# Paradigms of cognition

Flemming Topsøe

Department of Mathematical Sciences
University of Copenhagen

Presentation at the workshop

Mathematical Aspects of
Generalized Entropies
and their Applications

Kyoto, July 7-9, 2009

*What you will hear about*

Part I
- Nature
- Observer
- worlds
- situations
- truth
- belief
- knowledge
- interactor
- descriptor

Part II
- experiments
- observations
- games
- preparations
- exponential family
- equilibrium

# Part I

## *Nature and Observer*

Nature and Observer interact in a world;
Nature holds the truth;
Observer seeks the truth but is confined to belief;
A situation involves observations from an experiment;
In any situation, Observer strives for knowledge.

| |
|---|
| **A view and an assumption:** knowledge is the synthesis of extensive experience – and knowledge can be derived from truth and belief. |

Notation pertaining to any given situation:
$x$ for a truth instance
$y$ for a belief instance
$z$ for a knowledge instance

| |
|---|
| **In more detail, we assume:** There is a function $\Pi$, the (global) interactor such that $z = \Pi(x, y)$. |

A world, $\mathcal{W}$, is often characterized only by the interactor and we write $\mathcal{W} = \mathcal{W}_{\Pi}$.

## *Examples of worlds*

The interactor

$$\sqcap_1(x,y) = x$$

defines a classical world, $\mathcal{W}_1$. In this world, "truth is observable" or you may say that "truth is learnable". It may be called the Boltzmann-Gibbs-Shannon world. As another extreme,

$$\sqcap_0(x,y) = y$$

defines a black hole, $\mathcal{W}_0$. In this world "what you see is what you believe" (WYSIWYB). No reflection of a truth which lies outside you can be observed.

If the set of possible truth instances and the set of possible belief instances are both embeddable in the same linear space, we can also consider, for any real parameter $q$, the interactor

$$\sqcap_q(x,y) = qx + (1-q)y\,;$$

this defines the Tsallis world $\mathcal{W}_q$.

*control, descriptor, description effort*

The focus of Observer could be, either

□ speculative, directed at the truth: **"what could the truth be?"** Task: to determine, in a Bayesean way, say based on prior knowledge a good belief instance

or

□ constructive, directed at the question: **"What can I do about it?"** Task: design of experiments, aiming at only having to allocate a low effort on the way to knowledge when making observations associated with the suggested experiments.

Regarding the second task: key objects we call control instances or simply controls ($w$). They should tell Observer how he can "control" a given situation. Their determination depends on an overall strategy for description, which will, typically, be adapted to the world once and for all. (more details next slide)

. . . . . . . . .

NOTE: Other names could be codes, code length functions, description instances – reason for my choice will be clear later.

Key assumptions: Many overall strategies for observation are available. Each such strategy is called a descriptor ($\kappa$). Having selected $\kappa$ (by adapting it to the world), there is a bijection between belief instances and control instances, written

$$y \overset{\kappa}{\longleftrightarrow} w \quad \text{or} \quad w = \widehat{y} \quad \text{or} \quad y = \breve{w} \,.$$

For each descriptor, there is a function $\Phi$, description effort (or cost), which determines the effort required by Observer in any situation with truth- and belief instances $x$ and $y$ when using the overall strategy $\kappa$. This is denoted by either of the two expressions

$$\Phi_\Pi(x, y|\kappa) \quad \text{or} \quad \Psi_\Pi(x, w|\kappa) \,,$$

it being assumed that $y \overset{\kappa}{\longleftrightarrow} w$.

We ought to write $\mathcal{W}_{\Pi,\Phi}$ or $\mathcal{W}_{\Pi,\kappa}$ when characterizing the world. However, for examples considered, we shall see that $\kappa$ and then also $\Phi$ are uniquely determined in a natural way from the interactor $\Pi$. Thus we only need the notation $\mathcal{W}_\Pi$ (or $\mathcal{W}$ if $\Pi$ is understood).

*How can Observer select $\kappa$? PMP!*

All interactors which we will consider will be sound, i.e. $\sqcap(x,y) = x$ when belief matches truth ($y = x$).

**A variational principle:** Consider $\mathcal{W}_\sqcap$. Among all descriptors $\kappa$, Observer should choose one which satisfies the perfect match principle (PMP): *description effort should be the least when belief matches truth*, i.e. $\Phi_\sqcap(x,y|\kappa) \geq \Phi_\sqcap(x,x|\kappa)$ should hold.

Equivalently, PMP says $\Psi_\sqcap(x,\widehat{y}) \geq \Psi_\sqcap(x,\widehat{x})$. If $\exists\, \kappa$ unique s.t. PMP holds, $\kappa$ is the ideal descriptor for $\mathcal{W}_\sqcap$. Then write $\Phi_\sqcap(x,y)$ and $\Psi_\sqcap(x,w)$ (drop $\kappa$) and define entropy, divergence and redundancy by

$H_\sqcap(x) = \Phi_\sqcap(x,x)$ (minimal description effort),
$D_\sqcap(x,y) = \Phi_\sqcap(x,y) - H_\sqcap(x)$,
$R_\sqcap(x,w) = \Psi_\sqcap(x,w) - H_\sqcap(x)$.

For divergence, PMP says $D_\sqcap(x,y) \geq 0$. This is the fundamental inequality (FI). The world $\mathcal{W}_\sqcap$ is regular if $D_\sqcap(x,y) = 0 \Leftrightarrow y = x$,. If not, $\mathcal{W}_\sqcap$ is singular.

## *finally: examples!*

**Relativization in a geometrical model:** Simple models are obtained in a classical world when truth-, belief- and control-instamces are of the same type and the ideal descriptor is the identity map ($\widehat{y} = y$ for all $y$). As an example, take $x$- and $y$-instances as elements of a Hilbert space and let $\Phi$ measure squared distance relative to a prior $y_0$:

$$\Phi(x, y) = \|x - y\|^2 - \|x - y_0\|^2 \, .$$

(If you replace $y$ by $\widehat{y}$ and only consider surjective maps $y \mapsto \widehat{y}$ as descriptors, the identity map is eaily seen to be the only descriptor satisfying the PMP).

For this example,
$H(x) = -\|x - y_0\|^2$ and $D(x, y) = \|x - y\|^2$.
Maximizing entropy is the same as minimizing the distance to $y_0$.

## Probabilistic modeling, the world $\mathcal{W}_{\pi,\kappa}$

Now, $x$'s and $y$'s are probability distributions over an alphabet $\mathcal{A}$, $z$'s are functions over $\mathcal{A}$. A technical assumption: $x_i > 0 \Rightarrow y_i > 0$. The local interactor $\pi$ determines $\Pi$ by $\big(\Pi(x,y)\big)_i = \pi(x_i, y_i)$ for $i \in \mathcal{A}$. The descriptor $\kappa$ is smooth on $[0,1]$, $\kappa(1) = 0$ and $\kappa'(1) = -1$ (normalization). We assume that $\pi$ is sound, $(\pi(s,s) = s)$, finite on $[0,1]\times]0,1]$, smooth on $]0,1[\times]0,1[$, continuous on $[0,1]^2 \setminus \{(0,0)\}$. Below, $z_i = \pi(x_i, y_i)$ and $w_i = \kappa(y_i)$.

| $\mathcal{A}$ | $x$ truth | $y$ belief | $z$ knowledge | control $w$ (codelgth.) | local effort $\phi$ |
|---|---|---|---|---|---|
| · | · | · | · | · | · |
| $i$ | $x_i$ | $y_i$ | $z_i$ | $w_i$ | $z_i w_i$ |
| · | · | · | · | · | · |

$$\Phi(x, y|\kappa) = \sum_{i\in\mathcal{A}} z_i w_i = \sum_{i\in\mathcal{A}} \pi(x_i, y_i)\kappa(y_i)$$

Local effort is the function $\phi(s,t) = \pi(s,t)\kappa(t)$.

NOTE: No reference to coding!
Let us remind ourselves how things look *with coding* for the BGS-world and the standard descriptor (measuring in bits).

| **1.** | $i$ | $x$ | $y$ | code | $w$ | $\phi$ |
|---|---|---|---|---|---|---|
| | a | 1:2 | 1:4 | 00 | 2 bit | 1 |
| | b | 1:4 | 1:4 | 01 | 2 bit | 1/2 |
| | c | 1:8 | 1:4 | 10 | 2 bit | 1/4 |
| | d | 1:8 | 1:4 | 11 | 2 bit | 1/4 |

| **2.** | $i$ | $x$ | $y$ | code | $w$ | $\phi$ |
|---|---|---|---|---|---|---|
| | a | 1:2 | 1:2 | 0 | 1 bit | 1/2 |
| | b | 1:4 | 1:4 | 10 | 2 bit | 1/2 |
| | c | 1:8 | 1:8 | 110 | 3 bit | 3/8 |
| | d | 1:8 | 1:8 | 111 | 3 bit | 3/8 |

Note that (total) description effort, $\Phi$, is the average of the $w$'s and the sum of the $\phi$'s.

Case **1**: $\Phi = 2$.
Case **2**: $\Phi = \frac{7}{4}$, in fact, this is optimal, so $H = \frac{7}{4}$ when the true probability vector is as stated.
The above points to the fact that with descriptor $t \mapsto \log \frac{1}{t}$, PMP holds for the classical BGS-world.

## *PMP for $\mathcal{W}_\pi$*

There are many possible interactors, but typically, they fall in families. The $q$-family consists of interactors of the form:

$$\pi_q^\xi(s,t) = \xi^{-1}\Big(\pi_q\big(\xi(s), \xi(t)\big)\Big)$$

for some one-to-one function $\xi$. For $\xi$ the identity we find $\pi_q(s,t) = qs + (1-q)t$, for $\xi$ the natural logarithm we find $\pi_q^G(s,t) = s^q t^{1-q}$.

An interactor is consistent if $\sum_i z_i = 1$ for all probability vectors $x$ and $y$ with $z = \Pi(x,y)$.

| If $\pi$ is consistent, then $\pi \equiv \pi_q$ for some $q \in \mathbb{R}$ |
|---|

The interactors $\pi_q^\xi$ are $\xi$-consistent. (just a definition).

Theorem. Let $\pi$ be an interactor, denote by $\chi$ the function on $]0,1[$ defined by

$$\chi(t) = \frac{\partial \pi}{\partial t}(t,t)$$

and assume that $\chi$ is bounded in the vicinity of $t = 1$. Then, there can only be one descriptor such that PMP holds, viz., in $]0,1[$, the solution to the differential equation

$$\chi(t)\kappa(t) + t\kappa'(t) = -1$$

for which $\kappa(1) = \lim_{t\to 1} \kappa(t) = 0$.

**Proof** Assume $(\pi, \kappa)$ satisfies PMP. Put

$$f(t) = \chi(t)\kappa(t) + t\kappa'(t)\,.$$

Consider a fixed probability vector $x = (x_1, x_2, x_3)$ with positive point probabilities. By PMP, $F$ given by

$$F(y) = F(y_1, y_2, y_3) = \sum_1^3 \pi(x_i, y_i)\kappa(y_i)$$

on $]0,1[\times]0,1[\times]0,1[$ assumes its minimal value for the interior point $y = x$ when restricted to probability distributions. As standard regularity conditions are fulfilled, there exists a Lagrange multiplier $\lambda$ such that

$$\frac{\partial}{\partial y_i}\Big(F(y) - \lambda \sum_1^3 y_i\Big) = 0 \text{ for } i = 1, 2, 3$$

when $y = x$. This shows that $f(x_1) = f(x_2) = f(x_3)$.

Take $(x_1, x_2, x_3) = (\frac{1}{2}, x, \frac{1}{2} - x)$ for $x \in ]0, \frac{1}{2}[$ and conclude that $f$ is constant on $]0, \frac{1}{2}]$. Then consider a value $x \in ]\frac{1}{2}, 1[$ and the probability vector $(x, \frac{1}{2}(1 - x), \frac{1}{2}(1-x))$ and conclude that $f(x) = f(\frac{1}{2}(1-x))$. As $0 < \frac{1}{2}(1-x) < \frac{1}{2}$, we conclude that $f(x) = f(\frac{1}{2})$. Thus $f$ is constant on $]0, 1[$. By letting $t \to 1$ in the differential equation, we conclude that the value of the constant is $-1$. □

Given $(\pi,\kappa)$, the divergence generator is the function $\delta = \delta_{\pi,\kappa}$ given by

$$\delta(s,t) = \Big(\pi(s,t)\kappa(t) + t\Big) - \Big(s\kappa(s) + s\Big).$$

By the pointwise fundamental inequality, PFI, we understand that $\delta(s,t) \geq 0$ holds for every $(s,t) \in [0,1]\times]0,1]$. When so, then, for every $(x,y)$,

$$\begin{aligned}\Phi_\pi(x,y|\kappa) + 1 &= \sum_{\{i|y_i>0\}} \Big(\pi(x_i,y_i)\kappa(y_i) + y_i\Big)\\ &\geq \sum_{\{i|y_i>0\}} \Big(\pi(x_i,x_i)\kappa(x_i) + x_i\Big)\\ &= \sum_{i\in\mathbb{A}} \Big(\pi(x_i,x_i)\kappa(x_i) + x_i\Big)\\ &= \Phi_\pi(x,x|\kappa) + 1\,,\end{aligned}$$

and we conclude that PMP holds.

So PMP follows from PFI.

**Conjecture:** The converse is also true.

The close relation btw. PMP and PFI makes us define adjusted notions of local as well as total description effort:

$$\tilde{\phi}(s,t) = \phi(s,t) + t$$
$$\tilde{\Phi}(x,y) = \sum_{i \in \mathbb{A}} \tilde{\phi}(x_i, y_i)\,.$$

The added term, $t$, in $\tilde{\phi}$ is interpreted as the contribution to the total overhead due to a basic event with believed probability $t$. Total overhead is always $\sum y_i = 1$. In other words, the normalization $\kappa'(1) = -1$ implies that overhead cost is the unit we work with. Adjusting also the entropy function, one finds that adjusted entropy is always bounded below by the overhead cost, 1 nat.

**OBS:** Knowing the descriptor does not determine the world. Several interactors give the same descriptor. Without going into details, we mention that the interactors of the form $\pi_q^\xi$ determine the same ideal descriptor as $\pi_q$ ($q$ fixed $\geq 0$).

## The Tsallis family

The deformed logarithms $\ln_q$ are (Tsallis 1994):

$$\ln_q t = \begin{cases} \ln t \text{ if } q = 1 \\ \frac{1}{1-q}\left(t^{1-q} - 1\right) \text{ otherwise}. \end{cases}$$

**Theorem** Assume $\pi$ is consistent. Then, $\pi = \pi_q$ with $q = \pi(1,0)$, hence $\mathcal{W}_\pi = \mathcal{W}_q$. If $q < 0$, PMP fails whatever the descriptor. If $q \geq 0$, ideal descriptor is

$$\kappa_q(y) = \ln_q \frac{1}{y}.$$

All worlds $\mathcal{W}_q$ with $q > 0$ are regular, $\mathcal{W}_0$ is singular, in fact, $D_0 \equiv 0$.

**Proof** The first part we know already. Fix $q$. Then $\chi = 1 - q$. Solving the differential equation, we find that only $\kappa_q$ could work. If $q < 0$, PMP fails by easy examples. To check PMP when $q \geq 0$, we verify PFI. For $q = 0$, $\delta_0(s,t) = 0$ when $t > 0$, hence PFI holds (but $\delta_0(s,0) = -1$ for $s > 0$). Then assume $q > 0$. PFI for $q = 1$ is classical. For remaining cases, write

$$\delta_q(s,t) = \frac{q}{1-q} s t^{q-1} + t^q - \frac{1}{1-q} s^q \,.$$

Apply the GA-inequality and PFI follows (consider the cases $0 < q < 1$ and $q > 1$ separately and collect the two positive terms). □

**Key formulas for Tsallis case are really:**

$\pi_q(s,t) = qs + (1-q)t$ and $\kappa_q(t) = \ln_q \frac{1}{t}$

in connection with the general formulas:

$\Phi(x,y) = \sum_i \pi(x_i,y_i)\kappa(y_i)$, $H(x) = \sum_i x_i \kappa(x_i)$

$D(x,y) = \Phi(x,y) - H(x) = \sum_i \delta(x_i,y_i)$ with

$\delta(s,t) = \Big(\pi(s,t)\kappa(t) + t\Big) - \Big(s\kappa(s) + s\Big)$.

Of course, if you insist, here are the concrete formulas:

$$\Phi_q(x,y) = \sum_{i\in\mathbb{A}} \left(\tfrac{q}{1-q} x_i y_i^{q-1} + y_i^q - \tfrac{1}{1-q} x_i\right)$$

$$\mathrm{H}_q(x) = \tfrac{1}{1-q} \sum_{i\in\mathbb{A}} (x_i^q - x_i) = \tfrac{1}{1-q}\left(\sum_{i\in\mathbb{A}} x_i^q - 1\right)$$

$$D_q(x,y) = \sum_{i\in\mathbb{A}} \left(\tfrac{q}{1-q} x_i y_i^{q-1} + y_i^q - \tfrac{1}{1-q} x_i^q\right)$$

## *On the significance of $\kappa$*

For a Tsallis world, the ideal descriptor can be characterized in two ways:

• Direct view: Given $t \in [0,1]$, $\kappa(t)$ is the effort needed in nats in order to describe an event with probability $t$.

• Reversed view: Consider $\text{pr} : [0,\infty[ \to [0,1]$, defined as the inverse function of $\kappa$ (extended properly if $\kappa(0) < \infty$). Call this function either the $\kappa$-reciprocal-exponential or the $\kappa$-probability checker. Then, given $a \geq 0$, you can ask the question: "how complex events can I describe with access to $a$ nats?". The lower probability, the more complex the event. The answer is: You can describe any event with a probability $\geq \text{pr}(a)$.

# Part II

## *Nature versus Observer*

**The setting:**

A regular world $\mathcal{W}$ with interactor $\Pi$ and descriptor satisfying PMP. Description effort is $\Phi(x, y)$ or, better, in terms of controls, $\Psi(x, w)$.

Consider a two-person zero-sum game with Nature and Observer as players and with truth- and control instances as available strategies. They fight over the objective function, taken to be description effort $\Psi(x, w)$, with Nature as maximizer and Observer as minimizer.

The values of the game for, respectively Nature and Observer are

$$\sup_x \inf_w \Psi(x, w) \quad \text{and} \quad \inf_w \sup_x \Psi(x, w) \, .$$

But what are the sets of strategies over which sup's and inf's are taken?

We assume that control instances range over a fixed set, $K$, the Observer strategies, whereas the truth instances may vary over a set, the preparation $\mathcal{P}$, which depends on the situation. It is the strategy set for Nature. Note that the value for Nature is the maximum entropy value

$$H_{\text{max}}(\mathcal{P}) = \sup_{x \in \mathcal{P}} H(x)\,. \tag{1}$$

The value for Observer is the minimal risk value

$$R_{\text{min}}(\mathcal{P}) = \inf_{w} R(w) \quad \text{with} \quad R(w|\mathcal{P}) = \sup_{x \in \mathcal{P}} \Psi(x, w)\,.$$

Notation: $\gamma(\mathcal{P})$ for the game considered.

Note that $H_{\text{max}}(\mathcal{P}) \leq R_{\text{min}}(\mathcal{P})$, the minimax inequality. If "$=$" holds (and is finite), the game is in game theoretical equilibrium (GTE). An optimal strategy for Nature is a truth instance in $\mathcal{P}$ with maximal entropy. An optimal strategy for Observer is a control $w \in K$ with $R(w) = R_{\text{min}}$.

Another concept of equilibrium: A control $\varepsilon^*$ is robust if, for some $h \in \mathbb{R}$, $\Psi(x,\varepsilon^*) = h$ for all $x \in \mathcal{P}$; and $h$ is the level of robustness. Important connection:

**Robustness lemma** If $x^* \in \mathcal{P}$ and $\varepsilon^* = \widehat{x^*}$ is robust with level $h$, then GTE holds for $\gamma(\mathcal{P})$. The value of $\gamma(\mathcal{P})$ is $h$ and the Pythagorean inequalities hold:

$$\forall x \in \mathcal{P} \;:\; H(x) + R(x,\varepsilon^*) \le H_{\text{max}}(\mathcal{P})$$
$$\forall w \in K \;:\; R(w) \ge H_{\text{max}}(\mathcal{P}) + R(x^*, w)\,.$$

**Proof** [Really, an easy consequence of Nash's saddle value inequalities] By assumption, $R(\varepsilon^*) = h = \Psi(x^*,\varepsilon^*) = H(x^*)$, hence GTE holds with $x^*$ and $\varepsilon^*$ as optimal strategies. For any $x \in \mathcal{P}$, $H(x) \le H(x) + R(x,\varepsilon^*) = \Psi(x,\varepsilon^*) = h$ and, for the other inequality, $R(w) \ge \Psi(x^*,w)$, and result follows as $\Psi(x,w) = H(x) + R(x,w)$ (the linking identity). □

Standard form of Pythagorean inequality:
$H(x) + D(x,x^*) \le H_{\text{max}}(\mathcal{P})$ (Chentsov, Csiszár).

## *What* can *you know?*

"*which are the preparations Observer can realize i.e. enforce on Nature?*
"*which are the experiments, Observer can perform?*"

**Answer:** Basicly, Observer can choose one or more controls and select associated levels. With just one choice of $w$ and $h$, the preparation is the level set

$$L^w(h) = \{x|\Psi(x,w) = h\} = \{x|\Psi^w = h\}\,,$$

with $\Psi^w$ for the marginal function.

Roughly: Observer decides on a way of looking at the world via $w$. He uses $w$ to arrange an experimental set-up consisting of machinery, instruments and so on, including a special handle which he uses to fix the level of effort, $h$. This restricts the truth instances to the set $L^w(h)$. The scene is set, and observations can begin with the reading of measuring instruments etc. Observer may use the same experimental set-up for several experiments by using the handle to fix a desired level.

We take the non-empty finite intersections of level sets to constitute the family of feasible preparations. For simplicity we restrict attention to the level sets themselves (genus-1 type feasible preparations).

Why do the level sets play a central role? Because 1) they allow robustness considerations, 2) because sub-level sets do. These sets are defined by

$$\mathsf{SL}^w(h) = \{x|\Psi(x,w) \le h\} = \{\Psi^w \le h\}\,.$$

**maximal preparations** Assume that $\mathcal{W}_\Pi$ is regular. Consider $x^*$ and $w^*$. Then GTE holds for some $\gamma(\mathcal{P})$ with $x^*$ and $w^*$ as optimal strategies **iff** $h^* = \Psi(x^*,w^*) < \infty$ and $w^* = \widehat{x^*}$. If so, the largest such set is $\mathsf{SL}^{w^*}(h^*)$.

**Proof** By Nash' saddle value inequalities ,if $\mathcal{P}$ works,

$$\forall x \in \mathcal{P} \forall w \in K: \ \Psi(x, w^*) \leq \Psi(x^*, w^*) \leq \Psi(x^*, w)$$

and $\Psi(x^*, w^*)$ is finite. From right hand inequality and regularity of $\mathcal{W}_{\sqcap}$, $w^* = \widehat{x^*}$. The left hand inequality says that $\mathcal{P} \subseteq \mathsf{SL}^{w^*}(h^*)$. That all properties hold for $\mathcal{P} = \mathsf{SL}^{w^*}(h^*)$ follows from sufficiency of the saddle value inequalities and PMP. □

To make ideas precise, let $w$ be a control and denote by $\mathcal{L}^w$ the family of non-empty sets of the form $L^w(h)$. The associated exponential family , denoted $\widehat{\mathcal{E}}^w$ is the set of controls $\varepsilon$ which are robust for all preparations in $\mathcal{L}^w$. In terms of belief instances this is the family $\mathcal{E}^w$ of all belief instances $x^*$ which match one of the controls in $\mathcal{E}^{\mathrm{w}}$ ($x^* = \breve{\varepsilon}$ for some $\varepsilon \in \mathcal{E}^{\mathrm{w}}$).

From definitions and the robustness lemma you find:

Consider a preparation family $\mathcal{L}^w$. Let $x^*$ be a truth instance, put $\varepsilon^* = \widehat{x^*}$ and assume that $\varepsilon^* \in \widehat{\mathcal{E}}^w$. Put $h = \Psi(x^*, w)$. Then $\gamma(L^w(h))$ is in GTE and has $x^*$ and $\varepsilon^*$ as optimal strategies. In particular, $x^*$ is the maximum entropy strategy for the preparation $L^w(h)$.

**Example 1: geometry**

Recall: Hilbert space, classical world, prior $y_0$, $y \overset{\kappa}{\longleftrightarrow} w$ means $w = y$, $\Phi(x, y) = \|x - y\|^2 - \|x - y_0\|^2$, $H(x) = -\|x - y_0\|^2$ and $D(x, y) = \|x - y\|^2$.

Fix $w \neq y_0$. Then $\mathcal{L}^w$ consists of all hyperplanes with $wy_0$ as normal and $\widehat{\mathcal{E}}^w = \mathcal{E}^w$ consists of all $y$ on the line determined by $w$ and $y_0$. In this case our theorems give the standard results on projections on a hyperplane and the standard Pythagorean (in)equalities.

**Example 2: Tsallis world $\mathcal{W}_q$**

Below, the index $q$ is suppressed. Fix $y \stackrel{\kappa}{\longleftrightarrow} w$. Then $\mathcal{L}^w$ consists of all preparations for which $\Psi^w$ is constant, i.e. all preparations $\mathcal{P}$ of the form

$$\begin{aligned} \mathcal{P} &= \{x|\exists h : \Psi(x,w) = h\} \\ &= \{x|\exists h : \sum_i \Big(qx_i + (1-q)y_i\Big)w_i = h\} \\ &= \{x|\exists c : \Big(qx + (1-q)y\Big) \cdot w = c\} \\ &= \{x|\exists c : x \cdot w^* = c\} \end{aligned}$$

For $x \in \mathcal{P}$, also $x \cdot 1 = 1$ holds. Thus, for any $\alpha, \beta$, $x \cdot (\alpha + \beta w)$ is constant over $\mathcal{P}$.

Conclusion: all controls of the form $\alpha + \beta w$ are in $\widehat{\mathcal{E}}^w$

Let $\varepsilon = \alpha + \beta w$ and let $\rho \stackrel{\kappa}{\longleftrightarrow} \varepsilon$. We must insist that $\sum_i \rho_i = 1$, i.e. that

$$\sum_i \mathsf{pr}\left(\alpha + \beta w_i\right) = 1\,.$$

The value of $\alpha$ for which this is true (if...) we denote by $\alpha = \mathsf{LNZ}(\beta)$. Write

$\varepsilon(\beta) = \mathsf{LNZ}(\beta) + \beta w$ and adjust $\beta$ such that

$$\sum_i \mathsf{pr}\left(\mathsf{LNZ}(\beta) + \beta w_i\right) w_i$$

has the appropriate value . ...

## *What have we achieved?*

- provided a transparent interpretation of Tsallis entropy
- developed a basis for an abstract theory
- provided elements for establishing a bridge to information geometry (?)
- clarifyed role of FI via PMP; focus on PFI as the natural basis for establishing FI and hence PMP
- identified the unit of entropy as an overhead
- answered the question “what *can* we know”
- found good (*the* right ?) definition of an exponential family
- Stressed dual role of preparations and exponential families
- brought games into the picture, thereby showing how Nash’s general results pave the way to equilibrium and optimal strategies (even without introducing Lagrange multipliers)
- separated Nature from Observer in key expressions

## *... and what can we still ask?*

- interaction, how?
- how can we know the world we work in?
- control, how?
- coding interpretation possible ?
- ...