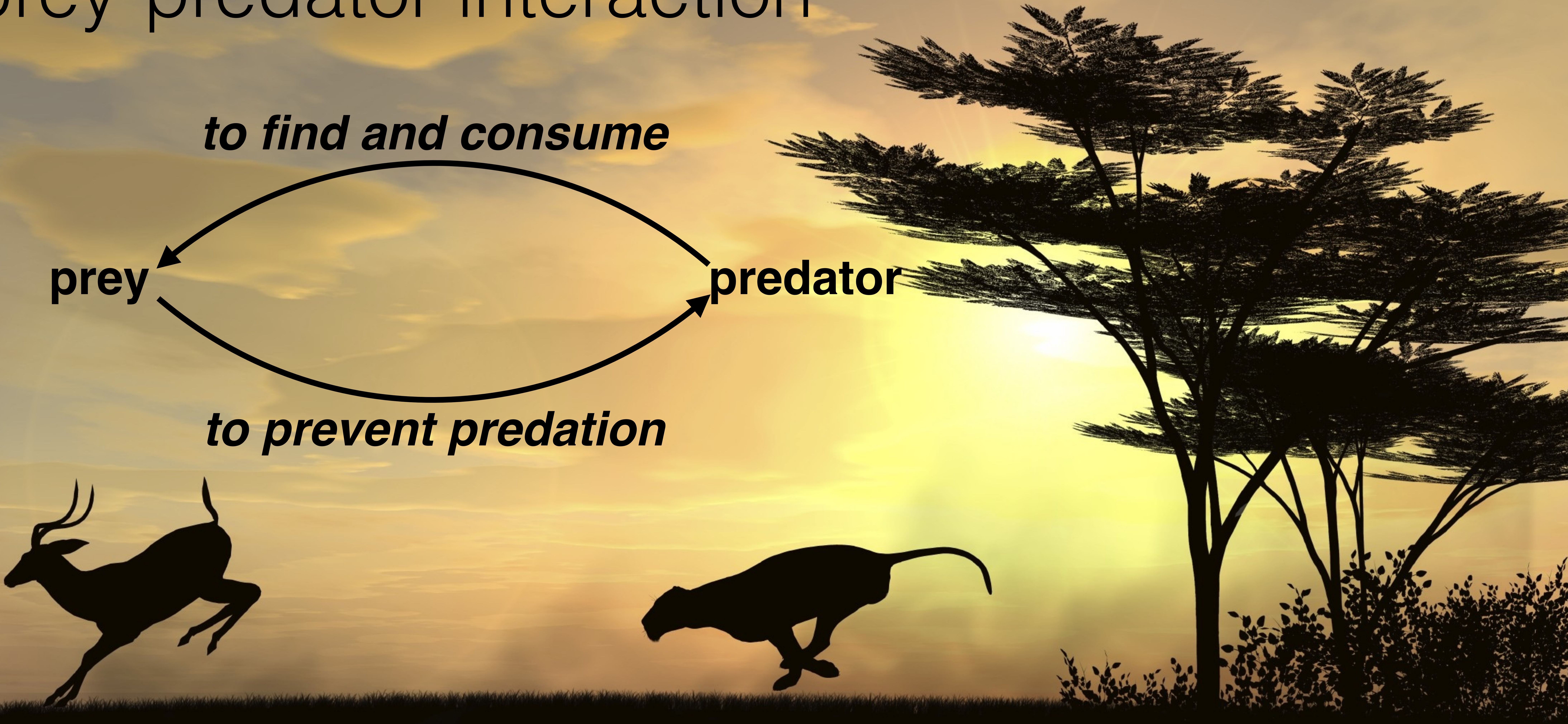
Defence in structured environments:
constructing knowledge to protect from danger

Błażej M. Bączkowski

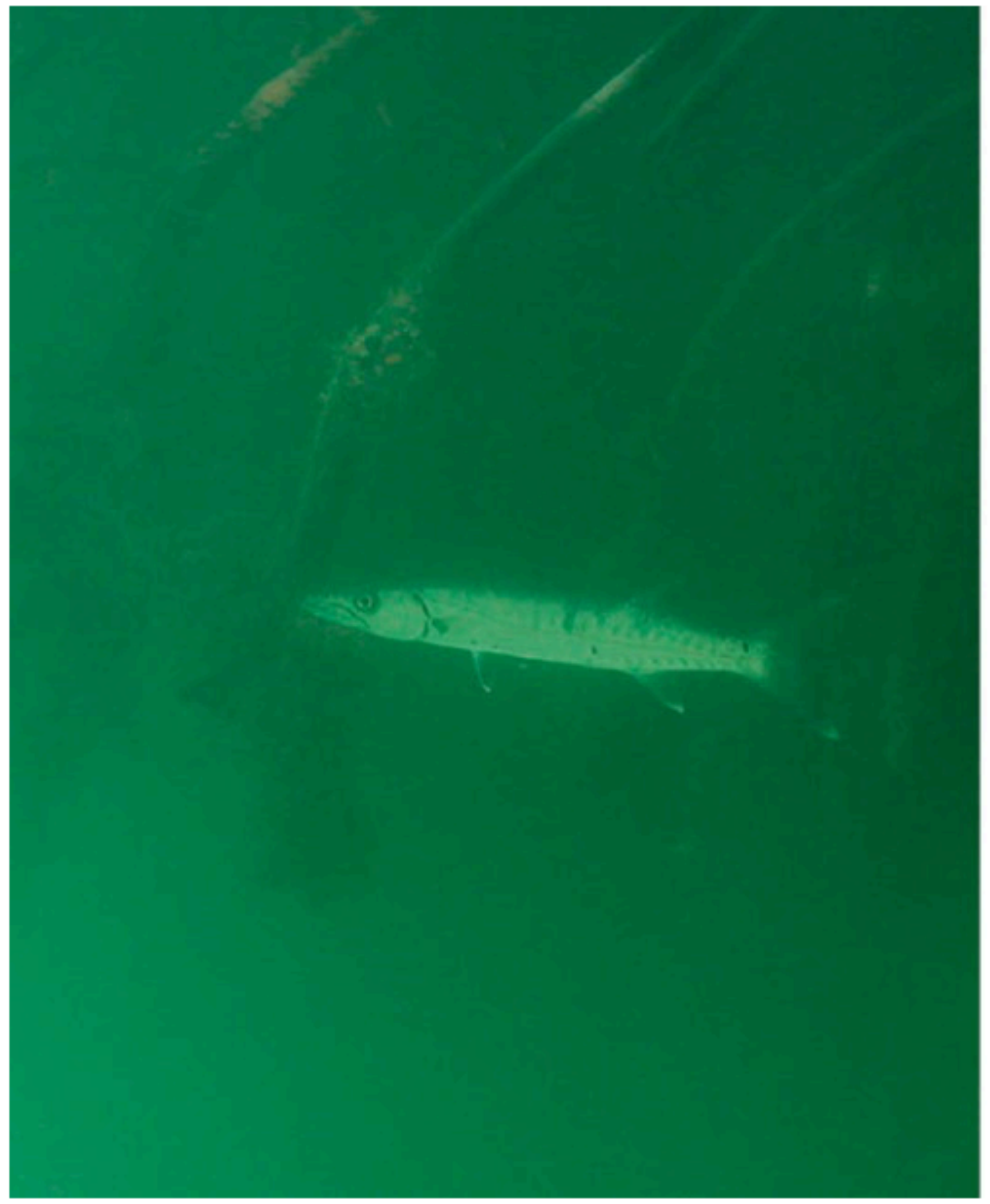
13.12.2022

Defence system evolved in the context of a prey-predator interaction

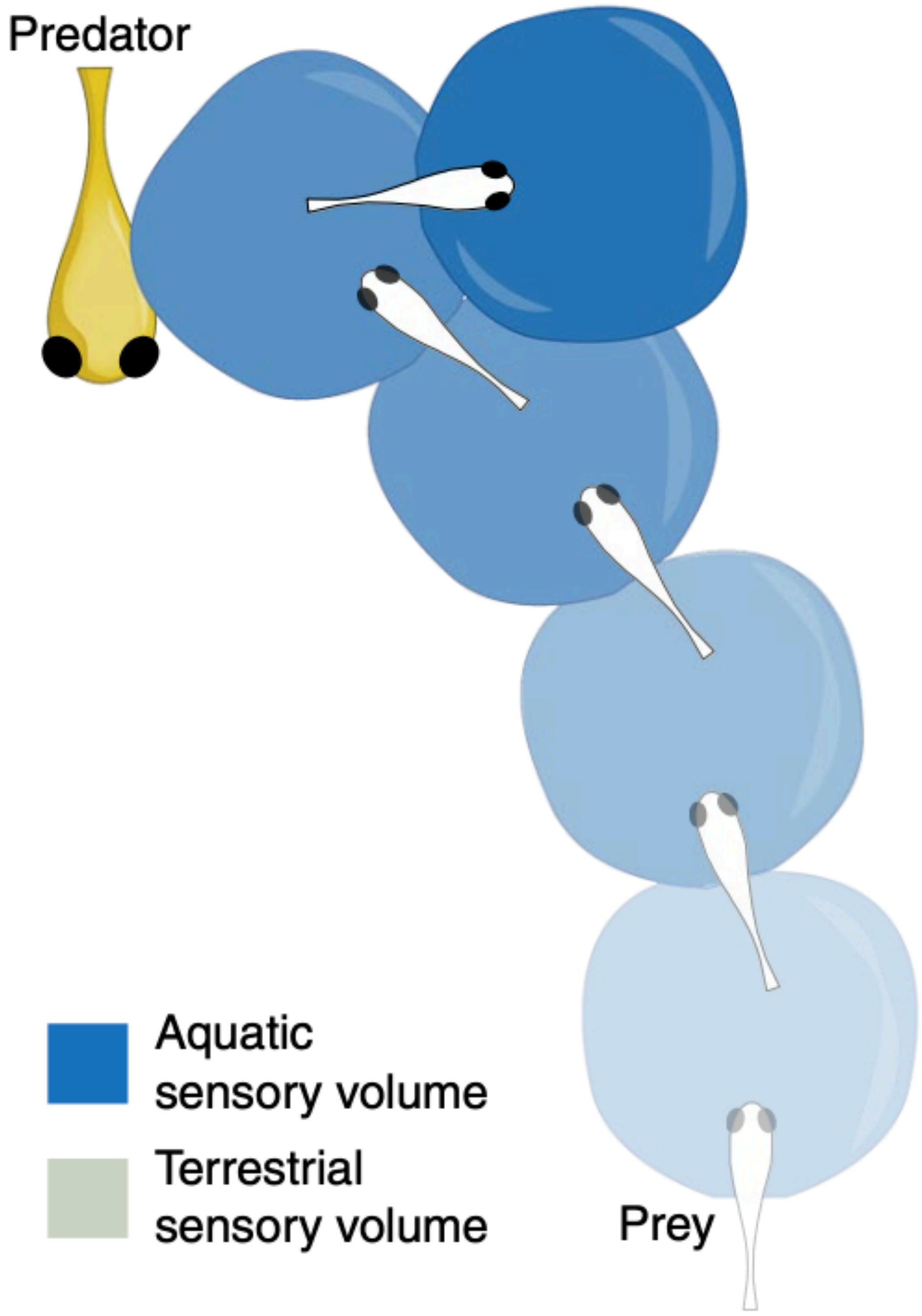


Avoiding predation in structured environments is temporally extended

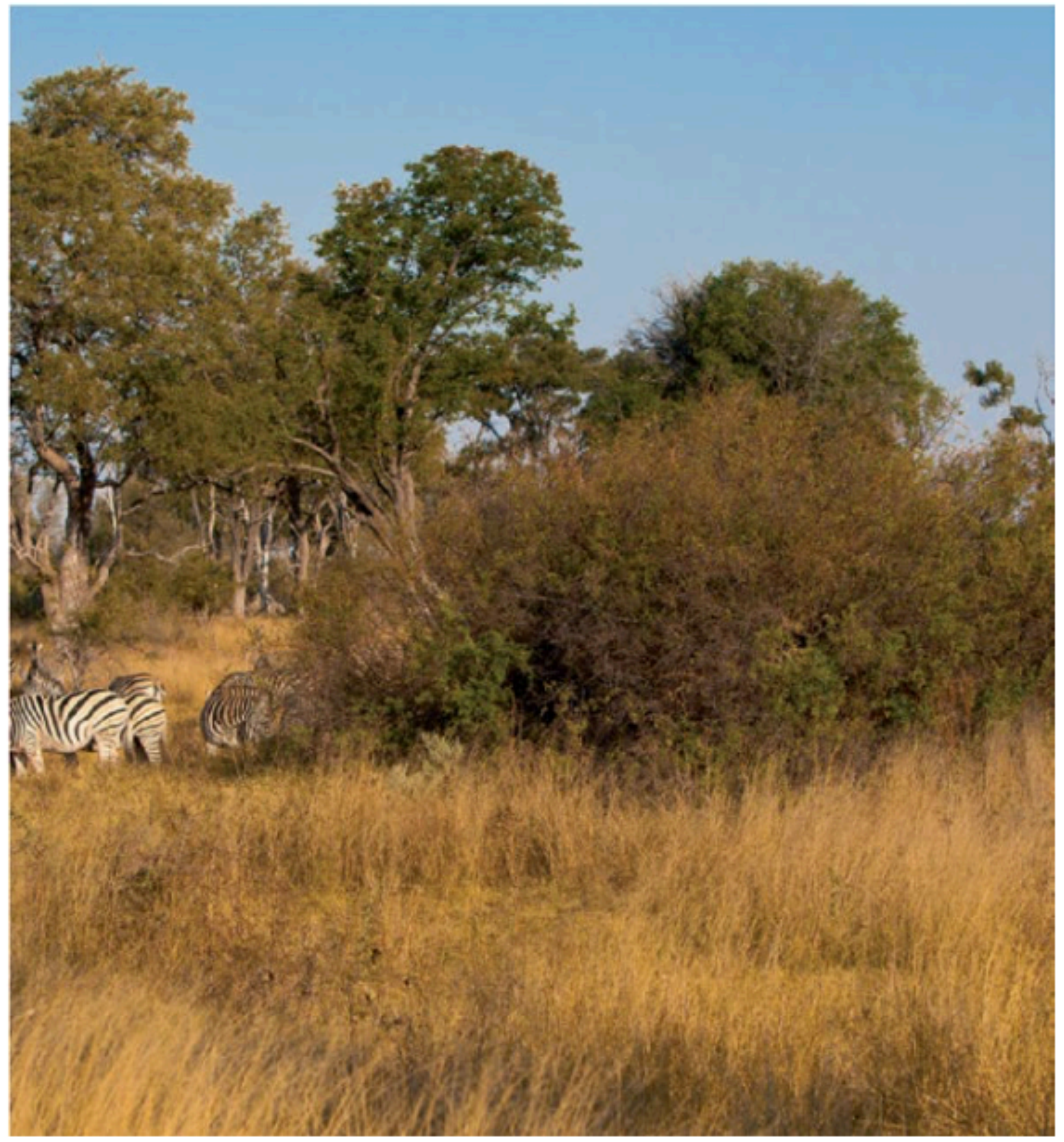
Aquatic visual scene



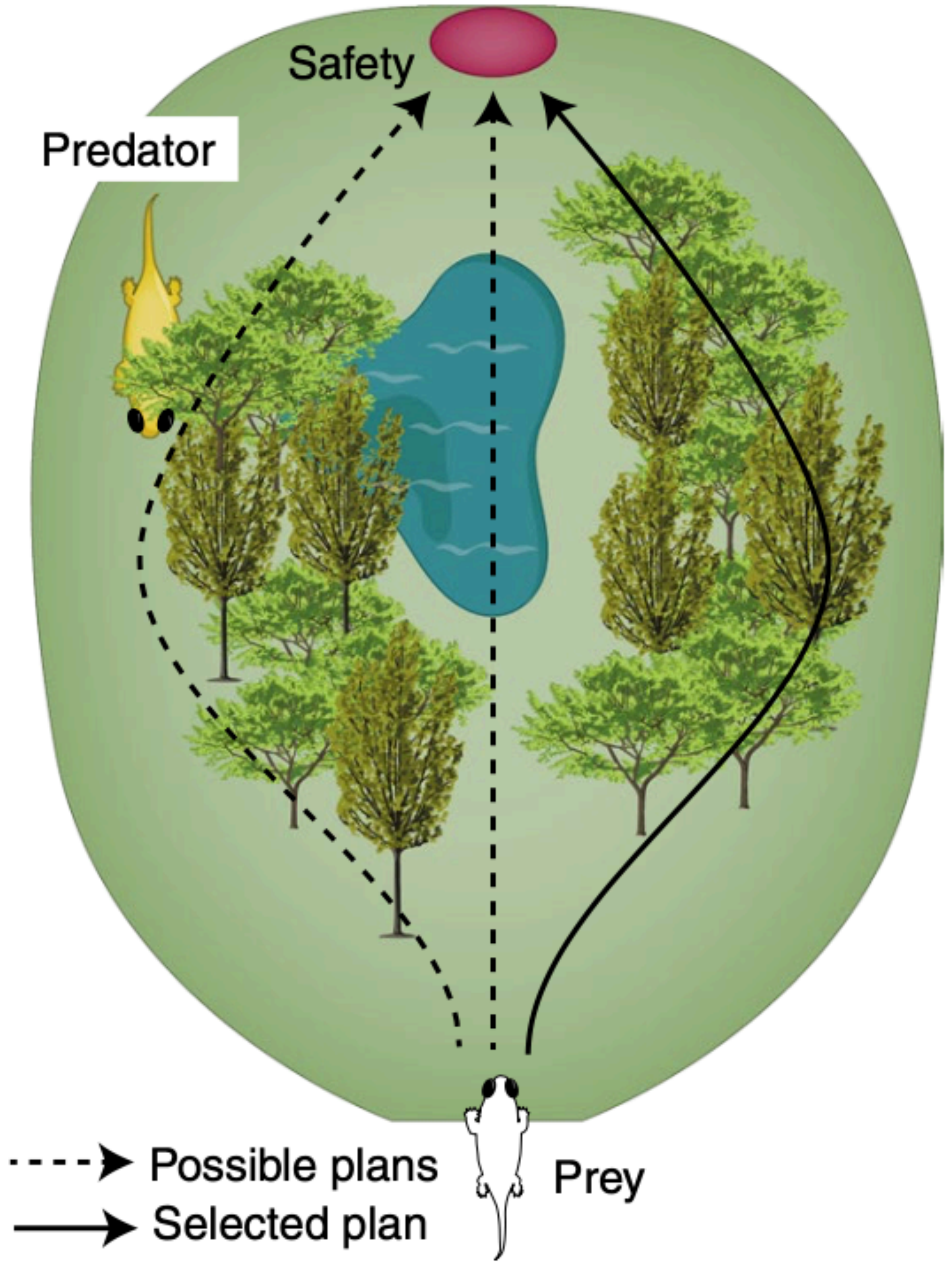
Short-range sensing (dynamic)



Terrestrial visual scene

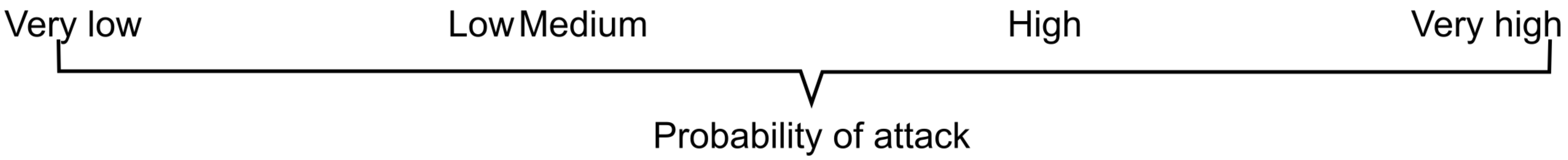


Long-range sensing (environment partially occluded)



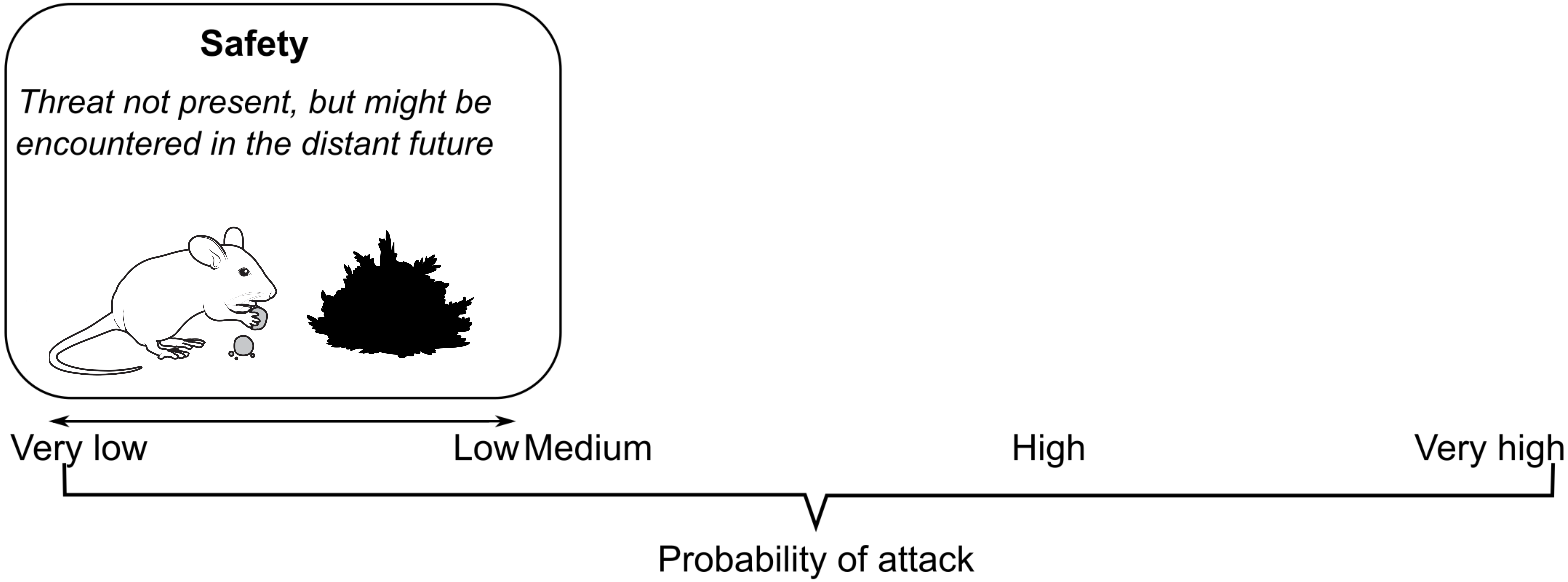
Hunt et al. (2021)

Structured environments shape the organisation of defence



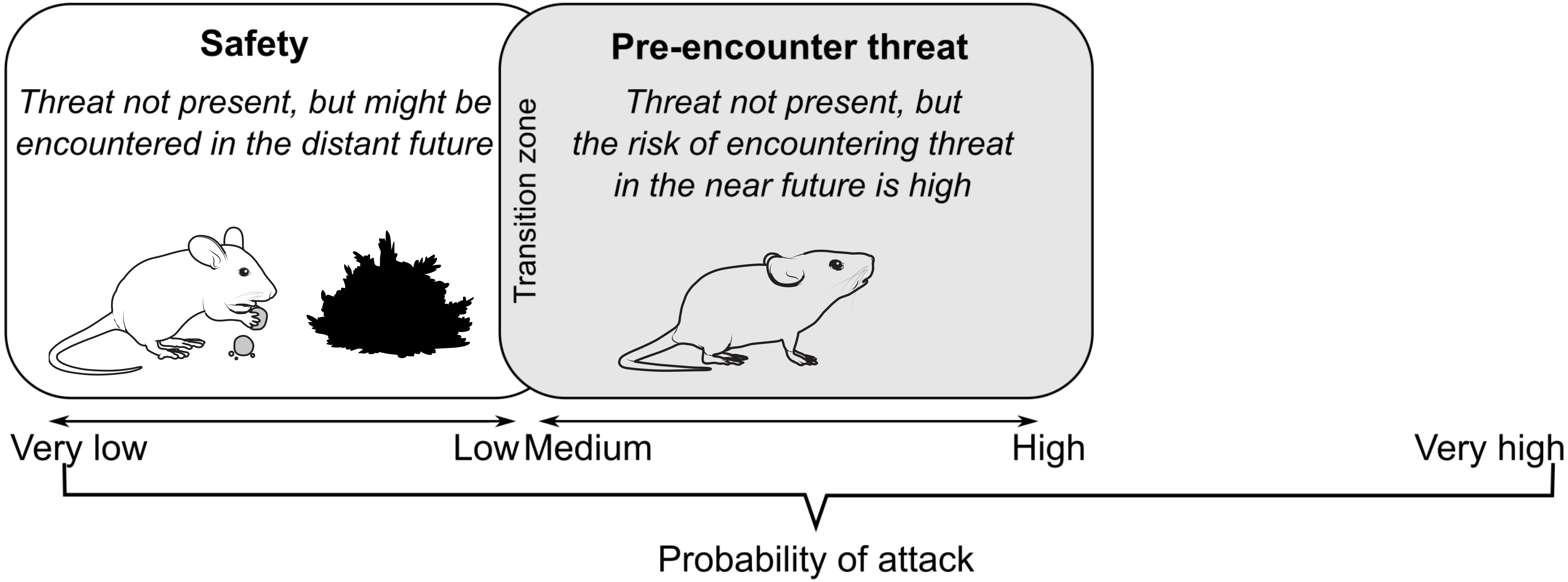
Mobbs et al., 2020; Fanselow & Lester, 1988

Structured environments shape the organisation of defence



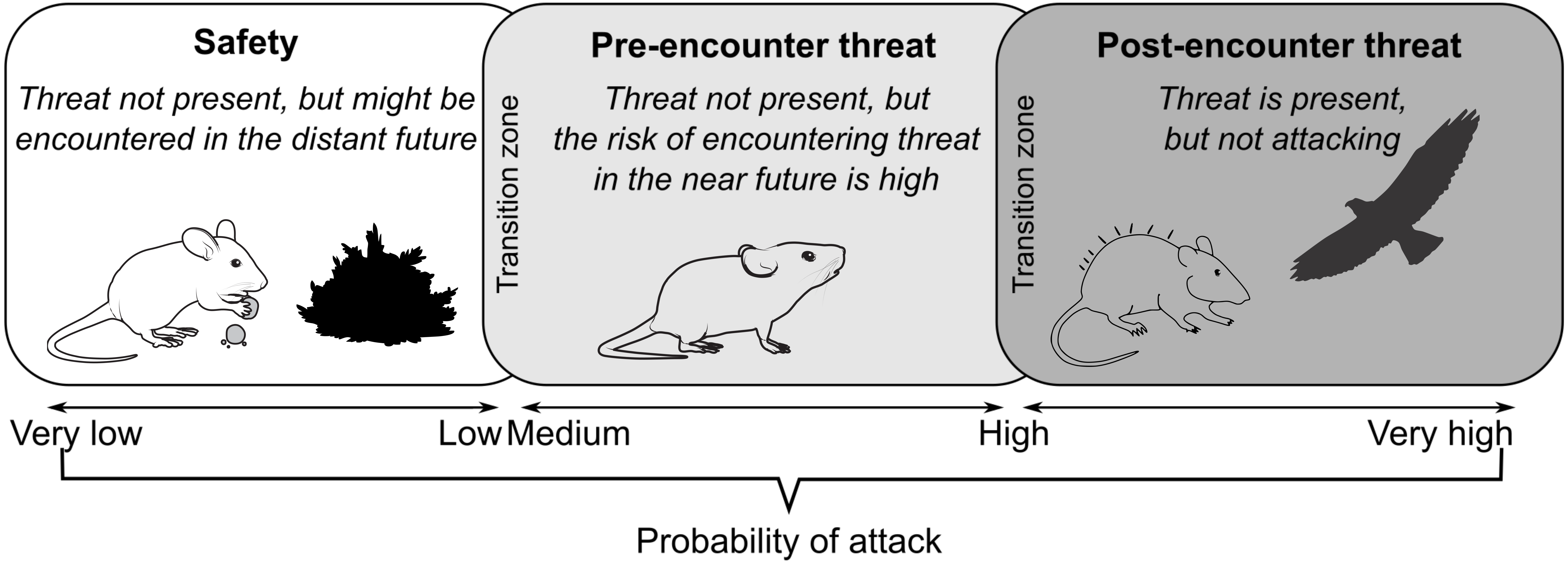
Mobbs et al., 2020; Fanselow & Lester, 1988

Structured environments shape the organisation of defence



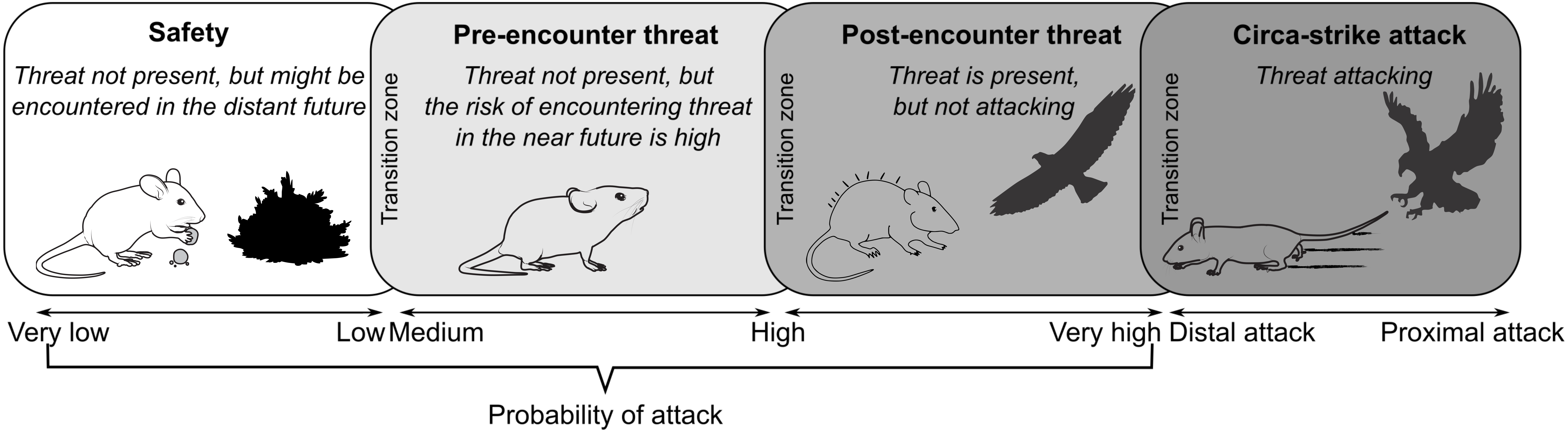
Mobbs et al., 2020; Fanselow & Lester, 1988

Structured environments shape the organisation of defence



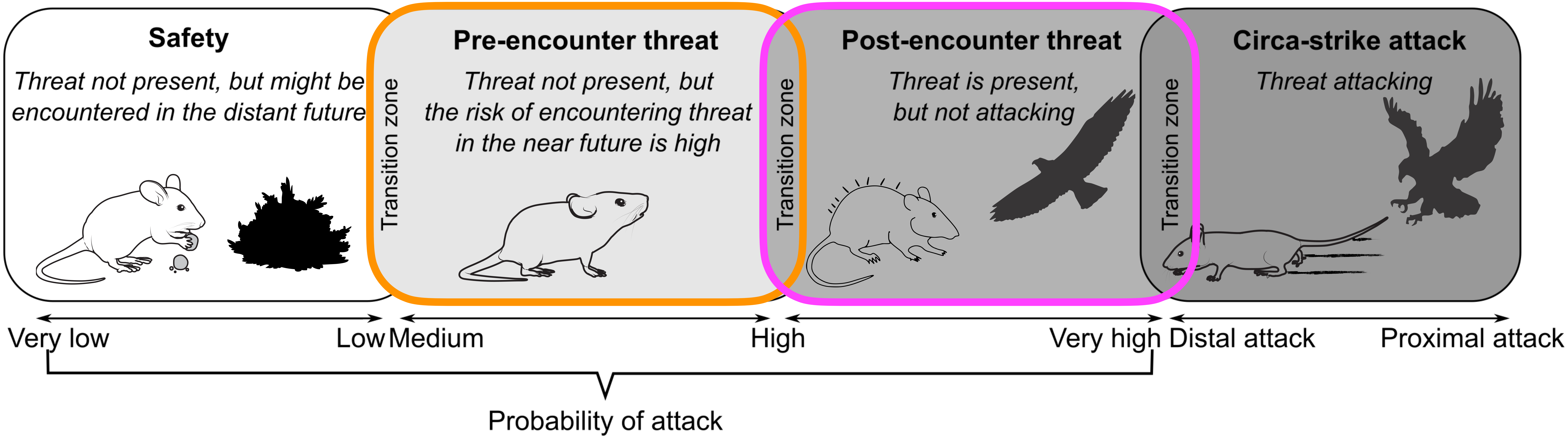
Mobbs et al., 2020; Fanselow & Lester, 1988

Structured environments shape the organisation of defence



Mobbs et al., 2020; Fanselow & Lester, 1988

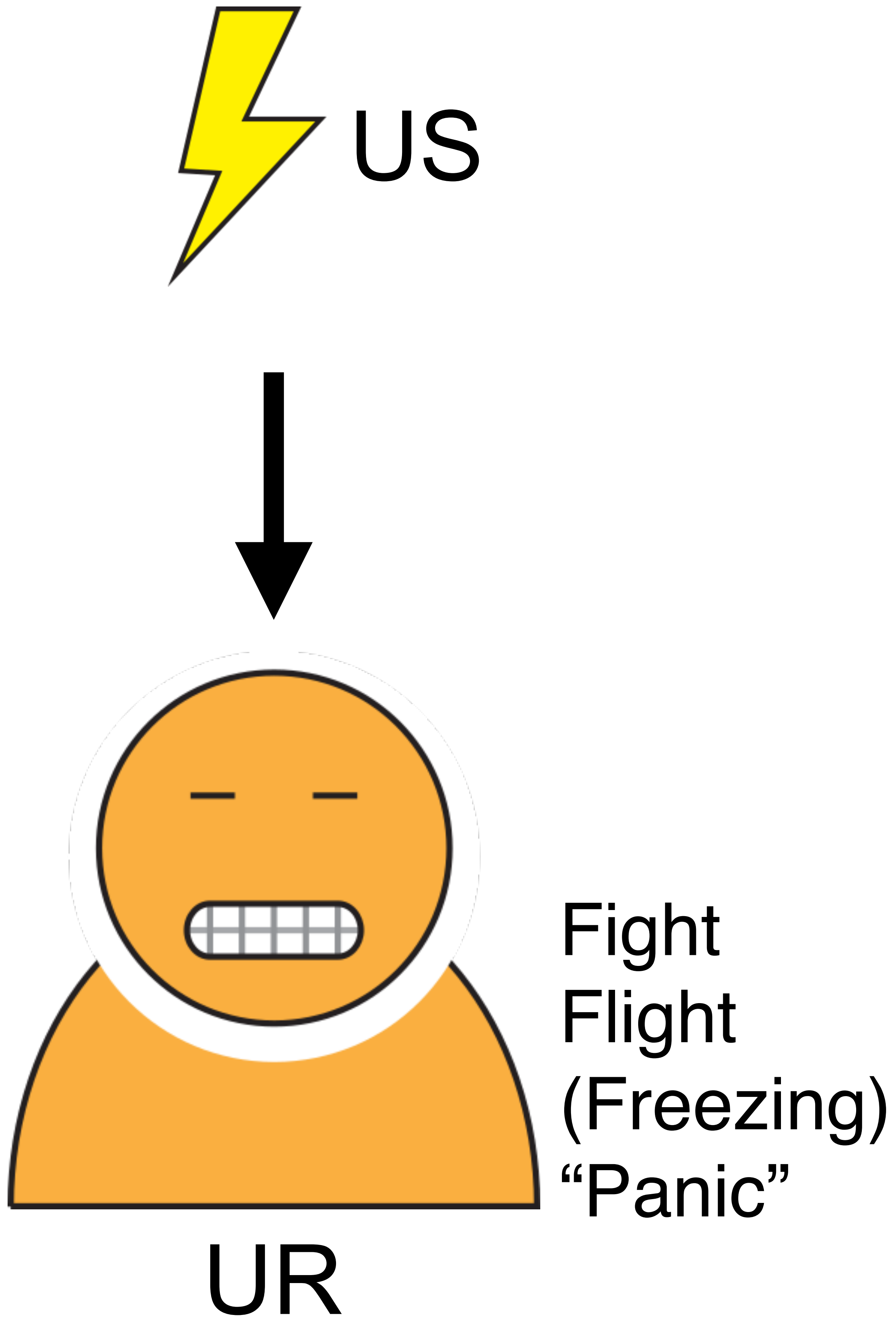
The organisation of defence rests upon threat prediction and detection



Mobbs et al., 2020; Fanselow & Lester, 1988

The ability to detect recurring threats comes from Pavlovian threat (fear) conditioning

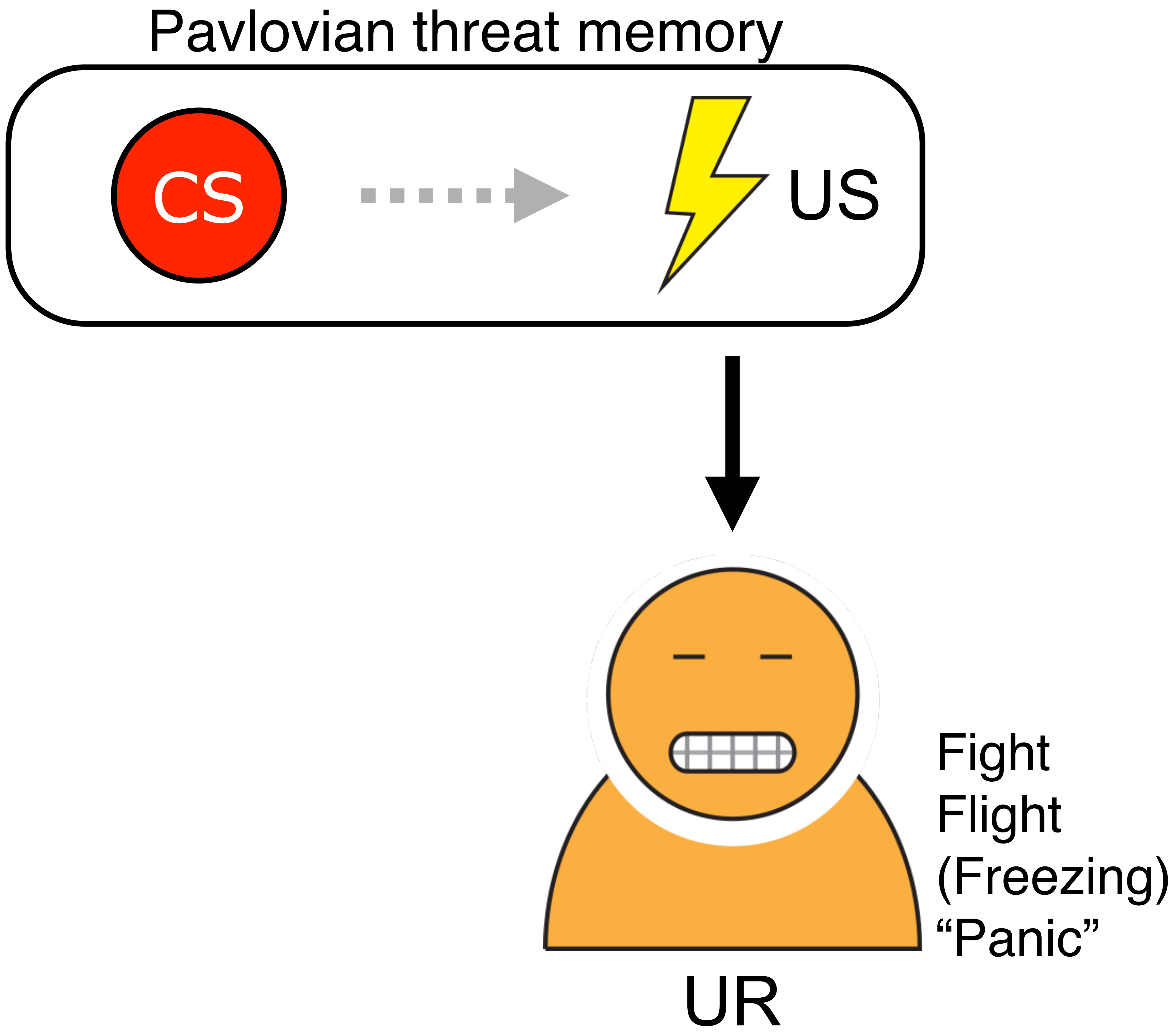
The ability to detect recurring threats comes from Pavlovian threat (fear) conditioning



US- unconditioned stimulus; UR- unconditioned response; CS- conditioned stimulus; CR- conditioned response

Fanselow & Wassum, 2016; Ledoux, 2000; 2014; Rescorla, 1988

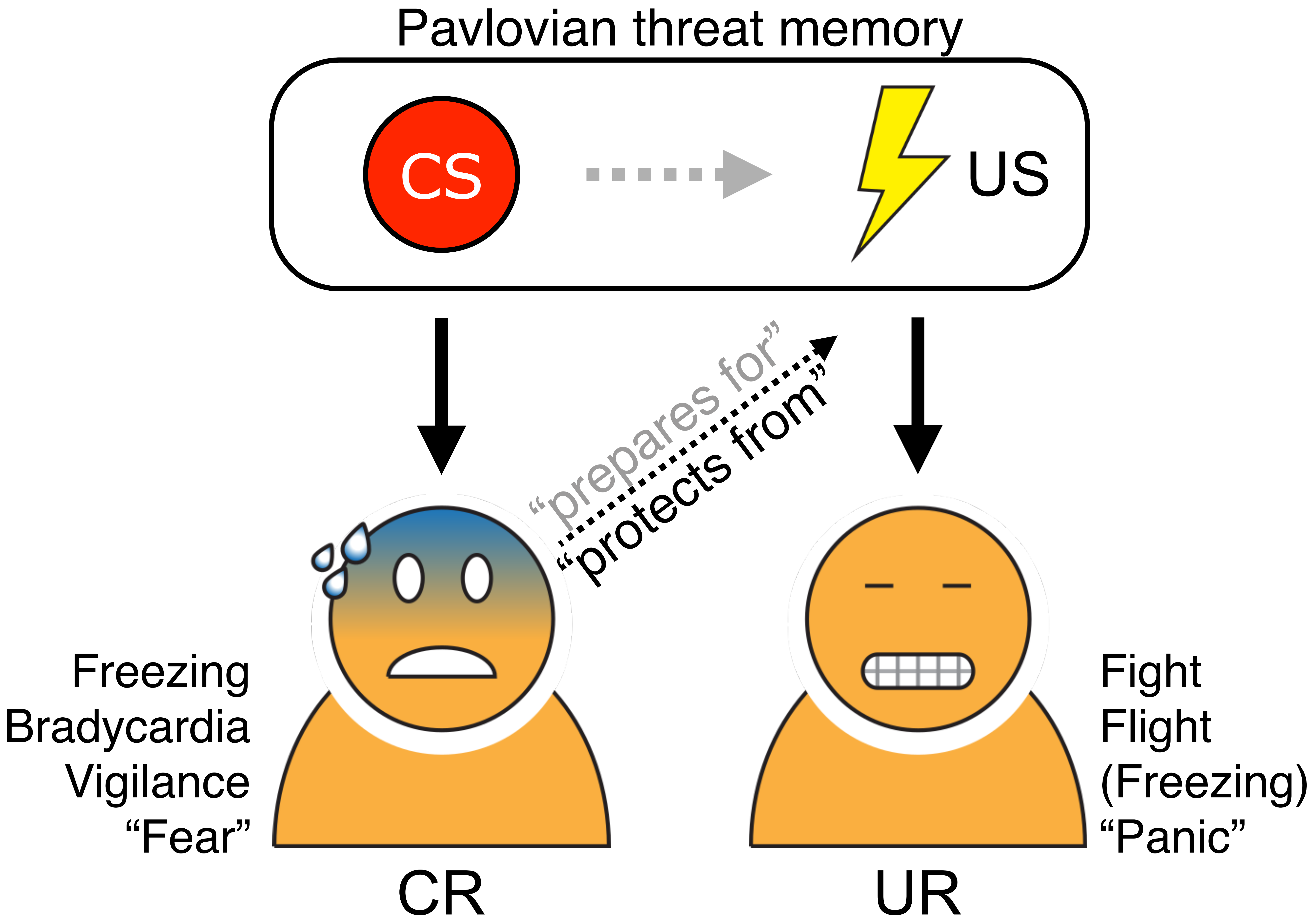
The ability to detect recurring threats comes from Pavlovian threat (fear) conditioning



US- unconditioned stimulus; UR- unconditioned response; CS- conditioned stimulus; CR- conditioned response

Fanselow & Wassum, 2016; Ledoux, 2000; 2014; Rescorla, 1988

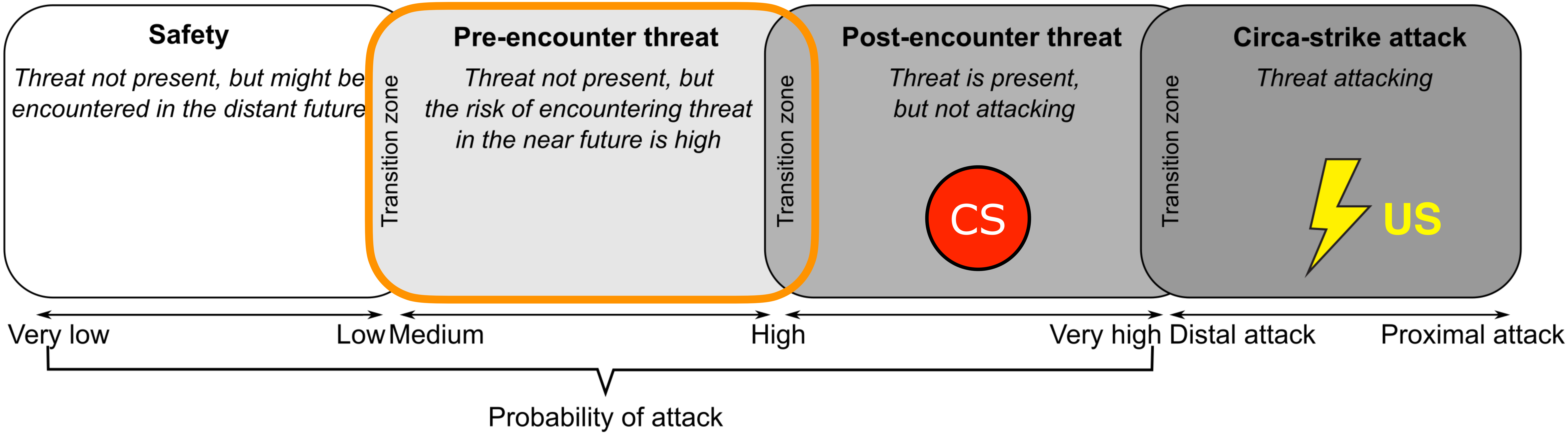
The ability to detect recurring threats comes from Pavlovian threat (fear) conditioning



US- unconditioned stimulus; UR- unconditioned response; CS- conditioned stimulus; CR- conditioned response

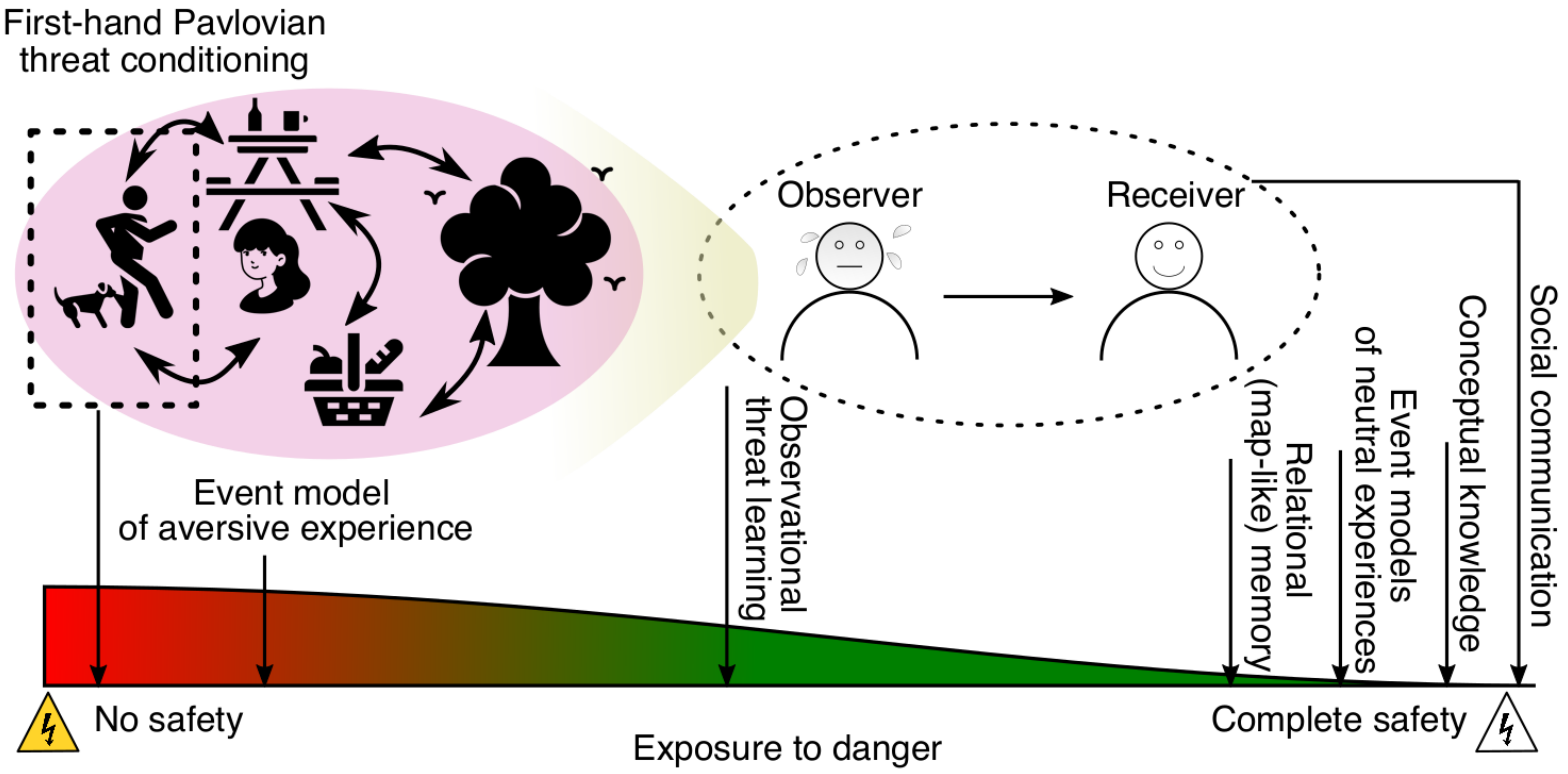
Fanselow & Wassum, 2016; Ledoux, 2000; 2014; Rescorla, 1988

Knowing what posits a threat is necessary for its detection but not sufficient for its prediction



Mobbs et al., 2020; Fanselow & Lester, 1988

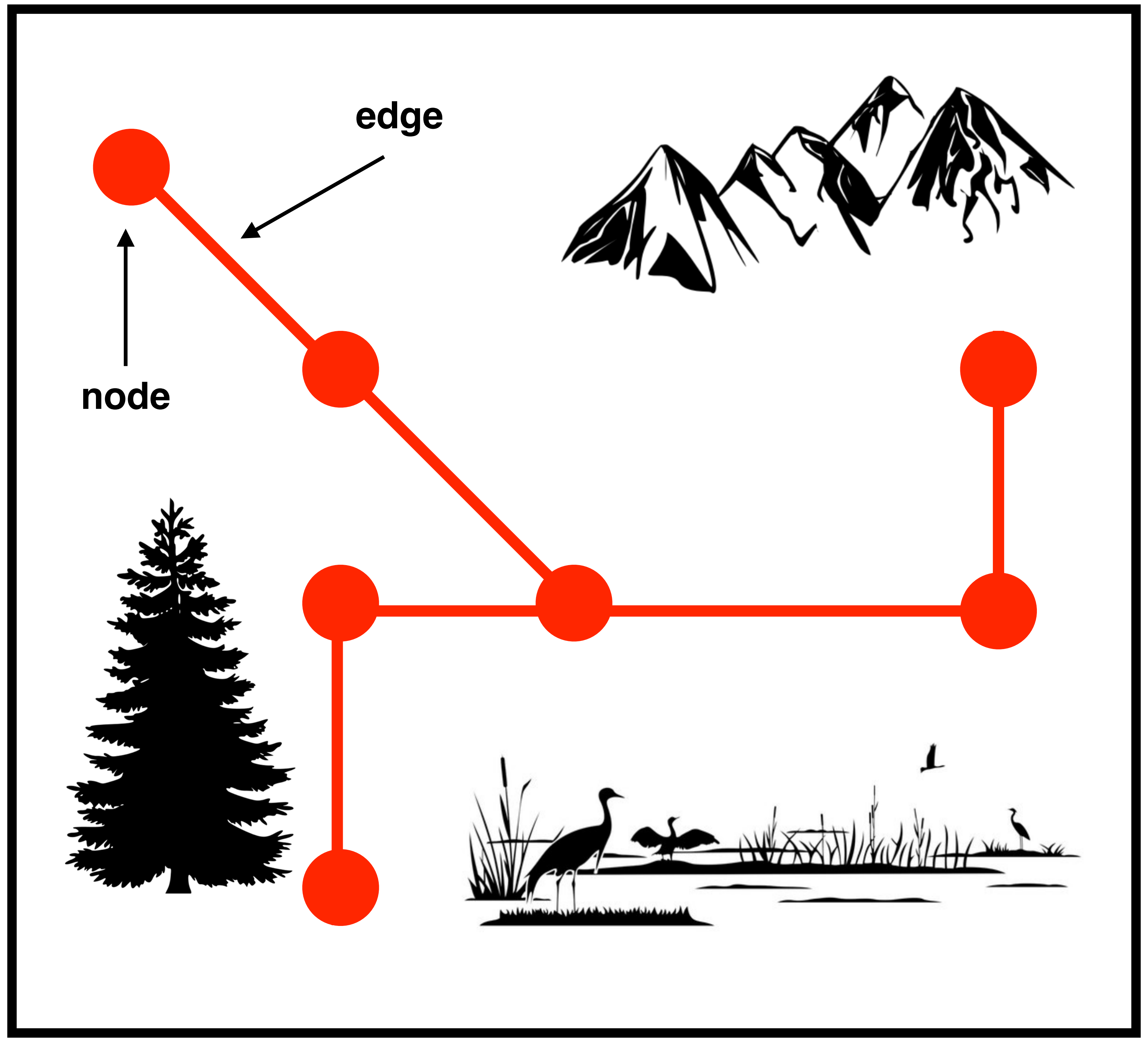
Individuals are equipped with a rich repertoire of mnemonic processes operating in safety



Structure of the environment is acquired in safety during spontaneous exploration (driven by curiosity)

Structure of the environment is acquired in safety during spontaneous exploration (driven by curiosity)

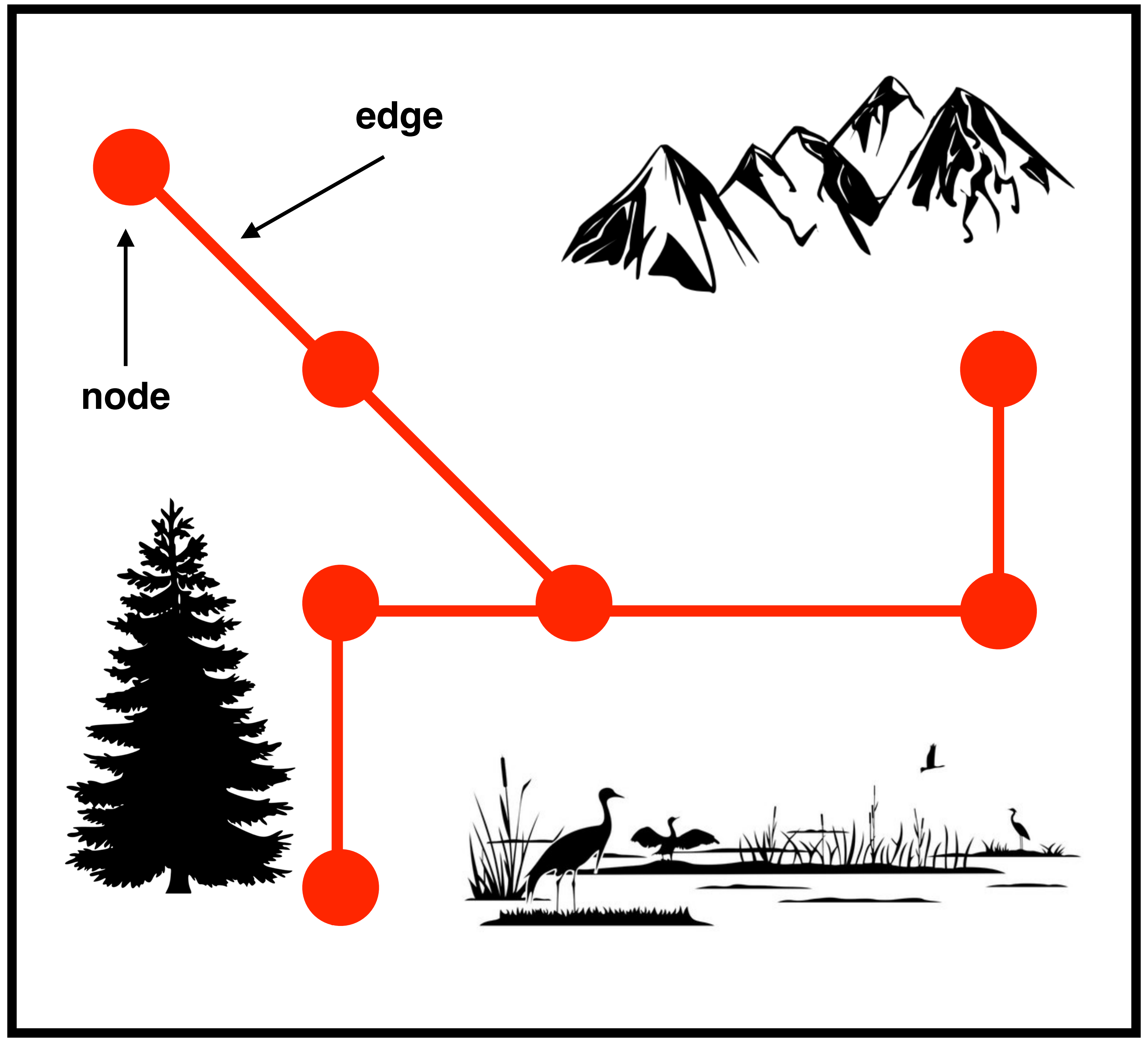
Cognitive map
as a graph



Adapted from: Peer et al., 2020

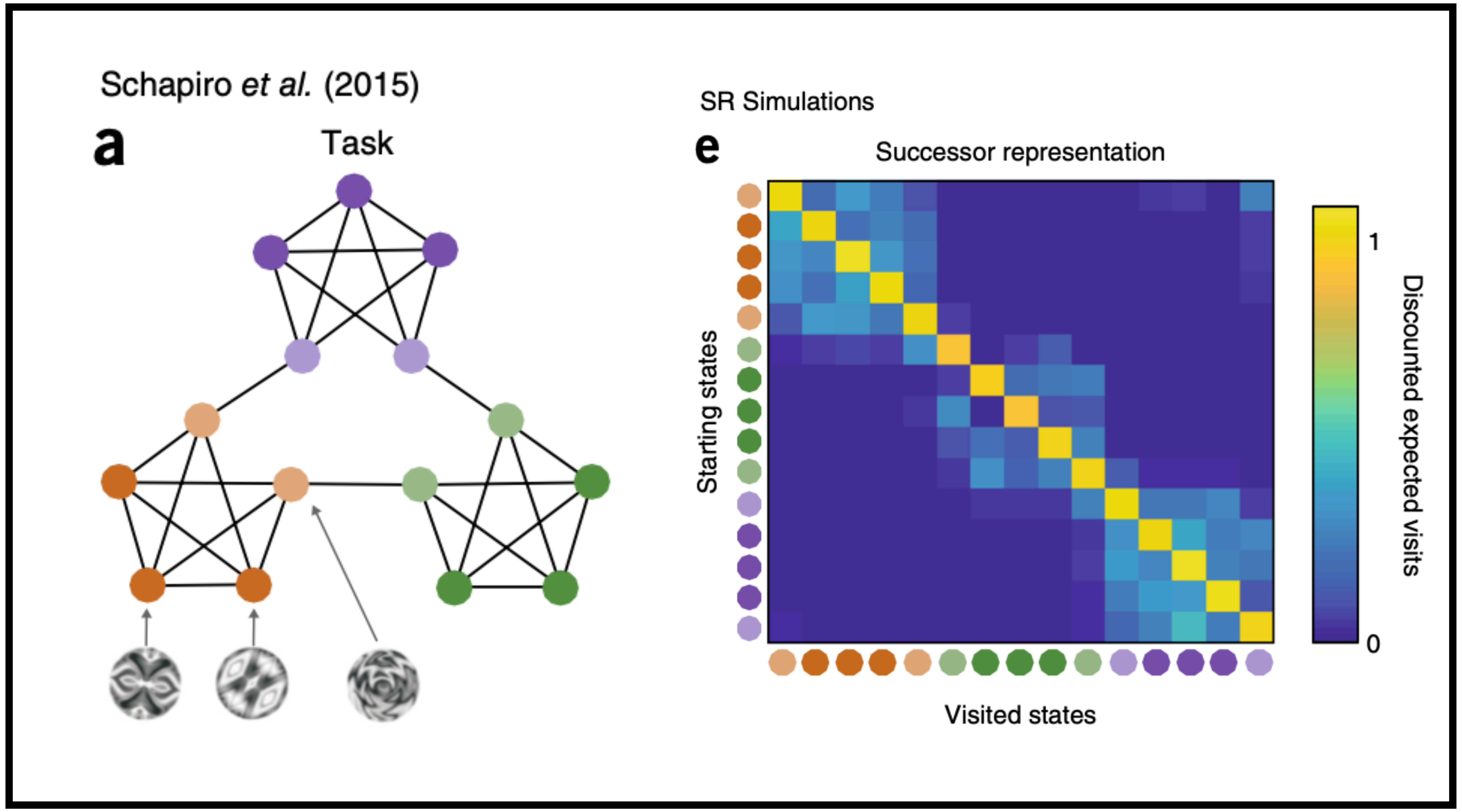
Structure of the environment is acquired in safety during spontaneous exploration (driven by curiosity)

Cognitive map
as a graph



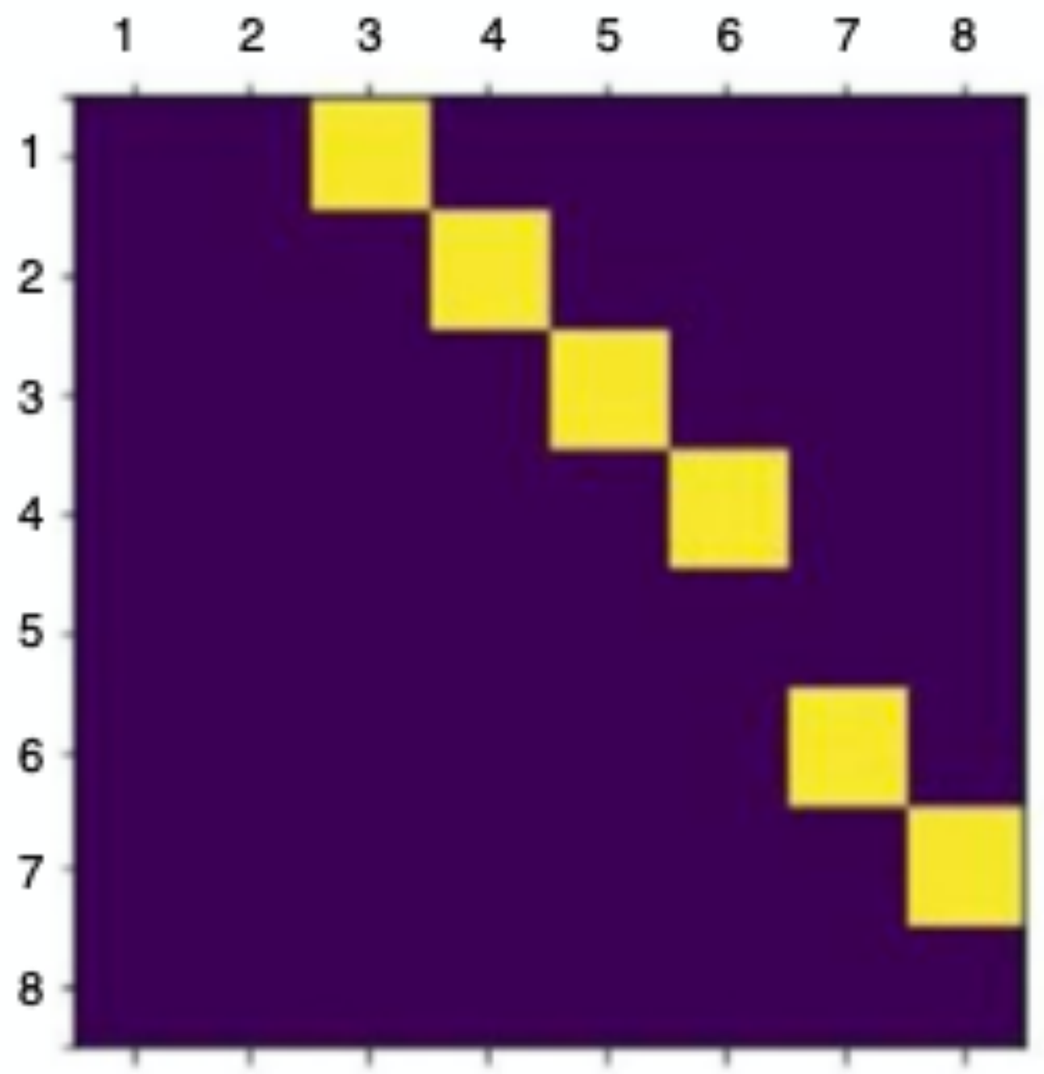
Adapted from: Peer et al., 2020

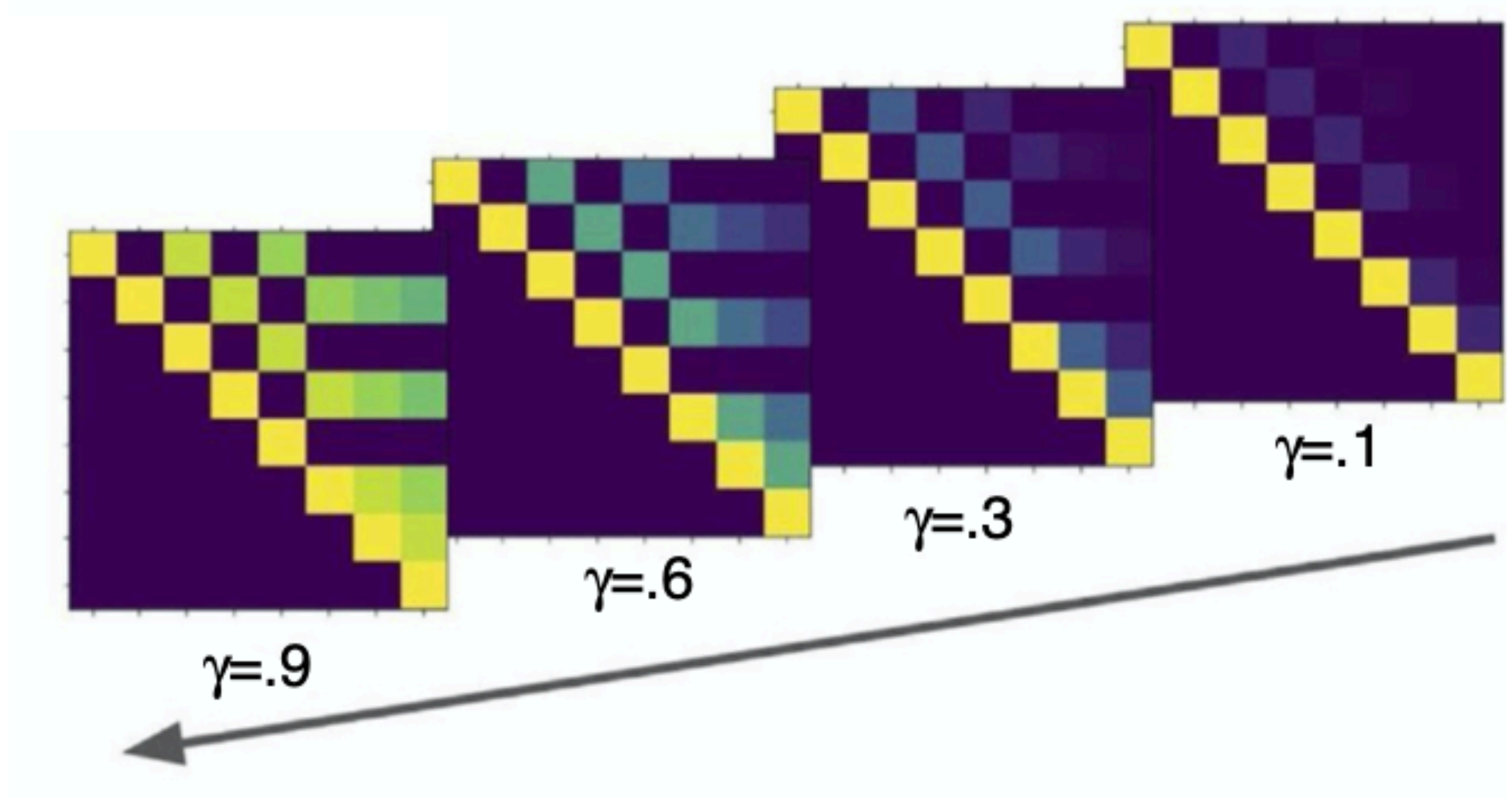
Cognitive map
as a successor representation



Stachenfeld et al. (2017)

Building successor representations





Policy-independent (random walk)

$$M = (I - \gamma T)^{-1}$$

$$0 < \gamma < 1$$

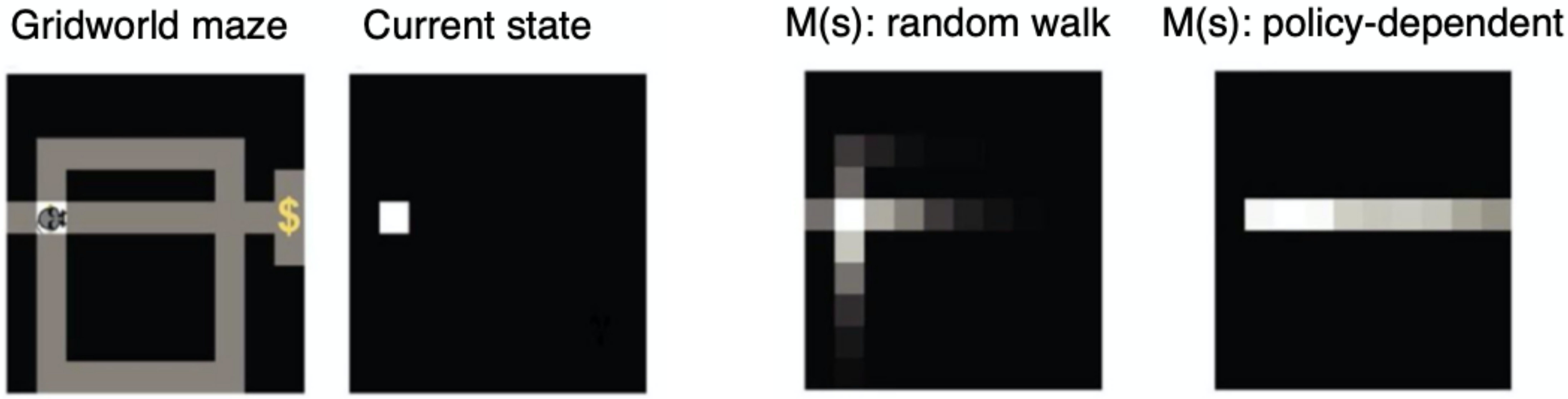
Policy dependent

$$M(s) = M(s) + \alpha(onehot(s_{new}) + \gamma M(s_{new}) - M(s))$$

$$0 < \alpha < 1$$

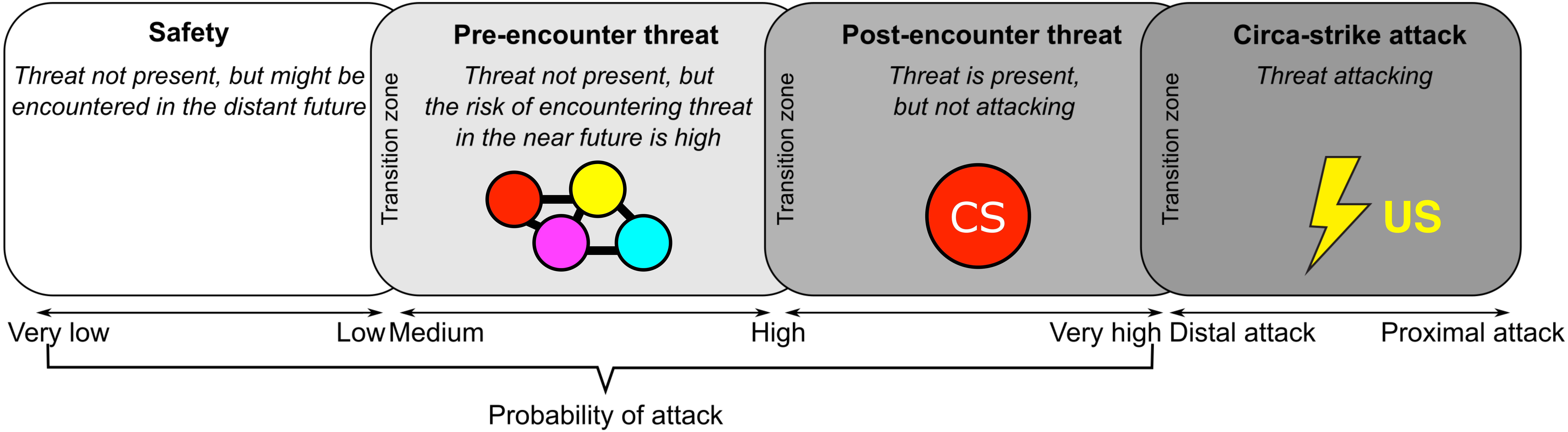
Momennejad (2020)

Building successor representations

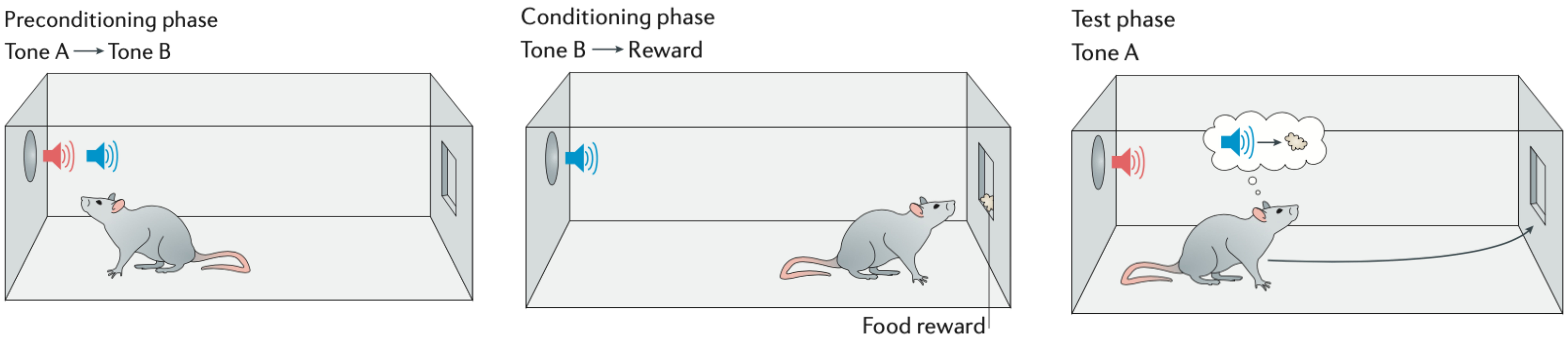


Russek et al. (2017)

To support threat prediction and detection, complementary memories acquired across time and space must interact

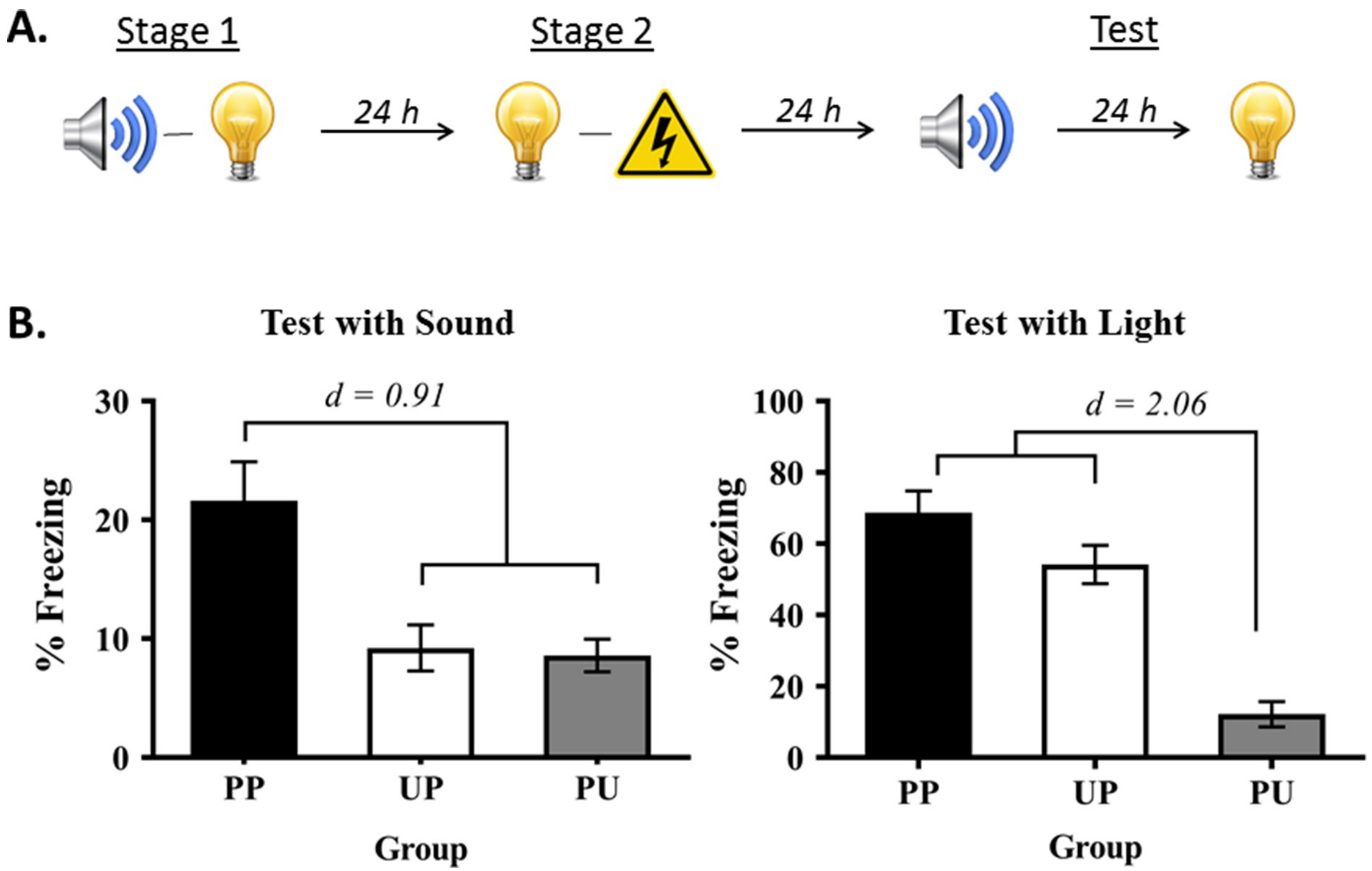


The phenomenon of sensory pre-conditioning



Wikenheiser & Schoenbaum, (2017)

The phenomenon in the aversive domain

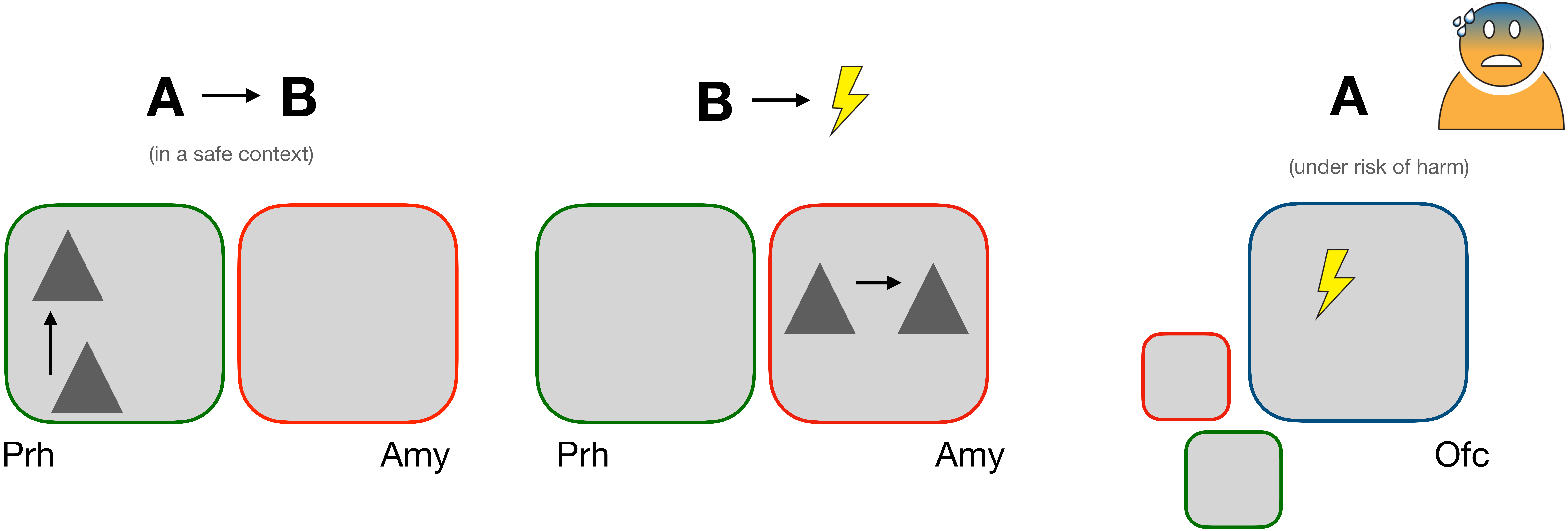


Pre-conditioning phase
Sound for 30 sec
Light for 10 sec
The offset of one is the onset of the other

Inference test
Stimuli presented alone
Sound for 30 sec
Light for 10 sec

Wong, Westbrook, Holmes, 2019 (eLife)

The machinery…

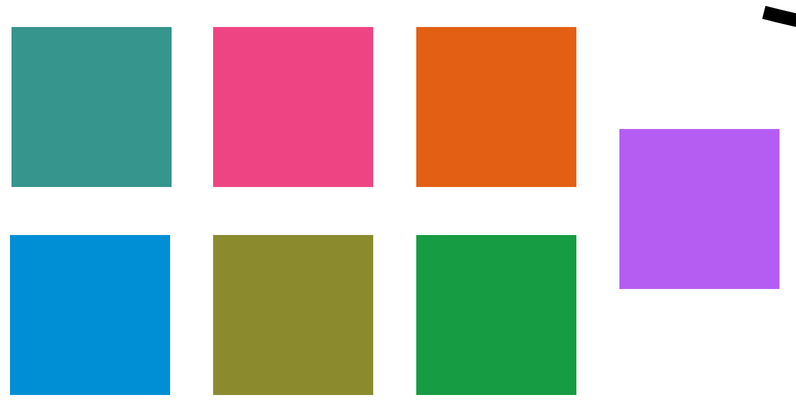


cf. Wong, Westbrook, Holmes, 2019 (eLife)
Jones et al., 2012 (Science)

Combining memories across context to estimate imminence of danger/harm

Procedure

Color stimuli familiarisation



Eye-tracker & skin conductance hook-up (~10 min)

(I) Sequence learning (~45 min)

Shock calibration (~10 min)

(II) Conditioning (~15 min)

"Inscapes" (~5 min)

(III) Inference task (~15 min)

(IV) 3-AFC task (~5 min)

Limesurvey (~10 min)

- Intolerance of uncertainty
- STAI-T

Sequence learning

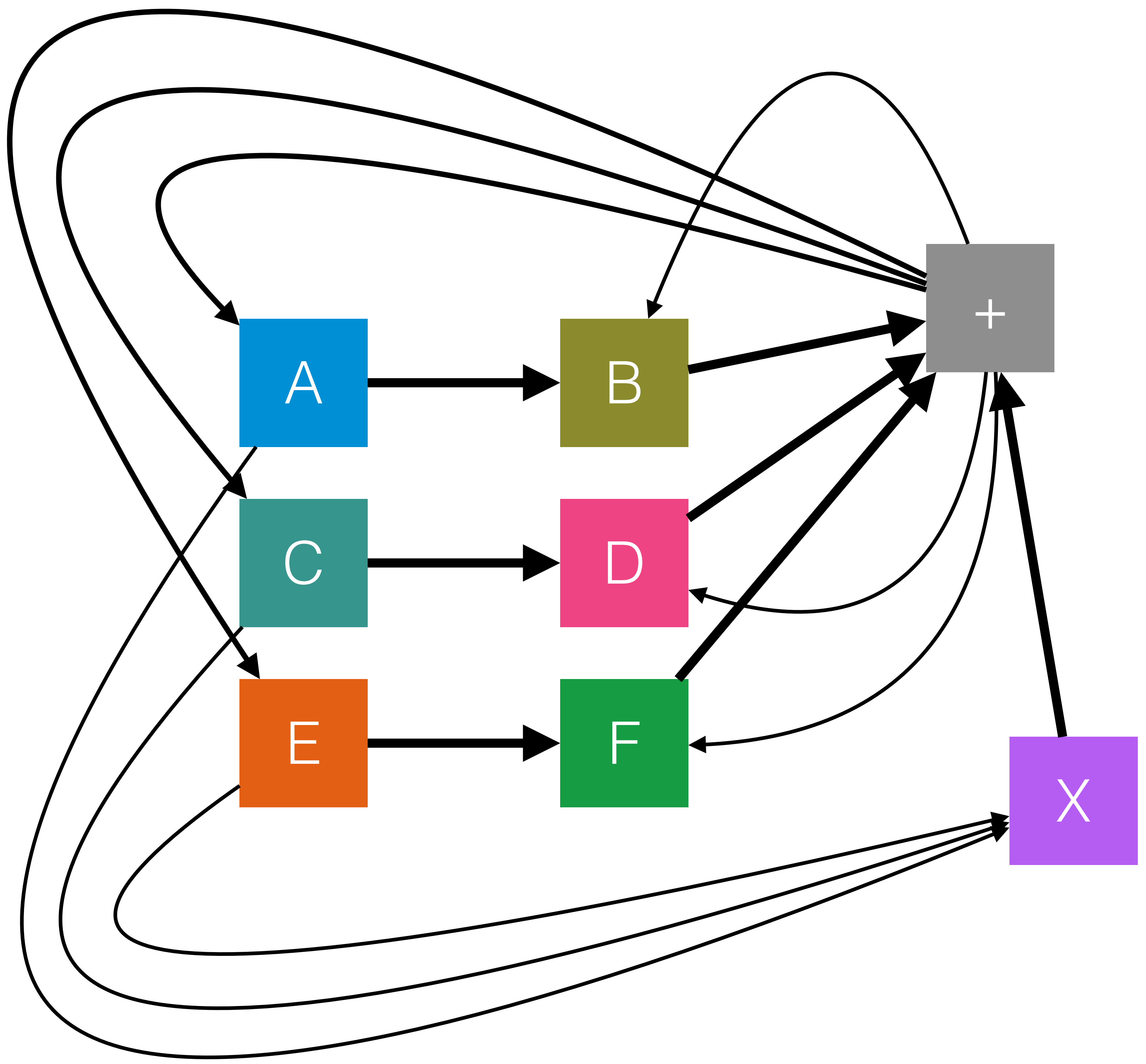
Contingencies

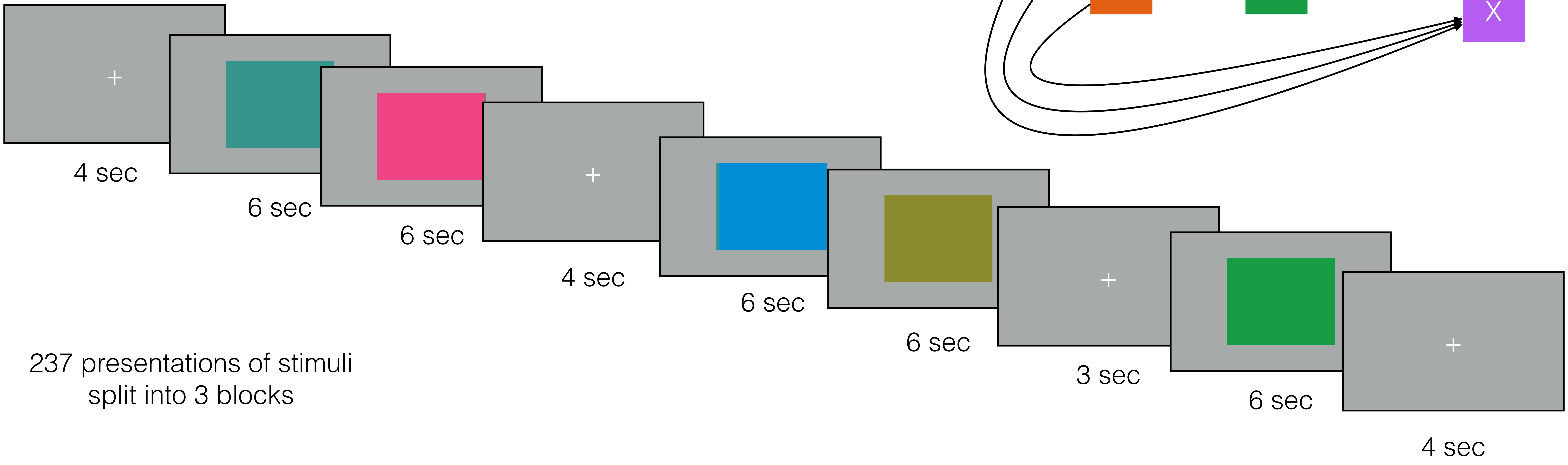
	A	B	C	D	E	F	X	+
A		.8					.2	
B		.						1
C				.8			.2	
D								1
E						.8	.2	
F								1
X								1
+	.28	.06	.28	.06	.28	.06		

Number of transitions

	A	B	C	D	E	F	X	+
A		24					6	
B		.						31
C				24			6	
D								31
E						24	6	
F								31
X								18
+	41	8	41	8	41	8		

Sequence learning

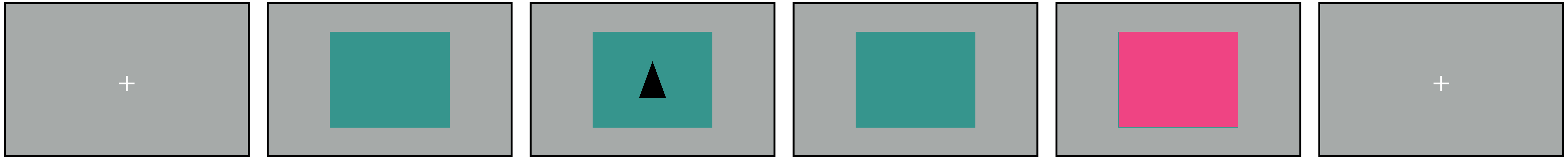




237 presentations of stimuli
split into 3 blocks

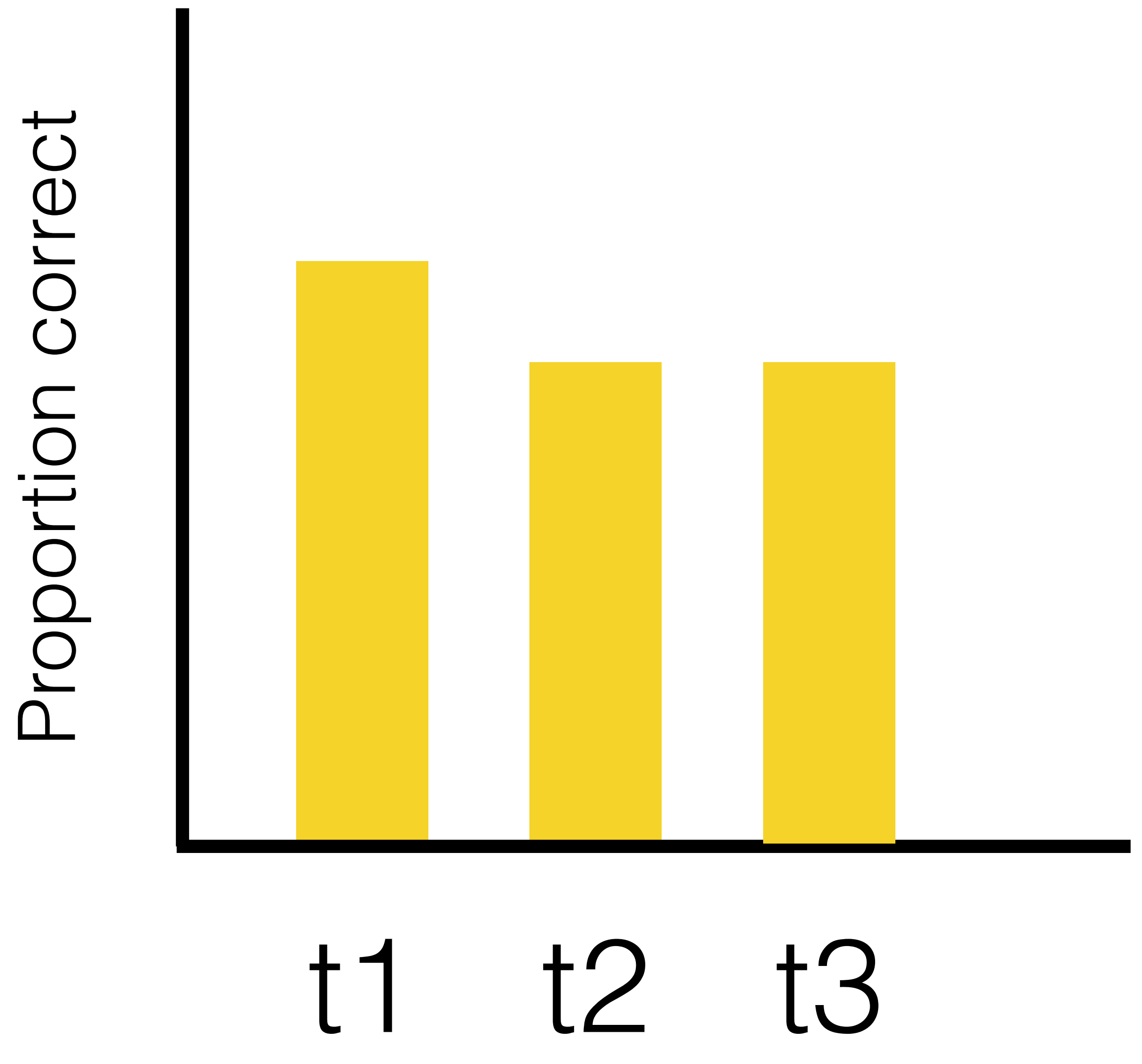
Sequence learning
- target detection sub-task

~50% of stimulus presentations
anytime between 0.5 and 5.5 sec



4 sec 4 sec 0.2 sec 1.8 sec 6 sec 4 sec

Descriptive model of accuracy



$$y_{ij} \sim Binomial(1, p_{ij})$$

$$logit(p_{ij}) = \beta_0 + \omega_{0j} + (\beta_1 + \omega_{1j}) \times x_{1i} + (\beta_2 + \omega_{2j}) \times x_{2i}$$

$$\beta_0 \sim Normal(0,10) \qquad \beta_1, \beta_2 \sim Normal(0,1)$$

$$\begin{bmatrix} \omega_{0j} \\ \omega_{1j} \\ \omega_{2j} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \right) \qquad \boldsymbol{S} = \begin{pmatrix} \sigma_{\omega_0} & 0 & 0 \\ 0 & \sigma_{\omega_1} & 0 \\ 0 & 0 & \sigma_{\omega_2} \end{pmatrix}$$

time: [t1, t2, t3]

$$\Sigma = \boldsymbol{S} \times \boldsymbol{R} \times \boldsymbol{S} \qquad \boldsymbol{R} \sim LKJcorr(2)$$

$$\sigma_{\omega_0}, \sigma_{\omega_1}, \sigma_{\omega_2} \sim HalfCauchy(0,1)$$

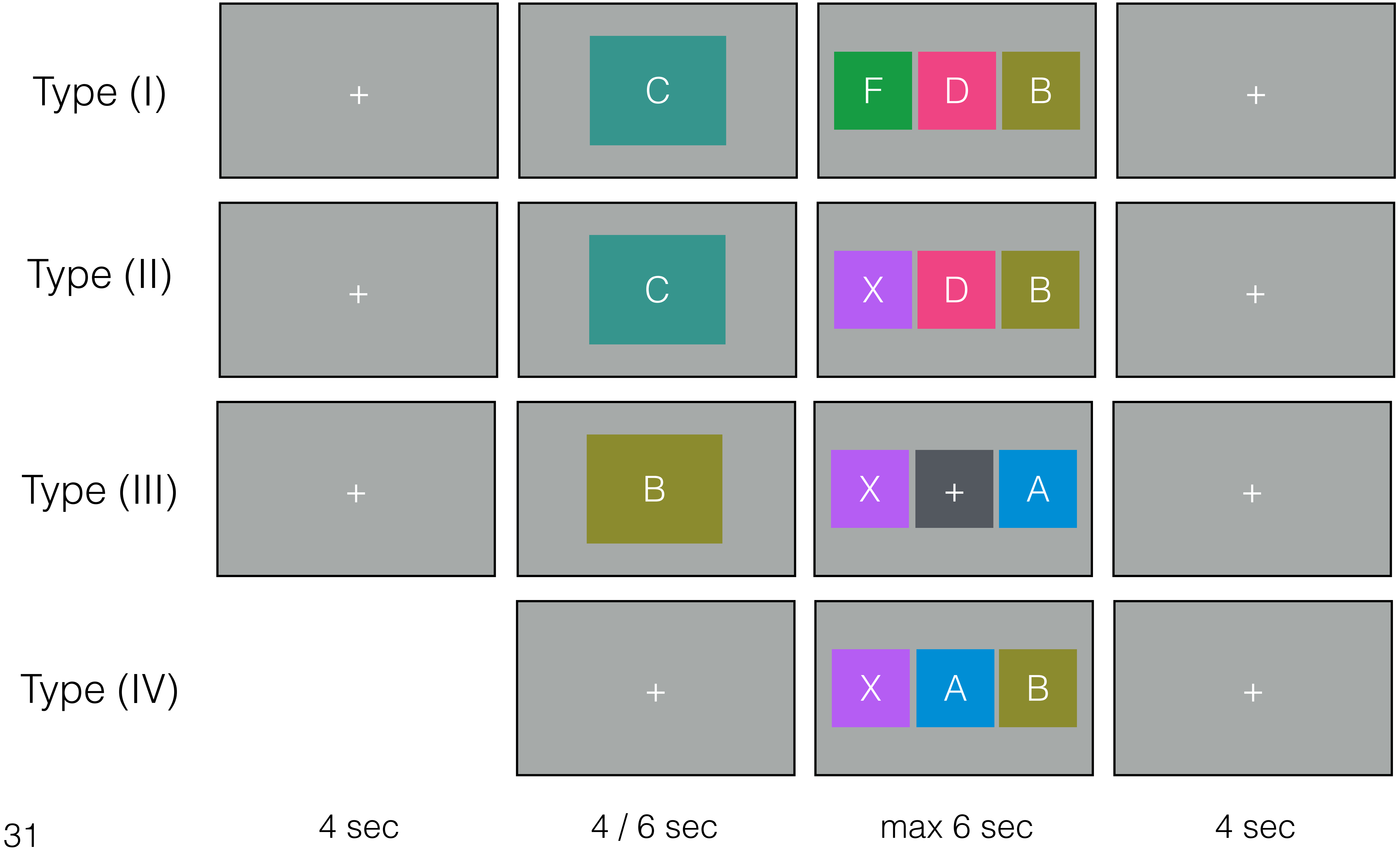
$\beta_0 = \beta_{gandmean}$

$\beta_1 = \beta_{t1-t2}$

$\beta_2 = \beta_{t2-t3}$

Sequence learning
- 3-AFC sub-task

42 questions in total
“What would have been the most likely now?”



	A	B	C	D	E	F	X	+
A		.8					.2	
B		.						1
C				.8			.2	
D								1
E						.8	.2	
F								1
X								1
+	.28	.06	.28	.06	.28	.06		

Cognitive model of SR and 3-AFC

$$y_{ij} \sim Categorical([\pi_1, \pi_2, \pi_3]) \qquad [\pi_1, \pi_2, \pi_3] = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad z_t = M_{tj}[x_{t-1}, q_t]$$

$$M[x_t, :]_{tj} = \begin{cases} M[x_{t-1}, :]_{i-1j} + \alpha_j(onehot(x_t) + \gamma_j M[x_t, :] - M[x_{t-1}, :]), & if\, x_t < 9, t > 0 \\ M[x_t, :]_{t-1j}, & if\, x_t = 9, t > 0 \\ M[zeros, zeros], & if\, t = 0 \end{cases}$$

$$\alpha_j \sim Normal(\mu_\alpha, \sigma_\alpha)T(0,1) \qquad \gamma_j \sim Normal(\mu_\gamma, \sigma_\gamma)T(0,1)$$

$$\mu_\alpha, \mu_\gamma \sim Uniform(0,1) \qquad \sigma_\alpha, \sigma_\gamma \sim HalfCauchy(0,1)$$

Testimonials:
Please describe in your own words what the rule was in the first task. In what order were the colors presented?

“1. first green, then blue 2. first purple , then pink 3. first olive green, then turquoise”

“the task was unclear”

“After the darker green comes turquois or orange.
After the blue comes pink or orange. After the purple comes green or orange”

“Two colors were usually connected.
Though I was unsure, when purple was introduced, when this appears.
Otherwise the sequence of pairs seemed very random to me”

“During the first sequence the color sequences were trained and queried.
I think the sequence was:
turquois, mustard yellow, +, green, magenta, +, purple, + ?.
But I am not so sure, because the memory of the first task is blurred together with the others.
I know, that in the first task no orange-red occurred.”

“no idea”

“I don’t know”

“Always two color sequences:
Orange-mud/Blue-purple/pink-green”

“I think the rule is, Orange-Blue, Green-pink,
but it seems to be, that the rule sometimes changes
and the option exists, that after one color more then one option exists."

“bit by bit the colors got stronger (related to the contrast)”

“So I think it follows the rule, that the location (”Stättung”) changes?
I am not sure, so maybe the same shade too or something in the direction.”

“I am not completely sure, but I think, that there were related color-pairs or -trios that were showed one after the other.
Beginning with a glaring color like pink or orange, after that a calmer color like blue, green etc. “

“according to gradation”

“you had to guess which color appears”

“I am not sure, but I would guess,
that you should specify the subsequent color according to the previous one”

“The colors were successively shown
and after that appeared repeatedly
and you had to estimate, which color is shown”

“always two same color pairs”

Testimonials:
What strategy did you use to learn and memorize the order of the colors in the first task?

“I memorized the first letter of the colors and formed words.”

“no strategy”

“I learned abbreviations of every colour, unfortunately that didn't help me much”

“counting”

“intuitive”

“none”

“I tried to say the colors in the sequence loud in my head e.g. pink, blue, purple etc.”

“I didn't use a deliberate strategy, I just realised that I read the colors to myself in my head”

“no special strategy. I tried for a few rounds to guess the second color, even if guessing wasn´t my task.”

“called the colors in my head and repeated them”

“just memorized the sequence. No special strategy”

“I repeated the names of the colors in my thoughts.
That was partly difficult because magenta and purple are quite similar”

“I tried to remember chronologically in wich sequence, the colors appeared”

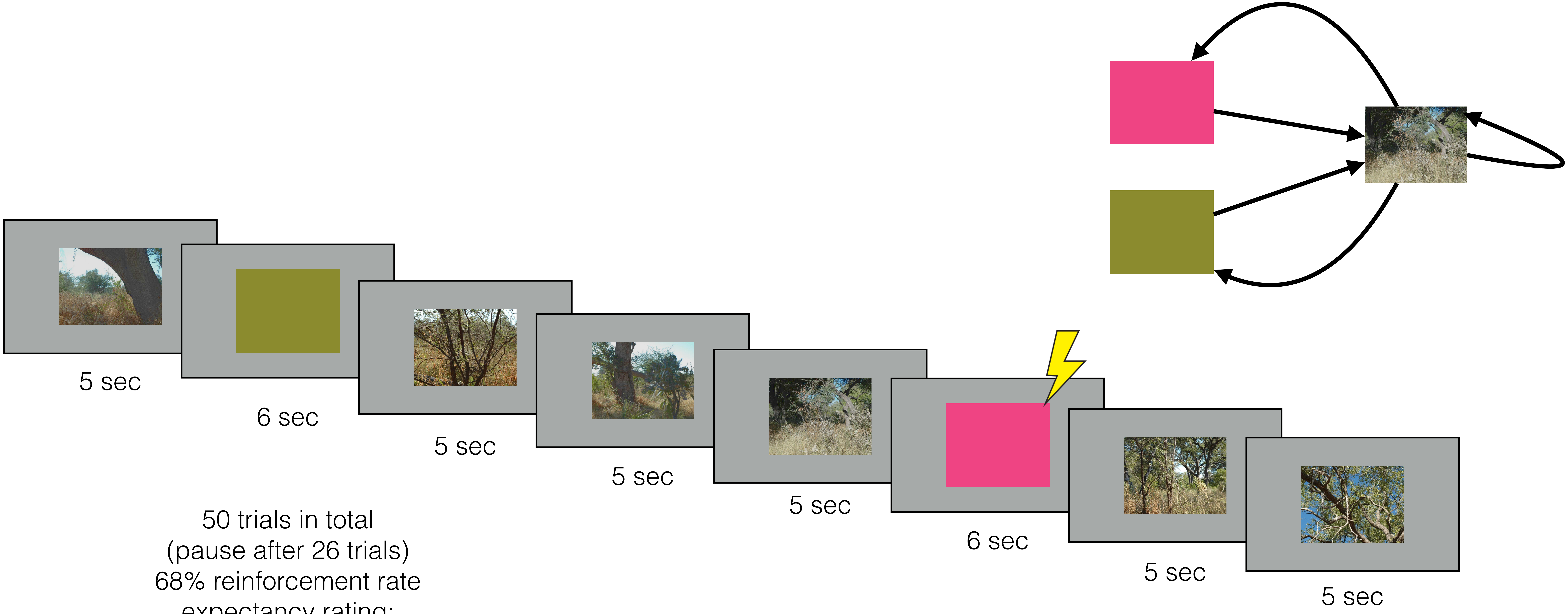
“I used no strategy, but tried to develop a feeling”

“simple remembering of color sequences”

“I have tried to remember the possible combinations color sequences.”

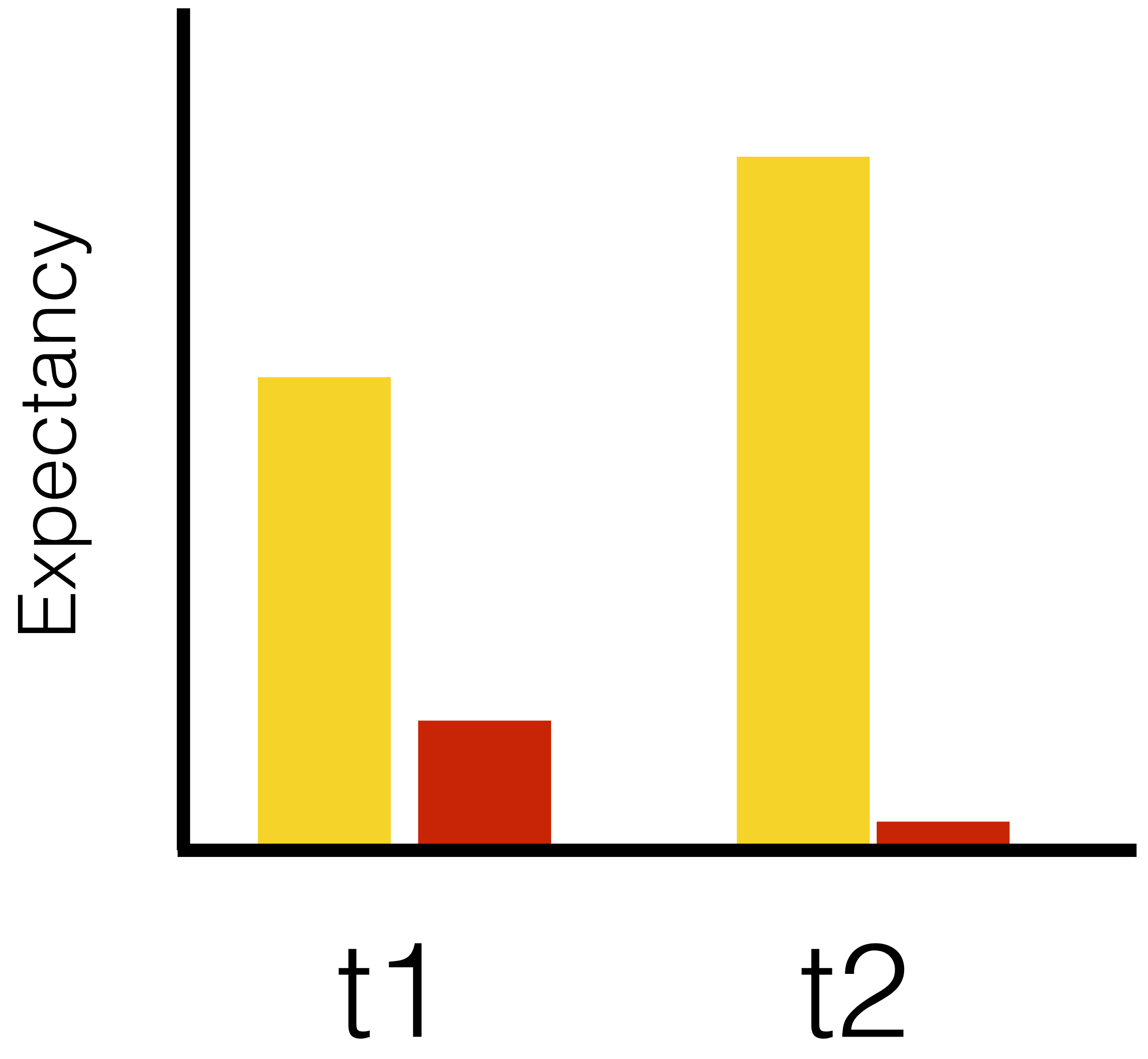
“I first tried to remember the first color, and to find what comes afterwards.
But this strategy isn't very good and correct , because I found, that there are multiple possibilities"

Pavlovian delay conditioning



50 trials in total
(pause after 26 trials)
68% reinforcement rate
expectancy rating:
dot -> square
between 1 and 3 sec after the stimulus onset

Descriptive model of expectancy nested effects



time: [t1, t2]

condition: [CS+, CS-]

$$\beta_0 = \beta_{gandmean}$$
$$\beta_1 = \beta_{time}$$
$$\beta_2 = \beta_{time1 \times condition}$$
$$\beta_3 = \beta_{time2 \times condition}$$

$$y_{ij} \sim Binomial(1, p_{ij})$$

$$logit(p_{ij}) = \beta_0 + \omega_{0j} + (\beta_1 + \omega_{1j}) \times x_{1i} + (\beta_2 + \omega_{2j}) \times x_{2i} + (\beta_3 + \omega_{3j}) \times x_{3i}$$

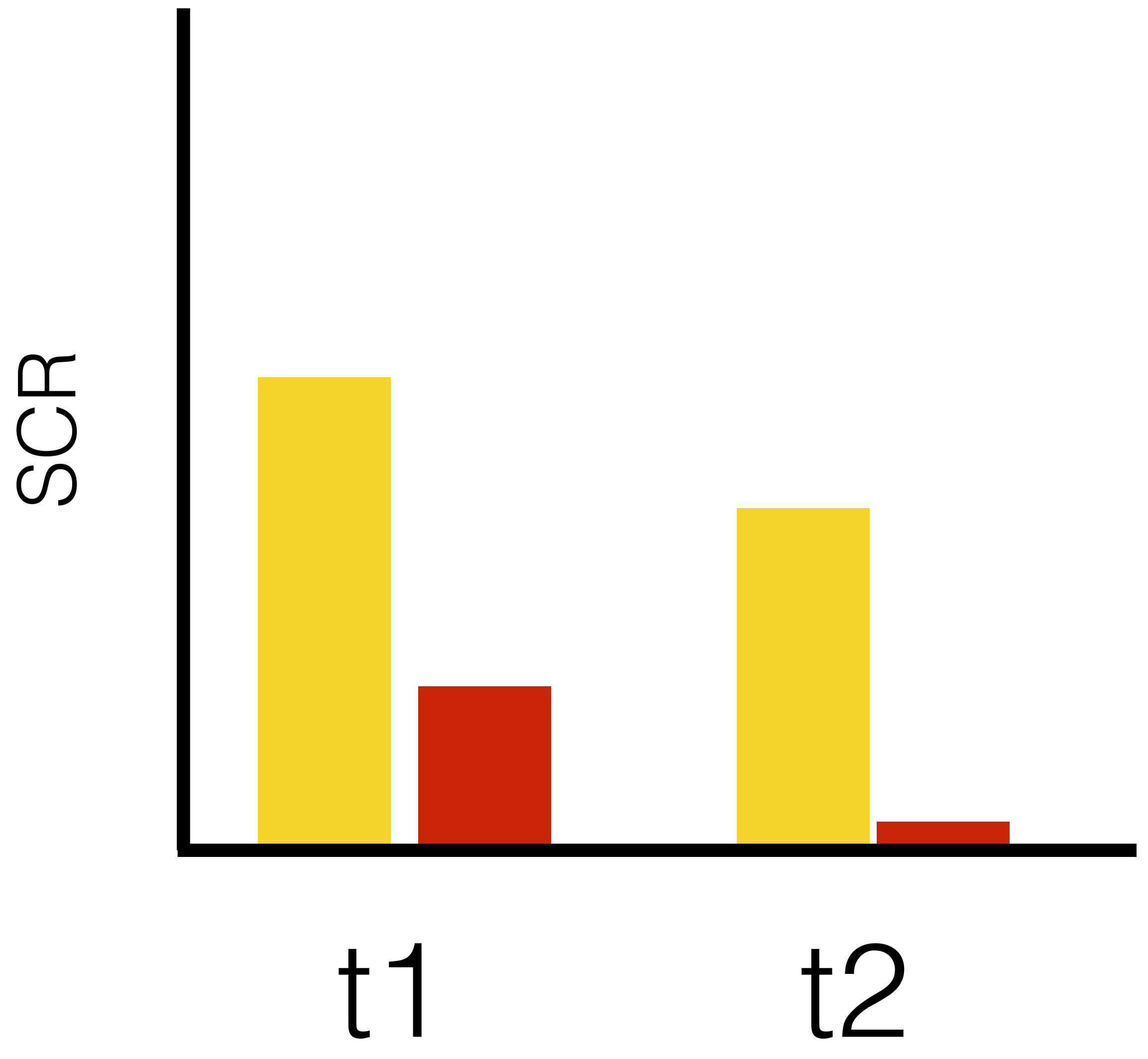
$$\beta_0 \sim Normal(0,10) \qquad \beta_1, \beta_2, \beta_3 \sim Normal(0,1)$$

$$\begin{bmatrix} \omega_{0j} \\ \omega_{1j} \\ \omega_{2j} \\ \omega_{3j} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \right) \qquad \boldsymbol{S} = \begin{pmatrix} \sigma_{\omega_0} & 0 & 0 & 0 \\ 0 & \sigma_{\omega_1} & 0 & 0 \\ 0 & 0 & \sigma_{\omega_2} & 0 \\ 0 & 0 & 0 & \sigma_{\omega_3} \end{pmatrix}$$

$$\Sigma = \boldsymbol{S} \times \boldsymbol{R} \times \boldsymbol{S} \qquad \boldsymbol{R} \sim LKJcorr(2)$$

$$\sigma_{\omega_0}, \sigma_{\omega_1}, \sigma_{\omega_2}, \sigma_{\omega_3} \sim HalfCauchy(0,1)$$

Descriptive model of SCR nested effects



time: [t1, t2]

condition: [CS+, CS-]

$\beta_0 = \beta_{gandmean}$
$\beta_1 = \beta_{time}$
$\beta_2 = \beta_{time1 \times condition}$
$\beta_3 = \beta_{time2 \times condition}$

$$y_{ij} \sim Gamma(k, k \div lp_{ij}) \quad k \sim Uniform(0,10)$$

$$exp(lp_{ij}) = \beta_0 + \omega_{0j} + (\beta_1 + \omega_{1j}) \times x_{1i} + (\beta_2 + \omega_{2j}) \times x_{2i} + (\beta_3 + \omega_{3j}) \times x_{3i}$$

$$\beta_0 \sim Normal(0,10) \quad \beta_1, \beta_2, \beta_3 \sim Normal(0,1)$$

$$\begin{bmatrix} \omega_{0j} \\ \omega_{1j} \\ \omega_{2j} \\ \omega_{3j} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \right) \qquad \boldsymbol{S} = \begin{pmatrix} \sigma_{\omega_0} & 0 & 0 & 0 \\ 0 & \sigma_{\omega_1} & 0 & 0 \\ 0 & 0 & \sigma_{\omega_2} & 0 \\ 0 & 0 & 0 & \sigma_{\omega_3} \end{pmatrix}$$

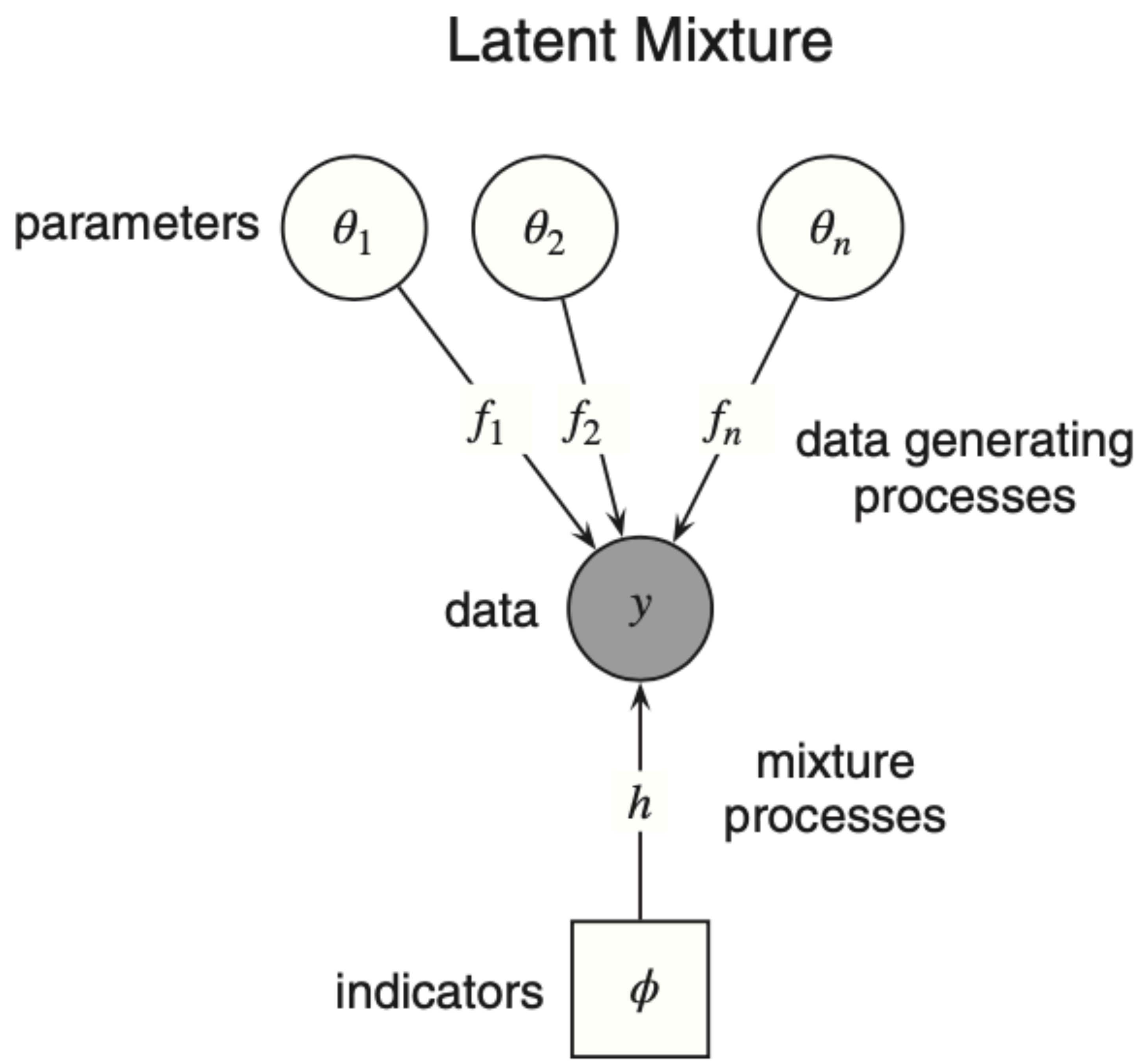
$$\Sigma = \boldsymbol{S} \times \boldsymbol{R} \times \boldsymbol{S} \quad \boldsymbol{R} \sim LKJcorr(2)$$

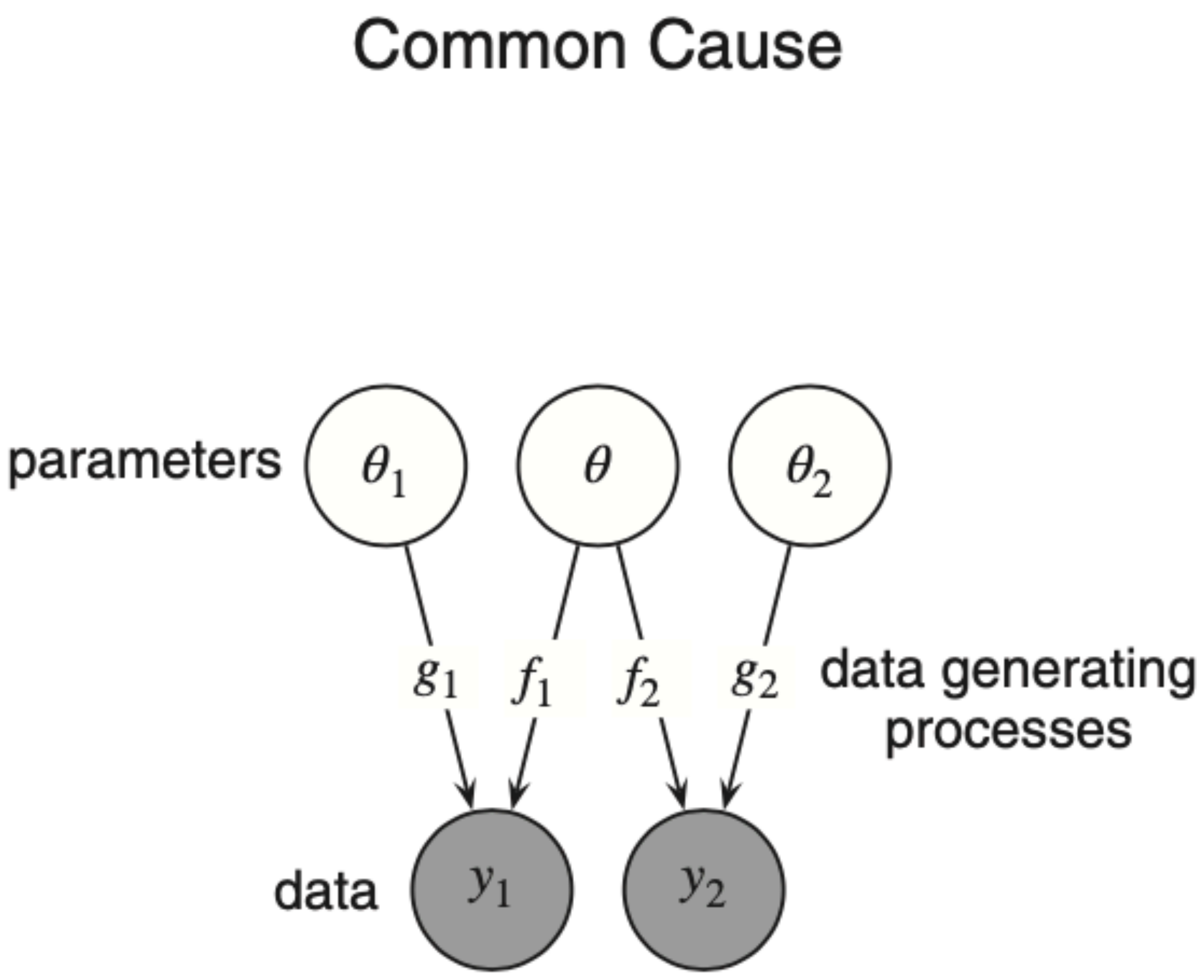
$$\sigma_{\omega_0}, \sigma_{\omega_1}, \sigma_{\omega_2}, \sigma_{\omega_3} \sim HalfCauchy(0,1)$$

Cognitive models of associative learning and (conditioned) responding

$$V_t[x_t] = V_{t-1}[x_t] + \alpha_j(US_{t-1} - V_{t-1}[x_t])$$

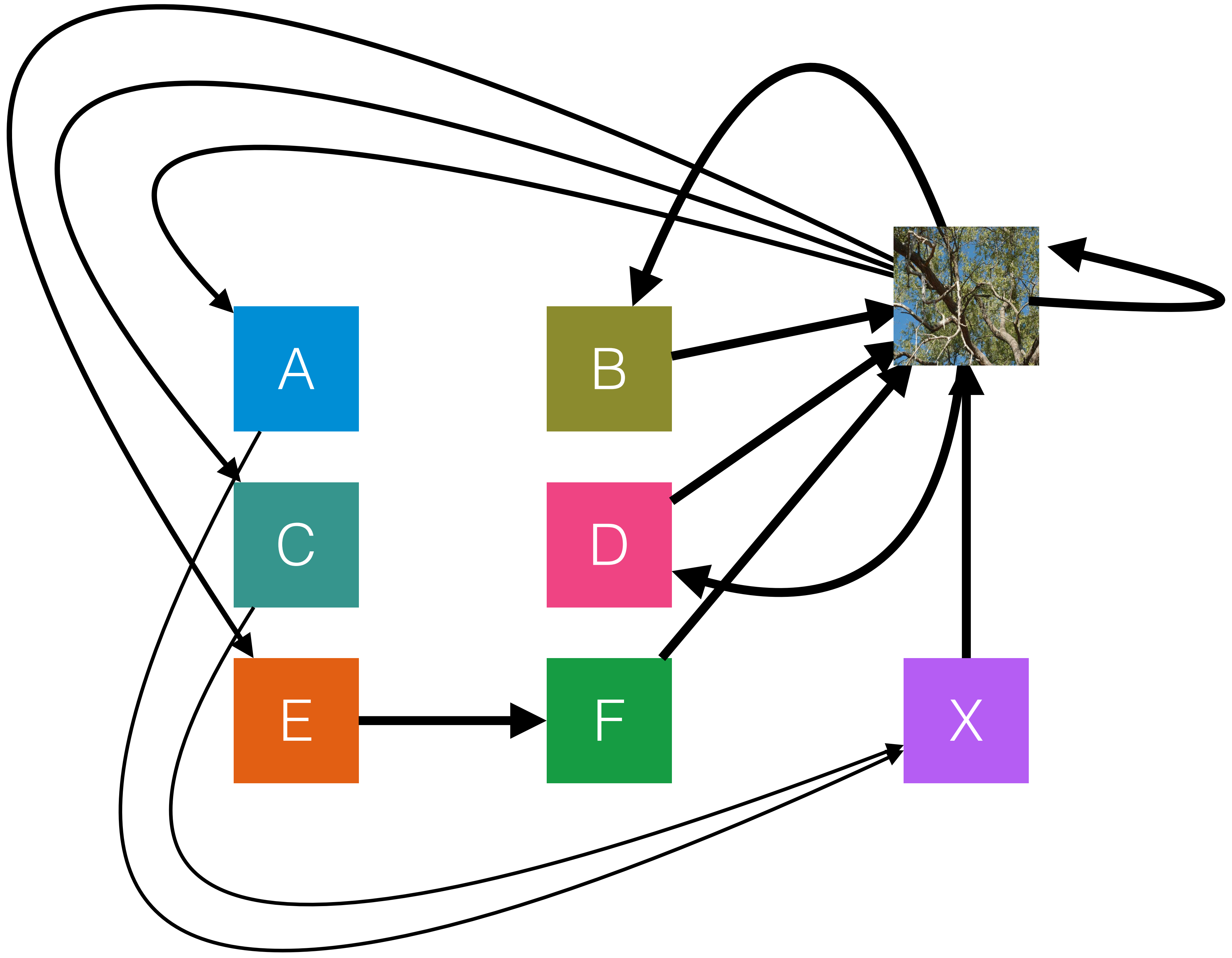
$$\alpha_j \sim Normal(\mu_\alpha, \sigma_\alpha) \quad \mu_\alpha \sim Uniform(0,1) \quad \sigma_\alpha \sim HalfCauchy(0,1)$$

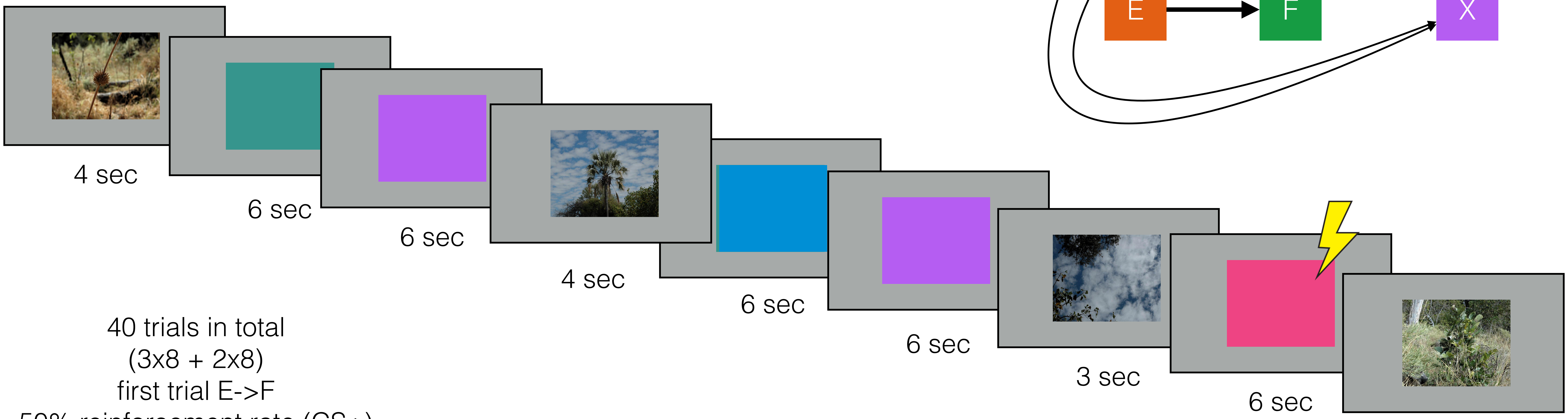




“Inscapes”

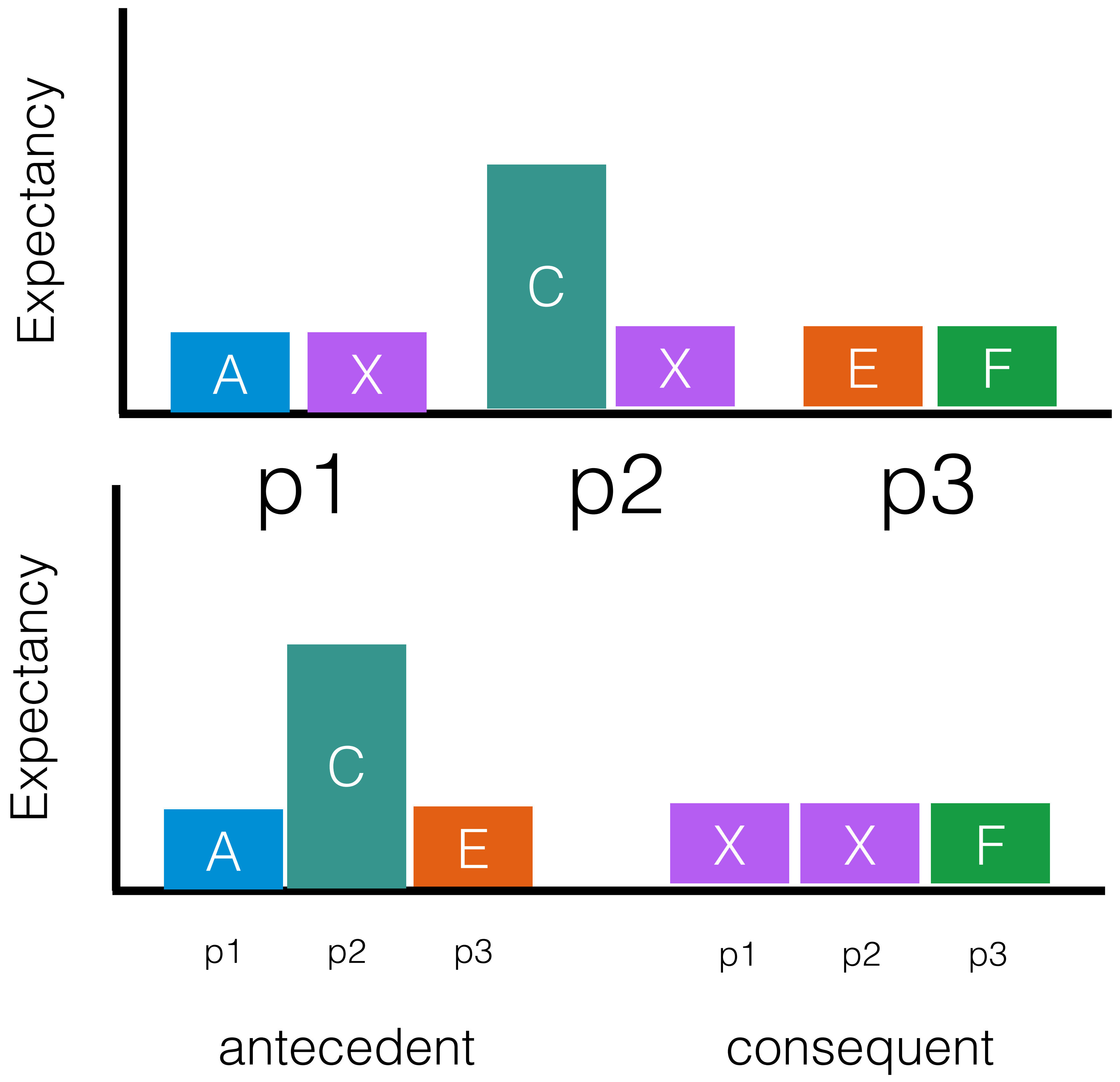
Inference test





40 trials in total
(3x8 + 2x8)
first trial E->F
50% reinforcement rate (CS+)

Descriptive model of expectancy nested effects



pair: [p1, p2, p3]

time: [ant, con]

$\beta_0 = \beta_{gandmean}$ $\beta_2 = \beta_{time1 \times p1vsp2}$ $\beta_4 = \beta_{time2 \times p1vsp2}$

$\beta_1 = \beta_{time}$ $\beta_3 = \beta_{time1 \times p1vsp3}$ $\beta_5 = \beta_{time2 \times p1vsp3}$

$$y_{ij} \sim Binomial(1, p_{ij})$$

$$logit(p_{ij}) = \beta_0 + \omega_{0j} + (\beta_1 + \omega_{1j}) \times x_{1i} + (\beta_2 + \omega_{2j}) \times x_{2i} + (\beta_3 + \omega_{3j}) \times x_{3i} + (\beta_4 + \omega_{4j}) \times x_{4i} + (\beta_5 + \omega_{5j}) \times x_{5i}$$

$$\beta_0 \sim Normal(0,10) \qquad \beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \sim Normal(0,1)$$

$$\begin{bmatrix} \omega_{0j} \\ \omega_{1j} \\ \omega_{2j} \\ \omega_{3j} \\ \omega_{4j} \\ \omega_{5j} \end{bmatrix} \sim MVNormal\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \right) \quad \boldsymbol{S} = \begin{pmatrix} \sigma_{\omega_0} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\omega_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\omega_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\omega_3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\omega_4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\omega_5} \end{pmatrix}$$

$$\Sigma = \boldsymbol{S} \times \boldsymbol{R} \times \boldsymbol{S} \qquad \boldsymbol{R} \sim LKJcorr(2)$$

$$\sigma_{\omega_0}, \sigma_{\omega_1}, \sigma_{\omega_2}, \sigma_{\omega_3}, \sigma_{\omega_4}, \sigma_{\omega_5} \sim HalfCauchy(0,1)$$

Cognitive models

- (Latent mixture) model on whether SR matrix is transferred to the inference task and affect the expectation of the shock
- Update SR and associative strength of the US

Testimonials:

Please describe in your own words what the rule was in the tasks with shocks. In the shock task with only two colors: When was the shock given? In the shock task with many different colors: When was the shock given?

"I can't answer that"

"don't know"

"orange - after two pictures"

"during green the shock was given, during purple not.
during the many colors I don't know but rather during green"

"The shock was given during turquois, not during pink. But not always during turquois"

"turquois"

"at purple"

"Green: Shock , during other colors: no shock"

"During the shock task with two colors the shock was given during the strong blue.
Only in seldom exceptions there were no shocks during the strong blue.
During the shock task with many colors it was the same."

"The shock only came during the orange color,
but it didn't come at every presentation of the color. This was the case in both tasks"

"I think during both tasks
it was the same strategy the shock came during the pink color"

"the shock was given during the dark yellow color the first task. In the second task too,
but not always. I don't know what made the difference"

"In the shocktask with two colors the shock was only given during the orange color,
in the task with multiple colors it was the same"

"1. Task: after two and three pictures respectively - purple.
2. Task: after two pictures respectively - purple"

"With the two colors: During turquoise a shock was usually given,
why sometime there was no shock during turquoise, I haven't found out.
during the other color there was no shock. With the many colors:
shocks were only during turquoise given, though also not every time and again I didn't know,
why there was sometimes no shock during turquoise.
During other colors there was never a shock."

"The shock always came like a surprise. I never knew exactly,
when I would get a shock. I was the same in the task with the many colors."

"the shocks only came during the purple.
In the task with two colors during the purple after two savanah pictures.
During the shock task with the many colors during every second purple coler"

"At the two colors: during mustard yellow there was never a shock. during purple in 85-90% there was a shock. When I was quick enough during purple and and clicked before the square that a shock comes, no shock came. Maybe this was only coincidence. When I did this again during purple (because I didn't want to get shocked), I got a shock again. During the task with the many colors, there were also only shocks during purple. The other colors promised safety. During purple, there were no shocks when you clicked left before the square came and during the next appearance of purple when you clicked left, during the time the square was there. Just like in the first case maybe this was also only a funny coincidence."

Testimonials:

In the tasks with shocks, what did you think about when you had to decide whether a shock might come (when the circle in the middle of the screen turned into a square)? Did you make your decision in the same way in both tasks (that is, in the task with only two colors and in the task with many different colors there)? Did you try to predict the shock in advance, i.e., that the shock might occur when the next color image was shown?

“when e.g. 2x there was no shock, should come then as the third, no correlation with colors/pictures recognized"

“I suspected, during the shock task with two colors every 4. color picture ‘Orange’ was left out, but this was not entirely correct (the 5. some times too). during the shocktask with multiple colors, I couldn´t derive when the shock was going to be presented.”

“I thought the shock come every second picture”

“If the color wasn't turquoise then I could decide quickly because shock wouldn't come. If the color was turquoise, then I could only guess.”

“To the colour. At first I was suspicious of whether colors other than turquoise sometimes shocked me. After a while I began to trust that this is not the case. So I tried to find a rule for Turquoise, since the shock is sometimes not used there. I couldn't see through this.”

“I tried to understand, how many savanah pictures I was shown beforehand. I have done it like this in the first shock task. In the second I tried to find out, whether or not I got a shock from the preceding purple color. If not, I pressed left. So I made the decision in different ways, but tried to predict the shock in advance.”

“I based myself on the color based on the previous shocks and made it dependent on whether a shock came or not.”

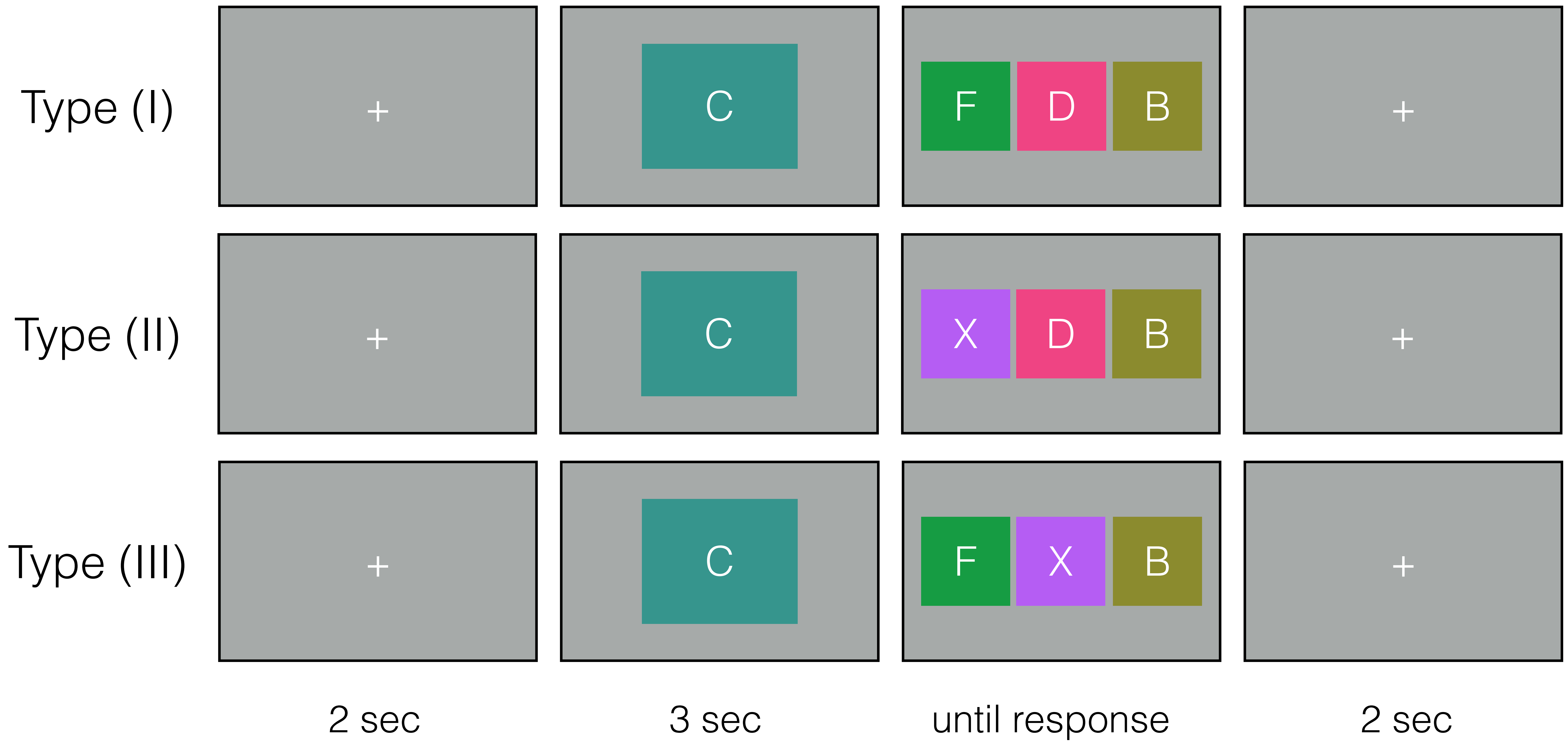
“With the ‘safe colors’ I only clicked when the square appeared. On purple I tried to anticipate and alternately clicked before the square flashed and then with the square appearing.”

“I quickly realised, that the shocks only comes during the pink color. Because of that I was a little bit relaxed, when a different color was shown on the screen.”

“I was more focused with the first task because there were a lot of shocks. With the second task, I got tired and I didn't recognize a concept either. It was often funny that during the green color there were often shocks, but not always, and I only had myself focused on the colors, the images were distracting”

“In the first task I think I first suppressed all the blues, but then realized that only the strong blue causes a shock. I made a slightly different decision in the second task. I think that because of the video shown during the break, I suspected that pink and purple would also give me a shock, based on their interaction with blue in the video. I set myself up for a shock and even when nothing came the first time Pink, I was still a little suspicious.”

3-AFC task
- 48 trials in total



original transition matrix

	A	B	C	D	E	F	X	+
A		.8					.2	
B								1
C				.8			.2	
D								1
E						.8	.2	
F								1
X								1
+	.28	.06	.28	.06	.28	.06		

Plans

- Refine cognitive models of individual tasks
  - latent mixture model on SR task (TD learning vs most recent/random item?)
- Find link function from associative models to psychophysiology index (SCR, pupil) and expectancy
  - ideal bayesian observer
  - other non-probabilistic models
- Consider common cause model for conditioning task with all modalities
  - dependency between SCR and pupil?
- Consider common cause model for multiple tasks
  - SR learning based on target detection or pupil data (attention)
- Follow-up experiments
  - multi-modal stimuli
  - “shorter” imminence
  - modify conditioning such that CS+ is preceded by another stimulus in conditioning