The role of glutamate receptor-dependent signaling in the dopamine system in reinforcement learning and adaptive decision-making

Przemysław Eligiusz Cieślak

INSTITUTE OF PHARMACOLOGY
POLISH ACADEMY OF SCIENCES

Two-armed bandit problem

LEFT choice

20%

RIGHT choice

80%

Choice: L L R R R R L L L R R R R L L L R R R

Outcome:

Learned expected value:

$\text{Value}_{\textit{LeftChoice}} < \text{Value}_{\textit{RightChoice}}$

Two-armed bandit problem

LEFT choice

80% $

RIGHT choice

20% $

Negative

!? prediction error

Positive

!? prediction error

Choice:	R	R	R	R	R	R	R	R	L	L	L	L	L	L	L	L
Outcome:	✗	✗	✗	$	✗	✗	✗	✗	$	$	$	$	✗	$	$	$

Updated expected value:

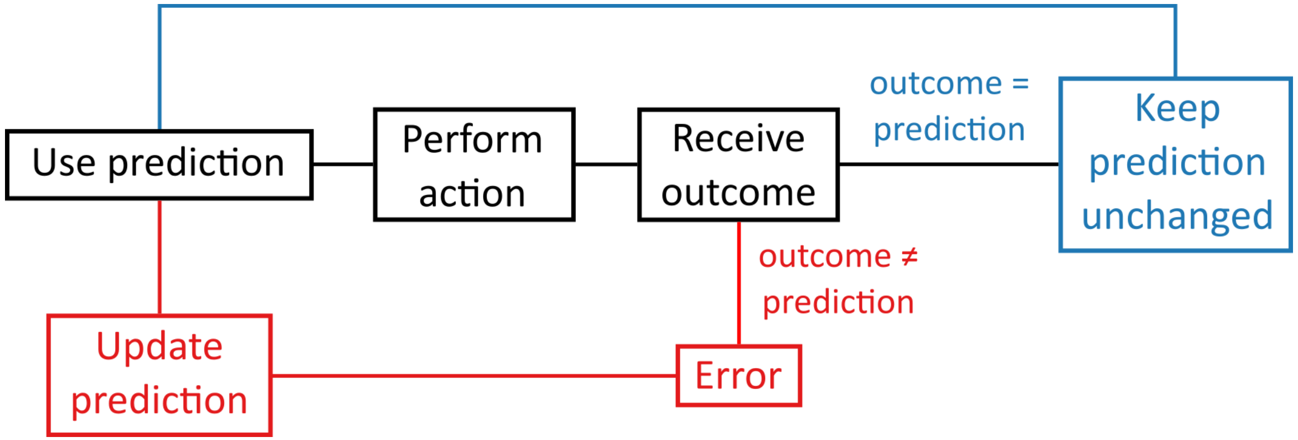
$\text{Value}_{LeftChoice} > \text{Value}_{RightChoice}$

What is reinforcement learning?

Reinforcement learning (RL):

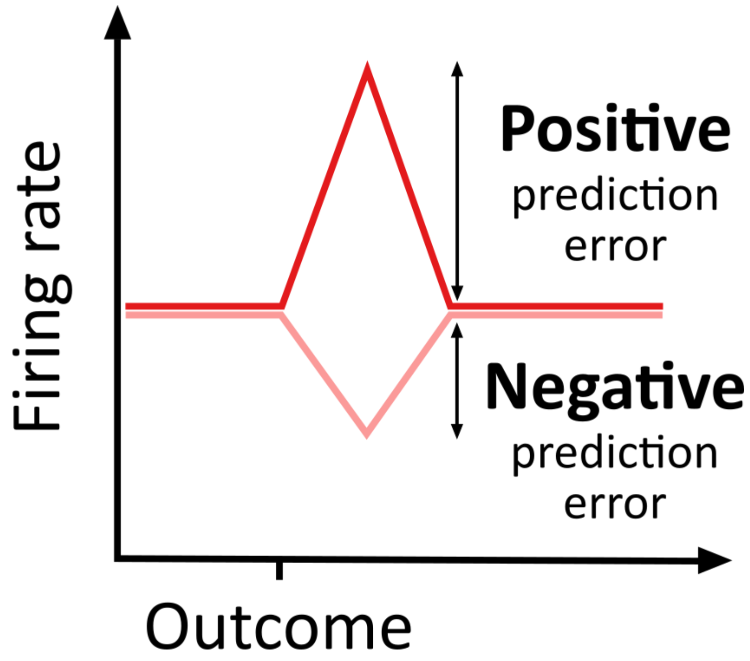
- Learning from the consequences of actions
- Actions are performed in the expectation of a predicted outcomes (expected value)
- Learning to select actions that maximize the accumulated reward over time (with higher value)
- The errors that occurs when the actual outcome differs from what had been predicted are used for updating predictions

- Framework for studying value-based decision-making

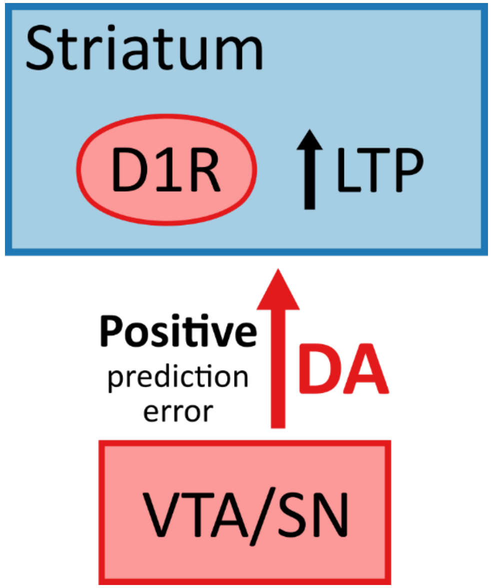


Dopamine system and reinforcement learning

- DA neurons code the discrepancy between the reward and its prediction

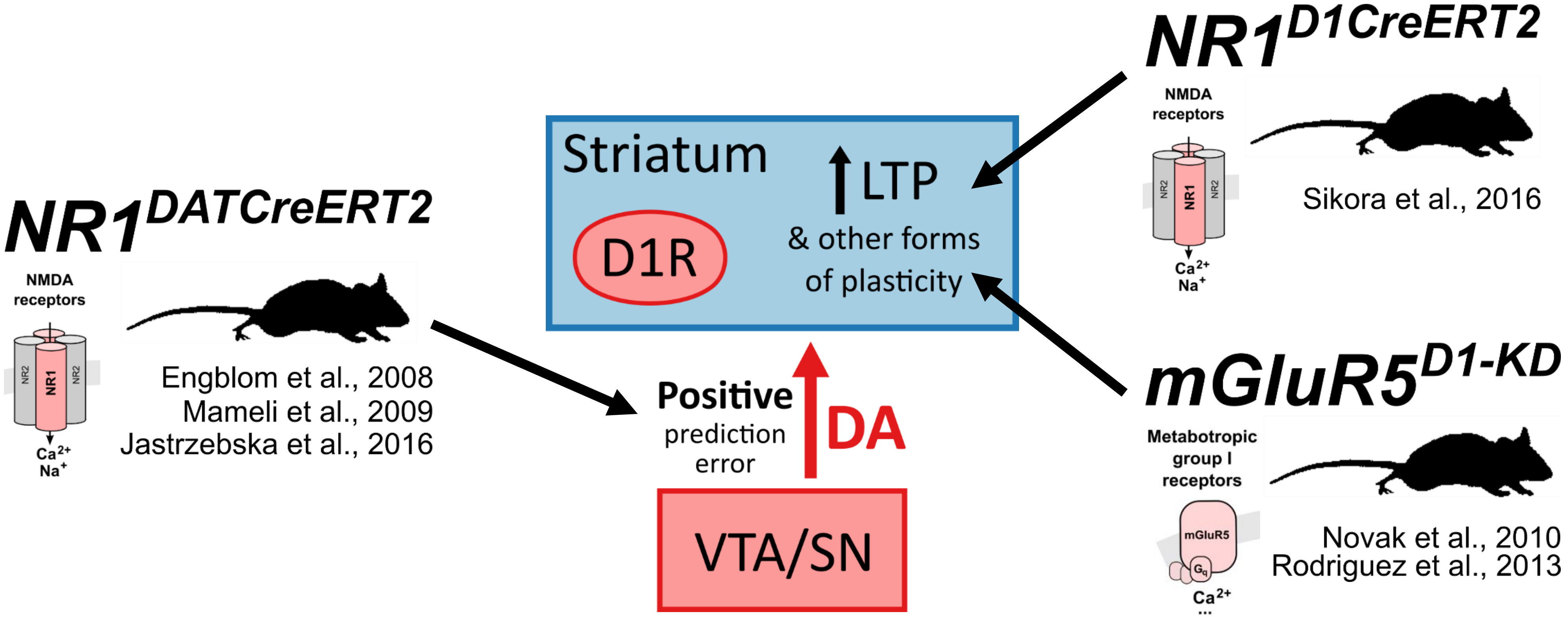


- The **RPE signal** is used for updating the action values stored by striatal neurons
- DA bursts associated with **Positive PEs** potentiate corticostriatal synapses that are active at the time of DA release via D1R

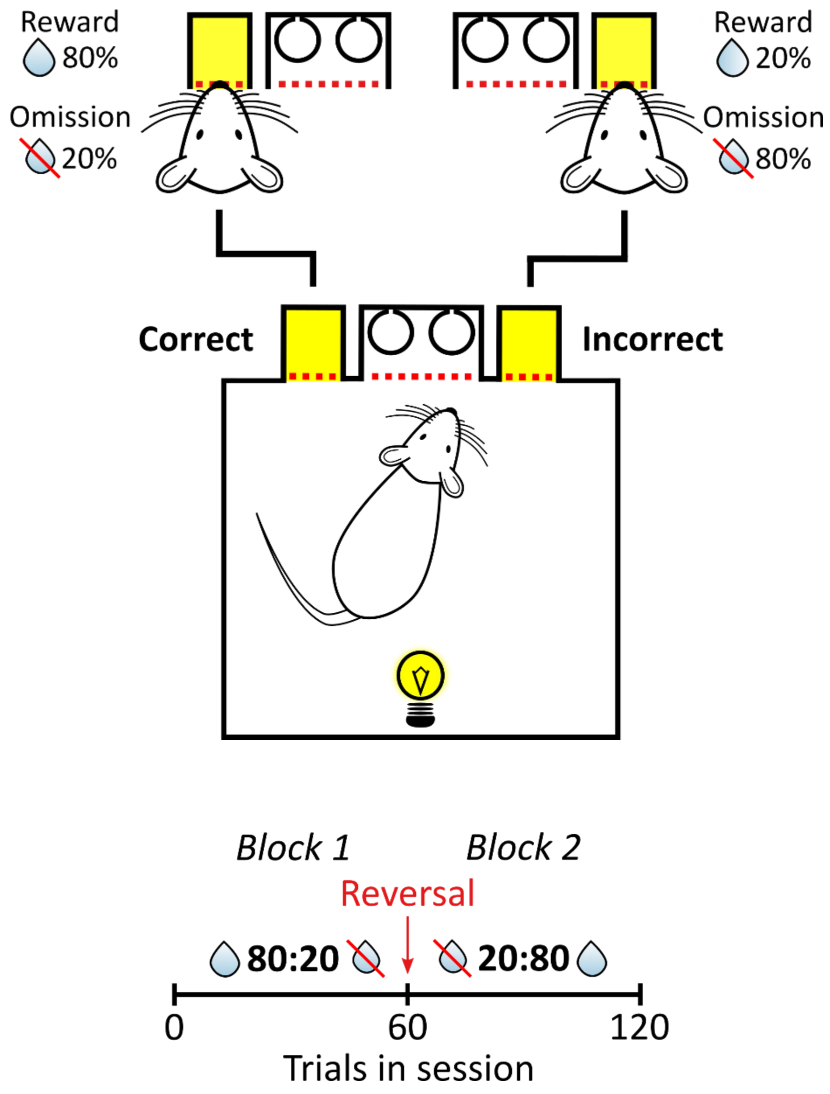


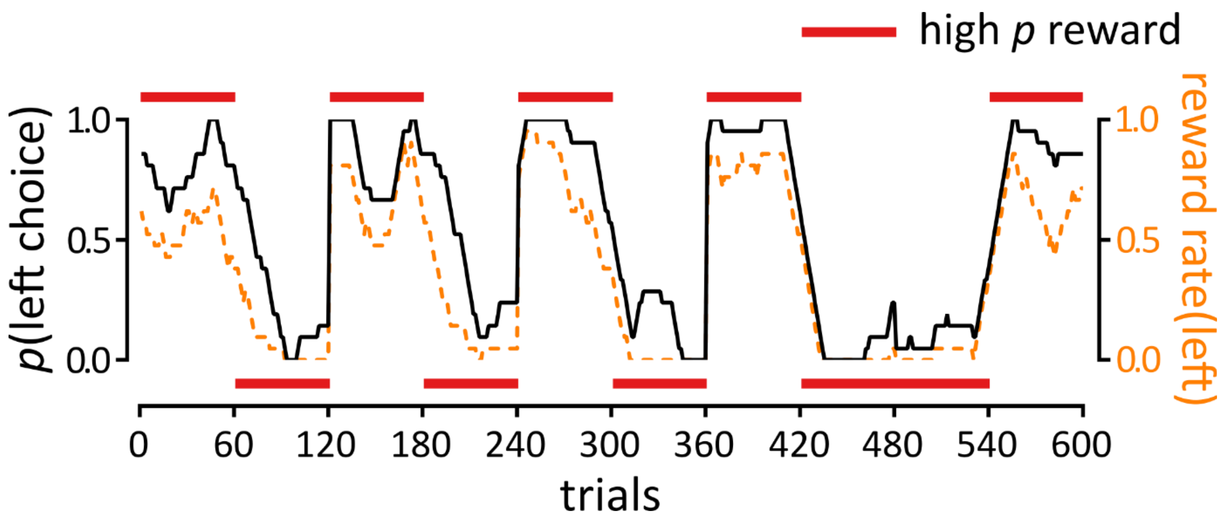
Aim

Use transgenic animals with ablation of Glu receptors in DA and D1R-expressing neurons to investigate the specific role of DA in reward-based learning



Rodent ‚two-armed bandit' task

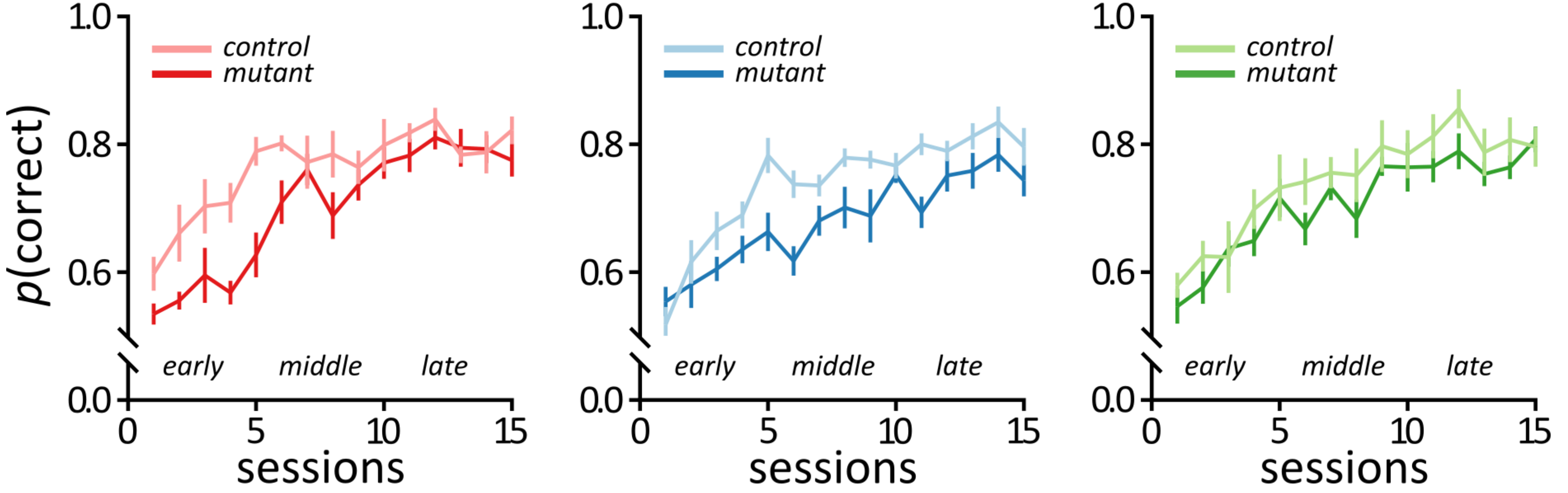




Representative example of one mouse's performance in the ‚TAB' task

General performance

$NR1^{DATCreERT2}$ **$mGluR5^{D1\text{-}KD}$** **$NR1^{D1CreERT2}$**



genotype $F_{1,12} = 11.5$, $p < 0.01$
session $F_{14,168} = 17.15$, $p < 0.001$
interaction $\mathbf{F_{14,168} = 1.90}$, $\boldsymbol{p < 0.05}$

genotype $\mathbf{F_{1,15} = 12.62}$, $\boldsymbol{p < 0.01}$
session $F_{14,210} = 20.69$, $p < 0.001$
interaction $F_{14,210} = 1.49$, *ns*

genotype $F_{1,13} = 1.80$, *ns*
session $F_{14,182} = 19.17$, $p < 0.001$
interaction $F_{14,182} = 0.53$, *ns*

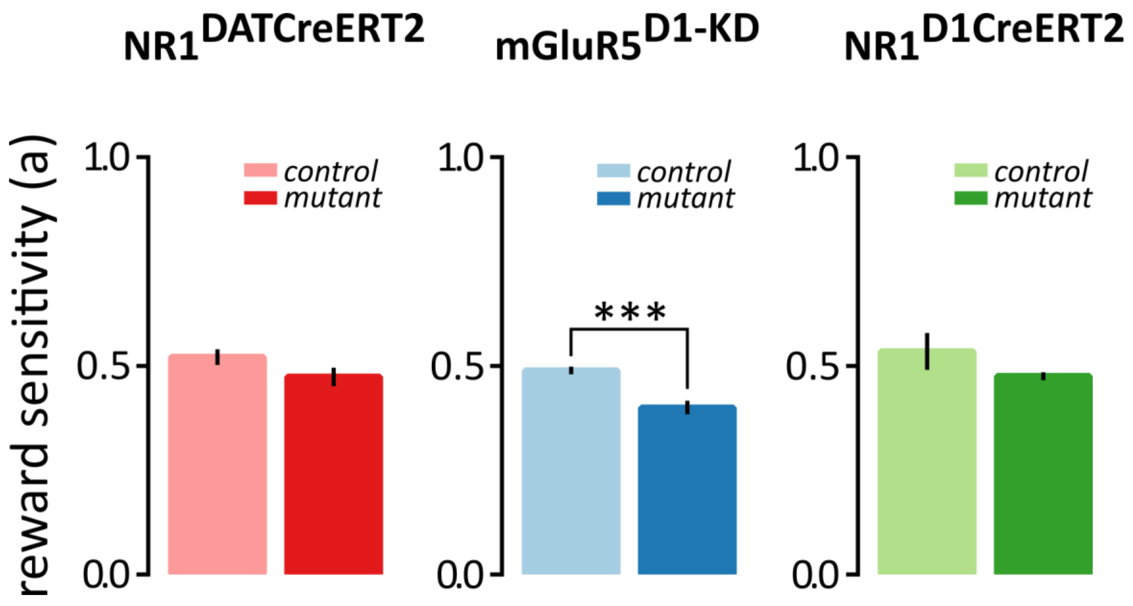
Influence of reward ratio on choice (reward sensitivity)

Generalized Matching Law

$$\log_2\left(\frac{C_L}{C_R}\right) = a \cdot \log_2\left(\frac{R_L}{R_R}\right) + \log_2 b$$

Choice ratio — Reward sensitivity — Reward ratio

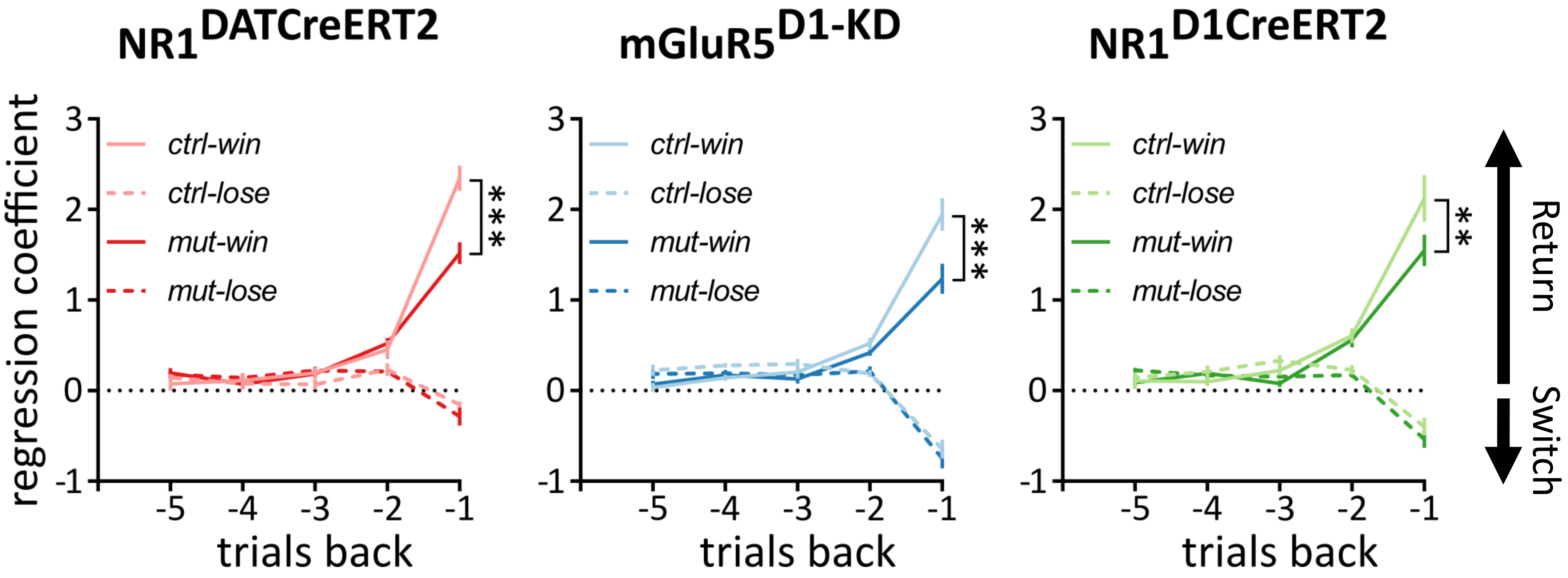
Reward sensitivity – reflects the degree to which the reward ratio actually impacts the choice ratio



Model fits

	NR1DATCreERT2	mGluR5D1-KD	NR1D1CreERT2
control	R^2 = 0,90	R^2 = 0,87	R^2 = 0,85
mutant	R^2 = 0,82	R^2 = 0,84	R^2 = 0,84

Influence of previous choices & outcomes on subsequent choice



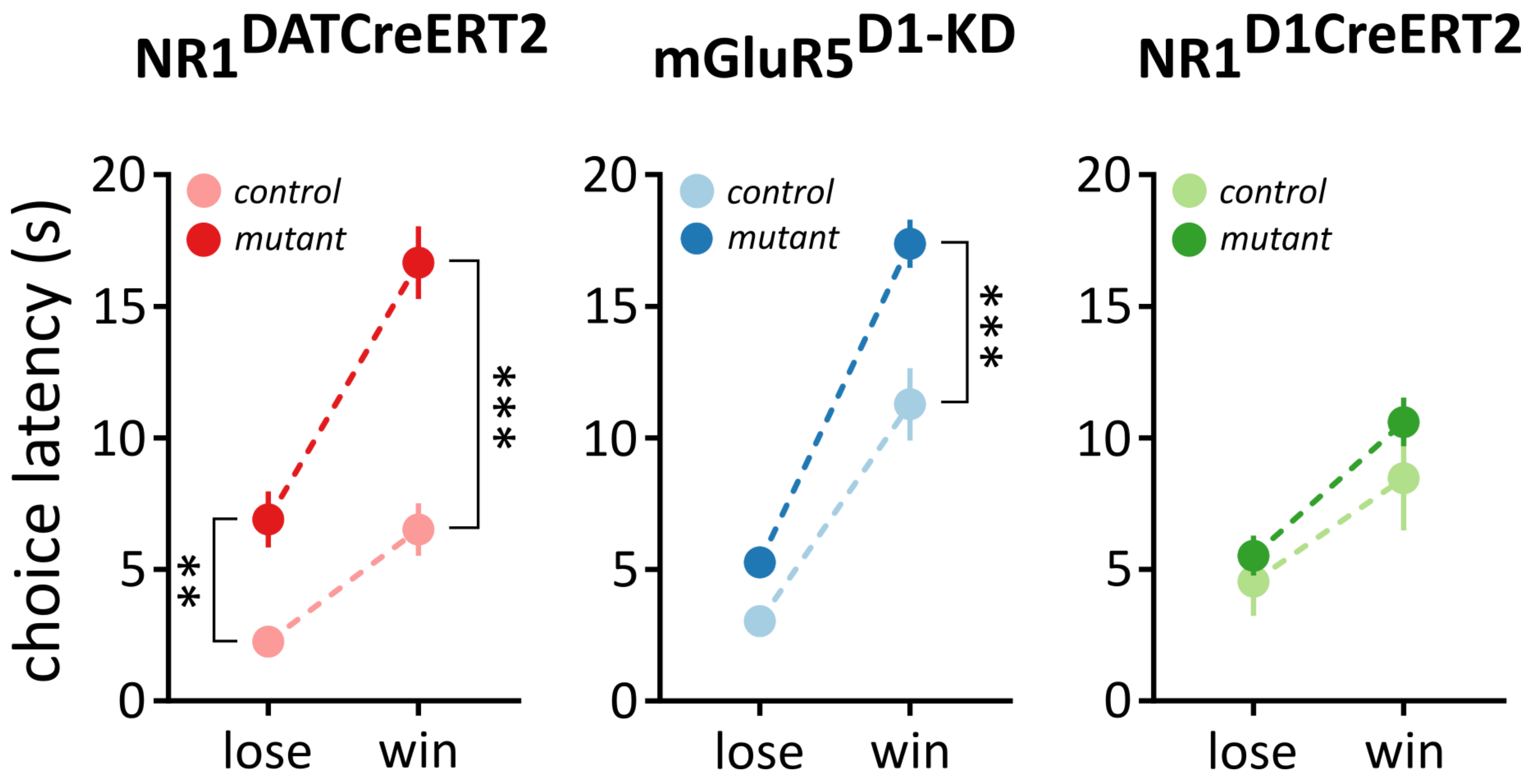
Model fits

ps-R^2 = 0,47	ps-R^2 = 0,41	ps-R^2 = 0,45
ps-R^2 = 0,32	ps-R^2 = 0,25	ps-R^2 = 0,32

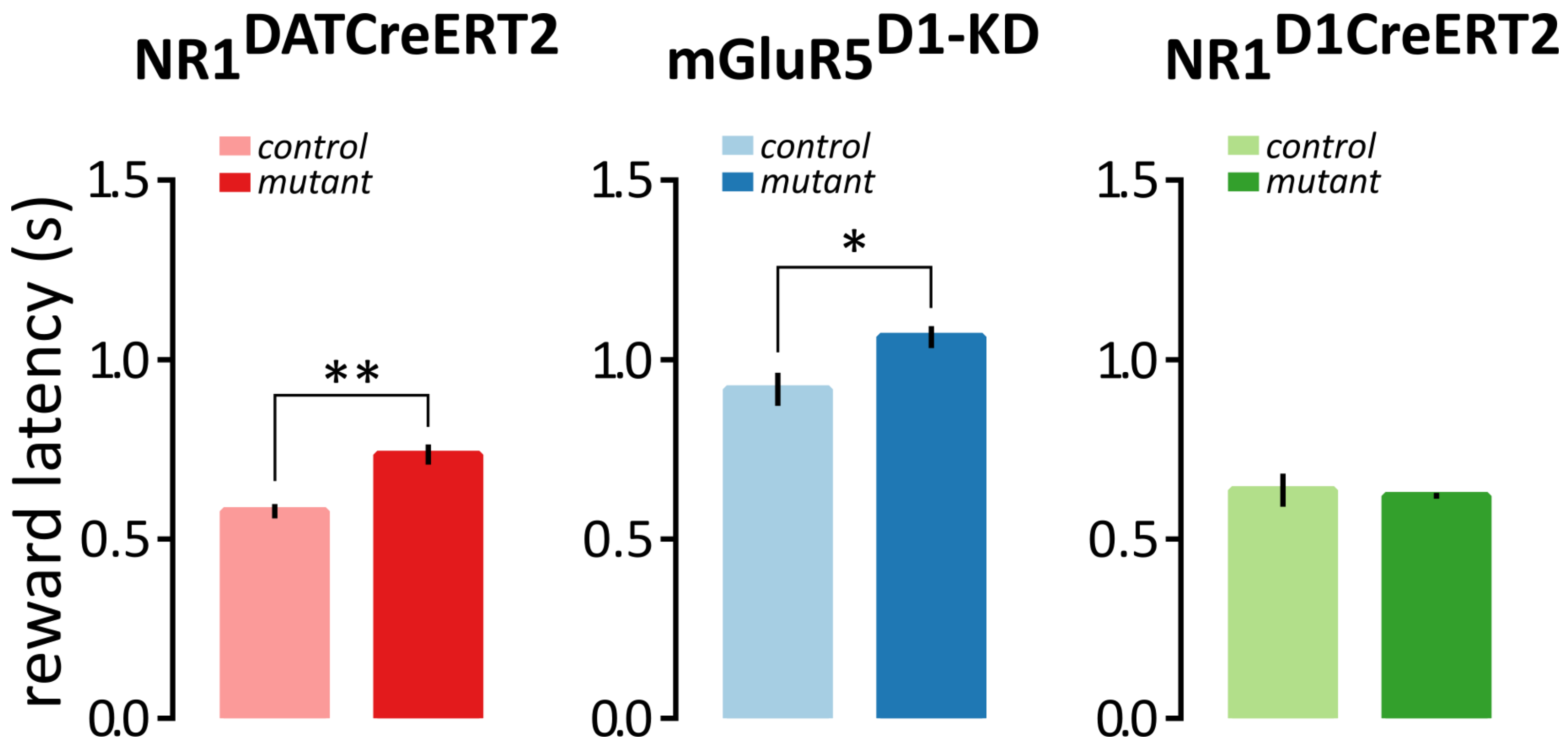
Logistic regression model

Parker et al., 2016 *Nat Neurosci*

Reaction time (choice)



Reaction time (reward)



Learning the value of actions & value-based action selection

(1) Value updating rule:

$$V_L \leftarrow V_L + \alpha \cdot RPE$$
$$RPE = R_L - V_L$$

where: V_L – expected value of ‚Left' option

R_L – outcome received from ‚Left' option

$\boldsymbol{\alpha}$ – **learning rate**, determines how rapidly the estimate of expected value is updated

(2) Action selection rule (softmax):

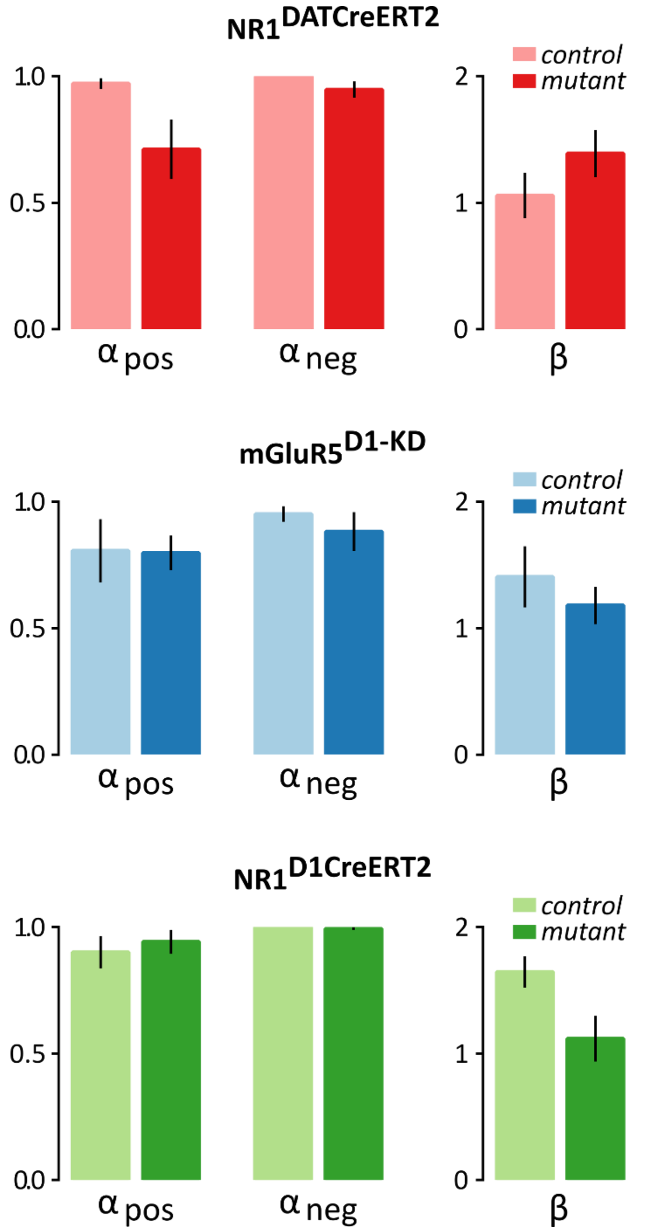
$$P_L = \frac{e^{\beta V_L}}{e^{\beta V_L} + e^{\beta V_R}}$$

where: P_L – probability of choosing ‚Left' option

$\boldsymbol{\beta}$ – **randomness in action selection:**

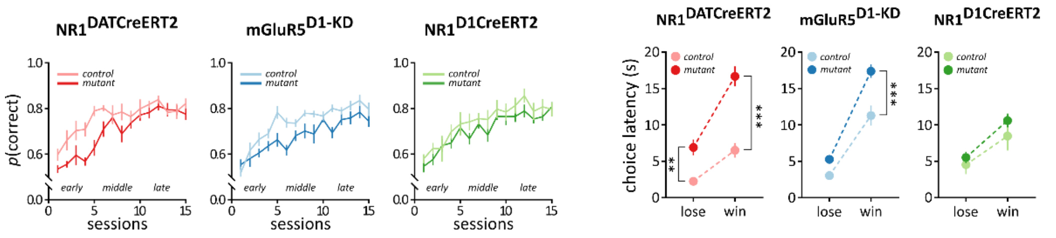
With β = 0, action selection is totally random,

As β is increased, the model is more likely to choose the action believed to have the maximum value

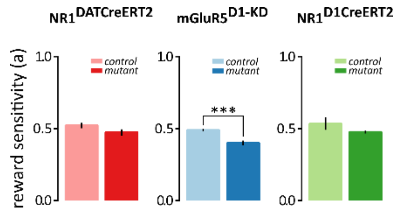


Summary & conclusion

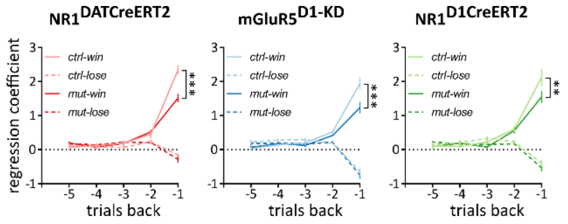
- Ablation of NMDA receptors in DA neurons and mGluR5 receptors in D1R-expressing neurons decreased the likelihood of choosing alternative with higher probability of reward and increased response latency



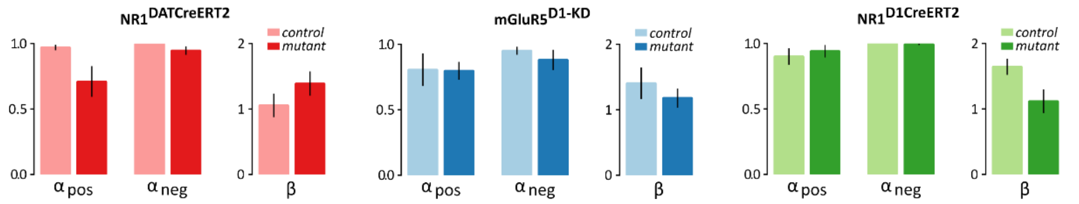
- Ablation of mGluR5 receptors in D1 neurons reduced sensitivity of mutant animals to changes in reward ratios



- Mutant animals were less likely to return to a choice which was previously rewarded



- Loss of NMDA receptors in DA neurons decreased the learning rate from positive outcomes, while ablation of Glu receptors in D1 neurons increased randomness in action selection



NATIONAL SCIENCE CENTRE
POLAND
PRELUDIUM 2014/15/N/NZ4/00761
ETIUDA 2016/20/T/NZ4/00503

Ministry of Science and Higher Education
Republic of Poland
KNOW Leading National Research Centre
2012-2017

Jan Rodriguez Parkitna

Thank You!

cieslak@if-pan.krakow.pl

if-pan.krakow.pl/en/departments/employees/208