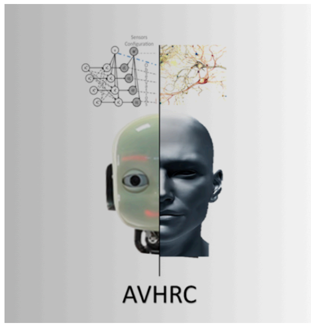
A probabilistic tour of visual attention and gaze shift computational models



Active Vision and perception in Human (-Robot) Collaboration (AVHRC 2020) September, 2020

Giuseppe Boccignone

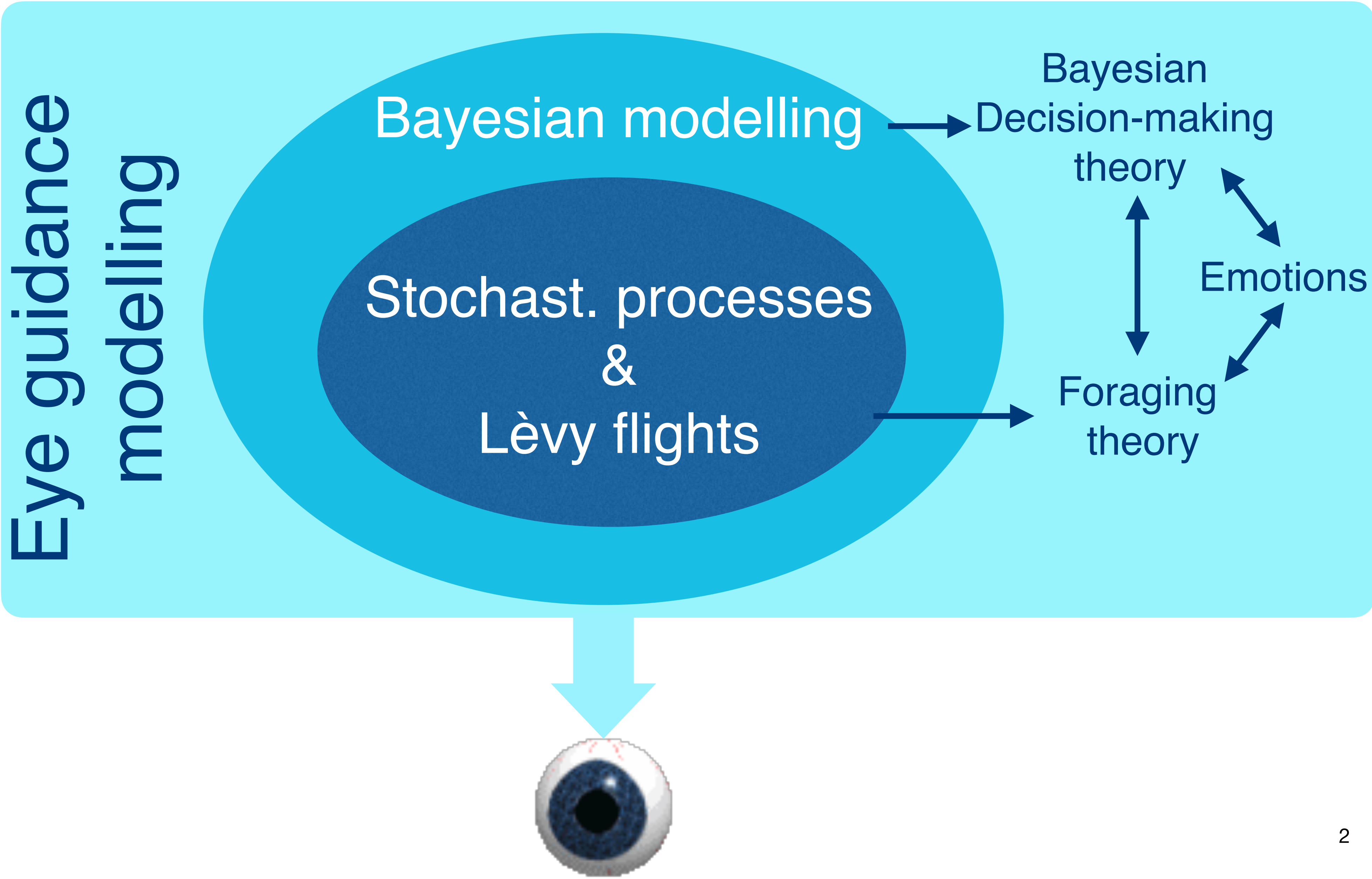
Dipartimento di Informatica
Università di Milano

Giuseppe.Boccignone@unimi.it

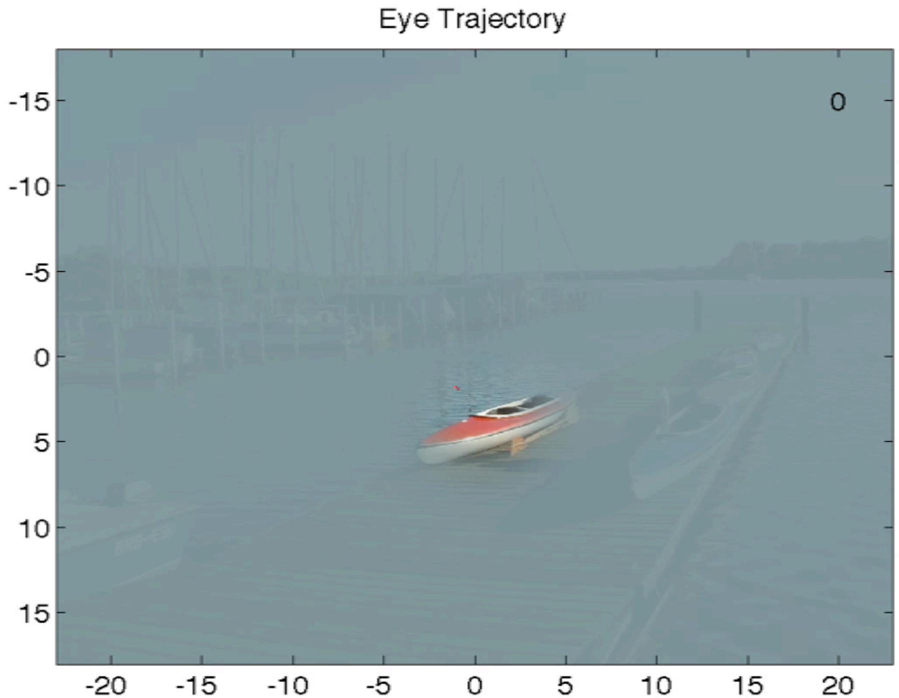
http://phuselab.di.unimi.it

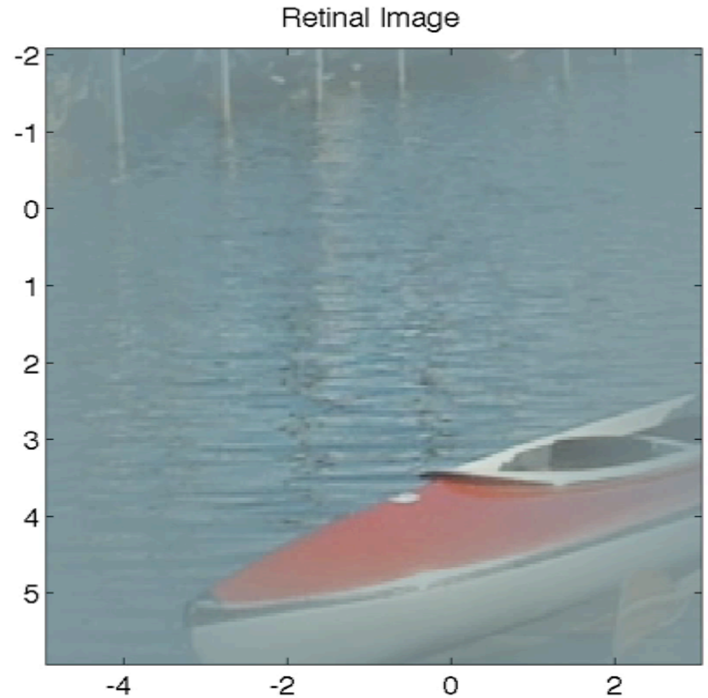
P HU SE Lab
Perceptual computing and HUman SEnsing

Some key points of this talk



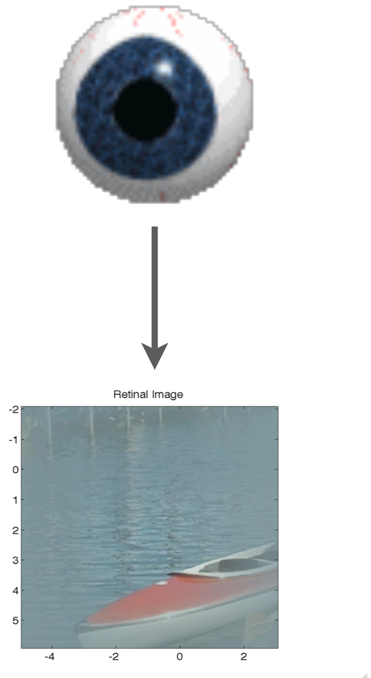
Yet, we have this wandering eye…
//at the heart of active sensing





the world **W(t)**
as we perceive it
at time t

Computational models of eye guidance
\\the bare essence



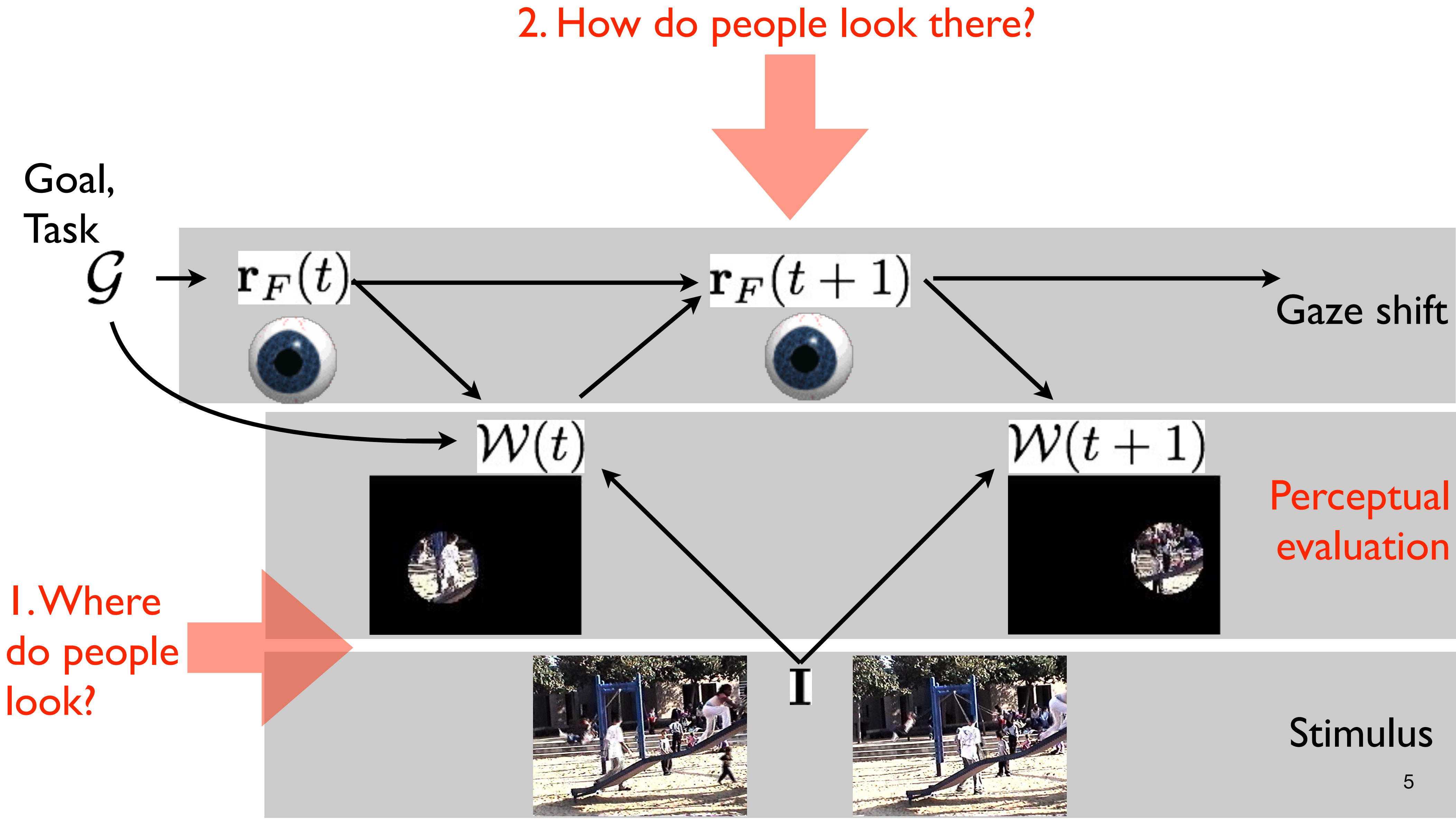
Computational Theory

1. Where do people look?

$$\mathbf{I} \underset{\mathbf{T}}{\longmapsto} \{\mathbf{r}_F(1), \mathbf{r}_F(2), \cdots\}$$

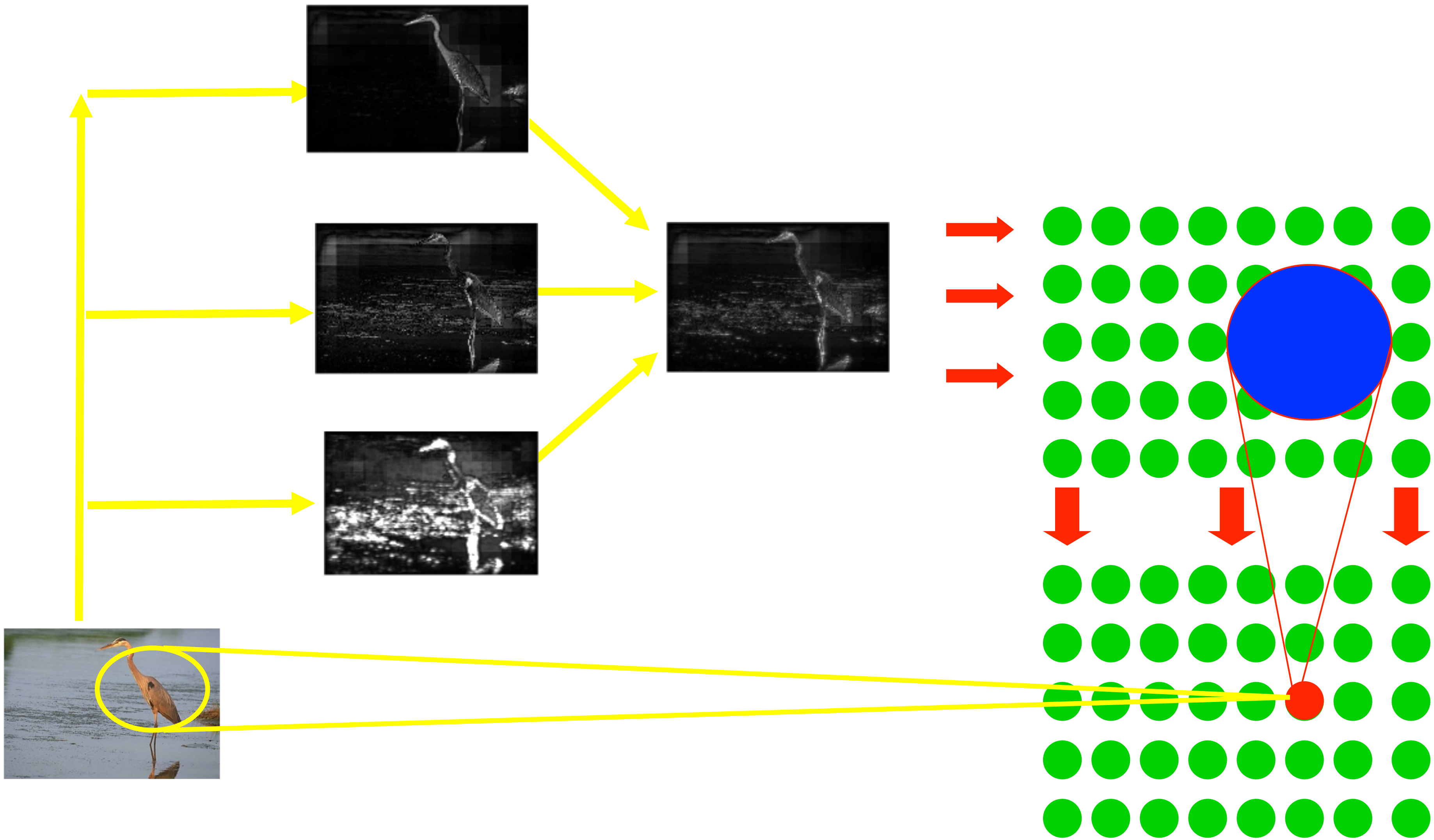
2. How do people look there?

Computational models of eye guidance
\\the bare essence



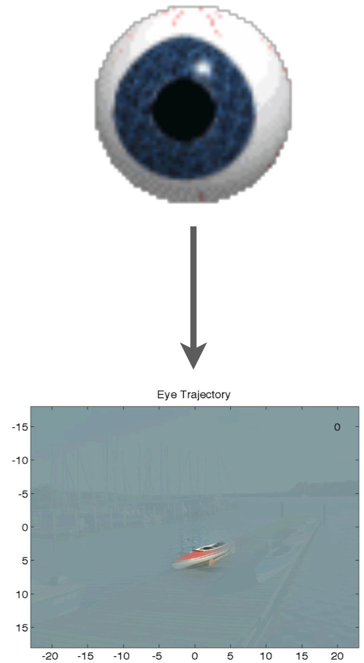
Computational models of eye guidance
//The historical baseline: Itti, Koch & Niebur model

1. Where do people look?



2. How do people look there?

Computational models of eye guidance
\\the bare essence



Computational Theory

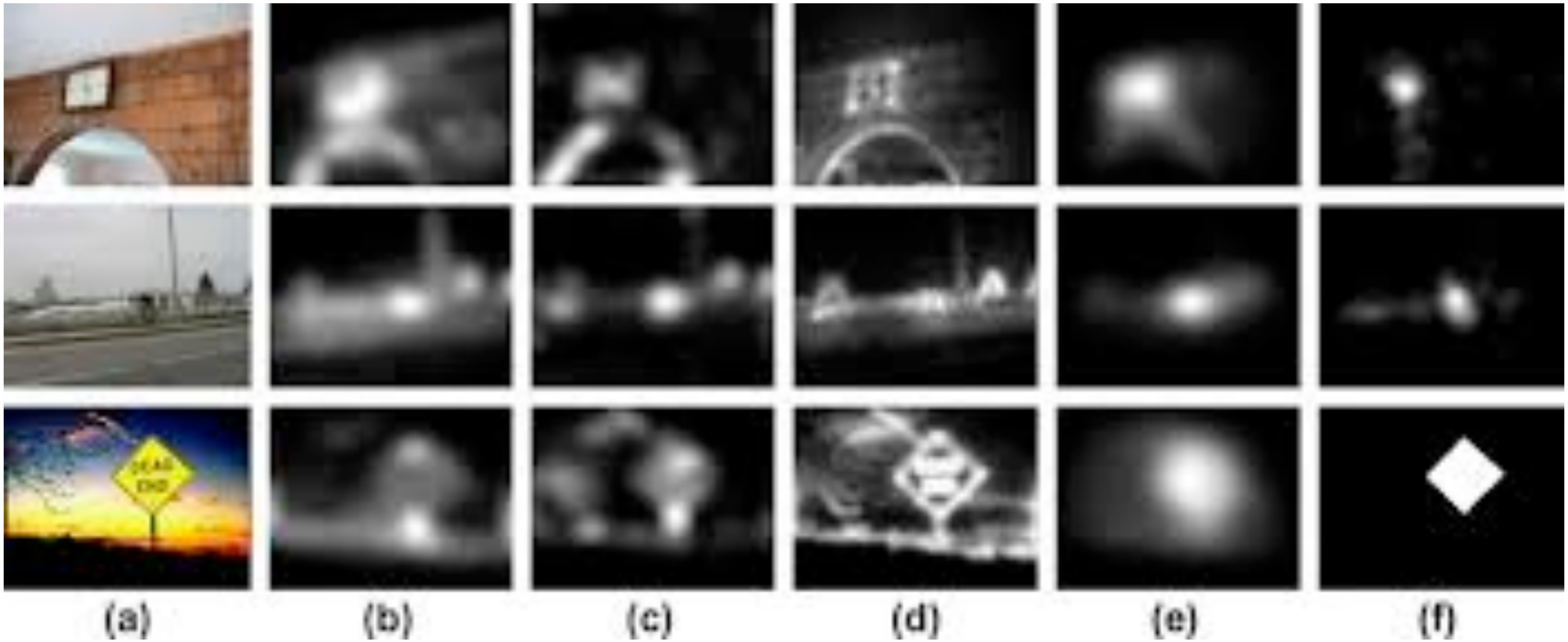
1. Where do people look?

$$\mathbf{I} \mapsto \mathcal{R}$$ (e.g., saliency map)

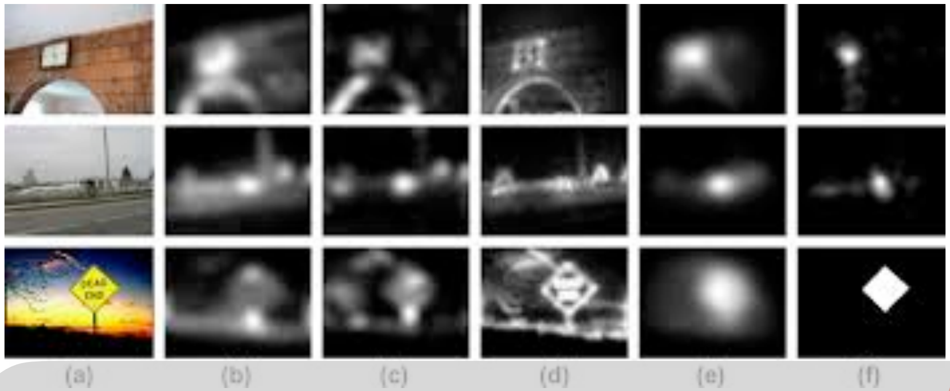
$$\mathcal{R} \mapsto \{\mathbf{r}_F(1), \mathbf{r}_F(2), \cdots\}$$

2. How do people look there?

Computational models of eye guidance
//anatomy & misery of saliency maps



Computational models of eye guidance //anatomy & misery of saliency maps



1. Where do people look?

$$\mathbf{I} \mapsto \mathcal{R}$$

(e.g., saliency map)

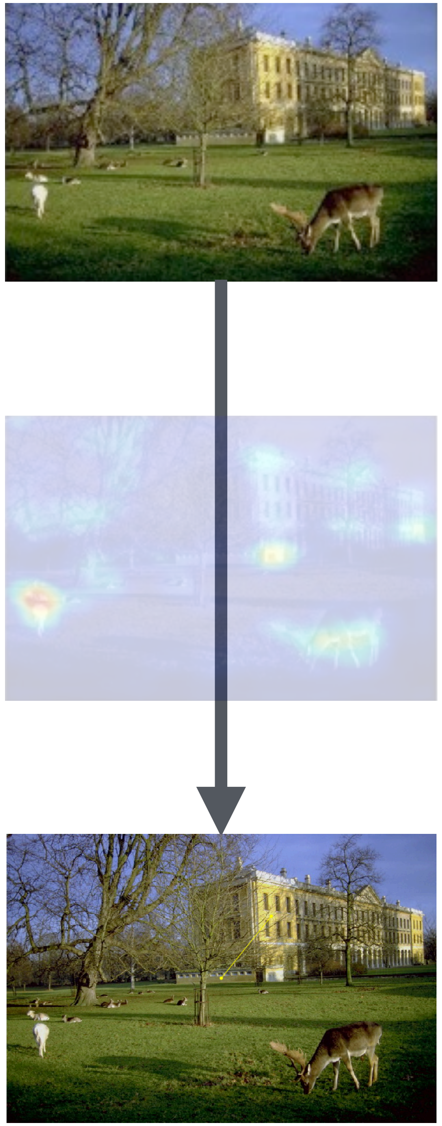
$$\cancel{\mathcal{R} \mapsto \{\mathbf{r}_F(1), \mathbf{r}_F(2), \cdots\}}$$

~~2. How do people look there?~~

Computational models of eye guidance //anatomy & misery of saliency maps

$$\overbrace{P(\mathbf{r} \mid \mathcal{W})}^{\text{posterior prob. of gaze shift}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r})}^{\text{data likelihood under the shift}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r})}^{\text{gaze shift prior}}$$

this is a shift



$\mathbf{r}_F(t) \rightarrow \mathbf{r}_F(t+1)$

Computational models of eye guidance
//anatomy & misery of saliency maps

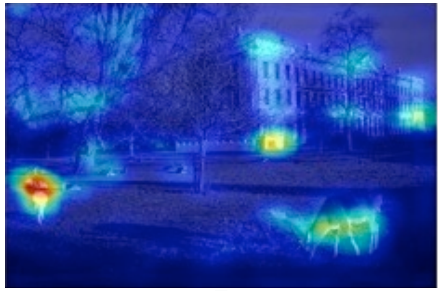
$$\overbrace{P(\mathbf{r} \mid \mathcal{W})}^{\text{posterior prob. of gaze shift}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r})}^{\text{data likelihood under the shift}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r})}^{\text{gaze shift prior}}$$

this is a shift

$$P(\mathbf{r}) = P(\mathbf{r}_F(t) - \mathbf{r}_F(t-1)) \simeq P(\mathbf{r}_F(t) \mid \mathbf{r}_F(t-1)) = P(\mathbf{r}_F(t))$$

$$\overbrace{P(\mathbf{r}_F \mid \mathcal{W})}^{\text{posterior prob. of gazing at}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r}_F)}^{\text{data likelihood under gaze at}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r}_F)}^{\text{prior prob. of gazing at}}$$

this is a point



$\mathbf{r}_F(t) \rightarrow \mathbf{r}_F(t+1)$

Computational models of eye guidance //anatomy & misery of saliency maps

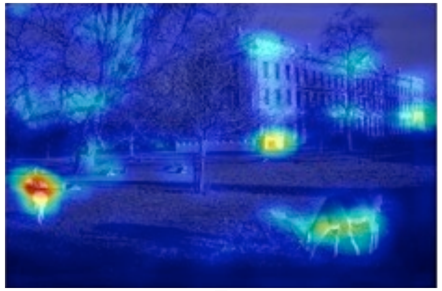
$$\overbrace{P(\mathbf{r} \mid \mathcal{W})}^{\text{posterior prob. of gaze shift}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r})}^{\text{data likelihood under the shift}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r})}^{\text{gaze shift prior}}$$

this is a shift

$$P(\mathbf{r}) = P(\mathbf{r}_F(t) - \mathbf{r}_F(t-1)) \simeq P(\mathbf{r}_F(t) \mid \mathbf{r}_F(t-1)) = P(\mathbf{r}_F(t))$$

$$\overbrace{P(\mathbf{r}_F \mid \mathcal{W})}^{\text{posterior prob. of gazing at}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r}_F)}^{\text{data likelihood under gaze at}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r}_F)}^{\text{prior prob. of gazing at}}$$

this is a point



$$\overbrace{P(\mathbf{L} \mid \mathbf{F})}^{\text{posterior prob. of selecting location L}} = \frac{\overbrace{P(\mathbf{F} \mid \mathbf{L})}^{\text{feature likelihood under location L}}}{P(\mathbf{F})} \overbrace{P(\mathbf{L})}^{\text{prior prob. of location L}}$$

this is a map

$$\mathbf{r}_F(t) \to \mathbf{r}_F(t+1)$$

Computational models of eye guidance
//anatomy & misery of saliency maps

$$\overbrace{P(\mathbf{r} \mid \mathcal{W})}^{\text{posterior prob. of gaze shift}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r})}^{\text{data likelihood under the shift}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r})}^{\text{gaze shift prior}}$$

this is a shift

$$P(\mathbf{r}) = P(\mathbf{r}_F(t) - \mathbf{r}_F(t-1)) \simeq P(\mathbf{r}_F(t) \mid \mathbf{r}_F(t-1)) = P(\mathbf{r}_F(t))$$

$$\overbrace{P(\mathbf{r}_F \mid \mathcal{W})}^{\text{posterior prob. of gazing at}} = \frac{\overbrace{P(\mathcal{W} \mid \mathbf{r}_F)}^{\text{data likelihood under gaze at}}}{P(\mathcal{W})} \overbrace{P(\mathbf{r}_F)}^{\text{prior prob. of gazing at}}$$

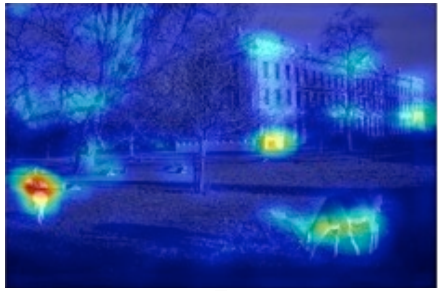
this is a point

$$\overbrace{P(\mathbf{L} \mid \mathbf{F})}^{\text{posterior prob. of selecting location L}} = \frac{\overbrace{P(\mathbf{F} \mid \mathbf{L})}^{\text{feature likelihood under location L}}}{P(\mathbf{F})} \overbrace{P(\mathbf{L})}^{\text{prior prob. of location L}}$$

this is a map $P(\mathbf{F} \mid \mathbf{L}) = P(\mathbf{L}) = const.$

$$\overbrace{P(\mathbf{L} \mid \mathbf{F})}^{\text{posterior prob. of selecting location L}} \propto \overbrace{\frac{1}{P(\mathbf{F})}}^{\text{salience at location L}} .$$

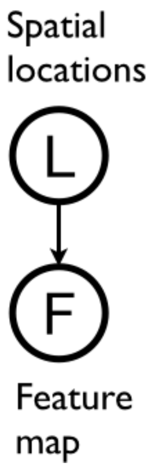
(Itti & Koch)



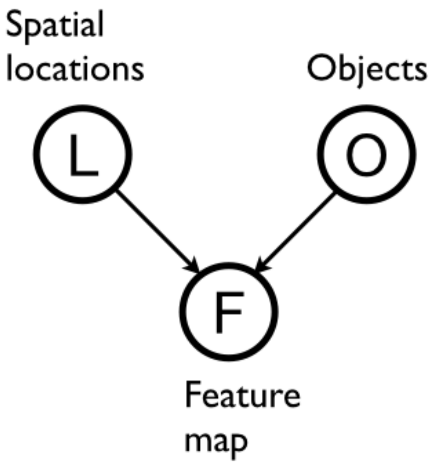
$\mathbf{r}_F(t) \rightarrow \mathbf{r}_F(t+1)$

Computational models of eye guidance
//anatomy & misery of saliency maps

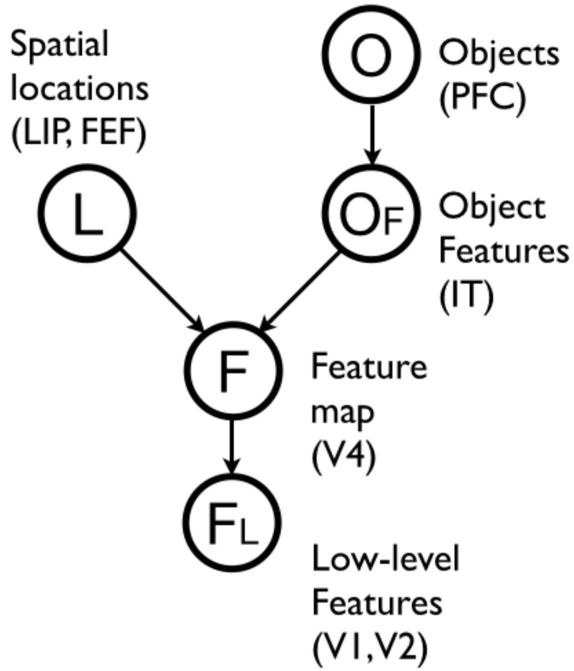
$$\overbrace{P(\mathbf{L} \mid \mathbf{F})}^{\text{posterior prob. of selecting location L}} = \frac{\overbrace{P(\mathbf{F} \mid \mathbf{L})}^{\text{feature likelihood under location L}}}{P(\mathbf{F})} \overbrace{P(\mathbf{L})}^{\text{prior prob. of location L}}$$



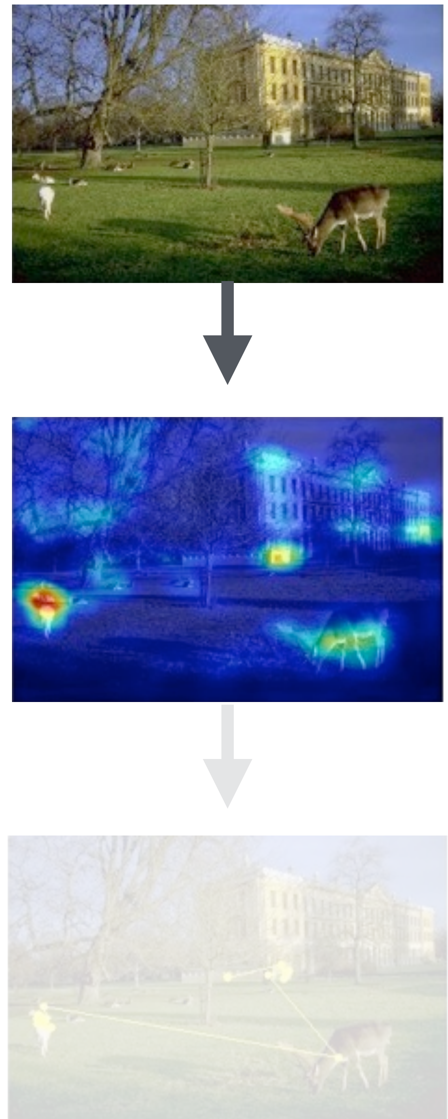
Torralba



Torralba
Perona



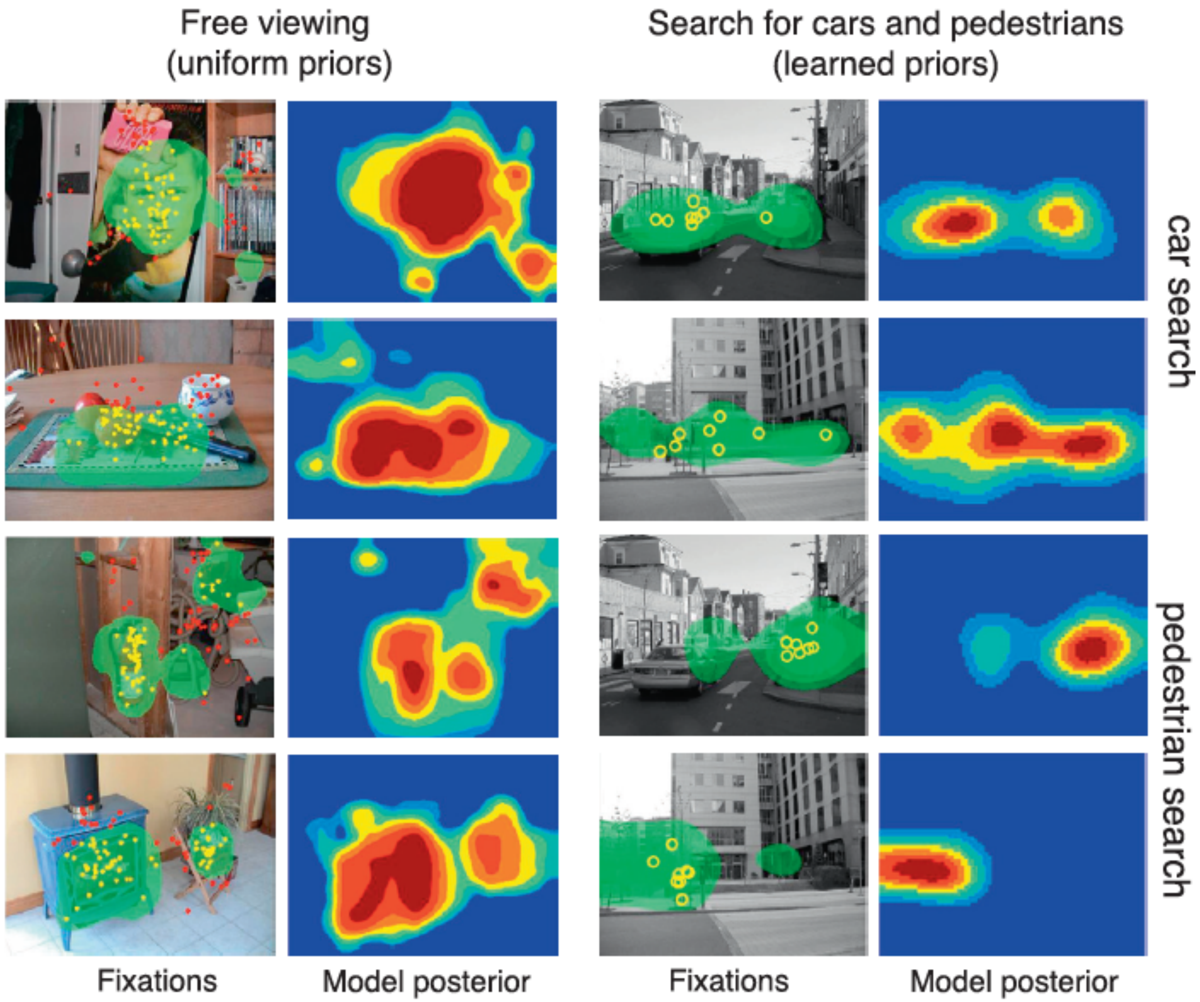
Poggio et al.

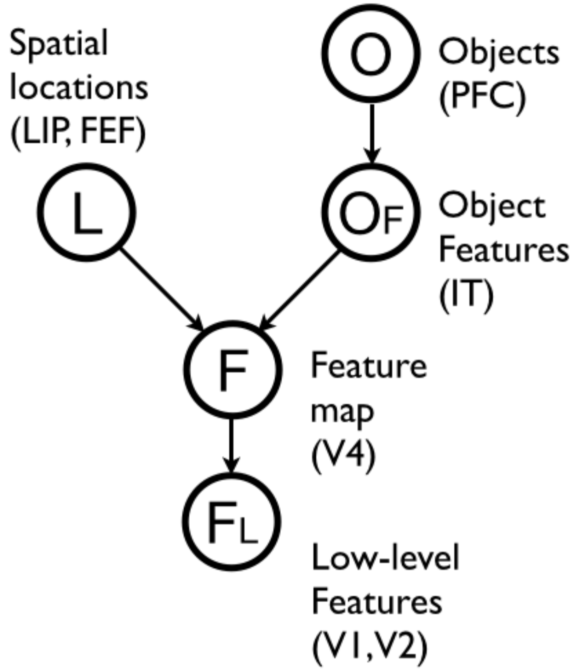


$\mathbf{x} = \mathbf{x}_F(t) - \mathbf{x}_F(t-1)$

Computational models of eye guidance
//anatomy & misery of saliency maps

$$\overbrace{P(\mathbf{L}\mid\mathbf{F})}^{\text{posterior prob. of selecting location L}} = \frac{\overbrace{P(\mathbf{F}\mid\mathbf{L})}^{\text{feature likelihood under location L}}}{P(\mathbf{F})}\;\overbrace{P(\mathbf{L})}^{\text{prior prob. of location L}}$$





Poggio et al.

To sum up…

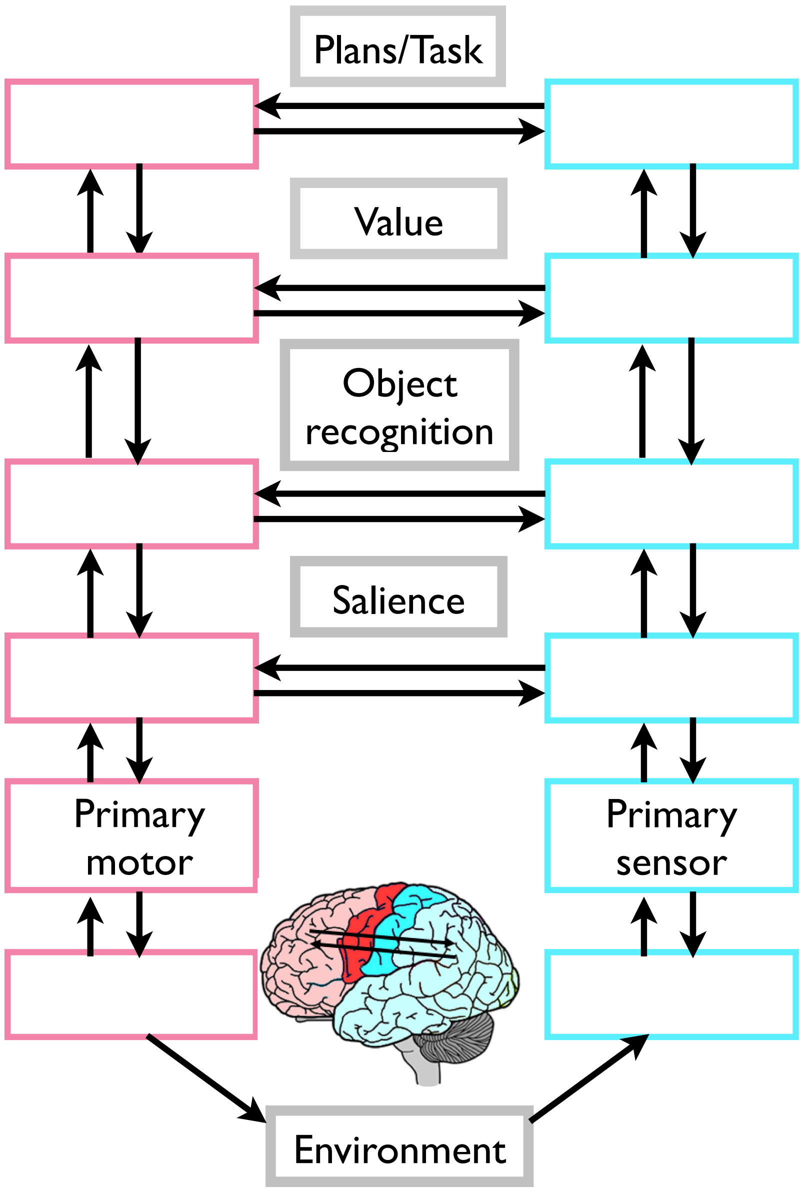
Journal of Vision (2011) 11(5):9, 1–30 http://www.journalofvision.org/content/11/5/9

Eye movements and perception: A selective review

Alexander C. Schütz Department of Psychology, Gießen University, Gießen, Germany

Doris I. Braun Department of Psychology, Gießen University, Gießen, Germany

Karl R. Gegenfurtner Department of Psychology, Gießen University, Gießen, Germany



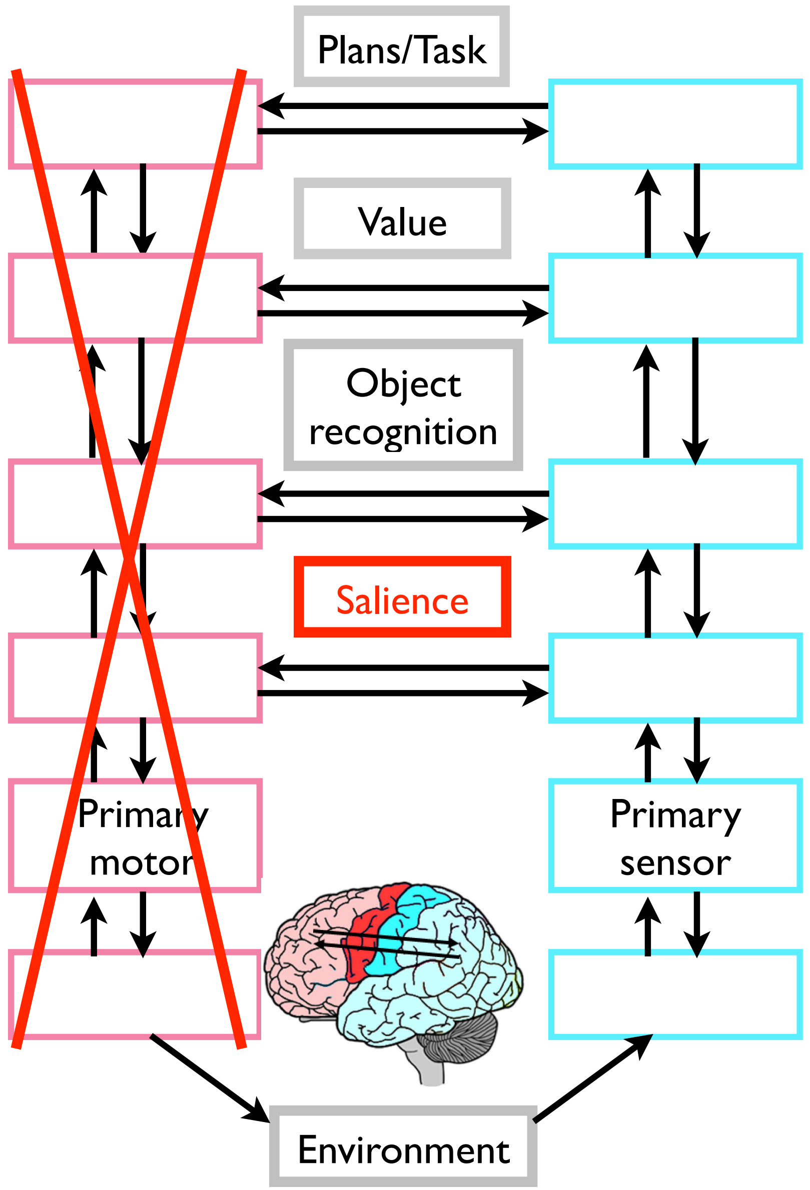
To sum up…

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 35, NO. 1, JANUARY 2013

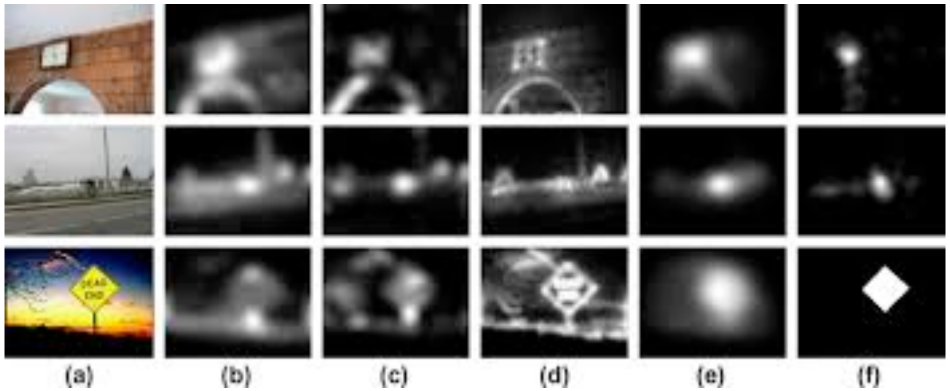
State-of-the-Art in Visual Attention Modeling

Ali Borji, *Member*, *IEEE*, and Laurent Itti, *Member*, *IEEE*

65 models:
variations,
variations of variations,
variations of variations of variations,......
on base schemes (Itti & Koch)



Computational models of eye guidance //anatomy & misery of saliency maps



deterministic gaze shift, no variability!

1. Where do people look?

$$\mathbf{I} \mapsto \mathcal{R}$$ (e.g., saliency map)

$$\mathcal{R} \mapsto \{\mathbf{r}_F(1), \mathbf{r}_F(2), \cdots\}$$

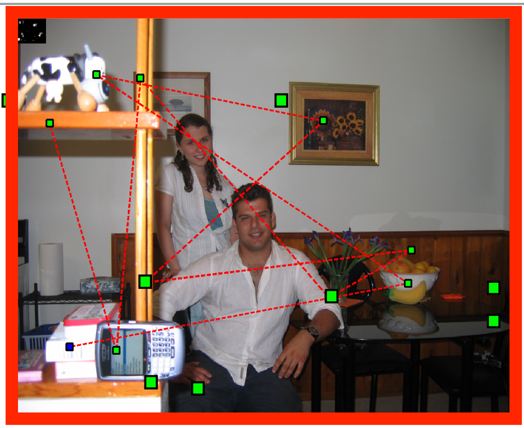
2. How do people look there?

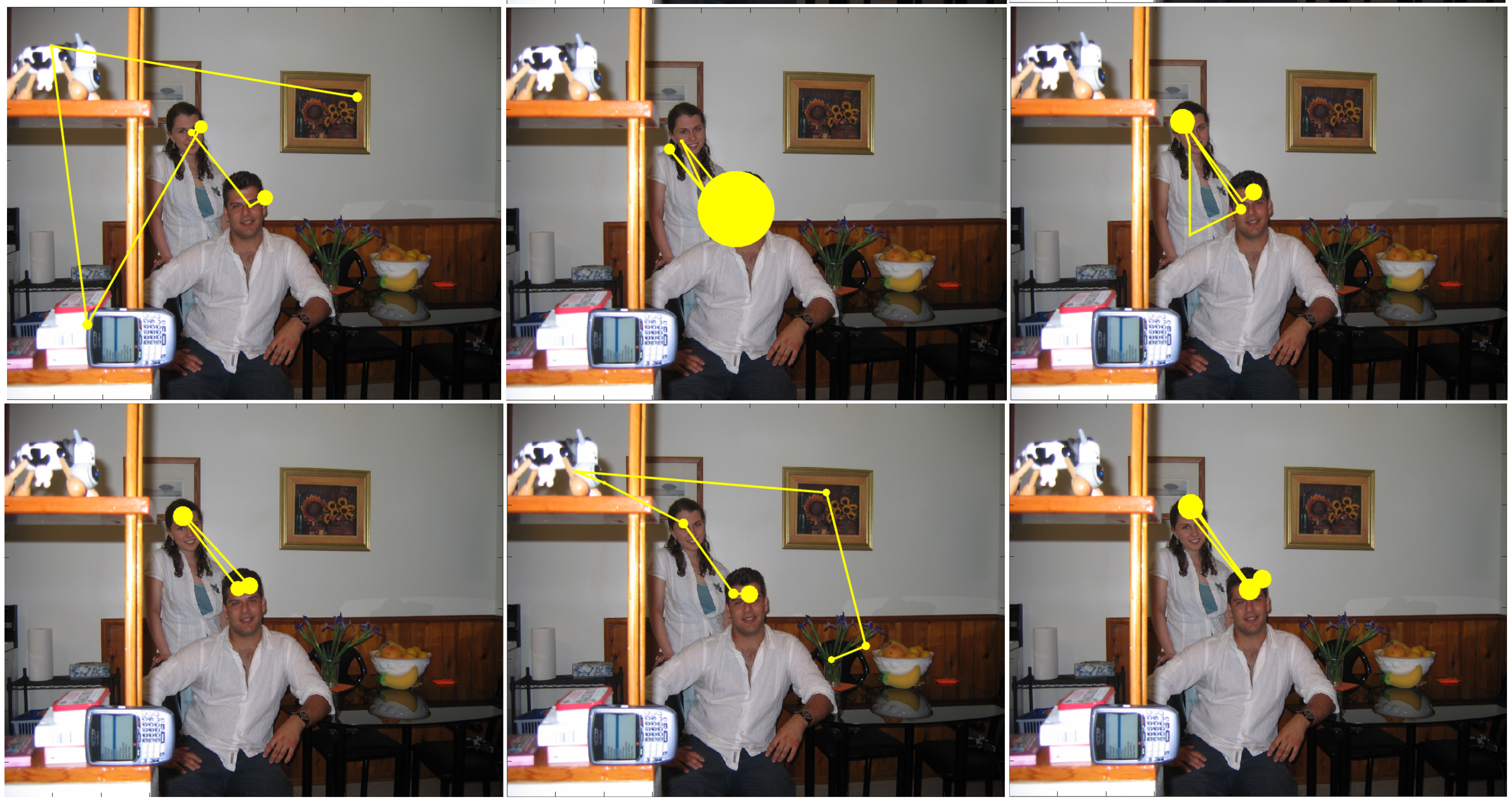
$$\arg\max \mathcal{R}$$

The problem of variability
// How random are gaze shifts?

The problem of variability
// How random are gaze shifts?

(Itti & Koch)





Human

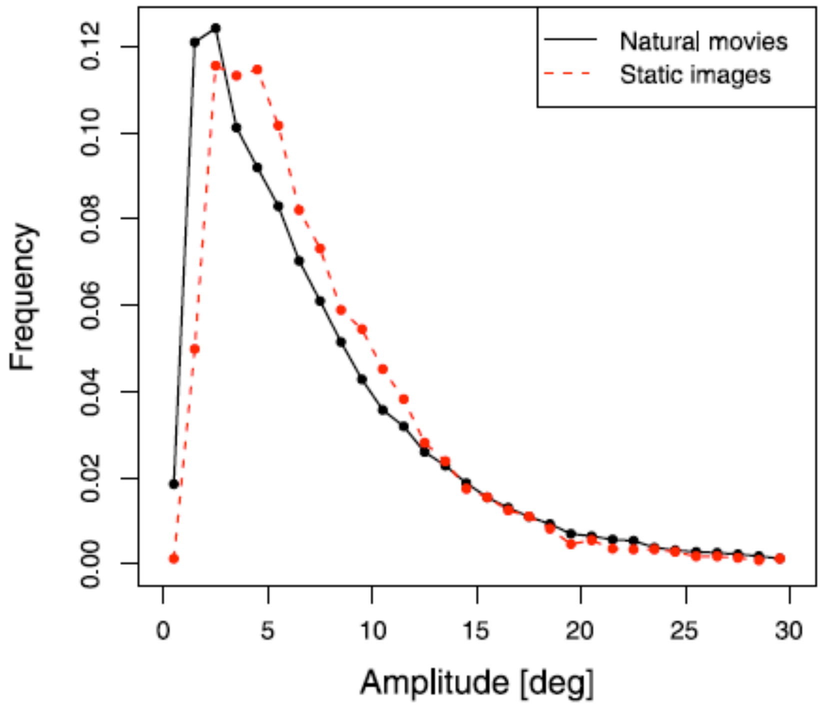
The problem of variability
//Oculomotor tendencies

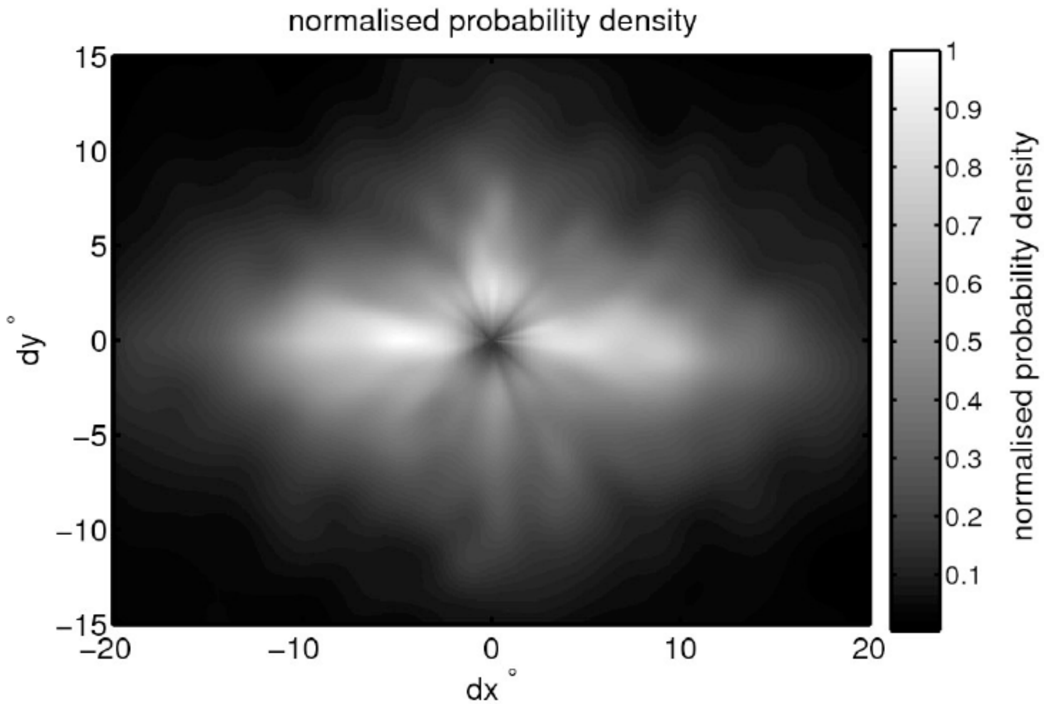
VISUAL COGNITION, 2009, 17 (6/7), 1029–1054

Psychology Press
Taylor & Francis Group

The prominence of behavioural biases in eye guidance

Benjamin W. Tatler and Benjamin T. Vincent
University of Dundee, Dundee, UK



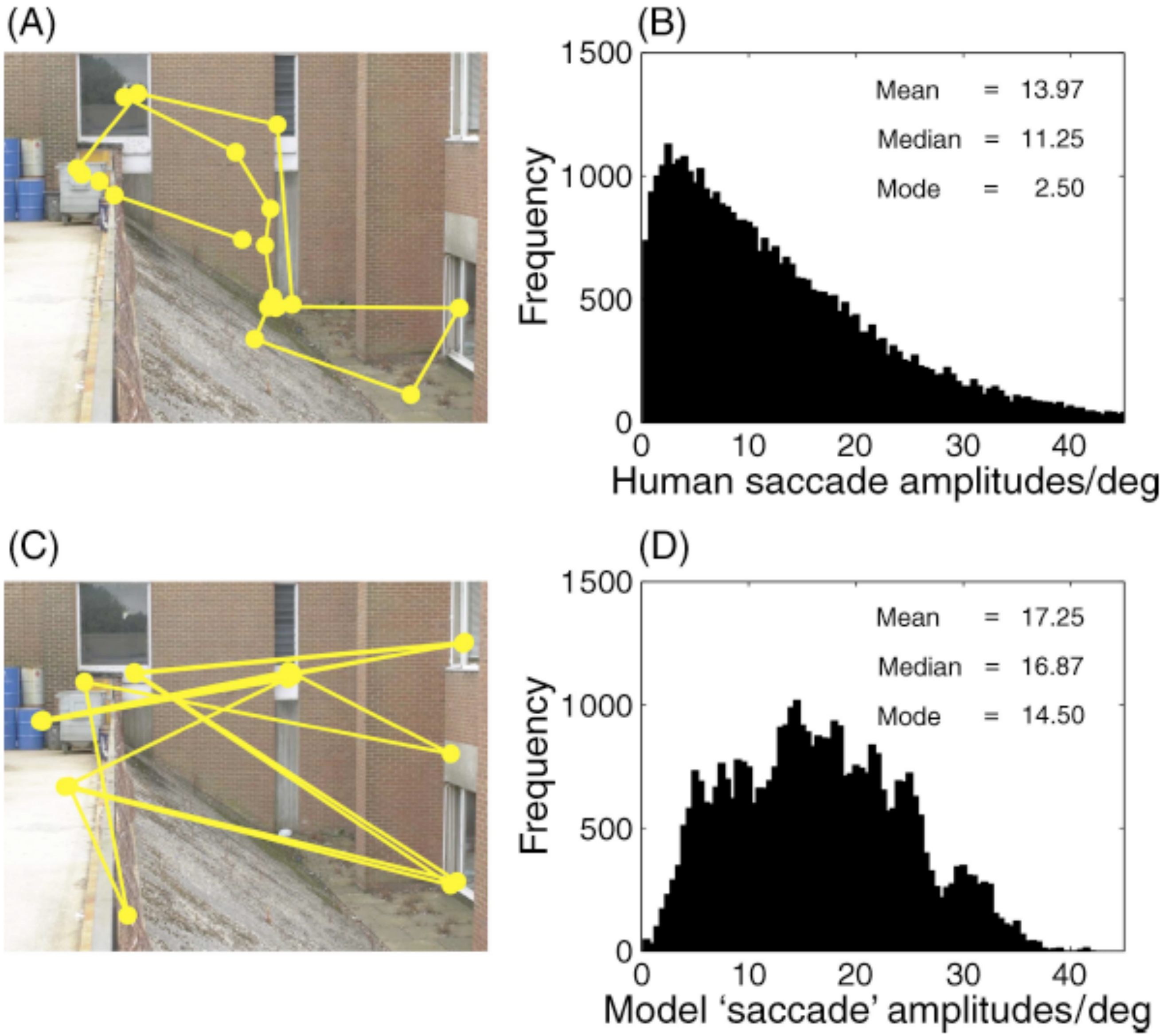


The problem of variability
//Oculomotor tendencies

Journal of Vision (2011) 11(5):5, 1–23 Tatler, Hayhoe, Land, & Ballard

Human

(Itti & Koch)



The problem of variability //Oculomotor tendencies

- Oculomotor tendencies:
  - regularities that are common across all instances of and manipulations to the behavior
  - Tatler & Vincent:
    - a model based on oculomotor biases alone performs better than the standard salience model

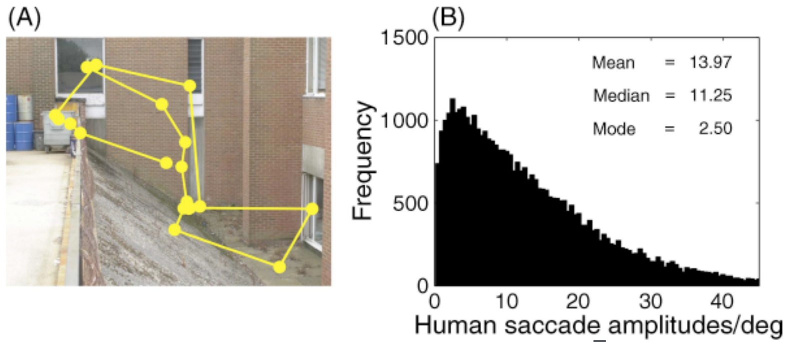
If one samples from prior only

$$\mathbf{r}(t) \sim P(\mathbf{r}(t)), \quad t = 1, 2, \cdots$$

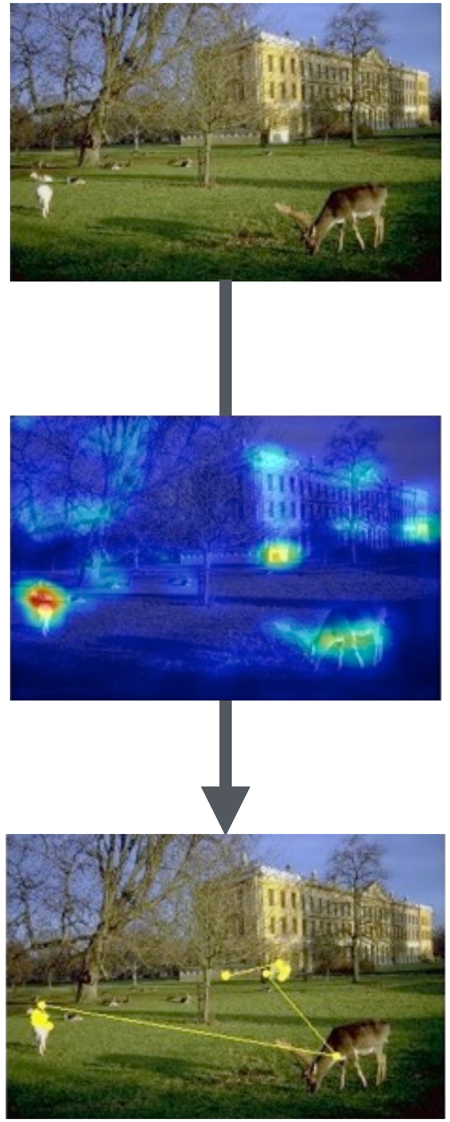
blind to visual information, out-performs feature-based accounts of eye guidance:

0. 648 area under the receiver operator curve (AUC) as opposed to 0. 593 for edge information and 0. 565 for salience information!

Computational models of eye guidance //bringing variability into the game

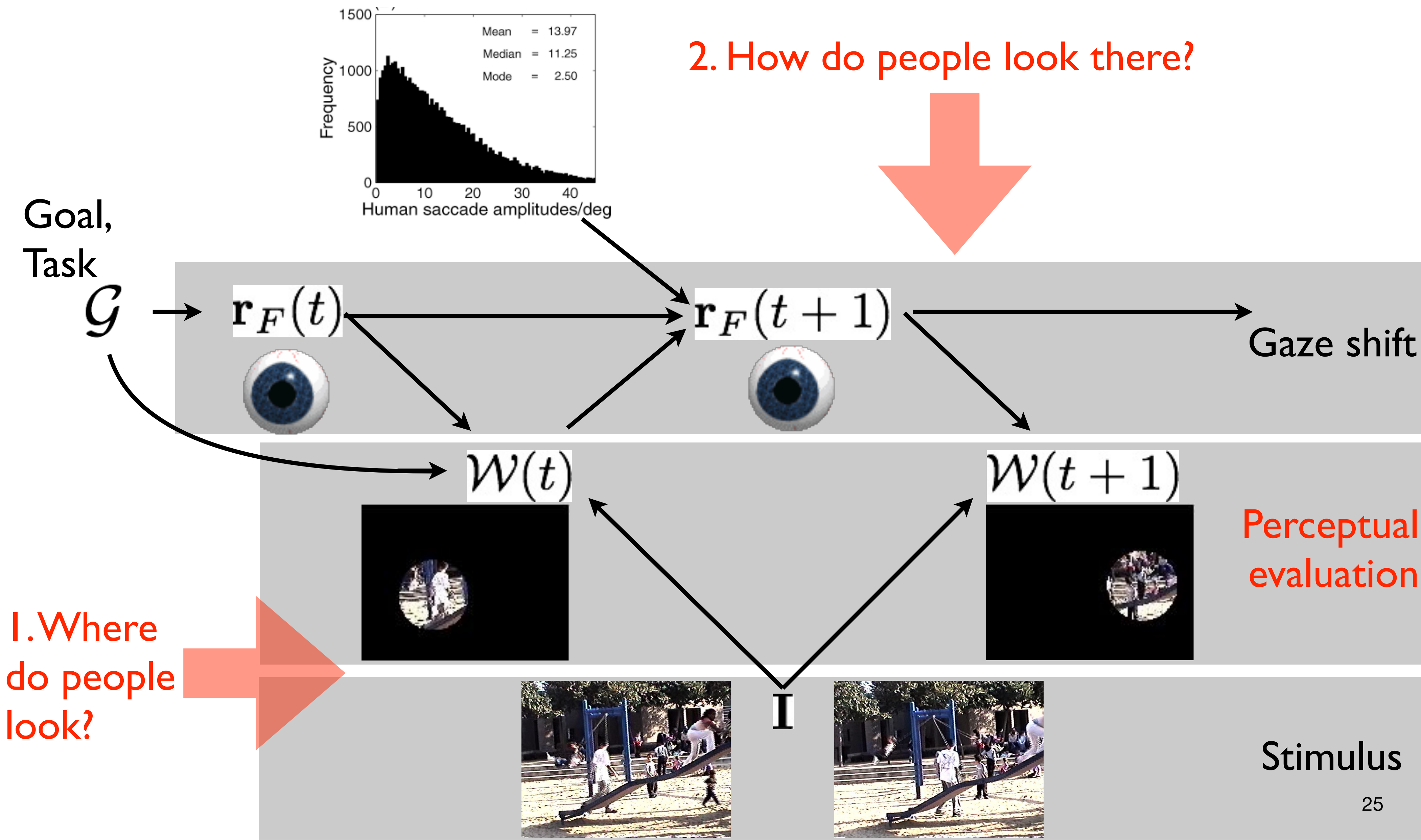


$$\overbrace{P(\mathbf{r}\mid\mathcal{W})}^{\text{posterior prob. of gaze shift}} = \frac{\overbrace{P(\mathcal{W}\mid\mathbf{r})}^{\text{data likelihood under the shift}}}{P(\mathcal{W})}\boxed{\overbrace{P(\mathbf{r})}^{\text{gaze shift prior}}}$$

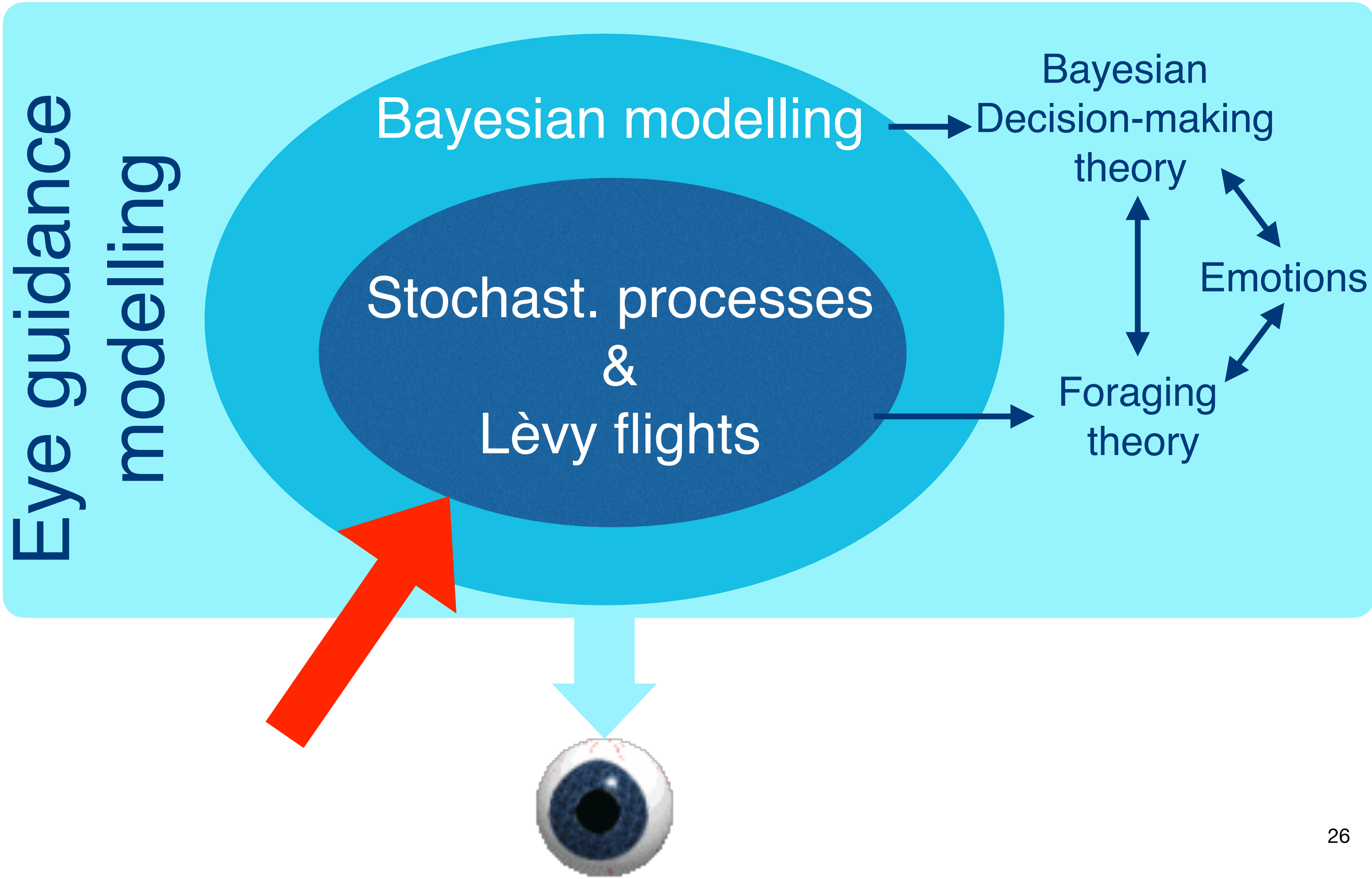


$$\mathbf{r}_F(t) \longrightarrow \mathbf{r}_F(t+1)$$

Computational models of eye guidance
\\the bare essence



Computational models of eye guidance //bringing variability into the game



Computational models of eye guidance
//bringing variability into the game

Neurocomputing 32–33 (2000) 643–650

The ecology of gaze shifts

Dirk Brockmann*, Theo Geisel

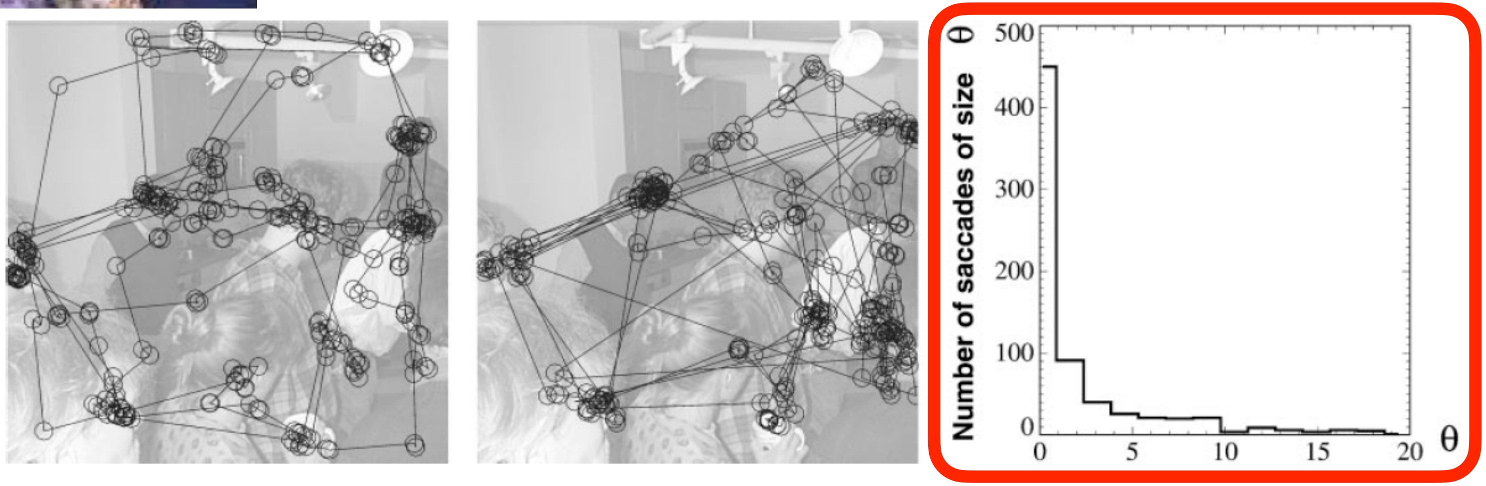


Fig. 1. Left, Center: Two typical scanpaths on different trials by the same subject. Each scanpath consists of approximately 350 saccades. Right: Saccadic magnitude histogram calculated from the scanpaths depicted. Θ denotes saccadic magnitude in degrees of visual angle.

Bringing variability into the game //anomalous walks

Brownian (Gaussian) walk

Cauchy walk

646 D. Brockmann, T. Geisel / Neurocomputing 32–33 (2000) 643–650

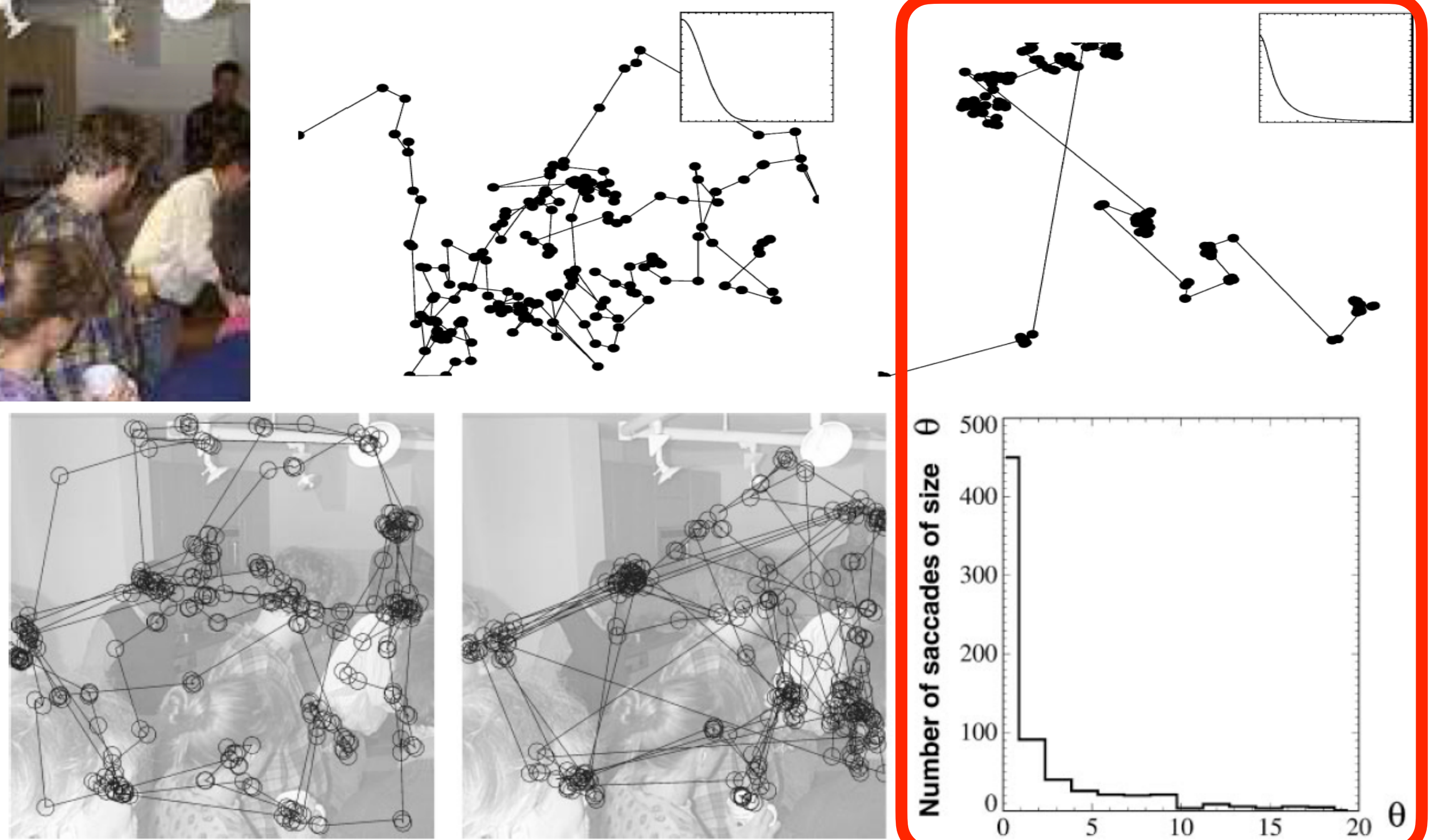
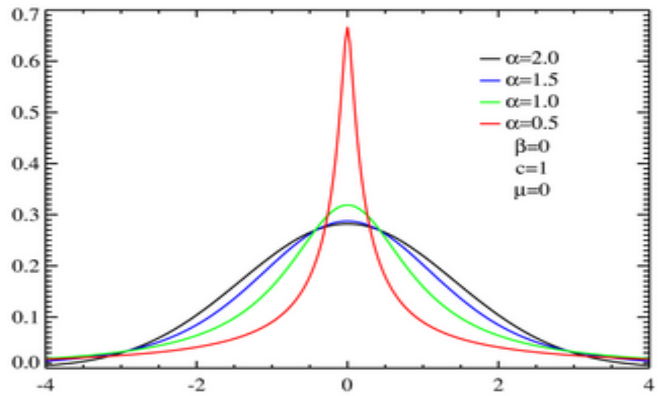
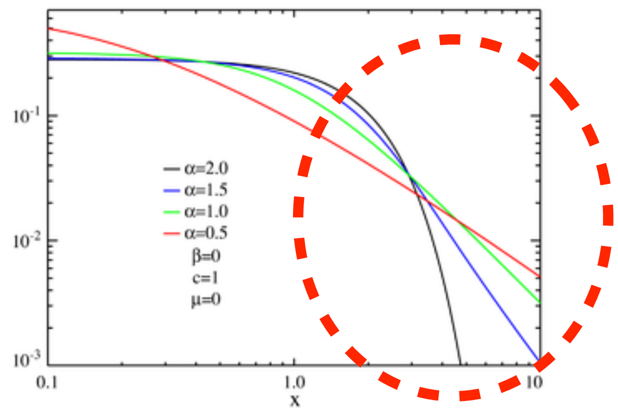


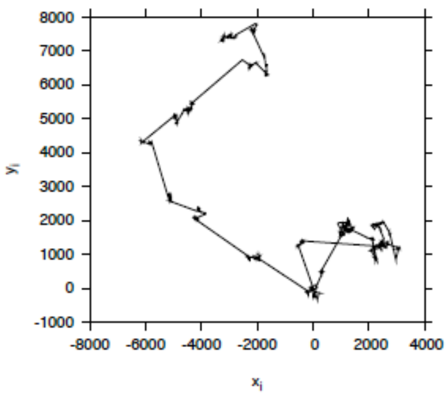
Fig. 1. Left, Center: Two typical scanpaths on different trials by the same subject. Each scanpath consists of approximately 350 saccades. Right: Saccadic magnitude histogram calculated from the scanpaths depicted. Θ denotes saccadic magnitude in degrees of visual angle.

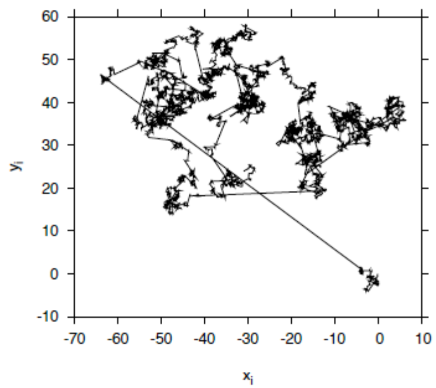
Bringing variability into the game //anomalous walks

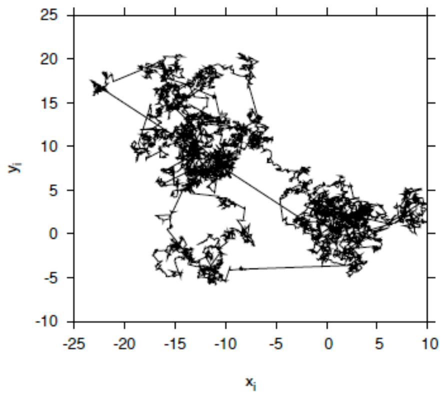
Lèvy alpha-stable distributions

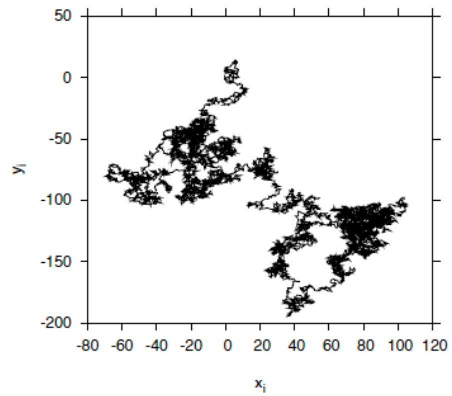










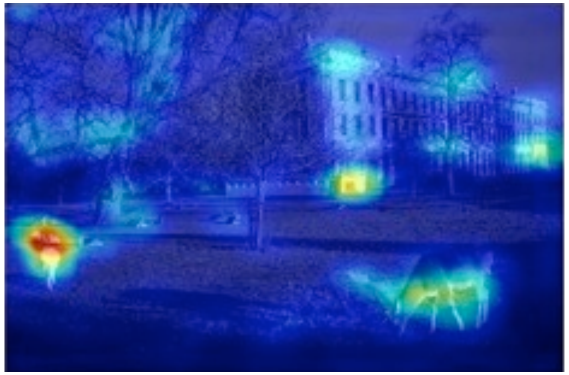


$\alpha = 1$ $\alpha = 1.5$ $\alpha = 1.8$ $\alpha = 2$

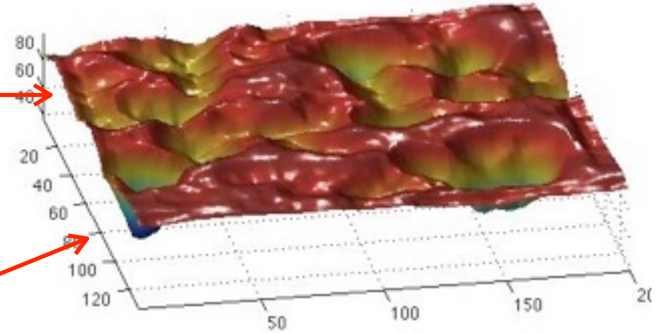
Cauchy walk

Gaussian walk

Gaze-shift as a constrained random walk
// Boccignone & Ferraro (Physica A, 2004)



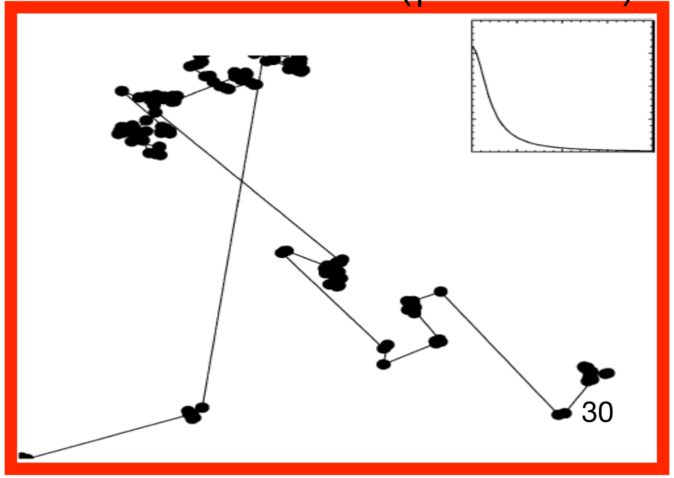
Deterministic component
(potential)



$$\mathbf{r}_{new}(t) = \mathbf{r}(t) - \nabla V + \boldsymbol{\eta}$$

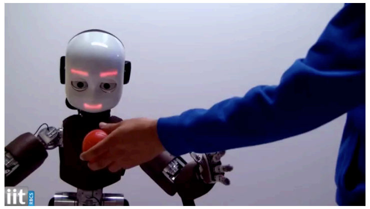
Random component
(potential)

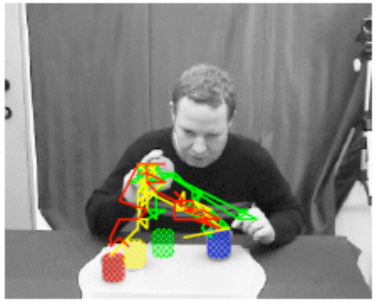
Constrained
Levy Search



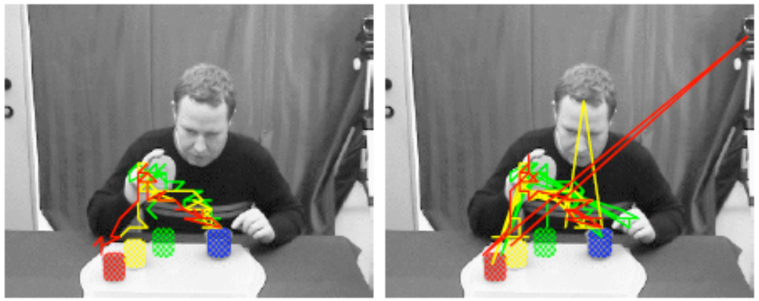
Gaze-shift as a constrained random walk // Boccignone & Ferraro (Physica A, 2004)

Successful Applications: Robot Action Learning for the iCub (Nagai. 2009)



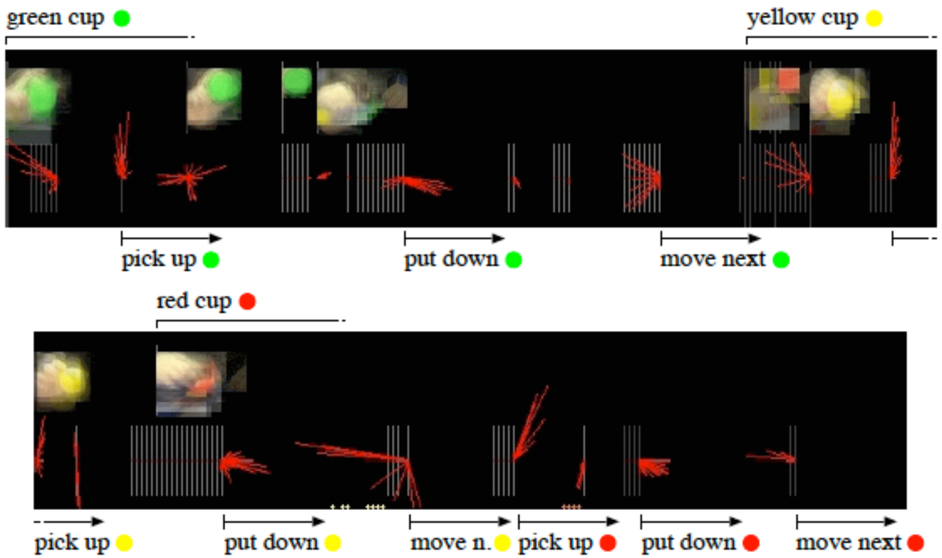


(a) Stochastic algorithm *with* retinal filter



(b) Winner-take-all algorithm *with* retinal filter (c) Winner-take-all algorithm *without* retinal filter

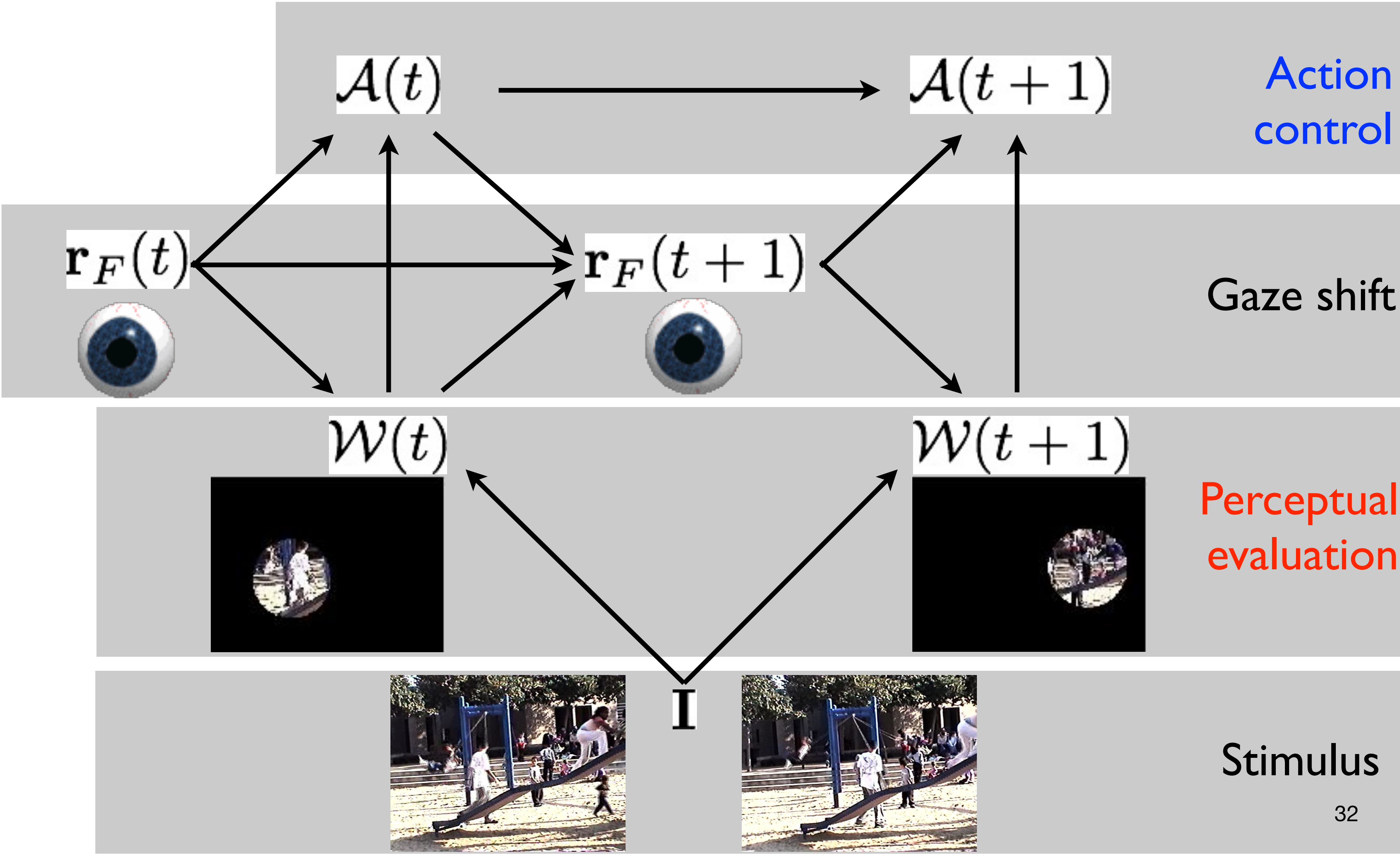
Fig. 7. Transition of attention of proposed model (a) and two comparative models, (b) and (c). The line color corresponds to the cup color.



(a) Object and motion chunks created through task demonstration

(b) Action map for moving green, yellow, and red cups

Computational models of eye guidance
//gaze shifts as actions

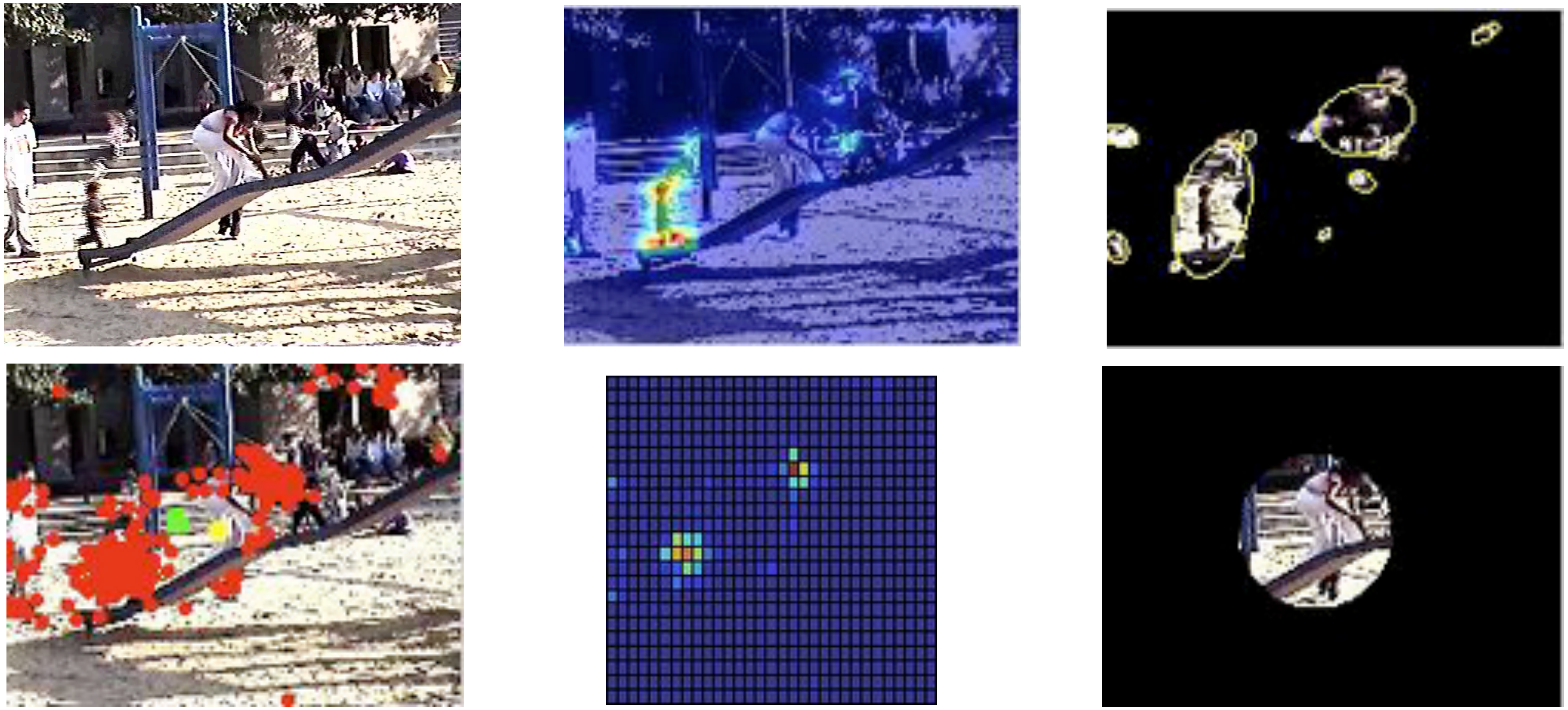


Computational models of eye guidance
//gaze shifts as actions

IEEE TRANSACTIONS ON CYBERNETICS, VOL. 44, NO. 2, FEBRUARY 2014

Ecological Sampling of Gaze Shifts

Giuseppe Boccignone and Mario Ferraro



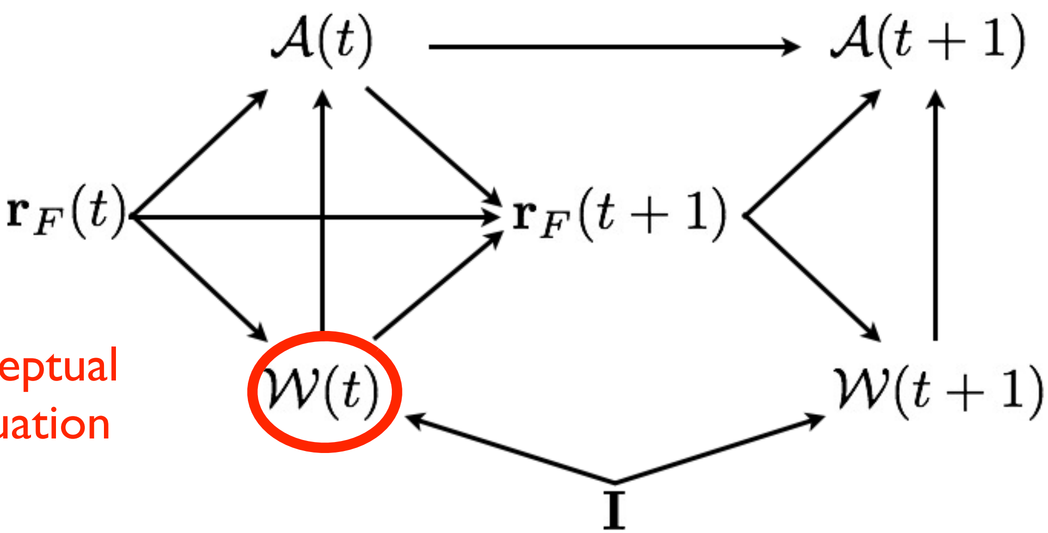
Matlab simulation code:

https://github.com/phuselab/EcoSampling

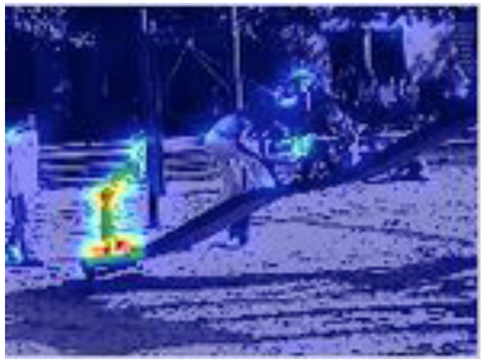
Ecological sampling of gaze shifts
//sampling the landscape

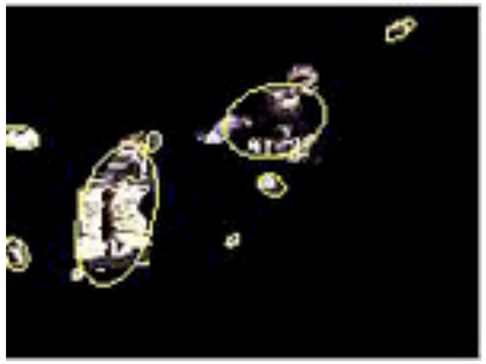
- Sampling the natural habitat

$$\mathcal{W}^*(t) \sim P(\mathcal{W}(t) | \mathbf{r}_F(t), \mathbf{F}(t), \mathbf{I}(t))$$



Perceptual evaluation

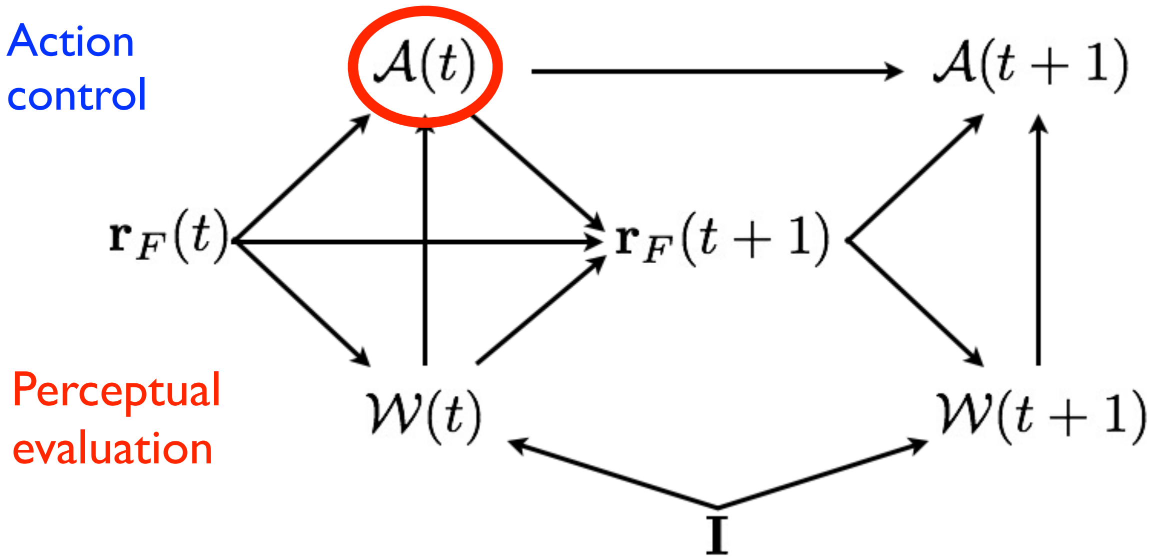


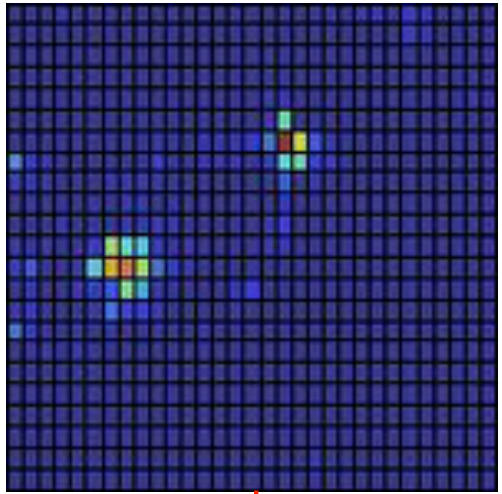


Ecological sampling of gaze shifts //sampling the oculomotor action

- Sampling the appropriate motor behavior

$$\mathcal{A}(t)^* \sim P(\mathcal{A}(t)|\mathcal{A}(t-1), \mathcal{W}^*(t))$$





$$\boxed{\mathcal{C}(t) = \Delta(t) \cdot \Omega(t)}$$

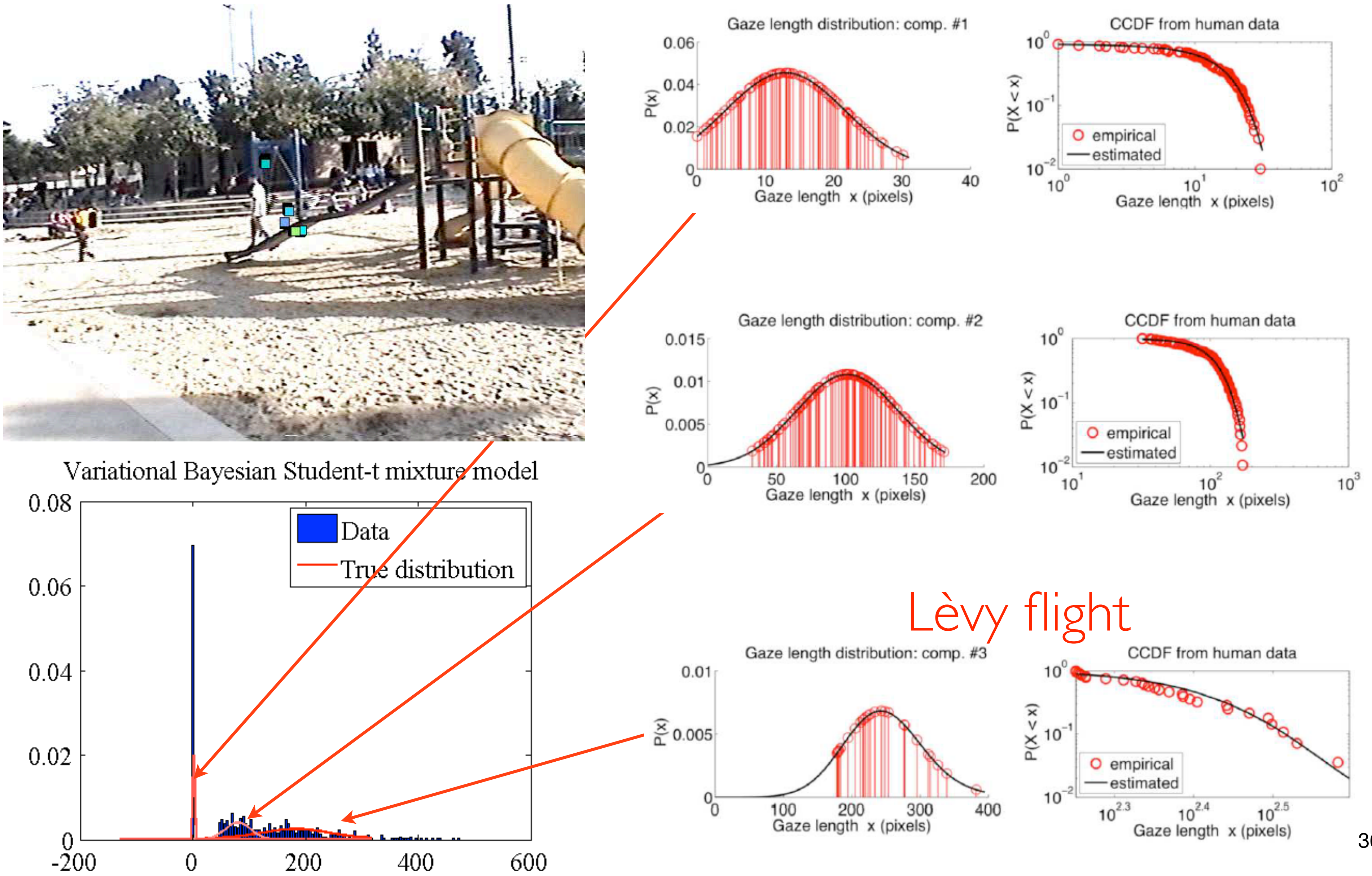
oculomotor prior

$$\pi^*(t) \sim Dir(\pi|\nu(O(t)))$$

$$z^*(t) \sim Mult(z(t)|\pi^*(t))$$

oculomotor choice

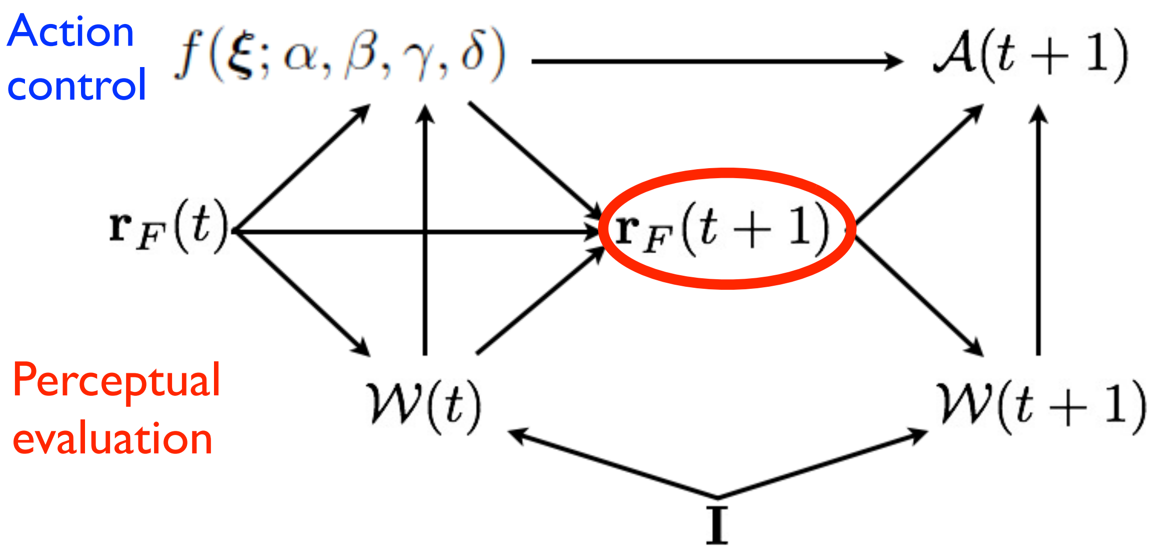
Ecological sampling of gaze shifts
//oculomotor actions: fixate, pursuit, saccade



Ecological sampling of gaze shifts //sampling the oculomotor action

- Sampling where to look next

$$\mathbf{r}_F(t+1) \sim P(\mathbf{r}_F(t+1) | \mathcal{A}(t)^*, \mathcal{W}^*(t), \mathbf{r}_F(t))$$



$$\boldsymbol{\xi} \sim f(\boldsymbol{\xi}; \alpha, \beta, \gamma, \delta)$$

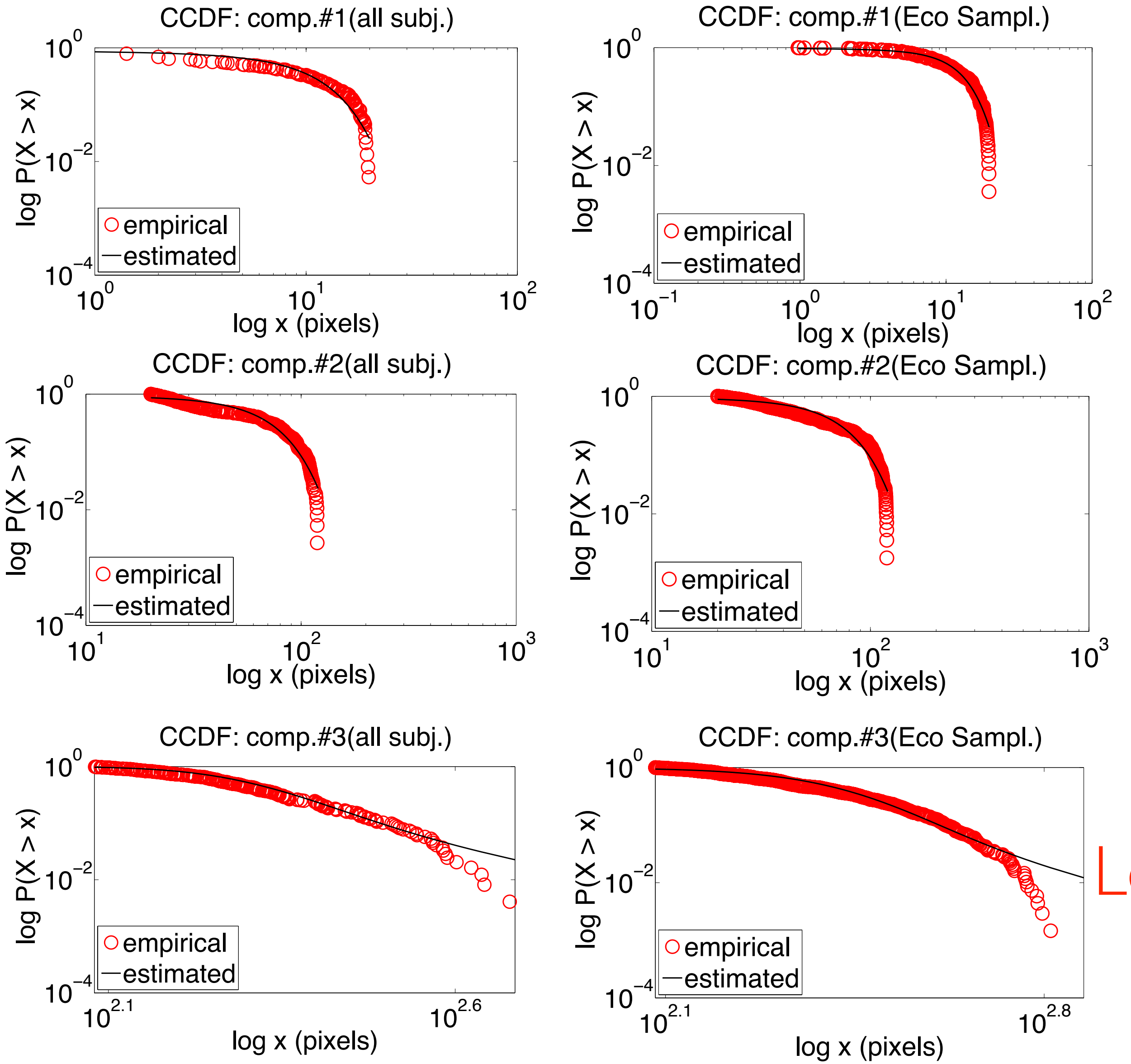
Alpha-stable (Lèvy flight)

$$\boxed{\mathbf{r}_F(t_{n+1}) \approx \mathbf{r}_F(t_n) - \sum_{p=1}^{N_V} (\mathbf{r}_F(t_n) - \mathbf{r}_p(t_n))\tau + \gamma_k \mathbb{I} \tau^{1/\alpha_k} \boldsymbol{\xi}_k.}$$

Ecological sampling of gaze shifts
//some results…

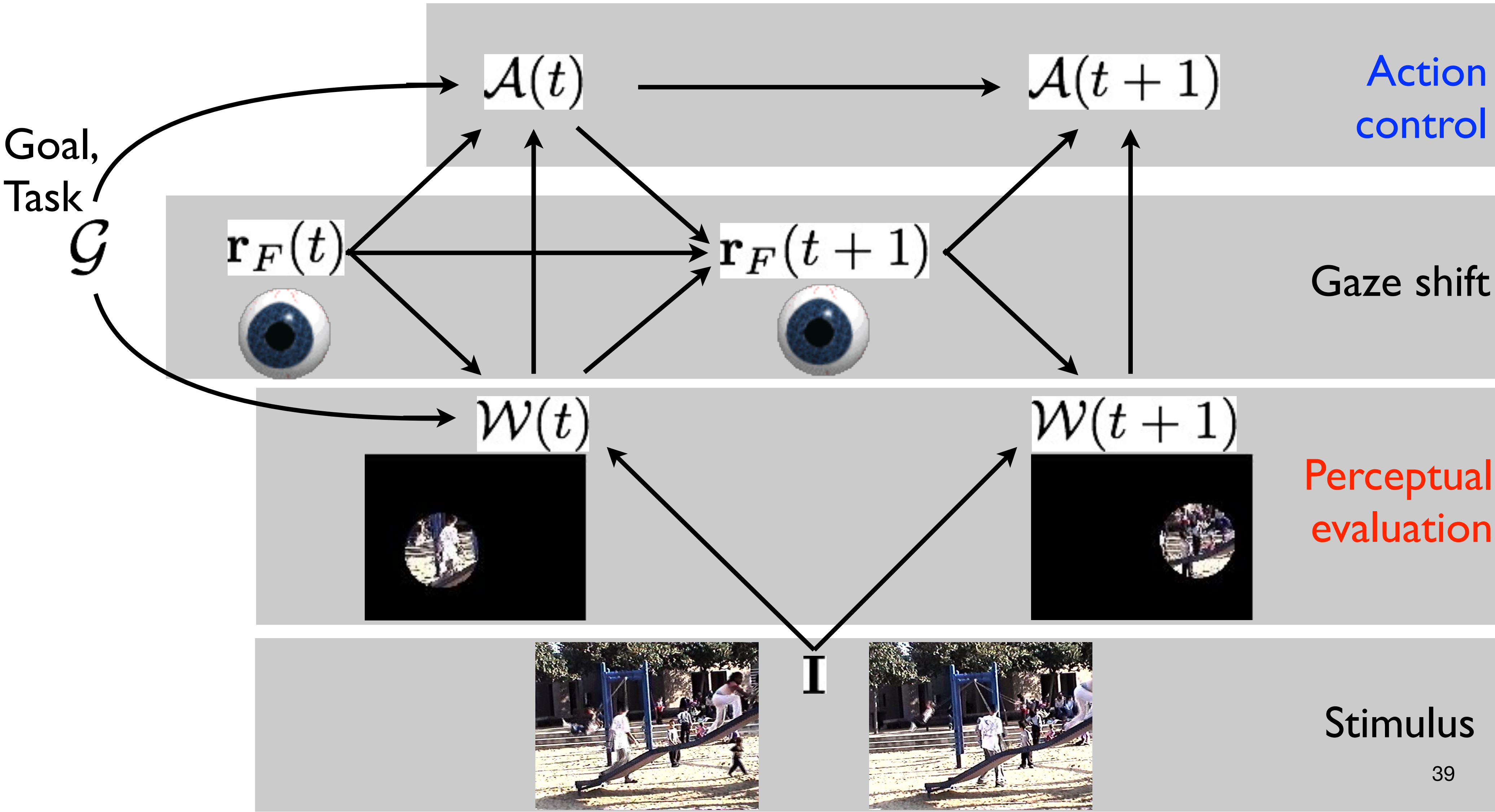
Human

Model

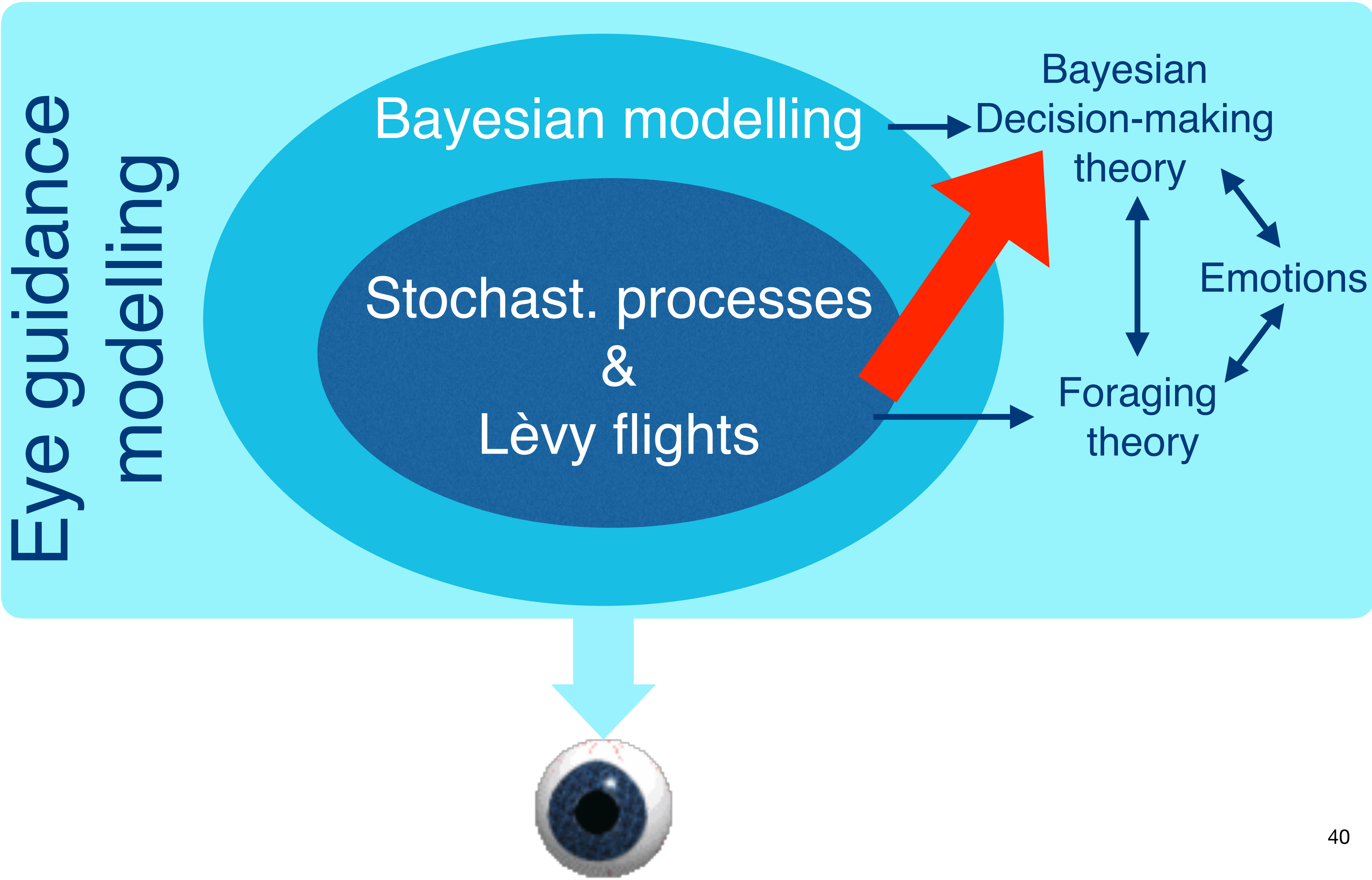


Lèvy flight

Computational models of purposive eye guidance // decisions on actions: considering task / goals

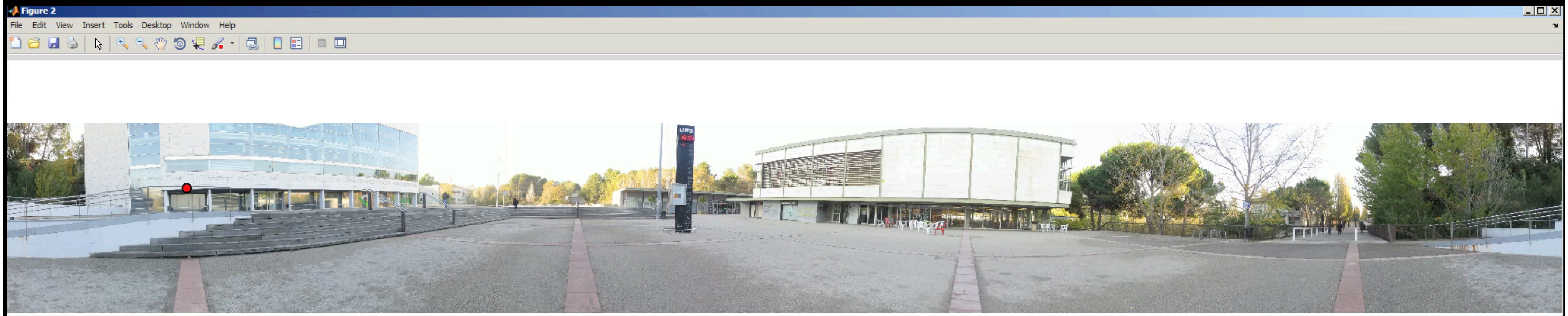


Some key points of this talk



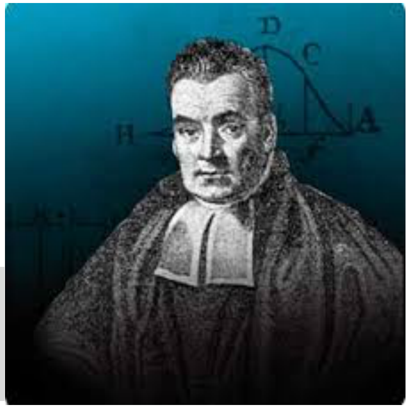
Computational models of purposive eye guidance // Considering task / goals

mobile
eye-tracking



homography derived panoramic image

Computational models of purposive eye guidance // Bayesian Decision Theory



Goal, Task $\mathcal{G}$

states, internal models

$$\mathbb{P}(\mathbf{x}|z_{0:t}, \mathrm{M}) \propto \mathbb{P}(z_{0:t}|\mathbf{x}, \mathrm{M})\mathbb{P}(\mathbf{x}|\mathrm{M})$$

stimulus

value, reward

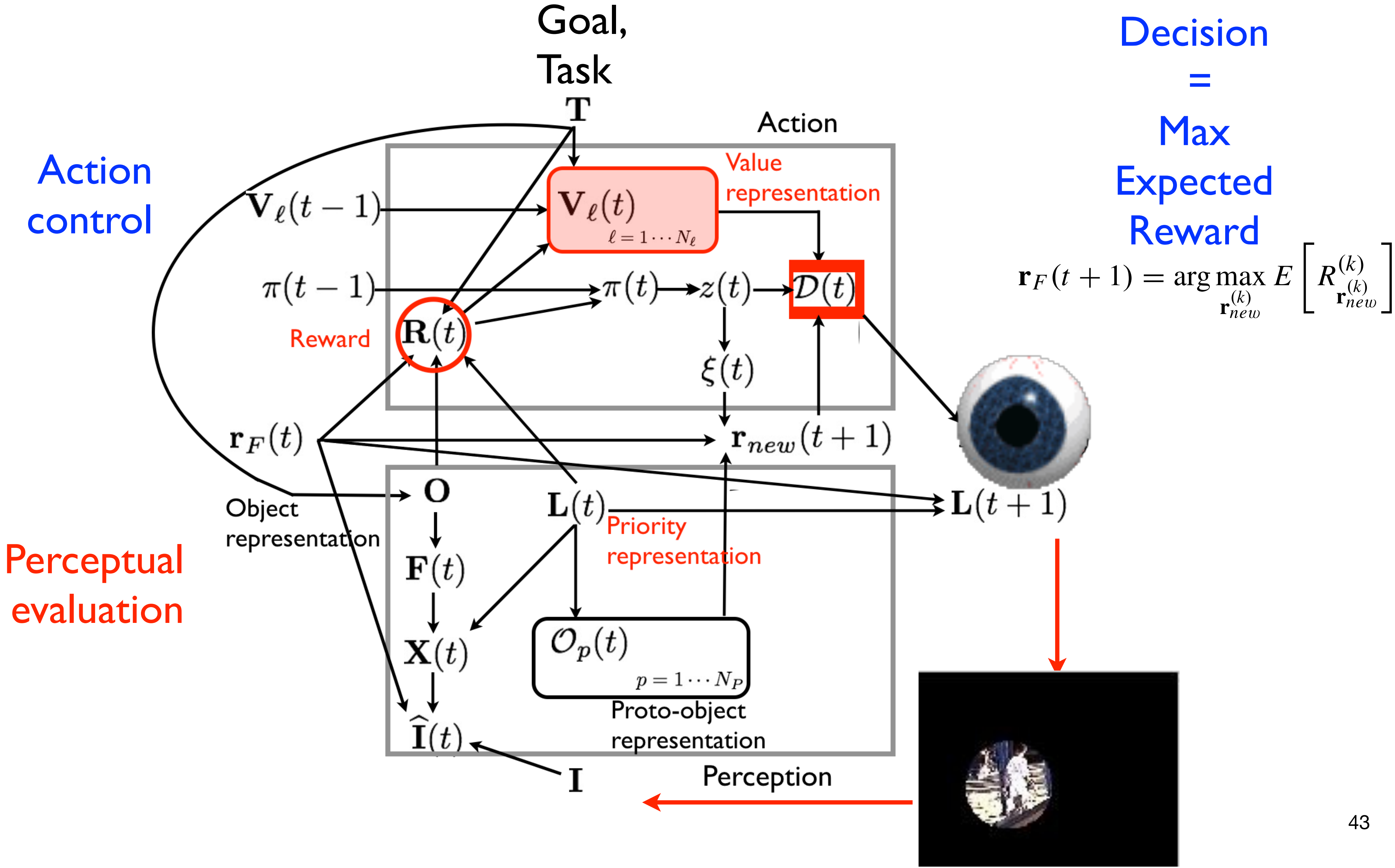
$$\mathcal{Q}_{\mathrm{now}}(a, \mathbb{P}(\mathbf{x}|z_{0:t}, \mathrm{M})) = \sum_{\mathbf{x}} \mathcal{R}(a, \mathbf{x})\mathbb{P}(\mathbf{x}|z_{0:t}, \mathrm{M})$$

action

Theoretical perspectives on active sensing
Scott Cheng-Hsin Yang[1], Daniel M Wolpert[1,3] and Máté Lengyel[1,2,3]

Current Opinion in Behavioral Sciences 2016, **11**:100–108

Computational models of purposive eye guidance // Considering task / goals



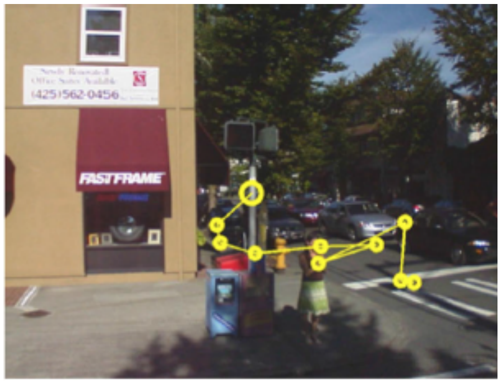
$$\mathbf{r}_F(t+1) = \arg\max_{\mathbf{r}_{new}^{(k)}} E\left[R^{(k)}_{\mathbf{r}_{new}^{(k)}} \right]$$

Computational models of purposive eye guidance // Considering task / goals

High level
of representation

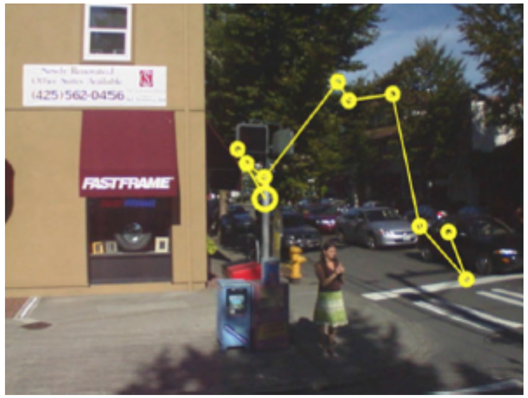
Low level
of representation

Value for text
(search task)



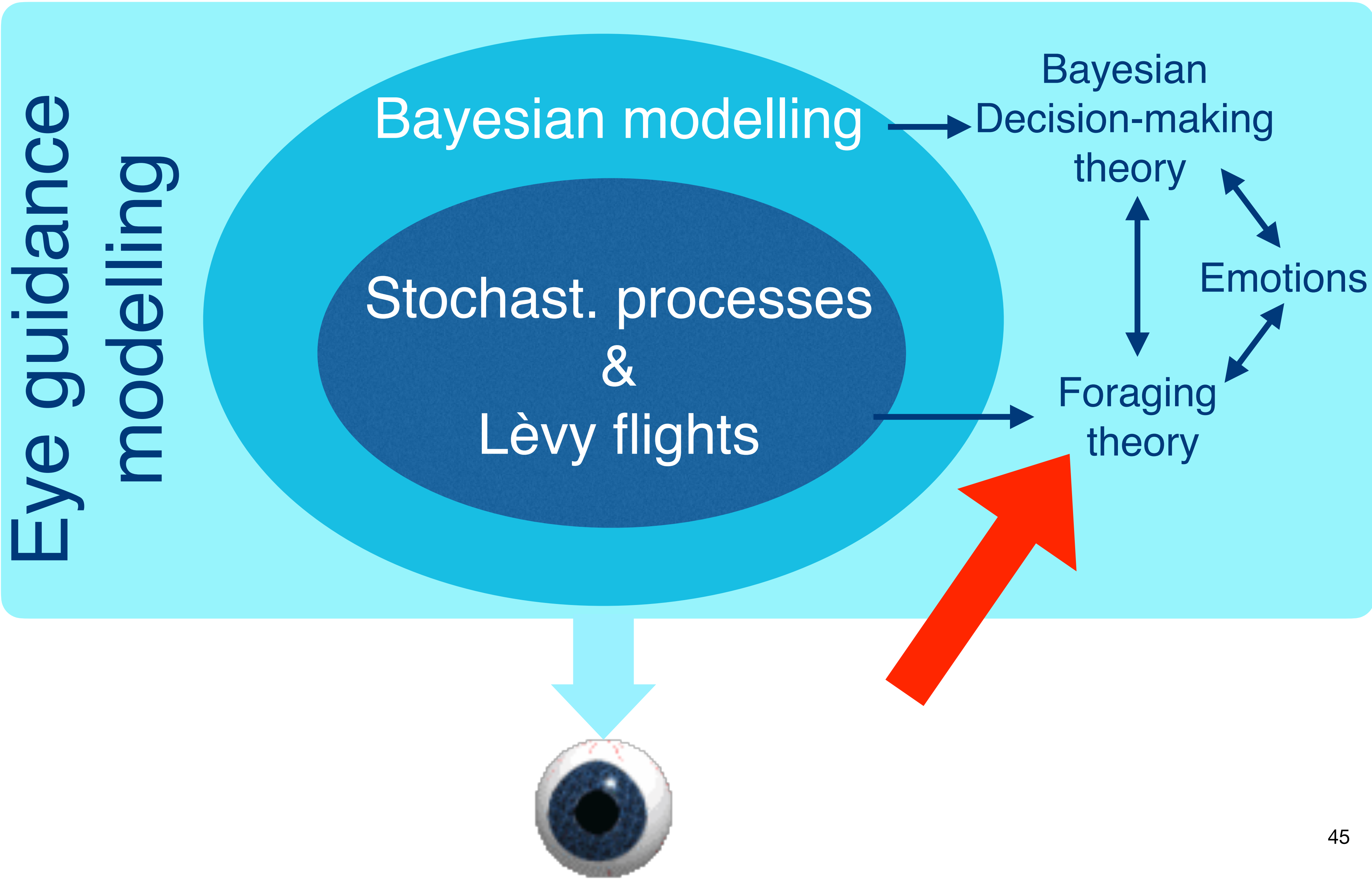
Value for people
(search task)

Object-based perception
(free viewing task)

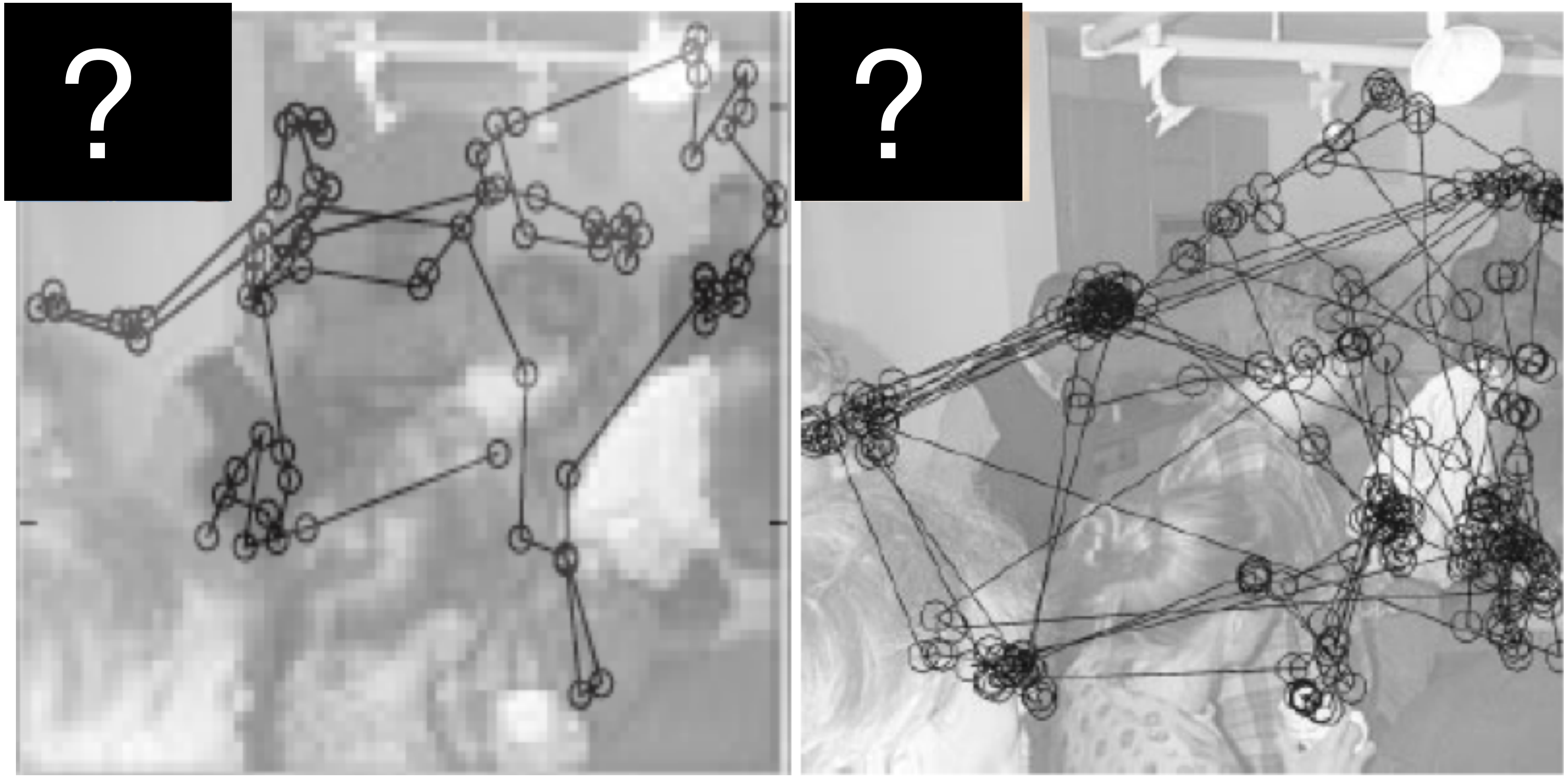


Early salience based
perception
(free viewing task)

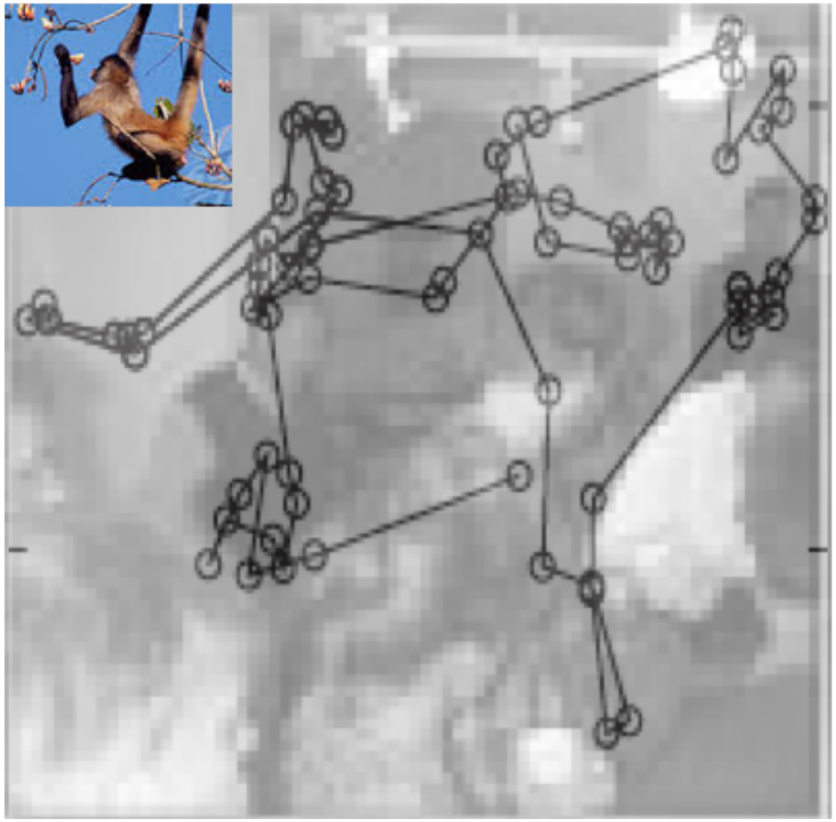
The foraging perspective

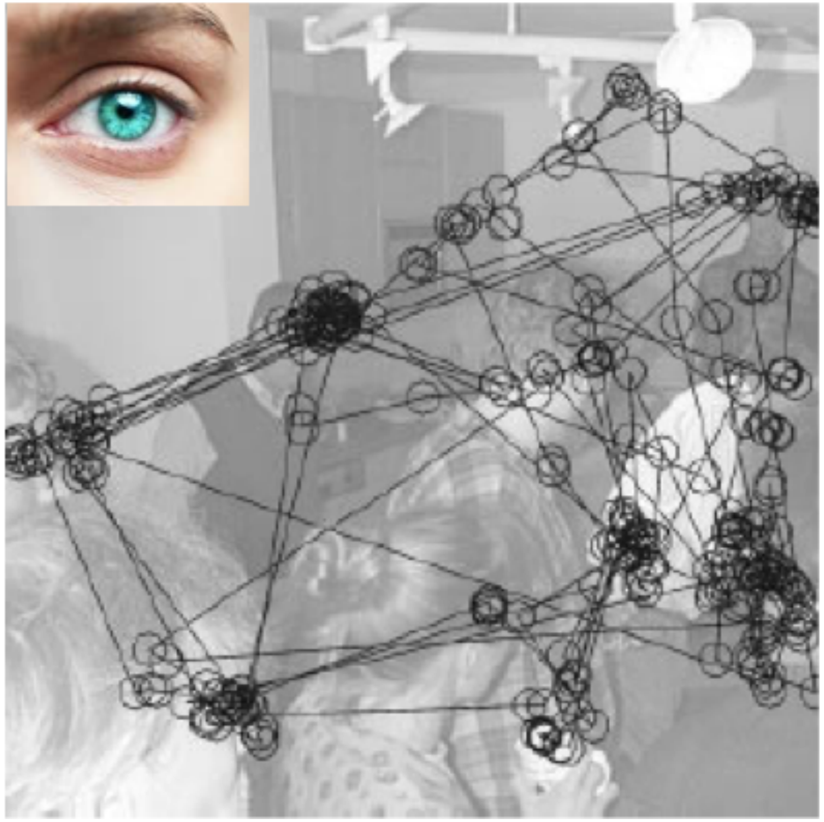


Back to the random walks….

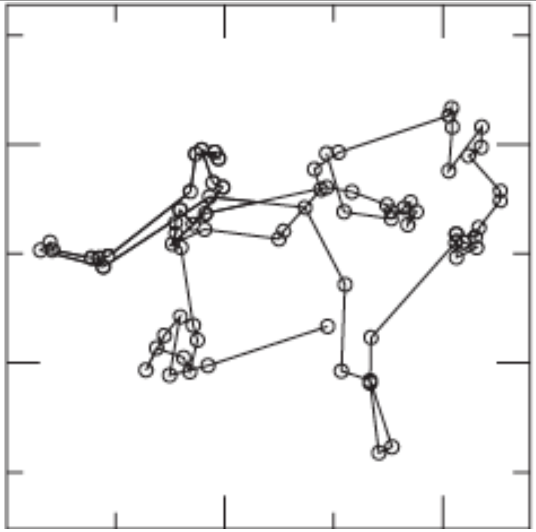


The foraging perspective

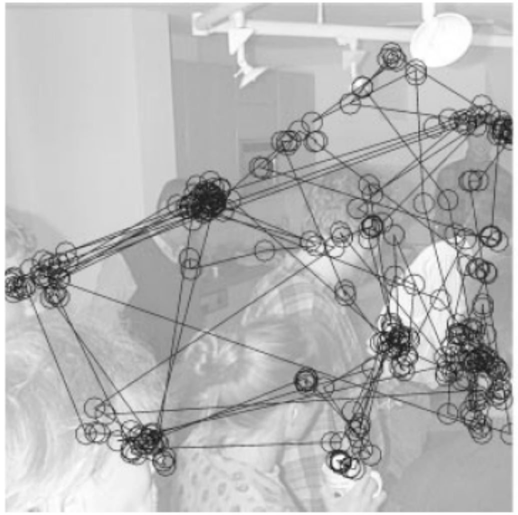




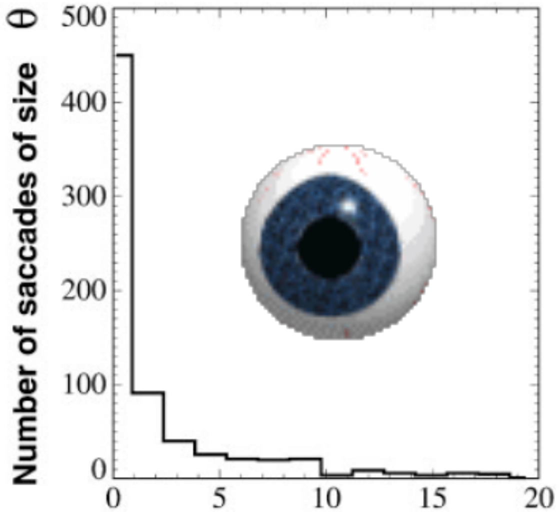
The foraging perspective



Foraging pattern of spider monkeys   in the Yucatan Peninsula



“Foraging pattern” of the eye



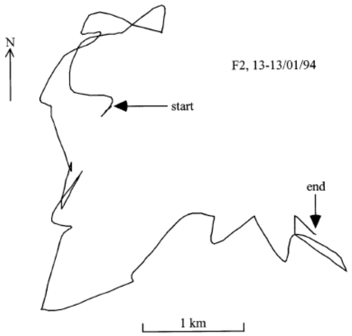
Long-tail distributions beyond the Central Limit Theorem

↕

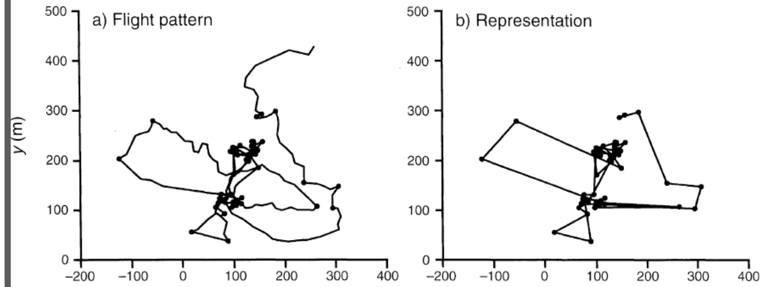
Anomalous diffusion (Cauchy walk)

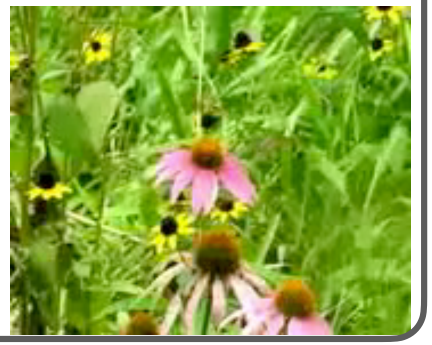
The foraging perspective

Jackal

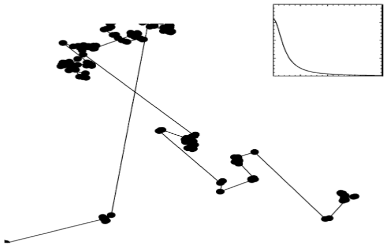


Bumblebee

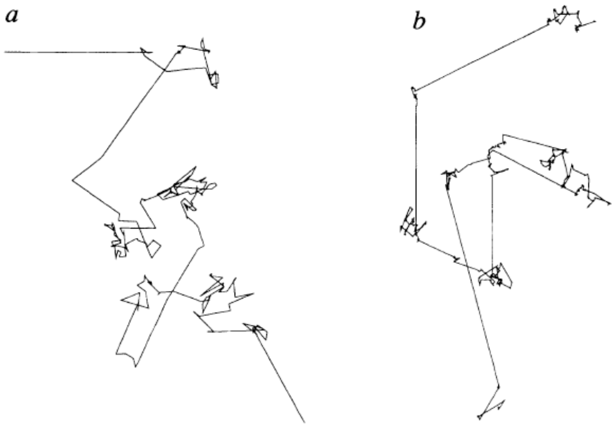




Eye



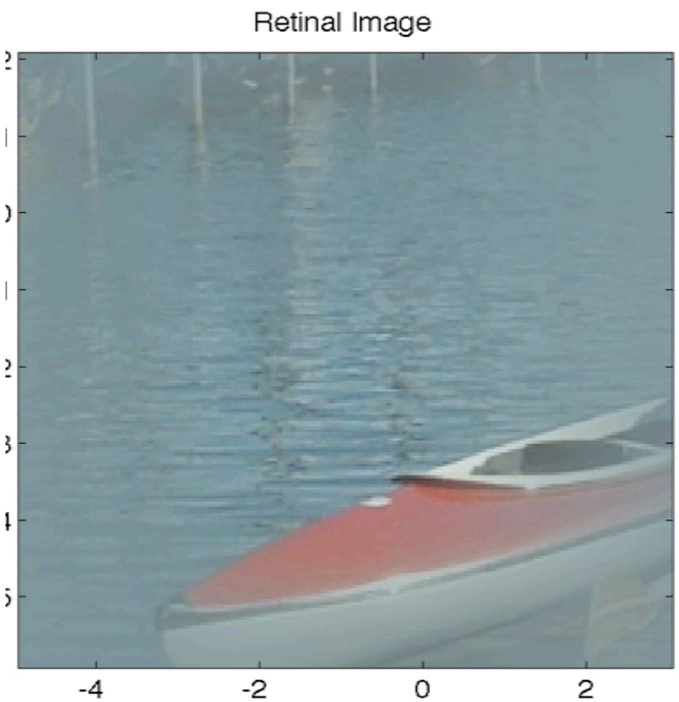
Albatros

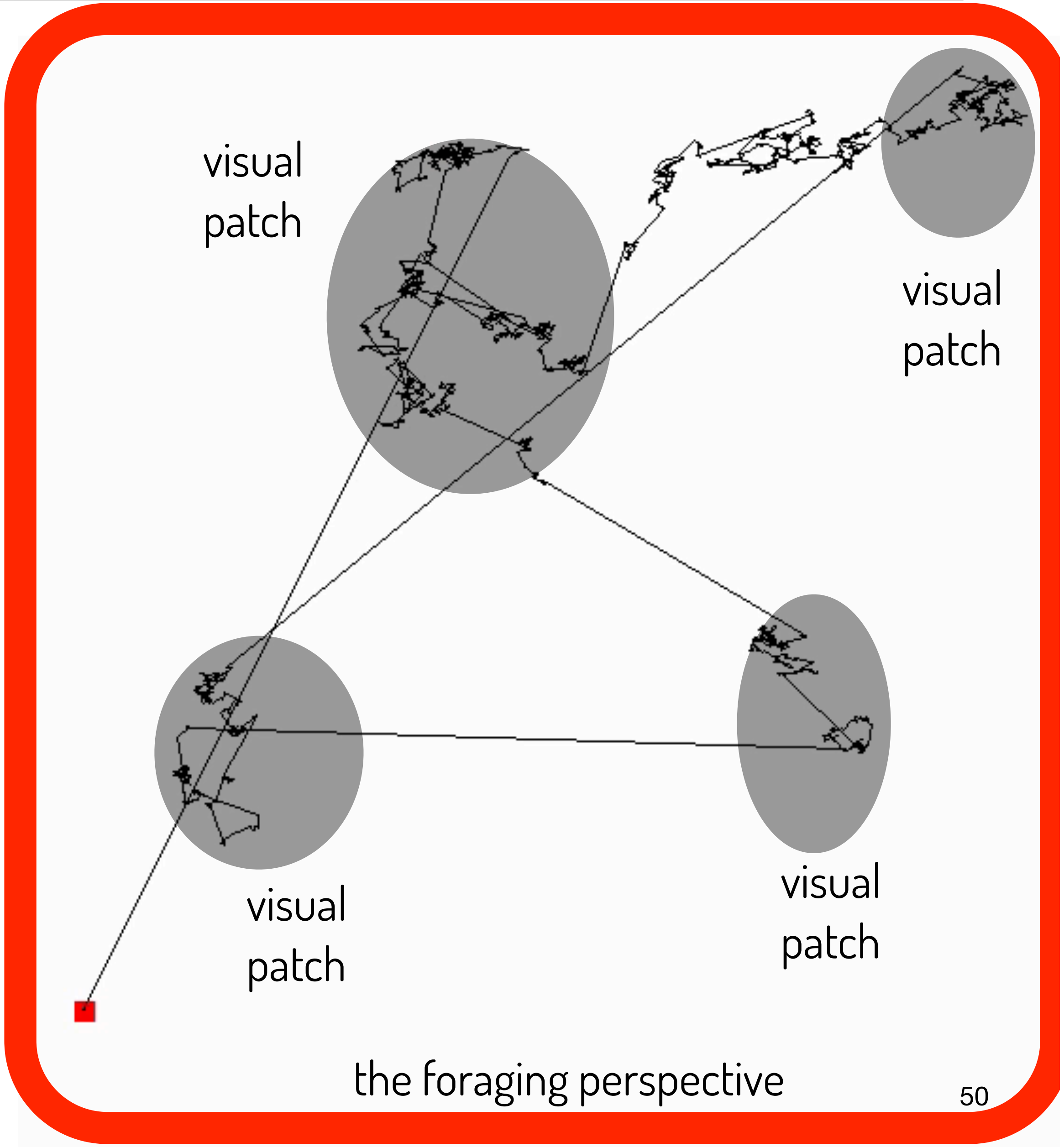


Anomalous diffusion

Lèvy Flights / Walks

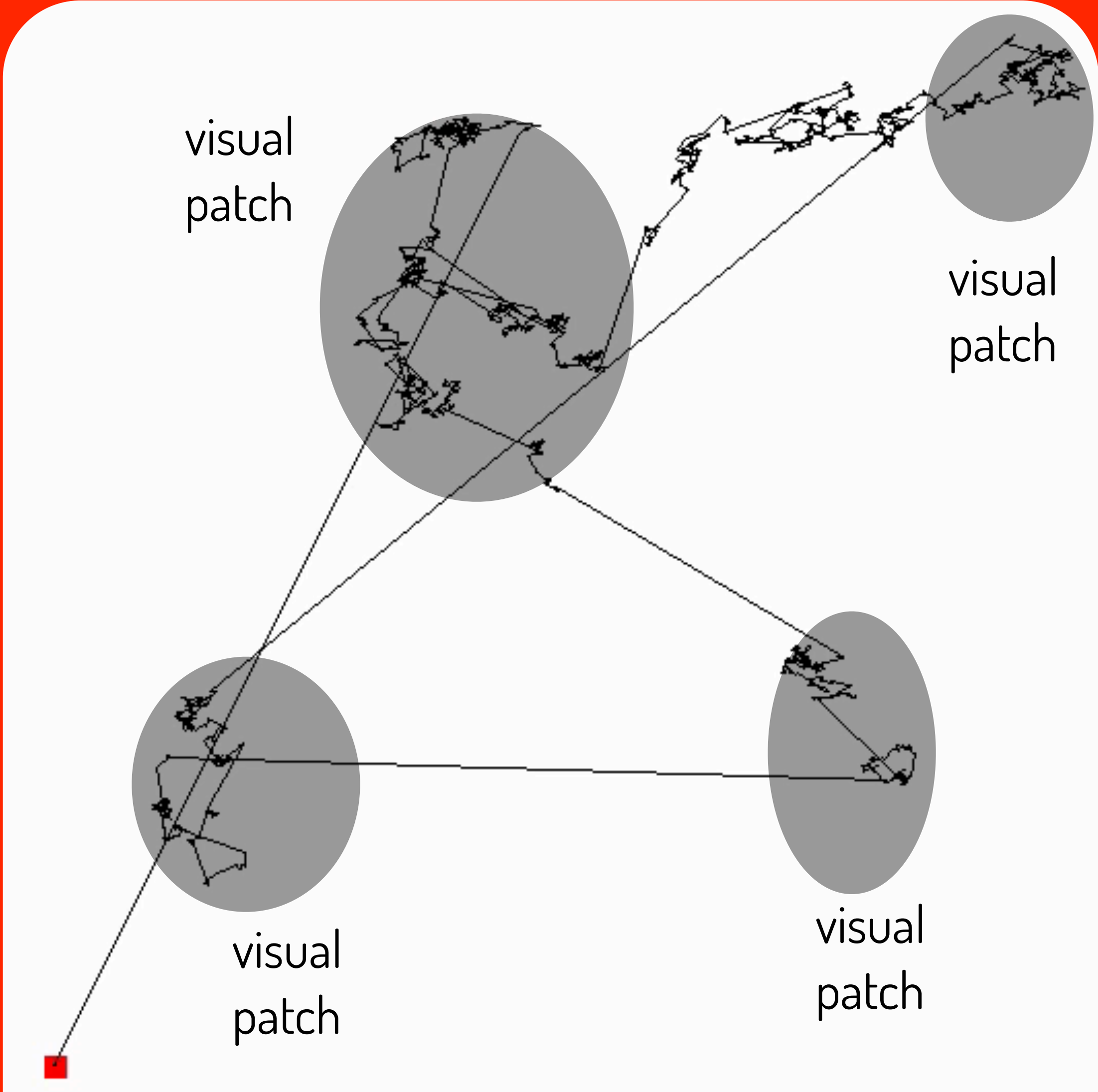
The foraging perspective





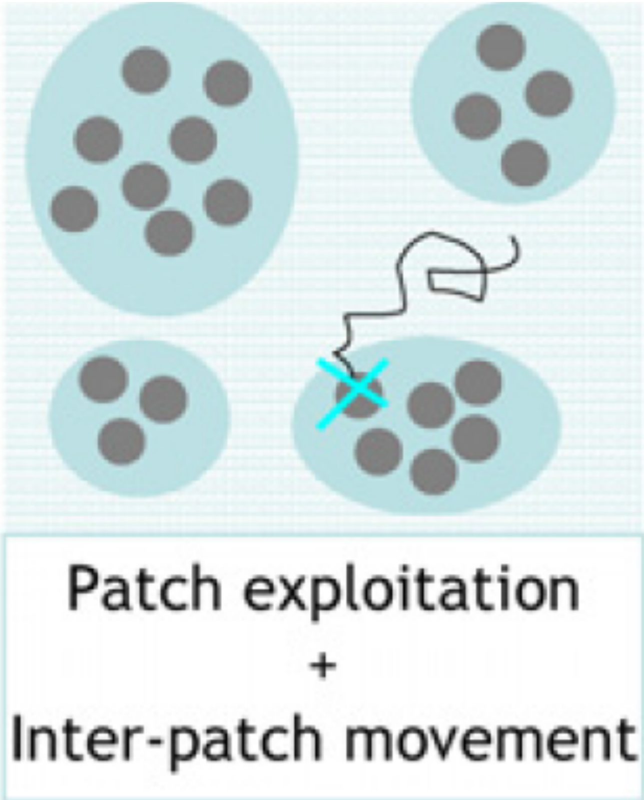
The foraging hypothesis

What was once
foraging in a physical space for tangible resources
became, over evolutionary time, foraging in cognitive space for information related to those resources

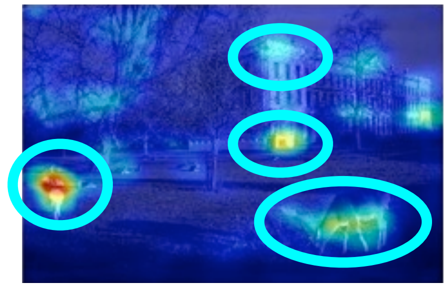


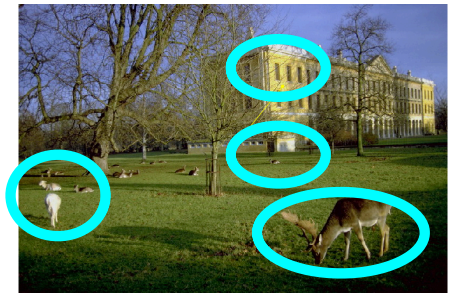
the foraging perspective

The foraging perspective
//optimal foraging theory

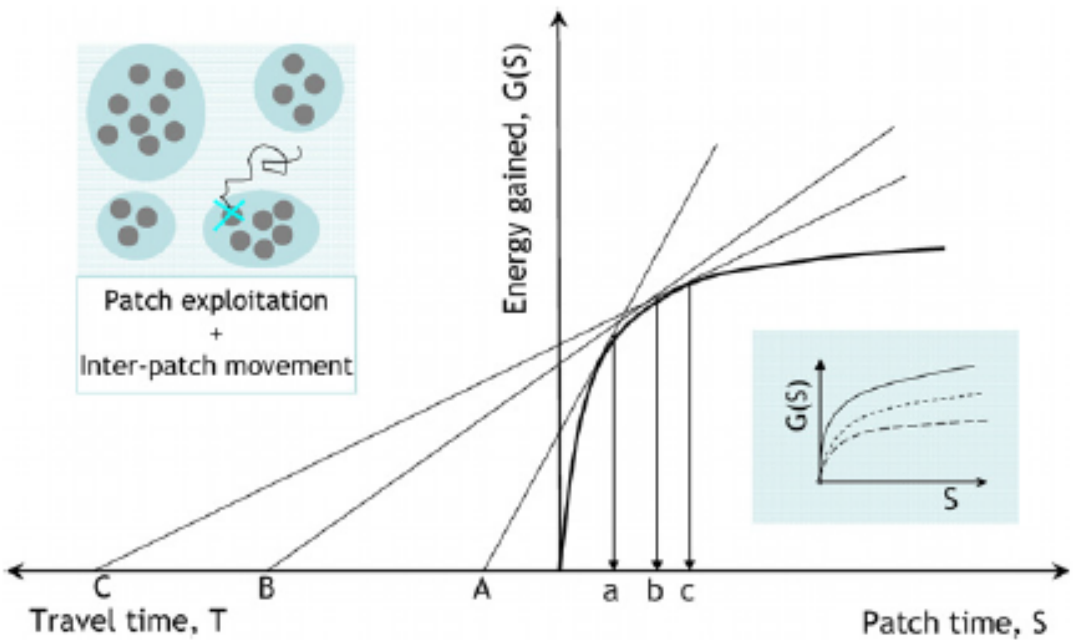


- What prey to take (optimal diet choice)
- What patch type to search (optimal patch choice)
- When to leave a patch (optimal giving up or departure times, GUT)
- How to move between patches (optimal movements)

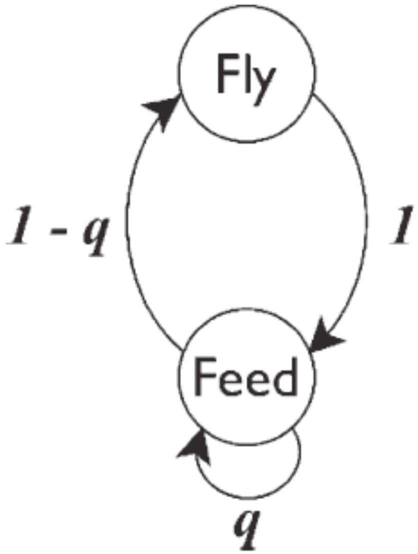




The foraging perspective //Charnov's Marginal Value Theorem



- When to leave a patch (optimal giving up or departure times, GUT)



How do we look at social scenes?

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

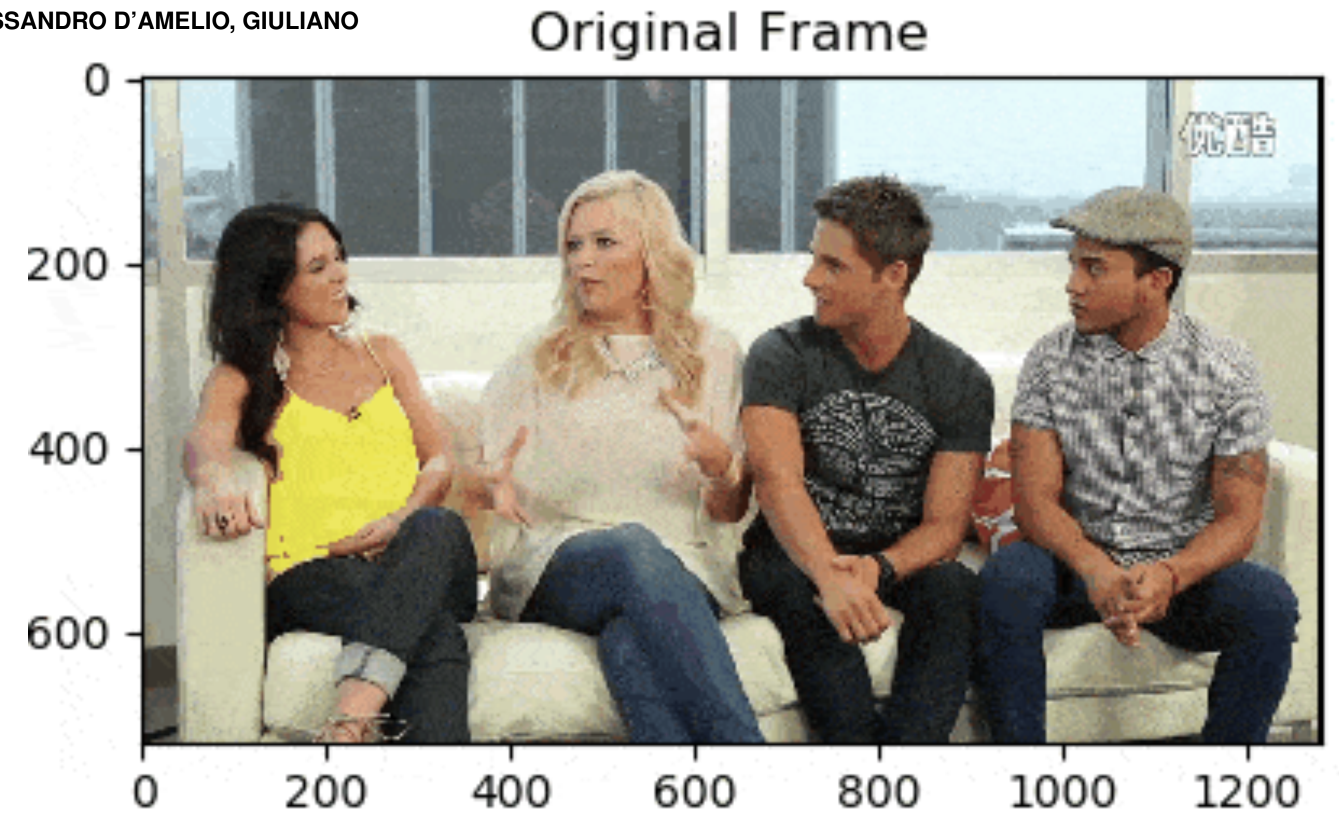
Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. (in press)

Digital Object Identifier 10.1109/ACCESS.2020.DOI

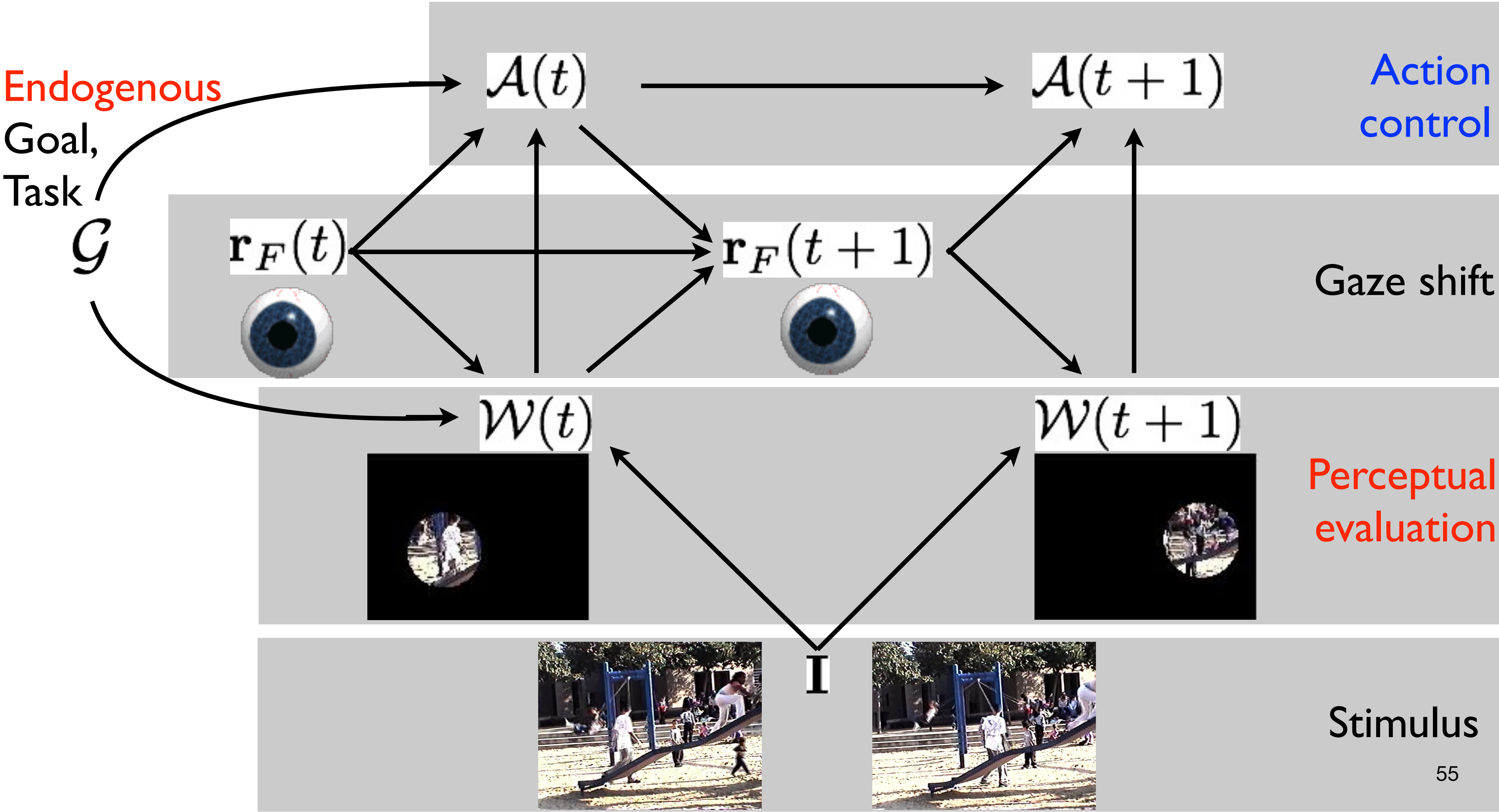
On gaze deployment to audio-visual cues of social interactions

GIUSEPPE BOCCIGNONE, VITTORIO CUCULO, ALESSANDRO D'AMELIO, GIULIANO GROSSI AND RAFFAELLA LANZAROTTI
PHuSe Lab - Dipartimento di Informatica, Universitá degli Studi di Milano,

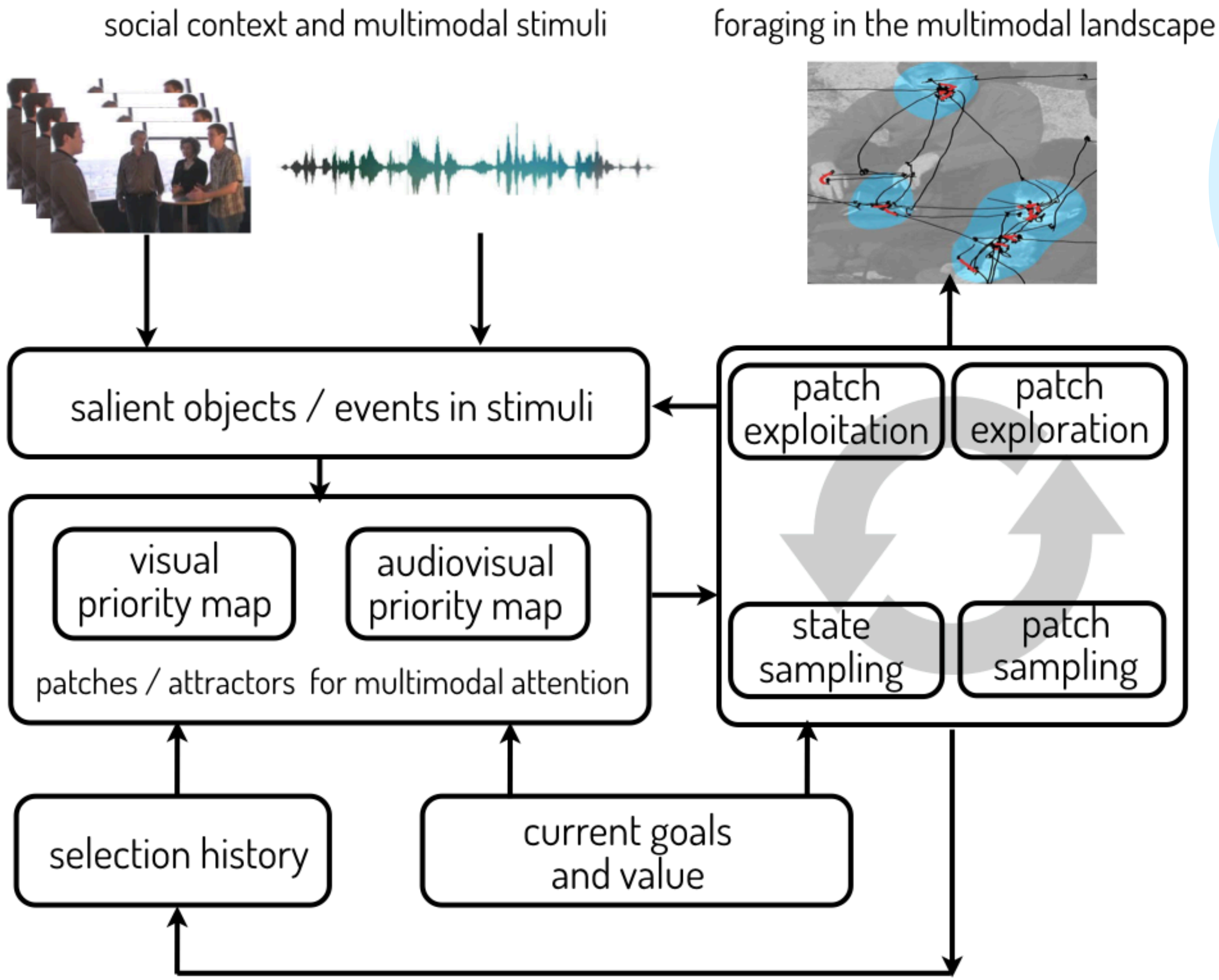
multimodal stimulus
(video + audio)



How do we look at social scenes?

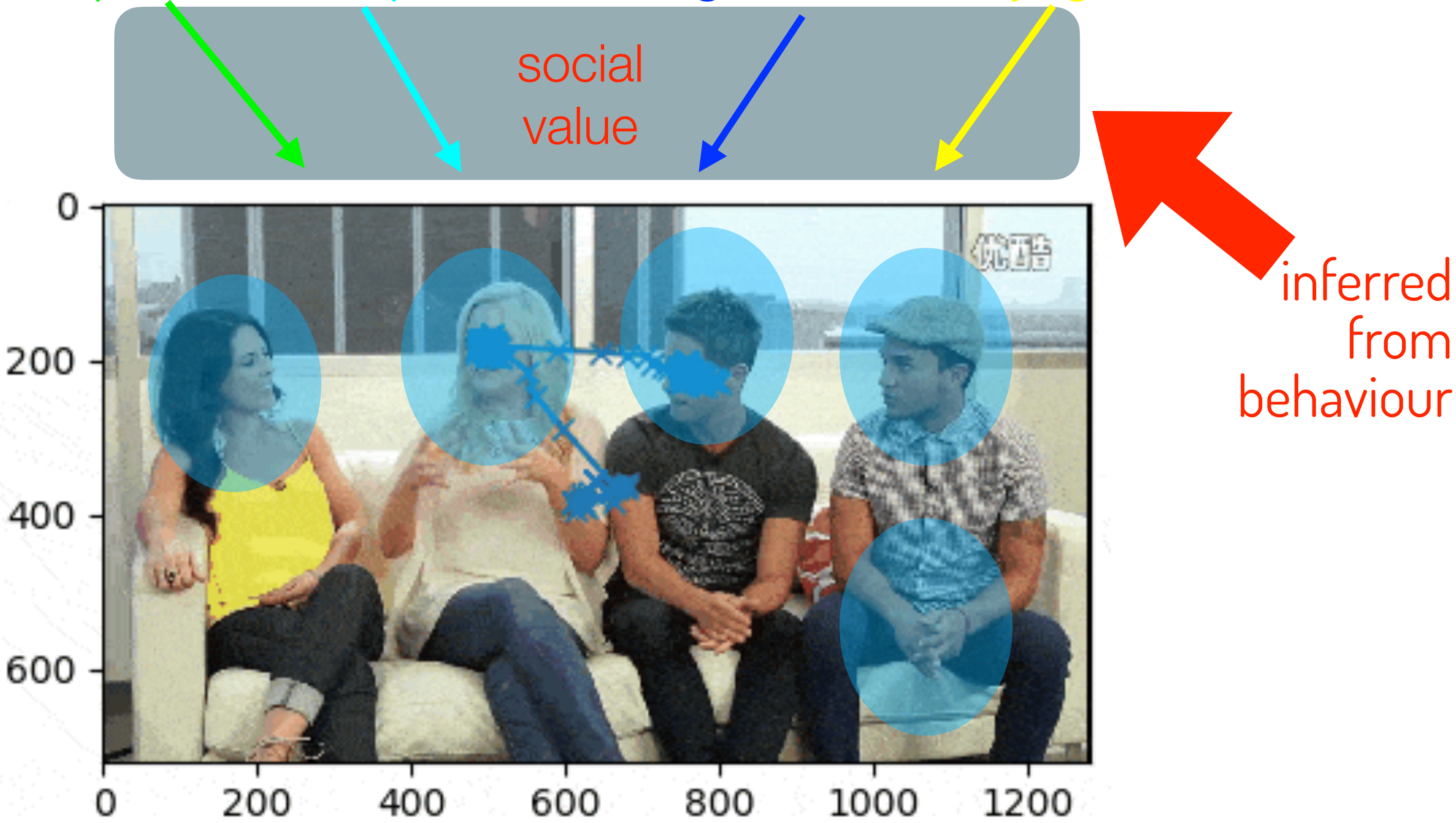


How do we look at social scenes?

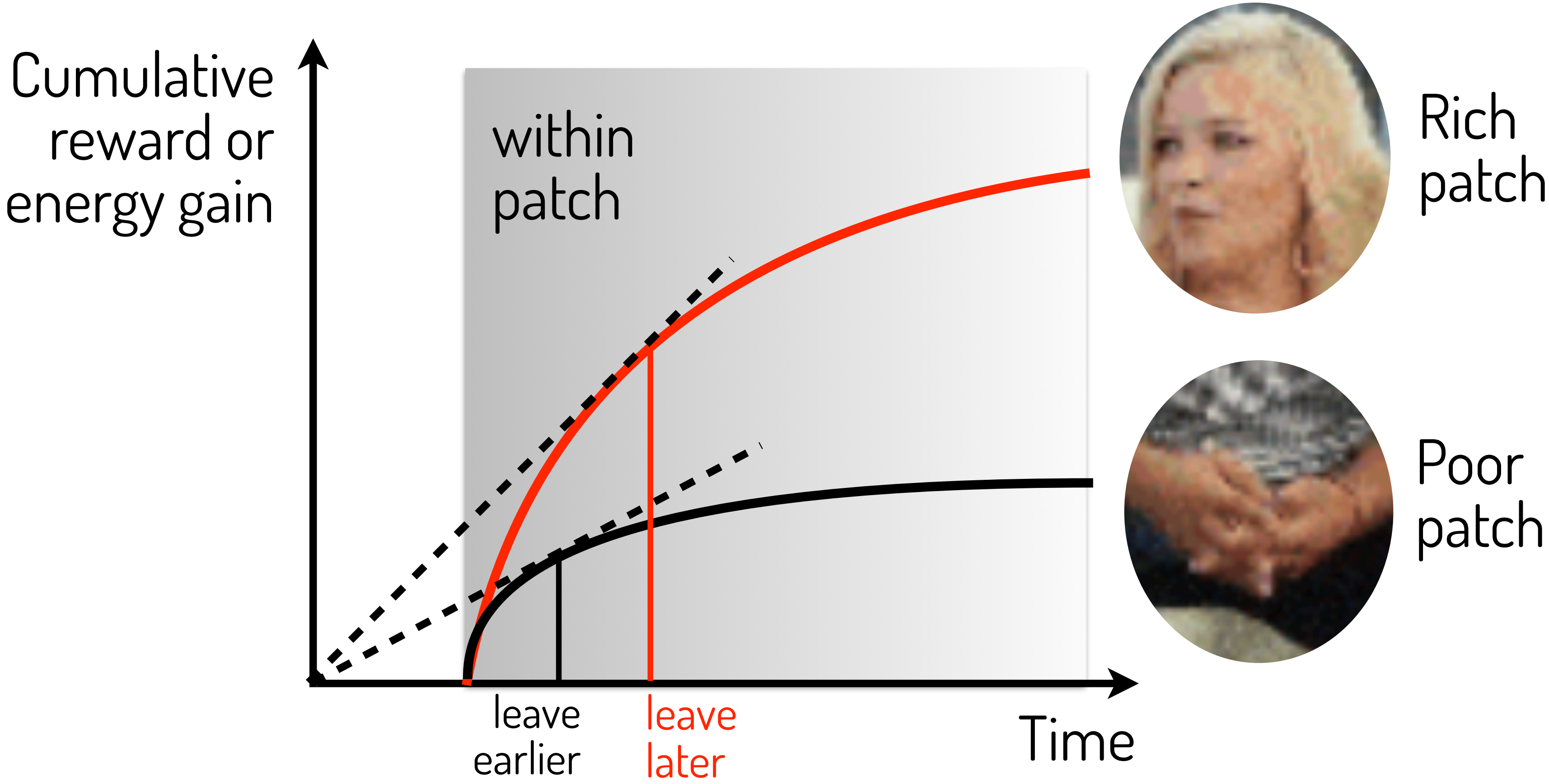


How do we look at social scenes?
//value-based patches

Patches = speaker, non speaker, hand (gestures), body (gestures), ...



How do we look at social scenes?
//giving-up time of a patch (stochastic Charnov)

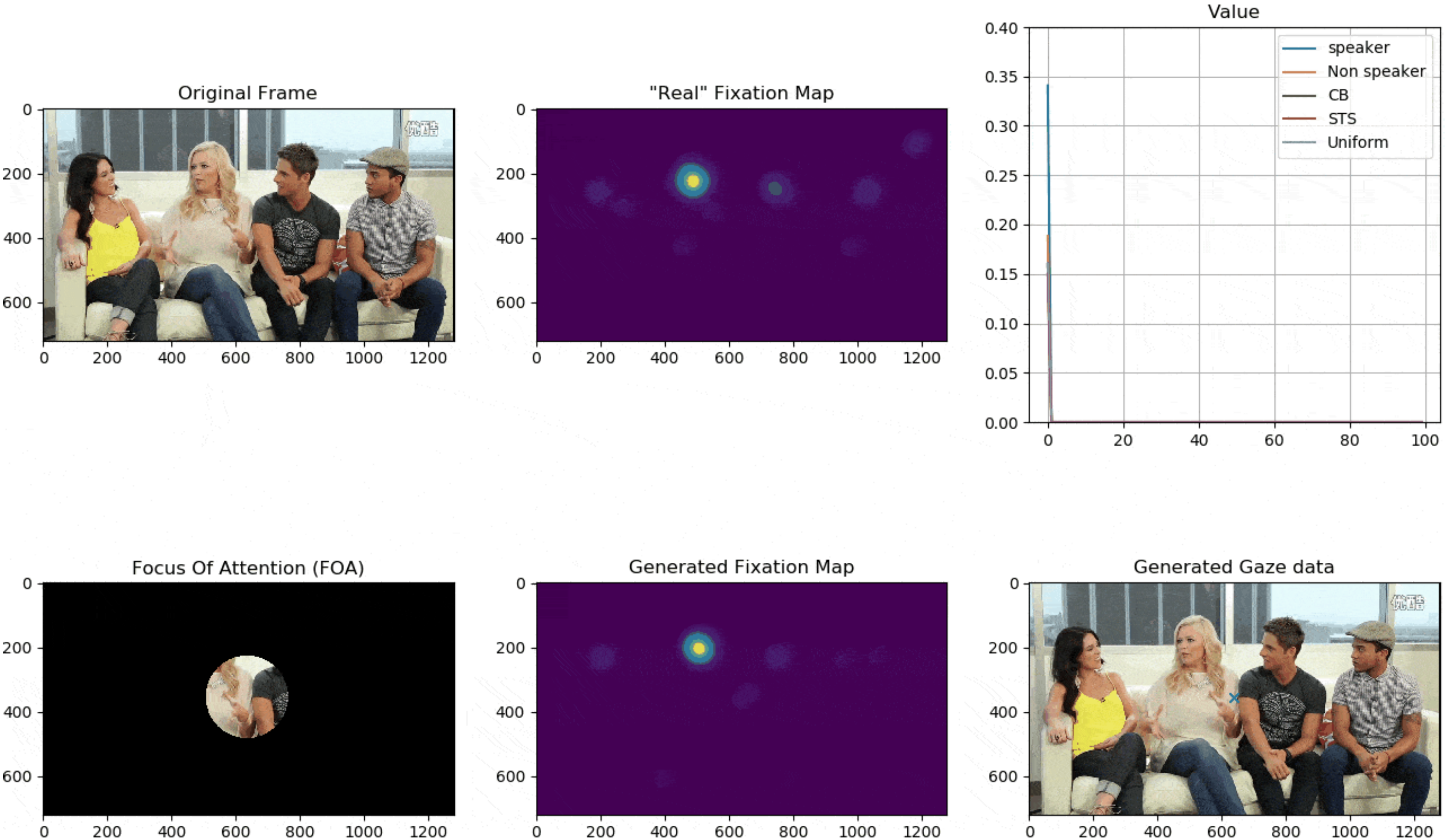


How do we look at social scenes?
//exploitation vs. exploration random walks

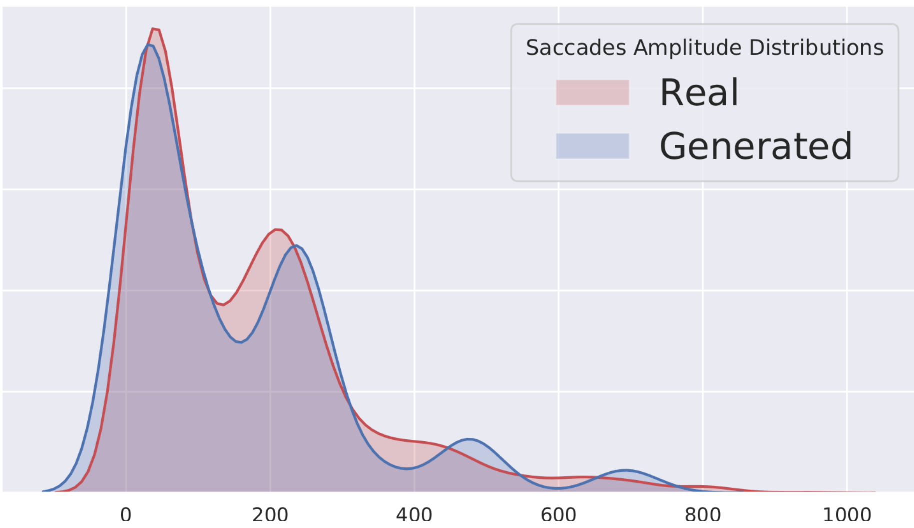
Exploitation within patch

Exploration between patches

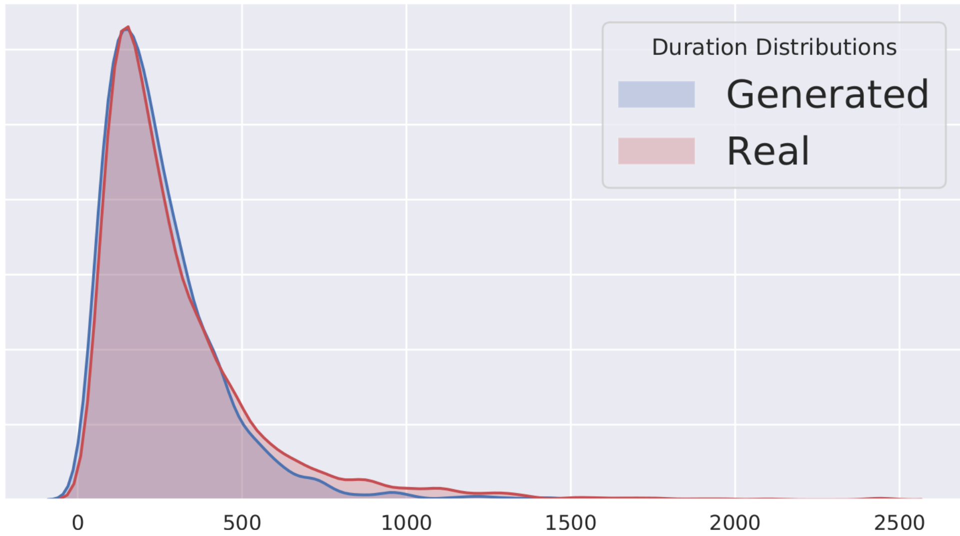
How do we look at social scenes?
//exploitation vs. exploration random walks



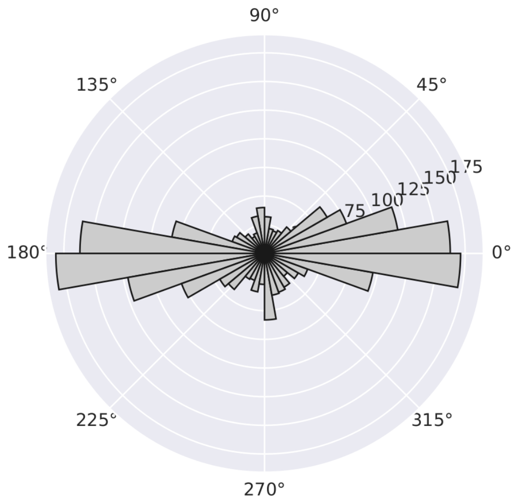
How do we look at social scenes?
//exploitation vs. exploration random walks



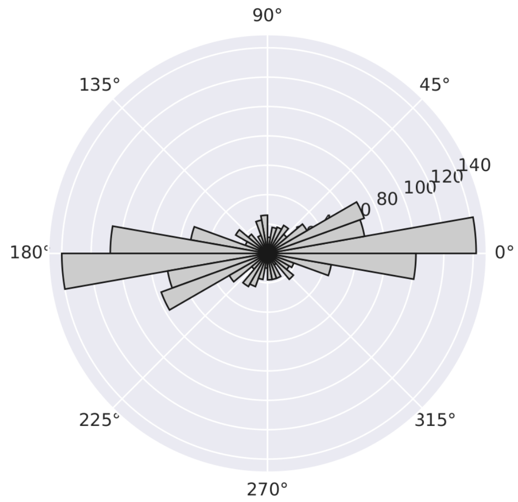
Shift amplitude



Fixation duration



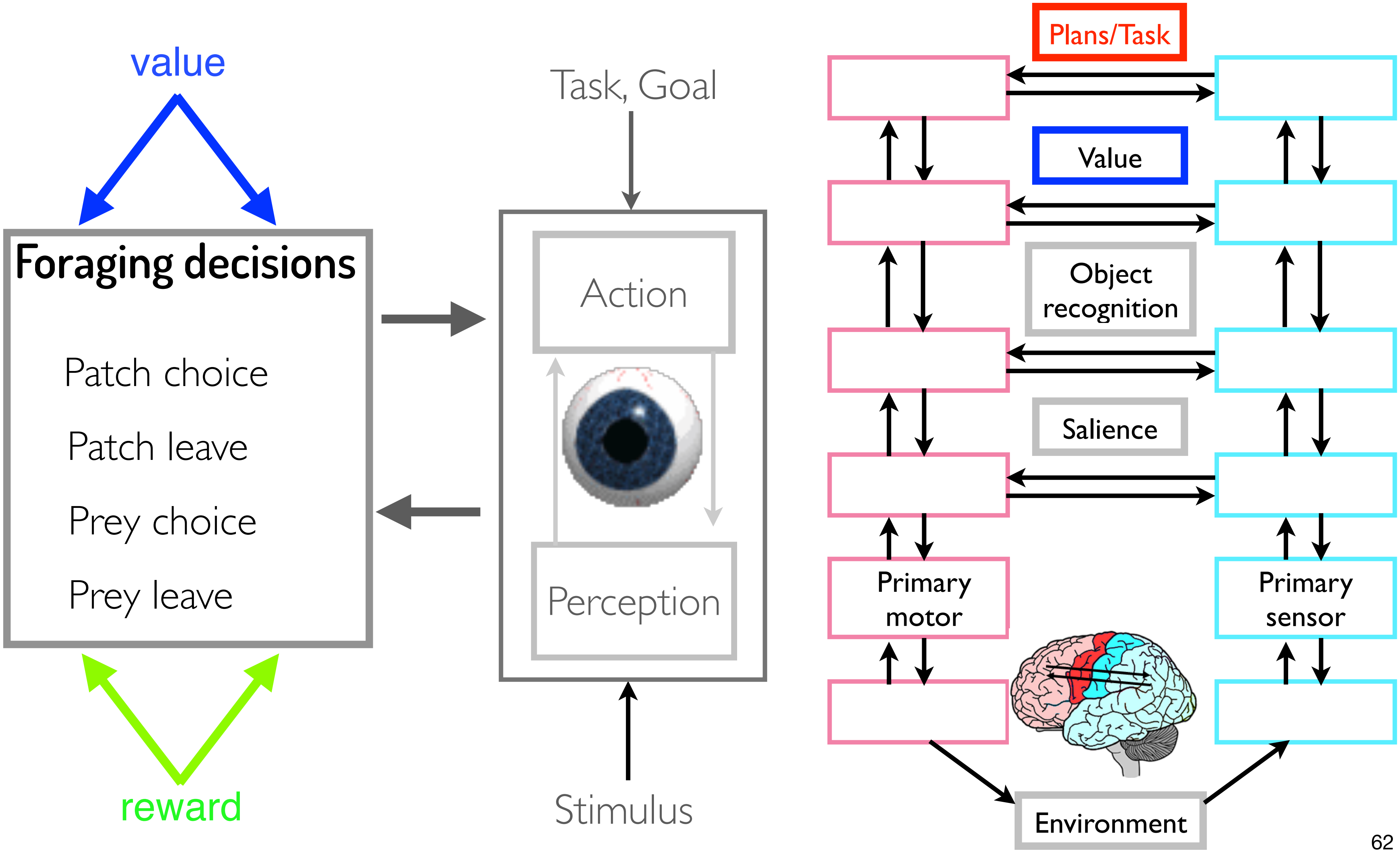
Real



Generated

Gaze direction

Computational models of purposive eye guidance // the value of “value” (and reward)



At the heart of purposive eye guidance //the dopamine hypothesis (Hills)

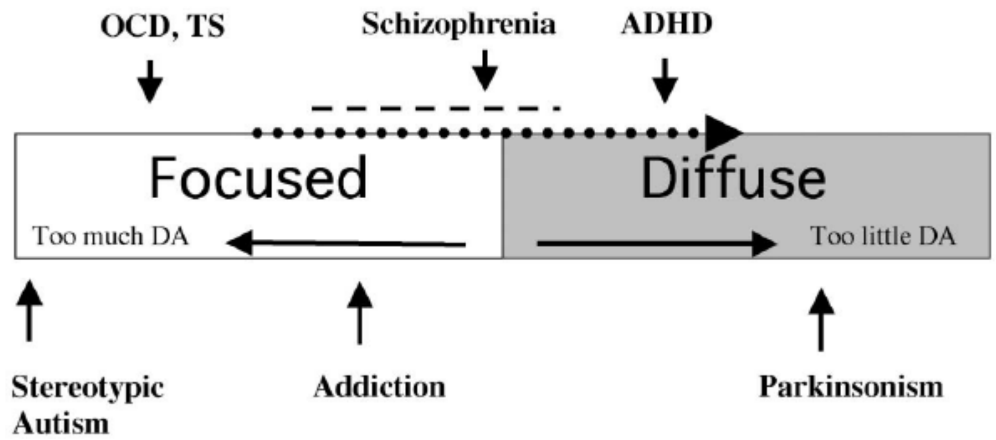
Cognitive Science 30 (2006) 3–41

Animal Foraging and the Evolution of Goal-Directed Cognition

Thomas T. Hills

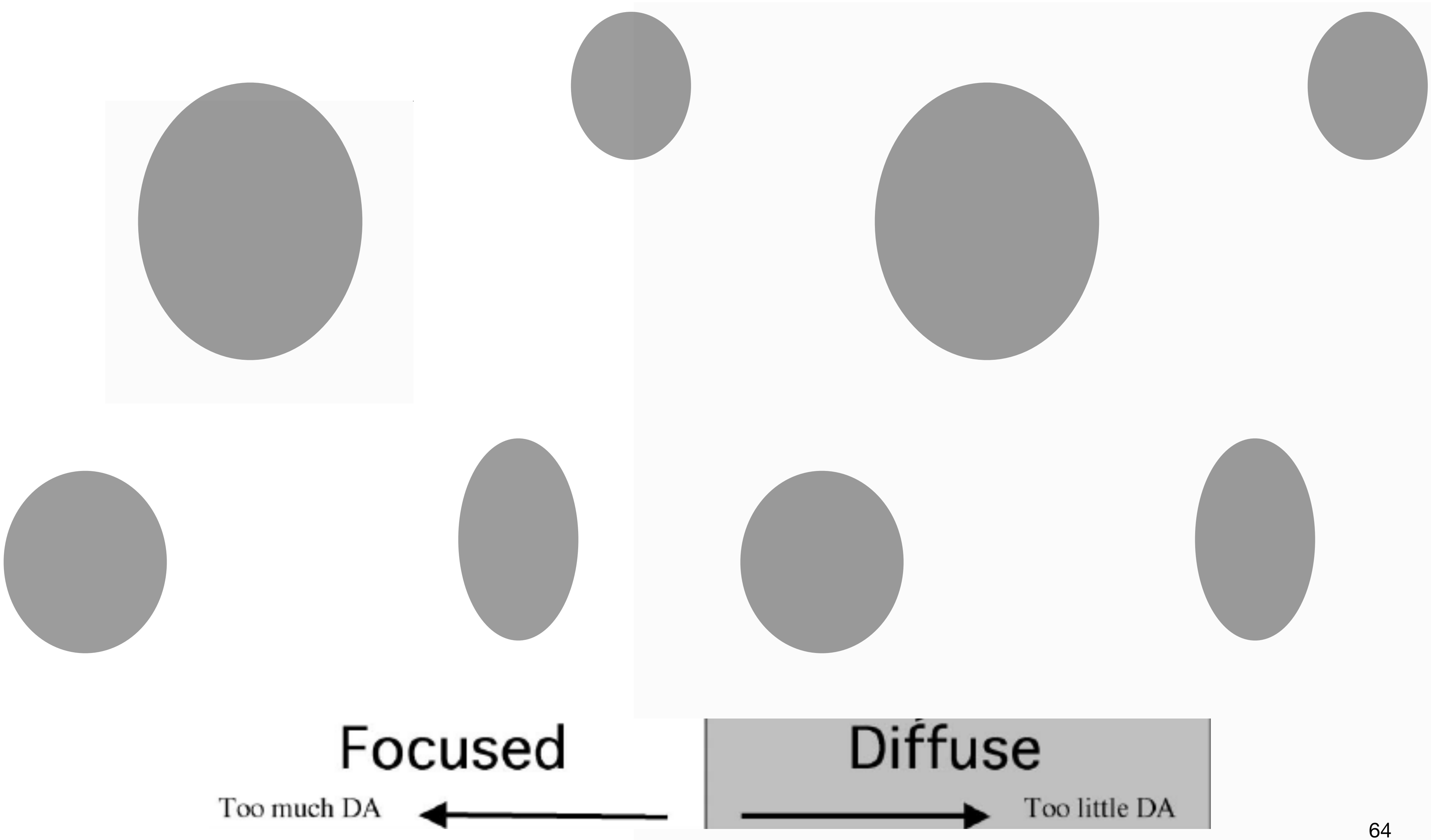
evolution of goal-directed cognition out of mechanisms
initially in control of spatial foraging but, through increasing cortical connections, eventually used to forage for information

Dopamine is a key component in foraging behaviors in invertebrates and vertebrates, in vertebrates dopamine is also associated with goal-directed cognition.

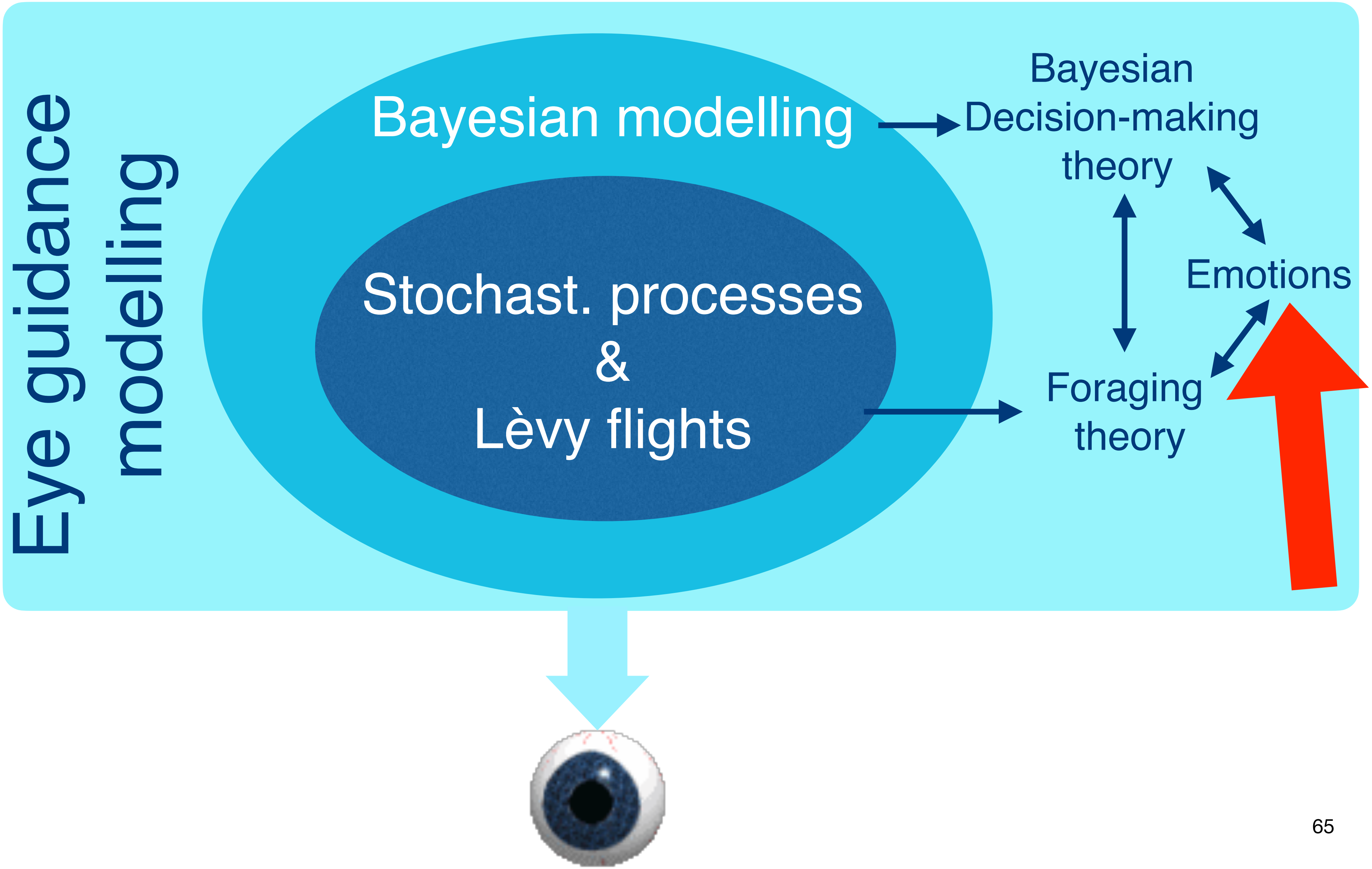


The evolutionary role of dopamine in the modulation of goal-directed behavior and cognition is further supported by pathologies of human goal-directed cognition, which have motor and cognitive dysfunction and organize themselves, with respect to dopaminergic activity, perseverative to unfocused.

At the heart of purposive eye guidance //the dopamine hypothesis (Hills)



Computational models of eye guidance //bringing emotions into the game

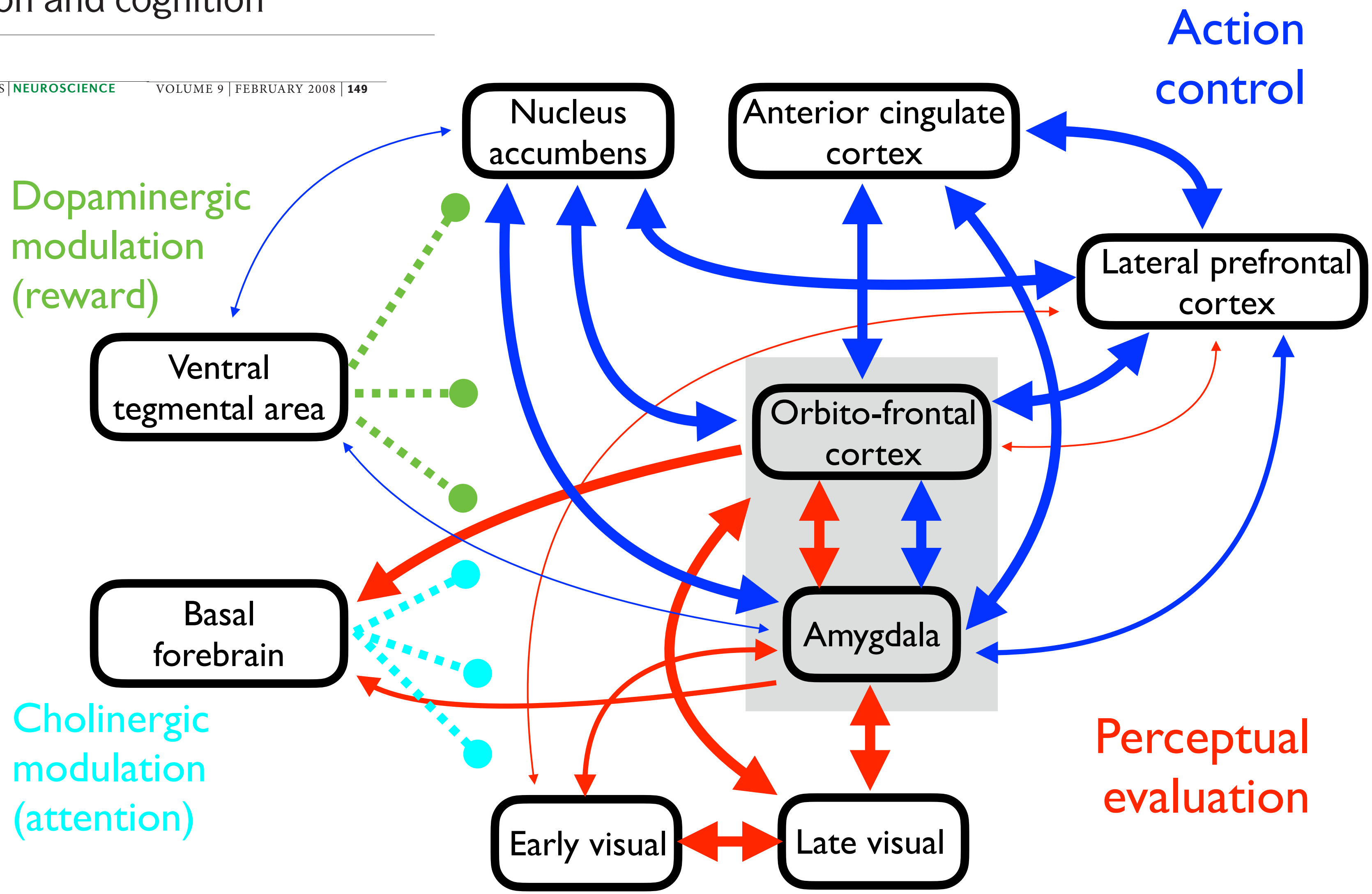


At the heart of purposive eye guidance //Value & reward: a doorway to emotions

On the relationship between emotion and cognition

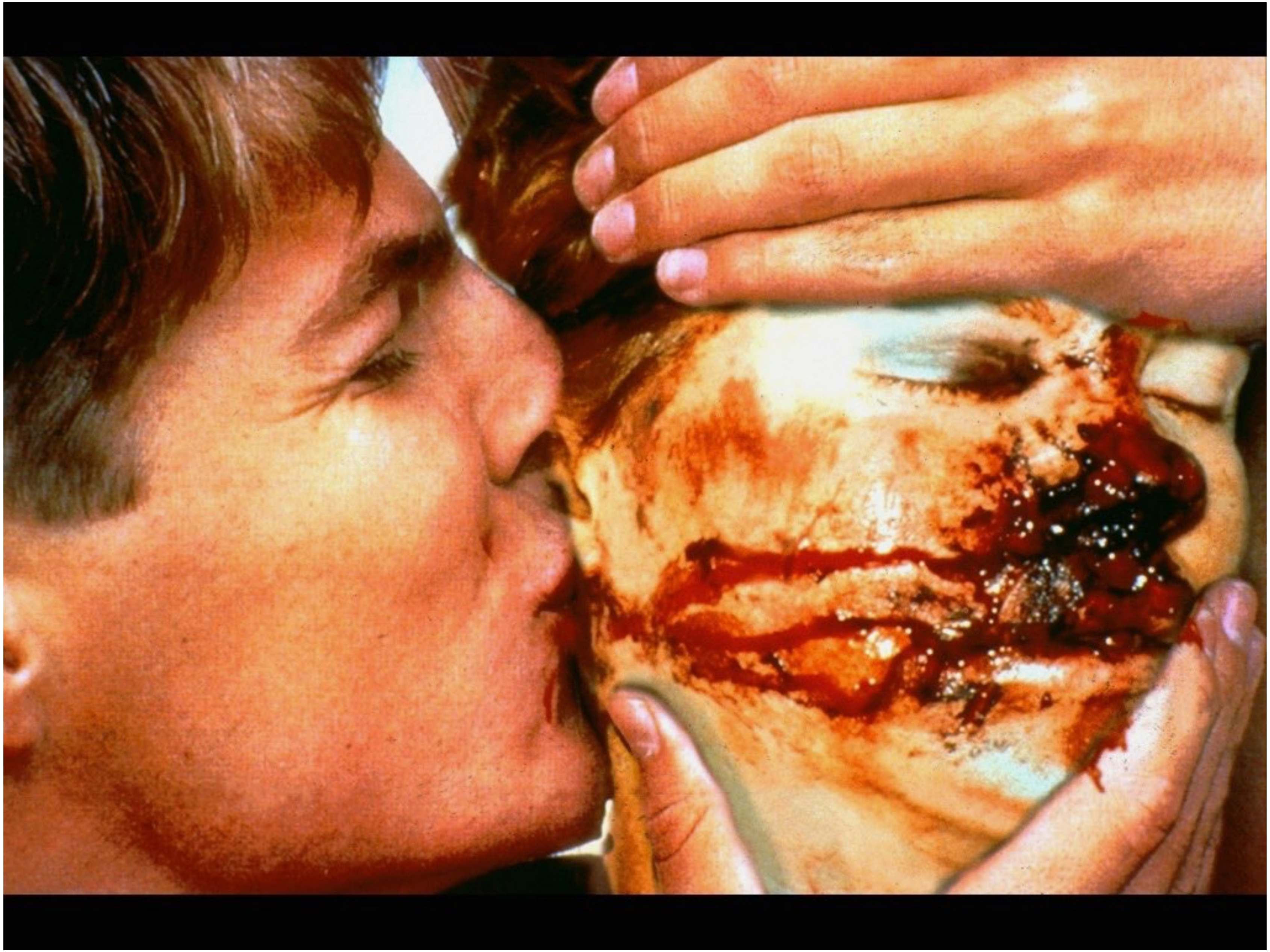
Luiz Pessoa

NATURE REVIEWS | NEUROSCIENCE VOLUME 9 | FEBRUARY 2008 | 149

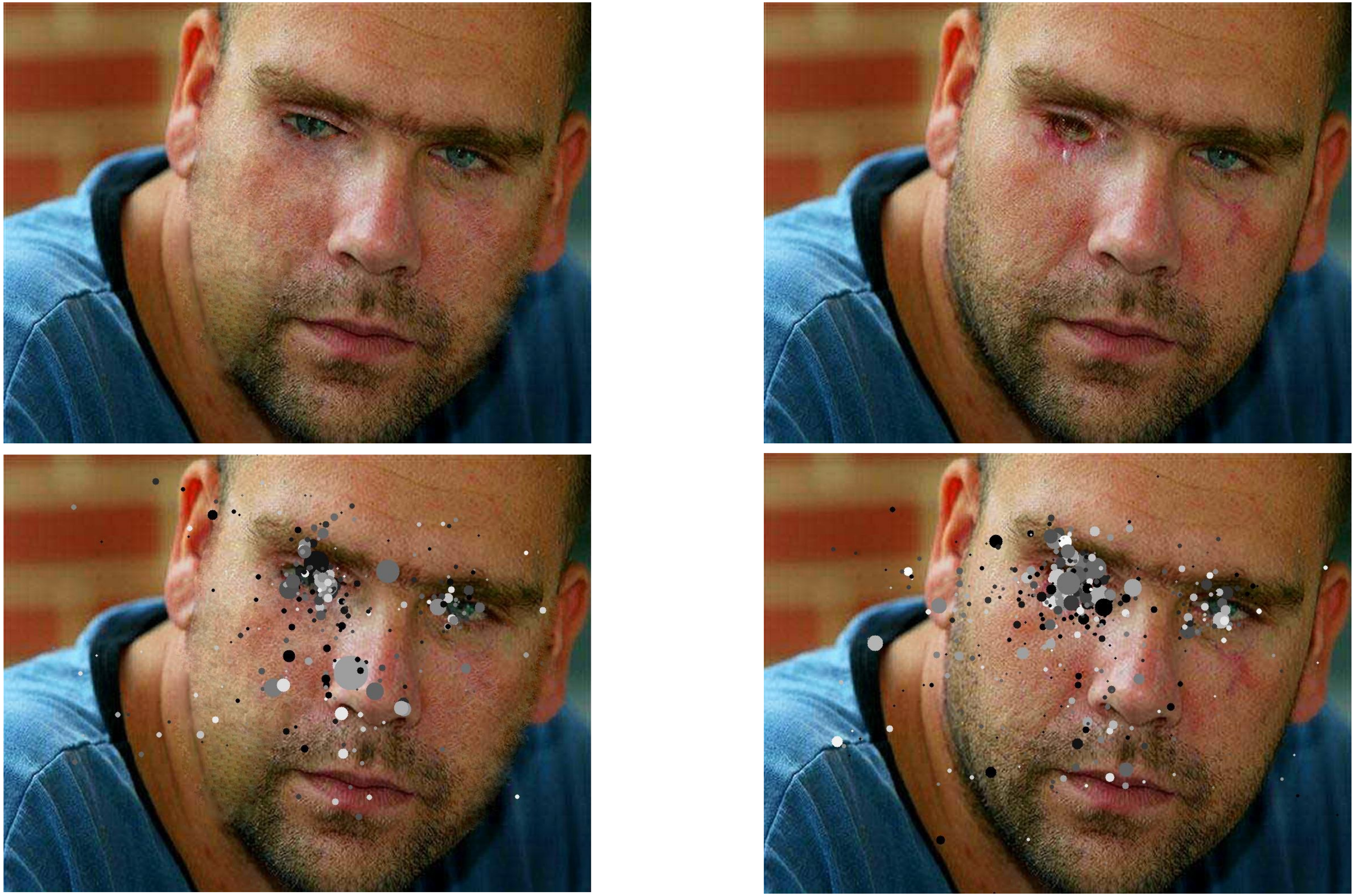


Future work
//bridging active sensing and emotions

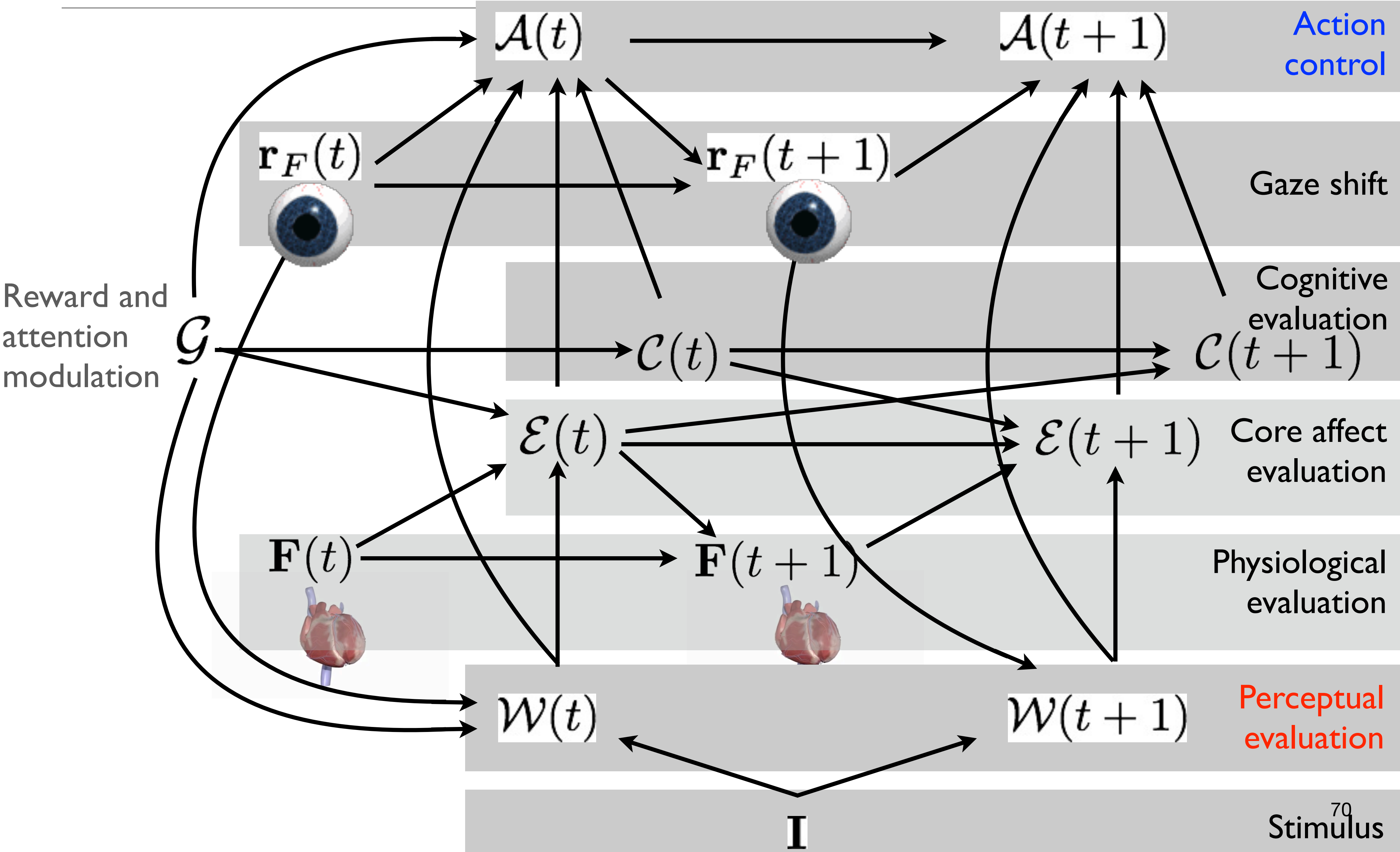
Future work
//bridging active sensing and emotions



Future work
//bridging active sensing and emotions



Future work
//bridging active sensing and emotions



Thank you!