Behavioural Brain Research 395 (2020) 112836

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Abnormal reinforcement learning in a mice model of autism induced by prenatal exposure to valproic acid

Meilin Wu[a,1], Yuanyuan Di[a,1], Zhijun Diao[a], Li Yao[a], Zhaoqiang Qian[a], Chunling Wei[a], Qiaohua Zheng[a], Yihui Liu[a], Jing Han[a], Zhiqiang Liu[a], Juan Fan[b], Yingfang Tian[b,*], Wei Ren[a,*]

[a] MOE Key Laboratory of Modern Teaching Technology, Center for Teacher Professional Ability Development, Shaanxi Normal University, Xi'an, 710062, China
[b] College of Life Sciences, Shaanxi Normal University, Xi'an, 710062, China

ARTICLE INFO

Keywords:
Autism spectrum disorder
Reinforcement learning
Valproic acid

ABSTRACT

Individuals with autism spectrum disorder (ASD) display dysfunction in learning from environmental stimulus that have positive or negative emotional values, posing obstacles to their everyday life. Unfortunately, mechanisms of the dysfunction are still unclear. Although early intervention for ASD victims based on reinforcement learning are commonly used, the mechanisms and characteristics of the improvement are also unknown. By using a mice model of ASD produced by prenatal exposure to valproic acid (VPA), the present work discovered a delayed response-reinforcer forming, and an impaired habit forming in a negative reinforcement learning paradigm in VPA exposure male offspring. But the extinction of the learned skills was found to become faster than normal male animals. Since escape action of nosepoking and the motility remain unchanged in the VPA male offspring, the impaired learning and the accelerated extinction are caused by deficits in higher brain functions underlying association between the animals' behavioral responses and the outcomes of such responses. The results further suggest that the rodent ASD model produced by prenatal exposure to VPA reproduces the deficits in reasoning or building the contingency between one's own behaviors and the consequent outcomes of the behavior seen in ASD patients.

1. Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disease with high prevalence, characterized by the impaired social communication, restricted interests, and repetitive stereotyped behaviors [1,2]. Besides the above core symptoms, individuals with ASD often display multiple dysfunctions in perceiving and reacting to environmental stimulus, ranging from aberrant sensory reactivities, such as excessively high or low reaction for different sensory stimulus [3,4], altered perceptual processing [5], to impaired executive functions including planning, attention and flexibility [6,7], which pose a huge obstacle to their adaptive behaviors in everyday life [8].

Learning is one of the most essential and vital adaptive behavior for all animal species, which yet often involves multiple levels of information processing, especially upon the rewarding or punishing signals [9]. Extensive evidence suggests that the reward system of ASD individuals are altered, resulting in abnormal reactivity to stimulus that have positive or negative emotional values [10,11]. Besides, researches using either animal models or human subjects have indicated that individuals with autism-like behaviors or ASD display impaired emotional learning [12,13]. Lacking effective drugs, behavior intervention treatments for ASD subjects based on learning tasks are getting popular [14,15]. Although early intervention based on reinforcement learning has been shown to improve the cognitive ability, language usage, and adaptive skills in children with ASD [1,16], the mechanisms and characteristics of the improvement are still largely unknown.

Reinforcement learning (instrumental learning), together with its extinction, are essential for decision-making and adaptation to environmental challenges [17]. Different with conditioning learning, reinforcement learning is realized by increasing probability of active behaviors to maintain rewarding or to reduce punishing stimulus [18], during which individuals must integrate information from internal and external sources [19], reason the contingency between their own action and the consequent outcomes, and then determine the following actions to gain further preferable outcomes. Previous studies have demonstrated that reinforcement learning consists of two phases, an initial

* Corresponding authors at: MOE Key Laboratory of Modern Teaching Technology, College of Life Sciences, Shaanxi Normal University, 199 South Chang'an Rd., Yanta District, Xi'an, 710062, China.
E-mail addresses: yingfang_tian@snnu.edu.cn (Y. Tian), renwei@snnu.edu.cn (W. Ren).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bbr.2020.112836
Received 28 April 2020; Received in revised form 3 July 2020; Accepted 24 July 2020
Available online 01 August 2020
0166-4328/

phase of a goal-directed reinforcer forming and a following one of habit forming [20,21]. The initial phase has a strong relation with the rewarding and complicated calculation progress [22,23], while the habit forming phase needs extended practice to acquire the insensitivity to outcome devaluation and contingency degradation [24]. Considering that ASD individuals often have difficulties in processing rewarding [10] and are abnormal in adaptative perception [25], there is a high probability that individuals with ASD have difficulties in completing these reinforcement learning tasks in both the reinforcer forming and the habit forming phases.

Valproic Acid (VPA) is a wildly used psychotropic drug for treatment of epilepsy, bipolar disorder, and a number of other neuropsychological disorders. Maternal exposure to VPA has been confirmed to be a high-risk factor of ASD [26]. With this association, prenatal exposure to VPA has been prevalently applied to produce a representative animal model of environmental risk factors-induced ASD [2,12]. In the present study, the performance in the two phases of negative reinforcement learning and in the extinction by mice prenatally exposed to VPA were evaluated systematically. The results indicate that prenatal exposure to VPA induced profound adverse effects on the negative reinforcement learning in VPA exposure male offspring, including a delayed response-reinforcer forming, and an impaired habit forming. In addition, exposure to VPA also accelerated extinction of the learned behaviors in male offspring.

2. Materials and methods

2.1. Animals

C57BL/6 J mice of 3- to 4-month-old were housed in standard Specific Pathogen Free facilities, under a 12 h light/dark cycle. Food and water were available ad libitum. Animals (one male mouse with two female mice) were mated, and in the next morning, if spermatozoa were found in vaginal secretion, the pregnancy days were calculated from this day on. After pregnancy, male mice were taken out from the cage, and the pregnant mice were housed in pairs. At E12.5 pregnant female mice were gently and randomly treated with an intraperitoneal injection of 500 mg/kg VPA (250 mg/mL in saline, Sigma, Oakville, CA) or the vehicle.

When the offspring were weaned at 21 days old, they were housed separately by sex. Before the marble burying experiment and the learning experiment, mice were gently handled for 5 days to minimize manipulation-related stress. All studies were performed from postnatal day 31 (P31) to P60 as illustrated in the schedule in Fig. 1. Two cohorts of experiment were performed, using male mice in one and the females in the other. In each cohort, 2–3 pups were randomly selected from each of 5 VPA exposed litters or 4 vehicle litters to form the VPA and the vehicle groups, respectively (N = 10 in each group).

All experiments were conducted during the light phase of the cycle and approved by the Animal Care and Use Committee in accordance with Governmental Regulations of Laboratory Animals of China. After the experiment, mice were used in other studies and finally euthanized by carbon dioxide asphyxiation.

2.2. Marble burying

Marble burying was performed to assess the repetitive stereotyped behaviors at P31 [27]. Empty transparent cages were filled with sawdust in depth of 10 cm, and 12 marbles were evenly spaced on the sawdust in a 3 × 4 grid. The time of burying marbles was recorded for 30 min. Numbers of marbles buried (approximate > 75 % of the size of the marble covered by bedding material) was assessed after every testing session.

2.3. Three-chamber test

Three-chamber tests were performed on P32-33 as previous described [28,29]. There were two sessions and each session lasted for 10 min. In the first session named sociability test, a novel mouse, which was randomly selected from normal mice, was placed under small plastic cage in one of the two side chambers in advance, and the other side chamber was kept empty during the test (Fig. 2D). The test mouse was then put into the central chamber and allowed to freely explore all three chambers. There was 5 min for every test mouse for habituation before the recording. Sociability was evaluated by the sociability index (SI) that was defined as the ratio between duration of test mouse in novel mouse side and that in the empty side. After the first session, a second novel mouse with same age and gender was put into the empty side chamber, and the second session, social preference test, was conducted immediately. During the second session, the first novel mouse was designated as the familiar mouse, and the second mouse as the novel one (Fig. 2G). Social preference to either the familiar mouse or the novel mouse was evaluated by social preference index (SPI) which was defined as the ratio calculated by using duration of test mouse in familiar mouse side and that in the novel side.

2.4. Negative reinforcement learning to escape footshocks

The learning behaviors were started when the mice were 57 days old, and all learning experiments were conducted in operant chambers (30 × 24 × 30, L × W × H in cm; MED-Associates, St. Albans, VT), which were placed in sound-attenuating boxes, equipped with a ventilation fan, a light (4 lx), a transparent door and two nosepokers located 2 cm above the metal grid floor (Fig. 3A). Footshocks were delivered through the metal grid floor, and one of the two nosepokers was randomly designated as “active” to terminate the footshocks, which was illuminated by a light-emitting diode (LED, 20 lx) light during the shock period and can be activated by a nosepoking. The inactive nosepoker was always not illuminated, and no programmed effects were associated with nosepoking on it [30].

Twenty-four hours before the negative reinforcement learning experiment, mice were individually placed in the chambers to acclimate to the experimental environment freely for 100 min with no shock. Negative reinforcement learning experiment consisted of 3 days of training, and each day of training comprised 50 trails. During each trial, mild electric shocks (0.15 mA) were delivered on the beginning of the experimental program, and were terminated when the active nosepokers were triggered at once, or at the time of the maximum shock duration (120 s) if animals failed to close the shocks. When the shock was terminated, the LED lights would be turned off, signaled with a 1.5 s tone (2.9 kHz, 65 dB). Between each trial, animals were allowed to rest for a pseudorandom period ranging from 30 s to 60 s. Learning process lasted for 3 days, and then the extinction process begun (Fig. 3B).

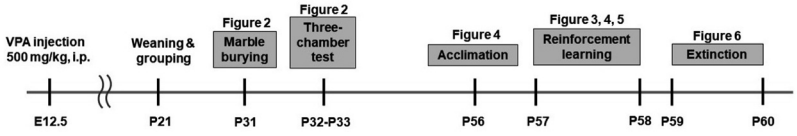

Fig. 1. The study schedule.

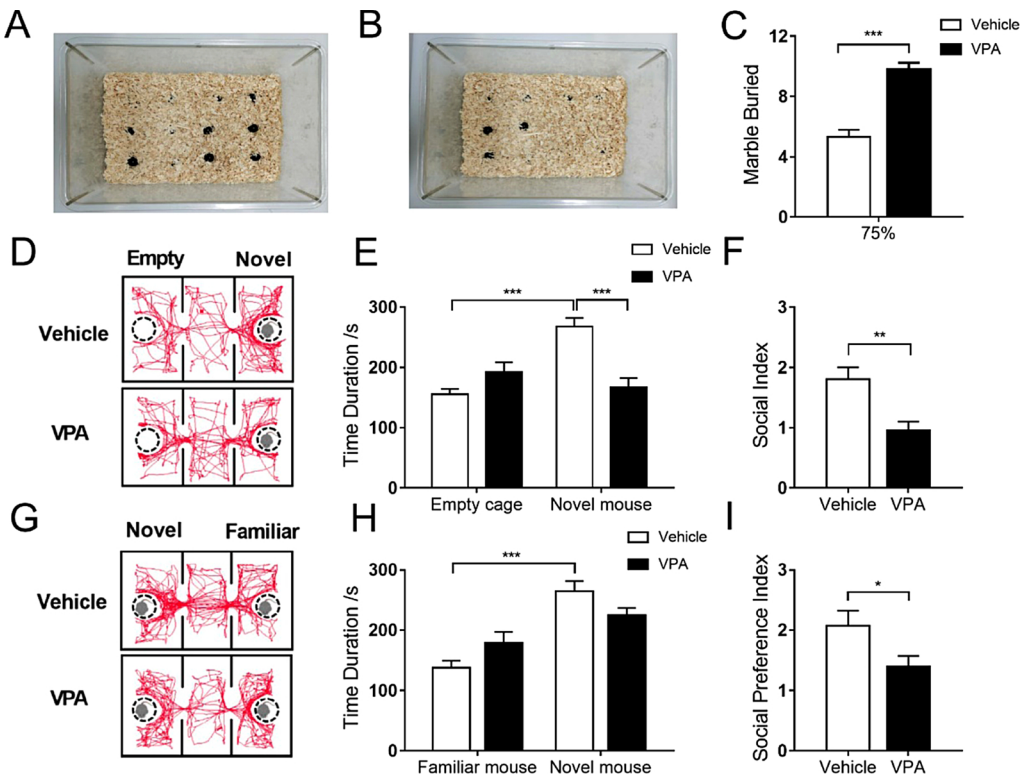

Fig. 2. Effects of prenatal exposure to VPA on autism-like behaviors in male offspring. (**A-B**) Typical performance of marble burying in male mice (A: vehicle group; B: VPA group). **(C)** The VPA male group demonstrated increased total numbers of buried marbles (with the criterion of 75 %) in relative to the vehicle male group. **(D and G)** Procedure of three-chamber test and typical trajectories of mice in the sociability session **(D)** and the social preference session **(G)** of the three-chamber test. (**E-F**) VPA exposed male offspring showed decreased social behaviors compared with the vehicle male group during the sociability session. (**H-I**) Male offspring exposed to VPA demonstrated decreased social preference for the novel mouse compared with the vehicle male group during the social preference session. Error bars represent SEM. RMANOVA and unpaired *t*-test: ***$p < 0.001$.VPA: Male offspring prenatally exposed to VPA; Vehicle: Male offspring prenatally exposed to saline. VPA: n = 10; Vehicle: n = 10.

2.5. Extinction of the learned operation behavior

Extinction experiments over 3 days started following the three days of negative reinforcement learning, which consisted of five extinction trials per day. During the extinction trials, the conditional probability of a termination following a nosepoking action is zero. That is, mice were exposed to a 10-min continuous shock in each extinction trial, which was followed by an intertrial interval of 10 min. For each extinction trial, mice were placed in the same chambers as reinforcement learning, whereas the previously active nose pokers were no longer effective. The total shock duration in each day was 50 min, and the total duration of experiment in each day was 100 min (Fig. 3C).

2.6. Statistical analysis

Repeated measures analysis of variance (RMANOVA) and unpaired

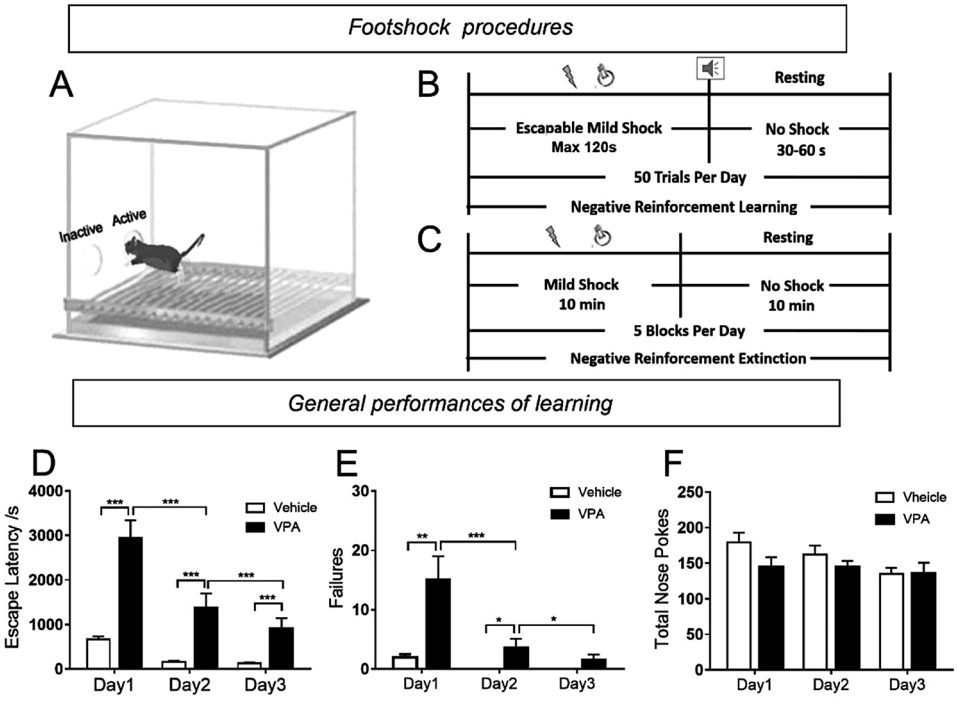

Fig. 3. Footshock procedures and general performances of learning in male mice. **(A)** The operant chamber used to conduct the learning and the extinction experiments. (**B-C**) Procedures in one day of the negative reinforcement learning **(B)** and the extinction process **(C)**. **(D)** Male offspring prenatal exposed to VPA demonstrated impaired learning in relative to the vehicle group while showed significant decreases in escape latencies during three days. **(E)** Male offspring in the VPA group displayed more failures during learning on Day1, Day2 but not on Day3, and showed significant decreases in failures during three days. **(F)** There are no significant differences of total nosepokes (the sum of active and inactive nosepokes) between male groups during the three Days. Error bars represent SEM. RMANOVA: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. VPA: Male offspring prenatally exposed to VPA; Vehicle: Male offspring prenatally exposed to saline. VPA: n = 10; Vehicle: n = 10.

t-test were used to determine the differences. Differences with $p < 0.05$ were considered statistically significant. All statistical analyses were conducted using SPSS version 20 (IBM Corp., Armonk, NY).

3. Results

3.1. Effects of prenatal exposure to VPA on the social and the repetitive stereotyped behaviors

The core autism-like behaviors, including social deficits and repetitive behaviors induced by prenatal exposure to VPA, were assessed before reinforcement learning test. In the typical performance in marble burying shown in Fig. 2A and B, the VPA group showed increased total numbers of buried marbles in relative to the vehicle group ($t_{18} = 6.775$, $p < 0.001$; Fig. 2C) in male offspring. In contrast, in the female offspring, the VPA group did not showed increased total numbers of buried marbles as compared with the vehicle group ($p < 0.05$, Fig. S1A). The results indicated that prenatally exposure to VPA only induced significant repetitive stereotyped behaviors in male offspring.

Three-chamber test was performed to measure social interaction. The VPA group demonstrated significantly decreased social exploratory behavior and social preference in male offspring (SI $t_{18} = 3.436$, $p < 0.01$; SPI $t_{18} = 2.222$, $p < 0.05$; Fig. 2F and I). During sociability test, VPA group exhibited a significant decrease in duration within the novel mouse side than the vehicle group in male offspring (simple interaction effect: $p < 0.001$), and showed no significant difference in duration between the empty cage side and the novel mouse side (simple interaction effect: $p = 0.311$), while the vehicle group showed a significant difference in duration between the two sides (Treatment $F_{(1,18)} = 10.125$, $p < 0.01$; Side $F_{(1,18)} = 6.557$, $p < 0.05$; Treatment × Side $F_{(1,18)} = 16.273$, $p < 0.001$; simple interaction effect: $p < 0.001$; Fig. 2E). Similarly, VPA group also showed no significant difference in duration between the familiar mouse and the novel mouse (simple interaction effect: $p = 0.105$), which is again in contrast to the vehicle group (Treatment $F_{(1,18)} = 0.006$, $p < 0.05$; Side $F_{(1,18)} = 20.735$, $p < 0.001$; Treatment × Side $F_{(1,18)} = 4.591$, $p < 0.05$; simple interaction effect: $p < 0.001$; Fig. 2H). For females, as expected, the social behaviors were not significantly different between the VPA group and the vehicle group (SI: $p < 0.05$; SPI: $p < 0.05$; Figs. S1B -1E; see supplemental materials for details). Only the male offspring prenatally exposed to VPA express significant impairments in social communication, including social exploration and social preference.

3.2. Prenatal exposure to VPA impaired negative reinforcement learning of male mice

3.2.1. Prenatal exposure to VPA impaired negative reinforcement learning without affecting the nosepoking action

Because severe autism-like behavioral deficits are only observed in male offspring after maternal VPA exposure [31–33], we predicted that prenatal exposure to VPA also impaired negative reinforcement learning in males. In the present study, the negative reinforcement learning lasted for three days. For the males, the VPA group demonstrated larger escape latency than the vehicle group (Treatment $F_{(1,18)} = 26.377$, $p < 0.001$; Fig. 3D) in all of the three days. In addition, the VPA group displayed more failures during the first and the second day of learning (Treatment $F_{(1,18)} = 11.141$, $p < 0.01$; Fig. 3E). The results showed that prenatal exposure to VPA indeed impaired negative reinforcement learning along the whole learning process in male offspring. On the other hand, there was no significant difference between the VPA group and the vehicle group in failures on Day 3 of learning, and the escape latency in the VPA group progressively showed significant decreases among days (Day $F_{(1,18)} = 48.639$, $p < 0.001$; Day × Treatment $F_{(1,18)} = 15.560$, $p < 0.001$; Fig. 3D), indicating that prenatal exposure to VPA impaired but not abolished the negative reinforcement learning in males. In parallel with the stepwise decrease in escape latency, there are significant differences in failures among days in the VPA male group (Day $F_{(1,18)} = 16.414$, $p < 0.01$; Day × Treatment $F_{(1,18)} = 9.002$, $p < 0.01$; Fig. 3E), further demonstrating that male offspring exposed to VPA are able to learn to terminate the footshocks by poking the active nosepoker, although their learning is slower than normal. We also performed the same negative reinforcement learning experiment using female offspring and observed no significant differences in the learning behavior between the VPA and the vehicle groups (escape latencies: Treatment $p < 0.05$, Fig. S2A; failures: Treatment $p < 0.05$, Fig. S2B; mistakes: Treatment $p < 0.05$, Fig. S2C; see supplemental materials for details).

Interestingly, there are no difference in the total numbers of nosepokes (sum of the active and the inactive nosepokes) between the male VPA group and the male vehicle group in all the three days of learning (Treatment $F_{(1,18)} = 1.553$, $p = 0.229$; Day × Treatment $F_{(1,18)} = 1.884$, $p = 0.167$; Day $F_{(1,18)} = 4.511$, $p < 0.01$; Fig. 3F), suggesting that the nosepoking action per se is not significantly affected by prenatal exposure to VPA. It has been repeatedly reported that the total moving distance in the open field tests was not significant changed by prenatal exposure to VPA [34,35], suggesting that the motility remains unaffected.

3.2.2. Prenatal exposure to VPA impaired response-reinforcer forming during negative reinforcement learning

To further determine the detailed changes in the negative reinforcement learning induced by exposure to VPA, the learning process on the first day was carefully analyzed. As no significant differences of learning were detected between the VPA female group and the control female group, the performances of male groups were mainly assessed in the following (including impaired response-reinforcer forming in Section 3.2.2 and habit forming in Section 3.2.3).

The results showed that there were no significant differences in the numbers of spontaneous nosepoking between the two nosepokers in male VPA and vehicle groups during the acclimation in the operant chambers, (Nosepoke $F_{(1,18)} = 0.524$, $p = 0.480$; Treatment $F_{(1,18)} = 0.001$, $p = 0.982$; Nosepoke × Treatment $F_{(1,18)} = 0.877$, $p = 0.363$; Fig. 4A). Then on the first day of negative reinforcement learning, the numbers of active nosepokes in the vehicle group were significantly more than the inactive nosepokes, indicating that mice in the vehicle group exhibit good learning performance (simple interaction effect: $p < 0.001$, Fig. 4B). Meanwhile, in the VPA group, there was no significant difference between numbers of active and inactive nosepokes, suggesting that mice in the VPA group didn't form the connection between termination of the shock and their nosepoking action at the active nosepoker (Nosepoke $F_{(1,18)} = 98.996$, $p < 0.001$; Day $F_{(1,18)} = 3.616$, $p < 0.05$; Treatment $F_{(1,18)} = 0.565$, $p = 0.462$; Day × Nosepoke × Treatment $F_{(1,18)} = 5.169$, $p < 0.05$; simple interaction effect: $p = 0.714$; Fig. 4B). Indeed, in the escape latency curves on Day 1 shown in Fig. 3C, latency of the vehicle group fell down to a platform level within 15 trials, while that of the VPA group didn't show an obvious decrease and maintained at a high level to the end of the whole learning period. In addition, mice exposed to VPA also demonstrated a significantly larger total escape latency (Fig. 3D) and more failures (Trials $F_{(1,18)} = 70.761$, $p < 0.001$; Treatment $F_{(1,23)} = 155.361$, $p < 0.001$; Treatment × Trials $F_{(1,18)} = 5.209$, $p < 0.01$; Fig. 4D) than the vehicle group on Day 1. The above results indicate that nosepoking at the active nosepoker was not successfully formed as a reinforcer on the first day of learning, in sharp contrast to that seen in the vehicle group.

3.2.3. Prenatal exposure to VPA impaired habit forming during negative reinforcement learning

On the second day of negative reinforcement learning, all male mice in the VPA group successfully learned to terminate the shocks. All of them were able to succeed over 10 consecutive trials (Fig. 5A).

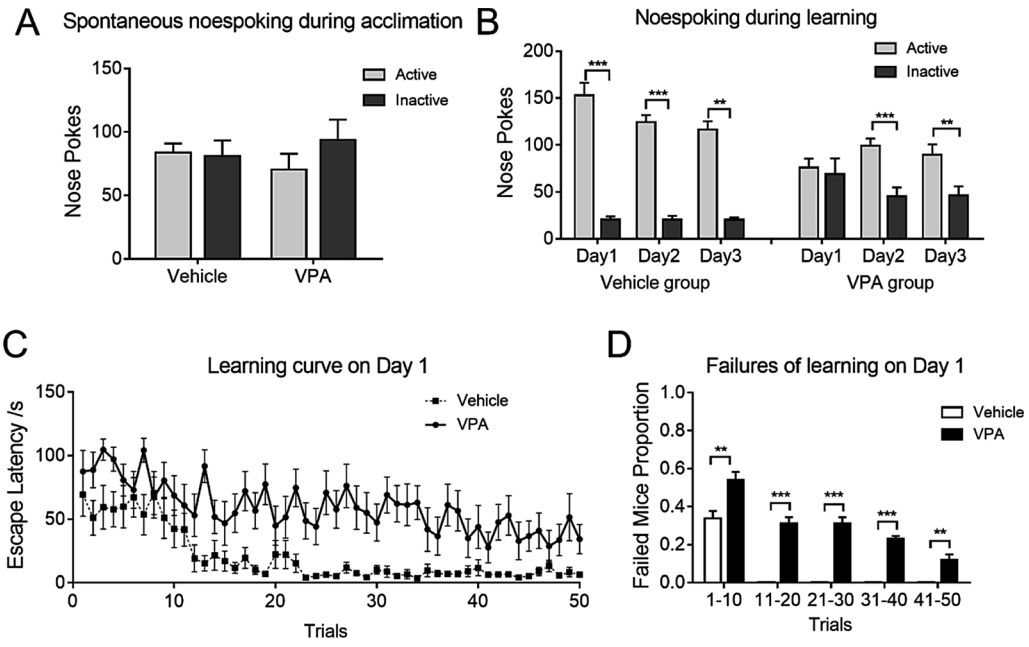

Fig. 4. Male offspring exposed to VPA demonstrated impaired stimulus-reinforcer forming during negative reinforcement learning. **(A)** There were no significant differences in the numbers of spontaneous nosepoking between the two nosepokers in both male groups. **(B)** The number of the active nosepokes in the VPA male group was significantly larger than that of the inactive ones on Day2 and Day3 but not on Day1, while the vehicle male group exhibited significantly larger numbers of the active nosepokes than those of the inactive ones during three days. **(C)** The learning curve of the vehicle male group on Day1 fell down to a plain within 15 trials, yet which of the VPA-exposed male group maintained at a high level during the whole learning period. **(D)** Male mice in the VPA group displayed more failures learning on Day1. Error bars represent SEM. RMANOVA: $^{**}p < 0.01$, $^{***}p < 0.001$. VPA: Male offspring prenatally exposed to VPA; Vehicle: Male offspring prenatally exposed to saline. VPA: n = 10; Vehicle: n = 10.

Meanwhile, the numbers of active nosepokes in the VPA group are significantly more than the inactive nosepokes on Day 2 (simple interaction effect: Day2 $p < 0.001$; Fig. 4B). According to the accepted criteria, the VPA group learned to escape from the shock and the nosepoking was formed as a reinforcer [36,37].

Even after this phase of reinforcer forming, the VPA group still made more mistakes (poking the inactive nosepoker during the shock) on Day 2 ($t_{18} = 3.081$, $p < 0.01$; Fig. 5D) and Day 3 ($t_{18} = 3.196$, $p < 0.01$; Fig. 5E) than the vehicle group. In a number of trials, mice in the VPA group firstly nosepoked at the inactive nosepoker and then at the active one on Day 2 and Day 3, but this phenomenon was never seen in the vehicle group. As a result, the escape latency curves of the VPA group exhibit much larger fluctuations than those of the vehicle group. The results clearly showed an impaired habit forming of negative reinforcement learning in male mice prenatally exposed to VPA (Fig. 5B, C).

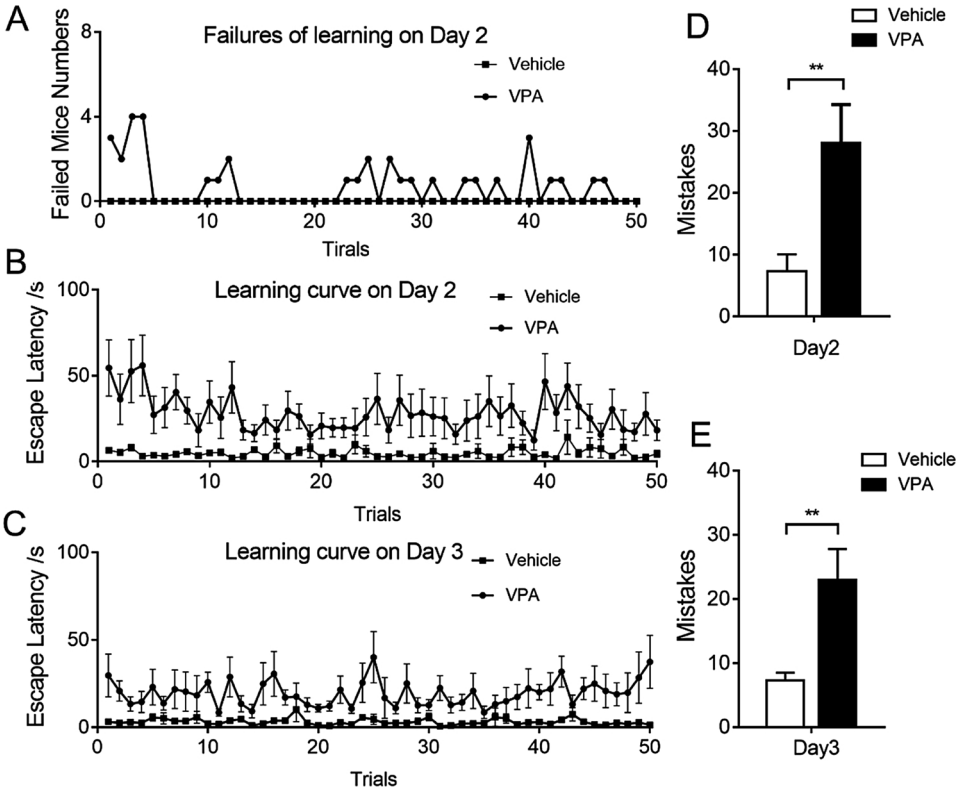

Fig. 5. Male mice exposed to VPA demonstrated impaired habit forming during negative reinforcement learning. **(A)** All of the male mice in the VPA group could terminate the shock by poking at the active nosepoker over 10 consecutive trials on Day2. (**B-C**) Escape latency curves of the VPA male group showed larger fluctuations than that of the vehicle male group on Day2 and Day3. (**D-E**) The VPA male group made more mistakes on Day2 and Day3 compared with the vehicle male group. Error bars represent SEM. Unpaired *t*-test: $^{**}p < 0.01$. VPA: Male offspring prenatally exposed to VPA; Vehicle: Male offspring prenatally exposed to saline. VPA: n = 10; Vehicle: n = 10.

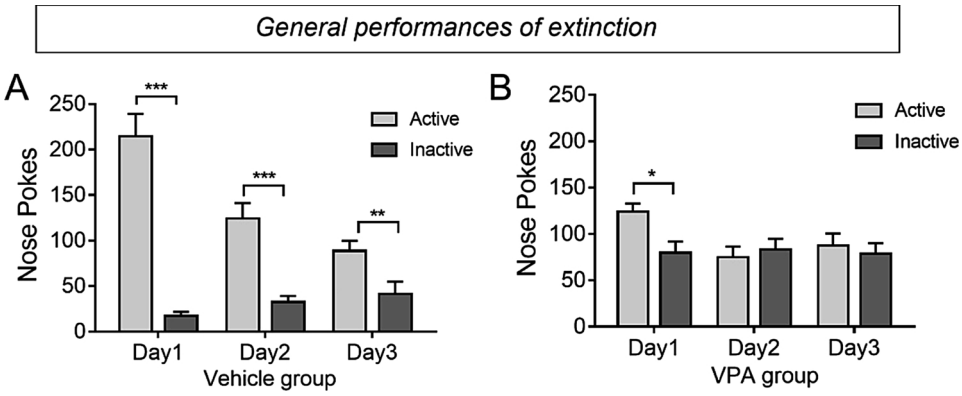

Fig. 6. The VPA male group exhibited accelerated extinction of learned operation behaviors in relative to the vehicle group. **(A)** The numbers of active nosepokes in the vehicle male group are significant more than their inactive nosepokes in all 3 days of extinction. **(B)** The number of active nosepokes is larger than that of the inactive ones in the VPA male group only on the first day of extinction. Error bars represent SEM. RMANOVA: $^{*}p < 0.05$; $^{***}p < 0.01$, $^{***}p < 0.001$. VPA: Male offspring prenatally exposed to VPA; Vehicle: Male offspring prenatally exposed to saline. VPA: n = 10; Vehicle: n = 10.

3.3. Prenatal exposure to VPA accelerated extinction of the learned escape behavior in male mice

After three days of learning, both groups learned to escape from shocks by nospoking the active nosepokers. The extinction process was then observed for three days. To our surprise, the VPA group showed a faster extinction than the vehicle group in males. On the first day of the extinction procedure, both groups showed significant differences between the two nosepokes (simple interaction effect: vehicle group $p < 0.001$; VPA group $p < 0.01$; Fig. 6A and B), indicating that animals in both male groups didn't give up or forget the learned skills. On Day 2 and Day 3, the numbers of active nosepokes in the vehicle group were still significantly more than the inactive ones (Nosepoke $F_{(1,9)} = 87.302$, $p < 0.001$; Day $F_{(1,9)} = 15.484$, $p < 0.01$; Day × Nosepoke $F_{(1,9)} = 26.982$, $p < 0.001$; simple effect: Day2 $p < 0.001$; Day3 $p < 0.01$; Fig. 6A). In contrast, there was no significant difference in the numbers of nosepoking between the two nosepokers in the VPA group on the second day and the third day (Nosepoke $F_{(1,9)} = 1.418$, $p = 0.264$; Day $F_{(1,9)} = 5.292$, $p < 0.05$; Day × Nosepoke $F_{(1,9)} = 2.242$, $p = 0.137$; simple effect: Day2 $p = 0.525$; Day3 $p = 0.858$; Fig. 6B). In the extinction process of the female offspring, there were no obvious differences between the VPA group and the vehicle group, because both groups showed significant differences between the two nosepokes in the first two days (Vehicle group: Nosepoke $p < 0.05$, Fig. S3A; VPA group: Nosepoke $p < 0.05$, Fig. S3B; see supplemental materials for details). These results indicated that only the male VPA group exhibited a faster extinction than the male vehicle group.

4. Discussion

Learning the association between one's behavior and its outcomes and shifting the reinforced behaviors into habits is vital for survival. Whereas growing researches considered reinforcement learning as an effective early behavior intervention to improve the cognitive ability, language capacity, and adaptive skills of ASD patients [1], individuals with ASD often display multiple dysfunctions of perceiving and reacting to environmental stimulus, which is highly related with deficits in response-reinforcer learning and habit forming. In this context, investigating the performance of reinforcement learning of ASD individuals using an animal model can not only help to understand how these dysfunctions influence the ability of ASD individuals to complete complex tasks, but also provide references to clinical treatment of ASD. In the present study, we presented novel results showing that the reinforcer forming and habit forming phases of negative reinforcement learning are impaired, but the extinction of the learned escape ability become faster, in a mice model of ASD produced by prenatal exposure to VPA. In the present study, significant autism-like behavioral deficits were only detected in male offspring prenatally exposed to VPA, while no significant social deficits and repetitive behavior were observed in the female offspring, which is consistent with the previous studies [33,38]. Our results also suggest that there is no significant difference in reinforcement learning in VPA females as compared with controls.

Since escape action of nosepoking and the motility remain unchanged in the VPA group, the impairment in learning and the accelerated extinction are likely associated with deficits in higher brain functions underlying the linking the animals' behavioral responses with the outcomes of such responses. The results further suggest that the rodent ASD model produced by prenatal exposure to VPA reproduces the deficits in reasoning or building the contingency between one's own behaviors and the consequent outcomes of the behavior seen in ASD patients.

It has been reported that the sensory process for aversive environmental signals in the rodent model of ASD produced by prenatal exposure to VPA remains normal [38]. In the present study, a paradigm of negative reinforcement learning was employed, in which the simulation per se is aversive. With the present of the aversive stimulation, termination of the stimulation by means of correctly nosepoking the active nosepoker is formed as a reinforcer. Positive evaluating of the reinforcer stimulus is often considered as a critical part of reinforcement learning [22]. Deficits in the reinforcer forming phase observed in the present work is in line with a number of previous researches, which discovered that individuals with ASD showed hypoactivation in the mesocorticolimbic circuitry in response to different types of reward [10], but exhibited enhanced fear conditioning and exaggerated fear memory [39].

With a remarkable delay, termination of the aversive stimulation could be successfully performed over 10 consecutive trials by mice prenatally exposed to VPA on the second day of learning, showing that terminating the aversive stimulation could be formed as a reinforcer. After that, during the habit forming process on the second day and the third day of learning mice prenatally exposed to VPA still manifest significantly more failures and mistakes than the vehicle group, suggesting a deficit in the transition from goal-direct behaviors into the habituated behaviors. This result is also consistent with previous researches with human subjects. Clinic researches showed that ASD patients exhibit reduced adaption to the familiar faces [25], which means ASD patients may have difficulties in habituating the repeated stimulus and forming behavior-outcome association.

Extinction, which is different with forgetting, is also an active learning process. The principle of extinction has long been considered and utilized in the cognitive-behavioral psychotherapeutic treatments, such as 'exposure therapy' [40]. Interestingly, our findings showed that mice prenatally exposed to VPA exhibited a faster extinction than the control mice, even though they displayed worse performances during the negative reinforcement learning. This result is apparently different with a number of previous studies, which reported that ASD individuals exhibited reduced extinction of fear memory [38,39]. The difference between the experimental paradigms may explain the difference in the results. During the extinction of fear memory described in previous

studies, the association between the context signals with aversion are reformed, while during the extinction of the negative reinforcement employed in the present work the association between the termination of aversion and a specific action, nosepoking, is reformed. During the fear conditioning and the fear extinction tasks, response of the animals is relatively simple, but in the negative reinforcement learning and its extinction tasks, the evaluated responses are initiative actions, which needs mobilization of diverse complicated cognitive and motivational processes. It has been widely reported that ASD individuals exhibit deficits in the cognitive and motivational functions [5,11], it is therefore possible that the extinction of the actively learned skills is also faster. This speculation needs to be investigated in future studies. On the other hand, one limitation in the present work may also contribute to the result. Because the training process of the negative reinforcement only lasted for three days, mice prenatally exposed to VPA might have not completed the consolidation of the learned skills, and this may also result in a faster extinction process. This issue should be clarified in future by suing a prolonged training paradigm.

Another limitation of the present work is lacking of mechanistic evidence. The abnormal performance observed in the present experiment must be underpinned by alternations in related brain structures, such as striatum, prefrontal cortex and anterior cingulate cortex. Thereinto, the striatum integrates excitatory afferents from the cortex, thalamus, and midbrain dopamine neurons, playing a critical role in mediating the motivation, motor planning, and reward evaluating functions [22,41]. In a series of genetic mice model of ASD, alterations in the synaptic transmission in the striatum and the dysfunction of dopaminergic regulation are found responsible for the impaired social communication and repetitive stereotyped behaviors. Thus, a reasonable hypothesis is that the dysfunctions of the synaptic transmission in the striatum may results in deficits in stimulus evaluating and contingency linking, thus lead to impairment in the negative reinforcement learning and extinction. In addition to that, prefrontal cortex as well as anterior cingulate cortex are also involved in both of the reinforcement learning [42] and the pathogenesis of ASD [43,44], which may jointly underline the impairment of reinforcement learning of ASD individuals. Meanwhile, it has been known that habit forming is highly related with the function of dorsal striatum [45,46], which yet is reported to be aberrant in ASD individuals [47]. Studying the role of dorsal striatum in the expression of reinforcement learning in mice exposed to VPA could be a promising topic in future researches.

In conclusion, we systematically investigated the performances of negative reinforcement learning as well as its extinction in ASD mice induced by prenatal exposure to VPA for the first time. We discovered that prenatal exposure to VPA induced profound adverse effects on the negative reinforcement learning in male offspring, including a delayed response-reinforcer forming, and an impaired habit forming. Besides, the extinction of the learned skills was found to be faster than normal males. Considering that the VPA group exhibited the normal sensory abilities, the impairment in learning and the accelerated extinction are likely caused by deficits in higher brain functions underlying associative learning. The novel results indicate that the rodent ASD model produced by prenatal exposure to VPA reproduces the deficits in reasoning or building the contingency between one's own behaviors and the consequent outcomes of the behavior seen in ASD patients. Previous studies have already shown that ASD individuals often display deficits in recognition and memory [1,5]. In addition, the rodent model produced by prenatal exposure to VPA also exhibits similar deficits in social and environmental recognition, indicating a good validity of the model [12,38]. The findings of the present study are consistent with such previous results, again demonstrating that acquisition, consolidation, and retrieval of the memory of association between nosepoking and termination of the environmental aversive stimulation are abnormal in the VPA model of ASD. Collectively, the above results strongly recommend further endeavors to use this model for a complete understanding and more effective intervention methods.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81971285, 11727813), and the Fundamental Research Funds for the Central Universities (GK202005001), Shaanxi Normal University.

References

[1] K.E. Sanchack, C.A. Thomas, Autism spectrum disorder: primary care principles, Am. Fam. Physician 94 (2016) 972–979.
[2] D.F. Mabunga, E. Gonzales, J. Kim, K.C. Kim, C. Shin, Exploring the validity of valproic acid animal model of autism, Exp. Neurobiol. 24 (2015) 285, https://doi.org/10.5607/en.2015.24.4.285.
[3] R.P. Lawson, C. Mathys, G. Rees, Adults with autism overestimate the volatility of the sensory environment, Nat. Neurosci. 20 (2017) 1293–1299, https://doi.org/10.1038/nn.4615.
[4] A. Posar, P. Visconti, Sensory abnormalities in children with autism spectrum disorder, J. Pediatr. (Rio. J) 94 (2018) 342–350, https://doi.org/10.1016/j.jped.2017.08.008.
[5] C. Cantio, S. White, G. Madsen, N. Bilenberg, J. Jepsen, Do cognitive deficits persist into adolescence in autism? Autism Res. 11 (2018), https://doi.org/10.1002/aur.1976.
[6] C. Berenguer, B. Roselló, C. Colomer, I. Baixauli, A. Miranda, Children with autism and attention deficit hyperactivity disorder. Relationships between symptoms and executive function, theory of mind, and behavioral problems, Res. Dev. Disabil. 83 (2018) 260–269, https://doi.org/10.1016/j.ridd.2018.10.001.
[7] B. Roselló, C. Berenguer, P. Navío, I. Baixauli, A. Miranda, Executive functioning, social cognition, pragmatics, and social interaction in attention deficit hyperactivity disorder and autism spectrum disorder, Curr. Dev. Disord. Reports. 4 (2017) 72–77, https://doi.org/10.1007/s40474-017-0114-1.
[8] C.E. Pugliese, L.G. Anthony, J.F. Strang, K. Dudley, G.L. Wallace, D.Q. Naiman, L. Kenworthy, Longitudinal examination of adaptive behavior in autism spectrum disorders: influence of executive function, J. Autism Dev. Disord. 46 (2016) 467–477, https://doi.org/10.1007/s10803-015-2584-5.
[9] J.R. Hollerman, W. Schultz, Dopamine neurons report an error in the temporal prediction of reward during learning, Nat. Neurosci. 1 (1998) 304–309, https://doi.org/10.1038/1124.
[10] G. Kohls, M. Schulte-Rüther, B. Nehrkorn, K. Müller, G.R. Fink, I. Kamp-Becker, B. Herpertz-Dahlmann, R.T. Schultz, K. Konrad, Reward system dysfunction in autism spectrum disorders, Soc. Cogn. Affect. Neurosci. 8 (2012) 565–572, https://doi.org/10.1093/scan/nss033.
[11] G. Kohls, L. Antezana, M.G. Mosner, R.T. Schultz, B.E. Yerys, Altered reward system reactivity for personalized circumscribed interests in autism, Mol. Autism 9 (2018) 9, https://doi.org/10.1186/s13229-018-0195-7.
[12] M. Fontes-Dutra, G. Della-Flora Nunes, J. Santos-Terra, W. Souza-Nunes, G. Bauer-Negrini, M.M. Hirsch, L. Green, R. Riesgo, C. Gottfried, V. Bambini-Junior, Abnormal empathy-like pro-social behaviour in the valproic acid model of autism spectrum disorder, Behav. Brain Res. 364 (2019) 11–18, https://doi.org/10.1016/j.bbr.2019.01.034.
[13] M. South, T. Newton, P. Chamberlain, Delayed reversal learning and association with repetitive behavior in autism spectrum disorders, Autism Res. 5 (2012), https://doi.org/10.1002/aur.1255.
[14] L. McClannahan, G. MacDuff, P. Krantz, Behavior analysis and intervention for adults with autism, Behav. Modif. 26 (2002) 9–26, https://doi.org/10.1177/0145445502026001002.
[15] T. Zane, C. Davis, F.R. Volkmar (Ed.), Differential Reinforcement Procedures of Other Behavior (DRO) BT - Encyclopedia of Autism Spectrum Disorders, Springer, New York, New York, NY, 2013, pp. 962–966, , https://doi.org/10.1007/978-1-4419-1698-3_1903.
[16] J. Cook, J. Rapp, K. Schulze, Differential negative reinforcement of other behavior to increase wearing of a medical bracelet, J. Appl. Behav. Anal. 48 (2015), https://doi.org/10.1002/jaba.228.
[17] B.W. Balleine, J.P. O'Doherty, Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action, Neuropsychopharmacology 35 (2010) 48–69, https://doi.org/10.1038/npp.2009.131.
[18] M.E. Bouton, Extinction of instrumental (operant) learning: interference, varieties of context, and mechanisms of contextual control, Psychopharmacology (Berl.) 236 (2019) 7–19, https://doi.org/10.1007/s00213-018-5076-4.
[19] M. Prsa, G.L. Galiñanes, D. Huber, Rapid integration of artificial sensory feedback during operant conditioning of motor cortex neurons, Neuron 93 (2017), https://doi.org/10.1016/j.neuron.2017.01.023 929-939.e6.
[20] B.W. Balleine, A. Dickinson, Goal-directed instrumental action: contingency and incentive learning and their cortical substrates, Neuropharmacology 37 (1998) 407–419, https://doi.org/10.1016/S0028-3908(98)00033-1.
[21] A. Gasbarri, A. Pompili, M. Packard, C. Tomaz, Habit learning and memory in

mammals: behavioral and neural characteristics, Neurobiol. Learn. Mem. 114 (2014), https://doi.org/10.1016/j.nlm.2014.06.010.
[22] A.V. Kravitz, A.C. Kreitzer, Striatal mechanisms underlying movement, reinforcement, and punishment, Physiology 27 (2012) 167–177, https://doi.org/10.1152/physiol.00004.2012.
[23] L. Gaytán-Tocavén, M.Á. López-Vázquez, M.Á. Guevara, M.E. Olvera-Cortés, Aberrant connections between climbing fibres and Purkinje cells induce alterations in the timing of an instrumental response in the rat, Exp. Brain Res. 235 (2017) 2787–2796, https://doi.org/10.1007/s00221-017-5014-4.
[24] A. Dickinson, P. Trans, R.S. Lond, Actions and habits: the development of behavioural autonomy, Philos. Trans. R. Soc. London. B, Biol. Sci. 308 (1985) 67–78, https://doi.org/10.1098/rstb.1985.0010.
[25] E. Pellicano, L. Jeffery, D. Burr, G. Rhodes, Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder, Curr. Biol. 17 (2007) 1508–1512, https://doi.org/10.1016/j.cub.2007.07.065.
[26] J. Christensen, T.K. Grønborg, M.J. Sørensen, D. Schendel, E.T. Parner, L.H. Pedersen, M. Vestergaard, Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism, JAMA 309 (2013) 1696–1703, https://doi.org/10.1001/jama.2013.2270.
[27] F. Huang, X. Chen, X. Jiang, J. Niu, C. Cui, Z. Chen, J. Sun, Betaine ameliorates prenatal valproic-acid-induced autism-like behavioral abnormalities in mice by promoting homocysteine metabolism, Psychiatry Clin. Neurosci. 73 (2019) 317–322, https://doi.org/10.1111/pcn.12833.
[28] J. Kang, E. Kim, Suppression of NMDA receptor function in mice prenatally exposed to valproic acid improves social deficits and repetitive behaviors, Front. Mol. Neurosci. 8 (2015) 17, https://doi.org/10.3389/fnmol.2015.00017.
[29] J.-W. Kim, H. Seung, K.C. Kim, E.L.T. Gonzales, H.A. Oh, S.M. Yang, M.J. Ko, S.-H. Han, S. Banerjee, C.Y. Shin, Agmatine rescues autistic behaviors in the valproic acid-induced animal model of autism, Neuropharmacology 113 (2017) 71–81, https://doi.org/10.1016/j.neuropharm.2016.09.014.
[30] L. Yao, Y. Li, Z. Qian, M. Wu, H. Yang, N. Chen, Y. Qiao, C. Wei, Q. Zheng, J. Han, Y. Tian, Z. Liu, W. Ren, Loss of control over mild aversive events produces significant helplessness in mice, Behav. Brain Res. 376 (2019), https://doi.org/10.1016/j.bbr.2019.112173 112173.
[31] K.C. Kim, P. Kim, H. Go, C. Choi, J. Park, S. Jeon, I. Dela Pena, S.-H. Han, J. Cheong, J. Ryu, C. Shin, Male-specific alteration in excitatory post-synaptic development and social interaction in pre-natal valproic acid exposure model of autism spectrum disorder, J. Neurochem. 124 (2013), https://doi.org/10.1111/jnc.12147.
[32] T. Schneider, A. Roman, A. Basta-Kaim, M. Kubera, B. Budziszewska, K. Schneider, R. Przewłocki, Gender-specific behavioral and immunological alterations in an animal model of autism induced by prenatal exposure to valproic acid, Psychoneuroendocrinology 33 (2008) 728–740, https://doi.org/10.1016/j.psyneuen.2008.02.011.
[33] C. Nicolini, M. Fahnestock, The valproic acid-induced rodent model of autism, Exp. Neurol. 299 (2018), https://doi.org/10.1016/j.expneurol.2017.04.017.
[34] S. Ha, H. Park, U. Mahmood, J.C. Ra, Y.-H. Suh, K.-A. Chang, Human adipose-derived stem cells ameliorate repetitive behavior, social deficit and anxiety in a VPA-induced autism mouse model, Behav. Brain Res. 317 (2017) 479–484, https://doi.org/10.1016/j.bbr.2016.10.004.
[35] U. Mahmood, S. Ahn, E.-J. Yang, M. Choi, H. Kim, P. Regan, K. Cho, H.-S. Kim, Dendritic spine anomalies and PTEN alterations in a mouse model of VPA-induced autism spectrum disorder, Pharmacol. Res. 128 (2017), https://doi.org/10.1016/j.phrs.2017.08.006.
[36] B.F. Skinner, Operant behavior, Am. Psychol. 18 (1963) 503–515, https://doi.org/10.1037/h0045185.
[37] M. Fox, M. Barense, M. Baxter, Perceptual attentional set-shifting is impaired in rats with neurotoxic lesions of posterior parietal cortex, J. Neurosci. 23 (2003) 676–681, https://doi.org/10.1523/JNEUROSCI.23-02-00676.2003.
[38] A. Banerjee, C.T. Engineer, B.L. Sauls, A.A. Morales, M.P. Kilgard, J.E. Ploski, Abnormal emotional learning in a rat model of autism exposed to valproic acid in utero, Front. Behav. Neurosci. 8 (2014) 387, https://doi.org/10.3389/fnbeh.2014.00387.
[39] K. Markram, T. Rinaldi, D. La Mendola, C. Sandi, H. Markram, Abnormal fear conditioning and amygdala processing in an animal model of autism, Neuropsychopharmacology 33 (2008) 901–912, https://doi.org/10.1038/sj.npp.1301453.
[40] S. Meir Drexler, T.C. Hamacher-Dang, O.T. Wolf, Stress before extinction learning enhances and generalizes extinction memory in a predictive learning task, Neurobiol. Learn. Mem. 141 (2017) 143–149, https://doi.org/10.1016/j.nlm.2017.04.002.
[41] A.C. Kreitzer, R.C. Malenka, Striatal plasticity and basal ganglia circuit function, Neuron 60 (2008) 543–554, https://doi.org/10.1016/j.neuron.2008.11.005.
[42] X. Liu, J. Hairston, M. Schrier, J. Fan, Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies, Neurosci. Biobehav. Rev. 35 (2011) 1219–1236, https://doi.org/10.1016/j.neubiorev.2010.12.012.
[43] S. Kataoka, K. Takuma, Y. Hara, Y. Maeda, Y. Ago, T. Matsuda, Autism-like behaviors with transient histone hyperacetylation in mice treated prenatally with valproic acid, Int. J. Neuropsychopharmacol. 16 (2011) 1–13, https://doi.org/10.1017/S1461145711001714.
[44] S. Nuvia, M. Bringas, M. Atzori, G. Flores, Prefrontal cortex, hippocampus, and basolateral amygdala plasticity in a rat model of autism Spectrum, Synapse 68 (2014), https://doi.org/10.1002/syn.21759.
[45] C. Nazzaro, B. Greco, M. Cerovic, P. Baxter, T. Rubino, M. Trusel, D. Parolaro, T. Tkatch, F. Benfenati, P. Pedarzani, R. Tonini, SK channel modulation rescues striatal plasticity and control over habit in cannabinoid tolerance, Nat. Neurosci. 15 (2012) 284–293, https://doi.org/10.1038/nn.3022.
[46] J. Goodman, M.G. Packard, The influence of cannabinoids on learning and memory processes of the dorsal striatum, Neurobiol. Learn. Mem. 125 (2015) 1–14, https://doi.org/10.1016/j.nlm.2015.06.008.
[47] H. Kim, Y. Lee, J.-Y. Park, J.-E. Kim, T.-K. Kim, J. Choi, J.-E. Lee, E.-H. Lee, D. Kim, K.-S. Kim, P.-L. Han, Loss of adenylyl cyclase type-5 in the dorsal striatum produces autistic-like behaviors, Mol. Neurobiol. 54 (2017) 7994–8008, https://doi.org/10.1007/s12035-016-0256-x.