Psychoneuroendocrinology 100 (2019) 27–31

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen

Short Communication

Reinforcement learning across the rat estrous cycle

Jeroen P.H. Verharen[a,b], Jiska Kentrop[a], Louk J.M.J. Vanderschuren[b,*,1], Roger A.H. Adan[a,*,1]

[a] Brain Center Rudolf Magnus, Department of Translational Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Department of Animals in Science and Society, Division of Behavioural Neuroscience, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Learning
Motivation
Estrous cycle
Hormones
Rats
Reinforcement earning

ABSTRACT

Reinforcement learning, the process by which an organism flexibly adapts behavior in response to reward and punishment, is vital for the proper execution of everyday behaviors, and its dysfunction has been implicated in a wide variety of mental disorders. Here, we use computational trial-by-trial analysis of data of female rats performing a probabilistic reward learning task and demonstrate that core computational processes underlying value-based decision making fluctuate across the estrous cycle, providing a neuroendocrine substrate by which gonadal hormones may influence adaptive behavior.

1. Introduction

Reinforcement learning is an essential mechanism for organisms to adapt to a dynamic environment, by allowing flexible alterations in behavior in response to positive and negative feedback, for example during foraging and social encounters (Sutton and Barto, 1998). As such, deficits in reinforcement learning have been implicated in several psychiatric conditions, including addiction and schizophrenia (Maia and Frank, 2011). Given the large gender differences in the prevalence of mental disorders, and the existence of cyclic changes in the severity of schizophrenia and sensitivity to drugs in women (Hendrick et al., 1996), we sought to determine how the estrous cycle of females affects the computational processes that underlie reinforcement learning. To this aim, we tested a cohort of female rats on a probabilistic reversal learning paradigm (Bari et al., 2010; Verharen et al., 2018), used computational modeling to extract the subcomponents of value-based decision making, and assessed how these components were affected by the estrous cycle.

2. Methods

2.1. Animals

Female, nulliparous Long-Evans rats (bred in-house; background Rj:Orl, Janvier labs, France; $n = 30$) weighing 180–220 g were used for the experiment. Animals were tested for 10 consecutive days, to ensure that we had at least one measurement of every cycle stage per animal. Eventually, 5 animals had to be excluded because the cycle could not reliably be estimated or not all stages of the cycle were captured due to unreliable vaginal smears, leaving a final group of $n = 25$. Animals were socially housed in groups of 2–4 and kept on a reversed day/night cycle (lights on at 8 A.M.), and behavioral experiments took place between 9 A.M. and 1 P.M.. During the training phase of the experiment, animals were kept on a food restriction regimen of 5 g chow per 100 g body weight, and during the 10 experimental days the animals were food restricted for 16 h prior to the behavioral task. For the male group of animals ($n = 18$), that is included for comparison, Long-Evans rats (bred in-house; background Rj:Orl, Janvier labs, France) of roughly the same age, weighing 310–390 g, were used. Animals had *ad libitum* access to water, except during behavioral experiments. The experiments were carried out in accordance with Dutch legislation (Wet op de Dierproeven, 2014), European Union guidelines (2010/63/EU), and approved by the Animal Welfare Body of Utrecht University and the Dutch Central Animal Testing Committee.

2.2. Behavioral task

The probabilistic reversal learning task (Fig. 1a) took place in operant conditioning chambers (Med Associates Inc., USA) equipped with a food receptacle (with infra-red entry detection) flanked by two retractable levers and two cue lights, a house light and an auditory tone generator. One lever was randomly assigned as the high-probability lever, responding on which was reinforced (i.e., delivery of a sucrose pellet) with an 80% probability and not reinforced (i.e., a time-out) with a 20% probability. The other lever was assigned as the low-probability lever, responding on which had a 20% chance of being

* Corresponding authors.
E-mail addresses: l.j.m.j.vanderschuren@uu.nl (L.J.M.J. Vanderschuren), r.a.h.adan@umcutrecht.nl (R.A.H. Adan).
[1] These authors contributed equally.

https://doi.org/10.1016/j.psyneuen.2018.09.016
Received 25 May 2018; Received in revised form 24 July 2018; Accepted 12 September 2018

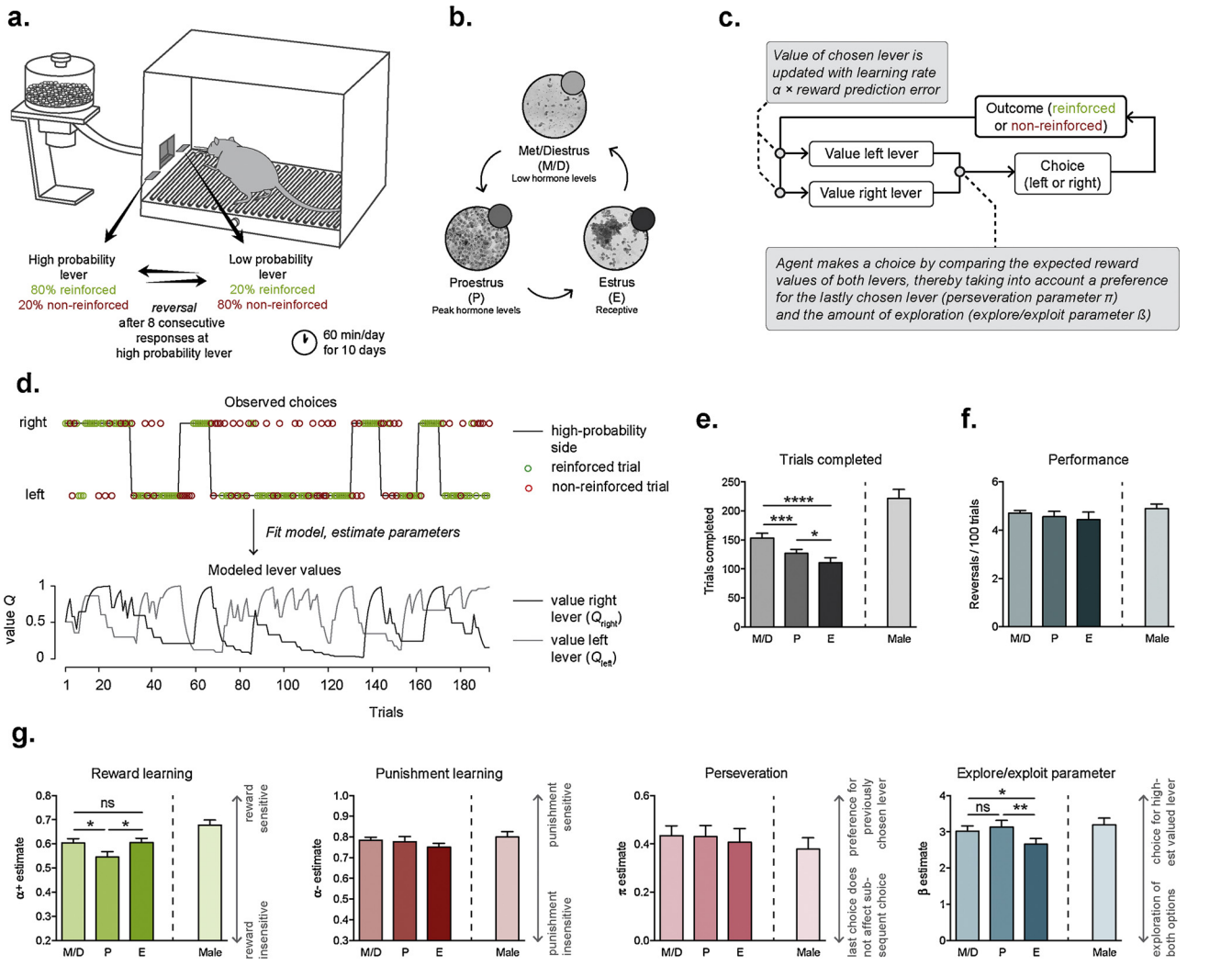

Fig. 1. a. Probabilistic reversal learning setup. Hungry female animals could respond on two levers, one of which delivered sucrose reward with a high probability (80%, high-probability lever), and the other lever with a low probability (20%, low-probability lever). Every time the animal made eight consecutive responses on the high-probability lever, a reversal in reinforcement contingencies occurred, so that the previously low-probability lever became the high-probability lever, and vice versa. In this way, animals had to track the outcome of responding on each of the two levers over a series of trials and based hereon make a choice between them. b. Example cytological images of samples from vaginal smears during the three stages of the estrous cycle. c. Computational model. d. Trial-to-trial data was fit to the computational model, and best-fit parameters were estimated. e. Total trials completed by the female animals ($n = 25$) in the 60-minute session was significantly affected by the estrous cycle (Repeated measures ANOVA, $F_{2,48} = 21.22$, $P < 0.0001$). Post-hoc tests: **** $P < 0.0001$, *** $P = 0.0002$, * $P = 0.0188$. Male data ($n = 18$) is shown for illustrative purposes; these data were not included in the statistical analyses. f. The total number of reversals was not affected by the cycle (ANOVA, $F_{2,48} = 0.48$, $P = 0.6209$). g. Best-fit computational model parameters per estrous cycle stage. Reward learning: ANOVA $F_{2,48} = 3.995$, $P = 0.0248$; post-hoc tests met/diestrus (M/D) vs proestrus (P), $P = 0.0198$, M/D vs estrus (E), $P = 0.9425$, P vs E, $P = 0.0166$. Punishment learning: ANOVA $F_{2,48} = 1.637$, $P = 0.2052$. Perseveration: ANOVA $F_{2,48} = 0.1349$, $P = 0.8741$. Explore/exploit: ANOVA $F_{2,48} = 5.201$, $P = 0.0090$; post-hoc tests M/D vs P, $P = 0.4444$, M/D vs E, $P = 0.0243$, P vs E, $P = 0.0033$. Male data is shown for illustrative purposes.

reinforced. Every single response on the high-probability and low-probability lever was reinforced with a 80% or 20% probability, respectively, irrespective of the outcome of the previous trials.

The session lasted for 60 min, and animals were constrained in the number of trials they could make only by the length of the session (maximum ~600 trials per session possible). A trial commenced by the illumination of the house light and the presentation of the two levers into the operant cage. After a lever press by the animal, the levers retracted and the house light was turned off. For reinforced trials, a 45 mg sucrose pellet (5TUL, TestDiet, USA) was delivered into the food port, and both cue lights that flanked the food receptacle were illuminated, and an auditory tone was played for 0.5 s. A new trial commenced directly when the animal entered the food port (detected by the infra-red movement detector); this was signaled to the animal by extinction of the cue lights, illumination of the house light and presentation of the two levers. On non-reinforced trials, no additional cues were presented, leaving the animals in the dark during a 10 s period.

Every time the animal made 8 consecutive responses on the high-probability lever, a reversal in reinforcement contingencies occurred, so that the high-probability and low-probability levers switched. This reversal was not signaled to the animal, so it had to infer this contingency switch from the outcomes of the trials.

The software automatically registered the responses and response times of the animals, as well as the outcome of the trial (reinforced or not), and the position of the high-probability lever.

2.3. Training

Animals first received lever press training, during which both levers were continuously presented, and a lever press was reinforced under a