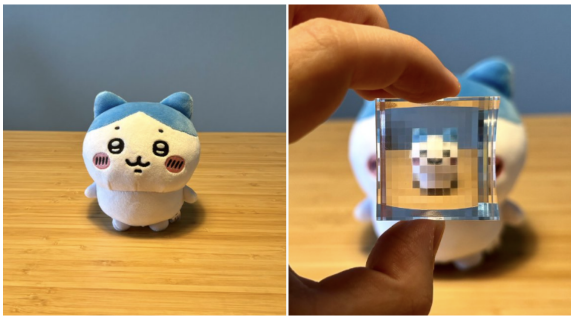
Macroscopic entropy and Bayesian inference: overview of recent results

Francesco Buscemi, Nagoya University

2025 Intl. Young Researchers Forum on Quantum Information Science
21-23 February 2025
Tainan, Taiwan



collaborators on this journey

- Clive Aw (CQT@NUS)
- Ge Bai (CQT@NUS)
- Kohtaro Kato (Nagoya)
- Teruaki Nagasawa (Nagoya)
- Arthur Parzygnat (MIT)
- Dominik Šafránek (IBS)
- Valerio Scarani (CQT@NUS)
- Joseph Schindler (UAB)
- Eyuri Wakakuwa (Nagoya)

a growing list:
The Observational Entropy Appreciation Club
(www.observationalentropy.com)

von Neumann entropy

For $\varrho = \sum_{x=1}^{d} \lambda_x |\varphi_x\rangle\langle\varphi_x|$ d-dimensional density matrix ($\lambda_x \geqslant 0$, $\sum_x \lambda_x = 1$),

$$S(\varrho) := -\operatorname{Tr}[\varrho \log \varrho] = -\sum_{x=1}^{d} \lambda_x \log \lambda_x$$

with the convention $0 \log 0 := 0$.

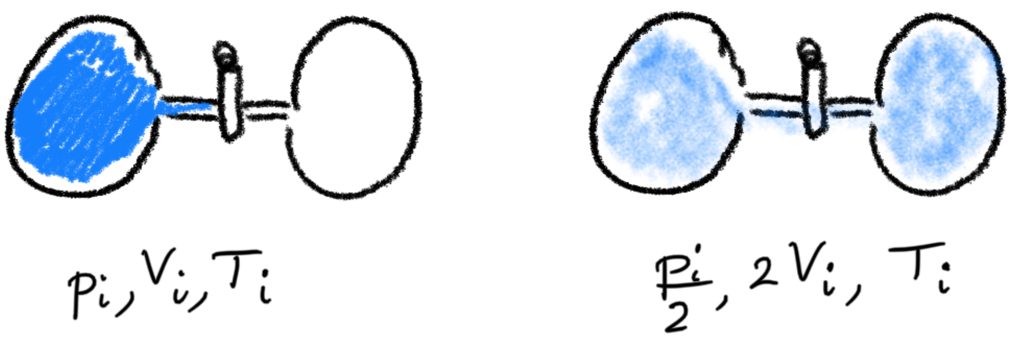
Unfortunately though:

> *"The expressions for entropy given by the author [previously] are not applicable here in the way they were intended, as they were computed from the perspective of an observer who can carry out all measurements that are possible in principle—i.e., regardless of whether they are macroscopic [or not]."*
>
> von Neumann, 1929; transl. available in `arXiv:1003.2133`

And again:

> *"Although our entropy expression, as we saw, is completely analogous to the classical entropy, it is still surprising that it is invariant in the normal [Hamiltonian] evolution in time of the system, and only increases with measurements—in the classical theory (where the measurements in general played no role) it increased as a rule even with the ordinary mechanical evolution in time of the system. It is therefore necessary to clear up this apparently paradoxical situation."*
>
> von Neumann, book (Math. Found. QM), 1932 (transl. 1955)

the paradox: free expansion of an ideal gas



$$\Delta S(\text{universe}) = nR \log 2 > 0$$

$\Longrightarrow$ there is net entropy increase in an isolated system's evolution, but von Neumann entropy remains constant

von Neumann's insight (inspired by Szilard's)

> *"For a classical observer, who knows all coordinates and momenta, the entropy is constant. [...]*
> *The time variations of the entropy are then based on the fact that the observer does not know everything—that he cannot find out (measure) everything which is measurable in principle."* von Neumann, 1932 (transl. 1955)

von Neumann recognizes that thermodynamic entropy should be a quantity *relative* to the observer's knowledge

von Neumann's proposal: macroscopic entropy

For a density matrix ϱ and an orthogonal resolution of identity (PVM) $\mathbf{\Pi} = \{\Pi_i\}_i$

$$S_{\mathbf{\Pi}}(\varrho) := -\sum_i p(i) \log \frac{p(i)}{\Omega(i)} , \qquad p(i) := \mathrm{Tr}[\varrho \, \Pi_i] \, , \ \Omega(i) := \mathrm{Tr}[\Pi_i] \, .$$

Modern version: observational entropy (OE)

For a positive operator-valued measure (POVM) $\mathbf{P} = \{P_i\}_i$

$$S_{\mathbf{P}}(\varrho) := -\sum_i p(i) \log \frac{p(i)}{V(i)} ,$$

where $p(i) := \mathrm{Tr}[\varrho \, P_i]$ and $V(i) := \mathrm{Tr}[P_i]$.

“observational” $:=$ “of the observer”

- von Neumann defines a **macro-observer** as a collection of simultaneously measurable quantities $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n, \dots\}$, where $\mathbf{X}_n = \{X_{j|n}\}_j$ are POVMs
- $\implies$ there exists one most-refined “parent” POVM $\mathbf{P} = \{P_i\}_i$ and a stochastic processing (i.e., cond. p.d.) μ such that

$$X_{j|n} = \sum_i \mu(j|n,i) P_i \ , \quad \forall j, n$$

- hence, a macro-observer for von Neumann is “just” a POVM (i.e., the most-refined parent POVM $\mathbf{P}$) from which all macroscopic quantities can be simultaneously inferred

the meaning of OE

Umegaki's quantum relative entropy

Definition

For density matrices ϱ, γ,

$$D(\varrho\|\gamma) := \begin{cases} \mathrm{Tr}[\varrho(\log\varrho - \log\gamma)] \ , & \text{if } \operatorname{supp}\varrho \subseteq \operatorname{supp}\gamma \ , \\ +\infty \ , & \text{otherwise.} \end{cases}$$

Useful properties:

- **monotonicity**: $D(\varrho\|\gamma) \geqslant D(\mathcal{E}(\varrho)\|\mathcal{E}(\gamma))$ for all channels (i.e., CPTP linear maps) $\mathcal{E}$ and all states ϱ, γ
- **parent quantity for micro-entropy**: $S(\varrho) = \log d - D(\varrho\|u)$ where $u := d^{-1}\mathbb{1}$
- **parent quantity for macro-entropy**: defining the quantum-to-classical measurement channel $\mathcal{P}(\bullet) := \sum_i \mathrm{Tr}[P_i \ \bullet] \ |i\rangle\langle i|$, it is easy to check that

$$S_{\mathbf{P}}(\varrho) = \log d - D(\mathcal{P}(\varrho)\|\mathcal{P}(u))$$

the fundamental bound

Theorem (NJP, 2023)

For any d-dimensional density matrix ϱ and any POVM $\mathbf{P} = \{P_i\}_i$,

$$S(\tilde{\varrho}_{\mathbf{P}}) - S(\varrho) \geqslant S_{\mathbf{P}}(\varrho) - S(\varrho) \geqslant D(\varrho\|\tilde{\varrho}_{\mathbf{P}}) \ ,$$

where

$$\tilde{\varrho}_{\mathbf{P}} := \sum_i \mathrm{Tr}[\varrho \ P_i] \, \frac{P_i}{V_i} \ .$$

In particular, $\log d \geqslant S(\tilde{\varrho}_{\mathbf{P}}) \geqslant S_{\mathbf{P}}(\varrho) \geqslant S(\varrho)$.

Remarks:

- the state $\tilde{\varrho}_{\mathbf{P}}$ only depends on the observer's knowledge
- in general, $[\varrho, \tilde{\varrho}_{\mathbf{P}}] \neq 0$
- in general, $S_{\mathbf{P}}(\varrho) \gtreqless H(\{p(i)\})$; but while $S_{\mathbf{P}}(\varrho)$ is monotonic under further postprocessings, $H(\{p(i)\})$ is not

OE tells us something about how much ϱ and $\tilde{\varrho}_{\mathbf{P}}$ "differ" from each other.

But what is the meaning of $\tilde{\varrho}_{\mathbf{P}}$?

Petz's transpose map

Petz (1986,1988)

Given a channel $\mathcal{E}$ and a prior state γ, the corresponding *transpose channel* is defined as

$$\mathcal{R}^{\gamma}_{\mathcal{E}}(\bullet) := \sqrt{\gamma}\, \mathcal{E}^{\dagger}\left[\mathcal{E}(\gamma)^{-1/2} \bullet \mathcal{E}(\gamma)^{-1/2}\right] \sqrt{\gamma}\ .$$

The "reconstructed" state

In terms of the measurement channel $\mathcal{P}(\bullet) := \sum_i \mathrm{Tr}[P_i\ \bullet]\, |i\rangle\langle i|$, it turns out that

$$\tilde{\varrho}_{\mathbf{P}} = [\mathcal{R}^{u}_{\mathcal{P}} \circ \mathcal{P}](\varrho) := d^{-1}\ \mathcal{P}^{\dagger}[\mathcal{P}(u)^{-1/2}\ \mathcal{P}(\varrho)\ \mathcal{P}(u)^{-1/2}]$$

So, the question to ask is: what is the meaning of Petz's transpose map?

exact recovery

Monotonicity: $D(\varrho\|\gamma) \geqslant D(\mathcal{E}(\varrho)\|\mathcal{E}(\gamma))$, for all $\mathcal{E}, \varrho, \gamma$.

Question: for which triples $(\varrho, \gamma, \mathcal{E})$ does the equality $D(\varrho\|\gamma) = D(\mathcal{E}(\varrho)\|\mathcal{E}(\gamma))$ hold?

Petz (1986,1988)

Answer: equality holds if and only if $[\mathcal{R}^{\gamma}_{\mathcal{E}} \circ \mathcal{E}](\varrho) = \varrho$. (The other equality $[\mathcal{R}^{\gamma}_{\mathcal{E}} \circ \mathcal{E}](\gamma) = \gamma$ is always satisfied *by construction*.)

But does Petz's transpose map also have a clear operational interpretation when $D(\varrho\|\gamma) > D(\mathcal{E}(\varrho)\|\mathcal{E}(\gamma))$?

Bayesian retrodiction

- consider a classical discrete noisy channel $P(i|x)$ and a prior $\gamma(x)$ on the input
- when the receiver reads a definite value i_0, (vanilla) Bayes' rule says that their posterior should be updated to $R_P^\gamma(x|i_0) := \dfrac{\gamma(x)P(i_0|x)}{[P\gamma](i_0)}$
- but what if the observation is noisy and returns some p.d. $\sigma(i)$ instead?

Theorem (Bayes–Jeffrey–Pearl retrodiction)

Given a channel $P(i|x)$ and a prior $\gamma(x)$, the result of a noisy observation $\sigma(i)$ is retrodicted to

$$\widetilde{\sigma}(x) := \sum_i R_P^\gamma(x|i)\sigma(i) .$$

The conventional Bayes' rule is recovered for $\sigma(i) = \delta_{i,i_0}$.

When everything commutes, Petz's transpose map coincides with the classical Bayes–Jeffrey–Pearl retrodiction rule.

But is this just a coincidence, or is there something deeper?

the principle of minimum change

"The updated belief should be consistent with the new information (the result of the observation), while deviating as little as possible from the initial belief."

Theorem (arXiv:2410.00319)

Given a qc-channel $\mathcal{P}(\bullet_{\rm in}) = \sum_i \mathrm{Tr}[P_i \bullet] \, |i\rangle\langle i|_{\rm out}$ *and a prior state* $\gamma_{\rm in} > 0$ *such that* $\mathcal{P}(\gamma) > 0$, *let* $Q^\gamma_{\mathcal{P}} := \sum_i |i\rangle\langle i|_{\rm out} \otimes \left(\sqrt{\gamma^T} P_i^T \sqrt{\gamma^T}\right)_{\rm in}$, *so that* $\mathrm{Tr}_{\rm in}[Q^\gamma_{\mathcal{P}}] = \mathcal{P}(\gamma)$ *and* $\mathrm{Tr}_{\rm out}[Q^\gamma_{\mathcal{P}}] = \gamma^T$.
Then, given any observation result $\sigma(i)$, *represented as* $\sigma_{\rm out} = \sum_i \sigma(i) |i\rangle\langle i|_{\rm out}$, *the optimization problem*

$$\max_{Q \geqslant 0 \text{ and } \mathrm{Tr}_{\rm in}[Q] = \sigma_{\rm out}} F(Q^\gamma_{\mathcal{P}}, Q) ,$$

where $F(A, B) := \left\| \sqrt{A}\sqrt{B} \right\|_1$ *is the (square-root) fidelity, has a unique solution* $\tilde{Q}$, *which in particular satisfies* $\mathrm{Tr}_{\rm out}\left[\tilde{Q}\right] = [\mathcal{R}^\gamma_{\mathcal{P}}(\sigma_{\rm out})]^T$.

immediate consequences

- Petz's transpose map is "the" analogue of Bayes' rule for quantum measurements
- $\tilde{\varrho}_{\mathbf{P}}$ is "the" quantum state to be retrodicted from the viewpoint of the macroscopic observer
- the difference between $S_{\mathbf{P}}(\varrho)$ and $S(\varrho)$ is a measure of "how retrodictable" ϱ is through $\mathbf{P}$, when the prior on the system is the uniform one

macroscopic = retrodictable

Definition

A state ϱ is macroscopic w.r.t. measurement **P** and prior γ whenever it can be perfectly retrodicted from them, i.e., whenever it belongs to the set

$$\mathfrak{M}_{\mathbf{P}}^{\gamma} = \{\varrho : \varrho = [\mathcal{R}_{\mathcal{P}}^{\gamma} \circ \mathcal{P}](\varrho)\} \; .$$

Theorem ($\star$)

A state ϱ is in $\mathfrak{M}_{\mathbf{P}}^{\gamma}$ if and only if there exists a PVM $\mathbf{\Pi} = \{\Pi_j\}_j$, with $\Pi_j = \sum_i \mu(j|i)P_i$, such that $[\Pi_i, \gamma] = 0$, together with coefficients $c_j \geqslant 0$, such that $\varrho = \sum_j c_j \Pi_j \gamma$.

Remark. The prior state is always macroscopic: $\gamma \in \mathfrak{M}_{\mathbf{P}}^{\gamma}$ for all POVMs **P**.

Remark. For uniform prior, i.e., $\gamma = u$, $\varrho \in \mathfrak{M}_{\mathbf{P}}^{u} \implies [\varrho, P_i] = 0$ for all i. (In general, it may be $[\gamma, P_i] \neq 0$.)

resolving the paradox of entropy increase in closed systems

- suppose that $\mathfrak{M}_{\mathbf{P}}^{u} \supsetneq \{u\}$ and let the initial state of the system at time $t = t_0$ be a macrostate $\varrho^{t_0} \neq u$
- the system evolves unitarily, i.e., $\varrho^{t_0} \mapsto \varrho^{t_1} = U\varrho^{t_0}U^\dagger$; thus,

$$\begin{aligned} S_{\mathbf{P}}(\varrho^{t_1}) &= -\sum_i \mathrm{Tr}\left[P_i\ (U\varrho^{t_0}U^\dagger)\right] \log \frac{\mathrm{Tr}\left[P_i\ (U\varrho^{t_0}U^\dagger)\right]}{\mathrm{Tr}[P_i]} \\ &= -\sum_i \mathrm{Tr}\left[(U^\dagger P_i U)\ \varrho^{t_0}\right] \log \frac{\mathrm{Tr}\left[(U^\dagger P_i U)\ \varrho^{t_0}\right]}{\mathrm{Tr}[U^\dagger P_i U]} \\ &= S_{U^\dagger \mathbf{P} U}(\varrho^{t_0}) \\ &\geqslant S(\varrho^{t_0}) = S_{\mathbf{P}}(\varrho^{t_0}) = S(\varrho^{t_1}) \end{aligned}$$

- summarizing: in general, $S_{\mathbf{P}}(\varrho^{t_1}) \geqslant S_{\mathbf{P}}(\varrho^{t_0})$, with equality if and only if $U\varrho^{t_0}U^\dagger \in \mathfrak{M}_{\mathbf{P}}^{u}$
- Corollary of Theorem ($\star$): $\varrho^{t_1} \in \mathfrak{M}_{\mathbf{P}}^{u} \implies [\varrho^{t_1}, P_i] = [U\varrho^{t_0}U^\dagger, P_i] = 0$ for all i
- hence, when the initial state is a macrostate $\varrho^{t_0} \neq u$, $S_{\mathbf{P}}(\varrho^{t_1}) > S_{\mathbf{P}}(\varrho^{t_0})$ *generically*

an "H-theorem" for OE

Theorem (PRR, 2025)

In a d-dimensional system, choose a state ϱ and a POVM $\mathbf{P} = \{P_i\}_i$ with a finite number of outcomes. Choose also a (small) value $\delta > 0$. For a unitary operator U sampled at random according to the Haar distribution, it holds:

$$\mathbb{P}_H \left\{ \frac{S_{\mathbf{P}}(U\varrho U^\dagger)}{\log d} \leqslant (1-\delta) \right\} \leqslant \frac{4}{\kappa(\mathbf{P})} e^{-C\delta\kappa(\mathbf{P})^2 d \log d} ,$$

where $\kappa(\mathbf{P}) = \min_i \mathrm{Tr}[P_i\, u]$ and $C \approx 0.0018$.

Remark. A similar statement holds for unitaries sampled from an approximate 2-design.

$\Longrightarrow$ in the eyes of the observer, the state of a randomly evolving system **quickly** becomes **indistinguishable** from the maximally uniform one, regardless of the system's initial state.

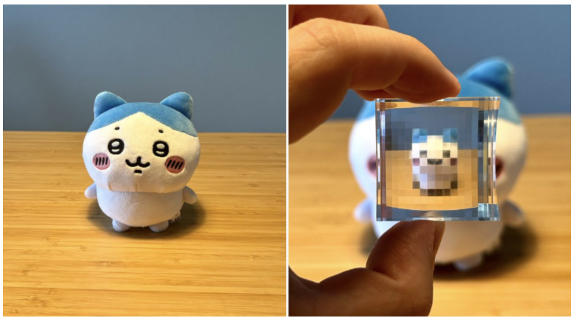
parenthesis: Watanabe's contention

"The phenomenological onewayness of temporal developments in physics is due to irretrodictability, and not due to irreversibility." Satosi Watanabe (1965)

- The second law is not about the arrow of time, but about the arrow of inference.
- The "mysterious" coarse-graining operation that appears in Gibbs' proof of the second law is nothing but Bayesian retrodiction done from the results of a macroscopic observation.

Conclusions

take-home messages



1. macroscopic entropy *emerges* from a fully operational/inferential scenario
2. Petz's transpose map *emerges* as the quantum Bayes rule, based on the principle of "minimum change"
3. the second law is about the generic loss of retrodictability

The End: Thank You!