Notes on the Margin to the Second Law of Thermodynamics

Francesco Buscemi, Nagoya University
`www.quantumquia.com`

the Second Law is "special"

"The law that entropy always increases holds, I think, the supreme position among the laws of Nature. [. . .] If your theory is found to be against the Second Law of Thermodynamics I can give you no hope; there is nothing for it to collapse in deepest humiliation." A.S. Eddington

"[. . .] the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown." A. Einstein

a journey into thermo: my Grand Tour

Companions on the journey: Clive Aw, Ge Bai, Kohtaro Kato, Shintaro Minagawa, Hamed Mohammady, Arthur Parzygnat, Dominik Šafránek, Kenta Sakai, Valerio Scarani, Joseph Schindler

my worry

if the Second Law is "special" and "cannot be overthrown", the argument of Maxwell's Demon must contain a fallacy

but where's the catch *exactly*? why do we *expect* that a fallacy must be there, in any demon-like argument?

traditional exorcisms assume particular models (trapdoors, pistons, ratchets, etc.)

I'm not satisfied with these: I want to see the Second Law and its "speciality" emerging from principles as a *logical necessity*

a contemporary example of a "quantum exorcism"

a line of research, initiated by Sagawa and Ueda in 2008 and still going strong within the stat-mech community, proposes a quantum exorcism called the Second Law of Information Thermodynamics:

- nonequilibrium free energy: $F_\beta^A(\varrho^A; H^A) := F_{\mathrm{eq},\beta}^A(H^A) + \beta^{-1} D(\varrho^A \| \gamma_\beta^A)$
- for β-isothermal processes, the Second Law reads $W_{\mathrm{ext}}^A \leqslant -\Delta F_\beta^A$
- in the presence of measurement and feedback (i.e., the Demon), the Second Law can be violated: $W_{\mathrm{ext}}^A \leqslant -\Delta F_\beta^A + \Delta$
- however, the work needed to do the measurement and erase it satisfies $W_{\mathrm{inj}}^{\mathrm{meas}} \geqslant \Delta$
- therefore the net work still obeys the Second Law: $W_{\mathrm{ext}}^A - W_{\mathrm{inj}}^{\mathrm{meas}} \leqslant -\Delta F_\beta^A$

the hope...

taken at face value, this approach is able to "prove" the Second Law as a consequence of the formalism of quantum measurement theory

...and the reality

in `arXiv:2308.15558` we look closely into these claims

- also in this case we found the "model fallacy": to prove the Second Law many (operationally unjustified and mutually inconsistent) assumptions are necessary—in particular, a restriction to von Neumann–Lüders-type measurements
- we removed as many assumptions as possible...
- in the end, we were able to remove them all, obtaining a universally valid Second Law of Information Thermodynamics, which is nice...
- ...but also found that it is logically equivalent to the conventional Second Law Thermo!

it seems that explanations of the Second Law "from within" are sound (i.e., tautologically true), but cannot be profound (cfr. Earman and Norton)

and so we're back to square one: I'm still worried

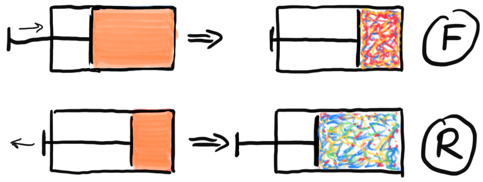
another exorcism: fluctuation relations

in the late 1990s, Jarzynski and Crooks discovered that the Second Law can be "proved" using two, strictly more powerful relations:

$$\langle W_{\mathrm{inj}} \rangle \geqslant \Delta F \;\Longleftarrow\; \left\langle e^{-\beta W_{\mathrm{inj}}} \right\rangle_F = e^{-\beta \Delta F} \;\Longleftarrow\; \frac{P_F(W_{\mathrm{inj}})}{P_R(-W_{\mathrm{inj}})} = e^{\beta(W_{\mathrm{inj}} - \Delta F)}$$

since these relations can prove the Second Law, then maybe they are the answer...

Crooks' proof



- stochastic thermodynamics: state, work, and energy are all random variables
- forward process: $P_F(x, y) = \gamma_\beta^{(0)}(x)\varphi_F(y|x)$ ($\gamma_\beta^{(0)}$: thermal for piston out)
- reverse process: $P_R(y, x) = \gamma_\beta^{(1)}(y)\varphi_R(x|y)$ ($\gamma_\beta^{(1)}$: thermal for piston in)
- microscopic reversibility: $\varphi_F(y|x) = \varphi_R(x|y)$
- $\implies \ln \frac{P_F(x,y)}{P_R(y,x)} = \ln \frac{\gamma_\beta^{(0)}(x)}{\gamma_\beta^{(1)}(y)} = \beta(F_\beta^{(0)} - \eta_x^{(0)} - F_\beta^{(1)} + \eta_y^{(1)}) = \beta(W_{\text{inj}} - \Delta F_\beta)$

again an assumption (microscopic reversibility) about the model is necessary
to say what the reverse process is

the inferential approach, or:
how I learned to stop worrying

a hint from Ed Jaynes

"To understand and like thermo we need to see it, not as an example of the n-body equations of motion, but as an example of the logic of scientific inference."

E.T. Jaynes (1984)

sufficiency of Bayesian retrodiction

- start from a forward process (statistical model) $\varphi_F(y|x)$
- fix a prior $\alpha(x)$ and compute the Bayes inverse $\varphi_R^\alpha(x|y) \propto \varphi_F(y|x)\alpha(x)$
- in particular, for a Hamiltonian process $\varphi_F(y|x) = \delta_{y,f(x)}$, the choice of α is immaterial: $\varphi_R(x|y) = \delta_{x,f^{-1}(y)}$ does not depend on α
- let $p(x)$ and $q(y)$, resp., be the initial states for the forward and the reverse processes
- $\implies \ln \frac{P_F(x,y)}{P_R^\alpha(y,x)} \equiv \ln \frac{p(x)\varphi_F(y|x)}{q(y)\varphi_R^\alpha(x|y)} = \ln \frac{p(x)}{q(y)} - \ln \frac{\alpha(x)}{\alpha'(y)}$
- choosing $p(x) = \gamma_\beta^{(0)}(x)$ and $q(y) = \gamma_\beta^{(1)}(y)$

$$\implies \ln \frac{P_F(x,y)}{P_R^\alpha(y,x)} = \beta(\Delta E - \Delta F_\beta) - \ln \frac{\alpha(x)}{\alpha'(y)}$$

- we get nonequilibrium potentials for free!

necessity of Bayesian retrodiction

- the log-ratio $\ln \frac{P_F(x,y)}{P_R(y,x)}$ plays a crucial role in stochastic thermodynamics (entropy production)
- it is itself a random variable, function of X and Y: $L \equiv \ell(X,Y) = \ln \frac{P_F(X,Y)}{P_R(Y,X)}$
- assume a form of "locality": $\ell(X,Y) = g(Y) - f(X)$ (note however that f and g can depend on the process φ, which is not a random variable)
- $\implies P_R(y,x) = P_R(y)\varphi_R^{\alpha}(x|y)$, for some prior $\alpha(x)$

if the reverse process is not a Bayesian retrodiction,
the entropy production is "nonlocal"

so: the Second Law is special because it is a law of logic, not physics

but now I'm worried about probabilities: where do they come from?

a hint from John von Neumann (inspired by Szilard)

"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)

Thus, von Neumann links the Second Law to an incomplete observation of the system

observational entropy

For

- ϱ density matrix,
- $\mathbf{P} = \{P_i\}_i$ POVM (i.e., $P_i \geqslant 0$, $\sum_i P_i = \mathbb{1}$),
- $p_i = \mathrm{Tr}[\varrho\ P_i]$,
- $V_i := \mathrm{Tr}[P_i]$,

The *macroscopic* or *observational* entropy of ϱ with respect to observer **P** is given by

$$S_{\mathbf{P}}(\varrho) := -\sum_i p_i \log \frac{p_i}{V_i}$$

first interpretation

Theorem

Given a POVM $\mathbf{P} = \{P_i\}_i$*, define the CPTP linear map* $\mathcal{P}(\bullet) := \sum_i \mathrm{Tr}[P_i\ \bullet]\ |i\rangle\langle i|$*. Then, for any state* ϱ*,*

$$\begin{aligned}\Sigma_{\mathbf{P}}(\varrho) :&= S_{\mathbf{P}}(\varrho) - S(\varrho)\\ &= D(\varrho\|u) - D(\mathcal{P}(\varrho)\|\mathcal{P}(u))\\ &\geqslant D(\varrho\|\widetilde{\varrho}_{\rm cg})\ ,\end{aligned}$$

where $u = d^{-1}\mathbb{1}$ *and* $\widetilde{\varrho}_{\rm cg} := \sum_i p_i P_i/V_i$ *is the coarse-graining of* ϱ *through* **P**. *If* $\varrho = \widetilde{\varrho}_{\rm cg}$*, the state* ϱ *is said to be macroscopic for observer* **P**.

Hence, the closer is $S_{\mathbf{P}}(\varrho)$ to the "true" $S(\varrho)$, the closer is $\widetilde{\varrho}_{\rm cg}$ to the "true" ϱ.

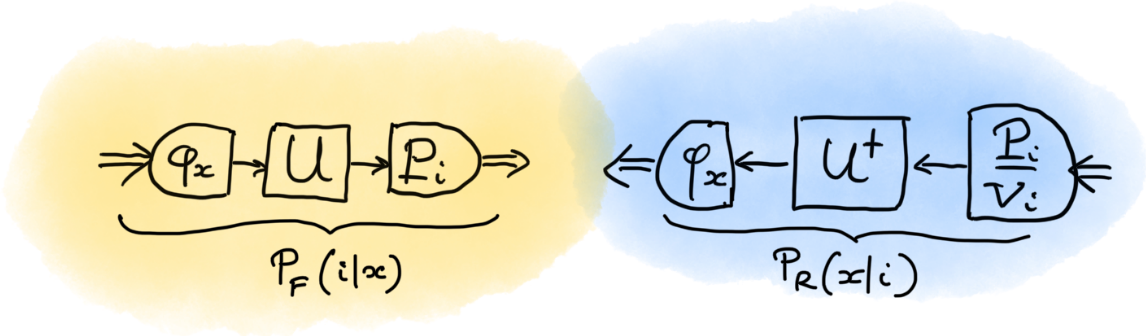
second interpretation

Theorem

Given a d-dimensional system, a density matrix ϱ with diagonalization $\{\lambda_x, |\varphi_x\rangle\}_{x=1}^d$, a unitary operator U, and a POVM $\mathbf{P} = \{P_i\}_i$, let us define two joint probability distributions:

$$P_F(x,i) := \lambda_x \underbrace{\mathrm{Tr}\left[U|\varphi_x\rangle\langle\varphi_x|U^\dagger \; P_i\right]}_{P_F(i|x)} , \qquad P_R^u(x,i) := P_F(i)\underbrace{\mathrm{Tr}\left[|\varphi_x\rangle\langle\varphi_x| \; \frac{U^\dagger P_i U}{V_i}\right]}_{P_R^u(x|i)} .$$

Then, $S_{\mathbf{P}}(U\varrho U^\dagger) - S(\varrho) = D(P_F \| P_R^u)$. Hence, the coarse-grained state is, in fact, the retrodicted 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)

generalization to non-uniform priors

Suppose that the retrodictor's uniform prior u is replaced with another state γ, but such that $[\varrho, \gamma] = 0$, i.e., predictor's and retrodictor's priors commute.

Then, everything goes through:

- define
$$S^{\mathsf{clax}}_{\mathbf{P},\gamma}(\varrho) := -\operatorname{Tr}[\varrho \, \log \gamma] + \sum_i p_i \log \frac{p_i}{q_i} \; ,$$
with $p_i := \operatorname{Tr}[\varrho \; P_i]$ and $q_i := \operatorname{Tr}[\gamma \; P_i]$
- then
$$S^{\mathsf{clax}}_{\mathbf{P},\gamma}(\varrho) - S(\varrho) = D(\varrho \| \gamma) - D(\mathcal{P}(\varrho) \| \mathcal{P}(\gamma)) = D(P_F \| P_R^{\gamma}) \; ,$$
with $P_F(x,i) = \lambda_x \langle \varphi_x | P_i | \varphi_x \rangle$ and $P_R^{\gamma}(x,i) = P_F(i) \frac{\gamma_x \langle \varphi_x | P_i | \varphi_x \rangle}{q_i}$

so: probabilities come from the interaction of a macro-observer with a micro-system

but now: is entropy "physical" or "bettabilitarian"?

thermodynamics: physics or beliefs?

- if the Second Law needs probabilities, and if probabilities need an observer, is thermodynamics about "physics" or "beliefs"?
- a more modest question: is there a "bettabilitarian" interpretation of the Second Law?

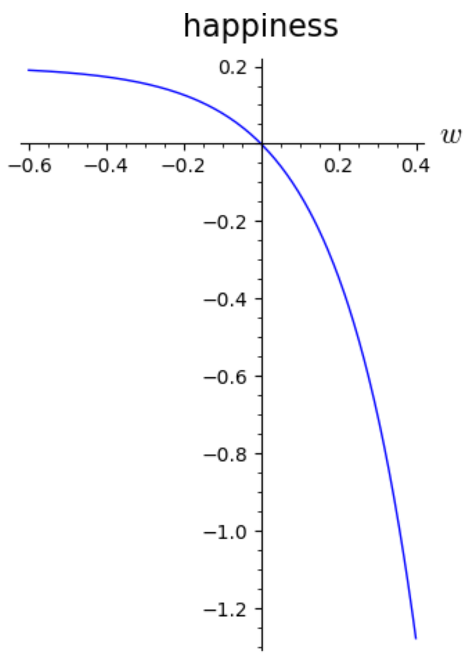
the setting

- an agent is forced to make a choice (Ginsberg's theorem): either activate a stochastic piston (like in Crooks' process) and pay the random value W_{inj}
- or walk away and pay a fixed amount of energy $\overline{W}$
- what is the "correct" price? it depends on the agent's risk-aversion
- in Expected Utility Theory, agents are characterized by their "utility function" $u : \mathbb{R} \to \mathbb{R}$ measuring the agent's "happiness" $u(w)$ associated with amount w
- risk-aversion is measured by the curvature of u
- in applications, often one resorts to utility functions having *constant absolute risk aversion* (CARA):

$$u_r(w) := \frac{1}{r}(1 - e^{rw})$$

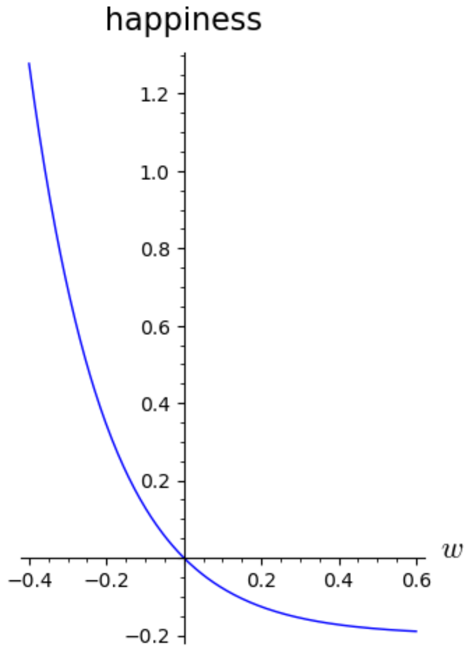
example: risk-averse agent ($r = 5$)

$$u(w) = \frac{1}{5}(1 - e^{5w})$$



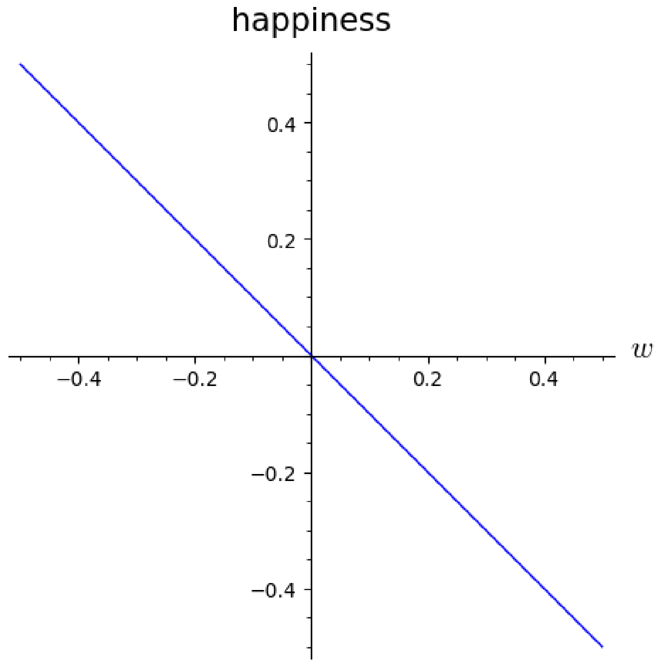
example: risk-seeking agent ($r = -5$)

$$u(w) = \frac{1}{5}(e^{-5w} - 1)$$



example: almost risk-neutral agent ($r = 0.001$)

$$u(w) = 1000(1 - e^{w/1000}) \xrightarrow{r \to 0} -w$$



entropy as "certainty-equivalent work"

- consider a stochastic piston: $\beta(W_{\rm inj} - \Delta F) \equiv w = \ln \frac{P_F(w)}{P_R(-w)}$
- the agent must either compress the piston and pay whatever value w occurs, or walk away and pay a fixed amount $\overline{w}$
- the "certainty-equivalent work" for agent $u_r(w)$ is given implicitly by

$$u_r(w_{\rm CE}^{(r)}) = \langle u_r(w)\rangle_F \iff w_{\rm CE}^{(r)} = u_r^{-1}\left[\langle u_r(w)\rangle_F\right]$$

- if $\overline{w} < w_{\rm CE}^{(r)}$, a player will pay and quit; otherwise they will gamble (if equality holds, the two options are equally preferable)

Theorem

For any $r \in [-\infty, +\infty]$,

$$w_{\rm CE}^{(r)} = D_{1+r}(P_F(w)\|P_R(-w)) ,$$

where $D_{1+r}(p\|q) := \frac{1}{r}\ln\langle (p/q)^r\rangle_p$. *(For* $r \in [-1, +\infty]$ *these are Rényi divergences.)*

special cases

- bears fear that the worst possible outcome may occur (Yunger-Halpern et al.)

$$w_{\rm CE}^{(+\infty)} = D_\infty(P_F(w)\|P_R(-w))$$

- bulls count on the fact that the best possible outcome may occur

$$w_{\rm CE}^{(-\infty)} = D_{-\infty}(P_F(w)\|P_R(-w))$$

- for $r = -1$, we get $w_{\rm CE}^{(-1)} = D_0(P_F(w)\|P_R(-w)) = \ln \sum_{w:P_F(w)>0} P_R(-w) = 0$: an agent so lazy that prefers to gamble as soon as $\overline{w} > 0$
- again, the Second Law corresponds to the case of a perfectly logical agent

$$w_{\rm CE}^{(0)} = D_{\rm KL}(P_F(w)\|P_R(-w)) = \beta(\langle W_{\rm inj}\rangle - \Delta F)$$

corollary: a generalized Jarzynski relation

Theorem

For any $r \in [-\infty, +\infty]$,

$$\left\langle e^{r\beta(W_{\mathrm{inj}}-\Delta F)}\right\rangle_F = e^{r w_{\mathrm{CE}}^{(r)}} .$$

The conventional case is recovered for $r=-1$, *for which* $w_{\mathrm{CE}}^{(-1)} = 0$[a].

[a] Assuming a normalized reverse process; otherwise this is the so-called *log-efficacy*.

conclusion

take home messages

- physics alone cannot explain the “special role” of the Second Law
- the Second Law is a statement about the agent’s stochastic inference and its logical consistence
- the inference in general is probabilistic, because the macro-observer cannot have a complete observation of the micro-system (not only in practice, but also in principle!)
- thus, there’s a notion of “agency” hiding in the Second Law and indeed one can “bet” on it

thank you for your attention