Individual Behavioral Variability Across Time and Contexts in *Dendrobates tinctorius* Poison Frogs

Katharina M Soto[1,2], Faith O Hardin[1], Harmen P Alleyne[1], Eva K Fischer[1,3*]

[1] Department of Evolution, Ecology, and Behavior; School of Integrative Biology, University of Illinois Urbana-Champaign, Urbana IL 61801, USA

[2] Program of Evolution, Ecology, and Conservation, School of Integrative Biology; University of Illinois Urbana-Champaign, Urbana IL 61801, USA

[3] Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana IL 61801, USA

*Corresponding author:

Eva K Fischer

Department of Evolution, Ecology, and Behavior

School of Integrative Biology

University of Illinois Urbana-Champaign

505 S Goodwin Ave

Urbana, IL 61801

efischer@illinois.edu

Abstract

Consistent individual differences in behavior ("animal personality") have consequences for individual fitness, evolutionary trajectories, and species' persistence. Such differences have been documented across a wide range of animals, though amphibians are generally underrepresented in this research area. The aim of our study was to examine consistent individual differences in Dyeing poison frogs, *Dendrobates tinctorius*. We evaluated repeatability in behaviors including activity, exploration, and boldness to assess consistency of behaviors across different temporal, experimental, and environmental contexts. We found repeatability in activity and exploration across time and contexts. In contrast, we observed context-specific behavior for our metrics of boldness, with consistent individual differences only for some measures. Further, while activity and exploration displayed consistent correlations across contexts, relationships between activity and boldness were context dependent. Our findings document the presence of consistent individual differences in behavior in *D. tinctorius* poison frogs, and also reveal context-dependent behavioral differences, highlighting the complex relationship between consistent individual differences and context-specific responses in animal behavior.

Statement of Significance

The concept of animal personality centers on the existence of consistent individual differences in behavior. However, behavioral responses can depend on context, and consistent individual differences in one context do not guarantee consistent differences in another. To address this question, we assessed activity, exploration, and boldness in captive-bred poison frogs (*Dendrobates tinctorius*) across time and environmental contexts. Our comprehensive approach revealed consistent individual differences in some behaviors and context-specificity in others. While activity and exploration were generally repeatable and correlated with one another,

boldness was not. Especially in view of the emphasis on measures of boldness in the animal personality literature, our findings emphasize the importance of reiterative and holistic approaches in the study of animal behavior.

Keywords: behavioral syndromes, animal personality, repeatability, individual variation, Dendrobatid

Introduction

There is growing evidence that consistent individual differences in behavior exist across taxa, and that these differences can have important consequences for individual fitness, adaptation, and species' persistence (Bell et al. 2009). For example, bold and aggressive individuals may be superior at territory defense but inferior at parental care (Duckworth 2006). Such trade-offs can lead to the establishment of alternative behavioral strategies within populations and potentially shape evolutionary trajectories (Smith and Blumstein 2008; Dochtermann and Dingemanse 2013). Moreover, certain personality types may fare better in the face of increasing human-induced global change, for example if they are better able to adapt to increasingly urban habitats (Pamela et al. 2015; Murgui and Hedblom 2017) or more likely to invade novel environments (Brodin et al. 2013; Carvalho et al. 2013). Increased recognition of the complex consequences of consistent individual behavioral differences have led to an increase in the study of animal personalities across taxa. Animal personality is increasingly invoked to explain behavioral differences in natural and captive populations, and to speculate on how these differences may influence survival and adaptation. However, many studies do so without confirming consistent individual differences across multiple contexts and over time. Assuming – rather than testing – that individual behavioral variation is stable across time and contexts can introduce bias, leading to adaptive inferences based on isolated incidents or context-specific behavior (Bell et al. 2009). Moreover, design and interpretation of behavioral assays that facilitate cross-species behavioral comparisons remain a challenge (Uher 2008; O'Connell and Hofmann 2011; Odom et al. 2021).

Animal personality is broadly defined as consistent individual differences in behavior (Réale et al. 2010; Carere and Maestripieri 2013). These differences are typically characterized by quantifying behavioral repeatability, defined as the proportion of variance explained by within-individual variation compared to between-individual variation (Bell et al. 2009; Bergmüller 2010; Garamszegi and Herczeg 2012). Animal personality encompasses an array of behavioral

traits, often with a particular focus on boldness:shyness, exploration:avoidance, and activity (Réale et al. 2007; Garamszegi and Herczeg 2012; Michelangeli et al. 2019). Further, these traits can become "packaged together" into behavioral syndromes (Sih et al. 2004). For instance, studies spanning diverse taxa such as common voles (*Microtus arvali*) (Herde and Eccard 2013), convict cichlids (*Amatitlania nigrofasciata*) (Mazué et al. 2015), sea turtles (*Chelonia mydas)* (Kudo et al. 2021), and great tits (*Parus major*) (Cole and Quinn 2011) document positive correlations between boldness, activity, and exploration.

While the presence of animal personalities and behavioral syndromes have been documented in animals from primates to beetles, there is bias in taxonomic representation. While there is extensive research in non-human animals including birds (e.g., Réale et al. 2007; Carere and Locurto 2011) and fish (e.g., Bell 2005; Smith and Blumstein 2010; Mazué et al. 2015), amphibians have received comparatively less attention. Indeed, recent cross-taxon reviews mention one or no studies on amphibians (Moiron et al. 2019; Cabrera et al. 2021). These biases likely arise in part from historical assumptions that amphibians were “largely instinctive bound machines controlled by external stimuli” (Maier and Schneirla 1964), yet it is increasingly clear that these creatures make complex, context dependent behavioral decisions (Burghardt 2013; Kelleher et al. 2018). Studying individual behavioral variation in amphibians is critical for advancing our understanding of how these animals navigate risks and opportunities, such as the trade-off between mating advertisement and predator avoidance (Stamps 2007; Sih and Del Giudice 2012). Importantly, this knowledge has broader implications for understanding evolutionary trajectories, ecological interactions, and conservation needs, which are of particular importance given the alarming global decline of amphibian populations (Kelleher et al. 2018).

Among anurans (frogs and toads), poison frogs (family Dendrobatidae) offer compelling study subjects for exploring individual variation and correlations in behavior. Poison frogs are known for their bright coloration, toxicity, and behavior. These species’ vivid hues are an honest signal of their toxicity and have coevolved with a diurnal lifestyle (Summers 2003; Maan and

Cummings 2012). Poison frogs exhibit an array of interesting behaviors, including anti-predator defenses (Breed 2008), aggressive territoriality (Crothers and Cummings 2015), parental care (Roland and O'Connell 2015; Fischer et al. 2019), spatial learning and navigation (Liu et al. 2016; Pašukonis et al. 2022), and behavioral plasticity (Ringler et al. 2017; Peignier et al. 2023). Taken together, these characteristics have garnered poison frogs increasing attention in the realm of animal personality and behavioral syndromes. Existing studies in poison frogs have examined consistent individual differences in activity (Sonnleitner et al. 2022), antipredator behavior (Peignier et al. 2023), and aggression (Chaloupka et al. 2022). Nonetheless, a knowledge gap remains regarding the extent of consistent individual differences across time and context in poison frogs specifically, and amphibians more generally.

Dyeing poison frogs, *Dendrobates tinctorius*, have long been popular in zoos, aquaria, and the pet trade, and have recently received increasing research attention in the wild (Rojas and Pašukonis 2019; Fouilloux et al. 2021; Pašukonis et al. 2022) and the lab (Fischer et al. 2019, Fischer and O'Connell 2020; Sonnleitner et al. 2022). *D. tinctorius* boast a variety of bright color morphs without sexual dimorphism in coloration. In the wild, males and females maintain home ranges of a similar size, have similar daily movement patterns, and exhibit aggression toward same- and opposite-sex individuals (Rojas and Endler 2013; Pašukonis et al. 2022). Unusually among frogs, males of the species do not produce conspicuous advertisement calls, though they do produce low amplitude courtship vocalization (Rojas and Pašukonis 2019). Like other Dendrobatids, *D. tinctorius* provide extended parental care that includes defense, hydration, and cleaning of terrestrially deposited embryos, and transportation of tadpoles piggyback to pools of water upon hatching (Stynoski et al. 2015; Rojas and Pašukonis 2019). In *D. tinctorius*, tadpole transport is typically performed by males, however, females will occasionally take over (Rojas and Pašukonis 2019; Fischer and O'Connell 2020). Alongside their fascinating traits, existing work on *D. tinctorius* in both the field and the lab, and the ability to ethically acquire animals make them an excellent study system.

We investigated personality and behavioral syndromes across contexts and timeframes in *D. tinctorius*. Through a set of three experiments, we repeatedly assessed activity, exploration, and boldness – in the same and in different assays – to test whether commonly used personality metrics are robust to differences in assay design and timing (Fig. 1). Following the methodologies of Kelleher et al. (2018), our experiments included an open field test (OFT), a threat stimulus test (TST), a novel environment test (NET), and an activity in a familiar environment test (AFET) (Fig. 1). All experiments were built upon the open field test, a widely utilized method for assessing activity levels, exploratory behavior, and boldness across taxa (Bell et al. 2009; Carere and Locurto 2011). Importantly, individuals were assayed multiple times in each experiment (i.e., experienced multiple 'rounds' of each 'trial type' - terminology we use throughout the manuscript). This design allowed us to assess repeatability of commonly used behavioral metrics within and across contexts, and in response to multiple stimuli.

We predicted consistent individual differences in activity, exploration, and boldness across multiple rounds and the three distinct experiments. Further, we anticipated a behavioral syndrome, in which frogs with higher activity levels would also display increased boldness and exploration. Finally, we expected frogs to perceive the threat stimulus and the novel environment as stressful and exhibit context-dependent behavioral responses. Our overarching goal was to provide a comprehensive, nuanced understanding of how consistent individual differences in behavior manifest across time and context, and to reflect on the implications of these patterns for extrapolation to ecological and evolutionary outcomes.

Methods

Experimental overview

In Experiment 1, we conducted an open field test (OFT) in an artificial environment to establish baseline activity, exploration, boldness, and the relationships among them (Fig. 1).

156 This test reflects the standard, basic design used for open field tests. In Experiment 2, we
157 transitioned to an open field test (OFT) in a naturalistic environment, additionally assessing
158 behavior before, during, and after exposure to a threat stimulus test (TST) (Fig. 1). In
159 Experiment 3, we maintained the naturalistic environment, but extended the duration of each
160 trial including the baseline open field test (OFT) and an activity in a familiar environment test
161 (AFET), and additionally performed an extended duration novel environment test (NET) in an
162 artificial environment (Fig. 1).

163 *D. tinctorius* were housed in our animal facility at the University of Illinois Urbana-
164 Champaign. Terraria contained soil, moss, leaf litter, live plants, coconut shelters, and a small
165 pond. For housing and all experiments, we maintained terraria and test arenas at 21°–23° C
166 and 80%–90% humidity on a 12L:12D light cycle. We fed frogs *Drosophila hydei* or *Drosophila*
167 *melanogaster* dusted with vitamins three times weekly.

168 All behavioral experiments were conducted between March and April 2022. We used a
169 total of 48 captive bred frogs (N= 20 adult males, N= 12 adult females, and N= 17 juveniles of
170 unknown sex). All frogs were assayed in Experiments 1 and 2. In Experiment 3, we used a
171 subset of 19 frogs (N= 6 adult males, N= 8 adult females, and N= 5 juveniles; see additional
172 details below). For Experiments 1 and 2, we ran frogs sequentially in groups of five per week,
173 excluding the last group which had only four frogs. We followed the same protocol in a second
174 round of trials with the same groups of frogs. In essence, each individual underwent two
175 identical rounds, with a 5-week interval in between. For Experiment 3, we ran frogs in groups of
176 nine per week through a series of sequential trials that lasted a combined 25 hours (details
177 below). These trials were repeated three times with a 3-week interval between each round.
178 Group size for all experiments was chosen to make all behavioral testing feasible within a given
179 daily time window and to balance individuals of different color morphs across groups. Group
180 sizes and time intervals in Experiment 3 were modified to allow for the longer testing time
181 course while still assaying the same frogs in all experiments within the same timeframe.

182 Although frogs varied in age, all individuals were sexually inexperienced. Previous studies have
183 demonstrated that fed or digesting amphibians exhibit decreased levels of activity compared to
184 fasted individuals (Andrade et al. 2005). To ensure that the frogs were not actively digesting
185 during the experiments, we fasted frogs for 24 hours before behavioral trials.
186
187 *Experiment 1: Artificial open field test*
188 In Experiment 1, we investigated exploration, activity, and boldness in an open field test
189 (OFT) in a novel, artificial environment (Fig. 1). The open field arena was a circular enclosure
190 with a diameter of 108 cm. To maintain humidity and moisture, we lined the arena with white
191 bench paper soaked in reversed-osmosis (RO) water. For each observation, we placed a single
192 frog at the center of the arena as a standardized starting point and recorded behavior for 15
193 minutes using an overhead camera (DVC, HDV-604S). Individuals were immediately returned to
194 their home terraria following trials. The 15-minute testing time was chosen based on previous
195 tests in amphibians and other species. We quantified exploration as the proportion of areas a
196 frog visited, activity as the number of times a frog crossed the line between areas, and boldness
197 as time spent in the center of area (Fig. 2) (additional details below).
198
199 *Experiment 2: Naturalistic open field and threat stimulus test*
200 In Experiment 2, we quantified activity and exploration in a naturalistic open field. The
201 behavioral arena was a 20 x 30 x 20 cm (L x W x H) plexiglass tank that mimicked the frogs'
202 home terraria, including soil substrate, moss, leaves, and a live plant positioned in one corner.
203 The arena was completely covered from the outside with white contact paper, excluding a 24 x
204 17 cm (W x H) exposed area opposite the plant for the presentation of the threat stimulus. As in
205 Experiment 1, we placed a single frog at the center of the arena at the beginning of each
206 observation and used an overhead camera to record behavior. We conducted three trials to
207 investigate behavior (1) pre-stimulus open field test (OFT) (15 minutes after introduction to

208 arena), (2) during threat stimulus test (TST) (3 minutes of looming stimulus), and (3) post-
209 stimulus open field test (for 15 minutes after the stimulus) (Fig.1). The 15-minute testing time
210 was chosen based on previous studies and to align with Experiment 1. The duration of the
211 looming stimulus was determined through pilot testing. Test trials on frogs not included in the
212 study demonstrated that frogs typically showed a reaction within 3 minutes or froze throughout
213 the entire stimulus presentation.

214 We assessed activity and exploration using the same methods as in Experiment 1 (Fig.
215 2). Additionally, we measured boldness by observing the frogs' responses to a looming stimulus
216 (Fig. 2). Previous studies using leopard frogs (*Lithobates pipiens*) have shown reliable
217 avoidance behavior in response to looming objects (Ingle 1990; Waldeck and Gruberg 1995).
218 To present the looming stimulus, we used an iPad (Apple, 6th generation). The iPad was
219 covered in white tape, except for the display screen, and placed in the exposed area of the
220 arena (as shown Fig. 1). The looming stimulus was an expanding black circle projected
221 repeatedly onto the iPad screen for a duration of three minutes. To prevent habituation and
222 predictability, random pauses ranging from 1 to 30 seconds were incorporated into the stimulus
223 presentation. During the pre- and post- exposure trials, the iPad displayed a white screen to
224 match the surrounding environment. During the looming stimulus, we recorded activity and
225 exploration, as well as boldness based on whether the frogs approached or displayed
226 avoidance behavior in response to the stimulus.

227

228 *Experiment 3: Naturalistic open field, familiar environment, and novel environment tests*

229 In Experiment 3, we increased the duration of each trial to capture more prolonged
230 patterns in the frogs' behavior. To conduct these longer trials concurrently with Experiments 1
231 and 2 while also maintaining our repeated measures design, we selected a subset of individuals
232 (N= 19) from the previous experiments and limited our observations to two behaviors (activity
233 and boldness). The behavioral arena was constructed using the same materials as Experiment

2. However, the dimensions of the arena were twice as long at 30 x 60 x 20 cm (L x W x H) and the arena split into two halves. One side of the arena mimicked the frogs' home terraria as in Experiment 2. The other side mimicked the novel, artificial environment of Experiment 1, with only white, moistened paper flooring. The two sides of the arena were separated by a white cardboard barrier, which could be removed without disturbing the frog or the soil structure of the natural chamber.

We conducted three trials to investigate the frogs' activity and boldness: (1) an open field test (OFT) in a novel, naturalistic environment (the first 90 minutes after introduction to the area), (2) activity in a familiar, natural environment test (AFET) (60 minutes after a 24 hour acclimation period in the arena), and (3) exposure to a novel, artificial environment test (NET) (60 minutes after the removal of the barrier between the familiar and novel environment) (Fig. 1). Time points were chosen to characterize differences based on longer observations (60 minutes = 4x the open field trial time in Experiments 1 and 2) and extended habituation to the test arena (and immediate 90-minute trial and a second trial after 24 hours). Trials were conducted sequentially, with each frog remaining in the arena for 25 hours while the trials progressed. We conducted two subsequent rounds of trials for each individual, with three weeks between each round. Due to the extended duration of each trial, we focused our analyses on activity and boldness. This decision was made to streamline data collection and because we expected – and found – a positive correlation between activity and exploration. We quantified activity using the same methods as in Experiment 1 and 2, number of line crosses. To assess boldness, we measured the frogs' latency to cross into the novel environment when the barrier was lifted (Fig. 2).

Data analysis

We describe general methods here with modifications for each experiment below. We coded videos using Behavioral Observation Research Interactive Software (BORIS) ver. 8.13

260 (Friard and Gamba 2016). Experiment type was not blind because assay identity was obvious
261 from videos, but all analyses were blind to individual identity. Behavioral data was analyzed with
262 R v.4.2.4 (R Core Team 2022) in Rstudio v.4.2.2. (RStudio Team 2021). For all response
263 variables that were normally distributed, we used the raw data for analysis. Response variables
264 that did not meet expectations of normality were transformed as described below. For all
265 analyses, outcomes were considered significant at $p <= 0.05$ and all posthoc comparisons were
266 Tukey adjusted for multiple hypothesis testing.

267 We used the lmer function in the lmerTest package (Kuznetsova et al. 2017) for all linear
268 models, and always included frog ID as a random effect and round (i.e., repeated
269 measurements in the same assay) as a fixed effect. Additional fixed effects varied based on
270 experiment (see below). We tested for the effects of sex on behavior and found no significant
271 results. Therefore, and because not all individuals could be sexed at the time of the experiment,
272 we did not include sex as a factor in the analyses described below. We ran Type III ANOVAs on
273 linear models to generate test statistics and p-values for main effects and interactions.
274 Subsequently, we used the emmeans package (Lenth 2021) to run pairwise post-hoc
275 comparisons where appropriate.

276 We performed repeatability analyses using the rptR package (Stoffel et al. 2017), with
277 95% confidence intervals (CI) based on 1000 bootstrapping runs. Frog ID was used as a
278 random effect for repeatability (R) among individuals. We used Pearson correlation for normally
279 distributed data and Kendall correlation for non-normally distributed data to assess relationships
280 between behaviors. The correlation coefficients were interpreted using these criteria: ≈0.1 a
281 small effect, ≈0.3 a moderate effect, and >0.5 a strong effect (Garamszegi et al. 2013).

282

283 *Analysis Experiment 1: Artificial open field test*

284 We coded videos in BORIS using a transparent overlay that divided the behavioral arena
285 into 40 areas: 24 on the ground and 16 on the wall. From videos, we quantified exploration as

286 the proportion of areas a frog visited and activity as the number of times a frog crossed a line
287 between areas. Movement between areas was recorded when the majority of the frogs' body
288 entered into a new area. We measured boldness as time spent in the center four areas of the
289 arena (Fig. 2). We log transformed the activity and boldness data to improve normality. We ran
290 linear models for each behavior (activity, exploration, boldness) as described above, including
291 round as a fixed effect and frog ID as a random effect. We performed repeatability analysis with
292 Gaussian distributions for exploration and Poisson distributions for activity and boldness (Stoffel
293 et al. 2017). Data from 11 individuals was lost due to equipment failure, leaving a final sample
294 size of N=38 individuals. Analyses and results are based on these samples.
295
296 *Analysis Experiment 2: Naturalistic open field threat stimulus test*
297 We coded videos in BORIS using a transparent overlay that divided the behavioral arena
298 into 32 areas: 16 on the ground and 16 on the wall. We quantified and analyzed activity and
299 exploration as in Experiment 1, with the addition of a fixed effect of trial (pre-stimulus, during
300 stimulus, and post-stimulus) to test for pre- and post- stimulus behavior differences. To compare
301 trials of different duration, we divided each behavior by the total trial time.
302 For response to stimulus, which we used as a measure of boldness, we divided the
303 arena into three sections: (1) at stimulus (the two wall areas overlaying the iPad, the two floor
304 areas directly in front the iPad, and one wall area immediately to the left of the iPad), (2) near
305 stimulus (the two wall areas and the four floor areas surrounding the "at stimulus" area), and (3)
306 away from stimulus (the remaining areas) (Fig. 2). We noted the frogs' position before and after
307 stimulus presentation. We used a Chi-Squared test in the Tidyverse package (Wickham et al.
308 2019) for pairwise comparisons of responses (moved away, moved toward, or froze) (R Core
309 Team 2022). We fit rptR models with Gaussian distributions to test for repeatability of each
310 behavior across rounds and across trials (Stoffel et al. 2017).
311

312 *Analysis Experiment 3: Naturalistic open field novel environment test*

313 We coded videos in BORIS using a transparent overlay that divided the behavioral arena
314 into 96 areas: 16 areas on both the natural and novel arena floor, and 16 areas on the walls of
315 each chamber. We recorded activity as the number of times a frog crossed the line between
316 areas as in Experiments 1 and 2. After the divider was removed, we recorded the time it took for
317 a frog to move into the novel area as latency to explore, which we interpreted as a measure of
318 boldness (Fig. 2). We recorded activity on both sides of the chamber after the barrier was
319 removed. We ran linear models for activity and latency to explore, including round (i.e., repeated
320 measurements in the assay), trial (baseline, after acclimation, in a novel environment), and trial:
321 round interaction as fixed effects, and frog ID as a random effect. As in Experiment 1, behaviors
322 were divided by trial duration when analyzing trial effects.

323 To determine whether activity predicted an individual's latency to cross into a new area,
324 we filtered the data to only include observations taken during the novel environment test. We
325 then fit a similar linear mixed effect model to predict latency to enter the novel area with fixed
326 effects of activity, round, and their interaction, and frog ID as a random effect. We fit rptR
327 models with Gaussian distributions to test for repeatability of each behavior across rounds and
328 across trials (Stoffel et al. 2017).

329

330 *Cross-experiment analysis of activity and exploration*

331 In addition to behavioral analyses within experiments, we were interested in the behavior
332 of individuals across experiments. To compare exploration between Experiments 1 and 2, which
333 had different arena sizes, we calculated the proportion of areas explored. We did not quantify
334 exploration in Experiment 3, and it was therefore excluded from this analysis (Fig. 2). We
335 compared the ratio of areas explored in Experiment 1 to both the pre-stimulus and post-stimulus
336 exploration in Experiment 2 using linear mixed models, with exploration rate and experiment as
337 fixed effects (Bates et al. 2015). We tested repeatability using a Gaussian distribution (Stoffel et

338 al. 2017).

339 In all experiments, we measured activity by counting the number of line crosses (Fig. 2).
340 To compare across experiments of different duration without introducing biases based on trial
341 time, we divided the number of lines crossed by the total trial time to calculate a line crossing
342 rate. We compared the subset of individuals (n= 19) in Experiment 3 with the same individuals
343 from Experiments 1 and 2. Baseline activity measurements were similar across all experiments,
344 taken prior to either the presentation of a stressor (threat stimulus test) or an environmental
345 stressor (novel environment test) (Fig. 1). We also compared the post-stimulus activity in
346 Experiment 2 to the activity observed during the novel environment test in Experiment 3,
347 reasoning that these trials were similar in representing responses following an environmental
348 change. We analyzed behavioral differences using activity rate as the response variable,
349 experiment as a fixed effect, and frog ID as a random effect (Bates et al. 2015). We calculated
350 post hoc comparisons and repeatability as before.

351

352

353 **Results**

354 *Experiment 1: Artificial open field test*

355 Frogs displayed repeatability in exploration (R= 0.278, p= 0.029) (Fig. 3a) and activity
356 (R= 0.279, p= 0.032) (Fig. 3b) and across both rounds of an open field test separated by five
357 weeks. However, time in the center of the arena, which we interpreted as a measure of
358 boldness, was not repeatable (R= 0.044, p= 0.395). We found no significant effect of round for
359 activity ($F_{1,37}$= 3.390, p= 0.074), exploration ($F_{1,37}$= 1.392, p= 0.245), or boldness ($F_{1,74}$= 0.628,
360 p= 0.431), suggesting that frogs' behavior did not change in response to previous experience in
361 the arena. Activity and exploration exhibited a strong positive correlation (corr= 0.767, p<
362 0.001). Boldness was not correlated with either activity (corr= -0.150, p= 0.196) or exploration
363 (corr= -0.119, p= 0.307) (Table 1).

Experiment 2: Naturalistic open field and threat stimulus test

Activity and exploration were repeatable across pre- and post- stimulus open field tests in both round 1 (activity: R= 0.593, p< 0.001; exploration: R= 0.421, p< 0.001) and round 2 (activity: R= 0.529, p< 0.001; exploration: R= 0.449, p< 0.001). We observed the same trend across rounds within trial (activity: R= 0.123, p= 0.006; exploration: R= 0.177, p< 0.001), aligning with the results from Experiment 1 (Table 1). In contrast, during the threat stimulus test, there was no repeatability of activity (R= 0.201, p= 0.069) or exploration (R= 0.175, p= 0.145) across rounds.

We found an effect of trial on exploration ($F_{2,230}$= 112.143, p< 0.001) (Fig. 4a) and activity ($F_{1,234}$= 170.128, p< 0.001) (Fig. 4b), with frogs decreasing behaviors during the threat stimulus test and then increasing post stimulus (Fig. 4). We observed no significant round effect for activity ($F_{1,232}$= 0.570, p= 0.451) or exploration ($F_{1,234}$= 0.292, p= 0.590), nor an interaction between round and trial ($F_{2,228}$= 0.506, p= 0.604; $F_{2,228}$= 0.216, p= 0.801), indicating that frogs behaved consistently across multiple experiences in the arena. Activity and exploration were highly correlated across all trials (pre-stimulus, during stimulus, and post-stimulus) (corr= 0.7–0.8, p< 0.001).

We evaluated boldness by quantifying frogs' responses to the threat (looming) stimulus. Frogs' responses to the looming stimulus were not repeatable across rounds (R= 0, p= 1). This could be in part because overall among individual variation was low, as the majority of frogs demonstrated a freezing response during stimulus presentation rather than moving away ($\chi^2_{1,96}$= 19.755, p< 0.001) or towards ($\chi^2_{1,96}$= 41.796, p< 0.001) the stimulus. No significant effect of round was observed for boldness ($\chi^2_{2,96}$= 22.429, p= 0.401).

We found significant correlations between boldness (response to stimulus) and frogs' activity (tau= 0.155, p= 0.002) and exploration (tau= 0.155, p= 0.002) during the pre- and post-stimulus trials. These trends suggest that frogs displaying higher levels of activity and

390 exploration prior to and after the stimulus were also more inclined to approach the stimulus
391 during the threat stimulus trial. These correlations between boldness, activity, and exploration
392 contrast with the results obtained in Experiment 1 (Table 1).
393
394 *Experiment 3: Naturalistic open field and novel environment test*
395 Frogs were repeatable for activity both across trials (R= 0.321, p<0.011) and across
396 rounds (R= 0.651, p= 0.001). We found that frogs moved more with repeated experience in the
397 arena across all trials ($F_{2,143}$= 3.919, p= 0.022), with round 1 having significantly less movement
398 than round 2 (t_{36}=-3.423, p=0.004) and Round 3 (t_{36}= -3.784, p= 0.002), but no significant
399 difference between rounds 2 and 3 (t_{36}= -0.360 p=0.931) (Fig. 5a). Similarly, we found that frogs
400 took less time to cross the boundary into the novel area during the novel environment test as
401 rounds progressed ($F_{2,36}$= 13.100, p< 0.001), with frogs taking longer to cross in round 1
402 compared to round 2 (t_{36}= 2.837, p= 0.019) and round 3 (t_{36}= 5.108, p< 0.001), but not in round
403 2 compared to round 3 (t_{36}= 2.270, p= 0.073) (Fig 5b). We found significant repeatability for
404 latency to explore across rounds (R= 0.388, p= 0.003) suggesting all frogs responded to
405 repeated experience similarly.
406 We use latency to cross into a novel environment as an indicator of boldness,
407 interpreting shorter latencies as bolder. We found a significant negative correlation between
408 activity and latency to cross into a novel environment during the novel environment test (corr= -
409 0.665, p< 0.001), indicative of a positive relationship between activity and this measure of
410 boldness (Fig 5c). Additionally, we found that boldness was influenced by both activity ($F_{1,38}$=
411 28.795, p<0.001) and round ($F_{2,38}$= 6.546, p= 0.004), indicating that more active frogs moved
412 into the novel area more quickly, and all frogs moved into the novel area more quickly with
413 repeated experience.
414
415 *Cross-experiment analysis of activity and exploration*

Frogs exhibited consistent individual differences in exploration (R= 0.171, p= 0.005) across Experiments 1 and 2. For activity, baseline activity during the open field tests were repeatable between the naturalistic environments of Experiment 2 and Experiment 3 (R= 0.207, p< 0.001), and between the artificial environment of Experiment 1 and the naturalistic environment of Experiment 3 (R= 0.171, p= 0.041). Activity was not repeatable between the artificial environment of Experiment 1 and the naturalistic environment of Experiment 2 (R= 0.029, p= 0.401). Activity was repeatable across stressor trials (Fig. 1) (i.e., threat stimulus test in Experiment 2 and novel environment test in Experiment 3) (R= 0.243, p= 0.003).

Discussion

We investigated animal personality and behavioral syndromes in captive-bred poison frogs (*Dendrobates tinctorius*) by characterizing behavioral differences across multiple environmental contexts and timepoints. We conducted three experiments to assess activity, exploration, and boldness in the same individuals and found repeatability in activity and exploration across time and contexts. In contrast, we observed context-specific behavior for our metrics of boldness, with consistent individual differences only for some measures. The stability of individual differences in behavior can have implications for how species respond to shifting environmental conditions (Wong and Candolin 2015; Merrick and Koprowski 2017), thereby shaping evolutionary trajectories and ecological interactions (Shaw 2020; Stiegler et al. 2022). Despite their declining populations, amphibians remain relatively understudied in behavioral research. The results of our work help to fill this gap, address broader issues in the field of "animal personality", and set the stage for future work in *D. tinctorius*.

Our study revealed widespread repeatability for activity and exploration, evident across time, trials, and experimental contexts. These findings suggest that measures of activity and exploration are indeed reliable, and measurements in one assay or at one timepoint are good

442 predictors of an individual's behavior in another context or at another time. We also recovered
443 the commonly described activity:exploration behavioral syndrome (Réale et al. 2007). While we
444 expected a positive relationship between activity and exploration (i.e., exploration is impossible
445 without movement), we note that there is room for variability in this relationship. For instance,
446 frogs might have displayed a lot of movement but only within a restricted area, leading to
447 heightened activity but limited exploration. The strong, significant relationships we found
448 (correlations > 0.76), suggest that these behaviors are indeed tightly linked and can therefore
449 reliably predict one another as well as some other behaviors (e.g., latency to explore a novel
450 area, see below).

451 The above outcomes align with prior research on animal personality across taxa (Réale
452 et al. 2007; Kelleher et al. 2018), and complement a recent investigation on another
453 Dendrobatid, *Allobates femoralis*, which demonstrated the efficacy of both artificial and
454 naturalistic setups in capturing consistent individual differences in behavior (Bégué et al. 2023).
455 However, a study on *D. tinctorius* froglets by Sonnleitner et al. (2022) yielded somewhat
456 different findings: the authors found significant repeatability at the family level but no significant
457 repeatability at the individual level. These results suggests that genetic factors influence
458 behavior, and this effect outweighs individual differences when related individuals are reared
459 under similar conditions. We were unable to control for relatedness in our study, but assume
460 genetic differences explain at least some of the behavioral variation we document. A non-
461 mutually exclusive explanation is that individuals in the study of Sonnleitner et al. (2022) were
462 much younger than those used here, and individual differences in behavior may become more
463 pronounced with age and experience (Bell et al. 2009). We note that our findings here and
464 those of Sonnleitner et al. (2022) are from captive-bred individuals. While *D. tinctorius* readily
465 exhibit behaviors of interest (e.g., parental care and aggression) under captive conditions,
466 influences of lab adaptation and genetic differences between wild and captive animals surely
467 also influence behavior. In any case, while consistent individual differences are being

468 increasingly documented across diverse taxa, contrasting results even in the same species
469 underscore the importance of considering various contexts and timepoints when extrapolating
470 from single studies to evolutionary and ecological outcomes.

471 While we observed a high degree of repeatability in activity and exploration across time
472 and contexts, the behaviors we considered as measures of boldness demonstrated contextual
473 dependence. Our first measure was the commonly used metric of time spent in the center of a
474 novel arena (Carter et al. 2012; Yuen et al. 2017; Baker et al. 2018; Rudolfová et al. 2022). We
475 did not observe consistent individual differences in this measure, nor was it correlated with other
476 behaviors. As our next measure of boldness, we assayed response to a threat stimulus in a
477 familiar environment, aligning with prior studies (Ingle 1990; Waldeck and Gruberg 1995; Carter
478 et al. 2012; Yuen et al. 2017). Once again, we did not find repeatability for boldness, although
479 we did observe a correlation between this measure of boldness, activity, and exploration. In our
480 final approach, we measured boldness as the latency to enter a novel environment, a less
481 common yet established method for evaluating boldness (White et al. 2013; Magnhagen et al.
482 2014; Theódórsson and Ólafsdóttir 2022). With this approach, we observed both repeatability
483 and a correlation between boldness and activity. Taken together, our results suggest that
484 commonly used boldness methodologies may not measure the same personality trait.

485 Despite the popularity of measuring boldness in animal personality studies, our findings
486 align with those of others who have explicitly compared outcomes across different metrics
487 categorized as 'boldness' and found conflicting results. For example, our approach and
488 outcomes are similar to that of Yuen et al. (2017) in their study on African striped mice
489 (*Rhabdomys pumilio*). The authors investigated whether distinct experimental methods were
490 indeed measuring boldness. Like our approach, they employed both artificial and natural
491 environments, and conducted two boldness tests, an open field, and a startle test. Yuen et al.
492 (2017) highlighted the distinction between these methods: the startle test evaluated risk-taking
493 behavior after threat exposure, potentially measuring anxiety rather than boldness, while the

open field test did not incorporate a preceding threat and proved to be a more reliable method for gauging boldness in their species. Carter et al. (2012) found related results when examining chacma baboons (*Papio ursinus*), concluding that threatening situations may measure anxiety, not boldness. Moreover, the use of artificial open field tests, particularly in evaluation of time spent in the center, may not consistently capture boldness. Indeed, a recent study on Zebrafish (*Danio rerio*) suggests that thigmotaxis may not effectively measure boldness or anxiety-like behavior in this species (Rajput et al. 2022).

Translating these observations to our own results, our open field test focused on evaluating time in center, the threat stimulus test observed behavior in the presence of a threat, whereas the novel environment test examined behavior in a new and unfamiliar environment. While we initially considered all tests as measures of boldness, our data suggest that the specific nature of the challenges presented in each experiment and the environmental context led to varying types and intensities of behavioral responses, and in turn distinct patterns of repeatability and behavioral correlations. Our findings also confirm the importance of considering the environment when measuring behavioral correlations: we observed correlations between boldness, activity, and exploration (i.e., behavioral syndromes) in naturalistic environments (Experiments 2 and 3) but not in artificial settings (Experiment 1) (Table 1). Alongside those of others, our findings highlight that the environment and assay in which individuals are assessed can exert a profound influence on their behavior (e.g., Niemelä and Dingemanse 2014; Bégué et al. 2023) and the necessity of evaluating individual variation across diverse contexts (Bell et al. 2009).

Taken together, these observations emphasize the challenges of adapting standard methodologies across closely and distantly related taxa. For example, poison frogs are diurnal unlike most frogs, and amphibians are generally less active than mammals. While behavioral and life-history differences between taxa are obvious and themselves of interest, they provide fundamental challenges for the design, implementation, and interpretation of comparative

studies (Uher 2008; O'Connell and Hofmann 2011; Odom et al. 2021). When findings align across taxa, we assume shared core principles; however, when patterns differ it is more difficult to understand whether organisms differ fundamentally in, for example, relationships between activity and boldness, or whether we have not "asked them the right question" with our chosen assay. Fortunately, with a growing literature that includes more species and taxa, these challenges are also opportunities to build our understanding through comparative data and thinking.

Conclusions

Our findings demonstrate consistent individual differences in behavior across time and context in *D. tinctorius*, while also highlighting the subtle complexities of studying animal personality. While we found strong evidence for individual consistency in activity and exploration, our measures of boldness were variable and context dependent. Our results demonstrate the value of assessing multiple behavioral traits and the relationships among them to build a comprehensive view of behavioral repeatability and variability within and across species (Sih et al. 2004; Bell et al. 2009; Sih and Del Giudice 2012). Broadly, our results add to the growing literature emphasizing the importance of evaluating context-dependent patterns of individual behavioral variation, even for assays commonly considered as measuring the same trait (Bell et al. 2009; Carter et al. 2012; Yuen et al. 2017). More specifically, our findings contribute to the growing recognition of behavioral complexity in anuran amphibians, with implications for amphibian research, including their role as ecological indicators amid habitat degradation and climate change (Grant et al. 2019). Our findings here pave the way for future research on the impact of consistent individual differences on reproductive success, survival, and community dynamics.

Acknowledgments

We thank the members of the Fischer Lab for feedback on previous versions of the manuscript and help with frog husbandry. For their insight on statistical analyses, we thank Drs. Jen Moss and Sarah Westrick. We thank two anonymous reviewers and the associate editor for comments that improved the manuscript. Our work would not be possible without the animal care support of the Division of Animal Resources (DAR) at UIUC.

Funding

This work is/was supported by the USDA National Institute of Food and Agriculture Hatch project 1026333 (ILLU-875-984 to KMS), a UIUC Graduate College Master's Fellowship (to KMS), a UIUC Illinois Distinguished Fellowship (to FOH), and UIUC Laboratory Start-up funds (to EKF).

Data availability

Data are included in the Supplementary information.

Author Contributions

KMS, FOH, HPA, and EKF conceived of and planned the experiments. KMS, FOH, and HPA carried out the experiments. KMS, FOH, and EKF contributed to data analysis and interpretation of the results. KMS wrote the manuscript with assistance from all authors.

Compliance with Ethical Standards

572 *Disclosure of potential conflicts of interest*

573 The authors declare no competing interests.

574

575 *Ethics approval*

576 All animal care and experimental procedures were approved by the University of Illinois Urbana-
577 Champaign Animal Care and Use Committee (protocol #20147) and followed all institutional and
578 national guidelines for the use of animals in research.

579

580

581 **References**

582 Andrade DV, Cruz-Neto AP, Abe AS, Wang T (2005) Specific dynamic action in ectothermic
583 vertebrates: a review of the determinants of postprandial metabolic response in fishes,
584 amphibians, and reptiles. In: Stark JM, Wang T (eds) Physiological and ecological
585 adaptations to feeding in vertebrates. Science Publishers, Enfield, pp 306-324

586 Baker MR, Goodman AC, Santo JB, Wong RY (2018) Repeatability and reliability of exploratory
587 behavior in proactive and reactive zebrafish, *Danio rerio*. Sci Rep 8:12114.
588 https://doi.org/10.1038/s41598-018-30630-3

589 Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using
590 lme4. J Stat Softw 67:1-48. https://doi.org/10.18637/jss.v067.i01

591 Bégué L, Peignier M, Ringler E (2023) The link between animal personality and habitat selection
592 in males of the Neotropical poison frog *Allobates femorali*s. Behaviour 160:257–274.
593 https://doi.org/10.1163/1568539X-bja10202

594 Bell AM (2005) Behavioural differences between individuals and two populations of stickleback
595 (*Gasterosteus aculeatus*). J Evol Biol 18:464–473. https://doi.org/10.1111/j.1420-
596 9101.2004.00817.x

597 Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis.

598 Anim Behav 77:771–783. https://doi.org/10.1016/j.anbehav.2008.12.022
599 Bergmüller R (2010) Animal personality and behavioural syndromes. In: Kappeler P (ed) Animal
600 Behavior: Evolution and Mechanisms. Springer, Berlin, pp 587-621
601 Breed M (2008) Warning colouration predicts behavioural and morphological differences in the
602 strawberry poison-dart frog. Thesis, Uppsala University
603 Brodin T, Lind MI, Wiberg MK, Johansson F (2013) Personality trait differences between
604 mainland and island populations in the common frog (*Rana temporaria*). Behav Ecol
605 Sociobiol 67:135–143. https://doi.org/10.1007/s00265-012-1433-1
606 Burghardt GM (2013) Environmental enrichment and cognitive complexity in reptiles and
607 amphibians: Concepts, review, and implications for captive populations. Appl Anim Behav
608 Sci 147:286–298. https://doi.org/10.1016/j.applanim.2013.04.013
609 Cabrera D, Nilsson JR, Griffen BD (2021) The development of animal personality across
610 Ontogeny: A cross-species review. Anim Behav 173:137–144.
611 https://doi.org/10.1016/j.anbehav.2021.01.003
612 Carere C, Locurto C (2011) Interaction between animal personality and animal cognition. Curr
613 Zool 57:491–498. https://doi.org/10.1093/czoolo/57.4.491
614 Carere C, Maestripieri D (2013) Animal personalities: behavior, physiology, and evolution.
615 University of Chicago Press, Chicago
616 Carter AJ, Marshall HH, Heinsohn R, Cowlishaw G (2012) How not to measure boldness: Novel
617 object and antipredator responses are not the same in wild baboons. Anim Behav 84:603–
618 609. https://doi.org/10.1016/j.anbehav.2012.06.015
619 Carvalho CF, Leitão AV., Funghi C, Batalha HR, Reis S, Mota PG, Lopes RJ, Cardoso GC
620 (2013) Personality traits are related to ecology across a biological invasion. Behav Ecol
621 24:1081–1091. https://doi.org/10.1093/beheco/art034
622 Chaloupka S, Peignier M, Stückler S, Araya-Ajoy Y, Walsh P, Ringler M, Ringler E (2022)
623 Repeatable territorial aggression in a neotropical poison frog. Front Ecol Evol 10:881387.

https://doi.org/10.3389/fevo.2022.881387

Cole, Ella F, Quinn, John L (2011) Personality and problem-solving performance explain competitive ability in the wild. Proc R Soc Lond B 279:1168–1175

Crothers LR, Cummings ME (2015) A multifunctional warning signal behaves as an agonistic status signal in a poison frog. Behav Ecol 26:560–568. https://doi.org/10.1093/beheco/aru231

Dochtermann NA, Schwab T, Sih A (2014) The contribution of additive genetic variation to personality variation: Heritability of personality. Proc R Soc B 282:20142201. https://doi.org/10.1098/rspb.2014.2201

Duckworth RA (2006) Behavioral correlations across breeding contexts provide a mechanism for a cost of aggression. Behav Ecol 17:1011–1019. https://doi.org/10.1093/beheco/arl035

Fischer EK, O'Connell LA (2020) Hormonal and neural correlates of care in active versus observing poison frog parents. Horm Behav 120:104696. https://doi.org/10.1016/j.yhbeh.2020.104696

Fischer EK, Roland AB, Moskowitz NA, Vidoudez C, Ranaivorazo N, Tapia EE, Trauger SA, Vences M, Coloma LA, O'Connell LA (2019) Mechanisms of convergent egg provisioning in poison frogs. Curr Biol 29:4145-4151.e3. https://doi.org/10.1016/j.cub.2019.10.032

Friard O, Gamba M (2016) BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol Evol 7:1325–1330. https://doi.org/10.1111/2041-210X.12584

Fouilloux CA, Serrano Rojas SJ, Carvajal-Castro JD, Valkonen JK, Gaucher P, Fischer M, Pašukonis A, Rojas B (2021) Pool choice in a vertical landscape: tadpole-rearing site flexibility in phytotelm-breeding frogs. Ecol Evol 11:9021-9038. https://doi.org/10.1002/ece3.7741

Garamszegi LZ, Herczeg G (2012) Behavioural syndromes, syndrome deviation and the within- and between-individual components of phenotypic correlations: When reality does not meet

650 statistics. Behav Ecol Sociobiol 66:1651–1658. https://doi.org/10.1007/s00265-012-1439-8
651 Garamszegi LZ, Markó G, Herczeg G (2013) A meta-analysis of correlated behaviors with
652 implications for behavioral syndromes: Relationships between particular behavioral traits.
653 Behav Ecol 24:1068–1080. https://doi.org/10.1093/beheco/art033
654 Grant EHC, Muths E, Schmidt BR, Petrovan SO (2019) Amphibian conservation in the
655 Anthropocene. Biol Conserv 236:543–547. https://doi.org/10.1016/j.biocon.2019.03.003
656 Herde A, Eccard JA (2013) Consistency in boldness, activity and exploration at different stages
657 of life. BMC Ecol 103:61–68. https://doi.org/10.1186/1472-6785-13-49
658 Ingle DJ (1990) Visually elicited evasive behavior in frogs. Bioscience 40:284–291.
659 https://doi.org/10.2307/1311265
660 Kelleher SR, Silla AJ, Byrne PG (2018) Animal personality and behavioral syndromes in
661 amphibians: a review of the evidence, experimental approaches, and implications for
662 conservation. Behav Ecol Sociobiol 72:79. https://doi.org/10.1007/s00265-018-2493-7
663 Kudo H, Nishizawa H, Uchida K, Sato K (2021) Boldness–exploration behavioral syndrome in
664 wild sub-adult green sea turtles caught at Oita, Japan. Appl Anim Behav Sci 236:105216.
665 https://doi.org/10.1016/j.applanim.2021.105216
666 Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed
667 effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/JSS.V082.I13
668 Lenth R (2021) emmeans: Estimated Marginal Means, aka Least-Squares Means. R package
669 version 1.8.5, https://github.com/rvlenth/emmeans.
670 Liu Y, Day LB, Summers K, Burmeister SS (2016) Learning to learn: Advanced behavioural
671 flexibility in a poison frog. Anim Behav 111:167–172.
672 https://doi.org/10.1016/j.anbehav.2015.10.018
673 Maan ME, Cummings ME (2012) Poison frog colors are honest signals of toxicity, particularly for
674 bird predators. Am Nat 179:E1-E14. https://doi.org/10.1086/663197
675 Magnhagen C, Wacker S, Forsgren E, Myhre LC, Espy E, Amundsen T (2014) Context

676 consistency and seasonal variation in boldness of male two-spotted gobies. PLoS ONE
677 9:e93354. https://doi.org/10.1371/journal.pone.0093354
678 Maier NR, Schneirla TC (1964) Principles of animal psychology. Dover, London
679 Mazué GPF, Dechaume-Moncharmont FX, Godin JGJ (2015) Boldness-exploration behavioral
680 syndrome: Interfamily variability and repeatability of personality traits in the young of the
681 convict cichlid (*Amatitlania siquia*). Behav Ecol 26:900–908.
682 https://doi.org/10.1093/beheco/arv030
683 Merrick MJ, Koprowski JL (2017) Should we consider individual behavior differences in applied
684 wildlife conservation studies? Biol Conserv 209:34–44.
685 https://doi.org/10.1016/j.biocon.2017.01.021
686 Michelangeli M, Chapple DG, Goulet CT, Bertram MG, Wong BBM (2019) Behavioral
687 syndromes vary among geographically distinct populations in a reptile. Behav Ecol 30:393–
688 401. https://doi.org/10.1093/beheco/ary178
689 Moiron M, Laskowski KL, Niemelä PT. 2019. Individual differences in behaviour explain
690 variation in survival: A meta-analysis. Ecol Lett 23:399–408.
691 https://doi.org/10.1111/ele.13438
692 Murgui E, Hedblom M (eds) (2017) Ecology and conservation of birds in urban environments.
693 Springer, Cham. https://doi.org/10.1007/978-3-319-43314-1
694 Niemelä PT, Dingemanse NJ (2014) Artificial environments and the study of “adaptive”
695 personalities. Trends Ecol Evol 29:245–247. https://doi.org/10.1016/j.tree.2014.02.007
696 O’Connell LA, Hofmann HA (2011) Genes, hormones, and circuits: An integrative approach to
697 study the evolution of social behavior. Front Neuroendocrinol 32:320–335.
698 https://doi.org/10.1016/j.yfrne.2010.12.004
699 Odom KJ, Araya-Salas M, Morano JL et al (2021) Comparative bioacoustics: A roadmap for
700 quantifying and comparing animal sounds across diverse taxa. Biol Rev 96:1135–1159.
701 https://doi.org/10.1111/brv.12695

Pamela Delarue EM, Kerr SE, Lee Rymer T (2015) Habitat complexity, environmental change and personality: A tropical perspective. Behav Process 120:101–110. https://doi.org/10.1016/j.beproc.2015.09.006

Pašukonis A, Serrano-Rojas SJ, Fischer MT et al (2022) Contrasting parental roles shape sex differences in poison frog space use but not navigational performance. eLife 11:e80483. https://doi.org/10.7554/eLife.80483

Peignier M, Araya-Ajoy YG, Bégué L, Chaloupka S, Dellefont K, Walsh P, Ringler M, Ringler E (2022) Exploring links between personality traits and their social and non-social environments in wild poison frogs. Behav Ecol Sociobiol 76:93. https://doi.org/10.1007/s00265-022-03202-9

Peignier M, Bégué L, Ringler M, Szabo B, Ringler E (2023) Regardless of personality, males show similar levels of plasticity in territory defense in a Neotropical poison frog. Sci Rep 13:3435. https://doi.org/10.1038/s41598-023-30546-7

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Rajput N, Parikh K, Kenney JW (2022) Beyond bold versus shy: Zebrafish exploratory behavior falls into several behavioral clusters and is influenced by strain and sex. Biol Open 11:bio05944. https://doi.org/10.1242/bio.059443

Réale D, Dingemanse NJ, Kazem AJN, Wright J (2010) Evolutionary and ecological approaches to the study of personality. Phil Trans R Soc B 365:3937–3946. https://doi.org/10.1098/rstb.2010.0222

Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. Biol Rev 82:291–318

Ringler E, Barbara Beck K, Weinlein S, Ludwig H, Ringler M (2017) Adopt, ignore, or kill? Male poison frogs adjust parental decisions according to their territorial status. Sci Rep 7:43544. https://doi.org/10.1038/srep43544

Rojas B, Endler JA (2013) Sexual dimorphism and intra-populational colour pattern variation in the aposematic frog *Dendrobates tinctorius*. Evol Ecol 27:739–753. https://doi.org/10.1007/s10682-013-9640-4

Rojas B, Pašukonis A (2019) From habitat use to social behavior: Natural history of a voiceless poison frog, *Dendrobates tinctorius*. PeerJ 7:e7648. https://doi.org/10.7717/peerj.7648

Roland AB, O'Connell LA (2015) Poison frogs as a model system for studying the neurobiology of parental care. Curr Opin Behav Sci 6:76–81. https://doi.org/10.1016/j.cobeha.2015.10.002

RStudio Team (2021) RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA

Rudolfová V, Petrásek T, Antošová E, Frynta D, Landová E, Valeš K, Nekovářová T (2022) Inter-individual differences in laboratory rats as revealed by three behavioural tasks. Sci Rep 12:9361. https://doi.org/10.1038/s41598-022-13288-w

Shaw AK (2020) Causes and consequences of individual variation in animal movement. Mov Ecol 8:12. https://doi.org/10.1186/s40462-020-0197-x

Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: An ecological and evolutionary overview. Trends Ecol Evol 19:372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih A, Del Giudice M (2012) Linking behavioural syndromes and cognition: A behavioural ecology perspective. Phil Trans R Soc B 367:2762–2772. https://doi.org/10.1098/rstb.2012.0216

Smith BR, Blumstein DT (2008) Fitness consequences of personality: A meta-analysis. Behav Ecol 19:448–455. https://doi.org/10.1093/beheco/arm144

Smith BR, Blumstein DT (2010) Behavioral types as predictors of survival in Trinidadian guppies (*Poecilia reticulata*). Behav Ecol 21:919–926. https://doi.org/10.1093/beheco/arq084

Sonnleitner R, Alanen E, Fouilloux C, Valkonen JK, Rojas B (2022) Genetic factors influence behavioural repeatability in juvenile poison frogs. bioRxiv, 2022.09.19.508512.

https://doi.org/10.1101/2022.09.19.508512

Stamps, Judy A (2007) Growth-mortality tradeoffs and ‘personality traits’ in animals. Ecol Lett 10:355–363. https://doi.org/10.1111/j.1461-0248.2007.01034.x

Stiegler J, Lins A, Dammhahn M, Kramer-Schadt S, Ortmann S, Blaum N (2022) Personality drives activity and space use in a mammalian herbivore. Mov Ecol 10:33. https://doi.org/10.1186/s40462-022-00333-6

Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol 8:1639–1644. https://doi.org/10.1111/2041-210X.12797

Stynoski JL, Schulte LM, Rojas B (2015) Poison frogs. Curr Biol 25:1026–1028. https://doi.org/10.1016/j.cub.2015.06.044

Summers K (2003) Convergent evolution of bright coloration and toxicity in frogs. P Natl Acad Sci USA 100:12533–12534. https://doi.org/10.1073/pnas.2335928100

Theódórsson BK, Ólafsdóttir GÁ (2022) Laboratory measures of boldness correlate with ecological niche in threespine stickleback. Ecol Evol 12:e9235. https://doi.org/10.1002/ece3.9235

Uher J (2008) Comparative personality research: Methodological approaches. Eur J Pers 22:427–455. https://doi.org/10.1002/per.680

Waldeck RF, Gruberg ER (1995) Studies on the optic chiasm of the leopard frog. Brain Behav Evol 46:84–94. https://doi.org/10.1016/j.anbehav.2022.10.005

White JR, Meekan MG, McCormick MI, Ferrari MCO (2013) A comparison of measures of boldness and their relationships to survival in young fish. PLoS ONE 8:e68900. https://doi.org/10.1371/journal.pone.0068900

Wickham H, Averick M, Bryan J et al (2019) Welcome to the Tidyverse. J Open Source Softw 4:1686. https://doi.org/10.21105/joss.01686

Wilson ADM, Godin J-GJ, Ward AJW (2010) Boldness and reproductive fitness correlates in the

780 eastern mosquitofish, *Gambusia holbrooki*. Ethology 116:96–104.
781 https://doi.org/10.1111/j.1439-0310.2009.01719.x
782 Wong BBM, Candolin U (2015) Behavioral responses to changing environments. Behav Ecol
783 26:665–673. https://doi.org/10.1093/beheco/aru183
784 Yuen CH, Schoepf I, Schradin C, Pillay N (2017) Boldness: are open field and startle tests
785 measuring the same personality trait? Anim Behav 128:143–151.
786 https://doi.org/10.1016/j.anbehav.2017.04.009

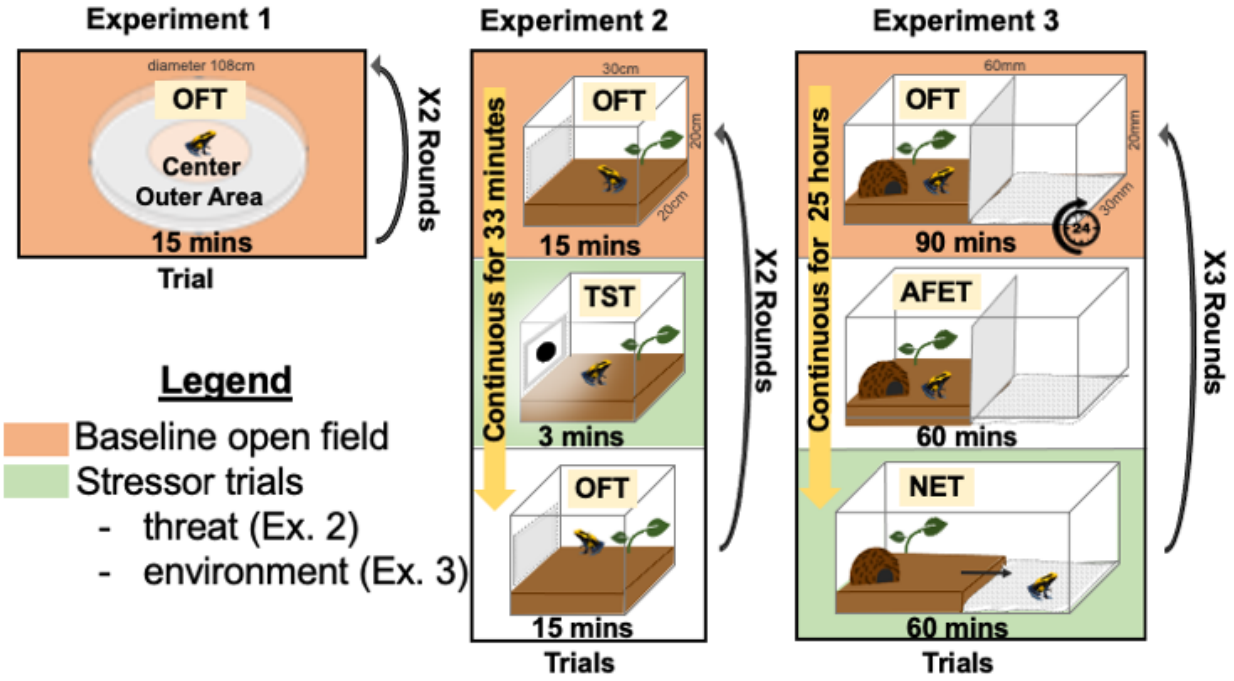
Tables

Table 1 Overview of statistical results for behavioral repeatability, correlations, and linear models. Exploration was not quantified in Experiment 3. Significance interpreted as: R= 0.2-0.4 moderate repeatability, corr≥0.5 strong correlation, p< 0.05 significant. Pearson correlation (corr) used for normal data, Kendall Rank correlation (tau) used for non-parametric data

Repeatability			
Behavior	**Experiment 1**	**Experiment 2**	**Experiment 3**
Activity	**R= 0.279, p= 0.032**	**R= 0.123, p= 0.006**	**R= 0.321, p= 0.011**
Exploration	**R= 0.278, p= 0.029**	**R= 0.177, p< 0.001**	NA
Boldness	R= 0.044, p= 0.395	R= 0, p= 1	**R= 0.388, p= 0.003**
Behavioral Syndromes			
Syndrome	**Experiment 1**	**Experiment 2**	**Experiment 3**
Activity:Exploration	**corr= 0.767, p< 0.001**	**corr= 0.911, p< 0.001**	NA
Activity:Boldness	corr= -0.150, p= 0.196	**tau= 0.155, p= 0.002**	**corr= -0.665, p< 0.001**
Exploration:Boldness	corr= -0.119, p= 0.307	**tau= 0.155, p= 0.002**	NA
ANOVAs			
Activity	**Experiment 1**	**Experiment 2**	**Experiment 3**
Trial	NA	**$F_{1,234}$= 170.128, p< 0.001**	**$F_{2,143}$= 18.387, p<0.001**
Round	$F_{1,37}$= 3.390, p= 0.074	$F_{1,232}$= 0.570, p= 0.451	**$F_{2,143}$=3.919, p= 0.022**
Trial:Round	NA	$F_{2,228}$= 0.216, p= 0.801	$F_{4,143}$ = 1.356, p= 0.252
Exploration	**Experiment 1**	**Experiment 2**	**Experiment 3**
Trial	NA	**$F_{2,230}$= 112.143, p< 0.001**	NA
Round	$F_{1,37}$= 1.392, p= 0.245	$F_{1,234}$= 0.292, p= 0.590	NA
Trial:Round	NA	$F_{2,228}$= 0.506, p= 0.604	NA
Boldness	**Experiment 1**	**Experiment 2**	**Experiment 3**
Round	$F_{1,74}$= 0.628, p= 0.431	$\chi^2_{2,96}$= 22.429, p= 0.401	**$F_{2,36}$ = 13.100, p< 0.001**

794 **Figures**

795



796

797 **Fig 1** Overview of experimental design. All experiments were built on the framework of an open

798 field test and expanded to explore differences in time and context. For each experiment, trial

799 type (i.e., different phases within each experiment), trial duration, and number of rounds (i.e.,

800 repeated experiences in the same assay) are shown. OFT = open field experiment, TST =

801 threat stimulus test, AFET = activity in a familiar environment test, NET = novel environment

802 test. Colors represent corresponding trial types that were compared across experiments (see

803 “Cross Analysis” Methods)

804

805

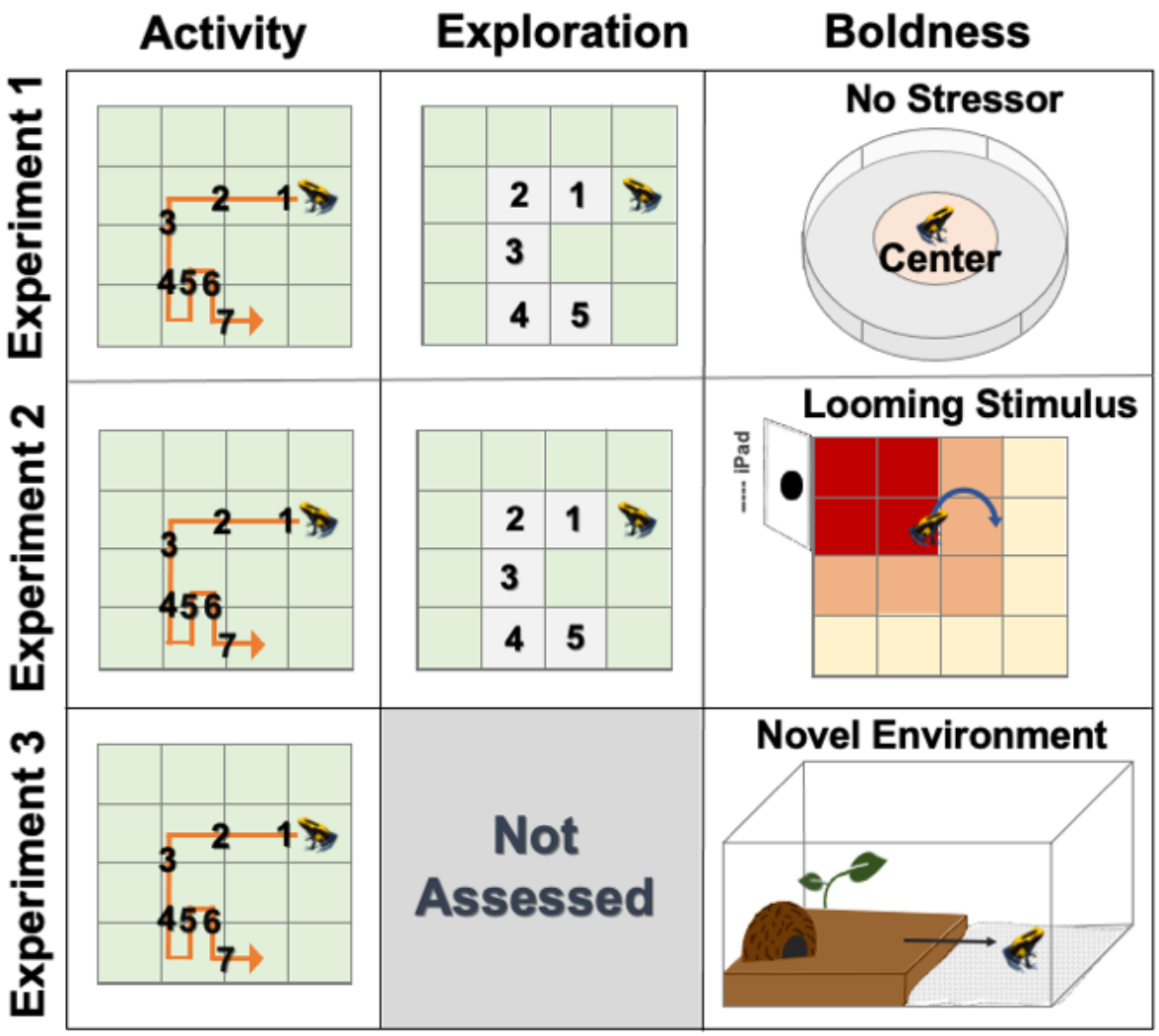


Fig 2 Overview of analysis methods. For all experiments, we quantified activity as the number of lines crossed from one area into the next and exploration as the total number of areas visited (not assessed in Experiment 3). Note that crossing back and forth from one area to another increases our measure of activity but not exploration. We quantified boldness as time in center (Experiment 1), response to a looming stimulus (Experiment 2), and latency to cross into a novel environment (Experiment 3)

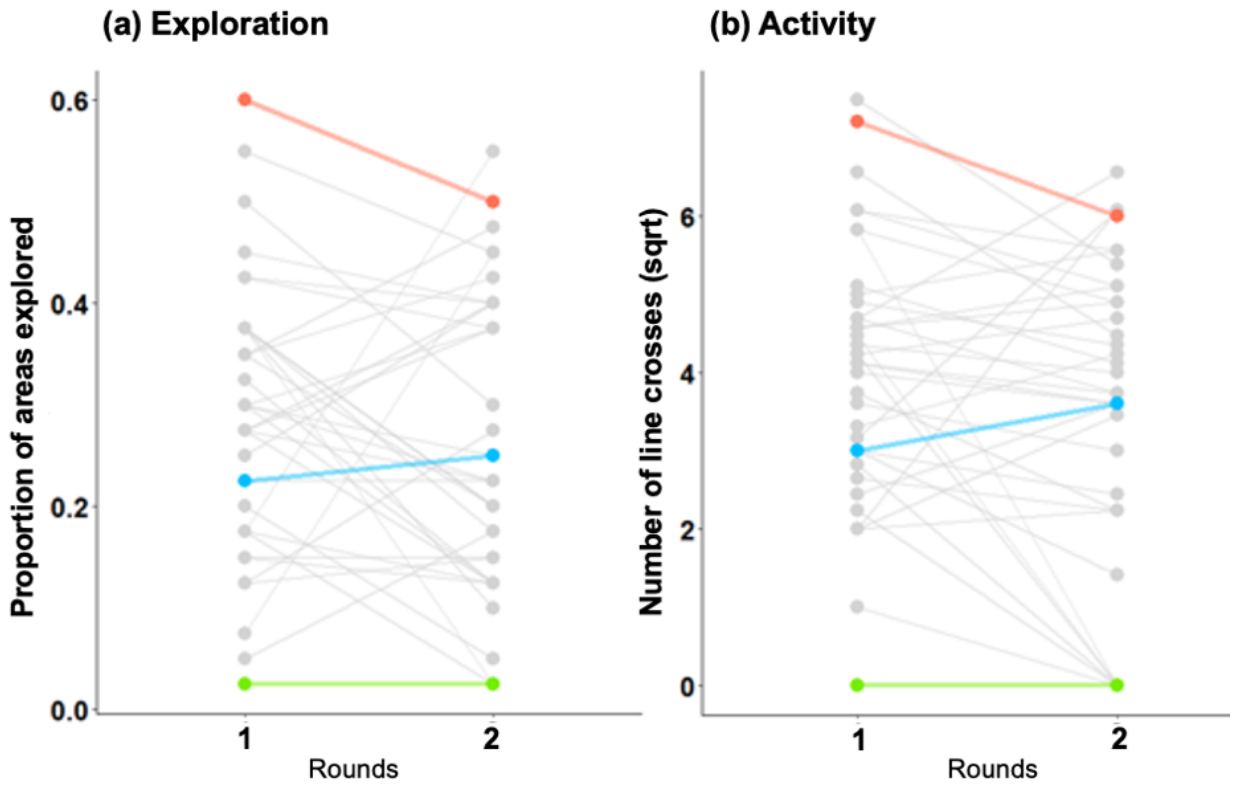


Fig 3 Behavior in an artificial environment open field test (Experiment 1). Frogs exhibited repeatability in (a) exploration (R= 0.278, p= 0.029) and (b) activity (R= 0.279, p= 0.032) across two rounds separated by five weeks. Three frogs are highlighted to showcase consistent individual differences in and the relationship between exploration and activity

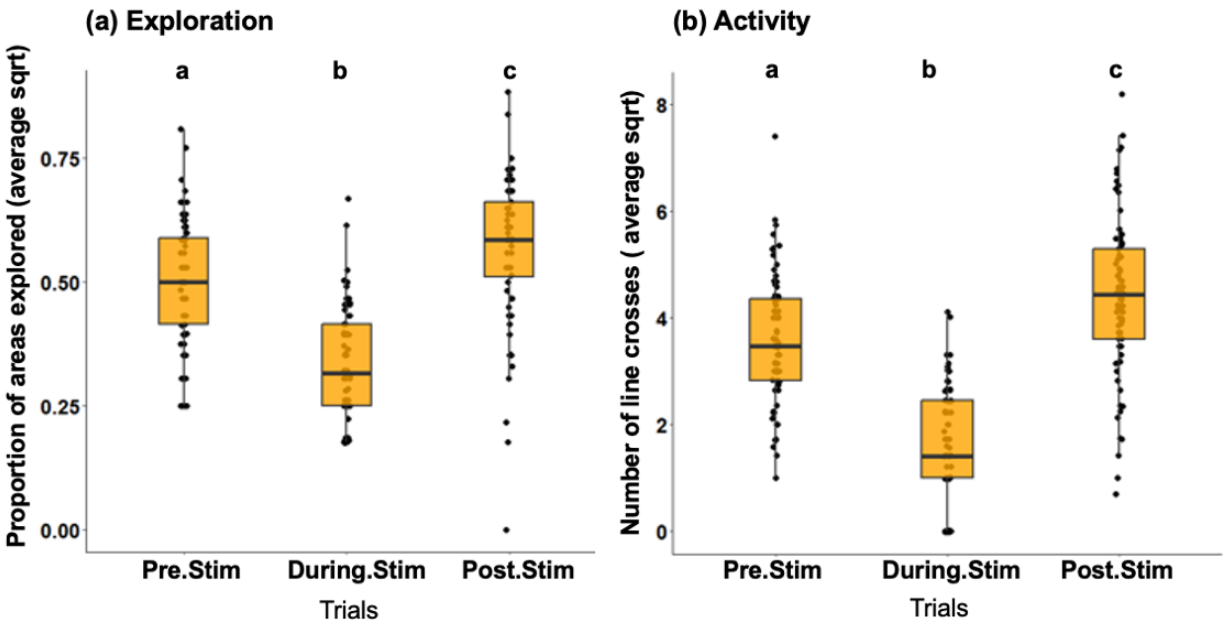


Fig 4 Behavior in a naturalistic open field before, during, and after a threat stimulus (Experiment 2). Exploration (a) and activity (b) decrease during stimulus presentation compared to pre-stimulus and post-stimulus ($F_{2,230}$= 112.143, p< 0.001; $F_{1,234}$= 170.128, p< 0.001), and these patterns were repeatable across the two rounds of the experiment. Data are collapsed across rounds, as no significant effect of round or a trial by round interaction were observed for exploration or activity. The boxes show the median, the first and third quartiles, and whiskers as 1.5x the interquartile range. Dots show individual data points.

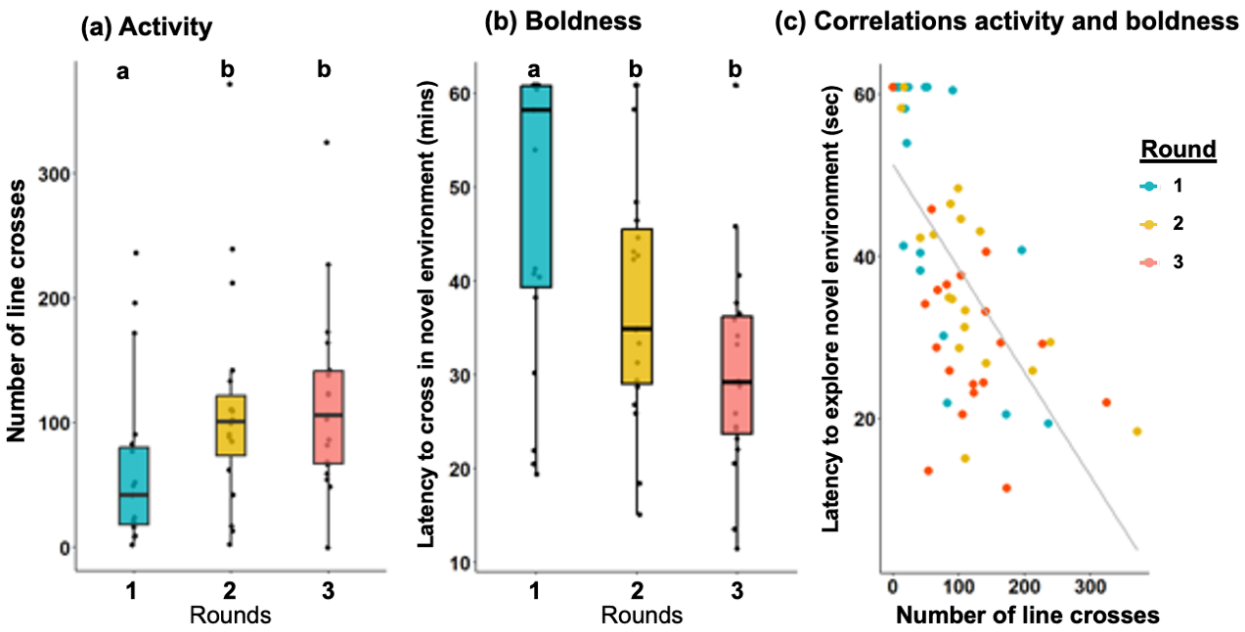


Fig 5 Behavior in a familiar, naturalistic environment and during the novel environment test (Experiment 3). (a) Activity increased with repeated experience in the assays (i.e., across rounds for the OFT, AFET, and NET; ($F_{2,143}$= 3.919, p= 0.022). (b) All frogs explored the novel area more quickly in subsequent rounds ($F_{2,36}$ = 13.100, p< 0.001), and (c) more active frogs explored the novel area sooner (corr = -0.655, p< 0.001). Letters in (a) and (b) indicate significant pair-wise differences between rounds. The boxes show the median, the first and third quartiles, and whiskers as 1.5x the interquartile range. Dots show individual data points.