The theory of statistical comparison

from majorization to quantum thermodynamics

Francesco Buscemi (Nagoya University)

Colloquium, Department of Mathematics, Osaka University
23 May 2022

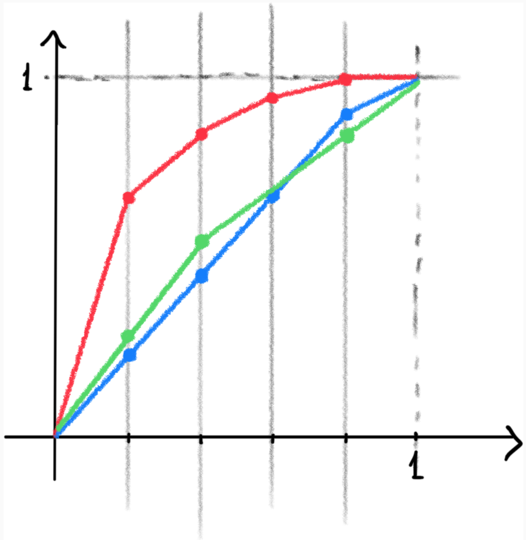
The precursor: majorization

Lorenz curves and majorization

- two probability distributions, $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$
- truncated sums $P(k) = \sum_{i=1}^{k} p_i^{\downarrow}$ and $Q(k) = \sum_{i=1}^{k} q_i^{\downarrow}$, for all $k = 1, \ldots, n$
- $\boldsymbol{p}$ majorizes $\boldsymbol{q}$, i.e., $\boldsymbol{p} \succ \boldsymbol{q}$, whenever $P(k) \geqslant Q(k)$, for all k
- minimal element: uniform distribution $\boldsymbol{e} = n^{-1}(1, 1, \cdots, 1)$

Hardy–Littlewood–Pólya, 1929

$\boldsymbol{p} \succ \boldsymbol{q} \iff \boldsymbol{q} = M\boldsymbol{p}$, for some bistochastic matrix M.



$(x_k, y_k) = (k/n, P(k)), \quad 1 \leqslant k \leqslant n$

Blackwell's extension

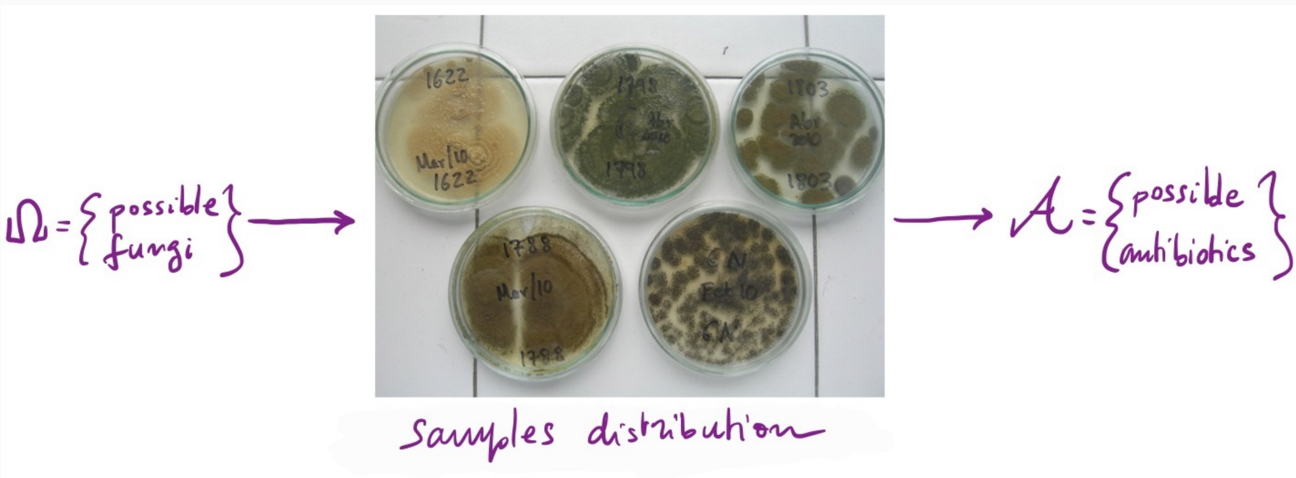
Statistical experiments

Lucien Le Cam (1924-2000)

"The basic structures in the whole affair are systems that Blackwell called experiments, and transitions between them. An experiment is a mathematical abstraction intended to describe the basic feature of an observational process if that process is contemplated in advance of its implementation."

Lucien Le Cam (1984)

A concrete example...



...and its abstract formulation

Definition (Statistical models and decision problems)

$$
\begin{tikzcd}
\Omega \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{A} \arrow[d, squiggly] \\
\omega \arrow[r, "w(x|\omega)"'] & x \arrow[r] & a
\end{tikzcd}
$$

- *parameter* set $\Omega = \{\omega\}$, *sample* set $\mathcal{X} = \{x\}$, *action* set $\mathcal{A} = \{a\}$
- a **statistical model/experiment** is a triple $\mathbf{w} = \langle \Omega, \mathcal{X}, w(x|\omega) \rangle$
- a **statistical decision problem/game** is a triple $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$, where $c : \Omega \times \mathcal{A} \to \mathbb{R}$ is a payoff function

Playing statistical games with experiments

- the experiment/model is the resource: it is given
- the decision is the transition: it can be optimized

$$
\begin{tikzcd}
\Omega \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{A} \arrow[d, squiggly] \\
\omega \arrow[r, "w(x|\omega)"'] & x \arrow[r, "d(a|x)"'] & a
\end{tikzcd}
$$

Definition

The **(expected) maximin payoff** of a statistical model $\mathbf{w} = \langle \Omega, \mathcal{X}, w \rangle$ w.r.t. a decision problem $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$ is given by

$$c^*_{\mathbf{g}}(\mathbf{w}) \stackrel{\text{def}}{=} \max_{d(a|x)} \min_{\omega} \sum_{a,x} c(\omega, a) d(a|x) w(x|\omega) \; .$$

Comparison of statistical models 1/2

$$\mathbf{w} = \langle \Omega, \mathcal{X}, w(x|\omega) \rangle$$

$$\begin{tikzcd}
\Omega \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{A} \arrow[d, squiggly] \\
\omega \arrow[r, "w(x|\omega)"'] & x \arrow[r, "d(a|x)"'] & a
\end{tikzcd}$$

$$\mathbf{w}' = \langle \Omega, \mathcal{Y}, w'(y|\omega) \rangle$$

$$\begin{tikzcd}
\Omega \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{Y} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{A} \arrow[d, squiggly] \\
\omega \arrow[r, "w'(y|\omega)"'] & y \arrow[r, "d'(a|y)"'] & a
\end{tikzcd}$$

For a fixed decision problem $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$, the payoffs $c^*_{\mathbf{g}}(\mathbf{w})$ and $c^*_{\mathbf{g}}(\mathbf{w}')$ can always be ordered (they are just real numbers).

Comparison of statistical models 2/2

Definition (Information Preorder)

If the model $\mathbf{w} = \langle \Omega, \mathcal{X}, w \rangle$ is better than model $\mathbf{w}' = \langle \Omega, \mathcal{Y}, w' \rangle$ *for all* decision problems $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$, that is,

$$c^*_{\mathbf{g}}(\mathbf{w}) \geqslant c^*_{\mathbf{g}}(\mathbf{w}'), \quad \forall \mathbf{g} \ ,$$

then we say that $\mathbf{w}$ is (always) more informative than $\mathbf{w}'$, and write

$$\mathbf{w} \succ \mathbf{w}' \ .$$

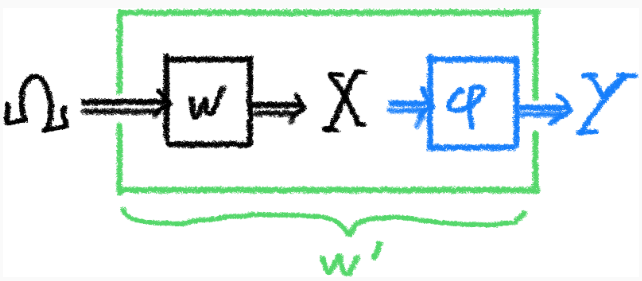
Can we visualize the information preorder more concretely?

Information preorder = statistical sufficiency

Theorem (Blackwell, 1953)

Given two statistical experiments $\mathbf{w} = \langle \Omega, \mathcal{X}, w \rangle$ *and* $\mathbf{w}' = \langle \Omega, \mathcal{Y}, w' \rangle$, *the following are equivalent:*

1. $\mathbf{w} \succ \mathbf{w}'$;
2. $\exists$ *cond. prob. dist.* $\varphi(y|x)$ *such that* $w'(y|\omega) = \sum_x \varphi(y|x) w(x|\omega)$ *for all* y *and* ω.



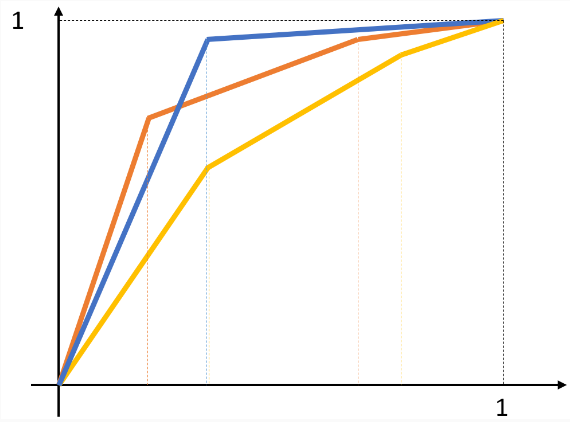
David Blackwell (1919-2010)

The case of dichotomies (a.k.a. relative majorization)

- for $\Omega = \{1, 2\}$, we compare two dichotomies, i.e., two pairs of probability distributions $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $(\boldsymbol{q}_1, \boldsymbol{q}_2)$, of dimension m and n, respectively
- relabel entries such that ratios p_1^i/p_2^i and q_1^j/q_2^j are nonincreasing
- construct the truncated sums $P_\omega(k) = \sum_{i=1}^k p_\omega^i$ and $Q_\omega(k) = \sum_{j=1}^k q_\omega^j$
- $(\boldsymbol{p}_1, \boldsymbol{p}_2) \succ (\boldsymbol{q}_1, \boldsymbol{q}_2)$ iff the relative Lorenz curve of the former is never below that of the latter

Blackwell, 1953

$(\boldsymbol{p}_1, \boldsymbol{p}_2) \succ (\boldsymbol{q}_1, \boldsymbol{q}_2) \iff \boldsymbol{q}_\omega = M\boldsymbol{p}_\omega$, for some stochastic matrix M.



$(x_k, y_k) = (P_2(k), P_1(k)), \quad 1 \leqslant k \leqslant n$

Quantum extensions

Quantum statistical decision theory (Holevo, 1973)

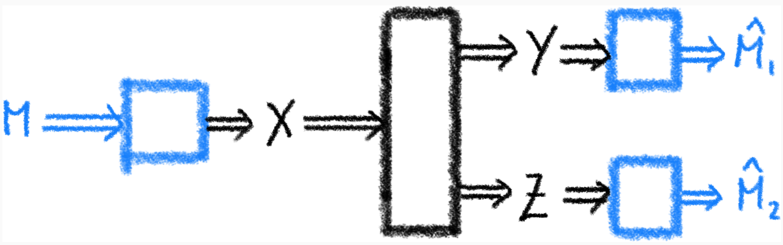
classical case	**quantum case**
• decision problems $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$	• decision problems $\mathbf{g} = \langle \Omega, \mathcal{A}, c \rangle$
• models $\mathbf{w} = \langle \Omega, \mathcal{X}, \{w(x\|\omega)\} \rangle$	• quantum models $\mathcal{E} = \langle \Omega, \mathcal{H}_S, \{\rho_S^\omega\} \rangle$
• decisions $d(a\|x)$	• POVMs $\{P_S^a : a \in \mathcal{A}\}$
• $c_\mathbf{g}^*(\mathbf{w}) = \max_{d(a\|x)} \min_\omega \cdots$	• $c_\mathbf{g}^*(\mathcal{E}) = \max_{\{P_S^a\}} \min_\omega \sum_a c(\omega, a) \operatorname{Tr}[\rho_S^\omega \; P_S^a]$

Quantum statistical comparison (FB, CMP 2012)

- let $\mathcal{E} = \langle \Omega, \mathcal{H}_S, \{\rho_S^\omega\} \rangle$ and $\mathcal{E}' = \left\langle \Omega, \mathcal{H}_{S'}, \{\sigma_{S'}^\omega\} \right\rangle$ be given
- information ordering: $\mathcal{E} \succ \mathcal{E}'$ iff $c_\mathbf{g}^*(\mathcal{E}) \geqslant c_\mathbf{g}^*(\mathcal{E}')$ for all $\mathbf{g}$
- *complete* information ordering: $\mathcal{E} \ggcurly \mathcal{E}'$ iff $\mathcal{E} \otimes \mathcal{F} \succ \mathcal{E}' \otimes \mathcal{F}$ for all ancillary models $\mathcal{F} = \left\langle \Theta, \mathcal{H}_A, \{\tau_A^\theta\} \right\rangle$
- **theorem**: $\mathcal{E} \succ \mathcal{E}'$ iff there exists a *quantum statistical morphism* $\mathcal{M} : \mathsf{L}(\mathcal{H}_S) \to \mathsf{L}(\mathcal{H}_{S'})$ such that $\mathcal{M}(\rho_S^\omega) = \sigma_{S'}^\omega$, for all $\omega \in \Omega$
- **theorem**: $\mathcal{E} \ggcurly \mathcal{E}'$ iff there exists a *completely positive trace-preserving linear map* $\mathcal{N} : \mathsf{L}(\mathcal{H}_S) \to \mathsf{L}(\mathcal{H}_{S'})$ such that $\mathcal{N}(\rho_S^\omega) = \sigma_{S'}^\omega$, for all $\omega \in \Omega$
- **theorem**: if $\mathcal{E}'$ is *commutative*, that is, if $[\sigma_{S'}^{\omega_1}, \sigma_{S'}^{\omega_2}] = 0$ for all $\omega_1, \omega_2 \in \Omega$, then $\mathcal{E} \ggcurly \mathcal{E}'$ iff $\mathcal{E} \succ \mathcal{E}'$

From the viewpoint of information theory

Classical broadcast channels



How to capture the idea that Y carries more information than Z?

(i) (stochastically) degradable: $\exists$ channel $Y \to Z$

(ii) less noisy: for all M, $H(M|Y) \leqslant H(M|Z)$

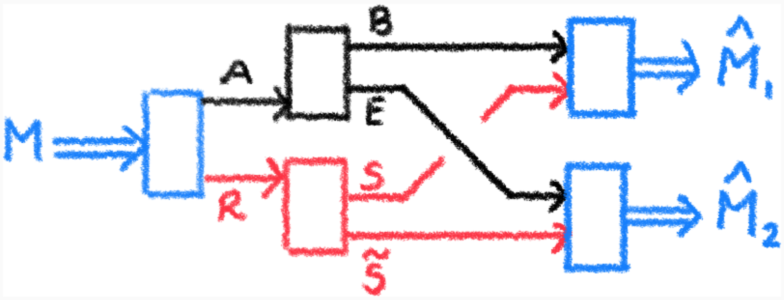
(iii) less ambiguous: for all M, $\max \mathbb{P}\{\hat{M}_1 = M\} \geqslant \max \mathbb{P}\{\hat{M}_2 = M\}$

(iv) less ambiguous (reformulation): for all M, $H_{\min}(M|Y) \leqslant H_{\min}(M|Z)$

Theorem (Körner–Marton, 1977; FB, 2016)

less noisy $\underset{\not\Rightarrow}{\Longleftarrow}$ *degradable* $\Longleftrightarrow$ *less ambiguous*

Quantum broadcast channels



(i) (CPTP) degradable: $\exists$ channel $B \to E$

(ii) *completely* less noisy: for all M and all symmetric side-channels $R \to S\tilde{S}$, $H(M|BS) \leqslant H(M|E\tilde{S})$

(iii) *completely* less ambiguous: for all M and all symmetric side-channels $R \to S\tilde{S}$, $H_{\min}(M|BS) \leqslant H_{\min}(M|E\tilde{S})$

Theorem (FB–Datta–Strelchuk, 2014)

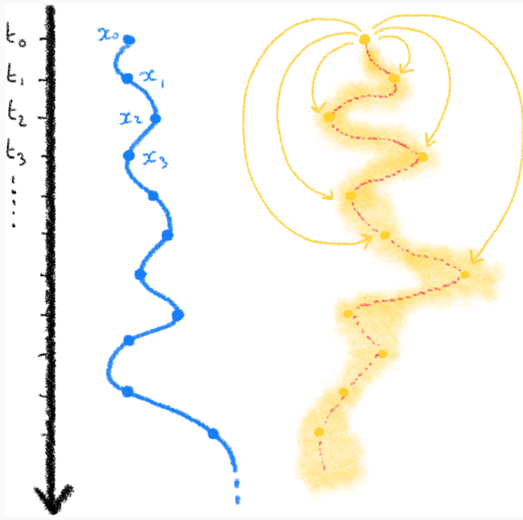
completely less noisy $\underset{\not\Rightarrow}{\Longleftarrow}$ *degradable* $\Longleftrightarrow$ *completely less ambiguous*

Applications in open quantum systems dynamics

Discrete-time stochastic processes

Formulation of the problem:

- for $i \in \mathbb{N}$, let x_i index the state of a system at time $t = t_i$
- given the system's initial state at time $t = t_0$, the process is fully predicted by the conditional distribution $p(x_N, \ldots, x_1 | x_0)$
- if the system evolving is quantum, we only have a quantum dynamical mapping $\left\{\mathcal{N}^{(i)}_{Q_0 \to Q_i}\right\}_{i \geqslant 1}$
- the process is divisible if there exist channels $\mathcal{D}^{(i)}$ such that $\mathcal{N}^{(i+1)} = \mathcal{D}^{(i)} \circ \mathcal{N}^{(i)}$ for all $i \geqslant 1$
- **problem**: to provide a *fully information-theoretic characterization* of divisibility

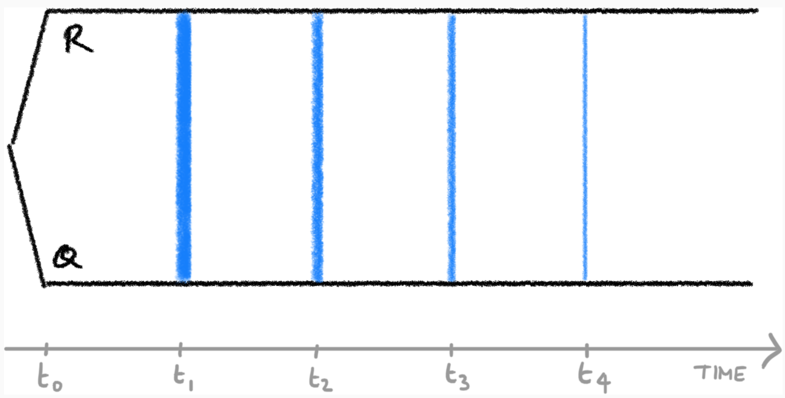


Divisibility as “information flow”

Theorem (FB–Datta, 2016; FB, 2018)

Given an initial open quantum system Q_0, a quantum dynamical mapping $\left\{\mathcal{N}^{(i)}_{Q_0 \to Q_i}\right\}_{i \geqslant 1}$ is divisible if and only if, for any initial state ω_{RQ_0},

$$H_{\min}(R|Q_1) \leqslant H_{\min}(R|Q_2) \leqslant \cdots \leqslant H_{\min}(R|Q_N) \ .$$



Applications in quantum thermodynamics

Quantum thermodynamics from relative majorization

Basic idea (FB, arXiv:1505.00535)

Thermal accessibility $\rho \to \sigma$ can be characterized as the statistical comparison between quantum dichotomies (ρ, γ) and (σ, γ), for γ thermal state

Two main problems:

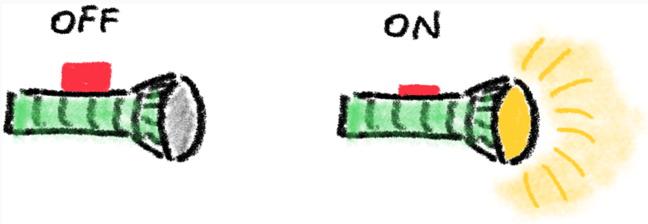
- for dimension larger than 2 and $[\sigma, \gamma] \neq 0$, we need a complete (i.e., extended) comparison
- moreover, Gibbs-preserving channels can create coherence between energy levels, while a truly thermal operation should not

Complete comparison of quantum dichotomies 1/2

Definition (ON/OFF channels)

Given a d-dimensional quantum dichotomy $\mathcal{E} = (\rho, \gamma)$, we define the corresponding ON/OFF channel $\mathcal{N}_\mathcal{E} : \mathscr{L}(\mathbb{C}^2) \to \mathscr{L}(\mathbb{C}^d)$ as

$$\mathcal{N}_\mathcal{E}(\cdot) := \gamma \langle 0| \cdot |0\rangle + \rho \langle 1| \cdot |1\rangle$$

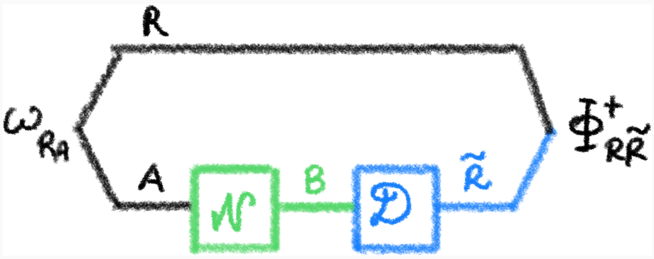


Complete comparison of quantum dichotomies 2/2

For a quantum channel $\mathcal{N} : A \to B$ and a state ω_{RA}, define the singlet fraction as

$$\Phi^*_\omega(\mathcal{N}) := \max_{\mathcal{D}:B\to\tilde{R}} \langle \Phi^+_{R\tilde{R}} | (\mathsf{id}_R \otimes \mathcal{D} \circ \mathcal{N})(\omega_{RA}) | \Phi^+_{R\tilde{R}} \rangle \, ,$$

where $\mathcal{D}$ is a decoding quantum channel with output system $R \cong \tilde{R}$



Theorem (FB, 2015)

Given two quantum dichotomies $\mathcal{E} = (\rho_1, \rho_2)$ and $\mathcal{F} = (\sigma_1, \sigma_2)$, let $\mathcal{N}_\mathcal{E}$ and $\mathcal{N}_\mathcal{F}$ the corresponding ON/OFF channels. Then, $\mathcal{E} \gg \mathcal{F}$ if and only if

$$\Phi^*_\omega(\mathcal{N}_\mathcal{E}) \geqslant \Phi^*_\omega(\mathcal{N}_\mathcal{F}) \, , \quad \forall \omega$$

Dealing with quantum coherence (sketch)

For quantum dichotomies $\mathcal{E} = (\rho, \gamma)$ and $\mathcal{F} = (\sigma, \gamma)$ and group $\mathscr{T} = \{e^{-it\log\gamma}\}_{t\in\mathbb{R}}$, we write $\mathcal{E} \succcurlyeq_{\mathscr{T}} \mathcal{F}$ iff $\exists$ CPTP linear $\mathcal{M}$ such that:

(i) $\mathcal{M}(\rho) = \sigma$ and $\mathcal{M}(\gamma) = \gamma$;

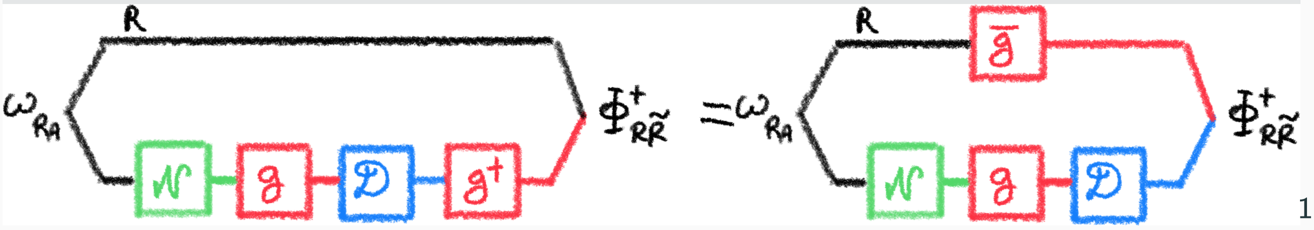
(ii) $\mathcal{M}(U_t \cdot U_t^\dagger) = U_t\mathcal{M}(\cdot)U_t^\dagger$, for all $t \in \mathbb{R}$

Theorem (Gour–Jennings–FB–Duan–Marvian, 2018)

$\mathcal{E} \succcurlyeq_{\mathscr{T}} \mathcal{F}$ if and only if

$$\widetilde{\Phi}^*_\omega(\mathcal{N}_\mathcal{E}) \geqslant \widetilde{\Phi}^*_\omega(\mathcal{N}_\mathcal{F})\ , \quad \forall\omega$$

(see picture below)



Conclusions

Conclusions

- the theory of statistical comparison studies morphisms (preorders) of one "statistical system" X into another "statistical system" Y
- equivalent conditions are given in terms of (finitely or infinitely many) monotones, e.g., $f_i(X) \geqslant f_i(Y)$
- such monotones quantify the resources at stake in the operational framework at hand, e.g.
    - the expected maximin payoff in decision problems for experiments
    - the information asymmetry for broadcast channels
    - the non-divisibility for open systems dynamics
    - the joint time-energy information for quantum thermodynamics

Thank you