Tyrosine Modulates Exploration Behaviour During Reinforcement Learning in a Gender-Balanced Sample

A. Brands[1], D. Mathar[1], K. Knauth[1], K. Lisner[1], T. Rödder[1], & J. Peters[1]
[1]Department of Psychology, Biological Psychology, University of Cologne, Cologne, Germany

correspondence: a.brands@uni-koeln.de

BACKGROUND

- adaptive behavior requires balancing exploration and exploitation
- this balance is influenced by Dopamine (DA)[1] and distortions are associated with several common psychiatric disorders[2]
- freely available supplements targeting DA promise performance enhancements, while their exact influences on basic decision processes remain unclear
- Mathar et al. (2022) [3], reported reduced reaction times and higher model-based control in a RL-task under the catecholamine-precursor **tyrosine** (2g)

→ Here we present partial results from a preregistered[4], double-blind, placebo-controlled, within-subject study, which set out to replicate and extend these findings in a gender-balanced sample.

METHODS

Sample & Procedure

N=64 healthy participants (32 self-identified female) completed a RL-task (4-armed restless bandit), following ingestion of 2g of **tyrosine** or **placebo**, on two testing days in counterbalanced order.

For the full study description see:
https://osf.io/4z39r/?view_only=c26d1419180646dab1af18999725e58f

Analyses used both model-agnostic approaches and comprehensive computational modeling comparing hierarchical Bayesian RL-models which consider various mechanisms that might drive directed exploration[1,4].

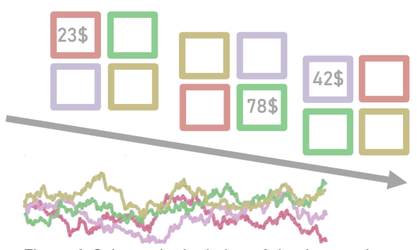

Figure 1. Schematic depiction of the 4-armed restless bandit task. Rewards for all bandits varied according to independent Gaussian random walks,

explore some more

Computational Modeling

Hierarchical Bayesian modeling in Rstan[5]
Base models differed in the learning mechanism (Q-learner vs. Bayesian learner; QL;BL) and accounted for choice repetition from the previous trial [1,4,6]
→ base models were combined with different accounts of directed exploration

Bayesian: participants track underlying reward walks incl. variation → **sigma** [1,6]
counter-heuristics: participants run counters as simplified uncertainty proxies[1,7]
how many trials ago bandit x was last chosen → **trial**
how many other bandits have been sampled since bandit x was last chosen → **bandit**

In total, seven models were fit to the empirical data and were compared with respect to their ability to account for subjects behaviour on the task.
→for full model description see Brands et al., 2023[8]; preregistration on OSF[4]

RESULTS

model-based

Predictive Accuracy

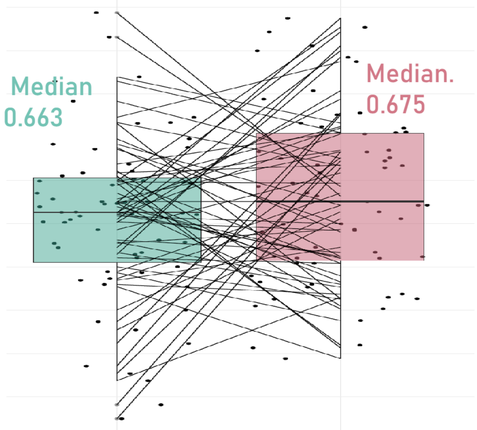

Figure 3. Correspondence of empirical choices with 8000 simulated data sets based on the winning model (BL+ Bandit).

Model Comparison

Model	elpd$_{diff}$	se$_{diff}$	rank	elpd$_{diff}$	se$_{diff}$	rank
BL+Bandit	0.0	0.0	1	0.0	0.0	1
BL+Sigma	-168.6	53.7	2	-258.5	65.5	3
BL+Trial	-237.1	63.2	3	-224.4	56.1	2
QL+Trial	-320.6	68.0	4	-340.0	58.9	4
BL	-572.4	55.1	5	-519.4	48.8	5
QL	-689.1	61.5	6	-674.8	50.6	6
QL+Bandit	-5043.0	282.4	7	4542.6	237.5	7

Table 4. Results from model comparison for both drug conditions separately. Turquoise = Placebo; Coral= Tyrosine. Elpd diff: difference in estimated log pointwise predictive density; lower absolute values reflect a superior fit. The difference in elpd is provided in reference to the winning model (which itself thus always has a -elpd diff of zero) se diff : standard error of this difference.

Posterior Density Estimates

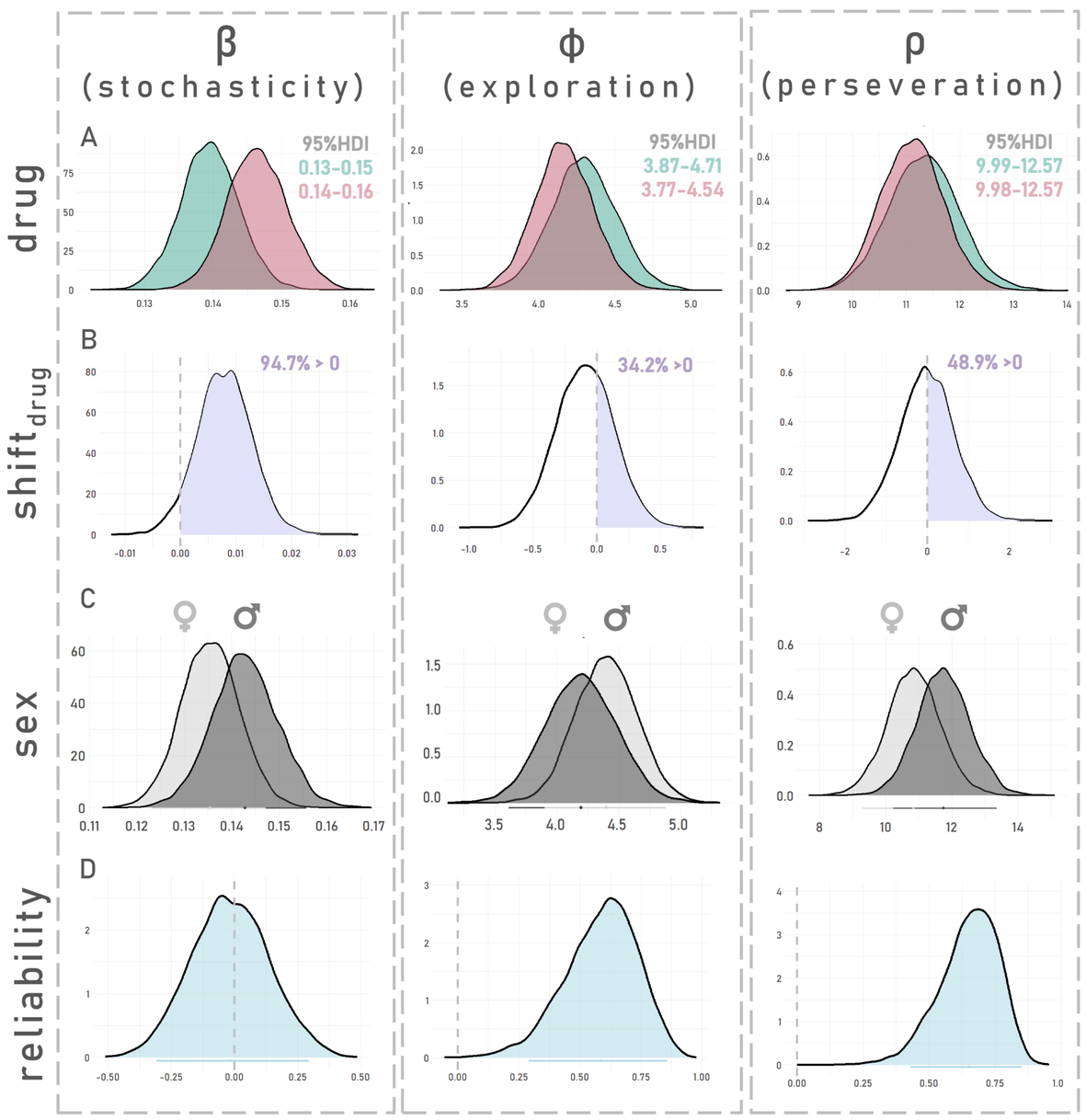

Figure 4. Posterior density estimates derived from the winning model (BL + Bandit). Columns (dashed boxes) refer to the group-level means of the three main parameters of interest (β/ϕ/ρ). Panel A/ drug: posterior densities derived from separate fits per condition (turquoise =placebo; coral =tyrosine). Panel B/ shift: shift parameters describing the group-level change from the placebo to tyrosine condition for β/ϕ/ρ each; c.f. preregistered modeling procedure[4]. Panel C/ sex: Posterior densities in the placebo condition derived from fitting the winning (shift-) model to data from both genders separately (light and dark gray: self-identified female and male respectively). Panel D: model-derived test-retest reliability estimates of each group-level parameter per condition.

model-agnostic

A %best choice

predictor	Estimate	95% CI	p
Intercept	**0.65**	0.64-.067	**<.001**
Drug	**0.02**	0.00-0.03	**.009**
Sex	0.00	-0.02-0,02	.781
Drug * Sex	0.00	-0.02-0,02	.994

B %switches

predictor	Estimate	95% CI	p
Intercept	**0.33**	0.03-0.36	**<.001**
Drug	**-0.3**	-0.04—0.01	**<.001**
Sex	**-0.04**	-0.08-0.00	**.044**
Drug * Sex	0.00	-0.02-0.02	.850

C Median RT

predictor	Estimate	95% CI	p
Intercept	**0.65**	0.64-.067	**<.001**
Drug	**0.02**	0.00-0.03	**.009**
Sex	0.00	-0.02-0,02	.781
Drug * Sex	0.00	-0.02-0,02	.994

Table 1,2,&3. Regression results (fixed effects) from separate linear mixed models with condition, sex, & their interaction as fixed effects and subjects as random effects for model-agnostic indices. Panel A: percentage of choices for bandit with highest reward. Panel B: percentage of choice alteration on two consecutive trials. Panel C: Median reaction times in seconds. Significant results are printed in bold.

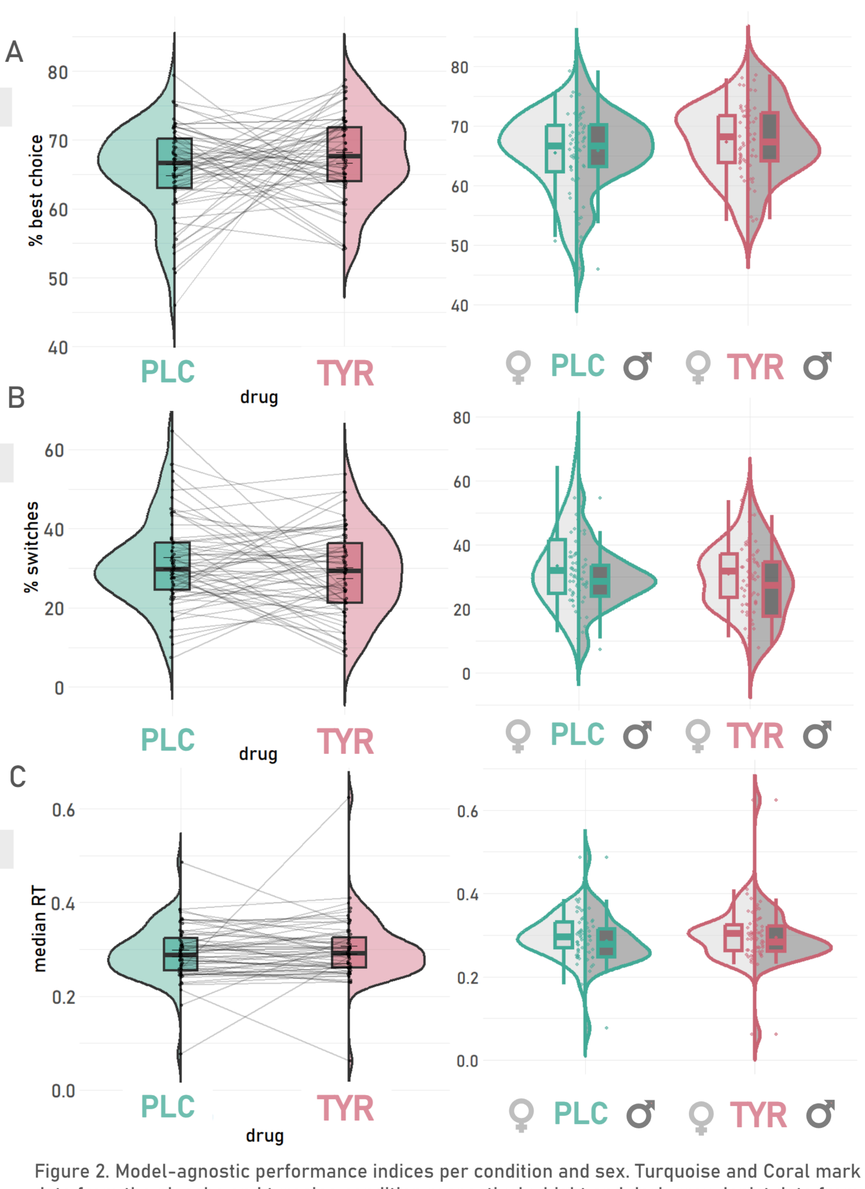

Figure 2. Model-agnostic performance indices per condition and sex. Turquoise and Coral mark data from the placebo and tyrosine condition respectively. Light and dark gray depict data from the self-identified female and male subjects respectively. Panel A: percentage of choices for bandit with highest reward. Panel B: percentage of choice alteration on two consecutive trials. Panel C: Median reaction times in seconds. Points represent individual subjects.

CONCLUSION

- subjects showed less switching, higher median RTs as well as a higher proportion of optimal choices under 2g tyrosine vs. placebo
- Bayesian learner models outperform Q-learner across both conditions
- heuristic-based exploration bonus captured subjects' behaviour better than full Bayesian uncertainty-tracking
- overall no apparent gender differences (despite switching behaviour)
- perseveration (ρ) and exploration (ϕ) parameters show medium to high test-retest reliability → stable, trait-like decision components
- while stochasticity (vs. value-guided decisions; β) varied across conditions → state-dependent component, more readily influenced by DA-processes
- modelling results are broadly in line with model-agnostic analyses

➔ potential shift of exploration-exploitation balance under tyrosine
➔ tyrosine facilitated more value-guided (less random) choices which may be accompanied by a decrease in directed exploration
➔ further investigation warranted

REFERENCES

1. Chakroun et al.(2020). eLife, 9, e51260. https://doi.org/10.7554/eLife.51260
2. Montague et al. (2012).Trends in Cognitive Sciences, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018
3. Mathar et al. (2022). PLoS computational biology, 18(12), e1010785. https://doi.org/10.1371/journal.pcbi.10107852.
4. Preregistration OSF: https://osf.io/4z39r/?view_only=c26d1419180646dab1af18999725e58f
5. Stan Development Team (2022). https://mc-stan.org/.
6. Daw et al. (2006). Nature, 441(7095), 876–879. https://doi.org/10.1038/nature04766
7. Fox et al. (2020). Current Opinion in Behavioral Sciences, 35, 104–111. https://doi.org/10.1016/j.cobeha.2020.08.004
8. Brands et al. (2023). bioRxiv. https://doi.org/10.1101/2023.05.22.541443