Published in final edited form as:

J Neurosci Methods. 2011 April 15; 197(1): 21–31. doi:10.1016/j.jneumeth.2011.01.019.

Integrated Behavioral Z-Scoring Increases the Sensitivity and Reliability of Behavioral Phenotyping in mice: Relevance to Emotionality and Sex

Jean-Philippe Guilloux^{1,2,*}, Marianne Seney^{1,*}, Nicole Edgar^{1,3}, and Etienne Sibille^{1,3,§}

¹Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213

²Univ Paris-Sud EA 3544, Fac. Pharmacie, Châtenay-Malabry cedex F-92296, France

³Center for Neuroscience, 3811 O'Hara street, BST W1643, University of Pittsburgh, Pittsburgh, PA 15213

Abstract

Defining anxiety- and depressive-like states in mice (“emotionality”) is best characterized by the use of complementary tests, leading sometimes to puzzling discrepancies and lack of correlation between similar paradigms. To address this issue, we hypothesized that integrating measures along the same behavioral dimensions in different tests would reduce the intrinsic variability of single tests and provide a robust characterization of the underlying “emotionality” of individual mouse, similarly as mood and related syndromes are defined in humans through various related symptoms over time. We describe the use of simple mathematical and integrative tools to help phenotype animals across related behavioral tests (syndrome diagnosis) and experiments (meta-analysis). We applied z-normalization across complementary measures of emotionality in different behavioral tests after unpredictable chronic mild stress (UCMS) or prolonged corticosterone exposure - two approaches to induce anxious-/depressive-like states in mice. Combining z-normalized test values, lowered the variance of emotionality measurement, enhanced the reliability of behavioral phenotyping, and increased analytical opportunities. Comparing integrated emotionality scores across studies revealed a robust sexual dimorphism in the vulnerability to develop high emotionality, manifested as higher UCMS-induced emotionality z-scores, but lower corticosterone-induced scores in females compared to males. Interestingly, the distribution of individual z-scores revealed a pattern of increased baseline emotionality in female mice, reminiscent of what is observed in humans. Together, we show that the z-scoring method yields robust measures of emotionality across complementary tests for individual mice and experimental groups, hence facilitating the comparison across studies and refining the translational applicability of these models.

Keywords

emotionality; anxiety; depression; mice; stress; ucms; corticosterone; normalization; behavior

© 2011 Elsevier B.V. All rights reserved.

§To whom correspondence should be addressed. sibilleel@upmc.edu.

* contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Introduction

Evaluation of behavioral and physiological parameters relating to emotion-like processes in animals is typically performed with several tests and without comprehensive analysis across paradigms. Mouse behavior is multimodal and full quantifiable assessment of emotionality (which covers anxiety-like and/or depressive-like behavior) is only possible when the same animal is exposed to multiple behavioral tests covering a wide range of behaviors over several days (Crawley et al., 1997; Crawley and Paylor, 1997). However, closely-related behavioral parameters that are specific to each test and that relate to aspects of emotionality (for instance, entries into open field center or into open arms of the elevated plus maze) do not necessarily agree within animals and/or across time, leading to behavioral noise that is difficult to interpret. This behavioral variability can be caused by the time of day, the experimenter and recent activity in the colony, or may represent false positive/negative results in experiments with small numbers of test subjects (less than 10). More often, the cause of the variability is unknown, but it is thought to reflect natural fluctuations over the underlying mean value. Thus, as mice can be in different emotional states within short periods of time (Ramos, 2008), correlation analyses of behavioral parameters obtained from different tests may result in lack of statistical power and affect principal component types of integrative analyses (Carola et al., 2002). Hence, to assess emotionality, we need simple and comprehensive tools that allow integration of behavioral parameters obtained in multiple (but complementary) behavioral tests.

It is important to note that convergent - rather than consistent - sets of symptoms are at the core of the clinical characterization of the human illness. Indeed, contrary to a putative “consistent” organ deficiency phenotype (i.e. muscle or liver function for instance), the manifestation of changes in emotionality can vary over time. This is one of the reasons why depression is diagnosed in humans by a set of variable symptoms (4–5 out of a list of 10) over time (2 weeks or more). It is not based on a single consistent behavior, but rather by a set of converging behavioral observations that together define a depressive syndrome. Here we are trying to provide a method to operationalize this approach to rodent studies to increase the translational value of the models.

Here, we z-normalized results from rodent behavioral tests, experiments and cohorts, with the goal of assessing the emotionality dimension of mice. Z-normalization is a methodology that standardize observations obtained at different times and from different cohorts, thus allowing their comparison and/or compilation. Its value is obtained by subtracting the average of observations in a population from an individual raw value and then dividing this difference by the population standard deviation. This type of normalization, compared to percentiles, allows data on different scales to be compared. Indeed, based on a translational application of the illness definition (i.e. a syndrome as a collection of variable symptoms), we actually may not expect systematically the same or “consistent” behavioral outputs, but we do expect converging results from emotionality measures over time. This may also be the reason why principal component analyses (PCA) have not been successful at summarizing emotionality behavioral data, as one of the assumptions under PCA is that “consistent” values should be systematically obtained (Carola et al., 2002; Milner and Crabbe, 2008). Instead, the proposed z-score approach relies on testing whether a particular experimental group deviates from mean behaviors in converging directions across tests and time.

Furthermore, taking example from clinical study meta-analyses, where z-normalization is used to compile related measures performed in different studies but that assess the same illness dimension, we evaluated the possibility of comparing integrated measures of emotionality across different rodent experiments. We first validated the approach using two common methods to induce anxious-/depressive-like states in mice - UCMS and chronic

corticosterone exposure - (David et al., 2009; Mineur et al., 2006) and then report its use in providing additional analytical opportunities, such as differentiating more subtle sex differences under baseline and induced high emotionality across studies.

2 Methods

2.1 Animals

Male and female C57BL/6NTac mice (Taconic, Hudson, NY) were used. Mice were maintained under standard conditions (12/12 h light/dark cycle, 22 ± 1 °C, food and water *ad libitum*, 4–5 animals/cage), and the protocol was approved by the University of Pittsburgh Institutional Animal Care and Use Committee (protocol #0801794, Animal Assurance # A3187-01). Two different cohorts were used for each model (UCMS and corticosterone exposure) for a total of 4 cohorts. Baseline sex differences were established in 3 cohorts (Figs 1, 2, 4, 5).

2.2 Estrous cycle

Estrous state was monitored in female mice by vaginal smears in selected tests (Goldman et al., 2007). Briefly, 10 μ l of saline was flushed into the vagina and then placed on a glass slide and coverslipped. Observation of stages of the estrous cycles was performed under light microscope with a 10X objective without staining. Vaginal smears were performed on the day of behavioral testing and on the day after to more accurately assess estrous stage.

2.3 Unpredictable Chronic mild stress (UCMS)

UCMS mimics the role of socio-environmental stressors in precipitating a depressive-like syndrome that shares characteristics with human depression, such as increased fearfulness/anxiety-like behavior, decreased consumption of palatable food and physiological changes (Mineur et al., 2006; Pothion et al., 2004; Santarelli et al., 2003). Importantly, the UCMS-induced syndrome is blocked and reversed by chronic antidepressant treatment (Surget et al., 2009). UCMS consisted of a 4-week regimen (or 6 weeks when fluoxetine was administered, see below) of pseudo-random unpredictable mild stressors: forced bath (~2 cm water in cage for 15 minutes), wet bedding, predator odor (1 hour exposure to fox urine), light cycle changes, social stress (rotate mice into previously occupied cage), tilted cage (45°), mild restraint (50mL Falcon tube with air hold for 15 minutes) and bedding changes (Joeyen-Waldorf et al., 2009; Surget et al., 2009).

2.4 Fluoxetine treatment

Fluoxetine (Sigma, St. Louis, MO) was dissolved and administered in the drinking water (18 mg/kg/d) for 4 weeks, 15 days after the onset of UCMS, in order to reverse and block the development of the depressive-like phenotype (Santarelli et al., 2003; Surget et al., 2009).

2.5 Corticosterone Treatment

Corticosterone (Sigma, St. Louis, MO) was dissolved in vehicle (0.45% β -cyclodextrin) and delivered (35 μ g/ml) in drinking water for 4 weeks, based on (David et al., 2009). Liquid consumption was monitored and bottles were changed every 3 days. This test models the elevated corticosteroid levels seen in some subjects with major depression (Antonijevic, 2006; Brouwer et al., 2005). Chronic antidepressant treatment reverses the corticosterone-induced elevated emotionality (David et al., 2009; Gourley and Taylor, 2009).

2.6 Behavior

Behavioral testing was performed using elevated plus-maze, open field and novelty suppressed feeding, three commonly used tests in the literature to measure components of

emotionality. Tests were performed 3–5 days apart to minimize the impact of a previous test on the response for the same animals. Tests were performed in the order described below:

2.6.1 Elevated Plus Maze Test (EPM)—Behavior in the EPM was measured using a cross maze with two open and two closed arms (30 × 5 cm arms). Time spent in the open arms and ratio of entries into the open arms (entries into open arms divided by total entries into any arm × 100) during a 10 min test measured anxiety-related behaviors (Sibille et al., 2000). The total number of arm entries was used as an index of locomotor activity.

2.6.2 Open Field Paradigm (OF)—The time and distance ratio spent in the center of a 43×43 cm open chamber were recorded for 10 min to evaluate anxiety-related behaviors (center was defined as a 32×32 cm central arena). Here, we report time in the center of the open field and ratio of distance traveled in the center (distance traveled in the center divided by the total distance traveled × 100). The total distance traveled was used as an index of locomotor activity (David et al., 2009).

2.6.3 The Novelty Suppressed Feeding test (NSF)—As an index of emotionality, the latency to start eating a food pellet was monitored in food-deprived animals in a brightly illuminated chamber. Briefly, animals were food-deprived for 16 hours prior to the test. Testing was performed in a 50×50 cm box covered with bedding and illuminated by a 70-watt lamp. Mice were tested individually by placing them in the box for a period of 10 minutes. The latency to eat was timed. Immediately afterwards, the animal was transferred to its home cage and the amount of food consumed in the subsequent 5 minutes was measured, serving as a control for change in appetite as a possible confounding factor.

2.7 Emotionality and Locomotion Z-score calculation

Z-scores are dimensionless mathematical tools that allow for mean-normalization of results within studies and for subsequent comparison of related data across studies. Z-scores are standardized scores (by the group mean and group standard deviation) and no normal assumption is made. They indicate how many standard deviations (σ) an observation (X) is above or below the mean of a control group (μ).

$$z = \frac{X - \mu}{\sigma}$$

X represents the individual data for the observed parameter. μ and σ represents the mean and the standard deviation for the control group, respectively. Here as we investigated stress and sex effects, the male control group was defined as the control group (except for Figure 2 where effects of antidepressant in females were monitored and thus, the control group was the unstressed female group). Z-score values were calculated for test parameters measuring emotionality and locomotor activity. The directionality of scores was adjusted so that increased score values reflected increased dimensionality (emotionality or locomotion). Standard measures of anxiety-/depressive-like behaviors (Crupi et al., 2010; Post et al., 2010) were used here, but the approach can be customized to other tests, based on each lab's expertise.

For instance, decreased normalized OF center activity and increased NSF latency were converted into positive standard deviation changes compared to group means indicating increased emotionality. To avoid any weighted effect of locomotion on anxious behavior in the OF and EPM, distance ratios (center/total distance in the OF; or open arm entry ratio in the EPM) are typically used (Crupi et al., 2010; Post et al., 2010), thus integrated parameters were normalized for the locomotor component. In the NSF, the time necessary to initially

approach the food pellet is orders of magnitude smaller than the time to overcome the conflict of the aversive environment, thus locomotor activity is typically not controlled for; rather appetite and food consumption are measured across groups. The selection of these specific dimensions was made based on the fact that these parameters are the most frequently used in the neuropsychopharmacology field so that readers could easily identify these components in their own studies. Furthermore, we selected for EPM and OF parameters that are associated in some principal component analyses studies (PCA) with "anxious behaviour" or "anxious locomotor activity" such as time in the open arms or time in the center (Carola, 2002); however other studies could not fully dissociate "unambiguously parameters fully reflecting 'activity' or 'anxiety'" (Milner and Crabbe, 2008). Finally, PCA analyses found that both these components had similar loads on anxious behaviour (Milner and Crabbe, 2008).

As an example, z-score in the open-field (Z_{OF}) was calculated for each animal using normalization of "time in the center" (TC) and "distance in periphery/total distance ratio" (DR) values.

$$Z_{OF} = \frac{\left(\frac{X-\mu}{\sigma}\right)TC + \left(\frac{X-\mu}{\sigma}\right)DR}{\text{Number of parameters}}$$

Similarly, in the elevated plus maze for each animal, Z_{epm} calculation was performed using normalization of "time in the open arms" (TOA) and "Open/Closed arms entries ratio" (ER) values.

Finally, in the novelty suppressed feeding, Z_{nsf} was calculated for each animal using normalization of the latency time to eat the pellet.

Individual emotionality scores were then calculated by averaging z-score values across tests, thus leveraging potential biases induced by a single test. An emotionality z-score was calculated for each animal based on 3 different tests:

$$\text{Emotionality Score} = \frac{Z_{OF} + Z_{EPM} + Z_{NSF}}{\text{Number of tests}}$$

Finally group emotionality score means (and standard deviations) were obtained by averaging individual values within each group for each experiment (Figures 1–3) and by integrating similar groups across experiments (Figures 4–5).

2.8 Statistical analysis

Based on the experiment, the number of groups and treatments applied, Student t-tests, one-way or two-way ANOVA (sex, treatment, estrous state as co-factor), followed by post-hoc PLSD (when main effects were observed significant) and χ^2 analysis, were performed.

3 Results

3.1 Z-score normalization confirmed elevated emotionality and identified robust sex differences in the UCMS model of depression

We employed emotionality z-scores to investigate the potential of combining results across different behavioral tests for anxiety- and depressive-like behaviors using the UCMS model, a validated paradigm to elicit anxious-/depressive-like behaviors. For this first analysis (Fig

1), results from independent tests were as follows: in the OF, stress exposure did not affect time and relative distance traveled in the center of the OF (Fig 1A); in the EPM, UCMS-exposed animals spent significantly less time ($p < 0.05$) and entered proportionately less often ($p < 0.05$) into the open arms compared to controls (Fig 1A); in the NSF, there was a trend for UCMS-exposed animals to have increased latencies to eat the pellet ($p = 0.09$; Fig 1A).

Z-score normalization was then performed, first, within the respective behavioral parameters, hence transforming absolute values to numbers of standard deviations from the control means (see methods). As described in the methods, the control group used was the male non-stressed group. Male and female results are pooled in figures 1A–D, which underlie the slightly positive value for the "no stress" group that combine mean of z-normalized values of both sex (See further characterization in §3.4). This first step yielded, as expected, the same statistical p-values as before normalization (Fig 1B). We then averaged these normalized behavioral parameter z-scores to obtain a single value per mouse and per behavioral test (Fig 1C). Analyses of test-specific z-scores indicated a significant effect of stress exposure on EPM ($p < 0.05$), with UCMS-exposed animals displaying higher z-scores than controls. There was no effect of stress on OF z-score ($p > 0.4$), but a trend for an effect of stress on NSF z-score ($p = 0.09$). Finally, these values were averaged to obtain a single "Emotionality Score" for each mouse, describing the integrated output of that experiment (Fig. 1D). Note that all three tests are weighted similarly, as within-test parameters were averaged at the prior step (Fig. 1C). Here, the analysis of the combined normalized measures of emotionality resulted in augmented statistical significance of the stress main effect ($p < 0.015$ versus $0.03 < p < 0.09$, depending on the test). We further compared males and females and show that the UCMS effect on emotionality was driven by a significant effect of stress exposure on emotionality score in females ($p < 0.05$), but not males ($p > 0.1$; Fig 1E). In summary, we showed that, in this particular experiment, z-scoring across complementary behavioral dimensions provided a more robust overall assessment of the effect of stress on emotionality (i.e. less sensitive to outlier values).

3.2 Antidepressant reversal of elevated emotionality z-scores

As the observed effect of stress was greater in females (Fig 1E), we studied the reversal effects of chronic fluoxetine administration on female mice with altered behavior induced by stress (Fig 2). In this second cohort of UCMS-exposed animals, female mice were exposed to similar stressors and behavioral testing, and an independent group was exposed to chronic fluoxetine at the onset of the UCMS syndrome (Surget et al., 2009). Emotionality z-scores were calculated as described above. Results from individual tests were as follows: in the OF, no significant effects of UCMS or fluoxetine were observed on parameters measured (Fig 2A); in the EPM, UCMS-exposed females spent significantly less time ($p < 0.001$) and entered proportionately less often ($p < 0.01$) into the open arms compared to controls, while chronic fluoxetine treatment blocked the development of those effects for both parameters ($p < 0.05$); in the NSF, fluoxetine-treated animals displayed lower latency to start eating the food pellet compared to saline treated UCMS-exposed mice ($p < 0.05$). These variable results are somewhat typical to behavioral studies, so to assess whether results reflected behavioral noise or fluctuations over a more stable underlying trend, we performed z-score normalization, first, within behavioral parameters (yielding the same statistical p-values as before normalization; Fig 2B), and, then, averaged results to obtain a single value per mouse and per behavioral test (Fig 2C). Fluoxetine-treated and UCMS-exposed mice did not differ from controls on measures of emotionality in the OF and in the NSF. No significant effect was observed in the OF (Fig. 2C). The final z-score integration revealed a significant effect of UCMS, suggesting a stable underlying effect, although modest in this case. Control unstressed mice were compared to fluoxetine-treated stressed mice and no difference were

observed for all experiments (Fig 2A–D). As expected, chronic SSRI treatment reversed the elevated stress-induced z-score measures of emotionality ($p < 0.01$, Fig 2D) (or blocked the development; see methods), bringing values back to baseline control levels. Together, this provides an additional example of using z-score normalization to extract a robust underlying trend out of more variable individual measures, and critically providing a pharmacological validation and a face validity of its application.

3.3 Elevated emotionality z-scores and increased statistical significance in the corticosterone-induced syndrome

To test the reliability of the z-normalization method across models, we then derived emotionality z-scores using behavioral results obtained in the chronic corticosterone model as an additional test case, since chronic exposure reliably increases emotionality in mice (David et al., 2009; Gourley and Taylor, 2009). In light of sex differences in the UCMS model, we present data analyzed by sex (Fig. 3). In the OF, corticosterone-exposed animals spent less time in the center than controls (main effect of corticosterone exposure, $p < 0.05$; Fig 3A). This result was driven by the fact that corticosterone-exposed males spent significantly less time in the open than control males ($p < 0.01$). There was also a significant effect of sex on time spent in the center ($p < 0.05$; Fig 3A), driven by a sex difference in corticosterone exposure, with treated males spending less time in the open than treated females ($p < 0.05$). For distance ratio in the OF, there was a main effect of treatment, with corticosterone-exposed animals having smaller distance ratios than controls ($p < 0.01$; Fig 3A). As for the time in the center of the OF, this result was driven by corticosterone-exposed males having smaller distance ratios than control males ($p < 0.01$). There was also a significant sex difference in distance ratio, driven by a sex difference in corticosterone exposure, with treated males having smaller ratios than treated females ($p < 0.01$; Fig 3A). In the EPM, corticosterone-exposed animals spent significantly less time in the open arms than controls (main effect of corticosterone exposure, $p < 0.05$; Fig 3A). Corticosterone-treated animals also had a smaller open arm entry ratio than controls (overall, $p < 0.01$; males, $p < 0.05$; females, $p < 0.05$). In the NSF, there was a significant sex difference in latency ($p < 0.01$; Fig 3A), driven by the fact that corticosterone exposed males had longer latencies than treated females ($p < 0.001$). There was also a trend for an effect of treatment on latency ($p = 0.09$), with corticosterone exposed males having longer latencies than control males ($p < 0.05$).

Using these results, z-score transformation was performed, first, within behavioral parameters, yielding, as expected, exactly the same statistical p-values as before normalization (Fig 3B). Z-scores were then averaged to obtain a single value per behavioral test, and group differences were assessed (Fig 3C). There was a significant main effect of treatment on OF z-score ($p < 0.01$), driven by the fact that corticosterone-treated males had higher z-scores than control males ($p < 0.01$). There was also a significant main effect of sex on OF z-score ($p < 0.05$), driven by corticosterone-treated males having higher OF z-scores than treated females ($p < 0.001$). There was a significant main effect of treatment on EPM z-score, with corticosterone-treated animals displaying higher z-scores than controls ($p < 0.01$; males, $p = 0.09$; females, $p < 0.05$). There was a trend for a main effect of treatment on NSF z-score ($p = 0.09$), driven by corticosterone-exposed males having higher NSF z-scores than untreated males ($p < 0.05$). There was also a significant main effect of sex on NSF z-score ($p < 0.01$), driven by a sex difference in corticosterone exposure, with treated males having higher NSF z-scores than treated females ($p < 0.001$). Finally, these values were averaged to obtain a single “emotionality score” per mouse, then per experimental group (Fig. 3D). Note that all three tests are weighted similarly, as within-test parameters were averaged at the prior step (Fig. 3C). Here, z-scoring confirmed that the smaller and variable effect sizes in

female mice reflected an overall less robust, although still significant, impact ($p < 0.05$) of corticosterone exposure in female mice.

Combining normalized measures in group emotionality z-scores augmented the overall statistical significance of the corticosterone main effect (Table 1, $p < 0.0001$ versus $0.008 < p < 0.09$, depending on the test), thus emphasizing the low but measurable convergence of behavior between tests, and confirming that individual mice displayed similar directionality of effects across tests, suggesting that integrated z-scores provide a robust assessment (i.e. less sensitive to outlier values) of the effect of corticosterone on emotionality.

Looking at the effect of sex, the emotionality z-score significance was lower in three out of five cases for the different parameters measured in the NSF, OF and EPM. The underlying cause of these more robust statistical parameters appears to rely on the fact that z-score normalization lowers the overall variance of the behavioral measurements (Table 1). This is consistent with the notion that the underlying changes in emotionality were similar in the three tests, but that test-specific variability in measures partly obscured its accurate measurement within individual tests. Hence, under these experimental conditions, emotionality z-scoring provided the best combination of low p-values, due to lower coefficient of variation of integrated values. Because the number of behavioral tests included in the analysis can also affect the overall statistical power and z-score values, we compared z-score values (Supplemental Figure 1 or Supplemental Table 1) and statistical results (Supplemental Table 2) obtained by averaging data from either 2 or 3 behavioral tests. As expected the lowest variability in measures was observed when averaging z-scores from 3 tests, as demonstrated by a combination of a lower coefficient of variation and a higher statistical significance.

3.4 Combining emotionality z-scores across experiments and models provided additional analytical opportunities across independent studies: quantitative differences and sexual dimorphism in the UCMS and corticosterone models of altered mood states

Results from the UCMS and corticosterone exposure studies suggested differences in opposite directions between males and females across the two models (Figs. 1, 3). To further investigate this potential sexual dimorphism, we took advantage of the fact that, similar to clinical meta-analysis approaches, normalized z-scores can allow for comparison and pooling of results across experiments, hence increasing sample size and analytical power. Indeed, in meta-analysis, the same measure [e.g. a scaled measure of depressive state for example] is used in different studies, while here the same measure, emotionality Z-score, [e.g. an equivalent of a scaled diagnosis of animal behavioral state] was derived in different experiments and subjects and compared across studies. Here, combined experimental group sizes ranged from 22 to 51 animals per sex for each model. Integrated emotionality z-scores from two experiments using the UCMS paradigm confirmed that stress increases emotionality in both sexes (male: $p < 0.01$; female: $p < 0.001$) and revealed a higher female response to UCMS (Fig. 4A, Female > Male, $p < 0.01$) (Dalla et al., 2005; Joeyen-Waldorf et al., 2009). On the other hand, integrating results from two independent corticosterone experiments confirmed the robust effect in males ($p < 0.001$), strengthened the conclusion of less robust female results ($p < 0.05$), and revealed a significant sex difference in increased emotionality (Fig. 4A, Male > Female, $p < 0.001$) while no group $\times$ sex interaction was observed ($p = 0.17$, Fig. 4A). No baseline sex difference was observed ($p = 0.31$), although more female mice displayed baseline emotionality scores greater than 0.5 ($p < 0.001$; see next section).

In Fig 4B we present an alternate use of z-scoring, where locomotion z-scores were derived from related locomotor parameters across two tests (total ambulatory distance in the OF and total entries in EPM). Integrated locomotion z-scores from these same experiments using the

UCMS and corticosterone exposure paradigms showed that (i) females had overall higher baseline locomotion activity compared to males ($p < 0.001$), (ii) corticosterone induced a decrease in locomotor activity in males ($p < 0.001$), but not in females ($p = 0.50$), and that (iii) chronic stress induced no effect on locomotion parameters in either sex (males: $p = 0.06$; females $p = 0.33$). Estrous state did not correlate with altered behavior in individual tests.

Together, these results provide examples of the application of z-scoring across experiments initially performed separately. Here, for instance, integrated z-scores across behavioral tests and experiments revealed significant sex differences that were at best at trend level in individual experiments.

3.5 Emotionality z-scores combined across cohorts revealed qualitative baseline sex differences

Elevated baseline emotionality was observed in female mice in some behavioral tests, but did not reach significance for individual experiments. Notably, highlighting consistent sex differences in mouse behavior can be difficult, as it requires a large group of animals, control for estrous state in females, and the direction of change can vary across different tests (Palanza, 2001; Voikar et al., 2001). Here, we speculated that integrating results across these tests may reveal baseline differences, either in mean group differences or in the distribution of z-scores within groups. We thus integrated emotionality z-scores over three experiments and focused on control animals ($n = 42$ males, 34 females; Fig. 5). Results revealed higher baseline emotionality in females (male, $z = 0.00$; female $z = 0.574$; $p < 0.001$). We next assessed the distributions of emotionality scores (“low”, scores below -0.5 ; “normal”, scores between -0.5 and $+0.5$; “high”, scores greater than $+0.5$). This alternate use of z-scores revealed a highly significant shift to higher emotionality in females ($\chi^2 = 16.8$, $df = 2$, $p < 0.001$), indicative of high baseline emotionality in 71% of female mice, but only in 24% of males. Notably, this difference did not correspond with estrous state in individual female mice, and in fact, represent integrated measures over a period of several days, hence encompassing most estrous states within individual mice.

4 Discussion

4.1 Principles of z-scoring methods adapted for behavioral measurements

To address inherent difficulties in behavioral phenotyping of mice over time and to obtain summarized results over tests and studies, we propose a method based on z-normalization principles for the quantification of behaviors in an integrative manner along coherent dimensions, such as shown here for emotionality. Indeed, it is often difficult to reconcile positive or intermediate findings across tests, especially for behavioral measures that are subject to known variability. We show that applying a z-normalization method across complementary behavioral measures related to aspects of emotionality can facilitate the “diagnosis” of an animal state. Emotionality in animal models is classically reflected by altered behavior monitored in different paradigms that can be restored after antidepressants (as performed here), by variations in physiological parameters (HPA axis, locomotor activity), and potentially through identification of brain region-specific genomic biomarkers of altered behavior (Krishnan et al., 2007; Sibille et al., 2009). Interestingly, since human mood is defined as an emotional state over time that is remote from proximal stimuli, we speculate that rodent emotionality z-scores may in fact represent the closest homolog of human mood. Indeed, they integrate behavioral states observed in various and multiple paradigms over several days of testing, including across various neuroendocrine states (i.e. sex hormones), hence capturing a more stable and enduring state of emotionality in mice. Similarly, the combined analysis of converging behavior can be assimilated to the clinical characterization of the human illness, which is diagnosed by a set of variable symptoms over

time. So it is not based on a single consistent behavior, but rather by a set of converging behavioral observations that together define a depressive syndrome. A recent study aimed at the same goal by combining different behavioral tests into a single apparatus ("triple test"; composed of OF, EPM and Light/Dark test physically linked together) to phenotype animal's behavior using a similar comprehensive strategy based on multiple testing (Fraser et al., 2010). The future value of such a test will need to be assessed in multiple studies. Notably, it is still based on a one-time assessment of animal's behavior, in contrast to our proposed analytical method for behavioral assessment over time.

Furthermore, emotionality z-scores – by allowing pooling of cohorts – can strengthen the reliability of effects and increase analytical opportunities. Specifically, we showed that emotionality z-scores reduced test-to-test variability for measures of dependent variables that are sensitive to multiple known (and unknown) environmental factors (time of day, animal facility-related events, experimenter, estrous phase, etc).

The rationale for using z-normalization, instead of, for instance, calculating percentage of control response for each parameter and averaging them across groups and cohorts, is that the standard deviations of z-normalization values are similar across parameters and tests. Thus, averaging z-values avoids weighted effects of one parameter or one test over another. Z-score methodology also differs from multivariate statistics, such as principal component analysis, which are performed to investigate whether behavioral measures assess a single and intangible entity. However, as discussed in the introduction, "emotionality" is by definition an underlying state that is vulnerable to timely fluctuations due to variable environmental and biological stimuli, and that may manifest as different behaviors, or "symptoms" over time. So we actually do not seek, and may not even expect, high correlation across tests, but rather we expect convergence of results obtained with integrated z-scores. Instead, we expect that a true underlying emotionality state will be revealed through similarities in effect size and directions over cohorts and tests over time. Similarly, other types of multivariate analysis, like MANOVA, assume linear relationships among dependent variables and covariates; therefore, when the relationship deviates from linearity – which might happen due to fluctuation in animal's behavior - the power of the analysis will be compromised. Similarly, multivariate analyses rely on similar assumptions of correlation rather than convergence, and therefore may not work as well. Finally, z-normalization within and across different behavioral tests results in a single score per mouse which may be seen as a quantitative "diagnosis" of their emotionality, a translational – and of course limited - equivalent to the way human depression is quantified by structured interviews, such as the Global Assessment of Functioning scale or the Hamilton Depression Rating Scale.

Defining new tools for behavioral analysis in neuropsychopharmacology necessitates assessing their validity. As emotionality is an integrative behavioral entity that is composed of different parameters, such as anxiety, depression, and fear of novel environment, that are measured over time, our methodology has a strong face validity as it combines these multiple aspects. Predictive validity of the z-score method has been tested here by looking at antidepressant reversal of stress induced-emotionality.

4.2 Proof of concept: Application of z-scoring methods to two different models of altered mood disorders and to behavioral sex differences

Here we applied behavioral z-scoring methods to the quantification of emotionality in two rodent models that are frequently used to induce higher anxiety- and depressive-like behavior in mice. UCMS is based on chronic psychosocial stress, while chronic corticosterone exposure relies on neuroendocrine dysfunction. In our studies, main effects of either UCMS or corticosterone exposure were observed in most, but not all, of the single

tests performed within individual cohorts (Fig 1A and 3A). Z-score normalization appeared to increase the robustness of the analyses by decreasing the variability of integrated measures (Fig 1D and 3D; Table 1). Combining emotionality z-scores across experiments revealed significant sex differences in response to stress or corticosterone exposures (Fig. 4), hence demonstrating the value of the approach at detecting effects that were either not significant or at trend levels in experiments performed separately. Notably, the goal of these integrated analyses is not to “increase statistical significance”, but rather to extract underlying trends out of apparently variable results. For instance, we showed that, compared to males, females were more sensitive to chronic stress, but less sensitive to chronic corticosterone administration. Greater female behavioral and physiological stress sensitivity has previously been reported (Dalla et al., 2005; Joeyen-Waldorf et al., 2009), associated with higher corticosterone levels after various stressors (Handa et al., 1994). Although corticosterone administration can induce high emotionality in males (David et al., 2009; Gourley et al., 2008; Murray et al., 2008; Zhao et al., 2008) and females (Ardayfio and Kim, 2006), sex differences had not yet been directly studied. Using emotionality z-scores, we were able to combine individual experiments and showed that females were overall less sensitive than males to corticosterone exposure, thus consolidating a large literature on sex-related differences in response to glucocorticoids and in HPA-axis dysregulation in rodents (Galea et al., 1997; Liu et al., 2006) and humans (Binder et al., 2009; Bremmer et al., 2007; Young and Ribeiro, 2006; Young et al., 2007). Of course, since both the rodent and human literature are mostly male-biased, an alternative interpretation is that males are more sensitive to corticosterone exposure and less to the effects of chronic stress. In summary, these results support our hypothesis that z-scoring normalization of related behavior can reveal consistent and stable changes in underlying emotionality in mice, despite apparent and often unexplained variability. Using this approach augmented the translational validity of the models by suggesting similar directions for sex differences that are observed in human subjects.

4.3 Application of behavioral z-scores

By definition, z-scores normalize results across tests, experiments and cohorts, as they take into consideration differences from mean group values in terms of numbers of standard deviations from the control mean (See methods). The approach is not new by itself, as it is commonly applied in clinical and epidemiological studies. An important feature of its application to behavioral data is to ensure conformity with the direction of effects. For instance, increased emotionality in mice is revealed by decreased values of dependent variables in some tests (OF and EPM) and by increased values in other tests (NSF), and thus all measures indicative of increased emotionality should be reflected by positive numbers of standard deviations from the control group mean. While our z-score calculation here was based on data extracted from three behavioral tests commonly used in neuropsychopharmacology, it could be extended to other behavioral tasks that measure other parameters related to emotionality, such as number of fecal boluses in a new environment, elevated O maze, marble burying, light/dark transition, etc. Here we use the term of emotionality to cover both anxiety-like and depressive-like behaviors, as they are difficult to fully dissociate in rodents, but the approach can be expanded to include more specific tests. However, while multiple testing in the same animal is necessary to robustly assess emotionality, experimenters should verify that response to one behavioral test was not altered by prior testing. Notably, the integrated approach does not detract from the analysis of distinct components of individual tests, which may reveal nuances in behavioral responses and changes. The potential application of behavioral z-scoring is quite extensive, from dissociating emotionality-related behavior in stressed/control animals, knockout or transgenics/wild-type (using combined group scores), to identify consistent outliers or segregate resilient from responder animals to environmental exposure or pharmacological

treatment (e.g. through score histograms), or to measure antidepressant-predictive behaviors or antidepressant reversal of induced behavioral syndromes.

Behavioral z-scores can also be applied to other behavioral dimensions (memory tests, addiction tests, etc.). Here we briefly showed a similar approach applied to locomotion. Indeed, while emotionality z-scores already include locomotor-controlled parameters extracted from each test, normalization of locomotion-specific parameters can further evaluate overall locomotor activity under baseline conditions, between males and females and after experimental manipulations for instance (i.e. UCMS or corticosterone exposure).

Some of the critical aspects and potential limitations that need to be further characterized include, amongst others: (i) reliable behavioral protocols across experiments (Wahlsten et al., 2006), (ii) combining results across strains (Milner and Crabbe, 2008; Yalcin et al., 2008), especially in the context of sex differences (Voikar et al., 2001), and (iii) careful consideration of behavioral dimensions to be integrated.

5 Conclusion

In summary, we suggest that using an easy-to-apply and "generalizable" z-score methodology can increase the reliability and comprehensiveness of behavioral testing from a variety of non-exclusive tasks, but along cohesive behavioral dimensions, for complex behaviors such as emotionality of animals. Here, the application of this method to quantify emotionality in mice demonstrated that mice display subtle baseline emotionality sex differences that are similar to those observed in humans (Brebner, 2003), support the use of chronic mild stress as a comprehensive model to induce an anxiety-like/depressive-like syndrome, and points to corticosterone exposure as a model for male neuroendocrine vulnerability to mood disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute of Mental Health (NIMH) MH084060 (ES), MH085111 (ES) and MH092984 (MS), and by the National Institute of Neurological Disorders and Stroke NS07391 (MS). The funding agency had no role in the study design, data collection and analysis, decision to publish and preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIMH or the National Institutes of Health.

We thank Dr. Denis David for providing β -cyclodextrin, Dr George C. Tseng for a critical statistical and both for their critical comments on the manuscript.

References

- Antonijevic IA. Depressive disorders -- is it time to endorse different pathophysiologies? Psychoneuroendocrinology. 2006; 31:1–15. [PubMed: 15950391]
- Ardayfio P, Kim KS. Anxiogenic-like effect of chronic corticosterone in the light-dark emergence task in mice. Behav Neurosci. 2006; 120:249–256. [PubMed: 16719689]
- Binder EB, Kunzel HE, Nickel T, Kern N, Pfennig A, Majer M, Uhr M, Ising M, Holsboer F. HPA-axis regulation at in-patient admission is associated with antidepressant therapy outcome in male but not in female depressed patients. Psychoneuroendocrinology. 2009; 34:99–109. [PubMed: 18829172]
- Brebner J. Gender and emotions. Pers Individ Differ. 2003; 34:387–394.

- Bremmer MA, Deeg DJ, Beekman AT, Penninx BW, Lips P, Hoogendijk WJ. Major depression in late life is associated with both hypo- and hypercortisolemia. *Biol Psychiatry*. 2007; 62:479–486. [PubMed: 17481591]
- Brouwer JP, Appelhof BC, Hoogendijk WJ, Huyser J, Endert E, Zuketto C, Schene AH, Tijssen JG, Van Dyck R, Wiersinga WM, Fliers E. Thyroid and adrenal axis in major depression: a controlled study in outpatients. *Eur J Endocrinol*. 2005; 152:185–191. [PubMed: 15745924]
- Carola V, D'Olimpio F, Brunamonti E, Mangia F, Renzi P. Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behav Brain Res*. 2002; 134:49–57. [PubMed: 12191791]
- Crawley JN, Belknap JK, Collins A, Crabbe JC, Frankel W, Henderson N, Hitzemann RJ, Maxson SC, Miner LL, Silva AJ, Wehner JM, Wynshaw-Boris A, Paylor R. Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl)*. 1997; 132:107–124. [PubMed: 9266608]
- Crawley JN, Paylor R. A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Horm Behav*. 1997; 31:197–211. [PubMed: 9213134]
- Crupi R, Cambiaghi M, Spatz L, Hen R, Thorn M, Friedman E, Vita G, Battaglia F. Reduced adult neurogenesis and altered emotional behaviors in autoimmune-prone B-cell activating factor transgenic mice. *Biol Psychiatry*. 2010; 67:558–566. [PubMed: 20185032]
- Dalla C, Antoniou K, Drossopoulou G, Xagoraris M, Kokras N, Sfikakis A, Papadopoulou-Daifoti Z. Chronic mild stress impact: are females more vulnerable? *Neuroscience*. 2005; 135:703–714. [PubMed: 16125862]
- David DJ, Samuels BA, Rainer Q, Wang JW, Marsteller D, Mendez I, Drew M, Craig DA, Guiard BP, Guilloux JP, Artymyshyn RP, Gardier AM, Gerald C, Antonijevic IA, Leonardo ED, Hen R. Neurogenesis-dependent and -independent effects of fluoxetine in an animal model of anxiety/depression. *Neuron*. 2009; 62:479–493. [PubMed: 19477151]
- Fraser LM, Brown RE, Hussin A, Fontana M, Whittaker A, O'Leary TP, Lederle L, Holmes A, Ramos A. Measuring anxiety- and locomotion-related behaviours in mice: a new way of using old tests. *Psychopharmacology (Berl)*. 2010; 211:99–112. [PubMed: 20454890]
- Galea LA, McEwen BS, Tanapat P, Deak T, Spencer RL, Dhabhar FS. Sex differences in dendritic atrophy of CA3 pyramidal neurons in response to chronic restraint stress. *Neuroscience*. 1997; 81:689–697. [PubMed: 9316021]
- Goldman JM, Murr AS, Cooper RL. The rodent estrous cycle: characterization of vaginal cytology and its utility in toxicological studies. *Birth Defects Res B Dev Reprod Toxicol*. 2007; 80:84–97. [PubMed: 17342777]
- Gourley SL, Taylor JR. Recapitulation and reversal of a persistent depression-like syndrome in rodents. *Curr Protoc Neurosci*. 2009; Chapter 9 Chapter 9: Unit 9 32.
- Gourley SL, Wu FJ, Kiraly DD, Ploski JE, Kedves AT, Duman RS, Taylor JR. Regionally specific regulation of ERK MAP kinase in a model of antidepressant-sensitive chronic depression. *Biol Psychiatry*. 2008; 63:353–359. [PubMed: 17889834]
- Handa RJ, Burgess LH, Kerr JE, O'Keefe JA. Gonadal steroid hormone receptors and sex differences in the hypothalamo-pituitary-adrenal axis. *Horm Behav*. 1994; 28:464–476. [PubMed: 7729815]
- Joeyen-Waldorf J, Edgar N, Sibille E. The roles of sex and serotonin transporter levels in age- and stress-related emotionality in mice. *Brain Res*. 2009; 1286:84–93. [PubMed: 19577546]
- Krishnan V, Han MH, Graham DL, Berton O, Renthall W, Russo SJ, Laplant Q, Graham A, Lutter M, Lagace DC, Ghose S, Reister R, Tannous P, Green TA, Neve RL, Chakravarty S, Kumar A, Eisch AJ, Self DW, Lee FS, Tamminga CA, Cooper DC, Gershenfeld HK, Nestler EJ. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell*. 2007; 131:391–404. [PubMed: 17956738]
- Liu HH, Payne HR, Wang B, Brady ST. Gender differences in response of hippocampus to chronic glucocorticoid stress: role of glutamate receptors. *J Neurosci Res*. 2006; 83:775–786. [PubMed: 16498614]
- Milner LC, Crabbe JC. Three murine anxiety models: results from multiple inbred strain comparisons. *Genes Brain Behav*. 2008; 7:496–505. [PubMed: 18182070]

- Mineur YS, Belzung C, Crusio WE. Effects of unpredictable chronic mild stress on anxiety and depression-like behavior in mice. *Behav Brain Res.* 2006; 175:43–50. [PubMed: 17023061]
- Murray F, Smith DW, Hutson PH. Chronic low dose corticosterone exposure decreased hippocampal cell proliferation, volume and induced anxiety and depression like behaviours in mice. *Eur J Pharmacol.* 2008; 583:115–127. [PubMed: 18289522]
- Palanza P. Animal models of anxiety and depression: how are females different? *Neurosci Biobehav Rev.* 2001; 25:219–233. [PubMed: 11378178]
- Post AM, Weyers P, Holzer P, Painsipp E, Pauli P, Wulsch T, Reif A, Lesch KP. Gene-environment interaction influences anxiety-like behavior in ethologically based mouse models. *Behav Brain Res.* 2010
- Pothion S, Bizot JC, Trovero F, Belzung C. Strain differences in sucrose preference and in the consequences of unpredictable chronic mild stress. *Behav Brain Res.* 2004; 155:135–146. [PubMed: 15325787]
- Ramos A. Animal models of anxiety: do I need multiple tests? *Trends Pharmacol Sci.* 2008; 29:493–498. [PubMed: 18755516]
- Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, Dulawa S, Weisstaub N, Lee J, Duman R, Arancio O, Belzung C, Hen R. Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science.* 2003; 301:805–809. [PubMed: 12907793]
- Sibille E, Pavlides C, Benke D, Toth M. Genetic inactivation of the Serotonin(1A) receptor in mice results in downregulation of major GABA(A) receptor alpha subunits, reduction of GABA(A) receptor binding, and benzodiazepine-resistant anxiety. *J Neurosci.* 2000; 20:2758–2765. [PubMed: 10751426]
- Sibille E, Wang Y, Joeyen-Waldorf J, Gaiteri C, Surget A, Oh S, Belzung C, Tseng GC, Lewis DA. A molecular signature of depression in the amygdala. *Am J Psychiatry.* 2009; 166:1011–1024. [PubMed: 19605536]
- Surget A, Wang Y, Leman S, Ibarguen-Vargas Y, Edgar N, Griebel G, Belzung C, Sibille E. Corticolimbic transcriptome changes are state-dependent and region-specific in a rodent model of depression and of antidepressant reversal. *Neuropsychopharmacology.* 2009; 34:1363–1380. [PubMed: 18536703]
- Voikar V, Koks S, Vasar E, Rauvala H. Strain and gender differences in the behavior of mouse lines commonly used in transgenic studies. *Physiol Behav.* 2001; 72:271–281. [PubMed: 11240006]
- Wahlsten D, Bachmanov A, Finn DA, Crabbe JC. Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc Natl Acad Sci U S A.* 2006; 103:16364–16369. [PubMed: 17053075]
- Yalcin I, Belzung C, Surget A. Mouse strain differences in the unpredictable chronic mild stress: a four-antidepressant survey. *Behav Brain Res.* 2008; 193:140–143. [PubMed: 18565601]
- Young EA, Ribeiro SC. Sex differences in the ACTH response to 24H metyrapone in depression. *Brain Res.* 2006; 1126:148–155. [PubMed: 16824495]
- Young EA, Ribeiro SC, Ye W. Sex differences in ACTH pulsatility following metyrapone blockade in patients with major depression. *Psychoneuroendocrinology.* 2007; 32:503–507. [PubMed: 17462829]
- Zhao Y, Ma R, Shen J, Su H, Xing D, Du L. A mouse model of depression induced by repeated corticosterone injections. *Eur J Pharmacol.* 2008; 581:113–120. [PubMed: 18184609]

Figure 1. Integrated emotionality z-scores in mice exposed to unpredictable chronic mild stress (A) Raw data obtained from three independent behavioral tests performed in the same animal in both males and females mice (OF, EPM and NSF; n=14–15/group/sex) (B) Normalization of data using z-score method was performed for each parameter as described in the methods using the control male group as the baseline. (C) Test z-values were then calculated by averaging individual z-scores, and (D) averaged to obtain Emotionality z-score. (E) Controls and stress groups were split by sex to investigate sex differences to stress exposure. Data represent mean $\pm$ SEM (n=14–15/group). A–E: * p<0.05, ** p<0.01, *** p<0.001 for effects of UCMS exposure compared to the no-stress group. # describe statistical trends (p<0.1).

Figure 2. Chronic antidepressant treatment blocks stress-induced increase in emotionality z-scores (n=14–15/group)

(A) Raw data obtained from three independent behavioral tests (OF, EPM and NSF; n=14–15/group) performed in the same animals. (B) Normalization of data using z-score method was performed for each parameter. (C) Test z-values were then calculated by averaging individual z-scores, and (D) averaged to obtain the Emotionality Score. Data represent mean $\pm$ SEM (n=14–15/group). **, p<0.01 ***; p<0.001 for effects of UCMS exposure compared to the no-stress group. \$, p<0.05 \$\$, p<0.01 for effects of 4-weeks fluoxetine treatment compared to the stressed group.

Figure 3. Integrated Emotionality z-score in corticosterone-treated male and female mice
(A) Raw data obtained from three independent behavioral tests (OF, EPM and NSF; n=14–22/group) performed in the same animals. **(B)** Normalization of data using the z-scoring method was performed for each parameter as described in the methods. **(C)** Test z-values were obtained by averaging individual z-scores, and then combined to obtain emotionality z-scores. Data represent mean $\pm$ SEM (n=14–22/group). * $p<0.05$, ** $p<0.01$, *** $p<0.001$ for effects of corticosterone exposure compared to the sex-matched non-stress group. § $p<0.05$, §§§ $p<0.001$ indicate sex differences within groups

Figure 4. Emotionality and locomotion scores in two animal models of anxiety/depression

Use of z-score normalization allowed pooling of various experiments and multiple cohorts (n=22–51 mice/group), highlighting sex differences after chronic stress or corticosterone exposure. **(A)** Sex differences in emotionality responses to either stress or corticosterone exposure. Specifically, females were significantly more sensitive to stress and less sensitive to corticosterone exposure compared to males. **(B)** Applying similar normalization to locomotor parameters extracted from different behavioral tests (total crosses in OF and in EPM) revealed baseline sex differences, and in response to stress and corticosterone exposure. Data represent mean $\pm$ SEM (n=22–51/group). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ for effects of corticosterone or stress exposure compared to the sex-matched control group. §§ $p < 0.01$, §§§ $p < 0.001$ indicate sex differences within groups.

Figure 5. Dissecting sex differences in baseline emotionality

Combining emotionality z-scores in control animals across several experiments shows that the distribution of baseline emotionality scores is significantly skewed towards higher values in females, as more females show higher emotionality states compared to males.

Emotionality scores were separated in “low” (scores below -0.5), “normal”, (scores between -0.5 and $+0.5$) and “high” (scores greater than $+0.5$) ($n=34-42$ mice/sex extracted from 3 different cohorts). Relative proportions of animals in each group are indicated within bars. χ^2 analysis on group distributions revealed sex differences (§§§ $p<0.001$).

Table 1

P-values for 2-way ANOVA main effects corresponding to data shown in Fig 3.

Fig.#	Test	Parameter Measured	P value for Main effect of sex	P value for Main effect of cort	Coefficient of variation
3A	OF	Distance Ratio	0.006	0.0046	0.55
3A	OF	Time in the center	0.0496	0.019	0.59
3C	OF	Averaged Z values	0.0170	0.0082	0.61
3A	EPM	Time in Open Arms	0.672	0.0334	1.56
3A	EPM	Ratio of entries	0.5998	0.0019	0.98
3C	EPM	Averaged Z values	0.6291	0.0076	0.69
3A	NSF	Latency to eat	0.0012	0.0921	0.42
3C	NSF	Z value	0.0012	0.0921	0.42
3D	EMOTIONALITY Z-SCORE	AVERAGED Z-VALUES	0.0188	<0.0001	0.16

The significant main effect of sex provides an integrated means to report results with test-to-test variability. Combining normalized measures, emotionality z-scores augmented the overall statistical significance of the corticosterone main effect compared to each test separately.