Behavioural Brain Research 331 (2017) 276–281

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Research report

Novel reinforcement learning paradigm based on response patterning under interval schedules of reinforcement

Christin Schifani[a,*], Ilya Sukhanov[b], Mariia Dorofeikova[b], Anton Bespalov[a,b]

[a] Department of Pharmacology, Neuroscience Research, AbbVie, Ludwigshafen, Germany
[b] Institute of Pharmacology, Pavlov Medical University, St. Petersburg, Russia

ARTICLE INFO

Chemical compounds studied in this article:
(+)-MK-801 maleate (PubChem CID: 53580)

Keywords:
Operant conditioning
Cognitive flexibility
Reinforcement learning
Response patterning
MK-801
Post-reinforcement pause

ABSTRACT

There is a need to develop cognitive tasks that address valid neuropsychological constructs implicated in disease mechanisms and can be used in animals and humans to guide novel drug discovery. Present experiments aimed to characterize a novel reinforcement learning task based on a classical operant behavioral phenomenon observed in multiple species – differences in response patterning under variable (VI) vs fixed interval (FI) schedules of reinforcement. Wistar rats were trained to press a lever for food under VI30 s and later weekly test sessions were introduced with reinforcement schedule switched to FI30 s. During the FI30 s test session, post-reinforcement pauses (PRPs) gradually grew towards the end of the session reaching 22–43% of the initial values. Animals could be retrained under VI30 s conditions, and FI30 s test sessions were repeated over a period of several months without appreciable signs of a practice effect. Administration of the non-competitive *N*-methyl-D-aspartate (NMDA) receptor antagonist MK-801 ((5*S*,10*R*)-(+)-5-Methyl-10,11-dihydro-5*H*-dibenzo[*a*,*d*]cyclohepten-5,10-imine maleate) prior to FI30 s sessions prevented adjustment of PRPs associated with the change from VI to FI schedule. This effect was most pronounced at the highest tested dose of MK-801 and appeared to be independent of the effects of this dose on response rates. These results provide initial evidence for the possibility to use different response patterning under VI and FI schedules with equivalent reinforcement density for studying effects of drug treatment on reinforcement learning.

1. Introduction

Cognitive deficits are at the core of a variety of psychopathologies, including schizophrenia, mood disorders, obsessive compulsive disorder, autism, and attention deficit hyperactivity disorder. In the recent years, there were a number of efforts invested into the development of cognitive tasks that can be used in both humans and preclinical species to guide novel drug discovery [1]. These efforts identified a number of cognition domains that were prioritized for detailed evaluation.

One of such domains is that of long-term memory which, besides relational encoding/retrieval and item encoding/retrieval, includes a construct of reinforcement learning defined as "acquire(d) behavior as a function of both positive and negative reinforcers including the ability to (a) associate previously neutral stimuli with value as in Pavlovian conditioning; (b) rapidly modify behavior as a function of changing reinforcement contingencies; and (c) slowly integrate over multiple reinforcement experiences to determine probabilistically optimal behaviors in the long run" [2].

There are a number of reinforcement learning tasks that were characterized with respect to psychometrics and involved neural systems and were applied in healthy and diseased individuals. An example is the probabilistic reward task that is "based on a differential reinforcement schedule that provides an objective assessment of participants' propensity to modulate behavior as a function of reward history" [3]. Under this task, appropriate answer to one of two stimuli is rewarded more often than to the other and these differences in reward probability generate a response bias that is present in healthy subjects but may be absent in some categories of patients [3]. Based on certain construct validity, initial information on underlying neuroanatomical substrates and evidence for pharmacological and behavioral modifiability of task performance, CNTRICS (Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia) initiative recommended this task for immediate development. There is an ongoing work to adapt this task for the use in animals [4].

Challenges associated with the translation of originally human probabilistic tasks into preclinical laboratory work are well illustrated

Abbreviations: CNTRICS, Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia; DMTP, delayed matching to position; DNMTP, delayed non-matching to position; FI, fixed interval; NMDA, *N*-methyl-D-aspartate; PRP, post-reinforcement pause; VI, variable interval

* Corresponding author. Present address: Research Imaging Centre, Centre for Addiction and Mental Health, 250 College Street, Toronto M5T 1R8, Canada.
E-mail address: christin.schifani@camh.ca (C. Schifani).

http://dx.doi.org/10.1016/j.bbr.2017.04.043
Received 24 January 2017; Received in revised form 21 April 2017; Accepted 22 April 2017
Available online 27 April 2017
0166-4328/

by studies on reversal learning phenomena. Besides difficulties with making the human and animal tasks truly symmetrical (e.g. deterministic vs probabilistic contingencies, nature of reward and punishment) [2], animal reversal learning tasks suffer from a practice effect that prevents longitudinal studies and within-subjects designs (see however [5]). It is not surprising that some human tasks are too difficult or too complex to be implemented in animals while others loose certain qualities when back translated. Therefore, one may also want to consider the development of a task in animals to assess a certain cognitive construct and then taking it forward into humans.

Reinforcement learning is the area of cognitive neuroscience that is very well researched in animals and there are a number of phylogenetically well-conserved phenomena that may serve as a basis for developing novel translatable tasks. It has been studied in animals for more than 100 years and various schedules of reinforcement have been extensively characterized. Even simple reinforcement schedules, ratio or interval, generate responding that differs markedly in terms of the response rate and response patterning. For example, variable interval schedules can be used to produce relatively high and steady response rates whereas fixed interval schedules produce lower response rates that accelerate towards the interval end and pause after the reinforcer delivery. Response patterning including the post-reinforcement pause (PRP) directly depends on the interval size (e.g. larger intervals produce larger PRPs) [6]. Most important, these differences between fixed and variable interval schedules exist even if the overall density of reinforcement is equal (i.e. rate of reinforcer delivery averaged across the session).

The present study attempted to take advantage of this knowledge and aimed to evaluate changes in the response patterning in animals, trained under variable interval (VI) schedule, when the schedule is switched to a fixed interval (FI) equivalent. Main questions to address were: a) how fast do these changes occur (e.g. observable within a single session or across multiple sessions), b) can the response patterning adjustment be studied repeatedly over long periods of time, and c) how is this adjustment affected by a drug known to induce cognitive deficits – the prototypical non-competitive *N*-methyl-D-aspartate (NMDA) receptor antagonist MK-801.

2. Materials and methods

2.1. Ethics

Experiments were approved by the AbbVie Animal Welfare Officer (Ludwigshafen, Germany) or by the Ethics Committee of Pavlov Medical University (St. Petersburg, Russia) and were performed in accordance with the European and German national guidelines (local Animal Welfare Act and the European Communities Council Directive of 24 November 1986 (86/609/EEC)) or the Helsinki Declaration as well as the recommendations and policies of the United States National Institutes of Health Principles of Laboratory Animal Care. Animal housing and experiments in Ludwigshafen were conducted in facilities with full accreditation by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). Reporting of the study complies with the ARRIVE (Animals in Research: Reporting *In Vivo* Experiments) guidelines [7].

2.2. Animals

Two groups of experimental-naïve animals (12 rats for the task development and 12 rats for MK-801 tests) were used at two different laboratories (AbbVie, Ludwigshafen, Germany and Pavlov Medical University, St. Petersburg, Russia, respectively) with essentially identical testing conditions. Male Wistar Han rats (Janvier, Le Genest St. Isle, France or Rappolovo, St. Petersburg, Russia; 9 weeks old at the delivery) were housed individually on non-coniferous woodchip bedding in Macrolon Type 3 cages (Tecniplast, Buguggiate, Italy) with environmental enrichment (one or two different items per cage) under standard laboratory conditions, 21 ± 1 °C and 40–70% humidity. During the entire experiments, the animals had access to drinking water *ad libitum* but were fed restrictively (15 g per day for the first two months followed by 17 g per day at AbbVie; 12–16 g per day taking into account the weight measurement results at St. Petersburg) to limit the body weight gain to 5–6 g per week.

All experiments were performed during the light phase (between 11 a.m. and 1 p.m. at AbbVie; between 2 p.m. and 5 p.m. in St. Petersburg) of a 12-h day/night cycle (lights on at 6 a.m. (AbbVie) or 8 a.m. (St. Petersburg)), Monday to Friday or Monday to Saturday (AbbVie and St. Petersburg, respectively).

2.3. Apparatus

Experiments in both laboratories were conducted in standard modular operant conditioning chambers for rats (MED Associates, St. Albans, VT, USA) enclosed in sound- and light-attenuating cubicles, connected to a computer through an interface, and controlled by MED-PC software. Each chamber was equipped with a grid floor, a house light, a response lever and a food dispenser that delivered 45-mg food pellets into a food receptacle located next to the lever.

2.4. Training and testing procedure

After one week of acclimatization, the rats were shaped to press a lever to receive food under a continuous reinforcement schedule until they reached a criterion of 100 lever presses during a 30-min session. After that, over several days, the reinforcement schedule was gradually changed to VI30 s (intervals ranging from 5 to 70 s). Training continued under the final schedule of VI30 s for at least 3 weeks. After that (AbbVie) or after reaching a criterion of stability (St. Petersburg; difference in the number of lever presses ≤ 20% during the last 3 experimental sessions), a testing procedure was introduced whereby "training" VI30 s schedule was operating during four of the five weekly sessions (five of the six session for the experiment made in St. Petersburg) and one "test" session was conducted under FI30 s. There was a maximum of 84 (AbbVie) or 120 (St. Petersburg) food pellets that could be earned during training and test sessions.

The task was first established at AbbVie and then was transferred to St. Petersburg where the experiments with MK-801 were performed. Drug testing did not commence until each rat was exposed to at least eight FI30 s test sessions.

2.5. Drugs

Solutions of MK-801 (dizocilpine, (5*S*,10*R*)-(+)-5-Methyl-10,11-dihydro-5*H*-dibenzo[*a*,*d*]cyclohepten-5,10-imine maleate; Tocris, Bristol, UK) were freshly prepared prior to testing. MK-801 was dissolved in 0.9% solution of sodium chloride and injected intraperitoneally in a volume of 1 ml/kg body weight 30 min before the test sessions. The pseudo-random sequence of drug dose testing (0, 0.025, 0.05 or 0.1 mg/kg MK-801) was based on a Latin square design. FI test sessions with drug pretreatment were conducted maximum once a week (varying between Wednesdays, Thursdays, and Fridays). Experimenters were not blinded to group assignment. The three doses of MK-801 were chosen based on former reports showing their cognition impairing effects [8–10].

2.6. Data analysis

During each session, every event (lever press responses and food deliveries) was recorded with the time stamp. Based on these raw data, the following parameters were calculated individually for each rat and session for subsequent analysis. No events or individual rat data were excluded from the analysis.