

## Original Article

**Cite this article:** Kazavchinsky L, Dahan S, and Einat H. (2020) Exploring test batteries for depression- and anxiety-like behaviours in female and male ICR and black Swiss mice. *Acta Neuropsychiatrica* 32:293–302. doi: [10.1017/neu.2020.20](https://doi.org/10.1017/neu.2020.20)

Received: 8 February 2020  
Revised: 27 April 2020  
Accepted: 30 April 2020  
First published online: 7 May 2020

**Key words:**

depression; affective disorders; animal models; validity; reproducibility

**Author for correspondence:**

Haim Einat, Email: [haimh@mta.ac.il](mailto:haimh@mta.ac.il)

<sup>a</sup>These authors had equal contribution to the study.

# Exploring test batteries for depression- and anxiety-like behaviours in female and male ICR and black Swiss mice

Lydmila Kazavchinsky<sup>1,2,a</sup>, Sofi Dahan<sup>2,a</sup> and Haim Einat<sup>2</sup> 

<sup>1</sup>School of Zoology, Tel-Aviv University, Tel-Aviv, Israel and <sup>2</sup>School of Behavioural Sciences, Tel Aviv-Yaffo Academic College, Tel-Aviv, Israel

**Abstract**

**Objective and rationale:** Animal models are critical for the study of mental disorders and their treatments but are repeatedly criticized for problems with validity and reproducibility. One approach to enhance validity and reproducibility of models is to use test batteries rather than single tests. Yet, a question regarding batteries is whether one can expect a consistent individual behavioural phenotype in mice across tests that can be presumed to be part of the same construct. This study was designed to explore the relationship between the behaviours of mice across tests in some variations of test batteries for depression- and anxiety-like behaviours. **Methods:** Female and male healthy, intact, and untreated mice from the ICR and black Swiss strains were used in four separate experiments. With some variations, mice were exposed to a battery of behavioural tests representing affective- and anxiety-like behaviours. Data were analysed for differences between sexes and for correlations between behaviours within and across the tests in the battery. **Results:** No differences were found between the sexes. With very few exceptions, we found correlations within tests (when one test has more than one measure or is repeated) but not across different tests within the battery. **Conclusions:** The results cast some doubt on the utility of behavioural test batteries to represent different facets of emotional behaviour in healthy intact outbred mice, without any interventions or treatments. Additional studies are designed to explore whether stronger relationship between the tests will appear after manipulations or drug treatments.

**Significant outcomes**

- In test batteries for affective-like behaviour, healthy outbred intact mice demonstrate within test but not across tests consistency of behaviour.
- For intact outbred mice, sex differences in tests for affective-like behaviours appear to be limited.
- Additional work is needed to explore consistency of behaviour in test batteries after manipulations or drug treatment.

**Limitations**

- The study tested only two strains of mice and only in a limited number of behavioural tests and models.
- The study is limited to animals that were not exposed to any manipulation or drug, and therefore, mice represent intact outbred mice population.
- Not all tests were used in all experiments, and they were not performed in the same order. This limits the possibility for comparisons across experiments.

**Introduction**

Exploring human psychopathology, especially affective disorders, is difficult due to many reasons, among others, the complexity of representing the mental symptoms and the underlying physiology in animal models (Flaisher-Grinberg & Einat, 2009; Einat, 2014; Nestler & Hyman, 2010). Animal models are repeatedly stated to be critical in exploring the biology of mental disorders and the development of new treatments but, at the same time, are repeatedly criticized for major problems of validity and reproducibility (Young & Einat, 2019; Kafkafi *et al.*, 2018; Belzung & Lemoine, 2011). Whereas advancement in genetic and biochemical methods can improve some aspects of validity in animal models, it is still difficult to embody the full complexity of behavioural symptoms that characterize psychopathologies, especially affective disorders



in one test or model (Kara & Einat, 2013; Einat, 2014). One approach that was suggested to enhance the strength, validity, and reproducibility of animal models was to examine animals in test batteries rather than in single tests (Crawley & Paylor, 1997; Crabbe *et al.*, 1999; Einat, 2006). The utilization of test batteries was suggested to result in more predictive models, as they can represent different facets of the pathological conditions and provide better understanding of the complexity of a disorder or treatment effect (van der Staay *et al.*, 2009; Flaisher-Grinberg & Einat, 2010). This approach resembles the standard methodology in clinical studies where disorders are diagnosed based on a list of symptoms (Sadock *et al.*, 2015), and effects of treatments are evaluated based on comprehensive measures that include different domains of the disorder, such as the Hamilton depression scale or the Beck depression inventory (Williams, 2001; Wang & Gorenstein, 2013). Similar approaches are also used in humans to study normal (not pathological) mood states such as with the “Positive and Negative Affect Schedule” (Watson *et al.*, 1988) or the Profile of Mood States (Pollock *et al.*, 1979). Moreover, test batteries are also a practical approach when the number of animals is limited as frequently happens when using mice with targeted mutations and is advantageous in the context of ethical use of animals in research, as it can reduce the total number of animals needed to answer scientific questions (Crawley & Paylor, 1997; Crawley, 2000). Altogether, test batteries became a very common tool in the modelling of neuropsychiatric disorders in general and affective disorders in particular. A special emphasis must be given to the testing of intact healthy animals (without interventions), and this is for two main reasons. (1) Intact mice from standard strains are used as background strains for targeted mutations, and significant literature examined the effects of these background strains on the outcomes related to the mutations (Bailey *et al.*, 2006). As mice with targeted mutations are essential tools in research, it is critical to understand the behaviour of the underlying healthy intact mice strains. (2) Many of the standard screening models in neuropsychopharmacology research are based on examining behaviour in intact mice. For example, tests, such as the forced swim test (FST) and the elevated plus maze (EPM), introduce animals to unique situations where their behaviour is measured with or without additional interventions. The test is based on the responses of intact animals that serve as the reference for any additional interventions. It is, therefore, of critical importance to gain in-depth understanding of the complexity of the behavioural profile of intact animals (Einat *et al.*, 2018).

Yet, the common practice in the analysis of results from test batteries is to analyse each test separately at the group level. There are a number of problems with this method that were already highlighted in previous work (Stukalin & Einat, 2019; Einat *et al.*, 2018). One concern that is a component of the issue of individual variability is the expectation of consistent behavioural phenotypes of mice across tests. The hidden assumption in test batteries is that model animals will demonstrate a relatively steady phenotype across tests, and that animals that tend to be more depressed-like in one test will also be more depressed-like in another test and vice versa (Einat *et al.*, 2007; Castro *et al.*, 2012). However, the relationship between the behaviour of individual mice across tests is hardly ever assessed, and most studies only explore group effects (Einat *et al.*, 2018).

In that context, this study was designed to explore the relationship between the behaviour of individual mice across tests in some variations of test batteries for depression- and anxiety-like behaviours, in males and females of two strains of mice. We hypothesized

that individual animals will demonstrate a consistent behavioural phenotype across tests that are suggested to represent the same (or similar) state. That animals showing a more depressed- and anxiety-like behaviour in one test will also do so in other relevant tests and that animals that demonstrate a more anxious behaviour in one test will also present it in other tests. Specifically, we hypothesized that there will be a correlation among the behaviours in the FST, the tail suspension test (TST), and the sweet solution preference (SSP) test, that there will be a correlation among the behaviours in the open field (OF), the EPM, and the novelty-induced hypophagia tests, that there will be a correlation between the behaviours in the OF and in the amphetamine-induced hyperactivity test, and that there could be a reverse correlation between the behaviours in the amphetamine-induced hyperactivity test and the test related to depression (FST, TST, and SSP). The aims of the study were, therefore, to (1) examine the relationship between the behaviours of ICR and black Swiss mice in a number of dissimilar test batteries for affective- and anxiety-like behaviours and (2) study sex differences in these strains and tests. It was our hypothesis that at least for some of the tests, especially ones representing similar constructs, there will be a significant correlation between the behaviours of individual mice in one test and the behaviour in the other test. We further hypothesized that sex differences will be limited to only a few tests.

## Methods

### *Animals and procedure*

ICR (CD-1<sup>®</sup>) mice (Experiments 1–3; Envigo, Israel) and black Swiss mice (Experiment 4; local colony based on mice from Taconic Labs, USA) were used in four separate experiments. We selected two outbred strains, as outbred strains are more heterogeneous compared with inbred mice and may be a better translational model for the genetically diverse human population (Tuttle *et al.*, 2018). ICR mice were selected because within outbred strains, they are very frequently used as model animals in relevant research related to affective and anxiety disorders (Messiha *et al.*, 1990; Sugimoto *et al.*, 2008; Willner *et al.*, 1996; Sade *et al.*, 2014). Black Swiss mice were selected, because they were suggested to be a model for a number of behavioural facets of mania (Hannah-Poquette *et al.*, 2011; Ene *et al.*, 2015; Flaisher-Grinberg & Einat, 2010; Kara *et al.*, 2018a). Because of their unique behavioural profile, we expected that these mice would demonstrate consistency in behaviour across specific tests related to affective- and anxiety-like behaviours. The number of animals per experiment was based on experiments in our laboratory where behaviour of mice was examined across tests and without drugs [e.g. (Kazavchinsky *et al.*, 2019; Shemesh *et al.*, 2018)]. However, Experiment 4 was an exception with only 10 mice per group. This is because supply of black Swiss mice in Israel is very complicated, and we had to rely on our small breeding colony.

All animals were singly housed in transparent Plexiglas cages (36.5 × 14 × 20 cm) with approximately 3 cm wood shavings bedding, cardboard roles, and cotton wool enrichment. Animals were maintained under constant conditions, wherein temperature was set to 22 ± 1 °C, *ad libitum* access to food and water, and 12-h light/dark cycle. Animals had at least 1-week habituation before the start of experiments and were not disturbed or handled during this period and the experimental period. Single housing was needed because of the SSP test (see in the following for details). Procedures and experiments were performed during the light

**Table 1.** Specific behavioural tests utilised in the different experiments

Test/Experiment	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Strain and sex	ICR, M	ICR, M/F	ICR, M/F	BS, M/F
Distance in open field	+	+	+	+
Centre time in open field	+			+
EPM	+	+	+	+
FST	+	+	+	+
TST	+			
Defensive marble burying test		+	+	
SSP test		+	+	
Novelty-induced hypophagia		+	+	
Amphetamine-induced hyperactivity		+	+	

phase of the light/dark cycle starting 1 h after “lights on” and ending no later than 1 h before “lights off.” All experimental procedures followed the Israeli Ministry of Health directives and were approved by the Tel Aviv-Yaffo Academic College IACUC (protocols MTA-2015-15-3 and MTA-2012-10-3). For all experiments, in each battery, tests were conducted by the same two experimenters, one test per day, in a dedicated testing room, except the SSP test (see in the following for details) that included 48 h of exposure to sweet solution in the home cages.

### Experiment 1

Sixty, 8–9 weeks old male ICR mice were tested in four consecutive behavioural tests: spontaneous activity in the OF; EPM; FST; and TST.

### Experiments 2 and 3

Forty, 8–9 weeks old ICR mice, 20 females, and 20 males were tested in seven consecutive behavioural tests: spontaneous activity in an OF; defensive marble burying; SSP; EPM; novelty-induced hypophagia; FST, and amphetamine-induced hyperactivity. Experiment 3 was intended to be a replication of Experiment 2, but, unfortunately, due to technical error, the order of tests was altered in Experiment 3 compared with Experiment 2. As the order of tests can influence the behavioural outcomes, we could not consider these experiments as replications but as two separate experiments.

### Experiment 4

Ten male and 10 female, 8–10 week old black Swiss mice were tested in three consecutive behavioural tests: spontaneous activity in the OF; EPM; and FST.

### Behavioural tests

The tests used in this study are all prototypic tests that are frequently used in the research related to the exploration of the underlying basis of affective and anxiety disorders and in the attempts to develop better treatments. Although, in general, these tests are not ideal for many reasons, they are heavily used. For example, a PubMed search for “EPM” during the last 5 years yields 2550

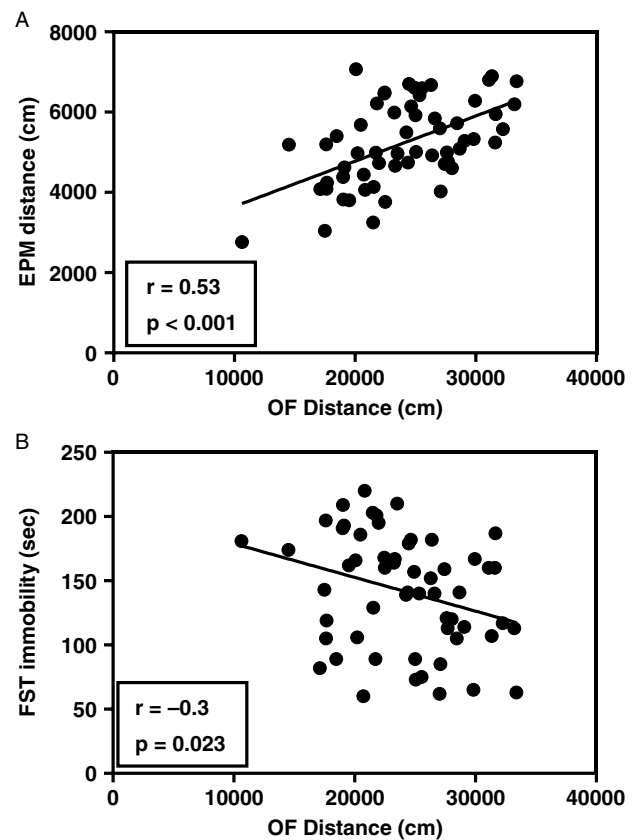
results, and similar searches for the other tests show hundreds to thousands of results, suggesting that many researchers are using them.

Specific tests that were used in each of the experiments are detailed in Table 1. All tests were conducted as detailed elsewhere (Ene *et al.*, 2015, 2016) with minor modifications: (1) in the marble burying test, marbles were placed on top of a 2 cm layer of wood chips bedding (and not 5 cm as described). (2) In the novelty-induced hypophagia test, animals had food restriction for 24 h and not 12–14 h as described earlier. In brief: (1) *Spontaneous activity* – the spontaneous activity test is a standard method to evaluate mice behaviour in the context of neuropsychiatric testing. Each mouse was individually placed in the centre of a small OF (Plexiglas box 38.5 × 38.5 cm with 35 cm walls) where behaviour was digitally recorded from above for 30 min. Recording and scoring were done automatically using a specialized software (Viewer; BioServe Bonn, Germany). Track length (measured in cm) and time in the centre (15 × 15 cm area) were recorded and analysed. At the end of session, the mouse was returned to its home cage and the arena cleaned with alcohol solution before the start of the next session. (2) *Sweet Solution Preference (SSP)* – the SSP is a non-intrusive test that can model aspects of reward seeking behaviour (Willner, 1997; Flaisher-Grinberg *et al.*, 2009). The SSP was conducted in the home cages of the mice. A bottle of 1% saccharin solution was presented in the cage for 48 h, on top of the regular water bottle and food. Bottles (saccharin solution and water) were weighed at the start of the experiment and every 24 h thereafter, resulting in three measures, time 0, time 24 h, and time 48 h. SSP was computed by dividing the amount of consumed saccharin solution by the amount of consumption of all liquids (saccharin solution + water). After 48 h, the saccharin bottles were removed. (3) *EPM* – the EPM is a standard model for anxiety-like behaviour (Lister, 1987). The EPM apparatus was a black Plexiglas plus shaped maze with two enclosed arms with walls and two open arms without walls. The maze was elevated 50 cm above the ground. The size of the arms was 35 cm long and 5 cm wide, with a centre area of 5 × 5 cm. The walls of the closed arms were 15 cm high, and the open arms had a 1 cm lip. Each mouse was placed in the centre of the maze for a 5 min session and was recorded from above using the Viewer software (BioServe Bonn, Germany). Time and frequency of visits to each arm were analysed by the software. At the end of session, each mouse was returned to its home cage and the maze cleaned with alcohol solution before the start of

the next session. (4) *Marble burying test* – the marble burying test is associated with anxiety-like behaviour (Nicolas *et al.*, 2006; Ene *et al.*, 2016). Each mouse was singly placed in a small cage with 5 cm layer of saw dust bedding and with 25 glass marbles placed in close contact in the middle of the cage. Mice were left with the marbles for a 30-min session after which they were placed back in their home cages, and the number of marbles that was more than two-thirds covered with bedding was counted. (5) *Novelty induced hypophagia test* – the hypophagia test examines an additional facet of anxiety or anhedonia (Deacon, 2011). Following 24 h with limited access to food, mice were placed individually in a plastic transparent cylinder with a 17 cm diameter and 20 cm height. Shredded sweet popcorn was spread across the cylinder's floor. Each mouse was placed in the arena for 2 min. In case the mouse ate from the popcorn, it was left in the arena for an additional 2 min (from the first bite). In case the mouse did not eat during the initial 2-min session, it was placed back in its home cage for a few minutes and then reintroduced into the arena, again for 2 min. If the mouse did not eat during the second exposure, the procedure was repeated for a third time. The latency to eat and the total duration of the eating activity was recorded. At the end of the experiment, mice were placed in their home cage and were once again provided with free access to food. (6) *FST* – the FST is one of the most frequently used tests for the screening of antidepressant activity (Porsolt *et al.*, 1977; Kara *et al.*, 2018b; Kazavchinsky *et al.*, 2019). Mice were placed individually in a transparent Plexiglas cylinder with a 45 cm height and a 20 cm diameter. The cylinder was filled with water at a temperature of 23–24 °C and a depth of 18 cm. Each mouse was placed in the water for a 6 min session during which behaviour was digitally recorded and scored for the duration of active (swim/struggle) and passive (floating or immobility) behaviours using a specialized software (FST, BioServe, Bonn, Germany). Immobility time during the last 4 min of each session served as the main measure of analysis. At the end of each session, mice were taken out of the water and placed in their home cages. Water in the cylinders was replaced every three trials. (7) *TST* – the TST is a standard test used for screening antidepressant activity (Steru *et al.*, 1985; Kara *et al.*, 2014) and was recently demonstrated to have relatively strong external validity (Stukalin *et al.*, 2020). Mice were suspended 50 cm above the floor with an adhesive tape placed approximately 1 cm from the tip of the tail for a 6-min session. Session was digitally recorded, and immobility time during the last 4 min of the test was scored manually from recordings. Mice were considered immobile only when they hung passively and completely motionless. (8) *Amphetamine-induced hyperactivity* – amphetamine-induced hyperactivity is frequently used to model manic-like behaviour (Kara *et al.*, 2014), although its validity in that context was recently questioned (Lan & Einat, 2019). Mice were placed in a small OF (Plexiglas box 38.5 × 38.5 cm with 35 cm walls) and their behaviour tracked as described earlier for the spontaneous activity test. Each session consisted of 60 min and amphetamine (1.0 mg/kg dissolved in saline to 10 ml/kg volume) was injected after the first 30 min, thereby allowing comparison of undrugged and drugged behaviour. At the end of session, the mouse was returned to its home cage, and the arena was cleaned with alcohol solution before the next session.

### Statistical analysis

Statistical analyses were performed using Statistica 13.0 software (Dell, Tulsa, OK). Pearson's correlation (two-tailed) was used to evaluate relationship of individual responses across and within



**Fig. 1.** (A) Correlation between distance travelled in the open field and distance travelled in the EPM in Experiment 1 with male ICR mice and (B) reverse correlation between distance travelled in the open field and immobility time in the FST in Experiment 1 with male ICR mice.

tests. When both females and males were used for an experiment, we performed an overall ANOVA for sex effect across all main behavioural measures (Stukalin & Einat, 2019). Power analysis for correlations was conducted using an online calculator at <https://www.masc.org.au/stats/PowerCalculator/PowerCorrelation>.

### Results

The main aim of the study was to examine the relationship between the behaviours of ICR and black Swiss mice in a number of dissimilar test batteries for affective- and anxiety-like behaviours for intact outbred mice (without interventions). Accordingly, the results are centred on correlations rather than group outcomes per-se, and group outcomes are presented in Supplementary Table 1.

#### Experiment 1 – male ICR mice

We found correlations within tests, between the behaviours of individual mice within the OF (distance and time in centre) and between the behaviours of individual mice in the EPM (open/closed time ratio and distance, open/closed time ratio and open/closed entries ratio, and open/closed entries ratio and distance). Moreover, there were correlations across tests between the distance in the open field and distance in the EPM (Fig. 1 A), and a reverse correlation between distance in the open field and immobility in the FST (Fig. 1 B). No other significant correlations were demonstrated (Table 2) including no correlation between behaviour in the

**Table 2.** Correlations within and across tests, Experiment 1

	OF Dist.	OF time centre	EPM time ratio	EPM entries ratio	EPM Dist.	FST immobility	TST immobility
OF Dist.	****	<b><math>r = 0.44</math></b> <b><math>p = 0.001</math></b>	$r = 0.15$ $p = 0.26$	$r = 0.19$ $p = 0.14$	<b><math>r = 0.53</math></b> <b><math>p &lt; 0.001</math></b>	<b><math>r = -0.3</math></b> <b><math>p = 0.024</math></b>	$r = 0.21$ $p = 0.12$
OF time centre	<b><math>r = 0.44</math></b> <b><math>p = 0.001</math></b>	****	$r = 0.13$ $p = 0.33$	$r = 0.0$ $p = 0.99$	<b><math>r = 0.28</math></b> <b><math>p = 0.03</math></b>	$r = 0.24$ $p = 0.07$	$r = 0.02$ $p = 0.91$
EPM time ratio	$r = 0.15$ $p = 0.26$	$r = 0.13$ $p = 0.33$	****	<b><math>r = 0.6</math></b> <b><math>p &lt; 0.001</math></b>	<b><math>r = 0.56</math></b> <b><math>p &lt; 0.001</math></b>	$r = 0.02$ $p = 0.86$	$r = 0.14$ $p = 0.3$
EPM entries ratio	$r = 0.19$ $p = 0.14$	$r = 0.0$ $p = 0.99$	<b><math>r = 0.6</math></b> <b><math>p &lt; 0.001</math></b>	****	<b><math>r = 0.38</math></b> <b><math>p = 0.003</math></b>	$r = 0.07$ $p = 0.57$	$r = 0.22$ $p = 0.12$
EPM Dist.	<b><math>r = 0.53</math></b> <b><math>p &lt; 0.001</math></b>	<b><math>r = 0.28</math></b> <b><math>p = 0.03</math></b>	<b><math>r = 0.56</math></b> <b><math>p &lt; 0.001</math></b>	<b><math>r = 0.38</math></b> <b><math>p = 0.003</math></b>	****	$r = 0.1$ $p = 0.44$	$r = 0.04$ $p = 0.76$
FST immobility	<b><math>r = -0.3</math></b> <b><math>p = 0.024</math></b>	$r = 0.24$ $p = 0.07$	$r = 0.02$ $p = 0.86$	$r = 0.07$ $p = 0.57$	$r = 0.1$ $p = 0.44$	****	$r = 0.03$ $p = 0.85$
TST immobility	$r = 0.21$ $p = 0.12$	$r = 0.02$ $p = 0.91$	$r = 0.14$ $p = 0.76$	$r = 0.22$ $p = 0.12$	$r = 0.04$ $p = 0.76$	$r = 0.03$ $p = 0.85$	****

OF Dist = distance travelled in the open field; OF time centre = time spent in the centre area of the open field; EPM time ratio = time in open arms divided by time in closed arms of the EPM; EPM Entries ratio = # of entries to open arms divided by # of entries to closed arms of the EPM; EPM Dist = distance travelled in the EPM; FST immobility = immobility (floating) time in the FST; TST immobility = immobility (hanging) time in the TST. **Bold** cells indicate significant correlations.

FST and behaviour in the TST. Power for all correlations was at the range of 0.80–0.85. These data suggest consistency of activity levels across tests but possibly not affective-like state.

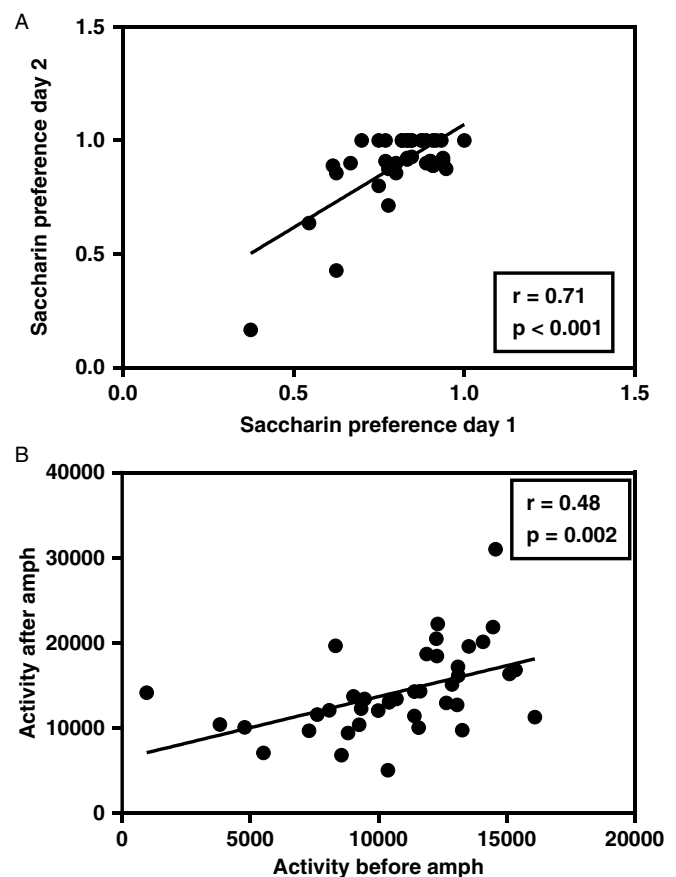
### Experiment 2

To evaluate a possible general effect of sex, we performed an overall ANOVA for sex effect across all main behavioural measures (Stukalin & Einat, 2019), including (1) open field distance; (2) marbles buried; (3) SSP day 1; (4) EPM open/closed time ratio; (5) novelty-induced hypophagia (time); (6) FST immobility time; (7) spontaneous activity before amphetamine; and (8) amphetamine-induced hyperactivity. Considering that we found no statistical difference between the sexes [ $F(8,31) = 0.8, p = 0.6$ ] and no differences in the homogeneity of variance (Levene's test, data not shown), we pooled the data for females and males.

Significant correlations were found within tests in the SSP (day 1 and day 2, Fig. 2A,  $r = 0.71, p < 0.001$ ) and in the amphetamine test (distance before amphetamine and distance after amphetamine, Fig. 2B,  $r = 0.48, p = 0.002$ ). Interestingly, there was no within-test correlation in the EPM between the open/closed time ratio and open/closed number of entries ratio ( $r = 0.21, p = 0.19$ ). Across tests, correlations (Table 3) were shown only between EPM time ratio and activity in the open field: (1) open field activity ( $r = 0.35, p = 0.03$ ) and (2) activity before amphetamine ( $r = 0.49, p < 0.001$ ). Power for all correlations was at the range of 0.80–0.92.

### Experiment 3

To evaluate a possible general effect of sex, we performed an overall ANOVA for sex effect across all main behavioural measures (Stukalin & Einat, 2019), including (1) open field distance; (2) marbles buried; (3) SSP day 1; (4) EPM open/closed time ratio; (5) novelty-induced hypophagia (time); (6) FST immobility time; (7) spontaneous activity before amphetamine; and (8) amphetamine-induced hyperactivity. Considering that we found no statistical difference between the sexes [ $F(8,31) = 1.42, p = 0.23$ ] and no differences in the homogeneity of variance except



**Fig. 2.** (A) Correlation between SSP in day 1 and day 2 of the test in Experiment 2 with female and male ICR mice and (B) correlation between activity before and after amphetamine injection in Experiment 2 with female and male ICR mice.

in the SSP measure (Levene's test, data not shown), we pooled the data for females and males.

As shown in Table 4, significant correlations were found within tests in the EPM (Fig. 3A; open/closed time ratio and open/closed

**Table 3.** Correlations within and across tests, Experiment 2

	OF Dist	SSP Mean	EPM time ratio	EPM entries ratio	Marbles	Hyponeophagia	FST	Amph before	Amph after
OF Dist	****	$r = 0.07$ $p = 0.66$	<b><math>r = 0.35</math></b> <b><math>p = 0.03</math></b>	$r = 0.21$ $p = 0.19$	$r = 0.02$ $p = 0.9$	$r = 0.1$ $p = 0.5$	$r = 0.22$ $p = 0.17$	$r = 0.15$ $p = 0.37$	$r = 0.19$ $p = 0.23$
SSP Mean	$r = 0.07$ $p = 0.66$	****	$r = 0.06$ $p = 0.7$	$r = 0.23$ $p = 0.15$	$r = 0.02$ $p = 0.89$	$r = 0.23$ $p = 0.12$	$r = 0.03$ $p = 0.82$	$r = 0.19$ $p = 0.23$	$r = 0.27$ $p = 0.09$
EPM time ratio	<b><math>r = 0.35</math></b> <b><math>p = 0.03</math></b>	$r = 0.06$ $p = 0.7$	****	$r = 0.21$ $p = 0.19$	$r = 0.1$ $p = 0.54$	$r = 0.04$ $p = 0.8$	$r = 0.06$ $p = 0.7$	<b><math>r = 0.49</math></b> <b><math>p &lt; 0.001</math></b>	$r = 0.22$ $p = 0.17$
EPM entries ratio	$r = 0.2$ $p = 0.23$	$r = 0.23$ $p = 0.15$	$r = 0.21$ $p = 0.19$	****	$r = 0.08$ $p = 0.61$	$r = 0.3$ $p = 0.06$	$r = 0.14$ $p = 0.41$	$r = 0.03$ $p = 0.88$	$r = 0.06$ $p = 0.71$
Marbles	$r = 0.02$ $p = 0.9$	$r = 0.02$ $p = 0.89$	$r = 0.1$ $p = 0.54$	$r = 0.08$ $p = 0.61$	****	$r = 0.16$ $p = 0.31$	$r = 0.2$ $p = 0.21$	$r = 0.03$ $p = 0.85$	$r = 0.02$ $p = 0.88$
Novelty-induced hypophagia	$r = 0.1$ $p = 0.5$	$r = 0.23$ $p = 0.12$	$r = 0.04$ $p = 0.8$	$r = 0.3$ $p = 0.06$	$r = 0.16$ $p = 0.31$	****	$r = 0.1$ $p = 0.52$	$r = 0.22$ $p = 0.17$	$r = 0.08$ $p = 0.63$
FST	$r = 0.22$ $p = 0.17$	$r = 0.03$ $p = 0.82$	$r = 0.06$ $p = 0.7$	$r = 0.14$ $p = 0.41$	$r = 0.2$ $p = 0.21$	$r = 0.1$ $p = 0.52$	****	$r = 0.07$ $p = 0.65$	$r = 0.05$ $p = 0.77$
Amph before	$r = 0.15$ $p = 0.37$	$r = 0.19$ $p = 0.23$	<b><math>r = 0.49</math></b> <b><math>p &lt; 0.001</math></b>	$r = 0.03$ $p = 0.88$	$r = 0.03$ $p = 0.85$	$r = 0.22$ $p = 0.17$	$r = 0.07$ $p = 0.65$	****	<b><math>r = 0.48</math></b> <b><math>p = 0.002</math></b>
Amph after	$r = 0.19$ $p = 0.23$	$r = 0.27$ $p = 0.09$	$r = 0.22$ $p = 0.17$	$r = 0.06$ $p = 0.71$	$r = 0.02$ $p = 0.88$	$r = 0.08$ $p = 0.63$	$r = 0.05$ $p = 0.77$	<b><math>r = 0.48</math></b> <b><math>p = 0.002</math></b>	****

OF Dist = distance travelled in the open field; EPM time ratio = time in open arms divided by time in closed arms of the EPM; EPM Entries ratio = # of entries to open arms divided by # of entries to closed arms of the EPM; Marbles = number of marbles in the defensive marble burying test; Hyponeophagia = latency to eat in the hyponeophagia test; FST immobility = immobility (floating) time in the FST; Amph before = distance travelled in the open field before injection of amphetamine; Amph after = distance travelled in the open field after the injection of amphetamine. Bold cells indicate significant correlations.

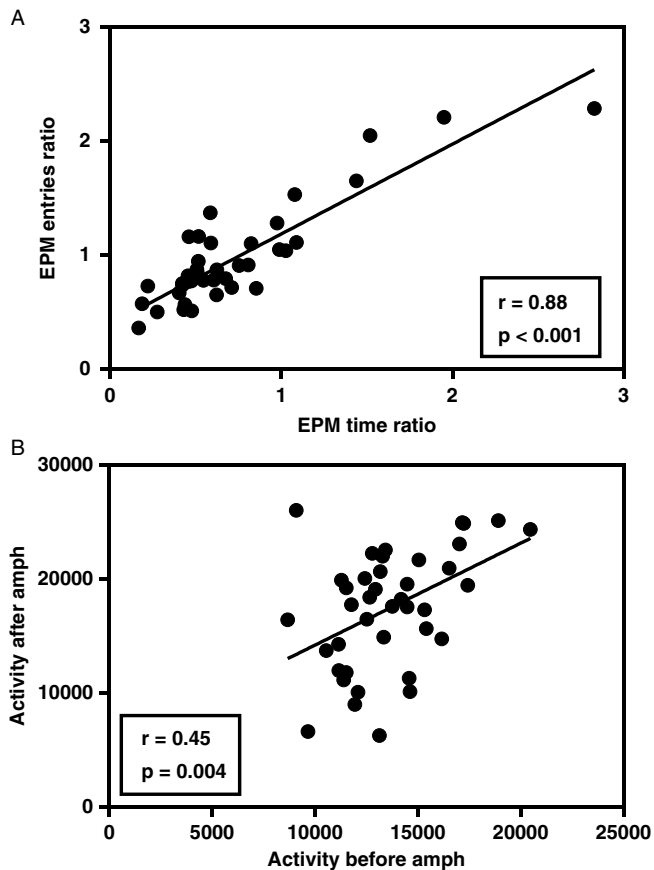
**Table 4.** Correlations within and across tests, Experiment 3

	OF Dist	SSP Mean	EPM time ratio	EPM entries ratio	Marbles	Hyponeophagia	FST	Amph before	Amph after
OF Dist	****	$r = 0.06$ $p = 0.72$	$r = 0.16$ $p = 0.34$	$r = 0.09$ $p = 0.57$	$r = 0.05$ $p = 0.77$	$r = 0.03$ $p = 0.87$	$r = 0.08$ $p = 0.62$	$r = 0.12$ $p = 0.47$	$r = 0.03$ $p = 0.81$
SSP Mean	$r = 0.06$ $p = 0.72$	****	$r = 0.11$ $p = 0.52$	$r = 0.09$ $p = 0.56$	$r = 0.12$ $p = 0.46$	$r = 0.09$ $p = 0.58$	$r = 0.06$ $p = 0.72$	$r = 0.27$ $p = 0.1$	$r = 0.09$ $p = 0.6$
EPM time ratio	$r = 0.16$ $p = 0.34$	$r = 0.11$ $p = 0.52$	****	<b><math>r = 0.88</math></b> <b><math>p &lt; 0.001</math></b>	$r = 0.01$ $p = 0.95$	$r = 0.23$ $p = 0.16$	$r = 0.24$ $p = 0.14$	$r = 0.06$ $p = 0.72$	$r = 0.001$ $p = 0.99$
EPM entries ratio	$r = 0.09$ $p = 0.57$	$r = 0.09$ $p = 0.56$	<b><math>r = 0.88</math></b> <b><math>p &lt; 0.001</math></b>	****	$r = 0.13$ $p = 0.42$	$r = 0.16$ $p = 0.33$	<b><math>r = 0.35</math></b> <b><math>p = 0.03</math></b>	$r = 0.01$ $p = 0.93$	$r = 0.03$ $p = 0.86$
Marbles	$r = 0.05$ $p = 0.77$	$r = 0.12$ $p = 0.46$	$r = 0.01$ $p = 0.95$	$r = 0.13$ $p = 0.42$	****	$r = 0.04$ $p = 0.8$	$r = 0.12$ $p = 0.44$	$r = 0.06$ $p = 0.72$	$r = 0.05$ $p = 0.75$
Novelty-induced hypophagia	$r = 0.03$ $p = 0.87$	$r = 0.09$ $p = 0.58$	$r = 0.23$ $p = 0.16$	$r = 0.16$ $p = 0.33$	$r = 0.04$ $p = 0.8$	****	$r = 0.14$ $p = 0.4$	$r = 0.14$ $p = 0.4$	$r = 0.03$ $p = 0.84$
FST	$r = 0.08$ $p = 0.62$	$r = 0.06$ $p = 0.72$	$r = 0.24$ $p = 0.14$	<b><math>r = 0.35</math></b> <b><math>p = 0.03</math></b>	$r = 0.12$ $p = 0.44$	$r = 0.14$ $p = 0.4$	****	$r = 0.15$ $p = 0.36$	$r = 0.19$ $p = 0.25$
Amph before	$r = 0.12$ $p = 0.47$	$r = 0.27$ $p = 0.1$	$r = 0.06$ $p = 0.72$	$r = 0.01$ $p = 0.93$	$r = 0.06$ $p = 0.72$	$r = 0.14$ $p = 0.4$	$r = 0.15$ $p = 0.36$	****	<b><math>r = 0.45</math></b> <b><math>p = 0.004</math></b>
Amph after	$r = 0.03$ $p = 0.81$	$r = 0.09$ $p = 0.6$	$r = 0.001$ $p = 0.99$	$r = 0.03$ $p = 0.86$	$r = 0.05$ $p = 0.75$	$r = 0.03$ $p = 0.84$	$r = 0.19$ $p = 0.25$	<b><math>r = 0.45</math></b> <b><math>p = 0.004</math></b>	****

OF Dist = distance travelled in the open field; EPM time ratio = time in open arms divided by time in closed arms of the EPM; EPM Entries ratio = # of entries to open arms divided by # of entries to closed arms of the EPM; Marbles = number of marbles in the defensive marble burying test; Hyponeophagia = latency to eat in the hyponeophagia test; FST immobility = immobility (floating) time in the FST; Amph before = distance travelled in the open field before injection of amphetamine; Amph after = distance travelled in the open field after the injection of amphetamine. Bold cells indicate significant correlations.

number of entries ratio:  $r = 0.88$ ,  $p < 0.001$ ) and in the amphetamine test (Fig. 3B; distance before amphetamine and distance after amphetamine:  $r = 0.45$ ,  $p = 0.004$ ). A non-significant trend was also observed for the correlation in the SSP between day 1

and day 2 of the test ( $r = 0.27$ ,  $p = 0.09$ ). Across tests correlation (Table 4) was shown only between EPM entries ratio and immobility in the FST ( $r = 0.35$ ,  $p = 0.03$ ). Power for most of the correlations in this experiment was high, ranging between 0.9 and 0.99,



**Fig. 3.** (A) Correlation between EPM open/closed time ratio and open/closed number of entries ratio in Experiment 3 with female and male ICR mice and (B) correlation between activity before and after amphetamine injection in Experiment 3 with female and male ICR mice.

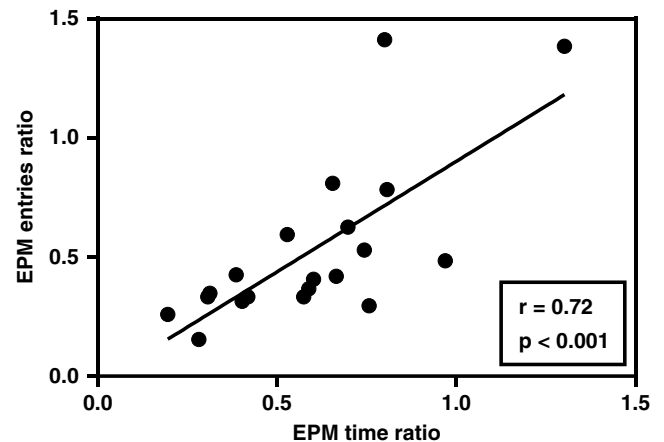
except the correlation between EPM entries ratio and FST at 0.51, and correlation between activity before and after amphetamine at 0.52.

#### Experiment 4

To evaluate a possible general effect of sex, we performed an overall ANOVA for sex effect across all main behavioural measures (Stukalin & Einat, 2019), including (1) open field distance; (2) open field centre time; (3) FST immobility; and (4) EPM open/closed time ratio. Considering that we found no statistical difference between the sexes [ $F(4,15) = 1.15, p = 0.37$ ] and no differences in the homogeneity of variance (Levene's test, data not shown), we pooled the data for females and males. A significant correlation was found within the EPM test (Fig. 4; open/closed time ratio and open/closed number of entries ratio:  $r = 0.72, p < 0.001$ ). No correlations were identified across tests (Table 5). Power for all correlations was at the range of 0.80–0.99.

#### Discussion

Within the attempts to develop better and more comprehensive animal models for neuropsychiatric disorders, one approach suggests the development and utilization of test batteries that will include separate tests for the different behavioural facets of the disorder that is being modelled (Bailey *et al.*, 2006). A background hypothesis for this approach is that a good model for a



**Fig. 4.** Correlation between EPM open/closed time ratio and open/closed number of entries ratio in Experiment 4 with female and male black Swiss mice.

neuropsychiatric disorder will respond in a similar way across dissimilar tests for the different facets of the disease (Einat, 2014). For example, one would expect that a good model for depression would demonstrate increased immobility in the FST, increased immobility in the TST, anhedonia in the SSP test, and similar changes in other tests related to depression. Increased attention to individual variability would further suggest that there should be some relationship between the behaviours of the individual animal in the different tests (Einat *et al.*, 2018). To explore this idea, this study followed individual mice exposed to a battery of behavioural tests related to affect and anxiety. We hypothesized that individual animals will demonstrate a consistent behavioural phenotype across tests, and that animals showing a more depressed- and anxiety-like behaviour in one test will also do so in other relevant tests. In contrast to our expectations, such consistent phenotypes were not demonstrated, as there were hardly any correlations between the behaviours of individual mice across tests, and even these few correlations can easily be attributed to statistical error due to multiple comparisons (Stukalin & Einat, 2019). No clear phenotypes is demonstrated in any of the strains (ICR and black Swiss), in both females and males, and in a number of variations in batteries. The only consistent correlations that were identified were within tests in some of the tests that include more than one measure.

The current findings, therefore, appear to cast some doubt on the benefits of using test batteries. Moreover, one cannot but wonder if these findings raise even broader questions regarding the validity of the tests themselves to model components of the disorders. The lack of correlation between the behaviour of individual mice across tests that are proposed to model different but close facets of a disorder presents a significant problem in the value of these tests. For example, both the FST and the TST are very similar screening tests for antidepressant activity and are based on a similar rationale related to the development of despair in depressed patients (Porsolt *et al.*, 1978; Steru *et al.*, 1985). Yet, no correlation between these behaviours in the two tests was found (Table 2). Interestingly, two recent meta-analysis studies, one regarding the FST (Kara *et al.*, 2018b) and the other regarding the TST (Stukalin *et al.*, 2020), may indicate that these tests are very different. The FST was found to have validity in qualitatively predicting antidepressant-like effects, whereas the TST was found to also quantitatively show effects of dose across different experiments. Similarly to expected correlation between the behaviour in the FST and TST, both the measure of

**Table 5.** Correlations within and across tests, Experiment 4

	OF Dist	OF Centre	EPM time ratio	EPM entries ratio	FST
OF Dist	****	$r = 0.01$ $p = 0.96$	$r = 0.37$ $p = 0.11$	$r = 0.29$ $p = 0.22$	$r = 0.004$ $p = 0.99$
OF Centre	$r = 0.01$ $p = 0.96$	****	$r = 0.13$ $p = 0.6$	$r = 0.02$ $p = 0.94$	$r = 0.21$ $p = 0.37$
EPM time ratio	$r = 0.37$ $p = 0.11$	$r = 0.13$ $p = 0.6$	****	<b><math>r = 0.72</math></b> <b><math>p &lt; 0.001</math></b>	$r = 0.05$ $p = 0.84$
EPM entries ratio	$r = 0.29$ $p = 0.22$	$r = 0.02$ $p = 0.94$	<b><math>r = 0.72</math></b> <b><math>p &lt; 0.001</math></b>	****	$r = 0.09$ $p = 0.7$
FST	$r = 0.004$ $p = 0.99$	$r = 0.21$ $p = 0.37$	$r = 0.05$ $p = 0.84$	$r = 0.09$ $p = 0.7$	****

OF Dist = distance travelled in the open field; OF Centre = time in the centre area of the open field; EPM time ratio = time in open arms divided by time in closed arms of the EPM; EPM Entries ratio = # of entries to open arms divided by # of entries to closed arms of the EPM; EPM total entries = number of entries to all arms of the EPM; FST immobility = immobility (floating) time in the FST. Bold cells indicate significant correlations.

“time in the centre” of the open field and “open/closed time” in the EPM are considered to reflect the balance between anxiety and exploratory drive (Cryan & Holmes, 2005), but no correlation was shown between them (Tables 2 and 5). It is, however, possible that the size of the arena used as open field was not large enough to detect anxiety-like behaviour, and testing in larger arenas would have resulted in different outcomes.

Unlike the results across tests, the results within tests suggest that animals do show consistent behaviour when repeatedly presented with the same task or two related tasks within the same test. In that way, there was a significant correlation between SSP in day 1 and day 2 of the tests in both Experiment 2 ( $r = 0.71$ ,  $p < 0.001$ ) and Experiment 3 ( $r = 0.37$ ,  $p = 0.02$ ). Similarly, there was a significant correlation between the “time ratio” measure and the “distance” measure in the EPM (Table 2) or between the “time ratio” measure and the “number of entries ratio” measure in the EPM (Tables 2–5). Finally, there was a significant correlation between activity levels within one session in the open field before and after an amphetamine injection (Tables 3 and 4). These results suggest that for each separate test, it is possible to expect consistent response. In that context, a recent study demonstrated consistency in the behaviour of mice across repeated exposures to the FST (Kazavchinsky et al., 2019).

Whereas the results regarding the behaviour in batteries may appear to be somewhat disappointing, it is important to remember that the mice tested in the current study were intact animals that did not experience any manipulation or treatment before testing. It is, therefore, possible that when we expose model animals to external interventions before testing, there would be a stronger correlation between the results in the different tests. It is, therefore, possible that if mice are exposed to manipulations that are related to the induction of depression-like behaviour such as stress, the mice that are more susceptible to the intervention will show consistent depression-like behaviour across tests, whereas the opposite will occur in animals that are more resilient. Similarly, drug treatment may also increase consistency of individual results across tests, as the major factor will be the response of the individual animals to the drug. Indeed, in a pilot study, we conducted with ICR mice, testing the individual variability of response to chronic oral lithium treatment, animals were exposed to the open field and the SSP tests and showed significant correlation between the individual behaviour in the two tests where mice that were more active in the open field showed higher reward seeking behaviour in the SSP

( $r = 0.4$ ,  $p = 0.03$ ). Similarly, in a recent published paper from our group, we examined the effects of a number of light and photo-period manipulations on behaviours related to depression and anxiety in the diurnal fat sand rats (*Psammomys obesus*) (Bilu et al., 2019). In that study, we report the effects of interventions of behaviours at the group level but re-analysing the data at the individual levels show that after interventions, there are correlations both within tests and across tests. Within tests, there is a correlation between the first and second “time to sink” measures in the modified FST ( $r = 0.64$ ,  $p < 0.001$ ) and between EPM “open time” and “open/closed time ratio” measures ( $r = 0.87$ ,  $p < 0.001$ ). Moreover, across tests, there is a correlation between “time to sink” in the FST and “open time” in the EPM ( $r = 0.37$ ,  $p = 0.01$ ), suggesting that the manipulations resulted in the development of individual phenotypes in the sand rats. It is, therefore, possible to suggest that although, in intact animals, the individual behaviour across tests does not reflect a coherent disease-relevant phenotype, when animals are exposed to interventions, the individual properties of the animal are exposed, either in the context of susceptibility and resilience to pathological-like interventions or in the context of response to drugs. Because the main objectives of using animal models are in the context of exploring underlying pathology or response to treatment, it is highly important at this point to design the appropriate studies that will directly evaluate the behaviour of animals in test batteries after some of the commonly used manipulations.

In summary, the current findings demonstrate within test but not across tests constancy of behaviour in intact, healthy mice, and additional work is now suggested to further explore the utility of test batteries in modelling affective- and anxiety-like behaviours.

**Supplementary material.** To view supplementary material for this article, please visit <https://doi.org/10.1017/neu.2020.20>

**Authors' contribution.** All authors were involved in the conception and design of the study. LK and SD performed the experiments under the supervision of HE. All authors took part in the analysis of the data. LK and SD wrote together the initial draft of the MS. HE wrote the final for submission.

**Financial support.** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Conflict of interest statement.** None.



**Ethical standards.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

## References

- Bailey KR, Rustay NR and Crawley JN (2006) Behavioral phenotyping of transgenic and knockout mice: practical concerns and potential pitfalls. *ILAR Journal* 47, 124–131.
- Belzung C and Lemoine M (2011) Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biology of Mood & Anxiety Disorders* 1, 9. doi:10.1186/2045-5380-1-9
- Bilu C, Einat H, Tal-Krivisky K, Mizrahi J, Vishnevskia-Dai V, Agam G and Kronfeld-Schor N (2019) Red white and blue – bright light effects in a diurnal rodent model for seasonal affective disorder. *Chronobiology International* 15, 1–8.
- Castro JE, Diessler S, Varea E, Marquez C, Larsen MH, Cordero MI and Sandi C (2012) Personality traits in rats predict vulnerability and resilience to developing stress-induced depression-like behaviors, HPA axis hyper-reactivity and brain changes in *perk1/2* activity. *Psychoneuroendocrinology* 37, 1209–1223. doi:10.1016/j.psyneuen.2011.12.014
- Crabbe JC, Wahlsten D and Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. *Science* 284, 1670–1672.
- Crawley JN (2000) What's Wrong with My Mouse? Behavioral Phenotyping of Transgenic and Knockout Mice. New York: Wiley-Liss.
- Crawley JN and Paylor R. (1997) A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Hormones and Behavior* 31, 197–211.
- Cryan JF and Holmes A (2005) The ascent of mouse: advances in modelling human depression and anxiety. *Nature Reviews Drug Discovery* 4, 775–790.
- Deacon RM (2011) Hyponeophagia: a measure of anxiety in the mouse. *Journal of Visualized Experiments* 51(51), 2613. doi:10.37971/2613.
- Einat H (2006) Modelling facets of mania – new directions related to the notion of endophenotypes. *Journal of Psychopharmacology* 20, 714–722.
- Einat H (2014) New ways of modeling bipolar disorder. *Harvard Review of Psychiatry* 22, 348–352. doi:10.1097/HRP.0000000000000059
- Einat H, Ezer I, Kara NZ and Belzung C. (2018) Individual responses of rodents in modelling of affective disorders and in their treatment: prospective review. *Acta Neuropsychiatrica* 18, 1–6.
- Einat H, Shaldubina A, Bersudsky Y and Belmaker RH (2007) Prospects for the development of animal models for the study of bipolar disorder. In: Soares JC and Young A (eds). *Bipolar Disorders: Basic Mechanisms And Therapeutic Implications*, 2nd Edn. New York: Taylor & Francis.
- Ene HM, Kara NZ, Barak N, Reshef Ben-Mordechai T and Einat H (2016) Effects of repeated asenapine in a battery of tests for anxiety-like behaviours in mice. *Acta Neuropsychiatrica* 28, 85–91. doi:10.1017/neu.2015.53
- Ene HM, Kara NZ and Einat H (2015) Introducing female black swiss mice: minimal effects of sex in a strain-specific battery of tests for mania-like behavior and response to lithium. *Pharmacology* 95, 224–228.
- Flaisher-Grinberg S and Einat H (2009) Mice models for the manic pole of bipolar disorder. In: Gould TD (ed.) *Mice Models For Mood And Anxiety Disorders*. Berlin: Springer Press.
- Flaisher-Grinberg S and Einat H (2010) Strain specific battery of tests for separate behavioral domains of mania. *Frontiers in Psychiatry* 1, 1–10.
- Flaisher-Grinberg S, Overgaard S and Einat H (2009) Attenuation of high sweet solution preference by mood stabilizers: a possible mouse model for the increased reward-seeking domain of mania. *Journal of Neuroscience Methods* 177, 44–50.
- Hannah-Poquette C, Anderson GW, Flaisher-Grinberg S, Wang J, Meinerding TM and Einat H (2011) Modeling mania: further validation for black swiss mice as model animals. *Behavioural Brain Research* 223, 222–226.
- Kafkafi N, Agassi J, Chesler EJ, Crabbe JC, Crusio WE, Eilam D, Gerlai R, Golani I, Gomez-Marin A, Heller R, Iraqi F, Jaljuli I, Karp NA, Morgan H, Nicholson G, Pfaff DW, Richter SH, Stark PB, Stiedl O, Stodden V, Tarantino LM, Tucci V, Valdar W, Williams R W, Wurbel H and Benjamini Y (2018) Reproducibility and replicability of rodent phenotyping in preclinical studies. *Neuroscience & Biobehavioral Reviews* 18, 30657-1.
- Kara NZ and Einat H (2013) Rodent models for mania: practical approaches. *Cell and Tissue Research* 354, 191–201. doi:10.1007/s00441-013-1594-x
- Kara NZ, Flaisher-Grinberg S, Anderson GW, Agam G and Einat H (2018a). Mood-stabilizing effects of rapamycin and its analog temsirolimus: relevance to autophagy. *Behavioural Pharmacology* 29, 379–384. doi:10.1097/FBP.0000000000000334
- Kara NZ, Karpel O, Toker L, Agam G, Belmaker RH and Einat H (2014) Chronic oral carbamazepine treatment elicits mood-stabilising effects in mice. *Acta Neuropsychiatrica* 26, 29–34. doi:10.1017/neu.2013.23
- Kara NZ, Stukalin Y and Einat H (2018b). Revisiting the validity of the mouse forced swim test: systematic review and meta-analysis of the effects of prototypic antidepressants. *Neuroscience & Biobehavioral Reviews* 84, 1–11. doi:10.1016/j.neubiorev.2017.11.003
- Kazavchinsky L, Dafna A and Einat H (2019) Individual variability in female and male mice in a test-retest protocol of the forced swim test. *Journal of Pharmacological and Toxicological Methods* 95, 12–15. doi:10.1016/j.vascn.2018.11.007
- Lan A and Einat H (2019) Questioning the predictive validity of the amphetamine-induced hyperactivity model for screening mood stabilizing drugs. *Behavioural Brain Research* 362, 109–113. doi:10.1016/j.bbr.2019.01.006
- Lister RG (1987) The use of a plus-maze to measure anxiety in the mouse. *Psychopharmacology (Berl)*. 92, 180–185.
- Messiha FS, Martin WJ and Bucher KD (1990) Behavioral and genetic inter-relationships between locomotor activity and brain biogenic amines. *General Pharmacology* 21, 459–464.
- Nestler EJ and Hyman SE (2010) Animal models of neuropsychiatric disorders. *Nature Neuroscience* 13, 1161–1169.
- Nicolas LB, Kolb Y and Prinszen EP (2006) A combined marble burying-locomotor activity test in mice: a practical screening test with sensitivity to different classes of anxiolytics and antidepressants. *European Journal of Pharmacology* 547, 106–115.
- Pollock V, Cho DW, Reker D and Volavka J (1979) Profile of mood states: the factors and their physiological correlates. *Journal of Nervous and Mental Disease* 167, 612–614. doi:10.1097/00005053-197910000-00004
- Porsolt RD, Bertin A and Jalfre M (1977) Behavioral despair in mice: a primary screening test for antidepressants. *Archives Internationales de Pharmacodynamie et de Therapie* 229, 327–336.
- Porsolt RD, Bertin A and Jalfre M (1978) “Behavioural despair” in rats and mice: strain differences and the effects of imipramine. *European Journal of Pharmacology* 51, 291–294.
- Sade Y, Kara NZ, Toker L, Bersudsky Y, Einat H and Agam G (2014) Beware of your mouse strain; differential effects of lithium on behavioral and neurochemical phenotypes in Harlan ICR mice bred in Israel or the USA. *Pharmacology, Biochemistry and Behavior* 124C 36–39. doi:10.1016/j.pbb.2014.05.007
- Sadock JB, Sadock VA and Ruiz P (2015) *Kaplan and Sadock's Synopsis of Psychiatry*. Philadelphia: Lippincott & Williams.
- Shemesh G, Kara N and Einat H (2018) Chronic stress may not be a factor in the behavioral response to chronic lithium in ICR mice. *Pharmacology* 102, 281–286. doi:10.1159/000492717
- Steru L, Chermat R, Thierry B and Simon P (1985) The tail suspension test: a new method for screening antidepressants in mice. *Psychopharmacology (Berl)* 85, 367–370.
- Stukalin Y and Einat H (2019) Analyzing test batteries in animal models of psychopathology with multivariate analysis of variance (MANOVA): one possible approach to increase external validity. *Pharmacology, Biochemistry and Behavior* 178, 51–55. doi:10.1016/j.pbb.2017.11.003
- Stukalin Y, Lan A and Einat H (2020) Revisiting the validity of the mouse tail suspension test: systematic review and meta-analysis of the effects of prototypic antidepressants. *Neuroscience & Biobehavioral Reviews* 29, 30544–30545.
- Sugimoto Y, Kajiwara Y, Hirano K, Yamada S, Tagawa N, Kobayashi Y, Hotta Y and Yamada J (2008) Mouse strain differences in immobility and sensitivity to fluvoxamine and desipramine in the forced swimming test:

- analysis of serotonin and noradrenaline transporter binding. *European Journal of Pharmacology* **592**, 116–122. doi:[10.1016/j.ejphar.2008.07.005](https://doi.org/10.1016/j.ejphar.2008.07.005)
- Tuttle AH, Philip VM, Chesler EJ and Mogil JS** (2018) Comparing phenotypic variation between inbred and outbred mice. *Nature Methods* **15**, 994–996. doi:[10.1038/s41592-018-0224-7](https://doi.org/10.1038/s41592-018-0224-7)
- Van Der Staay FJ, Arndt SS and Nordquist RE** (2009) Evaluation of animal models of neurobehavioral disorders. *Behavioral and Brain Functions* **5**, 11.
- Wang YP and Gorenstein C** (2013) Assessment of depression in medical patients: a systematic review of the utility of the Beck Depression Inventory-II. *Clinics (Sao Paulo)* **68**, 1274–1287. doi:[10.6061/clinics/2013\(09\)15](https://doi.org/10.6061/clinics/2013(09)15)
- Watson D, Clark LA and Tellegen A** (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* **54**, 1063–1070. doi:[10.1037//0022-3514.54.6.1063](https://doi.org/10.1037//0022-3514.54.6.1063)
- Williams JB** (2001) Standardizing the Hamilton Depression Rating Scale: past, present, and future. *E European Archives of Psychiatry and Clinical Neuroscience* **251**, 116–12.
- Willner P** (1997) Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. *Psychopharmacology (Berl)*, **134**, 319–29.
- Willner P, Moreau JL, Nielsen CK, Papp M and Sluzewska A** (1996) Decreased hedonic responsiveness following chronic mild stress is not secondary to loss of body weight. *Physiology & Behavior* **60**, 129–34.
- Young JW and Einat H** (2019) The importance and depth of reproducibility in rodent models of psychiatric diseases. *Pharmacology, Biochemistry and Behavior* **178**, 1–2. doi:[10.1016/j.pbb.2019.01.009](https://doi.org/10.1016/j.pbb.2019.01.009)