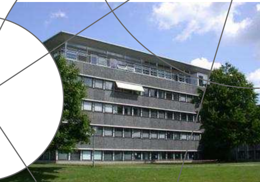
Philosophy of Information

Flemming Topsøe, topsoe@math.ku.dk
Department of Mathematical Sciences, University of Copenhagen



Institut Henri Poincare, Paris, 3-5. december, 2009

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability t, you *believe* the event has: $\kappa(t)$.

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability t, you *believe* the event has: $\kappa(t)$. κ is the individual effort (effort-function) or the descriptor.

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability t, you *believe* the event has: $\kappa(t)$. κ is the individual effort (effort-function) or the descriptor.

effort $\longleftrightarrow$ description ?

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability t, you *believe* the event has: $\kappa(t)$. κ is the individual effort (effort-function) or the descriptor.

effort $\longleftrightarrow$ description ?

Requirements: $\kappa(1) = 0$, κ is smooth (and decreasing).

the cost of information

What is the cost of information or, how much are you willing to pay – or *have* to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability t, you *believe* the event has: $\kappa(t)$. κ is the individual effort (effort-function) or the descriptor.

effort $\longleftrightarrow$ description ?

Requirements: $\kappa(1) = 0$, κ is smooth (and decreasing). Further, natural with normalization via the differential cost $\iota = -\kappa'(1)$. If $\iota = 1$, we obtain natural units, nats; if $\iota = \ln 2$, we measure in binary units, bits.

accumulated effort (corresp. to negative score)

Consider distributions over a discrete alphabet $\mathbb{A}$: $x = (x_i)_{i \in \mathbb{A}}$ representing truth , $y = (y_i)_{i \in \mathbb{A}}$ representing belief .

accumulated effort (corresp. to negative score)

Consider distributions over a discrete alphabet $\mathbb{A}$: $x = (x_i)_{i\in\mathbb{A}}$ representing truth , $y = (y_i)_{i\in\mathbb{A}}$ representing belief .
Accumulated effort (expected per observation) is

$$\Phi(x, y) = \sum_{i\in\mathbb{A}} x_i \kappa(y_i)\,.$$

accumulated effort (corresp. to negative score)

Consider distributions over a discrete alphabet $\mathbb{A}$: $x = (x_i)_{i \in \mathbb{A}}$ representing truth, $y = (y_i)_{i \in \mathbb{A}}$ representing belief.
Accumulated effort (expected per observation) is

$$\Phi(x, y) = \sum_{i \in \mathbb{A}} x_i \kappa(y_i) .$$

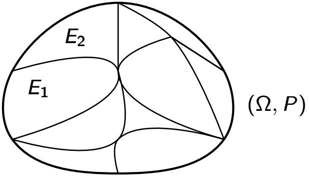
Theorem There is only one descriptor, the classical descriptor, for which the perfect match principle holds, i.e. for which

$$\Phi(x, y) \geq \Phi(x, x)$$

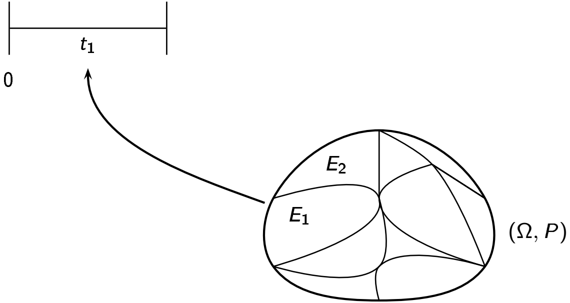
with equality only for $y = x$ (or $\Phi(x, x) = \infty$), viz. (nats)

$$\kappa(t) = \ln \frac{1}{t} .$$

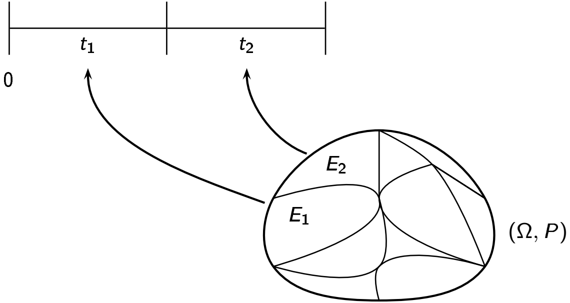
description for the classical descriptor



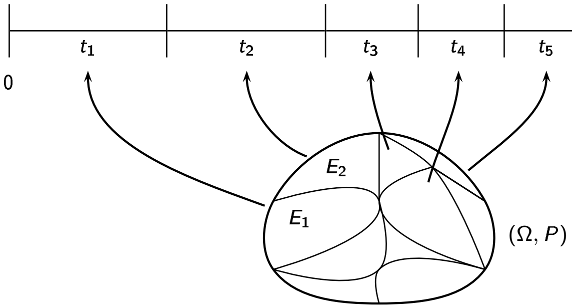
description for the classical descriptor



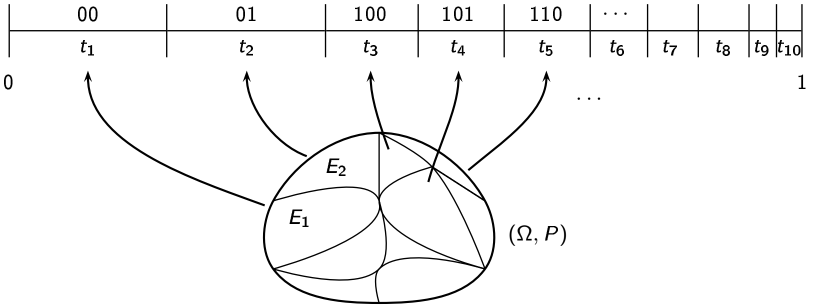
description for the classical descriptor



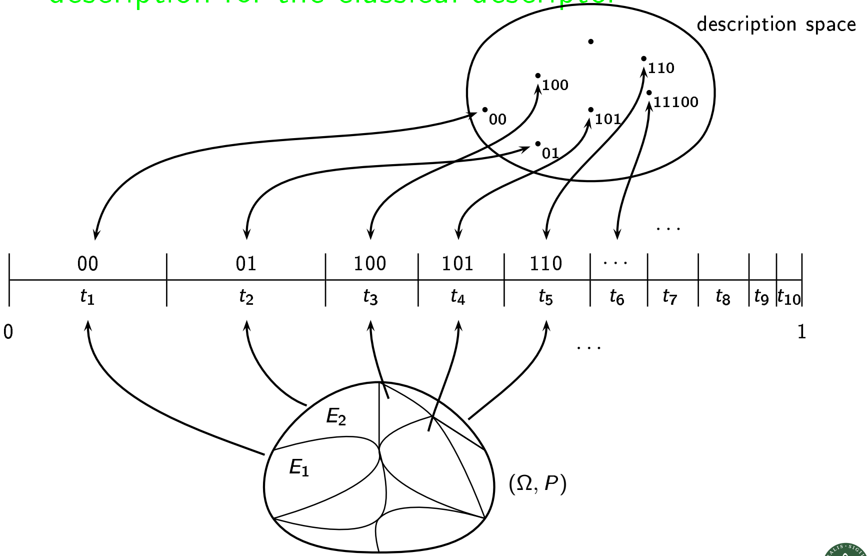
description for the classical descriptor



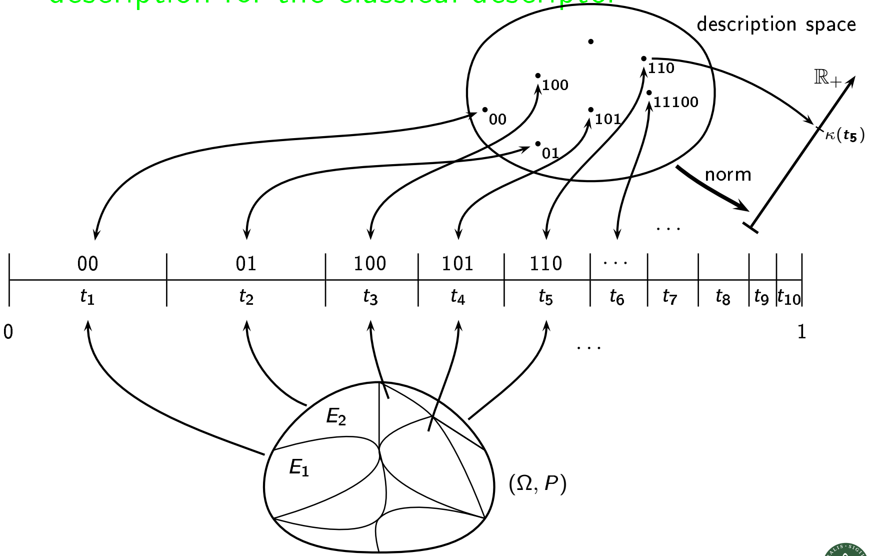
description for the classical descriptor



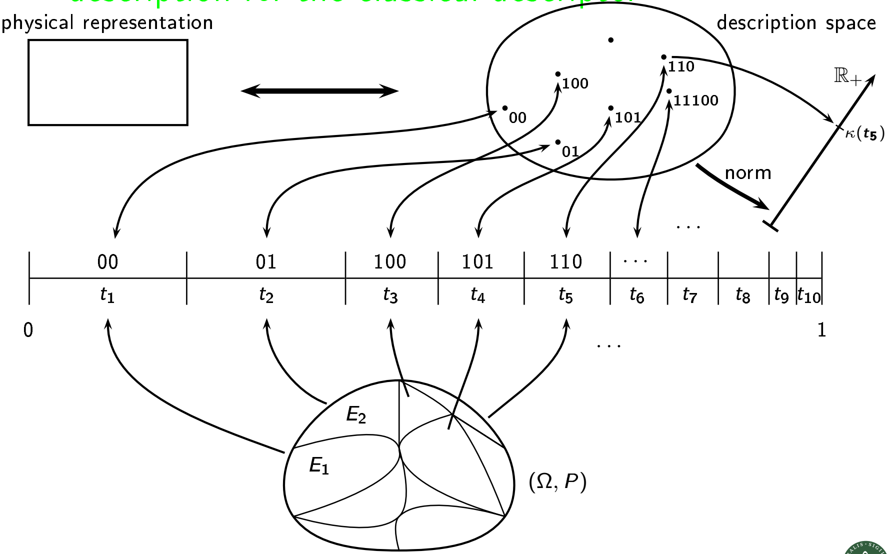
description for the classical descriptor



description for the classical descriptor



description for the classical descriptor



questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort **as seen by someone, who knows the truth** ...

questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort **as seen by someone, who knows the truth** ...

... but is this how **you** perceive accumulated effort?

questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort **as seen by someone, who knows the truth** ...

... but is this how **you** perceive accumulated effort?

What if the x_i's above are not what you perceive as truth?

questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort **as seen by someone, who knows the truth** ...

... but is this how **you** perceive accumulated effort?

What if the x_i's above are not what you perceive as truth?

... perhaps this also depends on what you believe – and Φ should rather be something like $\sum \pi(x_i, y_i)\kappa(y_i)$.

questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort **as seen by someone, who knows the truth** ...

... but is this how **you** perceive accumulated effort?

What if the x_i's above are not what you perceive as truth?

... perhaps this also depends on what you believe – and Φ should rather be something like $\sum \pi(x_i, y_i)\kappa(y_i)$.

Let's go philosophical:

the beginnings of a philosophy of information

The whole is the world , $\mathcal{V}$

the beginnings of a philosophy of information

The whole is the world, $\mathcal{V}$

Situations from the world involve Nature and you, Observer.

the beginnings of a philosophy of information

The whole is the world, $\mathcal{V}$

Situations from the world involve Nature and you, Observer.

Nature has no mind but holds the truth (x),

the beginnings of a philosophy of information

The whole is the world , $\mathcal{V}$

Situations from the world involve Nature and you, Observer .

Nature has no mind but holds the truth (x),

Observer has a creative mind,

- seeks the truth (x)
- is confined to belief (y)
- aims at knowledge (z).

the beginnings of a philosophy of information

The whole is the world, $\mathcal{V}$

Situations from the world involve Nature and you, Observer.

Nature has no mind but holds the truth (x),

Observer has a creative mind,
- seeks the truth (x)
- is confined to belief (y)
- aims at knowledge (z).

Knowledge is
- the synthesis of extensive experience
- an expression of how Observer perceives situations from $\mathcal{V}$
- how truth manifests itself to Observer, to you.

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_{\Pi}$.

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_\Pi$.

$\Pi_1 : (x, y) \mapsto x$ defines the classical world $\mathcal{V}_1$

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_\Pi$.

$\Pi_1 : (x, y) \mapsto x$ defines the classical world $\mathcal{V}_1$
$\Pi_0 : (x, y) \mapsto y$ defines a black hole $\mathcal{V}_0$

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_\Pi$.

$\Pi_1 : (x, y) \mapsto x$ defines the classical world $\mathcal{V}_1$
$\Pi_0 : (x, y) \mapsto y$ defines a black hole $\mathcal{V}_0$
$\Pi_q : (x, y) \mapsto qx + (1 - q)y$ defines mixtures, Tsallis' $\mathcal{V}_q$'s

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_\Pi$.

$\Pi_1 : (x, y) \mapsto x$ defines the classical world $\mathcal{V}_1$
$\Pi_0 : (x, y) \mapsto y$ defines a black hole $\mathcal{V}_0$
$\Pi_q : (x, y) \mapsto qx + (1 - q)y$ defines mixtures, Tsallis' $\mathcal{V}_q$'s

Associated with $\mathcal{V}_\Pi$ are (possibly many) effort functions, Φ's.
An effort function is proper if it satisfies the perfect match principle (PMP): $\Phi(x, y) \geq \Phi(x, x)$ with equality iff
$y = x$ (or $\Phi(x, x) = \infty$).

interaction and effort

Proposal: Knowledge depends on truth and belief via a characteristic interactor Π: $\boxed{z = \Pi(x, y)}$ $\mathcal{V} = \mathcal{V}_\Pi$.

$\Pi_1 : (x, y) \mapsto x$ defines the classical world $\mathcal{V}_1$
$\Pi_0 : (x, y) \mapsto y$ defines a black hole $\mathcal{V}_0$
$\Pi_q : (x, y) \mapsto qx + (1 - q)y$ defines mixtures, Tsallis' $\mathcal{V}_q$'s

Associated with $\mathcal{V}_\Pi$ are (possibly many) effort functions, Φ's. An effort function is proper if it satisfies the perfect match principle (PMP): $\Phi(x, y) \geq \Phi(x, x)$ with equality iff
$y = x$ (or $\Phi(x, x) = \infty$).

Thesis Given $\mathcal{V}_\Pi$, there is at most one proper Φ-function

digression: what if Nature can communicate?

Then we speak about an Expert .
You ask Expert for advice.
Expert's knowledge is x, advice given is y.

digression: what if Nature can communicate?

Then we speak about an Expert .
You ask Expert for advice.
Expert's knowledge is x, advice given is y.

Expert may be tempted to act in bad faith ($y \neq x$).

digression: what if Nature can communicate?

Then we speak about an Expert .
You ask Expert for advice.
Expert's knowledge is x, advice given is y.

Expert may be tempted to act in bad faith ($y \neq x$).

Problem: How to keep the expert honest?

A solution. If you know a proper Φ, you can avoid this and thus keep the expert honest: Fix a suitable downpayment in order to receive advice and then agree that Expert pays a penalty of $\Phi(x, y)$ as soon as the truth is known....

entropy, divergence, the fundamental inequality

Abstract modelling involves effort (Φ), entropy (H), and divergence (D). Φ is assumed proper.

entropy, divergence, the fundamental inequality

Abstract modelling involves effort (Φ), entropy (H), and divergence (D). Φ is assumed proper. Entropy is defined as

minimal effort, given the truth, divergence as excess effort:

$$H(x) = \Phi(x,x)\,;\;\; D(x,y) = \Phi(x,y) - H(x)\,.$$

(forget about possibility of infinite values)

entropy, divergence, the fundamental inequality

Abstract modelling involves effort (Φ), entropy (H), and divergence (D). Φ is assumed proper. Entropy is defined as minimal effort, given the truth, divergence as excess effort:

$$H(x) = \Phi(x,x)\,; \quad D(x,y) = \Phi(x,y) - H(x)\,.$$

(forget about possibility of infinite values)

The properness of Φ may be expressed in terms of D by the fundamental inequality of information theory (FI):

$$D(x,y) \geq 0 \quad \text{with equality iff } y = x\,.$$

Further notions and properties are best discussed for probabilistic modelling.

probabilistic modelling (discrete)

Truth-, belief- and knowledge instances are $x = (x_i)$, $y = (y_i)$ and $z = (z_i)$ (i ranging over an alfabet $\mathbb{A}$).
x and y are probability distributions, z just a function on $\mathbb{A}$.

probabilistic modelling (discrete)

Truth-, belief- and knowledge instances are $x = (x_i)$, $y = (y_i)$ and $z = (z_i)$ (i ranging over an alfabet $\mathbb{A}$).
x and y are probability distributions, z just a function on $\mathbb{A}$.

Interaction, Π, acts via the local interactor π:
$\big(\Pi(x,y)\big)_i = \pi(x_i, y_i)$. π is always assumed sound , i.e.
$\pi(s,t) = s$ if $t = s$ (perfect match).
π is weakly consistent if $\forall x \forall y : \sum z_i = 1$. Strong consistency requires that z is always a probability distribution.

Proposition: Only the π_q's given by $\pi_q(s,t) = qs + (1-q)t$ are weakly consistent; strong consistency requires $0 \leq q \leq 1$.

accumulated effort, the one and only

Accumulated effort always chosen among $\Phi_{\pi,\kappa}$ where κ is a descriptor and

$$\Phi_{\pi,\kappa}(x,y) = \sum_{i \in \mathbb{A}} \pi(x_i, y_i)\kappa(y_i)\,.$$

Theorem (modulo regularity conditions). Given $\pi = \pi(s,t)$, let $\pi_2' = \frac{\partial \pi}{\partial t}$ and put $\chi(t) = \pi_2'(t,t)$.
Only one among the $\Phi_{\pi,\kappa}$'s can be proper, viz. the solution to

$$t\kappa'(t) + \chi(t)\kappa(t) = -1\,,\ \ \kappa(1) = 0\,. \qquad (*)$$

accumulated effort, the one and only

Accumulated effort always chosen among $\Phi_{\pi,\kappa}$ where κ is a descriptor and

$$\Phi_{\pi,\kappa}(x,y) = \sum_{i \in \mathbb{A}} \pi(x_i, y_i)\kappa(y_i)\,.$$

Theorem (modulo regularity conditions). Given $\pi = \pi(s,t)$, let $\pi_2' = \frac{\partial \pi}{\partial t}$ and put $\chi(t) = \pi_2'(t,t)$.
Only one among the $\Phi_{\pi,\kappa}$'s can be proper, viz. the solution to

$$t\kappa'(t) + \chi(t)\kappa(t) = -1\,,\ \ \kappa(1) = 0\,. \qquad (*)$$

If π is consistent, hence one of the π_q's, then a proper $\Phi_{\pi,\kappa}$ exists iff $q > 0$ ($q = 0$ OK as a singular case, though).

accumulated effort, the one and only

Accumulated effort always chosen among $\Phi_{\pi,\kappa}$ where κ is a descriptor and

$$\Phi_{\pi,\kappa}(x,y) = \sum_{i \in \mathbb{A}} \pi(x_i, y_i)\kappa(y_i)\,.$$

Theorem (modulo regularity conditions). Given $\pi = \pi(s,t)$, let $\pi_2' = \frac{\partial \pi}{\partial t}$ and put $\chi(t) = \pi_2'(t,t)$.
Only one among the $\Phi_{\pi,\kappa}$'s can be proper, viz. the solution to

$$t\kappa'(t) + \chi(t)\kappa(t) = -1\,,\ \kappa(1) = 0\,. \quad (*)$$

If π is consistent, hence one of the π_q's, then a proper $\Phi_{\pi,\kappa}$ exists iff $q > 0$ ($q = 0$ OK as a singular case, though).
If so, the unique descriptor concerned is the one depending linearly on t^{q-1}, i.e. $\boxed{\kappa_q(t) = \ln_q \frac{1}{t}}$ (recall: $\ln_q u = \frac{1}{1-q}(u^{1-q}-1)$).

gross effort, pointwise fundamental inequality

Introduce gross (accumulated) effort and gross entropy by adding a term representing overhead cost (or effort):

gross effort, pointwise fundamental inequality

Introduce gross (accumulated) effort and gross entropy by adding a term representing overhead cost (or effort):

$$\text{gross effort: } \tilde{\Phi}(x,y) = \sum_{i\in\mathbb{A}} \big(\pi(x_i,y_i)\kappa(y_i) + y_i\big) = \Phi(x,y) + 1\,,$$

$$\text{gross entropy: } \tilde{H}(x) = \sum_{i\in\mathbb{A}} \big(x_i\kappa(x_i) + x_i\big) = H(x) + 1\,.$$

gross effort, pointwise fundamental inequality

Introduce gross (accumulated) effort and gross entropy by adding a term representing overhead cost (or effort):

$$\text{gross effort: } \tilde{\Phi}(x,y) = \sum_{i\in\mathbb{A}} \big(\pi(x_i,y_i)\kappa(y_i) + y_i\big) = \Phi(x,y) + 1\,,$$

$$\text{gross entropy: } \tilde{H}(x) = \sum_{i\in\mathbb{A}} \big(x_i\kappa(x_i) + x_i\big) = H(x) + 1\,.$$

Clearly, "gross divergencee"=divergence and, defining the divergence generator by
$\delta(s,t) = \big(\pi(s,t)\kappa(t) + t\big) - \big(s\kappa(s) + s\big)$, one has
$D(x,y) = \sum \delta(x_i,y_i)$.

gross effort, pointwise fundamental inequality

Introduce gross (accumulated) effort and gross entropy by adding a term representing overhead cost (or effort):

$$\text{gross effort: } \tilde{\Phi}(x,y) = \sum_{i \in \mathbb{A}} \big(\pi(x_i, y_i)\kappa(y_i) + y_i\big) = \Phi(x,y) + 1\,,$$

$$\text{gross entropy: } \tilde{H}(x) = \sum_{i \in \mathbb{A}} \big(x_i\kappa(x_i) + x_i\big) = H(x) + 1\,.$$

Clearly, “gross divergencee”=divergence and, defining the divergence generator by
$\delta(s,t) = \big(\pi(s,t)\kappa(t) + t\big) - \big(s\kappa(s) + s\big)$, one has
$D(x,y) = \sum \delta(x_i, y_i)$. We refer to the inequality $\delta \geq 0$ as the pointwise fundamental inequality (PFI). Clearly PFI $\Longrightarrow$ FI.

Conjecture Converse also true

In practice, PMP and FI are always proved via PFI !

given κ, which world are you in?

Given π, we insist, when possible, to choose κ such that the resulting function Φ is proper. This gives a unique choice, the ideal descriptor.

given κ, which world are you in?

Given π, we insist, when possible, to choose κ such that the resulting function Φ is proper. This gives a unique choice, the ideal descriptor. You determine κ from π, but

Warning: you cannot determine π from κ

given κ, which world are you in?

Given π, we insist, when possible, to choose κ such that the resulting function Φ is proper. This gives a unique choice, the ideal descriptor. You determine κ from π, but

Warning: you cannot determine π from κ

Thus knowing the entropy function does not reveal the world.

given κ, which world are you in?

Given π, we insist, when possible, to choose κ such that the resulting function Φ is proper. This gives a unique choice, the ideal descriptor. You determine κ from π, but

Warning: you cannot determine π from κ

Thus knowing the entropy function does not reveal the world.

Examples: Let $\pi = \pi_q$ $(q > 0)$ and consider π^ξ of the form

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

Then the differential equation (*) is unchanged, hence you find the same descriptor κ_q. E.g. for $\xi(u) = \ln u$, $\pi^\xi(s,t) = s^q t^{1-q}$; by PFI, the associated effort is proper.

given κ, which world are you in?

Given π, we insist, when possible, to choose κ such that the resulting function Φ is proper. This gives a unique choice, the ideal descriptor. You determine κ from π, but

Warning: you cannot determine π from κ

Thus knowing the entropy function does not reveal the world.

Examples: Let $\pi = \pi_q$ $(q > 0)$ and consider π^ξ of the form

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

Then the differential equation (*) is unchanged, hence you find the same descriptor κ_q. E.g. for $\xi(u) = \ln u$, $\pi^\xi(s,t) = s^q t^{1-q}$; by PFI, the associated effort is proper.

Problem which κ's are associated with (meaningful) π's?

e.g. $\kappa(t) = \frac{1}{2}(t^{-2} - 1)$?

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

To us, this is expressed via controls, w's. There is a bijection $y \leftrightarrow w$ ($w = \hat{y}$; $y = \check{w}$) defined by $w_i = \kappa(y_i)$; $i \in \mathbb{A}$.

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

To us, this is expressed via controls, w's. There is a bijection $y \leftrightarrow w$ ($w = \hat{y}$; $y = \check{w}$) defined by $w_i = \kappa(y_i)$; $i \in \mathbb{A}$.

Expressed via controls, the effort function is denoted Ψ: $\Psi(x, w) = \Phi(x, y)$ with $y \leftrightarrow w$.

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

To us, this is expressed via controls, w's. There is a bijection $y \leftrightarrow w$ ($w = \hat{y}$; $y = \check{w}$) defined by $w_i = \kappa(y_i)$; $i \in \mathbb{A}$.

Expressed via controls, the effort function is denoted Ψ:
$\Psi(x, w) = \Phi(x, y)$ with $y \leftrightarrow w$.

What can Observer do?

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

To us, this is expressed via controls, w's. There is a bijection $y \leftrightarrow w$ ($w = \hat{y}$; $y = \check{w}$) defined by $w_i = \kappa(y_i)$; $i \in \mathbb{A}$.

Expressed via controls, the effort function is denoted Ψ: $\Psi(x, w) = \Phi(x, y)$ with $y \leftrightarrow w$.

What can Observer do? Constrain the possible truth instances via control !

what *c a n* we know?

Setting: World $\mathcal{V}_\pi$ with ideal descriptor and effort fct. Φ.
I.J. Good (1952): $\boxed{\text{Belief is a tendency to act !}}$

To us, this is expressed via controls, w's. There is a bijection $y \leftrightarrow w$ ($w = \hat{y}$; $y = \check{w}$) defined by $w_i = \kappa(y_i)$; $i \in \mathbb{A}$.

Expressed via controls, the effort function is denoted Ψ:
$\Psi(x, w) = \Phi(x, y)$ with $y \leftrightarrow w$.

What can Observer do? Constrain the possible truth instances via control ! Constraints are expressed by preparations which are sets $\mathcal{P}$ of x's.
A feasible preparation is one which Observer can realize.

more on preparations

Typical example (of genus 1): Fix a control w and a level h. Set-up an experiment (!?) which constrains Natures possibilities to the preparation

$$\mathcal{P}(w,h) = \{x|\Psi(x,w) = h\}$$

$$\text{or variant } \mathcal{P}_{\leq}(w,h) = \{x|\Psi(x,w) \leq h\}\,.$$

more on preparations

Typical example (of genus 1): Fix a control w and a level h. Set-up an experiment (!?) which constrains Natures possibilities to the preparation

$$\mathcal{P}(w,h) = \{x | \Psi(x,w) = h\}$$
$$\text{or variant } \mathcal{P}_{\leq}(w,h) = \{x | \Psi(x,w) \leq h\}\,.$$

Finite non-empty intersections of such level sets (or sub-level sets) constitute the feasible preparations and shows what Observer can know !

games!

Fix a preparation $\mathcal{P}$ and consider the two-person zero-sum game $\gamma(\mathcal{P})$ between Nature and Observer with x's in $\mathcal{P}$ and controls w as available strategies and with objective function $\Psi(x, w)$. Nature is a maximizer, Observer a minimizer.

games!

Fix a preparation $\mathcal{P}$ and consider the two-person zero-sum game $\gamma(\mathcal{P})$ between Nature and Observer with x's in $\mathcal{P}$ and controls w as available strategies and with objective function $\Psi(x, w)$. Nature is a maximizer, Observer a minimizer.

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

$$\sup_{x \in \mathcal{P}} \inf_{w} \Psi(x, w) \,, \quad \text{respectively} \quad \inf_{w} \sup_{x \in \mathcal{P}} \Psi(x, w) \,.$$

games!

Fix a preparation $\mathcal{P}$ and consider the two-person zero-sum game $\gamma(\mathcal{P})$ between Nature and Observer with x's in $\mathcal{P}$ and controls w as available strategies and with objective function $\Psi(x, w)$. Nature is a maximizer, Observer a minimizer.

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

$$\sup_{x \in \mathcal{P}} \inf_{w} \Psi(x, w)\,, \quad \text{respectively} \quad \inf_{w} \sup_{x \in \mathcal{P}} \Psi(x, w)\,.$$

The value for Nature is the MaxEnt value

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

games!

Fix a preparation $\mathcal{P}$ and consider the two-person zero-sum game $\gamma(\mathcal{P})$ between Nature and Observer with x's in $\mathcal{P}$ and controls w as available strategies and with objective function $\Psi(x, w)$. Nature is a maximizer, Observer a minimizer.

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

$$\sup_{x\in\mathcal{P}} \inf_{w} \Psi(x, w)\,, \quad \text{respectively} \quad \inf_{w} \sup_{x\in\mathcal{P}} \Psi(x, w)\,.$$

The value for Nature is the MaxEnt value

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

The value for Observer is the minimal risk value

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

equilibrium, robustness

Note that $H_{\max}(\mathcal{P}) \leq R_{\min}(\mathcal{P})$, the minimax inequality. If "$=$" holds (and value is finite), the game is in equilibrium.

equilibrium, robustness

Note that $H_{\max}(\mathcal{P}) \leq R_{\min}(\mathcal{P})$, the minimax inequality . If "$=$" holds (and value is finite), the game is in equilibrium .
Optimal strategies: For Nature a MaxEnt strategy , an $x \in \mathcal{P}$ with $H(x) = H_{\max}(\mathcal{P})$; for Observer a control w with $R(w) = R_{\min}(\mathcal{P})$.

equilibrium, robustness

Note that $H_{\max}(\mathcal{P}) \leq R_{\min}(\mathcal{P})$, the minimax inequality . If "$=$" holds (and value is finite), the game is in equilibrium . Optimal strategies: For Nature a MaxEnt strategy , an $x \in \mathcal{P}$ with $H(x) = H_{\max}(\mathcal{P})$; for Observer a control w with $R(w) = R_{\min}(\mathcal{P})$.

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

equilibrium, robustness

Note that $H_{\max}(\mathcal{P}) \leq R_{\min}(\mathcal{P})$, the minimax inequality. If "$=$" holds (and value is finite), the game is in equilibrium.
Optimal strategies: For Nature a MaxEnt strategy, an $x \in \mathcal{P}$ with $H(x) = H_{\max}(\mathcal{P})$; for Observer a control w with $R(w) = R_{\min}(\mathcal{P})$.

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

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

$$\forall x \in \mathcal{P} \ : \ H(x) + D(x, x^*) \leq H_{\max}(\mathcal{P})$$
$$\forall w \ : \ R(w) \geq H_{\max}(\mathcal{P}) + D(x^*, \check{w}) \, .$$

Exponential families

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

maximal preparations Consider x^* and w^*. Then equilibrium holds for some $\gamma(\mathcal{P})$ with x^* and w^* as optimal strategies **iff** $h^* = \Psi(x^*, w^*) < \infty$ and $w^* = \hat{x^*}$. If so, the largest such set is the sublevel set defined from w^* and h^*.

Again, this follows by inspection of Nash' saddle value inequalities.

Exponential families, cont.

Let w be a control, let $\mathcal{L}^w$ be the preparation family of non-empty sets of the form $\mathcal{P}(w, h)$. The associated exponential family, denoted $\hat{\mathcal{E}}^w$ is the set of controls ε which are robust for all preparations in $\mathcal{L}^w$.

Exponential families, cont.

Let w be a control, let $\mathcal{L}^w$ be the preparation family of non-empty sets of the form $\mathcal{P}(w,h)$. The associated exponential family, denoted $\hat{\mathcal{E}}^w$ is the set of controls ε 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 $\mathcal{E}^w$ ($x^* = \check{\varepsilon}$ for some $\varepsilon \in \mathcal{E}^w$).

Exponential families, cont.

Let w be a control, let $\mathcal{L}^w$ be the preparation family of non-empty sets of the form $\mathcal{P}(w,h)$. The associated exponential family, denoted $\hat{\mathcal{E}}^w$ is the set of controls ε 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 $\mathcal{E}^w$ ($x^* = \check{\varepsilon}$ for some $\varepsilon \in \mathcal{E}^w$). From definitions and the robustness lemma you find:

> Consider a preparation family $\mathcal{L}^w$. Let x^* be a truth instance, put $\varepsilon^* = \hat{x^*}$ and assume that $\varepsilon^* \in \hat{\mathcal{E}}^w$. Put $h = \Psi(x^*, w)$. Then $\gamma(\mathcal{P}(w,h))$ is in equilibrium and has x^* and ε^* as optimal strategies. In particular, x^* is the MaxEnt distribution for $\mathcal{P}(w,h)$.

sketch of MaxEnt determination for $\mathcal{V}_q$

Consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to π_q with $q > 0$. Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.

sketch of MaxEnt determination for $\mathcal{V}_q$

Consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to π_q with $q > 0$. Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.
But $\Psi(x, w) = \sum \big(qx_i + (1-q)y_i\big)w_i$ so condition is equivalent to $\sum x_i w_i$ being constant over $\mathcal{P}$.

sketch of MaxEnt determination for $\mathcal{V}_q$

Consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to π_q with $q > 0$.
Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.
But $\Psi(x, w) = \sum \big(qx_i + (1-q)y_i\big)w_i$ so condition is equivalent to $\sum x_i w_i$ being constant over $\mathcal{P}$.
For fixed constants α and β this implies that $\sum x_i(\alpha + \beta w_i)$ is constant over $\mathcal{P}$.

sketch of MaxEnt determination for $\mathcal{V}_q$

Consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to π_q with $q > 0$.
Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.
But $\Psi(x, w) = \sum \big(qx_i + (1-q)y_i\big) w_i$ so condition is equivalent to $\sum x_i w_i$ being constant over $\mathcal{P}$.
For fixed constants α and β this implies that $\sum x_i(\alpha + \beta w_i)$ is constant over $\mathcal{P}$.
Now, if $\alpha + \beta w$ is a control, say w^*, $\sum x_i w_i^*$ is constant over $\mathcal{P}$, hence $\Psi(x, w^*)$ is constant over $\mathcal{P}$, i.e. $w^* \in \hat{\mathcal{E}}^w$ and the robustness lemma applies.

sketch of MaxEnt determination for $\mathcal{V}_q$

Consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to π_q with $q > 0$.
Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.
But $\Psi(x, w) = \sum \big(qx_i + (1-q)y_i\big) w_i$ so condition is equivalent to $\sum x_i w_i$ being constant over $\mathcal{P}$.
For fixed constants α and β this implies that $\sum x_i(\alpha + \beta w_i)$ is constant over $\mathcal{P}$.
Now, if $\alpha + \beta w$ is a control, say w^*, $\sum x_i w_i^*$ is constant over $\mathcal{P}$, hence $\Psi(x, w^*)$ is constant over $\mathcal{P}$, i.e. $w^* \in \hat{\mathcal{E}}^w$ and the robustness lemma applies.
Then, given β, try to adjust α so that $\alpha + \beta w$ is a control.
Classically, α is the logarithm of the partition function. .
Finally, adjust β ($\approx$ inverse temperature) to desired level ...

what have we achieved?

• found a reasonably transparent interpretation of Tsallis entropy
• developed a basis for an abstract theory
• clarified 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
• indicated dual role of preparations and exponential families
• exploited games and wisdom of Nash, enabled MaxEnt calculations without introducing Lagrange multipliers
• separated Nature from Observer in key expressions

what needs being done?

- interaction, how?
- description, how?
- control, how?
- expand, quantum setting ...
- link to information geometry
- ...

what needs being done?

- interaction, how?
- description, how?
- control, how?
- expand, quantum setting ...
- link to information geometry
- ...

thank you !