Information-optimal Coupling of Perception and Action through Lossy Compression

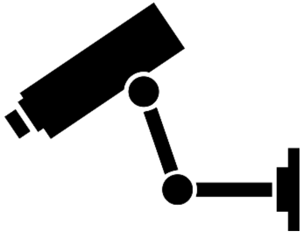
Perception → Action

Tim Genewein
Bosch Center for AI
1st of June 2017

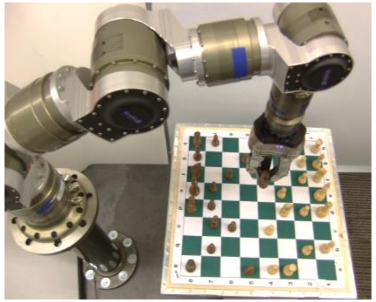
BOSCH

What is perception?

$$w \longrightarrow x \longrightarrow a$$



?



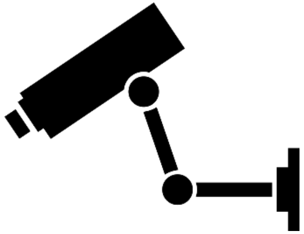
Sensory input (world-state)

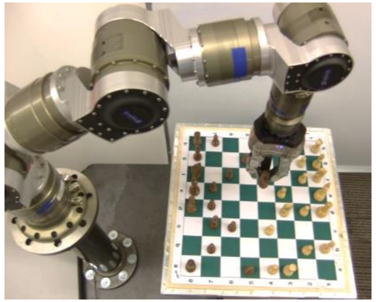
- Relevant information (entangled, latent)
- Irrelevant information (noise)

Action a

- Respond to world-state w
- Using internal representation x

$$w \xrightarrow{\text{Perception}} x \xrightarrow{\text{Action}} a$$





Sensory input (world-state)

- Relevant information (entangled, latent)
- Irrelevant information (noise)

Action a

- Respond to world-state w
- Using internal representation x

Perception = abstraction of world-state

- Extraction of relevant information for acting
- Formation of (disentangled) internal representation

$$w \xrightarrow{\text{Perception}} x \xrightarrow{\text{Action}} a$$

Camera

Semantic scene segmentation

Emergency breaking

Perception = abstraction of world-state

$$w \xrightarrow{\text{Perception}} x \xrightarrow{\text{Action}} a$$

Camera

Semantic scene segmentation

Emergency breaking

Perception = abstraction of world-state

Task-independent abstractions?

- Suboptimal
- Inefficient

Lane keeping?

Road sign recognition?

Headlight control?

…

$$w \xrightarrow{\text{Perception}} x \xrightarrow{\text{Action}} a$$

Camera

Perception and action should be tightly coupled!

Representation wish list:

- Capture (only) relevant information
- Robust to irrelevant variation (noise)
- Disentanglement (low redundancy)

- Suboptimal
- Inefficient

Emergency breaking

Lane keeping?

Road sign recognition?

Headlight control?

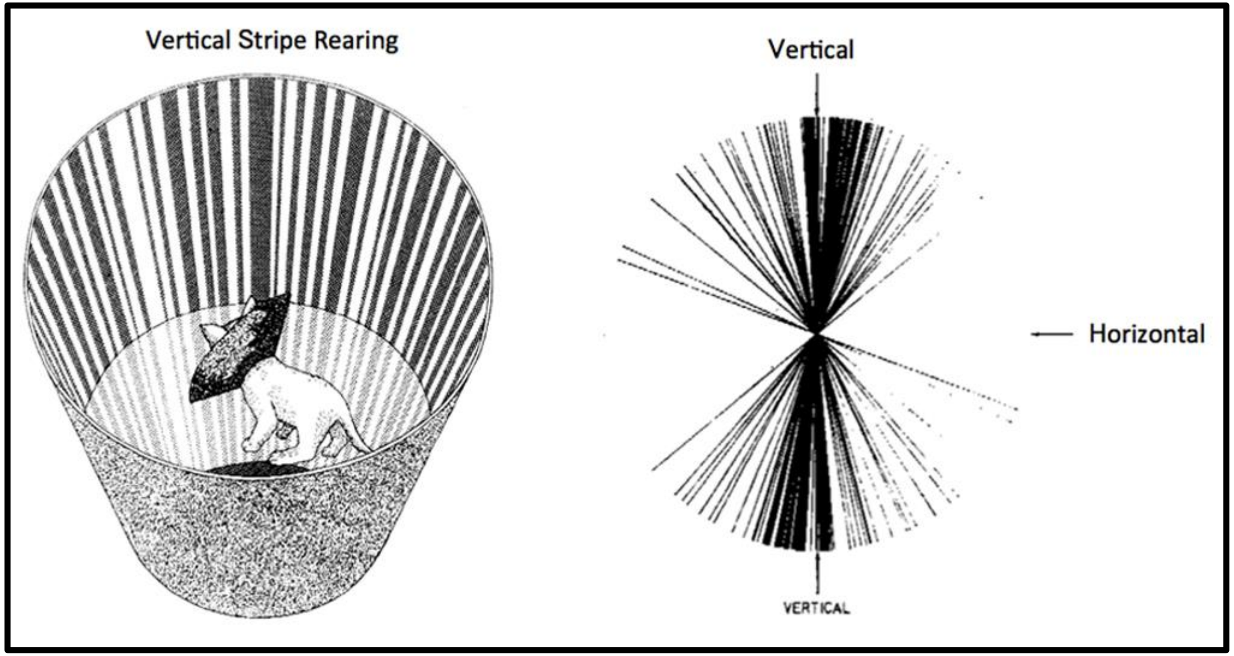
...

Perception acts as a "smart filter", leading to robustness and efficient information processing

*"Perception is **not** the **passive receipt** of* [sensory] *signals* [...] *Perception* [...] *is the organization, identification, and interpretation of sensory information in order to represent and understand the environment."* (Schacter 2011, Psychology. Worth Publishers)

"[...] *Evolutionary psychologists hold that the **primary purpose*** [of perception] ***is to guide action**. For example, depth perception seems to have evolved not to help us know the distances to other objects but rather to help us move around in space."* (Gaulin, McBurney 2003, Evolutionary Psychology. Prentice Hall)

Blakemore, Cooper 1970



Held, Hein 1963

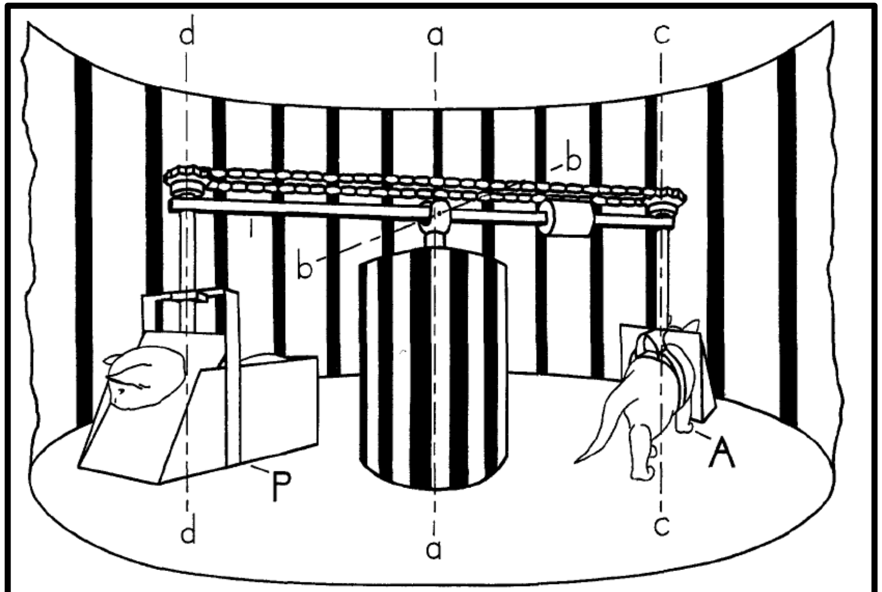
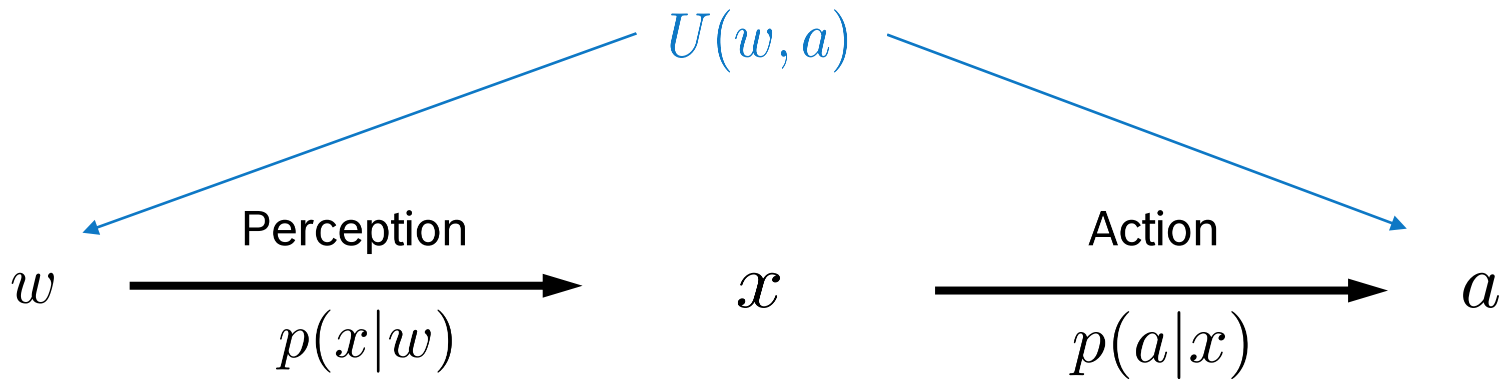


Fig. 1. Apparatus for equating motion and consequent visual feedback for an actively moving (A) and a passively moved (P) S.

Formalizing perception

Coupling sensory processing to action



Perception should enable (Bayesian) inference

$$p(w|x) = \frac{p(x|w)p(w)}{p(x)}$$

Action stage: world-state is not directly accessible

Maximize expected utility under **posterior belief**

$$U(x,a) = \sum_w p(w|x)U(w,a)$$

$$\begin{tikzcd}[column sep=huge]
 & U(w,a) \arrow[dl] \arrow[dr] & \\
w \arrow[r, "\text{Perception}", "p(x|w)"'] & x \arrow[r, "\text{Action}", "p(a|x)"'] & a
\end{tikzcd}$$

Perception should enable (Bayesian) inference

$$p(w|x) = \frac{\boxed{p(x|w)}p(w)}{p(x)}$$

Yes, but likelihood model $p(x|w)$ is undefined

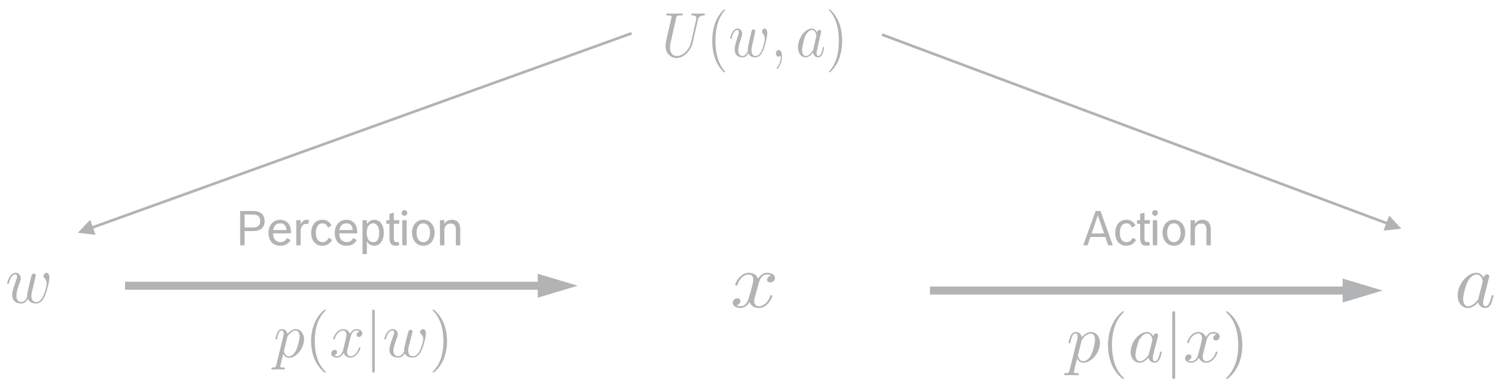
Action stage: world-state is not directly accessible

Maximize expected utility under **posterior belief**

$$U(x,a) = \sum_w p(w|x)U(w,a)$$

“x should represent w as faithfully as possible”

*“ perception should extract most **relevant information** for acting”*

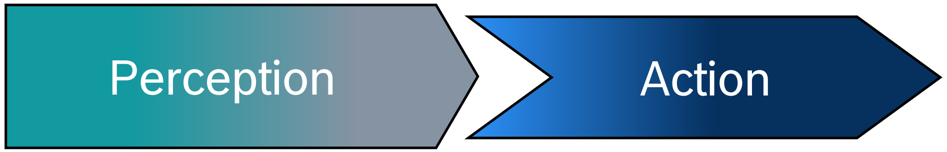


Perception should enable (Bayesian) inference

$$p(w|x) = \frac{\boxed{p(x|w)}\,p(w)}{p(x)}$$

Yes, but likelihood model $p(x|w)$ is undefined

"x should represent w as faithfully as possible"

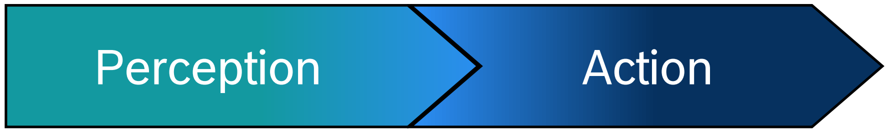


Action stage: world-state is not directly accessible

Maximize expected utility under **posterior belief**

$$U(x,a) = \sum_w p(w|x)U(w,a)$$

*" perception should extract most **relevant information** for acting"*



$U(w, a)$

Pe ple

Ye ef

An optimality principle for coupling perception and action

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A)$$

"x should represent w as faithfully as possible"

Perception → Action

" perception should extract most ***relevant information*** *for acting"*

Perception → Action

Bounded rational decision-making

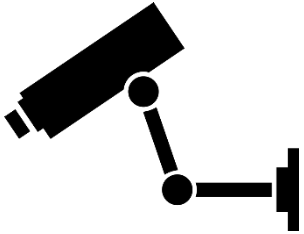
Decision-making

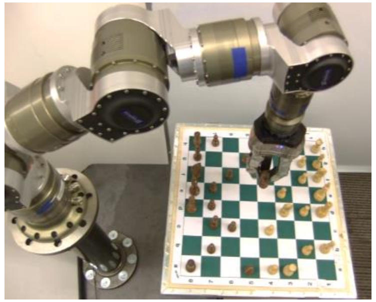
Decision-making:

Given some input (world-state), pick best action

Utility function quantifies desirability of action

$$w \xrightarrow{a_w^* = \arg\max_a U(w,a)} a$$





Decision-making

Decision-making:

Given some input (world-state), pick best action

Utility function quantifies desirability of action

$$w \xrightarrow{a^*_w = \arg\max_a U(w,a)} a$$

Problem:

Searching through vast set with limited computational capacity

Information-theoretic bounded rationality

Ortega et al. 2015: Information-Theoretic Bounded Rationality

Modify the optimality principle

Take the process of computation into account

Satisficing: find solution that are “good enough”

$$p^*(a|w) = \arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

\+ Expected utility

Inverse temperature

\- Computational effort

Information-theoretic bounded rationality

Modify the optimality principle

Take the process of computation into account

Satisficing: find solution that are “good enough”

Free energy optimization

$$p^*(a|w) = \arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w) || p_0(a)\right)$$

Closed-form solution

$$p^*(a|w) = \frac{1}{Z} p_0(a) e^{\beta U(w,a)}$$

Information-theoretic bounded rationality

Modify the optimality principle

Take the process of computation into account

Satisficing: find solution that are “good enough”

Free energy optimization

$$p^*(a|w) = \arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

Closed-form solution

$$p^*(a|w) = \frac{1}{Z} p_0(a) e^{\beta U(w,a)}$$

Special case: Bayes’ rule

$$U(w,a) = \log p(w|a)$$
$$\beta = 1$$

$$p^*(a|w) = \frac{p(w|a)p_0(a)}{Z}$$

From free energies to rate-distortion: the optimal prior

Genewein et al. 2015: Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle.

Free energy optimization

$$\arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w) || p_0(a)\right)$$

Free energy optimization

$$\arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

What is the optimal prior (on average)?

$$\arg\max_{p(a|w),p_0(a)} \sum_{w,a} p(w)p(a|w)[U(w,a)] - \frac{1}{\beta} \sum_w p(w) D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

Optimal prior is marginal $p_0^* = \sum_w p(w)p(a|w) = p(a)$

Free energy optimization

$$\arg\max_{p(a|w)} \mathbf{E}_{p(a|w)}[U(w,a)] - \frac{1}{\beta} D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

What is the optimal prior (on average)?

$$\arg\max_{p(a|w),p_0(a)} \sum_{w,a} p(w)p(a|w)[U(w,a)] - \frac{1}{\beta}\sum_w p(w) D_{\mathrm{KL}}\left(p(a|w)||p_0(a)\right)$$

Optimal prior is marginal $p_0^* = \sum_w p(w)p(a|w) = p(a)$

$$\boxed{\arg\max_{p(a|w)} \mathbf{E}_{p(w,a)}[U(w,a)] - \frac{1}{\beta} I(W;A)}$$

Lossy compression: rate-distortion theory

$$p^*(a|w) = \arg\max_{p(a|w)} \mathbf{E}_{p(w,a)}[U(w,a)] - \frac{1}{\beta} I(W;A)$$

Lossy compression: transmit information via channel of limited capacity

- Keep most relevant information
- Discard irrelevant information (noise)

Self-consistent solution

$$p^*(a|w) = \frac{1}{Z} p(a) e^{\beta U(w,a)}$$

$$p(a) = \sum_w p(w) p^*(a|w)$$

Lossy compression: rate-distortion theory

$$p^*(a|w) = \arg\max_{p(a|w)} \mathbf{E}_{p(w,a)}[U(w,a)] - \frac{1}{\beta} I(W;A)$$

Lossy compression: transmit information via channel of limited capacity

- Keep most relevant information
- Discard irrelevant information (noise)

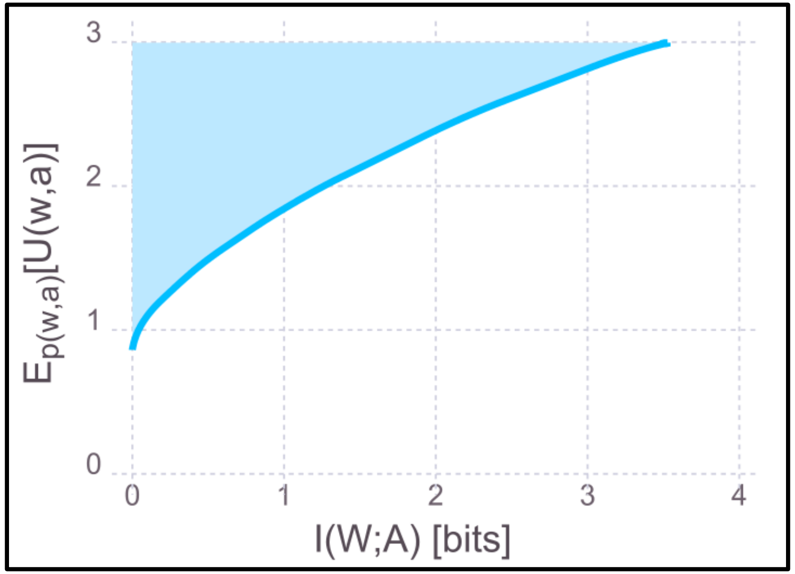
Mutual information $I(W;A) = H(A) - H(A|W)$

Reduction in uncertainty over A, given (the knowledge of) W

$$0 \xrightarrow{\hspace{6em}} H(A)$$

indep. RV $\quad I(W;A) \quad$ deterministic relation

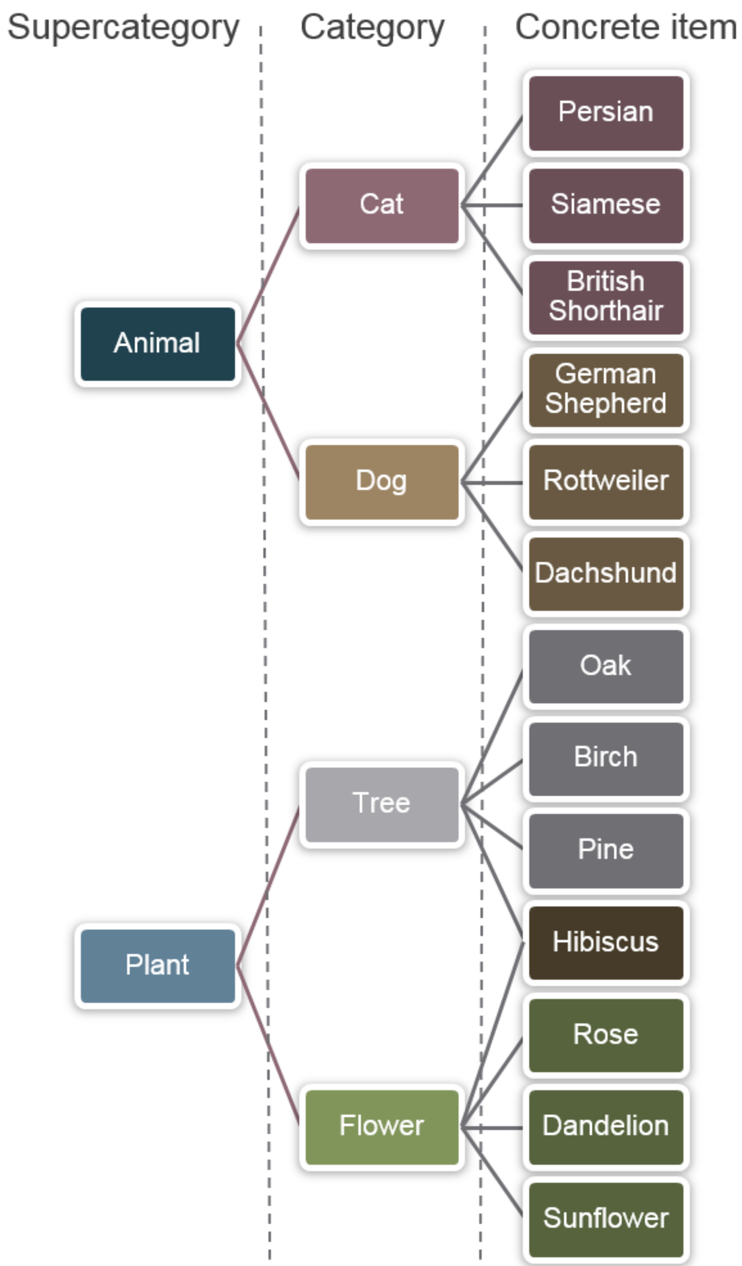
Rate-utility curve



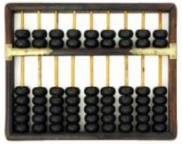
Simple taxonomy with **three layers of abstraction**

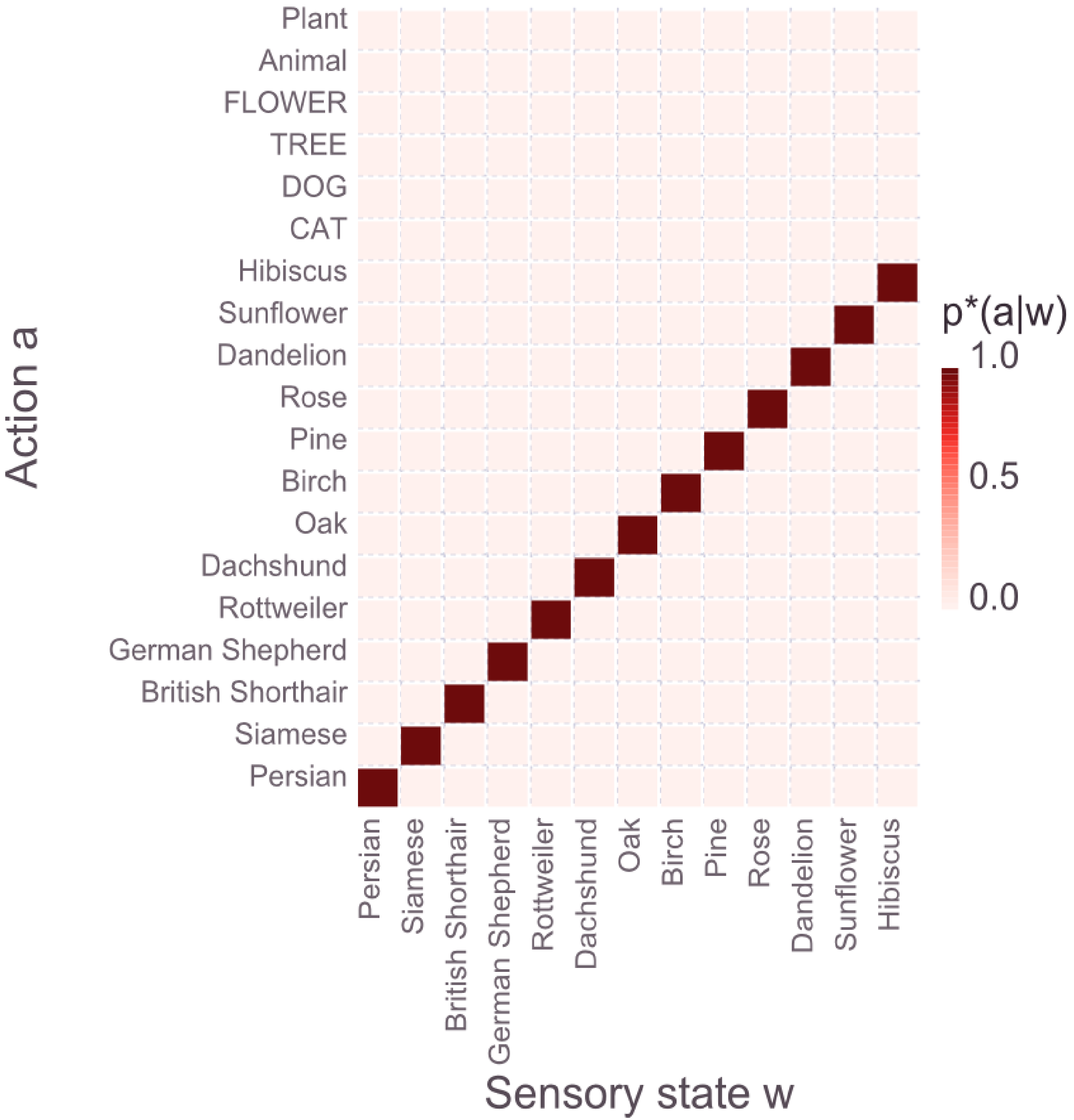
Sensory state $w \in \{concrete\ items\}$
Action $a \in \{concrete\ items, categories, supercategories\}$

Rewards/Utilities:
- 3€ if concrete item correct
- 2.2€ if category correct
- 1.6€ if supercategory correct

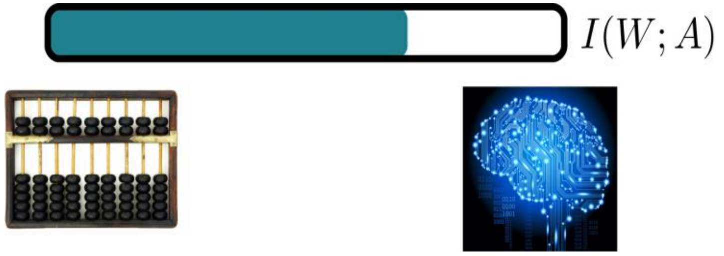


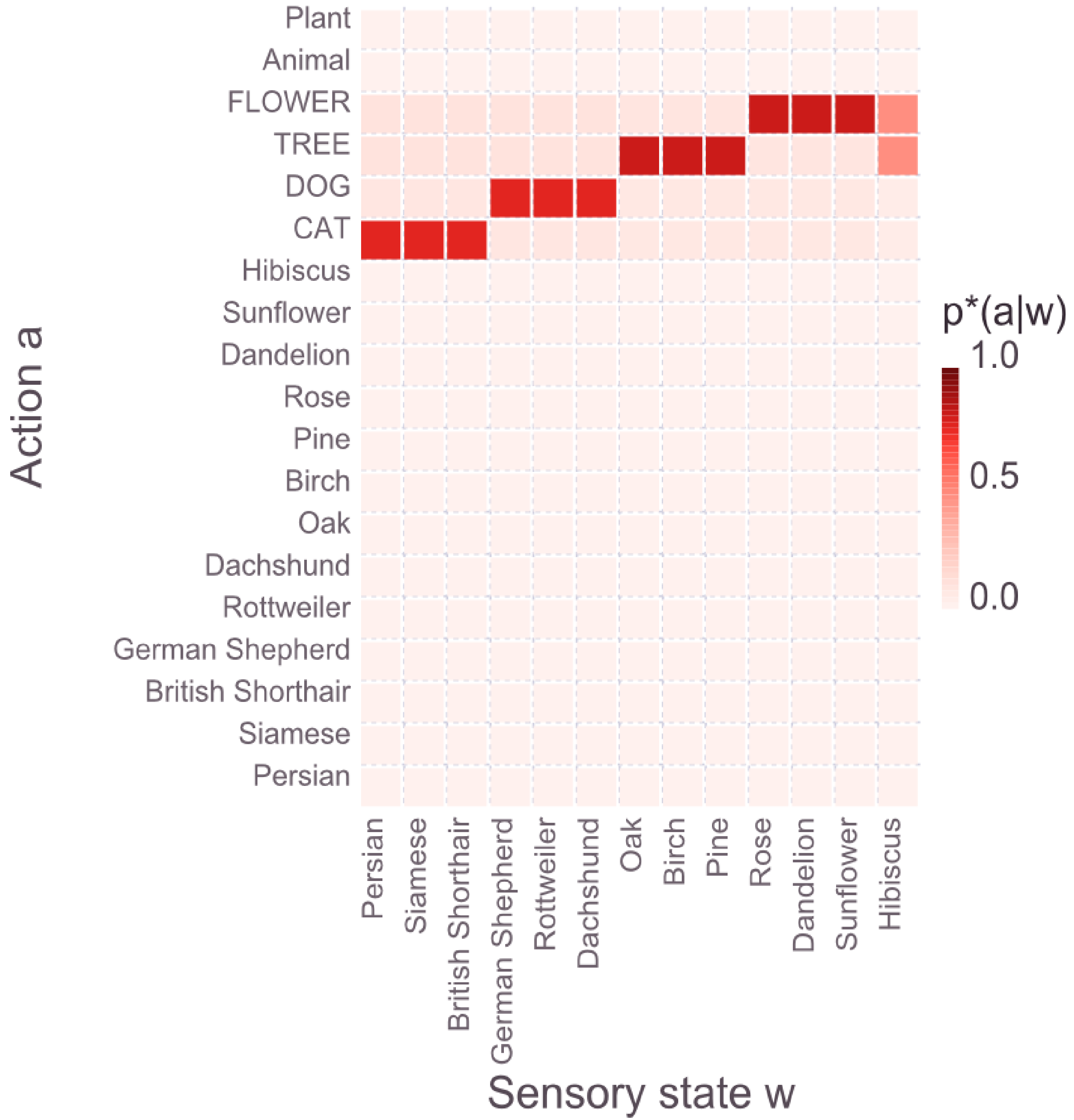
β	10	$\left[\text{bits}/_{€}\right]$
I	3.7	[bits]
$\mathbf{E}[U]$	3	[€]



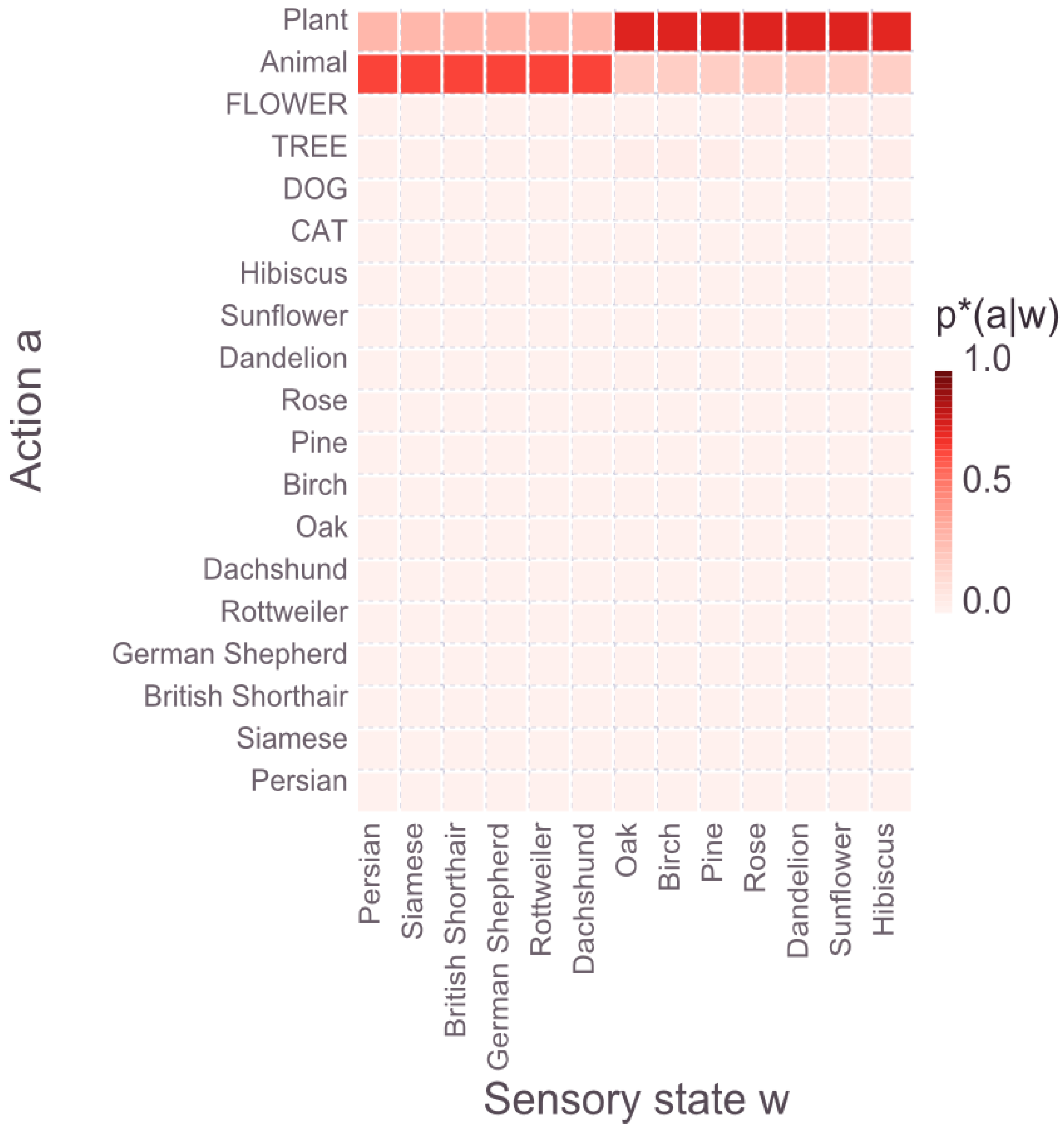


β	1.11	$\left[{}^{\text{bits}}/_{€}\right]$
I	0.9	[bits]
$\mathbf{E}[U]$	1.8	[€]



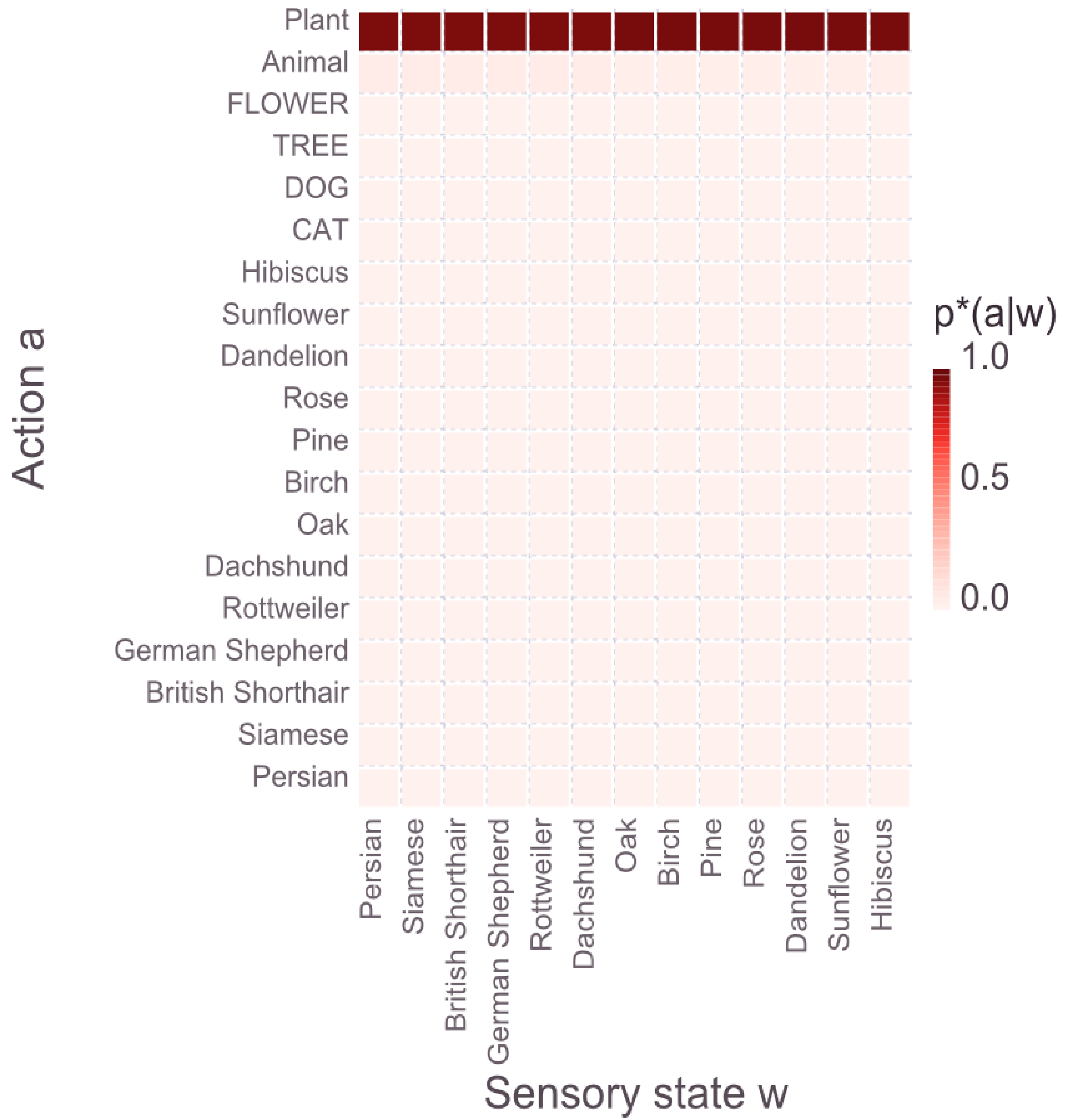


β	0.67	$\left[{}^{\text{bits}}/_{€}\right]$
I	0.18	[bits]
$\mathbf{E}[U]$	1.2	[€]

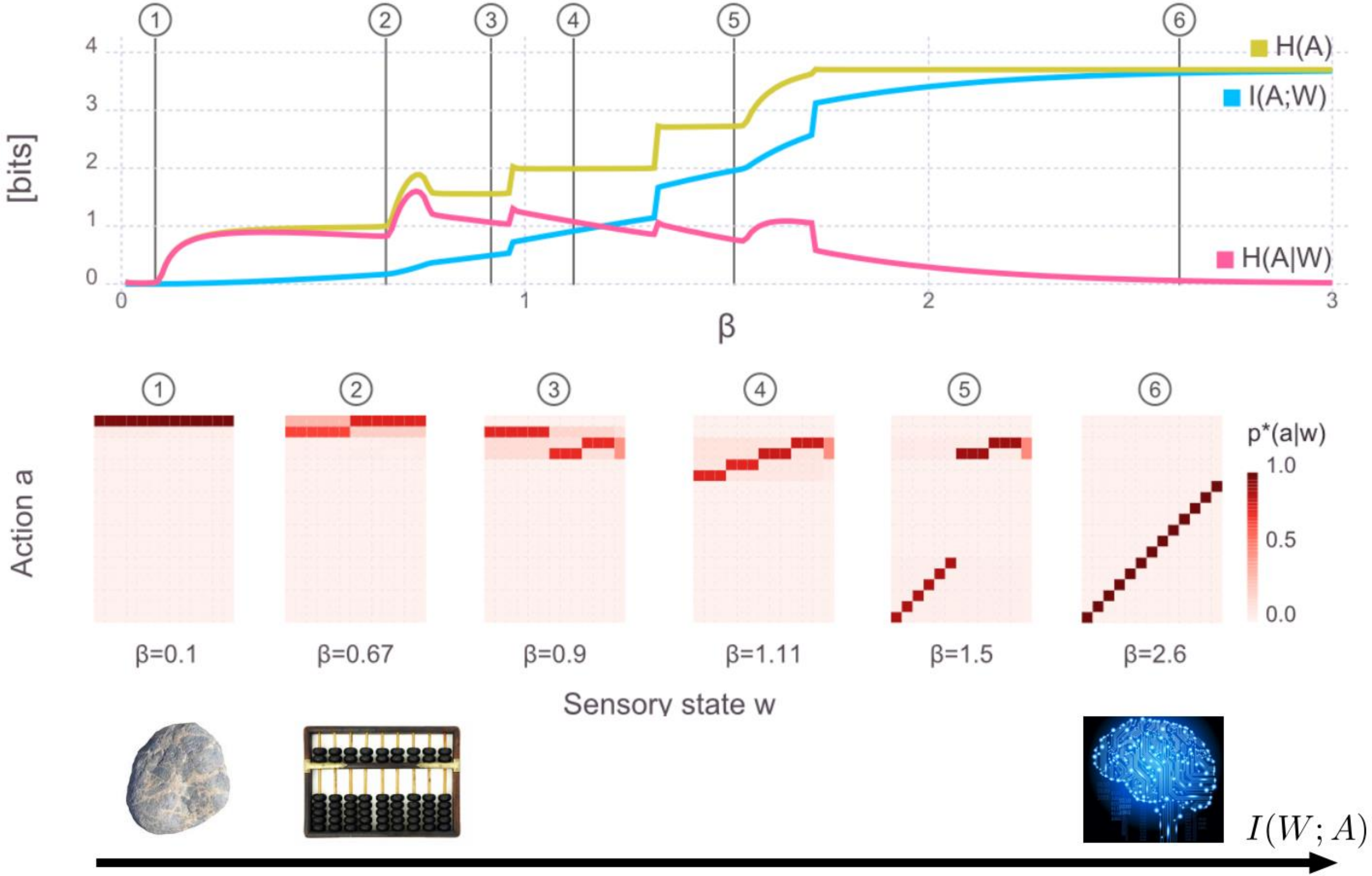



$I(W;A)$

β	0.1	$[\text{bits}/\text{€}]$
I	0	[bits]
$\mathbf{E}[U]$	0.86	[€]



$I(W;A)$

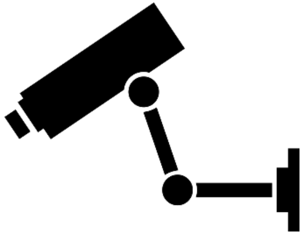


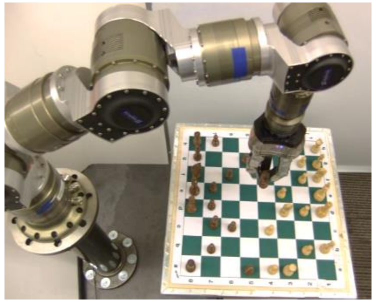
1
2
3
4
5
6
H(A)
I(A;W)
H(A|W)
[bits]
4
3
2
1
0
0
1
2
3
β
p*(a|w)
1.0
0.5
0.0
Action a
β=0.1
β=0.67
β=0.9
β=1.11
β=1.5
β=2.6
Sensory state w
I(W; A)

Two-stage system: two channels

Genewein et al. 2015: Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle.

$$w \xrightarrow[p(x|w)]{\text{Perception}} x \xrightarrow[p(a|x)]{\text{Action}} a$$





Perception should enable (Bayesian) inference

$$p(w|x) = \frac{p(x|w)p(w)}{p(x)}$$

“x should represent w as faithfully as possible”

Perception → Action

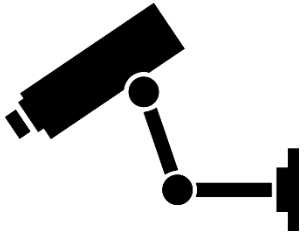
Action stage: maximize posterior expected utility

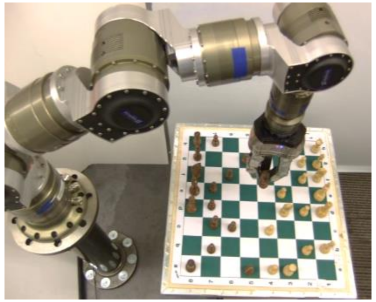
$$U(x,a) = \sum_w p(w|x)U(w,a)$$

“perception should extract most ***relevant information*** *for acting”*

Perception → Action

$$w \xrightarrow[p(x|w)]{\text{Perception}} x \xrightarrow[p(a|x)]{\text{Action}} a$$





Information-theoretic bounded rationality:

Trade off gains in expected utility against computational effort

Two channels with limited capacity: perception and action

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A)$$

An optimality principle for coupling perception and action

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A)$$

Self-consistent solution:

$$p^*(x|w) = \frac{1}{Z(w)} p(x) e^{\beta_1 \Delta F(w,x)}$$

Optimal perceptual likelihood

$$p^*(a|x) = \frac{1}{Z(x)} p(a) e^{\beta_2 U(x,a)}$$

Bounded rational maximization of posterior utility

$$U(x,a) = \sum_w p(w|x) U(w,a)$$

$$p(x) = \sum_w p(w) p^*(x|w)$$

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|x)$$

Priors are marginals

An optimality principle for coupling perception and action

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A)$$

Self-consistent solution:

$$p^*(x|w) = \frac{1}{Z(w)} p(x) e^{\beta_1 \Delta F(w,x)}$$

Optimal perceptual likelihood

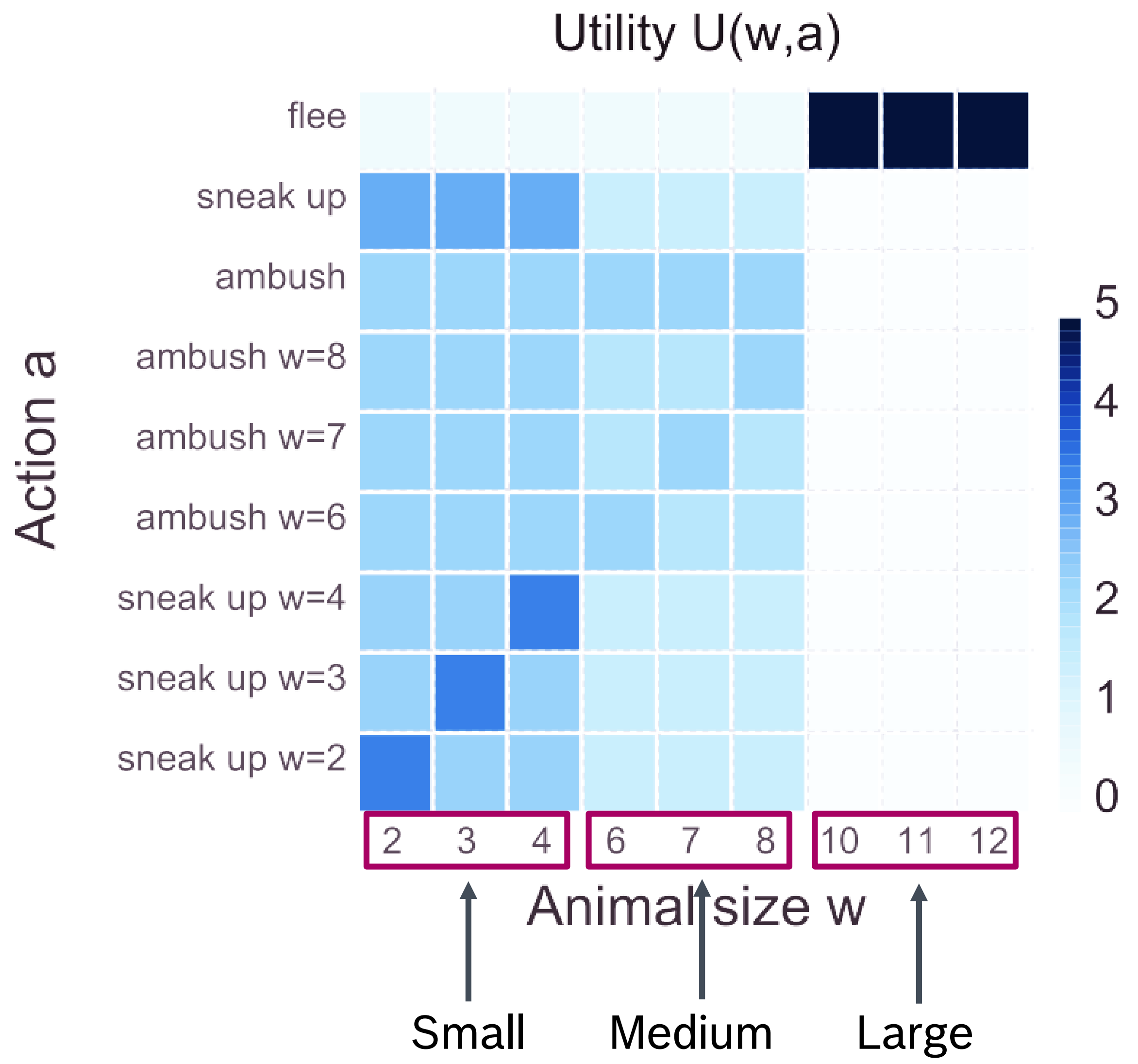
Free energy of action stage acts as utility for the perceptual stage!

$$\Delta F(w,x) := \mathbf{E}_{p^*(a|x)}[U(w,a)] - \frac{1}{\beta_2} D_{\mathrm{KL}}(p^*(a|x)||p(a))$$

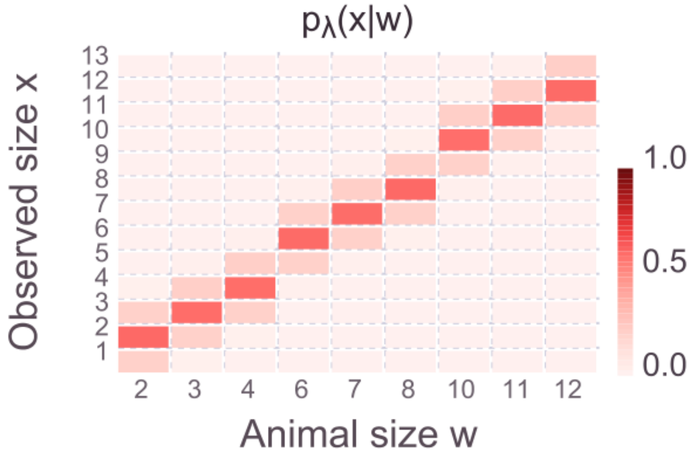
Priors are marginals

$$p(a) = \sum_{w,x} p(w) p^*(x|w) p^*(a|x)$$

Predator-Prey example



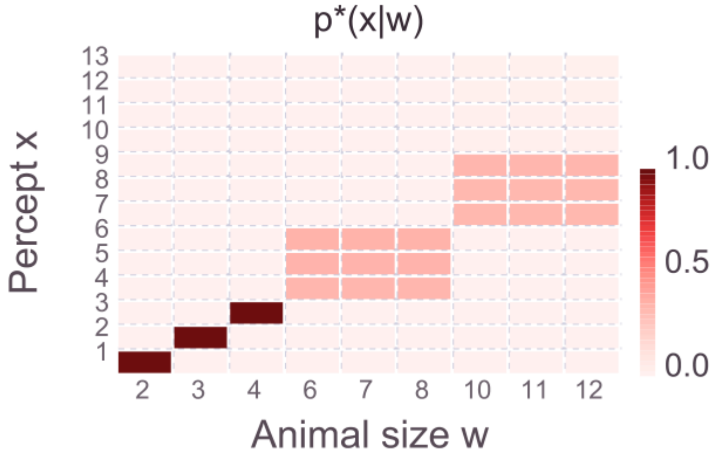
Predator-Prey example



Hand-crafted perceptual model

Gaussian likelihood $p_\lambda(x|w)$

$$x \sim \mathrm{round}[\mathcal{N}(w, \lambda)]$$

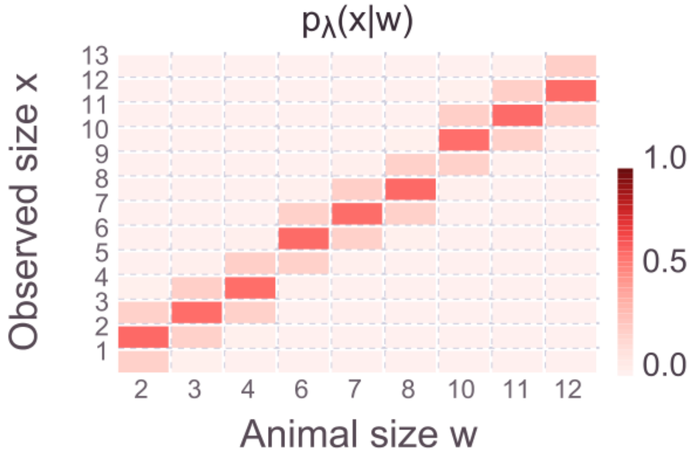


Perceptual model given by optimality principle

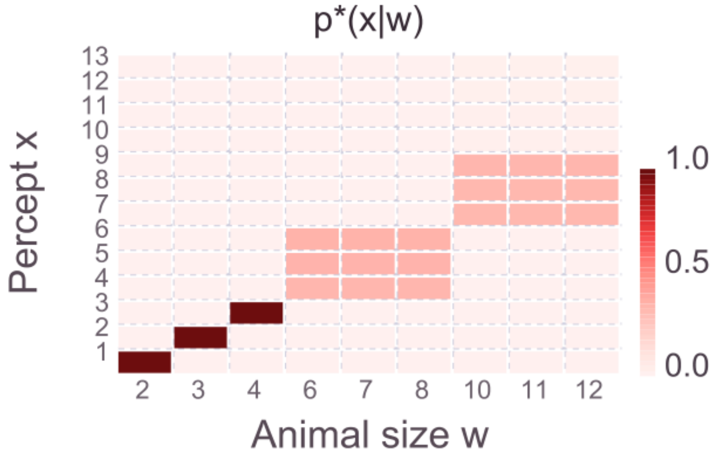
Information-optimal likelihood $p^*(x|w)$

$$\underset{p(x|w),p(a|x)}{\arg\max}\ \mathbf{E}_{p(w,x,a)}[U(w,a)] - \frac{1}{\beta_1} I(W;X) - \frac{1}{\beta_2} I(X;A)$$

Predator-Prey example

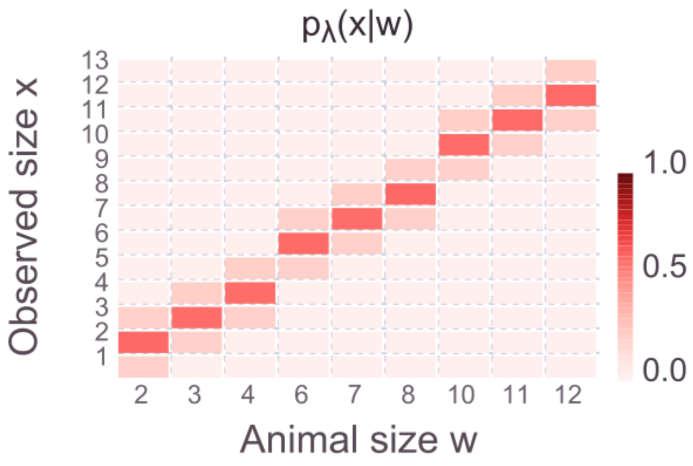


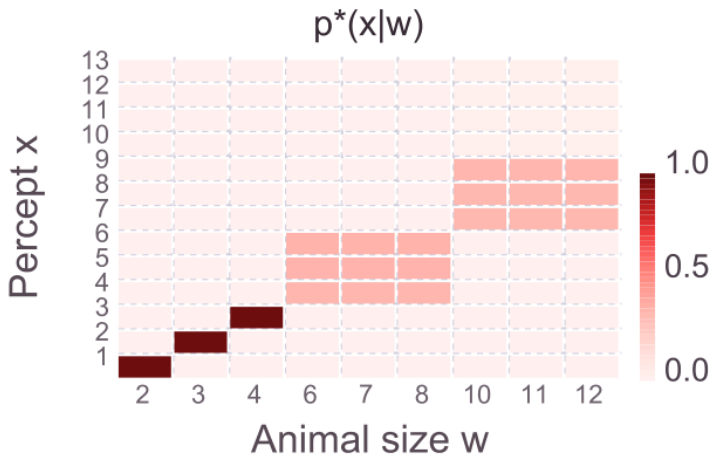
- N bits of information about animal size
- *passive receipt of sensory signals*



- N bits of **task-relevant** information about animal size
- *organization and interpretation of sensory information*

Predator-Prey example





Representations formed by optimality principle:

Efficient

Same number of bits captures more relevant information

Robust

Irrelevant information is discarded by design
(lossy compression)

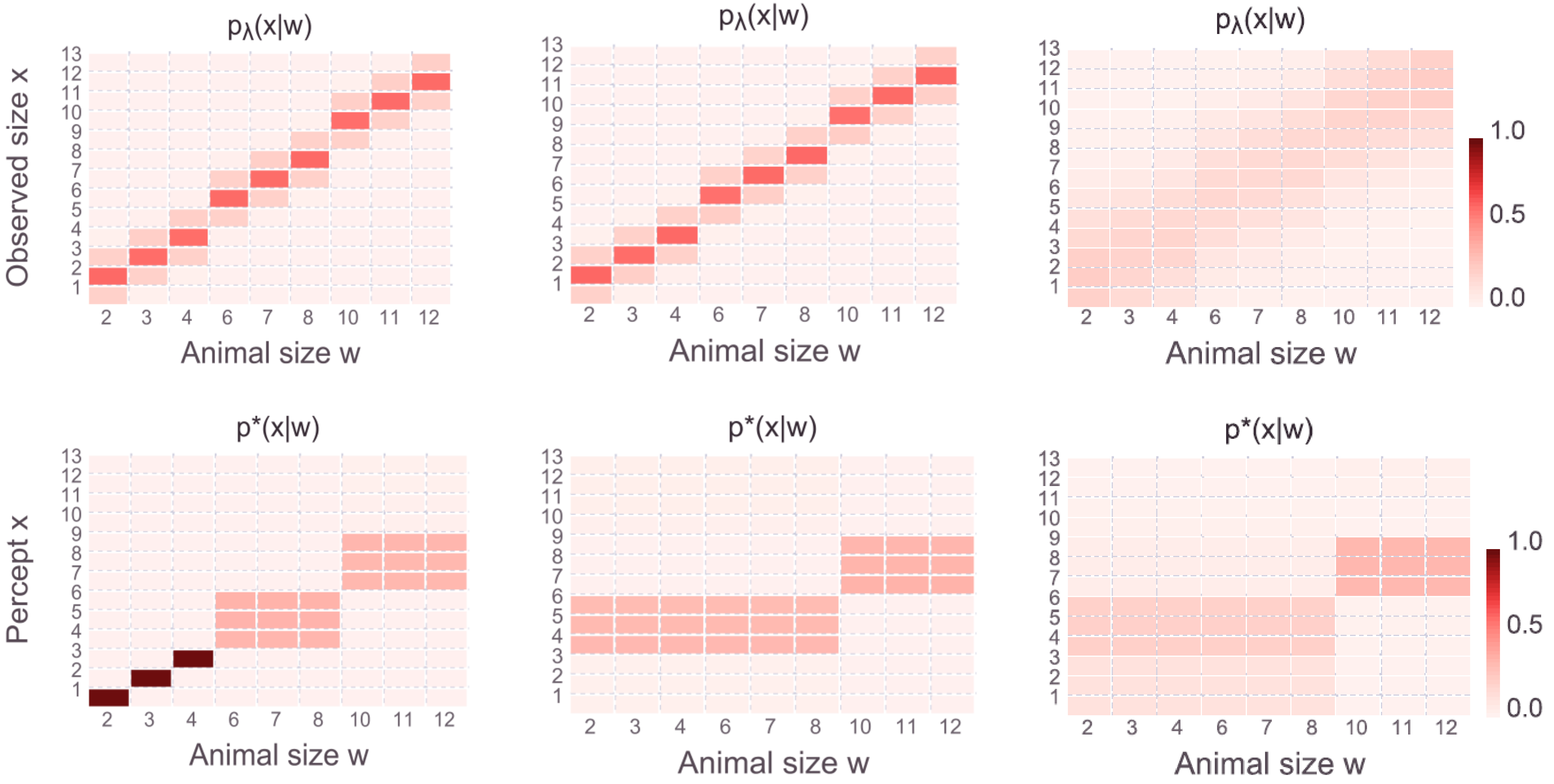
Interpretable

Representations capture task-relevant information

Representation can be interpreted in task-space

Capacity:

Perception	high	high	low
Action	high	low	high



Thanks!

Key-references:

- Information-theoretic bounded rationality: Ortega et al. 2015, *Information-theoretic bounded rationality*, arxiv:1512.06789
- Coupling perception and action: Genewein et al. 2015, *Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle*, Frontiers in Robotics and AI

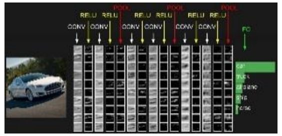
http://tim.inversetemperature.net/research/

Image credits:

https://pixabay.com/en/camera-cctv-security-cam-watching-156730/ (CC0)

http://chiara-robot.org/Challenge/images/gatech-chess.jpg (AAAI-10 Small-Scale Manipulation Challenge)

http://www.carnewscafe.com/2015/05/bosch-stereo-camera-enters-production-as-single-piece-solution-for-emergency-braking/



http://cs231n.github.io/convolutional-networks/

https://www.flickr.com/photos/27668445@N03/3191999216 (CC BY 2.0 https://creativecommons.org/licenses/by/2.0/)