Free Energy & Bounded Rationality

Pedro A. Ortega
Daniel A. Braun

Max Planck Institute for Intelligent Systems
Max Planck Institute for Biological Cybernetics

13th September 2012

Introduction

The mathematical foundation of

- economics,
- artificial intelligence,
- and control

is the **theory of (subjective) expected utility**, leading to the **maximum expected utility (MEU) principle**.

Introduction

The mathematical foundation of

- economics,
- artificial intelligence,
- and control

is the **theory of (subjective) expected utility**, leading to the **maximum expected utility (MEU) principle**.

However:

- **Exact** application of the MEU principle is **intractable** even for extremely simple systems.

$\Rightarrow$ We need a theory of **bounded rationality** that considers the **cost of choice**.

Caveat: Metareasoning does not work!

Most straightforward solution: penalize choice costs

- desired behavior: U
- reasoning about costs: $U' := U - C$

Caveat: Metareasoning does not work!

Most straightforward solution: penalize choice costs

- desired behavior: U
- reasoning about costs: $U' := U - C$
- reasoning about costs of costs: $U'' := U' - C'$

Caveat: Metareasoning does not work!

Most straightforward solution: penalize choice costs

- desired behavior: U
- reasoning about costs: $U' := U - C$
- reasoning about costs of costs: $U'' := U' - C'$
- ...

Caveat: Metareasoning does not work!

Most straightforward solution: penalize choice costs

- desired behavior: U
- reasoning about costs: $U' := U - C$
- reasoning about costs of costs: $U'' := U' - C'$
- ...

Problem of metareasoning:

- Unbounded meta-levels + growing solution spaces.
- Metareasoning is **not allowed** $\longrightarrow$ "interrupted" decision!

Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Our Answer: A **bounded rational** decision maker **can be thought of** as maximizing the negative free energy difference/KL control cost

- (one-step)

$$\sum_{x} p(x)\left\{ U(x) - \frac{1}{\alpha} \log \frac{p(x)}{q(x)} \right\}$$

- (multi-step)

$$\sum_{x_{\leq T}} p(x_{\leq T}) \sum_{t=1}^{T} \left\{ R(x_t|x_{<t}) - \frac{1}{\beta(x_{<t})} \log \frac{p(x_t|x_{<t})}{q(x_t|x_{<t})} \right\}$$

Bounded Rationality

Question: How do we **characterize** behavior when the decision maker is **bounded rational**, i.e. when his **processing resources are limited**?

Our Answer: A **bounded rational** decision maker **can be thought of** as maximizing the negative free energy difference/KL control cost

- (one-step)

$$\sum_x p(x)\left\{U(x) - \frac{1}{\alpha}\log\frac{p(x)}{q(x)}\right\}$$

- (multi-step)

$$\sum_{x_{\leq T}} p(x_{\leq T}) \sum_{t=1}^{T}\left\{R(x_t|x_{<t}) - \frac{1}{\beta(x_{<t})}\log\frac{p(x_t|x_{<t})}{q(x_t|x_{<t})}\right\}$$

Why? Result is based on an information-theoretic assumption about transformation costs, i.e. the cost of "changing".

The Cost of Transformations

Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

The Cost of Transformations

Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

Examples:

- Biologists infer behavior from anatomy. Energy-efficient behavior is more frequent than energy-inefficient behavior.
- Conversely, engineers design systems such that desirable behavior is cheaper than undesirable behavior.
- Every action/observation/interaction of a system **necessarily** transforms its **information state**, simply because "before" and "after" are distinguishable!

The Cost of Transformations

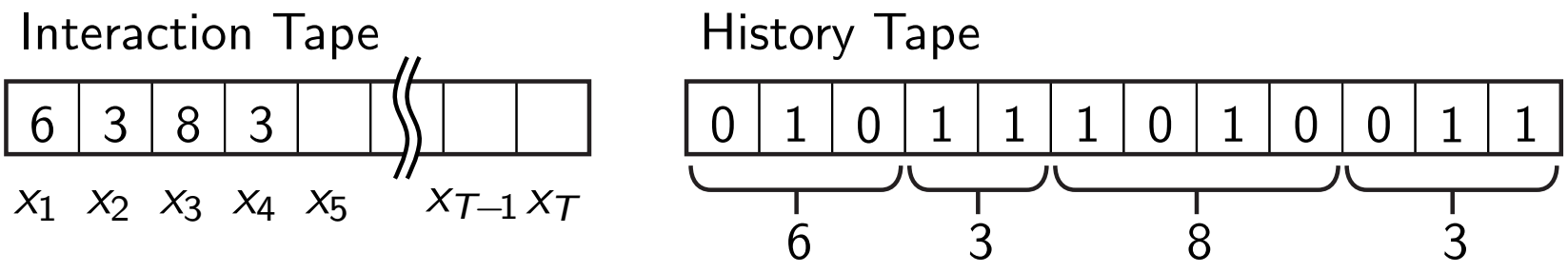
Our Fundamental Assumption: The difficulty of producing an event determines its probability ("probabilities encode costs").

Examples:

- Biologists infer behavior from anatomy. Energy-efficient behavior is more frequent than energy-inefficient behavior.
- Conversely, engineers design systems such that desirable behavior is cheaper than undesirable behavior.
- Every action/observation/interaction of a system **necessarily** transforms its **information state**, simply because "before" and "after" are distinguishable!

What is a Transformation? Chemical reaction, memory update, consulting a random number generator, changing location, advancing in time, ...

The Model of Information State

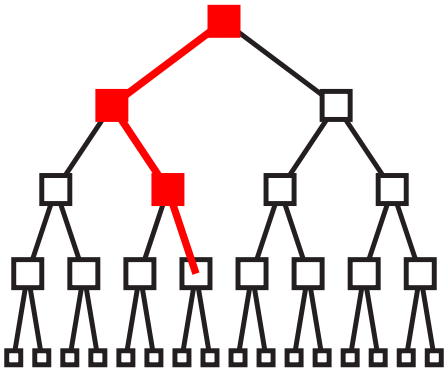


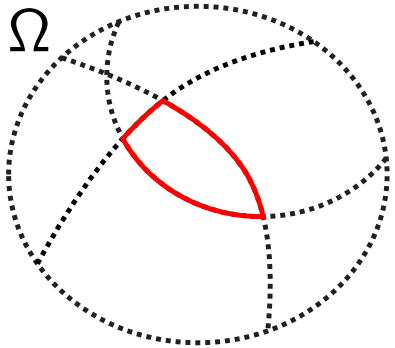
- Each interaction transforms the information state.
- Interactions are encoded (lossless) on a binary "history tape".
- No "jumps back in time" allowed.
- Tape consists of identical binary storage devices.
- Setting a bit costs the same in each cell.

$\Rightarrow$ Codeword lengths are **proxies** for transformation costs.

$\Rightarrow$ Codeword lengths have **associated** probabilities.

Measure-Theoretic Formalization of Transformations





- **Sequential realizations** are modeled as **filtrations**.
- An **information state** is a measurable set.
- A transformation is a **condition** on the information state:

$$\begin{array}{lcc} \text{State:} & A & \\ \text{Measure:} & P(S|A) & \end{array} \longrightarrow \text{“}B\text{ is true”} \longrightarrow \begin{array}{c} (A \cap B) \\ P(S|A \cap B) \end{array}$$

Axioms of Transformation Costs

Given:

- (Ω, Σ) measurable space
- $P(\cdot|\cdot) : (\Omega \times \Omega) \to [0, 1]$ conditional probability measure.

Then, $\rho(\cdot|\cdot) : (\Sigma \times \Sigma) \to \mathbb{R}^+$ is **transformation cost function** iff

A1. real-valued: $\exists f, \quad \rho(A|B) = f(P(A|B)) \in \mathbb{R}$

A2. additive: $\rho(A \cap B|C) = \rho(B|C) + \rho(A|B \cap C)$

A3. monotonic: $\rho(A|B) > \rho(C|D) \iff P(A|B) \lesseqgtr P(C|D)$

Axioms of Transformation Costs

Given:

- (Ω, Σ) measurable space
- $P(\cdot|\cdot) : (\Omega \times \Omega) \to [0, 1]$ conditional probability measure.

Then, $\rho(\cdot|\cdot) : (\Sigma \times \Sigma) \to \mathbb{R}^+$ is **transformation cost function** iff

A1. real-valued: $\exists f, \quad \rho(A|B) = f(P(A|B)) \in \mathbb{R}$

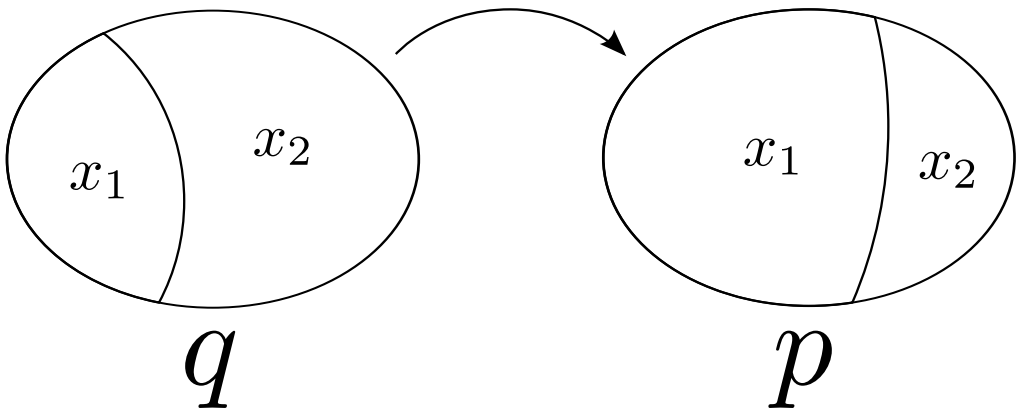
A2. additive: $\rho(A \cap B|C) = \rho(B|C) + \rho(A|B \cap C)$

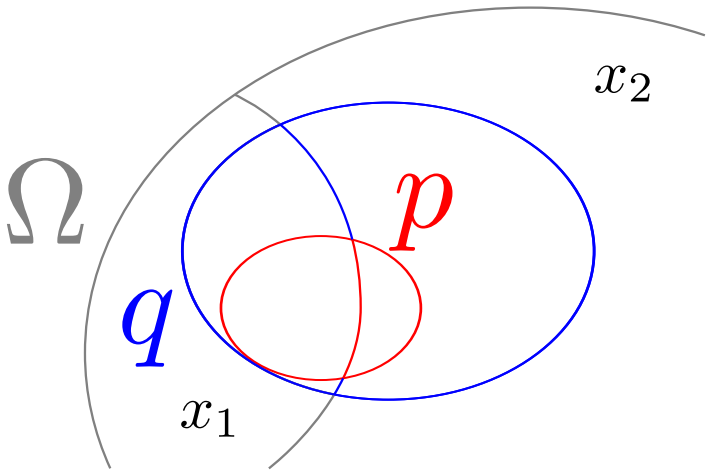
A3. monotonic: $\rho(A|B) > \rho(C|D) \iff P(A|B) \lesseqgtr P(C|D)$

Theorem: If f fulfills axioms A1–A3 for any (Ω, Σ, P), then f is of the form

$$\rho(A|B) = -\frac{1}{\alpha} \log(P(A|B)), \qquad \alpha \in \mathbb{R}.$$

Measure-Theoretic Formalization of Decisions





Decisions

Problem:

- Given $\mathcal{X}$, and $U(x)$, find $p(x)$ maximizing

$$\sum_{x \in \mathcal{X}} p(x) U(x).$$

Constraints:

- However, there are many candidate $p \in \mathcal{P}$, having probabilities & costs

$$P(p|q) \qquad \text{and} \qquad \rho(p|q)$$

  from some **reference information state** q.
- We define utility as **cost that is "saved"** (analogous to external work)

$$u(A|B) = -\rho(A|B)$$

Decisions (cont.)

- Identifying:

$$
\begin{aligned}
q(x) &= P(x|q) && \text{(Prior)} \\
p(x) &= P(x|q \cap p) && \text{(Posterior)} \\
U(x) &= u(p|x \cap q) - u(p|q) \\
&= u(x|q \cap p) - u(x|q) && \text{(Utility)}
\end{aligned}
$$

  we obtain (theorem)

$$
u(p|q) = \sum_x p(x)U(x) - \frac{1}{\alpha}\sum_x p(x)\log\frac{p(x)}{q(x)}.
$$

- This is the **negative free energy difference** (NFED).

Free Energy Principle

Let q be a probability distribution and U be a real-valued utility over $\mathcal{X}$. Given $\alpha \in \mathbb{R}$, the **negative free energy difference (NFED)** is given by

$$-\Delta F_\alpha[p] := \sum_x p(x)U(x) - \frac{1}{\alpha}\sum_x P(x)\log\frac{p(x)}{q(x)}.$$

Interpretation

- NFED = expected utility - transformation costs
- models **net utility gain** obtained in transforming q into p
- relative entropy models information content of transformation
- inverse temperature α models (transformation-) bits per utile
- higher inverse temperature $\longrightarrow$ higher net utility gain

Equilibrium Distribution

The solution to the NFED is the **equilibrium distribution**

$$p(x) = \frac{1}{Z(\alpha)} q(x) \exp\{\alpha U(x)\},$$

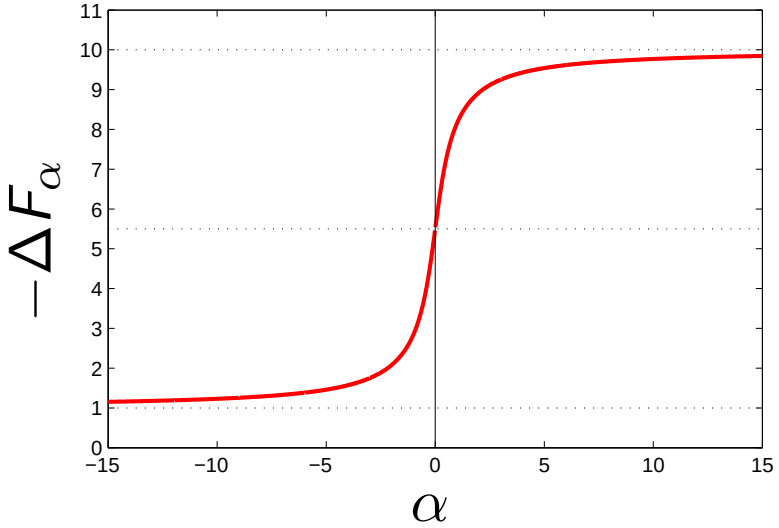
where $Z(\alpha)$ is the **partition function**

$$Z(\alpha) = \sum_x q(x) \exp\{\alpha U(x)\}.$$

The **NFED extremum** is

$$\frac{1}{\alpha} \log Z(\alpha) = \frac{1}{\alpha} \log \left(\sum_x q(x) \exp\left\{\alpha U(x)\right\} \right).$$

NFED Extremum



The inverse temperature α parameterizes the **degree of control**:

$$
\begin{array}{llllll}
\alpha \to \infty & : & \frac{1}{\alpha} \log Z & \longrightarrow & \max U(x) & \text{(maximum)} \\
\alpha \to 0 & : & \frac{1}{\alpha} \log Z & \longrightarrow & \mathbf{E}_x[U(x)] & \text{(expectation)} \\
\alpha \to -\infty & : & \frac{1}{\alpha} \log Z & \longrightarrow & \min U(x) & \text{(minimum)}
\end{array}
$$

Operational Interpretation of Inverse Temperature

Problem

- Let M be pmf over finite $\mathcal{X}$.
- Draw α i.i.d. samples $x_1, \ldots, x_\alpha$ from M.
- Pick the maximum $\max\{U(x_0), \ldots, U(x_\alpha)\}$.

Theorem

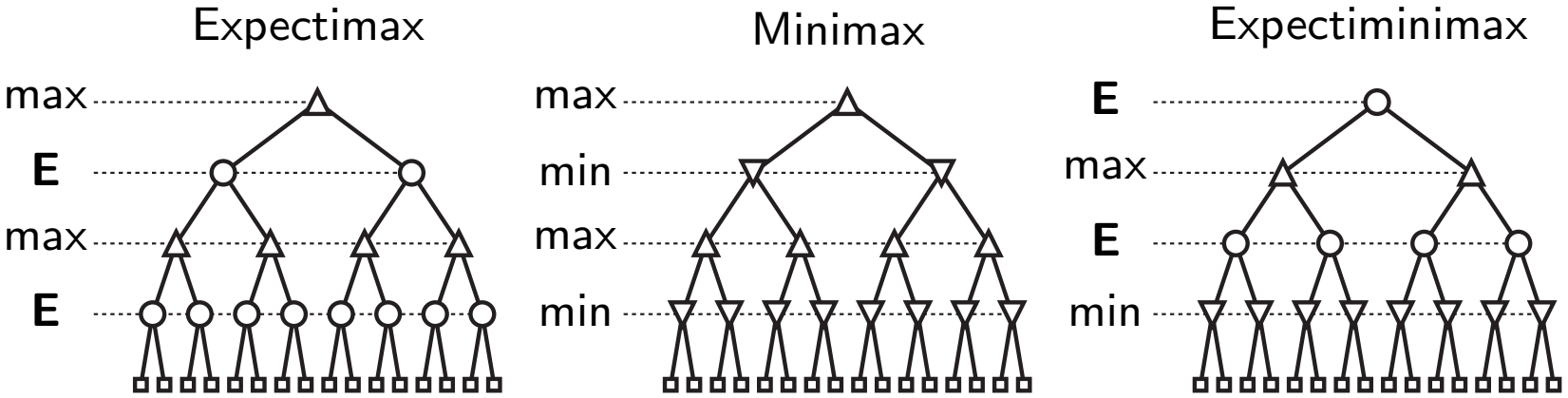
- *Let Q be pmf with same support as M.*
- *Let M_α be the pmf over the maximizing x after α draws.*
- *Then, there are $\delta > 0$ and $\xi > 0$ depending only on M such that for all α,*

$$\left| \frac{Q(x)e^{\alpha U(x)}}{\sum_{x'} Q(x')e^{\alpha U(x')}} - M_\alpha(x) \right| \leq e^{-(\alpha-\xi)\delta}.$$

Intuition

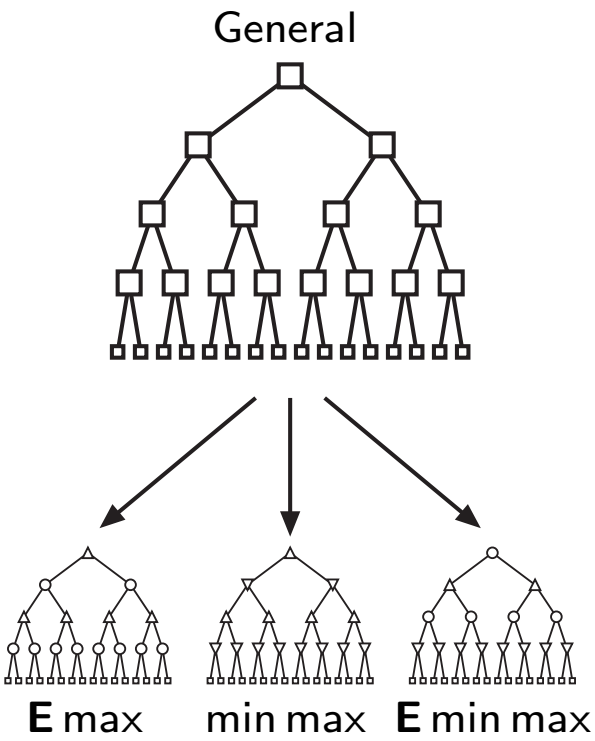
$$M_\alpha = \{\alpha \text{ iterations of “search algorithm”}\}.$$

Decision Trees



- Sequential decision problems are stated as decision trees and solved using backward induction.
- Decision rules depend on system: stochastic, cooperative, competitive, hybrid, . . .
- This intuitive distinction between "types of systems" is formally unsatisfactory.
- Decision rules can be reexpressed in a unified way using the free energy functional.

Goal: Generalized Decision Trees



- Different operators express different degrees of control (DoCs):
  - max $\Leftrightarrow$ full control
  - **E** $\Leftrightarrow$ no control
  - min $\Leftrightarrow$ full anti-control
- Goal: Find a generalized operator $\square$ that expresses
  - the 3 classical DoCs,
  - $+$ all the other DoCs in between.

Change of Temperature

Problem

Can we change the inverse temperature with constant reference and equilibrium distribution?

Theorem

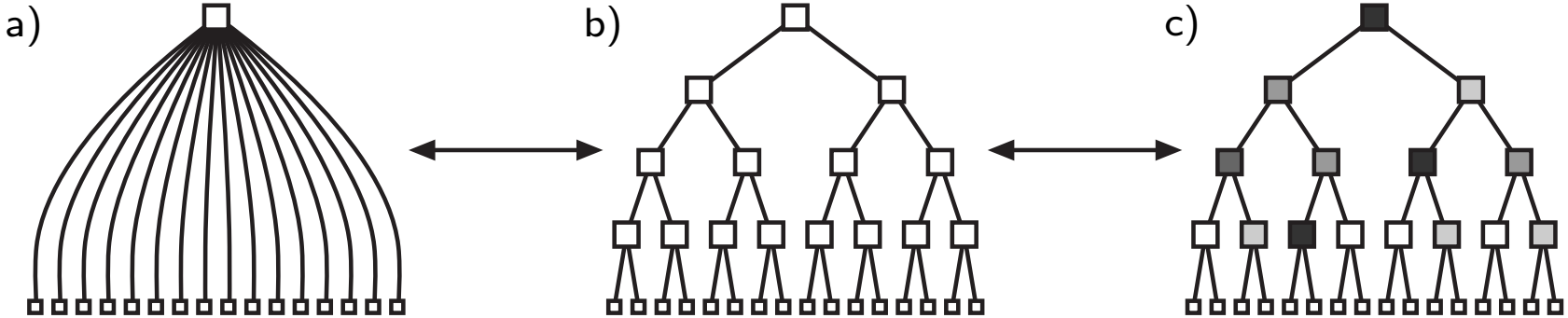
Let p be the equilibrium distribution given α, U and q.
If *α changes to β with fixed p and q,* **then** *U changes to V:*

$$V(x) = U(x) - \left(\frac{1}{\alpha} - \frac{1}{\beta}\right) \log \frac{p(x)}{q(x)}.$$

Intuition

Fix information costs: $\quad C(x) = \alpha U(x) = \beta V(x)$

Construction of Generalized Decision Trees



a) $q(x), U(x), \alpha$

$$\sum_x p(x)U(x) + \frac{1}{\alpha}\sum_x p(x)\log\frac{p(x)}{q(x)}$$

b) $q(x_t|x_{1:t-1}), S(x_t|x_{1:t}), \alpha$

$$\sum_{x_{\leq T}} p(x_{\leq T})\sum_{t=1}^{T}\left\{S(x_t|x_{<t}) + \frac{1}{\alpha}\log\frac{p(x_t|x_{<t})}{q(x_t|x_{<t})}\right\}$$

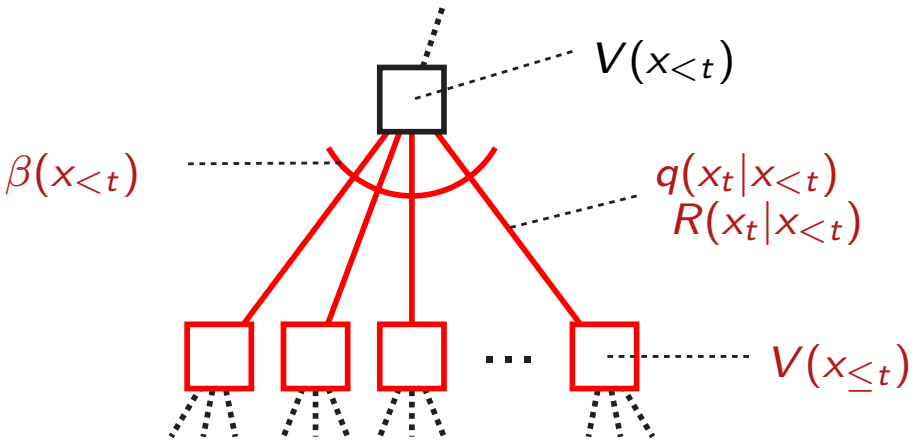
c) $q(x_t|x_{<t}), R(x_t|x_{<t}), \beta(x_{<t})$

$$\sum_{x_{\leq T}} p(x_{\leq T})\sum_{t=1}^{T}\left\{R(x_t|x_{<t}) + \frac{1}{\beta(x_{<t})}\log\frac{p(x_t|x_{<t})}{q(x_t|x_{<t})}\right\}$$

Generalized Optimality Equations

Given

Generalized decision problem $q(x_t|x_{<t})$, $R(x_t|x_{<t})$ and $\beta(x_{<t})$.



Generalized Value/Utility

$$V(x_{<t}) = \frac{1}{\beta(x_{<t})} \log\left\{ \sum_{x_t} q(x_t|x_{<t}) \exp\left\{ \beta(x_{<t}) \left[R(x_t|x_{<t}) + V(x_t) \right] \right\} \right\}$$

Conclusions

1. The free energy principle serves as an **axiomatic foundation** for bounded rational decision-making.
2. It formalizes a **trade-off** between the gains of maximizing the utility and the losses of transformation costs.
3. It establishes clear **links to** information theory and thermodynamics.
4. Inverse temperature **parameterizes** the resource limitations/degree of control.
5. It allows **generalizing** decision trees.

Open Questions

1. What are the **exact** relations to:
   - game theory,
   - search theory,
   - and computational complexity?
2. What are the implications for search algorithms?
3. What are the causal implications?

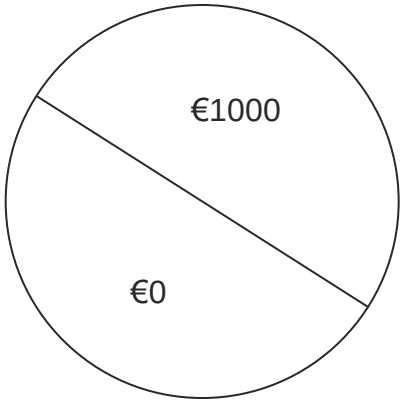
References

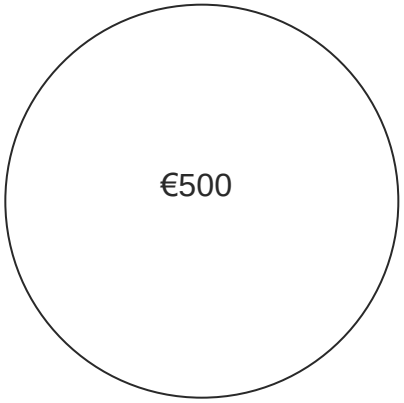
1. Ortega, P.A. and Braun, D.A. *Thermodynamics as a Theory of Decision Making with Information Processing Costs*. ArXiv:1204.6481, 2012.
2. Ortega, P.A. *Free Energy and the Generalized Optimality Equations*. European Workshop on Reinforcement Learning, 2012.
3. Braun, D.A., Ortega, P.A., Theodorou, E. and Schaal, S. *Path Integral Control and Bounded Rationality*, IEEE Symposium on adaptive dynamic programming and reinforcement learning, pp. 202–209, 2011.
4. Ortega, P.A. *A Unified Framework for Resource-Bounded Agents Interacting with Unknown Environments*, PhD Thesis, Department of Engineering, University of Cambridge, 2011.

The free energy principle in human sensorimotor control

Daniel Braun, Pedro Ortega

Risk in Decision-Making





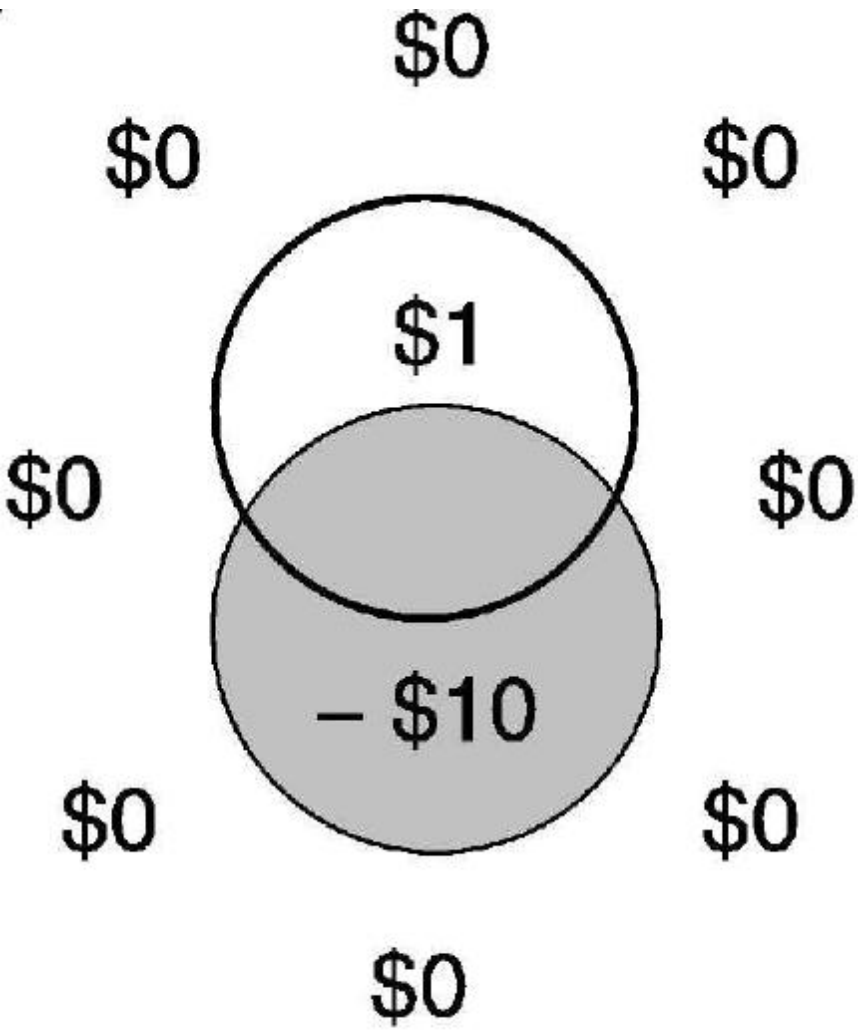
Decision Rule: Pick lottery with higher expected value

Motor Control and Maximum Expected Gain

Implicit probabilities through motor variability

Trommershauser et al. 2003

Motor Control and Maximum Expected Gain

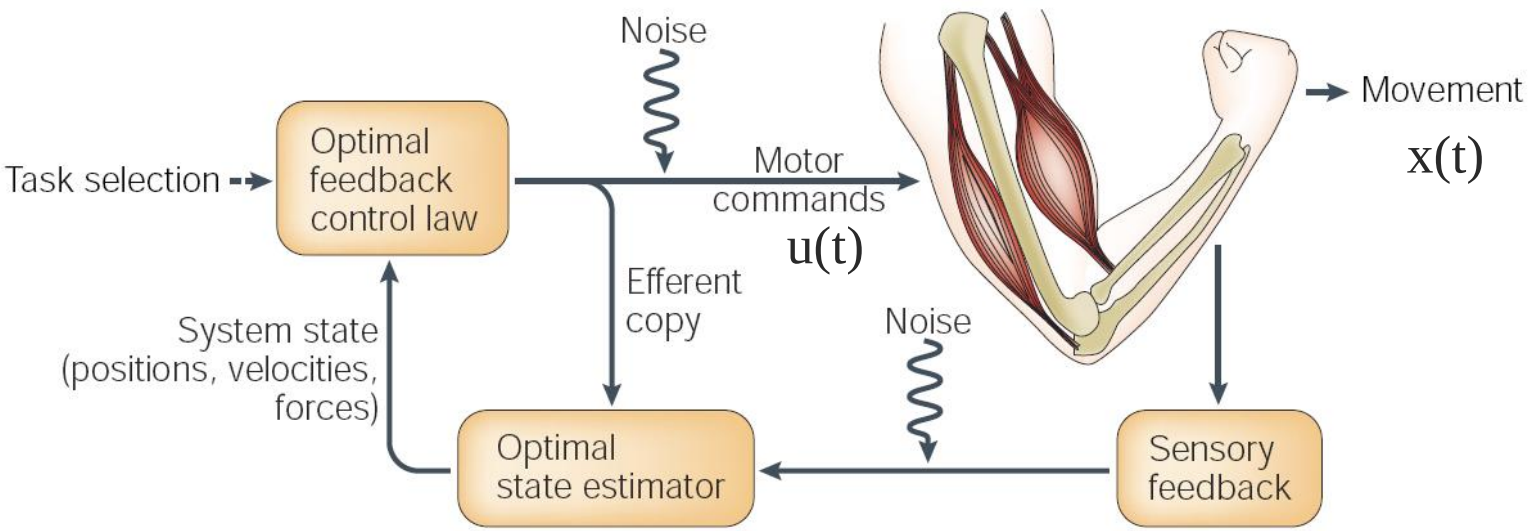


$$L(S) = \sum_{i=0}^{N} C_i P(R_i|S)$$

"explicit costs"

Trommershauser et al. 2003

Optimal Feedback Control



Dynamic System $$dx = f(x,u)dt + dw$$

x(t): state
u(t): control

Cost Function $$J = E\left[\int c(x,u)dt + \phi(x_T)\right]$$

Todorov & Jordan, *Nat. Neurosci.* (2002)

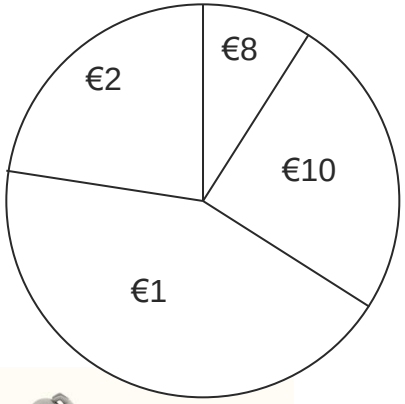
Variational Principle

$$-\Delta F = \sum_{x \in \mathcal{X}} \mathbf{P}_f(x)\mathbf{U}_*(x) - \alpha \sum_{x \in \mathcal{X}} \mathbf{P}_f(x) \log \frac{\mathbf{P}_f(x)}{\mathbf{P}_i(x)}$$

Negative free energy is maximized in equilibrium

- Estimation: Maximum Entropy principle given constraints on mean utility
- Control: Maximum Utility principle given constraints on relative entropy

Certainty-equivalent



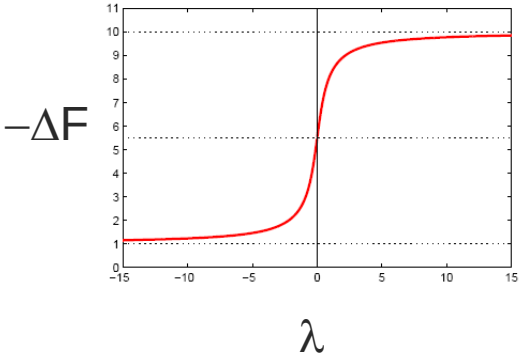
Value of the Lottery

$$-\Delta F = \frac{1}{\lambda}\log\left(\sum_j p_j^0 \exp\left(\lambda U_j\right)\right)$$

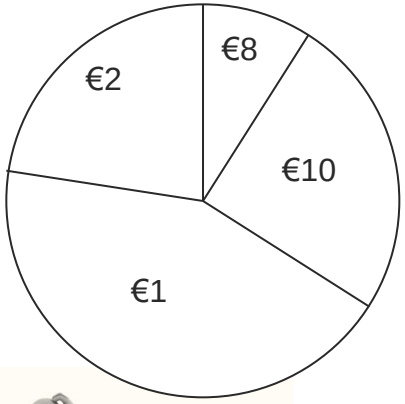
$\lambda \to \infty$ $\quad -\Delta F = \max_j U_j$

$\lambda \to -\infty$ $\quad -\Delta F = \min_j U_j$

$\lambda \to 0$ $\quad -\Delta F = \sum_j p_j^0 U_j$



Equilibrium distribution

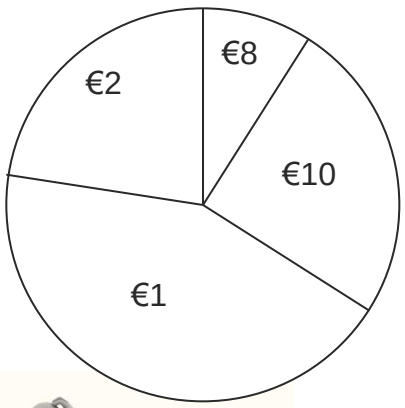


$$p_i = \frac{p_i^0 \exp\left(\frac{1}{\alpha} U_i\right)}{\sum_j p_j^0 \exp\left(\frac{1}{\alpha} U_j\right)}$$

Probabilities of the Lottery

Utilities of the Lottery

Equilibrium distribution



$$p_i = \frac{p_i^0 \exp\left(\frac{1}{\alpha} U_i\right)}{\sum_j p_j^0 \exp\left(\frac{1}{\alpha} U_j\right)}$$

Action Lotteries

- p0 is default policy
- α measures bounded rationality

Observation Lotteries

- p0 is default model
- α anticipates rationality of environment (model uncertainty, ambiguity)

Risk-sensitivity and model uncertainty

$$
\begin{aligned}
f &= max_p \sum_x p(x)U(x) - \frac{\theta}{2}\sum_x p(x)\log\frac{p(x)}{p_0(x)} \\
&= \frac{2}{\theta}\log\mathbb{E}[e^{\frac{1}{2}\theta U}] \\
&\approx \mathbb{E}[U] - \theta\mathbb{VAR}[U]
\end{aligned}
$$

Bias towards best-case outcome: $\theta < 0$

Bias towards worst-case outcome: $\theta > 0$

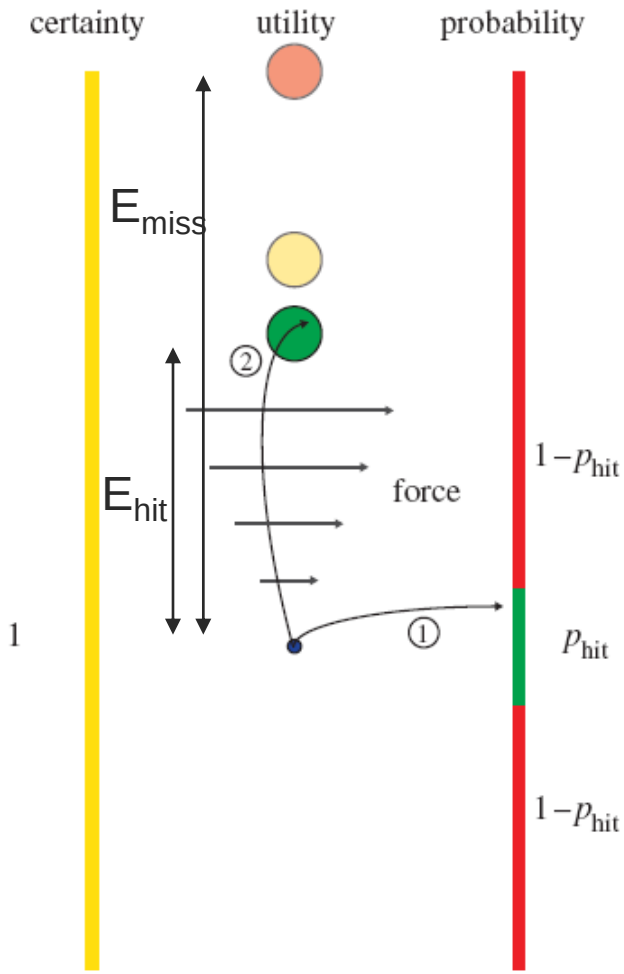
Experimental Studies

- Study 1: Mean-variance trade-off
- Study 2: Biasing of control gains
- Study 3: Biasing of Bayesian learning

Study 1

- The mean-variance trade-off

Experimental Setup



$$\mu = p_{\text{hit}} \cdot E_{\text{hit}} + p_{\text{miss}} \cdot E_{\text{miss}}$$

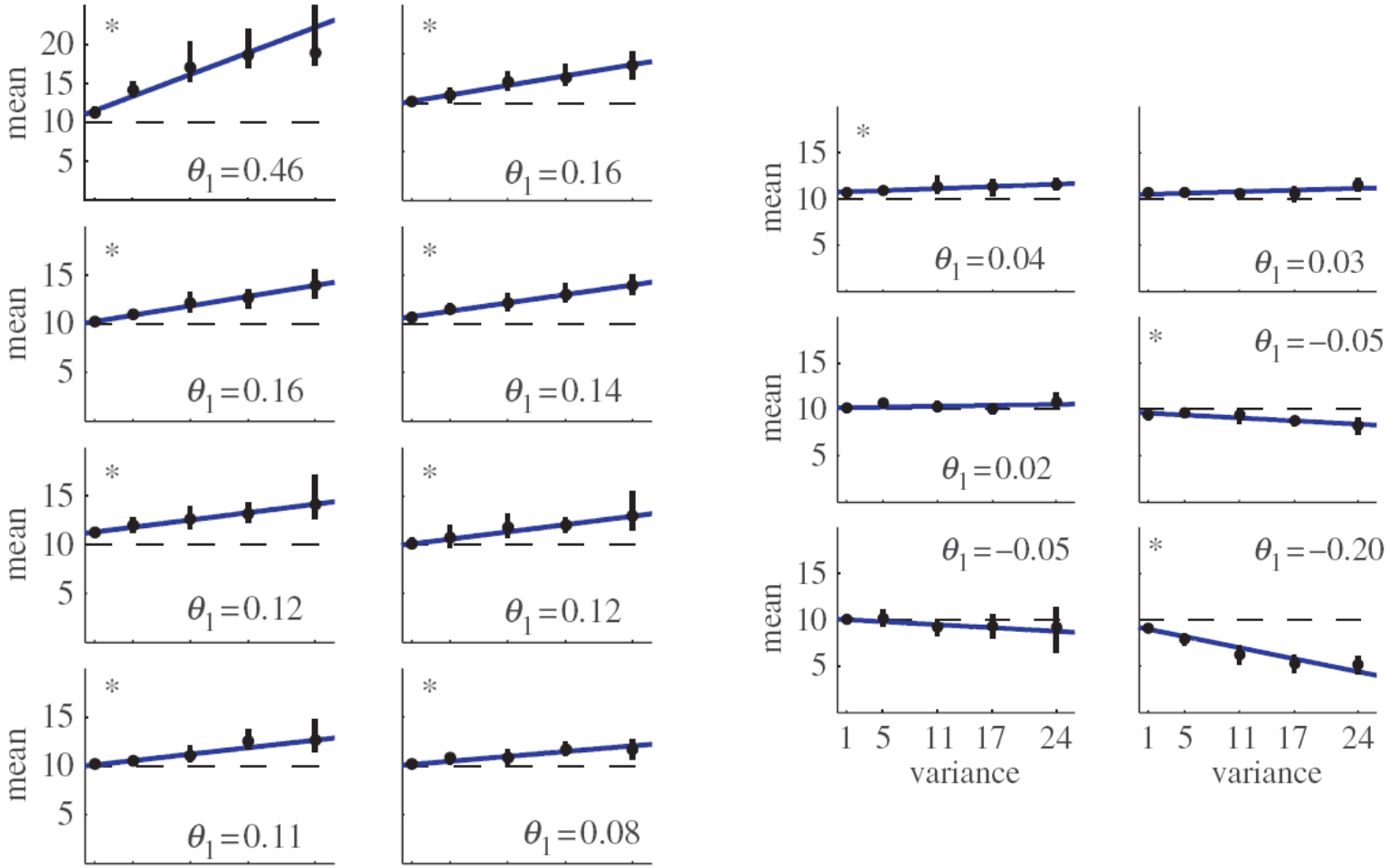
$$\sigma^2 = (E_{\text{hit}} - \mu)^2 \cdot p_{\text{hit}} + (E_{\text{miss}} - \mu)^2 \cdot p_{\text{miss}}$$

Model fit

- Sure bet $U_1^{\mathrm{s}} = -\mathrm{E}(10) = -10$
- Risky bet $U_1^{\mathrm{r}}(x) = -\mathbb{E}(x) + \theta_1 Var(x)$
- Curve of indifference points

$$\mathbb{E}(x) = \theta_1 Var(x) + 10$$

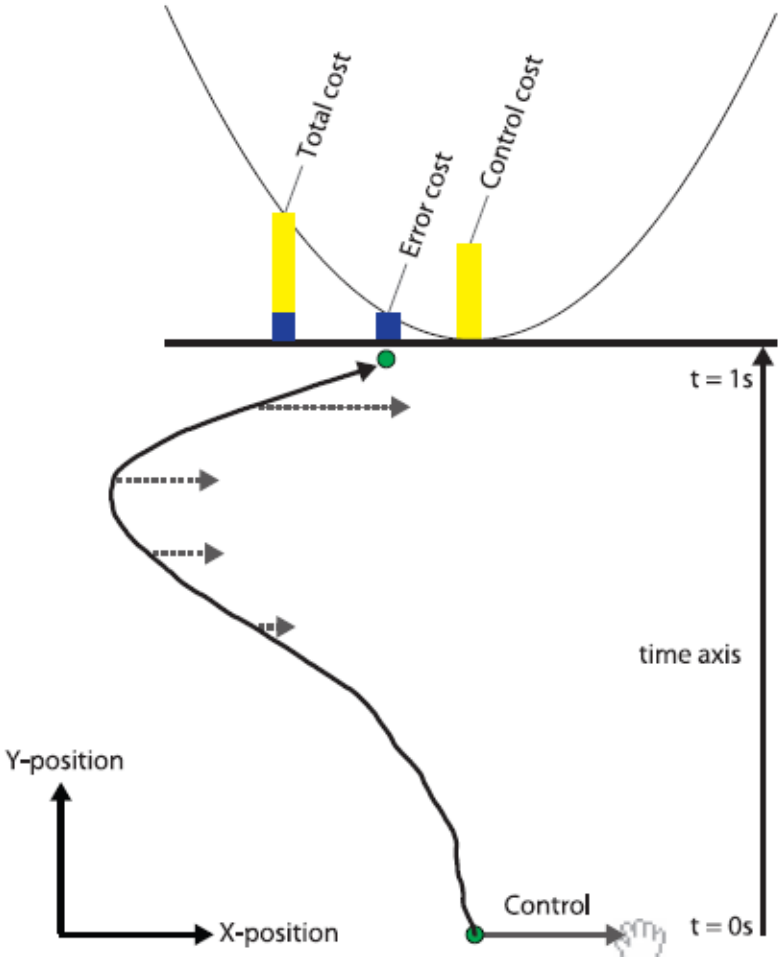
Results


mean
20
15
10
5
*
$\theta_1 = 0.46$
$\theta_1 = 0.16$
$\theta_1 = 0.16$
$\theta_1 = 0.14$
$\theta_1 = 0.12$
$\theta_1 = 0.12$
$\theta_1 = 0.11$
$\theta_1 = 0.08$
$\theta_1 = 0.04$
$\theta_1 = 0.03$
$\theta_1 = 0.02$
$\theta_1 = -0.05$
$\theta_1 = -0.05$
$\theta_1 = -0.20$
1 5 11 17 24
variance

Study 2

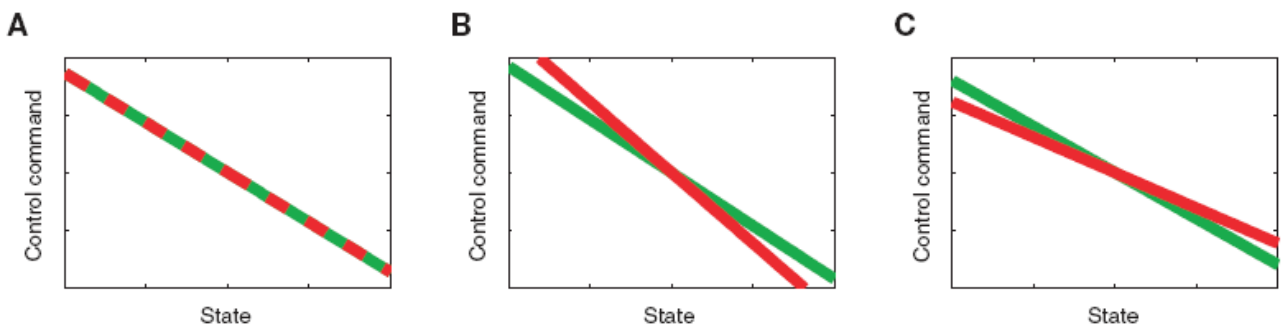
- Biasing of control gains

Experimental Setup

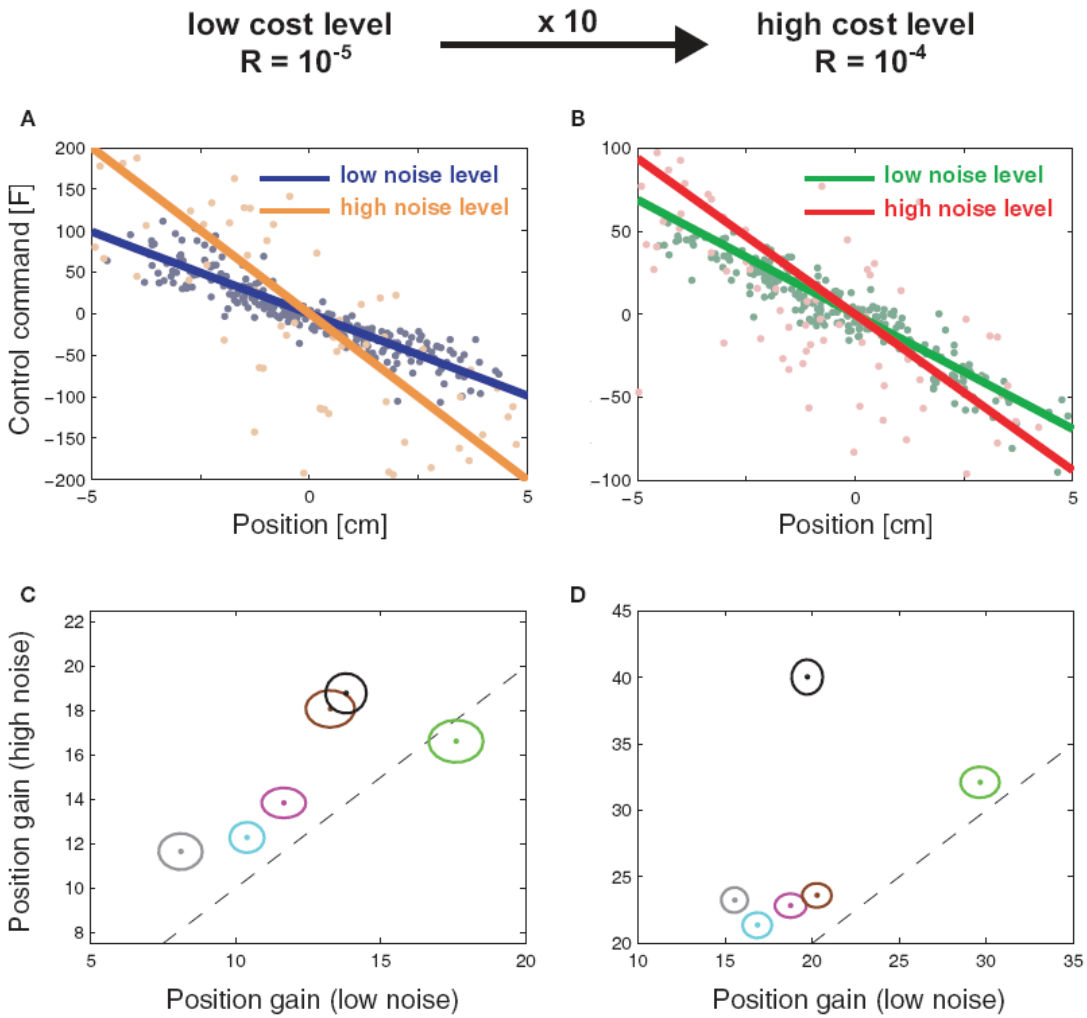


Model Prediction

		Risk-neutral	Risk-averse $\theta < 0$	Risk-seeking $\theta > 0$
Optimality criterion		$\min(\mathbb{E}(\mathbb{C}))$	$\gamma(\theta) = -2\theta^{-1} \ln \mathbb{E}[e^{-\frac{1}{2}\theta \mathbb{C}_{total}}]$	
Motor command	Low noise level	$\mathrm{u}_t = K_t \mathrm{x}_t$	$\mathrm{u}_t = K_t(N_{low}, \theta)\mathrm{x}_t$	
Motor command	High noise level	$\mathrm{u}_t = K_t \mathrm{x}_t$	$\mathrm{u}_t = K_t(N_{high}, \theta)\mathrm{x}_t$	



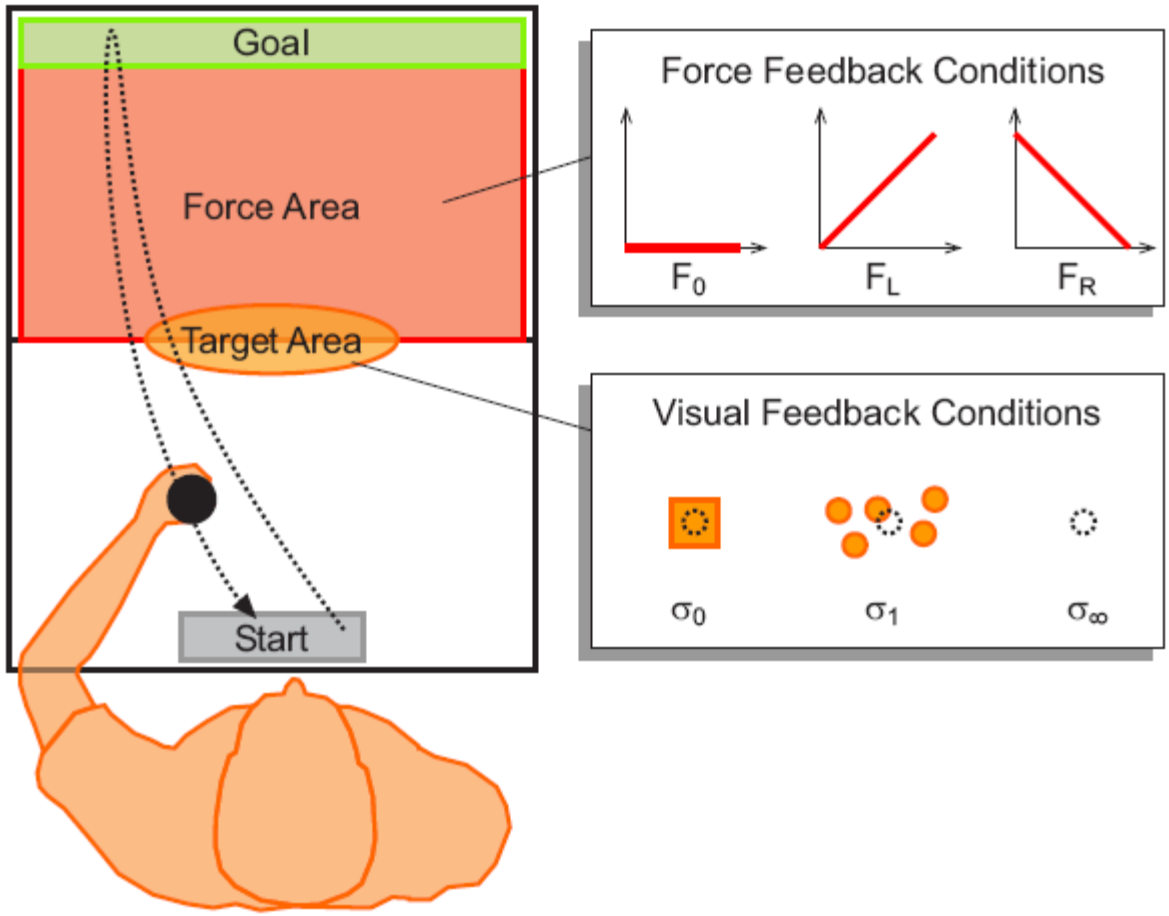
Results


low cost level
R = 10^-5
x 10
high cost level
R = 10^-4
A
B
C
D
low noise level
high noise level
low noise level
high noise level
Control command [F]
Position [cm]
Position [cm]
Position gain (high noise)
Position gain (low noise)
Position gain (low noise)

Study 3

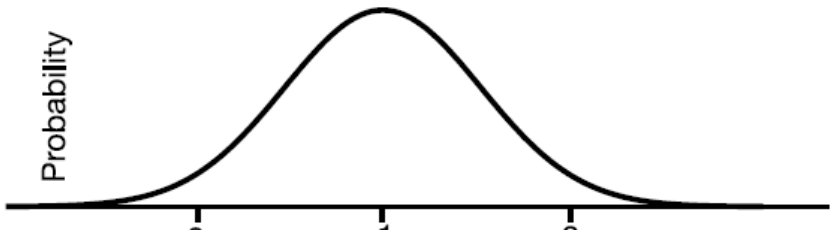
- Biasing of sensorimotor estimation

Experimental Setup

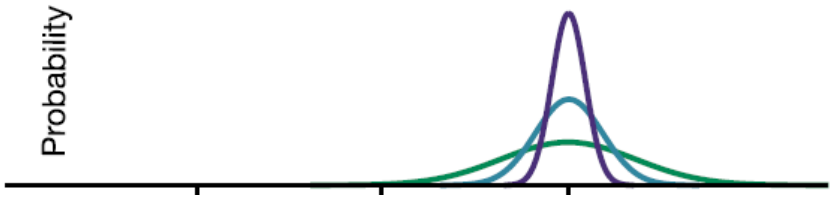


Bayesian Sensorimotor Integration

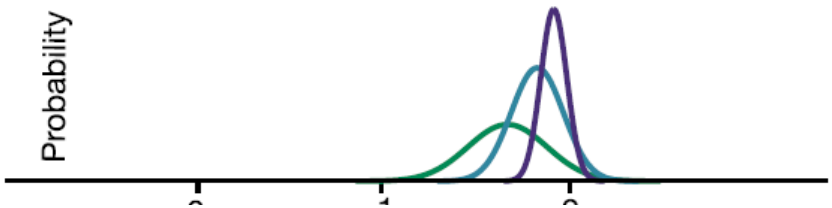
P(x)

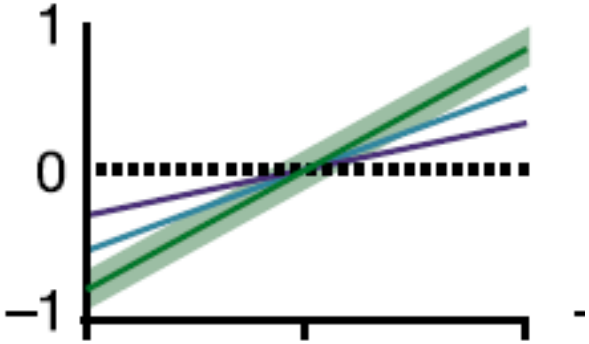


P(y|x)



P(x|y)





Körding & Wolpert, Nature, 2004

Model Prediction

Risk-neutral estimator

$$
\begin{aligned}
u^{opt} &= \arg\min_u \int_{-\infty}^{+\infty} dx\, p(x|y)\left[Q(x-u)^2 + c(u)\right] \\
&= \frac{\sigma_p^2}{\sigma_p^2 + \sigma_i^2}\, y - \frac{a_j}{2Q}.
\end{aligned}
$$

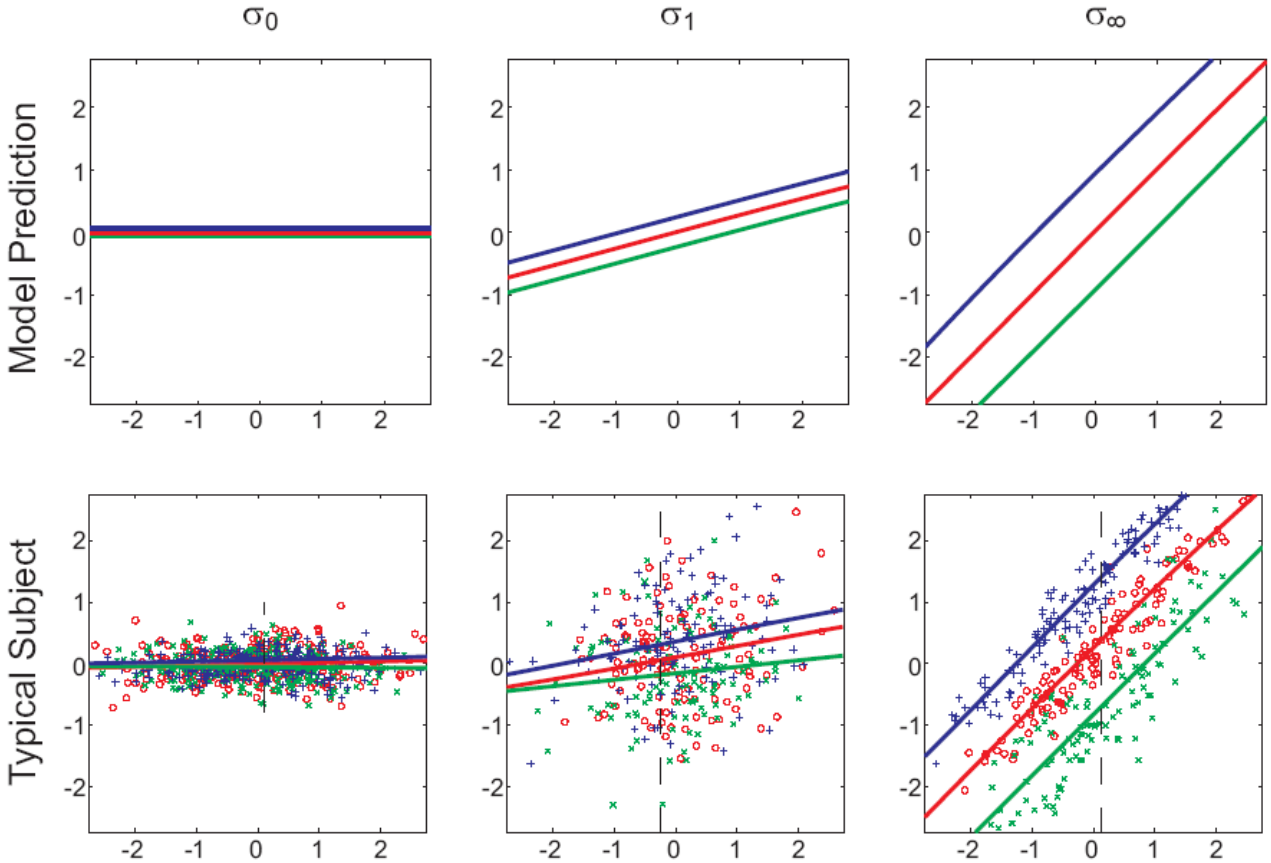
Risk-sensitive estimator

$$
\begin{aligned}
u^{opt} &= \arg\min_u -\frac{2}{\theta} \int_{-\infty}^{+\infty} dx\, p(x|y) e^{-\frac{\theta}{2}\left[Q(x-u)^2 + c(u)\right]} \\
&= \frac{\sigma_p^2}{\sigma_i^2 + \sigma_p^2}\, y - \frac{a_j}{2Q} - \frac{\sigma_i^2 \sigma_p^2}{\sigma_i^2 + \sigma_p^2}\, \theta\, a_j.
\end{aligned}
$$

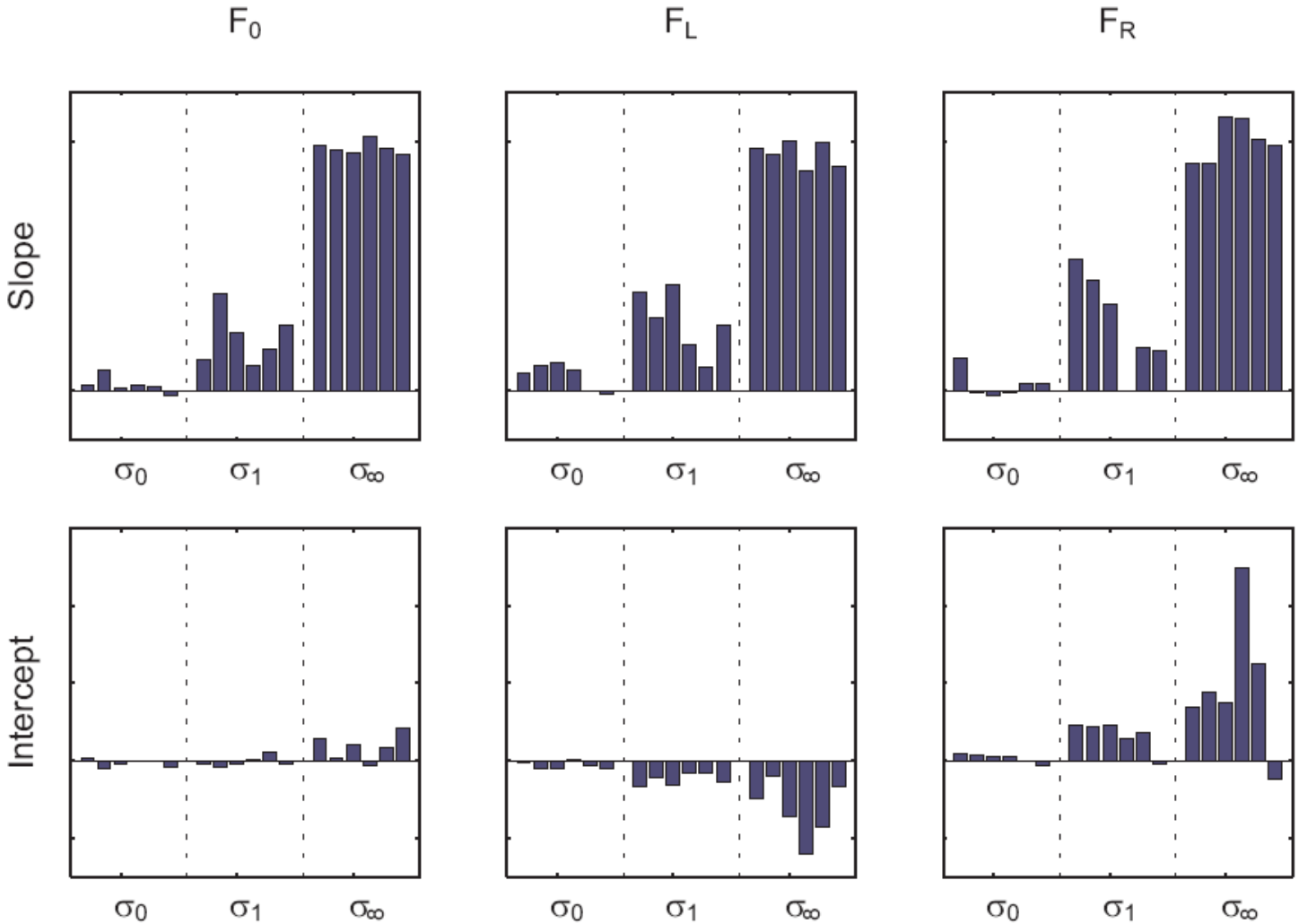
with cost function $c(h) = a_j\, h + b_j$

Results

$$\frac{\sigma_p^2}{\sigma_i^2 + \sigma_p^2}\,\mathbf{y} - \frac{\sigma_i^2 \sigma_p^2}{\sigma_i^2 + \sigma_p^2}\,\alpha\, a_j$$



Results



Conclusion

- Humans show deviations from risk-neutral behavior in motor control
- Risk-sensitivity implies a mean-variance trade-off
- Risk-sensitivity implies changes in control gains for different levels of uncertainty
- Sensorimotor learning can be described by risk-sensitive Bayesian models

Acknowledgments

Arne Nagengast

Jordi Grau Moya

Daniel Wolpert

Frontiers in Human Neuroscience (2011) 5:1-10
Proceedings of the Royal Society B (2011) 278(1716):2325-32.
PLoS Computational Biology (2010) 6(7): e1000857
PLoS Computational Biology (2012) in press