum models of computation. Its consequence is that any QAIXI would need to adopt a more sophisticated model of reality than classical AIXI. It cannot assume that the environment possesses a set of definite, non-contextual properties that are merely uncovered by measurement. Instead, QAIXI must learn and operate in a world where measurement outcomes are co-created by the interaction between the agent's choice of measurement context and the quantum system.

F   Bell Non-Locality and No-Cloning

Beyond contextuality, other foundational quantum principles like Bell non-locality and the no-cloning theorem impose significant constraints and offer unique characteristics to QAIXI compared to its classical counterpart. We expand upon the dicussion above in the subsections below.

F.1   Bell Non-Locality

Bell's theorem [6] demonstrates that quantum mechanics predicts correlations between spatially separated systems (that were previously entangled) which cannot be explained by any theory based on local hidden variables (LHVs). The implications for QAIXI are indicative of the differences between quantum and classical environments. If the true environment $Q^\star$ subsists in entangled states and allowing QAIXI to perform measurements on these subsystems at space-like separation, then QAIXI's universal mixture Ξ_Q (Eq. (10)) must accommodate models Q that are inherently non-local:

1. Classical AIXI, relying on Solomonoff induction over classical Turing machines $\mathcal{M}_{sol}$, constructs its universal prior ξ_U (Eq. (2)) from environments that are, by their classical nature, local. Such a prior would assign zero probability to observing correlations that violate Bell inequalities (e.g., the CHSH inequality).

2. QAIXI, by summing over quantum environments $\mathcal{Q}_{sol}$, can, in principle, learn and adapt to a non-local $Q^\star$. The term $2^{-K_Q(Q)}\rho_E^Q(a_{1:m})$ must include Qs whose $\rho_E^Q(a_{1:m})$ can lead to Bell-violating statistics upon appropriate measurements.

For example, consider a QAIXI agent designed with two components, Alice and Bob, who are spatially separated.

(a) The environment $Q^\star$ repeatedly sends Alice and Bob a pair of qubits in an entangled Bell state, e.g., $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$.

(b) At each cycle t, Alice receives a classical random bit $x \in \{0,1\}$ and Bob receives $y \in \{0,1\}$ (these could be part of the agent's internal state or from an external prompter, effectively part of the action setup).

(c) Alice chooses a measurement setting A_x (e.g., measuring her qubit along one of two directions) and obtains outcome $o_A \in \{+1, -1\}$. Bob similarly chooses B_y and gets $o_B \in \{+1, -1\}$.

(d) The action for QAIXI could be $a_t = $ (setting choice for A_x, setting choice for B_y), and the percept $e_t = (o_A, o_B, r_t)$. The reward r_t could be 1 if $o_A \cdot o_B = (-1)^{xy}$ (CHSH game condition for certain settings) and 0 otherwise.

If $Q^\star$ is quantum, Alice and Bob can choose their measurement settings such that they win the CHSH game with a probability that is impossible for any classical LHV strategy. A classical AIXI, whose models $\nu \in \mathcal{M}_{sol}$ are constrained by LHV, would never be able to predict or achieve this quantum level of success. QAIXI, with its quantum prior Ξ_Q, could learn to implement the optimal quantum strategy and understand that $Q^\star$ is non-local. The presence of non-local correlations means that the percept $e_t = (o_A, o_B, r_t)$ contains components o_A and o_B that are correlated in a way not explainable by any shared information in their common past light-cone (beyond the initial entanglement). This challenges classical notions of conditional independence typically used in agent-based convergence proofs. It can also break the martingale structure assumed in Theorem 1 unless over non-local Q are accounted for. More detail can be found in literature on quantum game theory [13, 29, 30]

F.2   No-Cloning Theorem and QAIXI's Sample Complexity

The no-cloning theorem states that it is impossible to create an identical copy of an arbitrary unknown quantum state. This has consequences for how QAIXI can learn about its environment.

Learning   Firstly, because percepts upon which QAIXI learning is based are QTC channels, the no cloning theorem gives rise to differences in how QAIXI learning would occur:

1. When QAIXI performs an action a_t (especially if it's a measurement) on the environment state $\rho_E^Q(a_{1:t})$, this interaction alters the state. The post-measurement state is different, and the specific instance of $\rho_E^Q(a_{1:t})$ is consumed in yielding the outcome o_t.

2. Classical AIXI can, in principle, take a history $h_{<t}$ and test many hypothetical continuations with a given model ν without altering the data $h_{<t}$. Classical information can be copied and reused.

3. QAIXI cannot do this with quantum states. To know how $Q^\star$ would have responded to a different action a_t' from the *exact same* instance of $\rho_E^\star(a_{1:t-1})$, is impossible. It would need $Q^\star$ to produce that state (or an identical one) again.

Sample Complexity This one-shot nature of quantum measurement on unknown states impacts the sample complexity of learning by a QAIXI agent:

1. To distinguish between different hypotheses Q about the environment, or to estimate the expected outcome/reward for different actions, QAIXI needs to observe the environment's response multiple times.
2. Since each observation of a specific state instance is unique and unrepeatable, QAIXI effectively needs $Q^\star$ to prepare a new instance of a comparable state for each piece of information it wants to gather about a particular type of situation or action.
3. If QAIXI wants to learn the full characteristics of $\rho_E^\star(a_{1:t})$ through measurements, it is performing a form of quantum state tomography. Full tomography of an n-qubit state generally requires a number of measurement settings and repetitions that scales exponentially with n. While QAIXI's goal is not necessarily full tomography but rather optimal action selection, its ability to learn the relevant features of $Q^\star$ is still constrained by the information extractable per (unclonable) interaction.
4. As noted above, this means the learning rate is limited by state-preparation resources. If $Q^\star$ itself has computational costs or time delays associated with preparing states, this directly translates into a slower learning rate for QAIXI in terms of real-time or computational steps of $Q^\star$.

The informational completeness condition (C2) for Theorem 1 ensures that measurements are informative. However, no-cloning dictates that this information is gathered sequentially, one unclonable sample at a time. The $1/t$ convergence rate for relative entropy and $\mathcal{O}(t^{-1/2})$ for trace distance are asymptotic statements about the number of interactions t. The practical time or resources needed to achieve a certain level of accuracy will be higher in quantum scenarios where each sample is precious and unrepeatable, compared to classical scenarios where data can be exhaustively analyzed. Bell non-locality thus expands the class of environments QAIXI must consider beyond classical capabilities, while the no-cloning theorem imposes a fundamental restriction on the efficiency with which QAIXI can extract information from its quantum world, directly impacting its sample complexity for learning.

G   Notation

Glossary of Symbols	
a_t	Action chosen by the agent at cycle t (classical label).
$o_t,\ r_t$	Observation and reward components of the percept.
$e_t = (o_t, r_t)$	Percept at cycle t.
$h_{<t}$	Complete history up to but not including cycle t.
$\mathcal{A},\ \mathcal{O}$	Action and observation alphabets.
m	Fixed lifetime / planning horizon.
γ	Discount factor in $[0, 1)$.
μ, ν	Classical (semi-computable) environments.
$\mathrm{K}(\nu)$	Classical Kolmogorov complexity of ν.
ξ_U	Classical Solomonoff prior $\displaystyle\sum_\nu 2^{-\mathrm{K}(\nu)} \nu(\,\cdot\,)$.
$\mathcal{H}_A,\ \mathcal{H}_E$	Hilbert spaces of the agent and environment registers.
$\rho_A^{(t)}, \rho_E^{(t)}$	Agent / environment density operators for cycle t.
$\rho_{AE}^{(t)}$	Joint state of agent and environment.
$\Phi_{U_{a_t}}$	Unitary CPTP channel implementing a coherent action a_t.
$\mathcal{I}_{a_t} = \{\mathcal{E}_k^{a_t}\}_k$	Quantum instrument representing a measurement action.
Γ_{obs}	Finite outcome alphabet of the instrument.
$M_k^{a_t}$	Kraus / POVM operators of branch k.
$\mathcal{Q}_{\mathrm{sol}}$	Set of chronological, semi-computable quantum environments.
Q	Individual quantum environment (CPTP channel family).
$|Q\rangle$	Choi–Jamiołkowski purification of an environment channel.
$K_Q(Q)$	Quantum Kolmogorov complexity (Eq. 4).
$\Xi_Q(a_{1:m})$	*Operator-valued* universal mixture (semi-density op.)
$\xi_Q(e_{1:m}\|a_{1:m})$	*Scalar* probability obtained by projecting Ξ_Q onto the instrument POVM.
$U_{\mathbf{univ}}$	Fixed universal quantum Turing machine.
V_Q^π	Discounted value of policy π in environment Q.
$V_{\Xi_Q}^\pi$	Value averaged under the universal mixture (Eq. 8).
π^{QAIXI}	Optimal policy that maximises $V_{\Xi_Q}^\pi$ (Eq. 9).
Λ_t^Q	*Likelihood operator*: state after applying the first t measurement branches to ρ_E^Q (Sec. 4).
$D(\rho\|\sigma)$	Umegaki quantum relative entropy $\mathrm{Tr}[\rho(\log \rho - \log \sigma)]$.

Table 1. Summary of notation used.