Bayesian learner model of reversal learning

in rodents self-administering cocaine

Modelling Choices:

Bayesian Learner with a Binomial likelihood link:

q_t: likelihood of the right side being rewarded on trial t, $Q_t(R)$

1-q_t: likelihood of the left side being rewarded on trial t, $Q_t(L)$

S_t: Reward sequence on trials t (e.g. R, R, noR, R, R, noR, R, R …)

Value representation

$$P(q_t|S_t) \sim (1-H) * P(S_t|q_t) * P(q_t) + H * P(q)$$

Posterior distribution of q

Likelihood

Prior

Uniform prior distribution over q

"Hazard" or leak rate H

Softmax decision rule: explore vs exploit based on q value above:

Decision probability

$$P(c_t = L|Q_t(L), Q_t(R), L_{t-1}, R_{t-1}) =$$

$$= \frac{exp(Q_t(L)/\beta + \kappa * L_{t-1})}{exp(Q_t(L)/\beta + \kappa * L_{t-1}) + exp(Q_t(R)/\beta + \kappa * R_{t-1})}$$

Model 1: Bayesian Learner | 3 parameters

Trial-wise updating of the expected reward value:

$$P(q_t|S_t) \sim (1-H) * P(S_t|q_t) * P(q_t) + H * P(q)$$

[1] Binomial Likelihood for a reward sequence S_t that has the length n and the number of rewards k:

$$P(S_t|q_t) = \frac{n!}{k! * (n-k)!} * q^k * (1-q)^{n-k}$$

On any trial *t*, a model with a memory size m will have access to the last m trials. In the simplest case, a model will have access only to the last trial (m=1), at which point the likelihood function is a simple Bernoulli distribution, $P(q_t)=q$

[2] Markovian dependency of $P(q_t)$ on $P(q_{t-1})$ is captured by the prior being equal to the posterior of the previous trial:

$$P(q_t) = P(q_{t-1}|S_{t-1})$$
$$P(q_1) = P(q)$$

Model 1: Bayesian Learner | 3 parameters

Trial-wise updating of the expected reward value:

$$P(q_t|S_t) \sim (1-H) * P(S_t|q_t) * P(q_t) + H * P(q)$$

[3] Hazard parameter *H*: determines how much the posterior will change based on the data-driven likelihood vs be resampled from a prior distribution P(q).

Here P(q) was treated as uniform, but it could e.g. be a beta function with [alpha, beta] hyperparameters that could be fitted to data.

Two versions of the model were tested, one where *H* changes on every trial for each subject (Bayesian_Learner_trialwiseH.m) and one where *H* is fitted individually for every subject but stays constant throughout the session (Bayesian_Learner_fit2data.m).

Low *H* will result in quicker adjustment to contingency changes but will also make the model less stable to random perturbations.

[4] Expected value of *q* based on the distribution $P(q_t|S_t)$:

for continuous PDF(q): $E(q_t) = \int q_t * P(q_t) dq$

and for discrete PDF(q): $E(q_t) = \sum_q q_t * P(q_t)$

$E(q_t) = Q_t(L)$

$1 - E(q_t) = Q_t(R)$

Model 1: Bayesian Learner | 3 parameters

Q-Learning (model-free) aka Rescorla Wagner model:

[I] Probability of choice c_t being rewarded *P(q)* at trial *t:*

$$P(q_t|S_t) \sim (1-H) * P(S_t|q_t) * P(q_t) + H * P(q)$$

[II] Probability of choosing c_t at trial *t* (softmax):

$$P(c_t = L|Q_t(L), Q_t(R), L_{t-1}, R_{t-1}) = \frac{exp(Q_t(L)/\beta + \kappa * L_{t-1})}{exp(Q_t(L)/\beta + \kappa * L_{t-1}) + exp(Q_t(R)/\beta + \kappa * R_{t-1})}$$

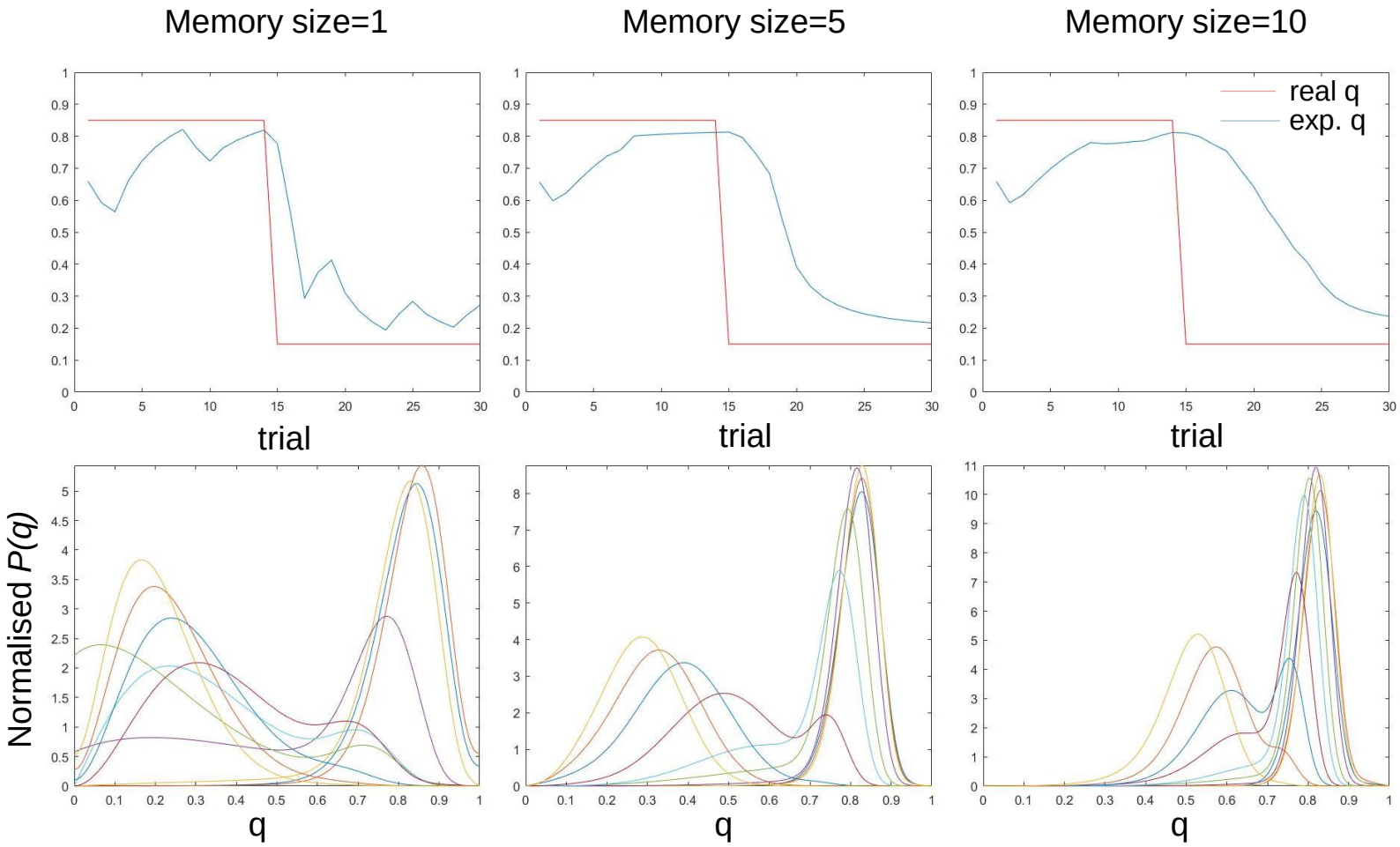
[III] Probability of observing *data D* (a sequence of choices and rewards) = product of the individual probabilities from [II]

$$P(Data\ D|Model\ M, parameters\ \theta) = P(D|M,\theta) = \prod P(c_t|Q_t(L), Q_t(R))$$

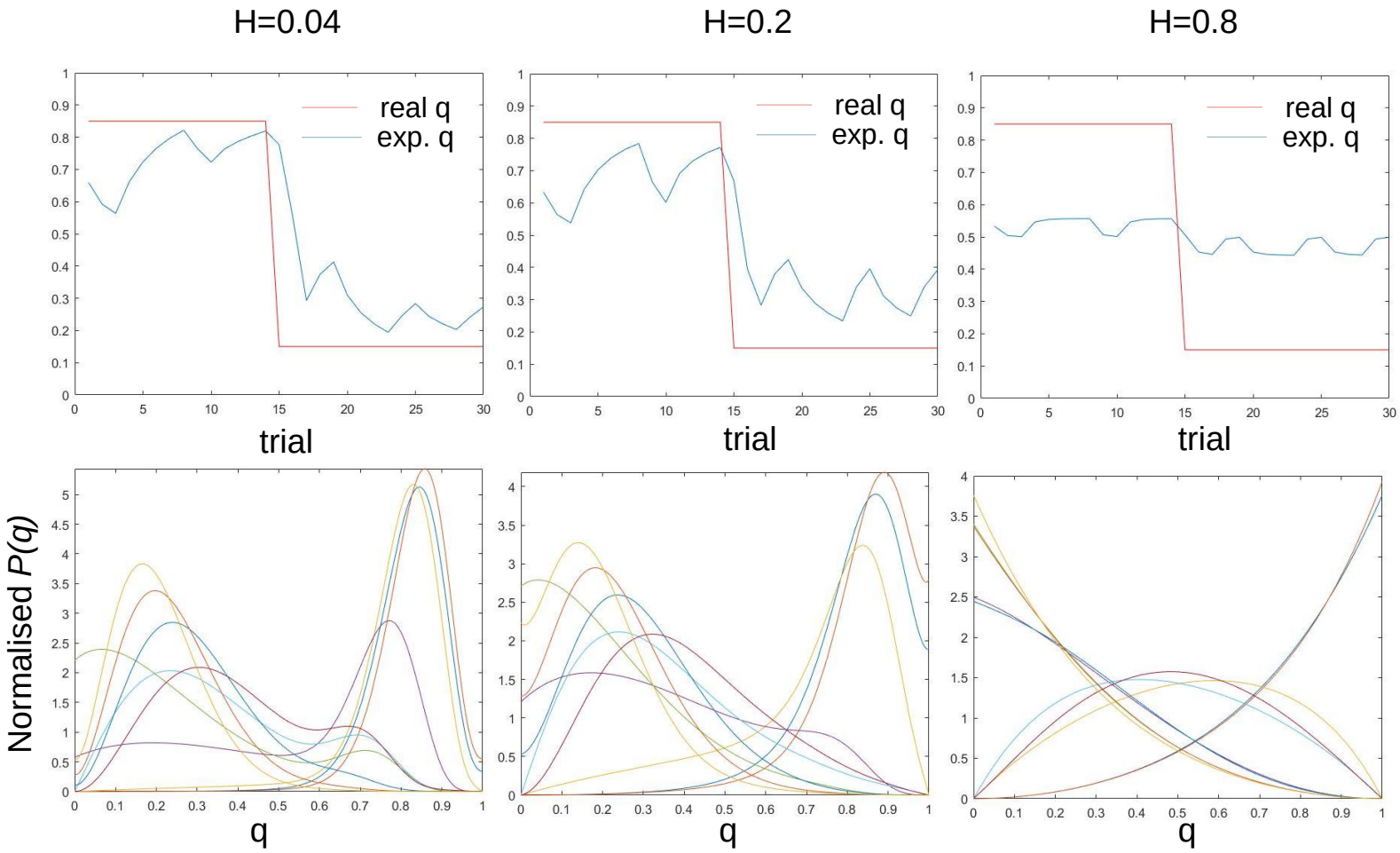
[IV] Fitting parameters [H,β, κ]=θ to achieve maximum likelihood of *data D:*

$$arg\max_{\theta} P(D|M,\theta)$$

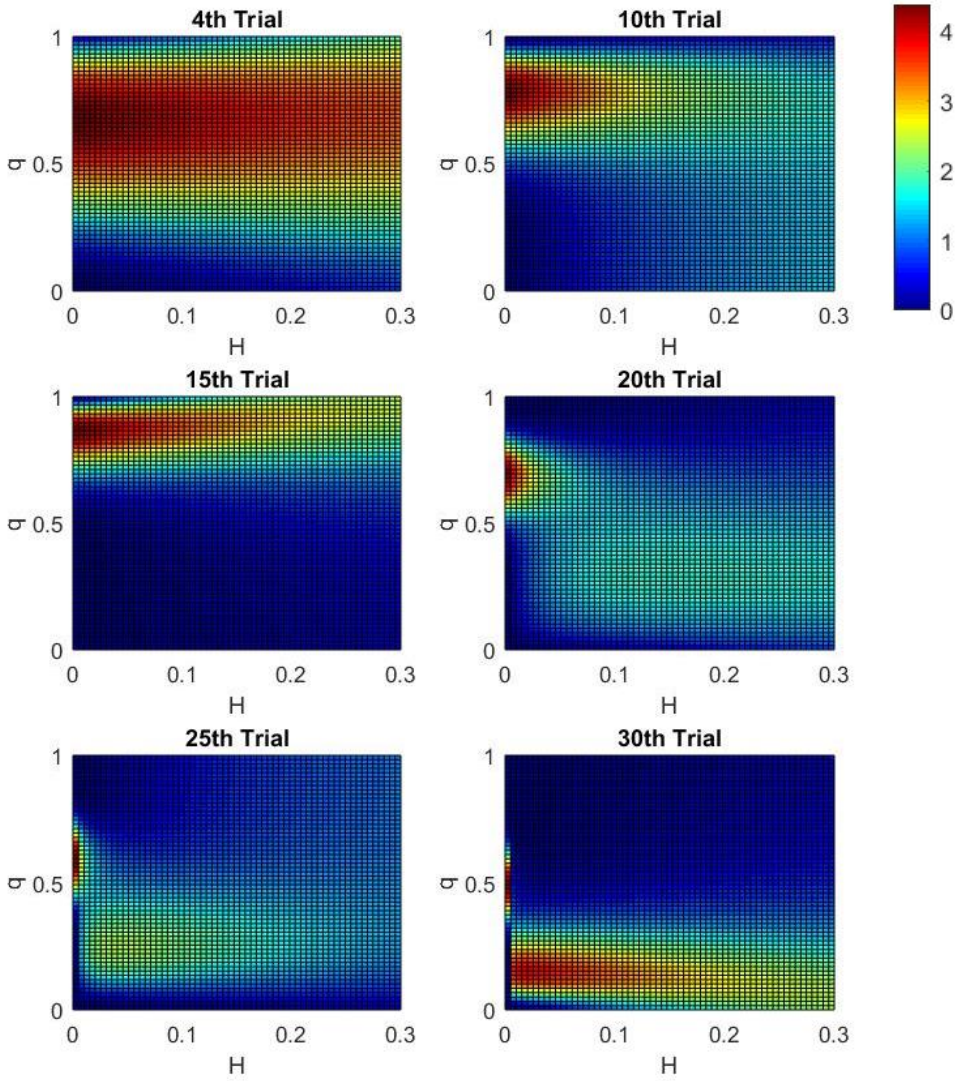
Results: evolution of $P(q_t)$ using different memory sizes:



Results: evolution of $P(q_t)$ using different hazard rates (H):



Results: trial-wise evolution of $P(q_t)$ and H_t:



Normalised posterior probability of q and H

- On the 4th trial the posterior of q contains a lot of uncertainty
- On trial 15, there is a fairly stable expectation of q and expected H is relatively small; a reversal occurs
- Around trial 20, the model is adjusting: the uncertainty or hazard rate is going up; q estimate is shifting from >0.5 to <0.5, i.e. from left to right
- Around 30th trial, a stable expectation of q>>0.5 has formed.

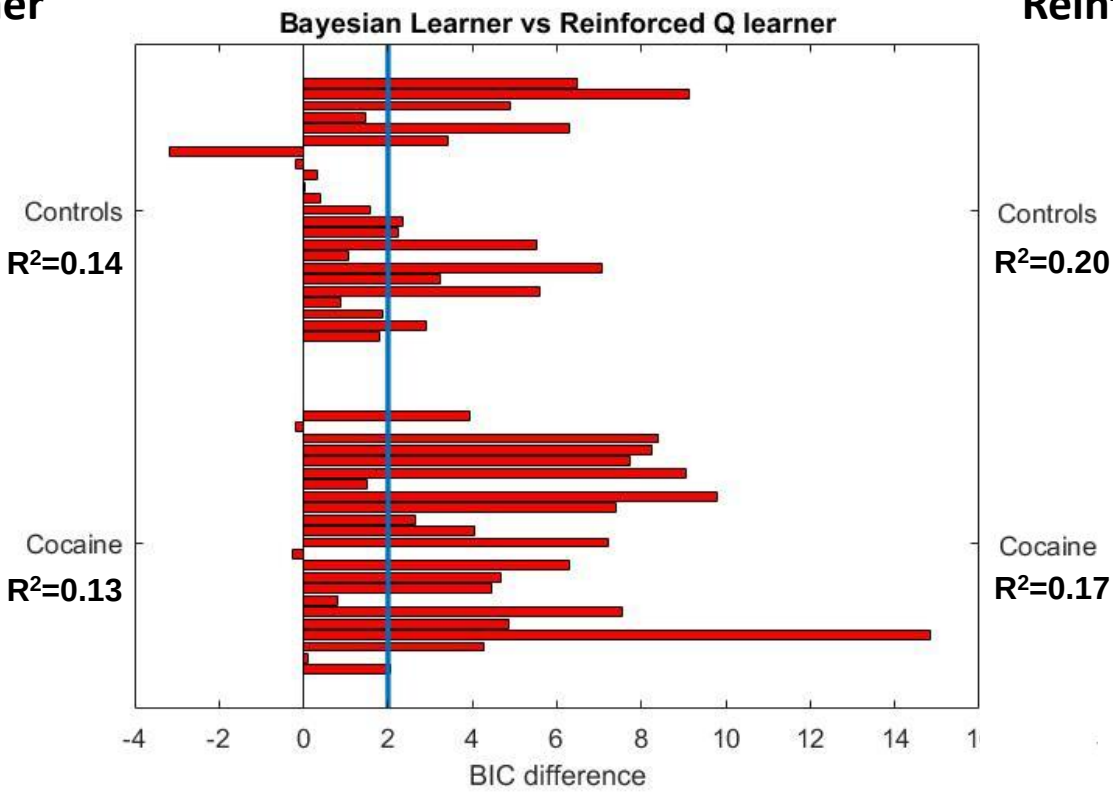
Bernoulli likelihood, i.e. memory size = 1 trial; flat prior

Application:
Rodent reversal data

following cocaine self-administration

Bayesian Learner vs Reinforced Q Learner

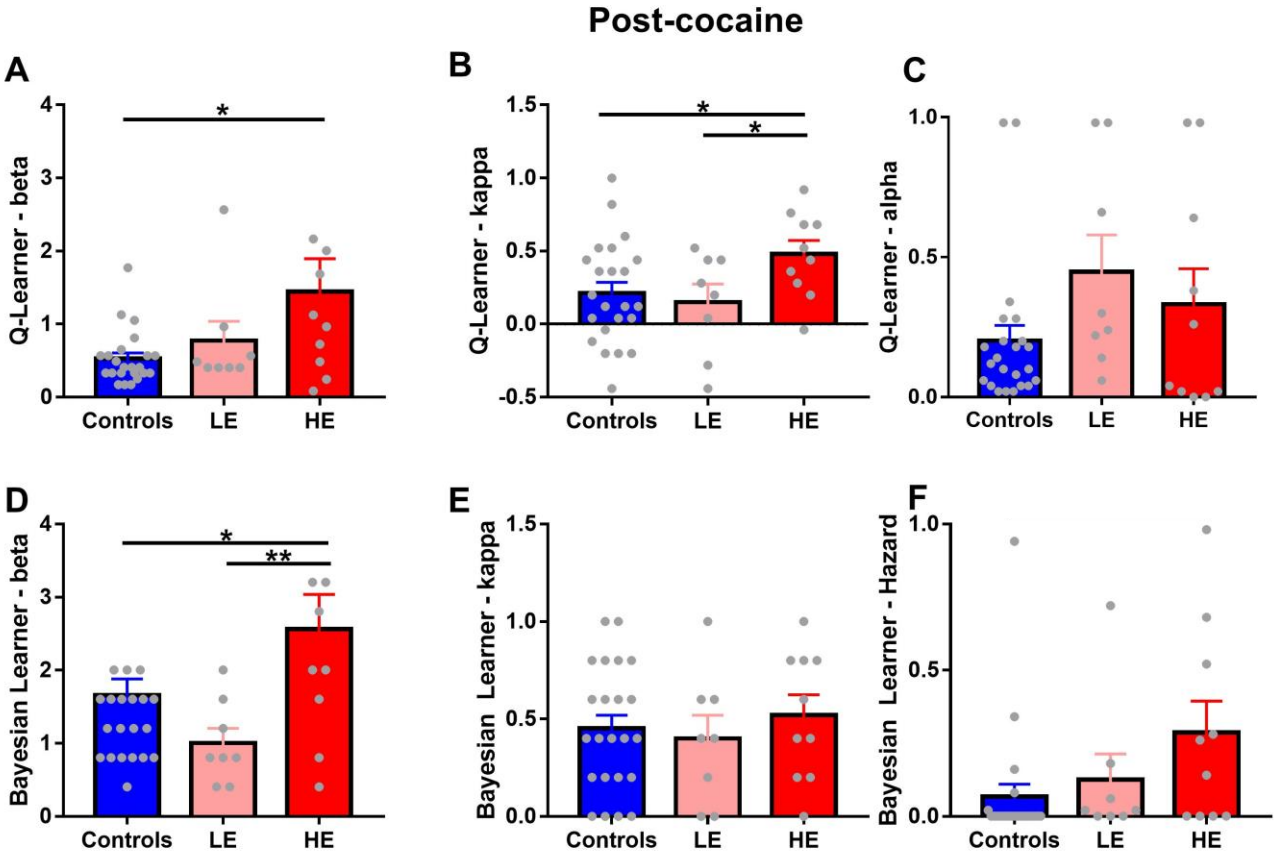
Bayesian Learner



Reinforced Learner

Both models have one free parameter (Hazard *H*, learning rate *alpha*) in the learning part and two free parameters (*beta, kappa*) in the observation part of the model

Model parameter comparison:



Both models show that cocaine escalation produces differences in the observation part of the model (*beta*), suggesting an inability to exploit the learned reward value