# Cognition and Games - an approach to Information

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

# Some key features

- The approach is abstract, devoid of references to probability
- Emphasis is on philosophy, on interpretations

# Some key features

- The approach is abstract, devoid of references to probability
- Emphasis is on philosophy, on interpretations

- From cognition: Truth, belief, knowledge, perception
- More specifically: Description, effort, control, preparations
- inference: MaxEnt thm $\supseteq$ projection thm.; exponential fam.
- A pythagorean thm. containing classical and inf.th. versions

## Some key features

- The approach is abstract , devoid of references to probability
- Emphasis is on philosophy , on interpretations

- From cognition: Truth, belief, knowledge, perception
- More specifically: Description, effort, control, preparations
- inference: MaxEnt thm $\supseteq$ projection thm.; exponential fam.
- A pythagorean thm. containing classical and inf.th. versions

- Key tool: Study of conflict situations via games .

# Some key features

- The approach is abstract , devoid of references to probability
- Emphasis is on philosophy , on interpretations

- From cognition: Truth, belief, knowledge, perception
- More specifically: Description, effort, control, preparations
- inference: MaxEnt thm $\supseteq$ projection thm.; exponential fam.
- A pythagorean thm. containing classical and inf.th. versions

- Key tool: Study of conflict situations via games .

**Ingarden & Urbanik 1962:** "... information seems intuitively a much simpler and more elementary notion than that of probability ... [it] represents a more primary step of knowledge than that of cognition of probability ..."

# Some key features

- The approach is abstract, devoid of references to probability
- Emphasis is on philosophy, on interpretations

- From cognition: Truth, belief, knowledge, perception
- More specifically: Description, effort, control, preparations
- inference: MaxEnt thm $\supseteq$ projection thm.; exponential fam.
- A pythagorean thm. containing classical and inf.th. versions

- Key tool: Study of conflict situations via games.

**Ingarden & Urbanik 1962:** "... information seems intuitively a much simpler and more elementary notion than that of probability ... [it] represents a more primary step of knowledge than that of cognition of probability ..."

**Kolmogorov $\approx$ 1970:** "Information theory must preceed probability theory and not be based on it"

# Modelling, basic elements I

# Modelling, basic elements I

| | |
|---|---|
| $X$ : | state space. $x \in X$ : truth instance or state |
| $Y \supseteq X$ : | belief reservoir. $y \in Y$ : a belief |
| $Z \supseteq Y$ : | knowledge space or set of potential perceptions |
| | $z \in Z$ : a knowledge element or a perception |
| $X \otimes Y$ : | relation of domination $(y \succ x)$ |
| $\Pi : X \otimes Y \rightarrow Z$ : | the interaction |
| $W \longleftrightarrow Y$ : | action space. $w \in W$ : a control. |

# Modelling, basic elements I

| | |
|---|---|
| $X$ : | state space. $x \in X$ : truth instance or state |
| $Y \supseteq X$ : | belief reservoir. $y \in Y$ : a belief |
| $Z \supseteq Y$ : | knowledge space or set of potential perceptions |
| | $z \in Z$ : a knowledge element or a perception |
| $X \otimes Y$ : | relation of domination $(y \succ x)$ |
| $\Pi : X \otimes Y \to Z$ : | the interaction |
| $W \longleftrightarrow Y$ : | action space. $w \in W$ : a control. |

· $\Pi$ defines the world: $\mathcal{W} = \mathcal{W}_\Pi$
· action $\approx$ control $\approx$ description
· Good (1952): belief is a tendency to act

# Modelling, basic elements II

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume: $\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume:
$\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

Introduce "players", Nature (chooses $x$) and Observer (seeks $x$ but is confined to belief, $y$).

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume:
$\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

Introduce "players", Nature (chooses $x$) and Observer (seeks $x$ but is confined to belief, $y$).

What do they fight over? Description effort, $\Phi$!

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume:
$\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

Introduce "players", Nature (chooses $x$) and Observer (seeks $x$ but is confined to belief, $y$).

What do they fight over? Description effort, $\Phi$!
This is a function $\Phi : X \otimes Y \to ]-\infty, \infty]$. Assume that $\Phi$ is calibrated i.e. $\Phi(x, y) = 0$ for situations of certain belief.

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume:
$\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

Introduce "players", Nature (chooses $x$) and Observer (seeks $x$ but is confined to belief, $y$).

What do they fight over? Description effort, $\Phi$!
This is a function $\Phi : X \otimes Y \to ]-\infty, \infty]$. Assume that $\Phi$ is calibrated i.e. $\Phi(x, y) = 0$ for situations of certain belief.

$\Phi$ (or $-\Phi$) are the score functions of other authors (Good, Fischer, ..., Dawid, Lauritzen,...).

# Modelling, basic elements II

A situation $(x, y) \in X \otimes Y$ is a perfect match if $y = x$ and a certain belief if $y \in$ some given non-empty set $Y_{det}$. Assume:
$\forall x : x \succ x, \Pi(x, x) = x$ (soundness); $\exists y \forall x : y \succ x$.

Introduce "players", Nature (chooses $x$) and Observer (seeks $x$ but is confined to belief, $y$).

What do they fight over? Description effort, $\Phi$!
This is a function $\Phi : X \otimes Y \to ]-\infty, \infty]$. Assume that $\Phi$ is calibrated i.e. $\Phi(x, y) = 0$ for situations of certain belief.

$\Phi$ (or $-\Phi$) are the score functions of other authors (Good, Fischer, ..., Dawid, Lauritzen,...).

**Key principle** $\Phi$ satisfies the perfect match principle (or is proper) if, for fixed $x$, $\Phi$ is minimized under a perfect match and not otherwise (unless $\Phi(x, x) = \infty$).

# Ideal description for a world $\mathcal{W}_\Pi$

There are worlds *without* associated proper descriptions but:

> **Thesis** Given the world, there exists at most one proper description modulo equivalence $(\Phi_1 \equiv \Phi_2 \therefore \exists c > 0 : \Phi_1 = c\Phi_2)$.

# Ideal description for a world $\mathcal{W}_\Pi$

There are worlds *without* associated proper descriptions but:

**Thesis** Given the world, there exists at most one proper description modulo equivalence $(\Phi_1 \equiv \Phi_2 \therefore \exists c > 0 : \Phi_1 = c\Phi_2)$.

If a proper description exists, there is often a natural choice among equivalent ones. Amounts to a choice of unit.
Assume that a unique proper description exists corresponding to such a choice, the ideal description.

# Ideal description for a world $\mathcal{W}_\Pi$

There are worlds *without* associated proper descriptions but:

**Thesis** Given the world, there exists at most one proper description modulo equivalence $(\Phi_1 \equiv \Phi_2 \therefore \exists c > 0 : \Phi_1 = c\Phi_2)$.

If a proper description exists, there is often a natural choice among equivalent ones. Amounts to a choice of unit.
Assume that a unique proper description exists corresponding to such a choice, the ideal description.

**Warning** Knowing the description, you may not know the world!

**Claim**
Ideal description
$\leftrightarrow$fundamental inequality of information theory
$\leftrightarrow$2.nd law of thermodynamics.

# Elements of information

# Elements of information

Full information means that the truth instance is revealed to Observer, e.g. by an informer. Notation: "$x$".

# Elements of information

Full information means that the truth instance is revealed to Observer, e.g. by an informer. Notation: "$x$".

Quantitatively, information is saved effort

Thus, $\Phi(x, y) =$ value to Observer of the information "$x$" in a situation with belief $y$.

# Elements of information

Full information means that the truth instance is revealed to Observer, e.g. by an informer. Notation: "$x$".

Quantitatively, information is saved effort

Thus, $\Phi(x, y) =$ value to Observer of the information "$x$" in a situation with belief $y$.

The unit of description is then also a unit of information.

# Elements of information

Full information means that the truth instance is revealed to Observer, e.g. by an informer. Notation: "$x$".

Quantitatively, information is saved effort

Thus, $\Phi(x, y)$ = value to Observer of the information "$x$" in a situation with belief $y$.

The unit of description is then also a unit of information.

Entropy $\mathrm{H}(x)$ = minimal effort required: $\mathrm{H}(x) = \Phi(x, x)$.
Divergence $\mathrm{D}(x, y)$ is excess description effort:
$\mathrm{D}(x, y) = \Phi(x, y) - \mathrm{H}(x)$.

# Elements of information

Full information means that the truth instance is revealed to Observer, e.g. by an informer. Notation: "$x$".

Quantitatively, information is saved effort

Thus, $\Phi(x, y)$ = value to Observer of the information "$x$" in a situation with belief $y$.

The unit of description is then also a unit of information.

Entropy $\mathrm{H}(x)$ = minimal effort required: $\mathrm{H}(x) = \Phi(x, x)$.
Divergence $\mathrm{D}(x, y)$ is excess description effort:
$\mathrm{D}(x, y) = \Phi(x, y) - \mathrm{H}(x)$.

$(\Phi, \mathrm{H}, \mathrm{D})$ is an information triple. Basic axioms:
$\Phi(x, y) = \mathrm{H}(x) + \mathrm{D}(x, y)$ (linking identity), $\mathrm{D} \geq 0$ with equality iff there is a perfect match (fundamental inequality, FI).

# Relativization, updating

Given an information triple $(\Phi, \mathrm{H}, \mathrm{D})$, we define updating gain from prior $y_0$ to posterior $y$ by (modulo $\infty - \infty$ problems):
$\Xi(x, y_0 \rightsquigarrow y) = \Phi(x, y_0) - \Phi(x, y) = \mathrm{D}(x, y_0) - \mathrm{D}(x, y)$.

# Relativization, updating

Given an information triple $(\Phi, \mathrm{H}, \mathrm{D})$, we define updating gain from prior $y_0$ to posterior $y$ by (modulo $\infty - \infty$ problems):
$\Xi(x, y_0 \rightsquigarrow y) = \Phi(x, y_0) - \Phi(x, y) = \mathrm{D}(x, y_0) - \mathrm{D}(x, y)$.

Note: The information triple $(\Phi(x, y), \mathrm{H}(x), \mathrm{D}(x, y))$ is transformed into the new information triple for updating

$$\big(-\Xi(x, y_0 \rightsquigarrow y), -\mathrm{D}(x, y_0), \mathrm{D}(x, y)\big).$$

# Relativization, updating

Given an information triple $(\Phi, \mathrm{H}, \mathrm{D})$, we define updating gain from prior $y_0$ to posterior $y$ by (modulo $\infty - \infty$ problems):
$\Xi(x, y_0 \rightsquigarrow y) = \Phi(x, y_0) - \Phi(x, y) = \mathrm{D}(x, y_0) - \mathrm{D}(x, y)$.

Note: The information triple $(\Phi(x, y), \mathrm{H}(x), \mathrm{D}(x, y))$ is transformed into the new information triple for updating

$$\big(-\Xi(x, y_0 \rightsquigarrow y), -\mathrm{D}(x, y_0), \mathrm{D}(x, y)\big).$$

Also note: With only D given (s.t. FI holds) such updating triples can be formed (under finiteness conditions). General results for information triples (with emphasis on MaxEnt) give results for updating! Leads to models where divergence is minimized (projection theorems).

# Relativization, updating

Given an information triple $(\Phi, \mathrm{H}, \mathrm{D})$, we define updating gain from prior $y_0$ to posterior $y$ by (modulo $\infty - \infty$ problems):
$\Xi(x, y_0 \rightsquigarrow y) = \Phi(x, y_0) - \Phi(x, y) = \mathrm{D}(x, y_0) - \mathrm{D}(x, y)$.

Note: The information triple $(\Phi(x, y), \mathrm{H}(x), \mathrm{D}(x, y))$ is transformed into the new information triple for updating

$$\big(-\Xi(x, y_0 \rightsquigarrow y), -\mathrm{D}(x, y_0), \mathrm{D}(x, y)\big).$$

Also note: With only D given (s.t. FI holds) such updating triples can be formed (under finiteness conditions). General results for information triples (with emphasis on MaxEnt) give results for updating! Leads to models where divergence is minimized (projection theorems).

**Example** Updating model in Hilbert space:
$\Xi(x, y_0 \rightsquigarrow y) = \|x - y_0\|^2 - \|x - y\|^2$ corresponding to triple

$$\Big(\|x - y\|^2 - \|x - y_0\|^2, \; -\|x - y_0\|^2, \; \|x - y\|^2\Big).$$

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation and these are finite intersections of basic strict, resp. basic slack feasible preparations

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation and these are finite intersections of basic strict, resp. basic slack feasible preparations which, in turn, are $\Phi$-level, resp. $\Phi$-sublevel sets of the form ($b \in Y$, $h \in \mathbb{R}$)

$$\mathcal{P}^b(h) = \{x | \Phi(x, b) = h\}\,; \quad \mathcal{P}^b(h^\bullet) = \{x | \Phi(x, b) \leq h\}\,.$$

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation and these are finite intersections of basic strict, resp. basic slack feasible preparations which, in turn, are $\Phi$-level, resp. $\Phi$-sublevel sets of the form ($b \in Y$, $h \in \mathbb{R}$)

$$\mathcal{P}^b(h) = \{x | \Phi(x,b) = h\}\,; \quad \mathcal{P}^b(h^\bullet) = \{x | \Phi(x,b) \le h\}\,.$$

Motivation? Later! (– or pretty clear?)
With $\mathbf{b} = (b_1, \cdots, b_n)$ and $\mathbf{h} = (h_1, \cdots, h_n)$, we put

$$\mathcal{P}^{\mathbf{b}}(\mathbf{h}) = \bigcap_{i \le n} \mathcal{P}^{b_i}(h_i)\,; \quad \mathcal{P}^{\mathbf{b}}(\mathbf{h}^\bullet) = \bigcap_{i \le n} \mathcal{P}^{b_i}(h_i^\bullet)\,.$$

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation and these are finite intersections of basic strict, resp. basic slack feasible preparations which, in turn, are $\Phi$-level, resp. $\Phi$-sublevel sets of the form ($b \in Y$, $h \in \mathbb{R}$)

$$\mathcal{P}^b(h) = \{x|\Phi(x,b) = h\}\,; \quad \mathcal{P}^b(h^\bullet) = \{x|\Phi(x,b) \le h\}\,.$$

Motivation? Later! (– or pretty clear?)
With $\mathbf{b} = (b_1, \cdots, b_n)$ and $\mathbf{h} = (h_1, \cdots, h_n)$, we put

$$\mathcal{P}^{\mathbf{b}}(\mathbf{h}) = \bigcap_{i \le n} \mathcal{P}^{b_i}(h_i)\,; \quad \mathcal{P}^{\mathbf{b}}(\mathbf{h}^\bullet) = \bigcap_{i \le n} \mathcal{P}^{b_i}(h_i^\bullet)\,.$$

By $\mathbb{P}^{\mathbf{b}}$ we denote the preparation family of all strict preparations of the form $\mathcal{P}^{\mathbf{b}}(\mathbf{h})$. We define $\mathbb{P}^{\mathbf{b}^\bullet}$ similarly.

# Intermezzo: preparations, feasible ones, families...

A preparation is a non-empty subset $\mathcal{P}$ of $X$.

A feasible preparation is either a strict or a slack feasible preparation and these are finite intersections of basic strict, resp. basic slack feasible preparations which, in turn, are $\Phi$-level, resp. $\Phi$-sublevel sets of the form ($b \in Y$, $h \in \mathbb{R}$)

$$\mathcal{P}^b(h) = \{x | \Phi(x,b) = h\}\,; \quad \mathcal{P}^b(h^\bullet) = \{x | \Phi(x,b) \leq h\}\,.$$

Motivation? Later! (– or pretty clear?)
With $\mathbf{b} = (b_1, \cdots, b_n)$ and $\mathbf{h} = (h_1, \cdots, h_n)$, we put

$$\mathcal{P}^{\mathbf{b}}(\mathbf{h}) = \bigcap_{i \leq n} \mathcal{P}^{b_i}(h_i)\,; \quad \mathcal{P}^{\mathbf{b}}(\mathbf{h}^\bullet) = \bigcap_{i \leq n} \mathcal{P}^{b_i}(h_i^\bullet)\,.$$

By $\mathbb{P}^{\mathbf{b}}$ we denote the preparation family of all strict preparations of the form $\mathcal{P}^{\mathbf{b}}(\mathbf{h})$. We define $\mathbb{P}^{\mathbf{b}^\bullet}$ similarly.

**Illuminating Example:** Updating model in Hilbert space ...

# the games $\gamma(\mathcal{P})$ for general $\mathcal{P}$, basic notions

The game $\gamma(\mathcal{P}) = \gamma(\Phi, \mathcal{P})$ .·. $\Phi$ objective function, Nature maximizer, Observer minimizer. Nature strategies: $x$'s in $\mathcal{P}$. Observer strategies: beliefs $y \succ \mathcal{P}$ ($\forall x \in \mathcal{P} : y \succ x$).

# the games $\gamma(\mathcal{P})$ for general $\mathcal{P}$, basic notions

The game $\gamma(\mathcal{P}) = \gamma(\Phi, \mathcal{P})$ .·. $\Phi$ objective function, Nature maximizer, Observer minimizer. Nature strategies: $x$'s in $\mathcal{P}$. Observer strategies: beliefs $y \succ \mathcal{P}$ ($\forall x \in \mathcal{P} : y \succ x$).

MaxEnt is value for Nature , MinRisk value for Observer:
$\mathrm{H}_{\max}(\mathcal{P}) = \sup_{x\in\mathcal{P}} \mathrm{H}(x) = \sup_{x\in\mathcal{P}} \inf_{y\succ x} \Phi(x, y)$.
$\mathrm{Ri}_{\min}(\mathcal{P}) = \inf_{y\succ\mathcal{P}} \mathrm{Ri}(y) = \inf_{y\succ\mathcal{P}} \sup_{x\in\mathcal{P}} \Phi(x, y)$.
Note: $\mathrm{Ri}(y) = \mathrm{Ri}(y|\mathcal{P})$.

# the games $\gamma(\mathcal{P})$ for general $\mathcal{P}$, basic notions

The game $\gamma(\mathcal{P}) = \gamma(\Phi, \mathcal{P})$ ∴ $\Phi$ objective function, Nature maximizer, Observer minimizer. Nature strategies: $x$'s in $\mathcal{P}$. Observer strategies: beliefs $y \succ \mathcal{P}$ ($\forall x \in \mathcal{P} : y \succ x$).

MaxEnt is value for Nature, MinRisk value for Observer:
$\mathrm{H}_{\max}(\mathcal{P}) = \sup_{x \in \mathcal{P}} \mathrm{H}(x) = \sup_{x \in \mathcal{P}} \inf_{y \succ x} \Phi(x, y)$.
$\mathrm{Ri}_{\min}(\mathcal{P}) = \inf_{y \succ \mathcal{P}} \mathrm{Ri}(y) = \inf_{y \succ \mathcal{P}} \sup_{x \in \mathcal{P}} \Phi(x, y)$.
Note: $\mathrm{Ri}(y) = \mathrm{Ri}(y|\mathcal{P})$.

$x^* \in \mathcal{P}$ optimal for Nature ∴ $\mathrm{H}(x^*) = \mathrm{H}_{\max}(\mathcal{P})$.
$y^* \succ \mathcal{P}$ optimal for Observer ∴ $\mathrm{Ri}(y^*) = \mathrm{Ri}_{\min}(\mathcal{P})$.

If $\mathrm{H}_{\max}(\mathcal{P}) = \mathrm{Ri}_{\min}(\mathcal{P})$ is finite, $\gamma(\mathcal{P})$ is in equilibrium .

# the games $\gamma(\mathcal{P})$ for general $\mathcal{P}$, basic notions

The game $\gamma(\mathcal{P}) = \gamma(\Phi, \mathcal{P})$ .·. $\Phi$ objective function, Nature maximizer, Observer minimizer. Nature strategies: $x$'s in $\mathcal{P}$. Observer strategies: beliefs $y \succ \mathcal{P}$ ($\forall x \in \mathcal{P} : y \succ x$).

MaxEnt is value for Nature, MinRisk value for Observer:
$\mathrm{H}_{\max}(\mathcal{P}) = \sup_{x\in\mathcal{P}} \mathrm{H}(x) = \sup_{x\in\mathcal{P}} \inf_{y\succ x} \Phi(x, y)$.
$\mathrm{Ri}_{\min}(\mathcal{P}) = \inf_{y\succ\mathcal{P}} \mathrm{Ri}(y) = \inf_{y\succ\mathcal{P}} \sup_{x\in\mathcal{P}} \Phi(x, y)$.
Note: $\mathrm{Ri}(y) = \mathrm{Ri}(y|\mathcal{P})$.

$x^* \in \mathcal{P}$ optimal for Nature .·. $\mathrm{H}(x^*) = \mathrm{H}_{\max}(\mathcal{P})$.
$y^* \succ \mathcal{P}$ optimal for Observer .·. $\mathrm{Ri}(y^*) = \mathrm{Ri}_{\min}(\mathcal{P})$.

If $\mathrm{H}_{\max}(\mathcal{P}) = \mathrm{Ri}_{\min}(\mathcal{P})$ is finite, $\gamma(\mathcal{P})$ is in equilibrium.

Strategies $(x^*, y^*)$ is a Nash equilibrium pair (NE-pair) if $\forall x \in \mathcal{P} \forall y \succ \mathcal{P} : \Phi(x, y^*) \le \Phi(x^*, y^*) \le \Phi(x^*, y)$.

# the games $\gamma(\mathcal{P})$ for general $\mathcal{P}$, basic notions

The game $\gamma(\mathcal{P}) = \gamma(\Phi, \mathcal{P})$ .∴ $\Phi$ objective function, Nature maximizer, Observer minimizer. Nature strategies: $x$'s in $\mathcal{P}$. Observer strategies: beliefs $y \succ \mathcal{P}$ ($\forall x \in \mathcal{P} : y \succ x$).

MaxEnt is value for Nature, MinRisk value for Observer:
$\mathrm{H}_{\max}(\mathcal{P}) = \sup_{x \in \mathcal{P}} \mathrm{H}(x) = \sup_{x \in \mathcal{P}} \inf_{y \succ x} \Phi(x, y)$.
$\mathrm{Ri}_{\min}(\mathcal{P}) = \inf_{y \succ \mathcal{P}} \mathrm{Ri}(y) = \inf_{y \succ \mathcal{P}} \sup_{x \in \mathcal{P}} \Phi(x, y)$.
Note: $\mathrm{Ri}(y) = \mathrm{Ri}(y|\mathcal{P})$.

$x^* \in \mathcal{P}$ optimal for Nature .∴ $\mathrm{H}(x^*) = \mathrm{H}_{\max}(\mathcal{P})$.
$y^* \succ \mathcal{P}$ optimal for Observer .∴ $\mathrm{Ri}(y^*) = \mathrm{Ri}_{\min}(\mathcal{P})$.

If $\mathrm{H}_{\max}(\mathcal{P}) = \mathrm{Ri}_{\min}(\mathcal{P})$ is finite, $\gamma(\mathcal{P})$ is in equilibrium.

Strategies $(x^*, y^*)$ is a Nash equilibrium pair (NE-pair) if $\forall x \in \mathcal{P} \forall y \succ \mathcal{P} : \Phi(x, y^*) \le \Phi(x^*, y^*) \le \Phi(x^*, y)$.

$y^* \succ \mathcal{P}$ is robust for $\gamma(\mathcal{P})$ if $\exists h < \infty \forall x \in \mathcal{P} : \Phi(x, y^*) = h$, the level of robustness.

## the games $\gamma(\mathcal{P})$, general results

The center of $\mathcal{P}$ is the set $\text{cen}(\mathcal{P}) = \{x^* \in \mathcal{P} | x^* \succ \mathcal{P}\}$.

# the games $\gamma(\mathcal{P})$, general results

The center of $\mathcal{P}$ is the set $\text{cen}(\mathcal{P}) = \{x^* \in \mathcal{P} | x^* \succ \mathcal{P}\}$.

**Identification** Let $(x^*, y^*)$ be strategies with $x^* \in \text{cen}(\mathcal{P})$ and $\text{H}(x^*) < \infty$. Then $\gamma(\mathcal{P})$ is in equilibrium with $(x^*, y^*)$ optimal strategies *iff* $(x^*, y^*)$ is a NE-pair. For this, $y^* = x^*$ must hold.

**Pythagorean inequalities** Let $(x^*, y^*)$ be strategies with $y^* = x^*$, $x^* \in \text{cen}(\mathcal{P})$, $\text{H}(x^*) < \infty$ and assume that $\forall x \in \mathcal{P} : \Phi(x, y^*) \leq \Phi(x^*, y^*)$.
Then $\gamma(\mathcal{P})$ is in equilibrium with $x^*$ and $y^* = x^*$ as unique optimal strategies ($x^*$ is the bioptimal strategy). Furthermore:
$\forall x \in \mathcal{P} : \text{H}(x) + \text{D}(x, y^*) \leq \text{H}_{\text{max}}(\mathcal{P})$ and
$\forall y \succ P : \text{Ri}_{\text{min}}(\mathcal{P}) + \text{D}(x^*, y) \leq \text{Ri}(y|\mathcal{P})$.

**Robustness** Assume that $y^*$ is robust with level of robustness $h$. Put $x^* = y^*$ and assume that $x^* \in \mathcal{P}$. Then $\gamma(\mathcal{P})$ is in equilibrium with $(x^*, y^*)$ as unique optimal strategies. Furthermore, $\forall x \in \mathcal{P} : \text{H}(x) + \text{D}(x, y^*) = \text{H}_{\text{max}}(\mathcal{P})$.

# main results reformulated

Inspection reveals significance of the previously introduced basic strict and basic slack feasible preparations. Expressed in terms of these sets we find that:

**The Pythagorean theorem, reformulated** Assume that $x^* \in \mathcal{P} \subseteq \mathcal{P}^{x^*}(h^\bullet)$ with $h = \mathrm{H}(x^*)$.

Then $x^*$ is the MaxEnt strategy, $\mathrm{H}_{\max}(\mathcal{P}) = h$ and,

$\forall x \in \mathcal{P} : \mathrm{H}(x) + \mathrm{D}(x, x^*) \leq h$.

(... plus more, bioptimality of $x^*$ ...).

If $\mathcal{P} \subseteq \mathcal{P}^{x^*}(h)$, equality holds above.

# main results reformulated

Inspection reveals significance of the previously introduced basic strict and basic slack feasible preparations. Expressed in terms of these sets we find that:

**The Pythagorean theorem, reformulated** Assume that $x^* \in \mathcal{P} \subseteq \mathcal{P}^{x^*}(h^\bullet)$ with $h = \mathrm{H}(x^*)$.

Then $x^*$ is the MaxEnt strategy, $\mathrm{H}_{\max}(\mathcal{P}) = h$ and,

$\forall x \in \mathcal{P} : \mathrm{H}(x) + \mathrm{D}(x, x^*) \leq h.$

(... plus more, bioptimality of $x^*$ ...).

If $\mathcal{P} \subseteq \mathcal{P}^{x^*}(h)$, equality holds above.

This as an abstract version of the Pythagorean (in)equality! To realize this, consider the updating model in Hilbert space ...

# Exponential families

**Idea** Given a preparation family $\mathbb{P}$, the associated exponential family $\mathcal{E}$ is the set of all "naturally occuring" candidates to (bi)optimal strategies for one of the preparations $\mathcal{P} \in \mathbb{P}$.

# Exponential families

**Idea** Given a preparation family $\mathbb{P}$, the associated exponential family $\mathcal{E}$ is the set of all "naturally occuring" candidates to (bi)optimal strategies for one of the preparations $\mathcal{P} \in \mathbb{P}$.

As preparation families take the families $\mathbb{P}^{\mathbf{b}}$ of strict feasible preparations (not the slack ones as ...) and as "naturally occuring" candidates we take the robust strategies. Thus:

The exponential family $\mathcal{E}^{\mathbf{b}}$ is the set of $y^* \in X$ which are robust for all preparations in $\mathbb{P}^{\mathbf{b}}$. By robustness theorem:

# Exponential families

**Idea** Given a preparation family $\mathbb{P}$, the associated exponential family $\mathcal{E}$ is the set of all "naturally occuring" candidates to (bi)optimal strategies for one of the preparations $\mathcal{P} \in \mathbb{P}$.

As preparation families take the families $\mathbb{P}^{\mathbf{b}}$ of strict feasible preparations (not the slack ones as ...) and as "naturally occuring" candidates we take the robust strategies. Thus:

The exponential family $\mathcal{E}^{\mathbf{b}}$ is the set of $y^* \in X$ which are robust for all preparations in $\mathbb{P}^{\mathbf{b}}$. By robustness theorem:

**Theorem** Assume that $x^* \in \mathcal{E}^{\mathbf{b}}$. For $i \leq n$, put $h_i = \Phi(x^*, b_i)$. Then $\gamma(\mathcal{P}^{\mathbf{b}}(\mathbf{h}))$ is in equilibrium and has $x^*$ as bioptimal strategy. In particular, $x^*$ is the MaxEnt strategy for $\mathcal{P}^{\mathbf{b}}(\mathbf{h})$.

# Exponential families

**Idea** Given a preparation family $\mathbb{P}$, the associated exponential family $\mathcal{E}$ is the set of all "naturally occuring" candidates to (bi)optimal strategies for one of the preparations $\mathcal{P} \in \mathbb{P}$.

As preparation families take the families $\mathbb{P}^{\mathbf{b}}$ of strict feasible preparations (not the slack ones as ...) and as "naturally occuring" candidates we take the robust strategies. Thus:

The exponential family $\mathcal{E}^{\mathbf{b}}$ is the set of $y^* \in X$ which are robust for all preparations in $\mathbb{P}^{\mathbf{b}}$. By robustness theorem:

**Theorem** Assume that $x^* \in \mathcal{E}^{\mathbf{b}}$. For $i \leq n$, put $h_i = \Phi(x^*, b_i)$. Then $\gamma(\mathcal{P}^{\mathbf{b}}(\mathbf{h}))$ is in equilibrium and has $x^*$ as bioptimal strategy. In particular, $x^*$ is the MaxEnt strategy for $\mathcal{P}^{\mathbf{b}}(\mathbf{h})$.

**Example:** Updating model in Hilbert space ...

# Limits to information

What *can* we know?

Full information ("$x$") normally not feasible.

partial information "$x \in \mathcal{P}$" could be.

So, which are the feasible preparations?

Answer (again!): Level (or sublevel) sets and their finite intersecions!

This is partly justified by previous results.

For further motivation recall: "Belief is a tendency to act".

Action through experiments.

Experiments require control.

Control depends on description.

**Postulate** Belief can be transformed into new objects, controls by a bijective correspondance $y \longleftrightarrow w$ between $Y$ and a new set, the action space $W$. We write $w = \hat{y}$ or $y = \check{w}$.

# Exponential families as a set of controls

Controls are technically superfluous but convenient!
Description effort is transformed to $\Psi$ given by
$\Psi(x, w) = \Phi(x, \check{w})$. Corresponding games: $\gamma(\Psi, \mathcal{P})$.

# Exponential families as a set of controls

Controls are technically superfluous but convenient!
Description effort is transformed to $\Psi$ given by $\Psi(x, w) = \Phi(x, \check{w})$. Corresponding games: $\gamma(\Psi, \mathcal{P})$.

Preparation families are then given in terms of controls $\mathbf{w} = (w_1, \cdots, w_n)$:

$$^{\Psi}\mathbb{P}^{\mathbf{w}} = \{^{\Psi}\mathcal{P}^{\mathbf{w}}(\mathbf{h}) = \bigcap_{i \leq n} {}^{\Psi}\mathcal{P}^{w_i}(h_i) | \mathbf{h} \cdots \}$$

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

# Exponential families as a set of controls

Controls are technically superfluous but convenient!
Description effort is transformed to $\Psi$ given by $\Psi(x, w) = \Phi(x, \check{w})$. Corresponding games: $\gamma(\Psi, \mathcal{P})$.

Preparation families are then given in terms of controls $\mathbf{w} = (w_1, \cdots, w_n)$:

$${}^{\Psi}\mathbb{P}^{\mathbf{w}} = \{{}^{\Psi}\mathcal{P}^{\mathbf{w}}(\mathbf{h}) = \bigcap_{i \le n} {}^{\Psi}\mathcal{P}^{w_i}(h_i) | \mathbf{h} \cdots \}$$

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

The exponential family for ${}^{\Psi}\mathbb{P}^{\mathbf{w}}$ in terms of controls is the set of $w^* \in W$ which are robust for all games $\gamma(\Psi, \mathcal{P})$ with $\mathcal{P} \in {}^{\Psi}\mathbb{P}^{\mathbf{w}}$. By robustness:

# Exponential families as a set of controls

Controls are technically superfluous but convenient!
Description effort is transformed to $\Psi$ given by $\Psi(x,w)=\Phi(x,\check{w})$. Corresponding games: $\gamma(\Psi,\mathcal{P})$.

Preparation families are then given in terms of controls $\mathbf{w}=(w_1,\cdots,w_n)$:

$${}^{\Psi}\mathbb{P}^{\mathbf{w}}=\{{}^{\Psi}\mathcal{P}^{\mathbf{w}}(\mathbf{h})=\bigcap_{i\le n}{}^{\Psi}\mathcal{P}^{w_i}(h_i)|\mathbf{h}\cdots\}$$

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

The exponential family for ${}^{\Psi}\mathbb{P}^{\mathbf{w}}$ in terms of controls is the set of $w^*\in W$ which are robust for all games $\gamma(\Psi,\mathcal{P})$ with $\mathcal{P}\in{}^{\Psi}\mathbb{P}^{\mathbf{w}}$. By robustness:

> Let $x^*\in X$, assume $w^*=\hat{x^*}$ is in the exponential family for ${}^{\Psi}\mathbb{P}^{\mathbf{w}}$ For $i\le n$, put $h_i=\Psi(x^*,w_i)$. Then $\gamma\big(\Psi,{}^{\Psi}\mathcal{P}^{\mathbf{w}}(\mathbf{h})\big)$ is in equilibrium and has $x^*$ and $w*$ as optimal strategies. In particular, $x^*$ is the MaxEnt strategy for ${}^{\Psi}\mathcal{P}^{\mathbf{w}}(\mathbf{h})$.

# Example: Probabilistic models, discrete case

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 $\mathbb{A}$.

# Example: Probabilistic models, discrete case

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

# Example: Probabilistic models, discrete case

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

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

We require description to be accumulated effort:

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

where $\kappa$, the descriptor gives the *cost of information*.

# accumulated effort, the one and only

Requirements: $\kappa$ is smooth, $\kappa(1) = 0$, $\kappa'(1) = -1$. Gives natural units , nats.

# accumulated effort, the one and only

Requirements: $\kappa$ is smooth, $\kappa(1) = 0$, $\kappa'(1) = -1$. Gives natural units, nats.For "any" monotone function on $[0, 1]$, a linear combination defines a descriptor. The power hierarchy is defined from $t \mapsto t^{q-1}$ with $q = 1$ special case, giving $\kappa_1(t) = \ln \frac{1}{t}$.

**Theorem** Given $\pi = \pi(s, t)$, there is at most *one* descriptor $\kappa$ which can define a proper accumulated description effort.

## accumulated effort, the one and only

Requirements: $\kappa$ is smooth, $\kappa(1) = 0$, $\kappa'(1) = -1$. Gives natural units, nats.For "any" monotone function on $[0, 1]$, a linear combination defines a descriptor. The power hierarchy is defined from $t \mapsto t^{q-1}$ with $q = 1$ special case, giving $\kappa_1(t) = \ln \frac{1}{t}$.

**Theorem** Given $\pi = \pi(s, t)$, there is at most *one* descriptor $\kappa$ which can define a proper accumulated description effort.

If $\pi$ is consistent, hence one of the $\pi_q$'s, then there exists a descriptor which generates a proper description effort iff $q > 0$ ($q = 0$ OK as a singular case, though).

# accumulated effort, the one and only

Requirements: $\kappa$ is smooth, $\kappa(1) = 0$, $\kappa'(1) = -1$. Gives natural units , nats.For "any" monotone function on $[0, 1]$, a linear combination defines a descriptor. The power hierarchy is defined from $t \mapsto t^{q-1}$ with $q = 1$ special case, giving $\kappa_1(t) = \ln \frac{1}{t}$.

**Theorem** Given $\pi = \pi(s, t)$, there is at most *one* descriptor $\kappa$ which can define a proper accumulated description effort.

If $\pi$ is consistent, hence one of the $\pi_q$'s, then there exists a descriptor which generates a proper description effort iff $q > 0$ ($q = 0$ OK as a singular case, though).

If so, the descriptor is the one in the power hierarchy, i.e. $\kappa_q(t) = \ln_q \frac{1}{t} = \frac{t^{q-1}-1}{1-q}$. The associated information triple is the power triple . The power entropies are the Tsallis entropies , and the power divergences are Bregman divergences.

## gross effort, pointwise fundamental inequality

**Claim:** The unit of information we chose is an overhead, connected with experiments/observations. Why?

## gross effort, pointwise fundamental inequality

**Claim:** The unit of information we chose is an overhead , connected with experiments/observations. Why?

Introduce gross $\Phi$ and gross H by adding the overhead:

# gross effort, pointwise fundamental inequality

**Claim:** The unit of information we chose is an overhead , connected with experiments/observations. Why?

Introduce gross $\Phi$ and gross H by adding the overhead:

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

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

# gross effort, pointwise fundamental inequality

**Claim:** The unit of information we chose is an overhead, connected with experiments/observations. Why?

Introduce gross $\Phi$ and gross $\mathrm{H}$ by adding the overhead:

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

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

Clearly, $\tilde{\mathrm{D}} = \mathrm{D}$, 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 $\mathrm{D}(x,y) = \sum \delta(x_i, y_i)$.

## gross effort, pointwise fundamental inequality

**Claim:** The unit of information we chose is an overhead , connected with experiments/observations. Why?
Introduce gross $\Phi$ and gross H by adding the overhead:

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

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

Clearly, $\tilde{\mathrm{D}} = \mathrm{D}$, 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
$\mathrm{D}(x,y) = \sum \delta(x_i,y_i)$.

The inequality $\delta \geq 0$ is the pointwise fundamental inequality (PFI). Clearly PFI $\Longrightarrow$ FI. **Conjecture** Converse also true

# sketch of MaxEnt determination for $\mathcal{W}_q$

Consider the world $\mathcal{W} = \mathcal{W}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then ${}^{\Psi}\mathbb{P}^{w}$ consists of all $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.

# sketch of MaxEnt determination for $\mathcal{W}_q$

Consider the world $\mathcal{W} = \mathcal{W}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then ${}^{\Psi}\mathbb{P}^{w}$ consists of all $\mathcal{P}$ 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 $\mathcal{W}_q$

Consider the world $\mathcal{W} = \mathcal{W}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then ${}^{\Psi}\mathbb{P}^{w}$ consists of all $\mathcal{P}$ 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 $\alpha$ and $\beta$ this implies that $\sum x_i(\alpha + \beta w_i)$ is constant over $\mathcal{P}$.

# sketch of MaxEnt determination for $\mathcal{W}_q$

Consider the world $\mathcal{W} = \mathcal{W}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then ${}^{\Psi}\mathbb{P}^w$ consists of all $\mathcal{P}$ 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 $\alpha$ and $\beta$ 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. $w^* \in {}^{\Psi}\mathcal{E}^w$ and robustness applies.

# sketch of MaxEnt determination for $\mathcal{W}_q$

Consider the world $\mathcal{W} = \mathcal{W}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then ${}^{\Psi}\mathbb{P}^{w}$ consists of all $\mathcal{P}$ 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 $\alpha$ and $\beta$ 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. $w^* \in {}^{\Psi}\mathcal{E}^{w}$ and robustness applies.
Then, given $\beta$, try to adjust $\alpha$ so that $\alpha + \beta w$ is a control.
Classically, $\alpha$ is the logarithm of the partition function. .
Finally, adjust $\beta$ ($\approx$ inverse temperature) to desired level ...

Similarly, the updating models are handled ...

# Conclusion

A theory of information freed from a tie to probability *is* possible – and useful. Probabilistic models appear as important examples.

# Conclusion

A theory of information freed from a tie to probability *is* possible – and useful. Probabilistic models appear as important examples.

Than you!