Behavioral Ecology

The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2021), 32(4), 686–694. doi:10.1093/beheco/arab019

Original Article

Spontaneous quantity discrimination in a family-living lizard

Birgit Szabo,[a,b] Daniel W. A. Noble,[c] Kaitlin J. McCloghry,[b] Marco E. S. Monteiro,[b] and Martin J. Whiting[b]

[a]Division of Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Wohlenstrasse 50a, 3032 Bern, Switzerland, [b]Department of Biological Sciences, Macquarie University, North Ryde, Sydney, NSW 2109, Australia, and [c]Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

Received 9 June 2020; revised 8 February 2021; editorial decision 16 February 2021; accepted 17 February 2021; Advance Access publication 28 April 2021.

While foraging or during social interactions, animals may benefit from judging relative quantity. Individuals may select larger prey or a patch with more food and, likewise, it may pay to track the number and type of individuals and social interactions. We tested for spontaneous quantity discrimination in the gidgee skink (*Egernia stokesii*), a family-living lizard. Lizards were presented with food quantities differing in number or size and were able to select the larger quantity of food items when they differed in number but not when items differed in size. We show, for the first time, superior spontaneous discrimination of items differing in number over size in a lizard species, which contrasts with previous findings. Our simple method, however, did not include controls for the use of continuous quantities, and further tests are required to determine the role of such information during quantity discrimination. Our results provide support for the use of the parallel individuation system for the discrimination of small quantities (four or fewer items). Lizards might, however, still use the approximate number system if items in larger quantities (more than four) are presented. Overall, we uncovered evidence that species might possess specific cognitive abilities potentially adapted to their niche with respect to quantity information (discrete and/or continuous) and the processing system used when judging quantities. Importantly, our results highlight the need for testing multiple species using similar testing procedures to gain a better understanding of the underlying causes leading to differences across species.

Key words: **cognition, number sense, relative quantity judgment, reptile, squamate**

INTRODUCTION

While foraging and interacting with conspecifics, animals may benefit from an ability to judge relative quantities (number sense)—a cognitive process involving counting and/or set size discrimination (Krebs et al. 1974; Agrillo 2015; Giaquinto 2018). Having a numerical sense is beneficial for optimizing food intake by choosing the larger quantity of food or during social interactions to track individuals. A numeical sense is also useful in competitive environments or predatory interactions to evaluate the number of opponents, competitors, and/or the threat level (Krebs et al. 1974; Agrillo 2015; Nieder 2020). In a foraging context, when alone, individuals might choose solely based on food quantity, while, in a social context, an individual's choice should also be affected by the number and identity of individuals competing for the same food source (ideal free choice; Harper 1982; Kennedy and Gray 1993). Similarly, during aggressive interactions with conspecifics, group size determines the behavior (attack or retreat) of individuals. Chimpanzees (Wilson et al. 2001), lionesses (McComb et al. 1994), and even ants (Tanner 2006) are more willing to enter contests if they outnumber their competitors.

Studies on spontaneous quantity discrimination provide evidence for such a number sense. From these studies, we can distinguish two nonsymbolic cognitive systems for number representation (Hyde 2011): the "parallel individuation system" (i.e., object-file system), in which the magnitude is represented as discrete quantities limited by a maximum set size that can be recognized (usually 3–4), and the "approximate number system," in which magnitudes are estimated and discriminated based on their ratio. The approximate number system is characterized by a ratio (Weber's law) and distance effect; the larger the ratio or the smaller the distance between two quantities, the harder they are to discriminate (Hyde 2011). The presence of such effects is evidence for the use of the approximate number system, while their absence is used as evidence for the use of the parallel individuation system (Agrillo 2015). In humans, these systems are encoded separately and, most often, it is assumed that they are mutually exclusive (Feigenson et al. 2004). Recent work

Address correspondence to B. Szabo. E-mail: birgit.szabo@gmx.at.

in humans has shown, however, that visual processing, attention, and working memory capacity determine which of the two systems is engaged during numerical discriminations (Hyde 2011).

Numerical abilities have mainly been studied in mammals and birds with much less investigation into the numerical competency of other vertebrates, such as fishes (Agrillo and Bisazza 2017), and only a handful of studies have looked at spontaneous quantity discrimination in amphibians and reptiles. Red-backed salamanders (*Plethodon cinereus*), for example, spontaneously chose the larger quantity of flies (*Drosophila virilis*) in a discrimination between one versus two (ratio 0.5) and two versus three (ratio 0.67) but not between three versus four (ratio 0.75) or six versus four (ratio 0.67; Uller et al. 2003). Red-legged (*Plethodon shermani*) and gray-cheeked (*Plethodon metcalfi*) salamanders were able to spontaneously discriminate between groups of live or videos of live crickets (*Gryllus bimaculatus*) differing in ratios of 0.5 (8 vs. 16) but failed to do so when the ratio was larger (0.67; 8 vs. 12) or when crickets were computer animated controlling for movement (Krusche et al. 2010). Fire-bellied toads (*Bombina orientalis*) spontaneously discriminated between mealworm larvae (*Tenebrio molitor*) when one versus two, two versus three, three versus six, and four versus eight larvae were presented but not when three versus four or four versus six larvae were used even when continuous physical properties (movement, volume, weight, and surface) were controlled for (Stancher et al. 2015). Thus far, only two species of reptiles have been tested on their ability to spontaneously discriminate between quantities: the Italian wall lizard (*Podarcis sicula*) and the Hermann's tortoise (*Testudo hermanni*). Wall lizards were able to select the larger-sized dead fly larvae (*Musca domestica*) differing in ratios 0.25, 0.5, 0.67, and 0.75. When groups of larvae (discrimination based on number) were used, lizards were unable to discriminate between quantities (Miletto Petrazzini et al. 2017). Hermann's tortoises were tested using a similar methodology but, instead of dead fly larvae, they were discriminating between differently sized, and groups of, tomato slices. Contrary to the lizards, tortoises chose more over less in both tests (Gazzola et al. 2018). In both these studies, animals' accuracy followed Weber's law such that smaller ratios were easier to discriminate compared to larger ratios (Hyde 2011). From these few studies, it becomes clear that, to date, we are unable to determine what cognitive system is used by amphibians or reptiles and why species differ in performance. Collecting data on additional species is an important step toward understanding what might cause differences across species when judging relative quantities. The observed similarities in performance between the tortoises and lizards could be based on convergent evolution or represent an evolutionary, deeply conserved ability which can only be determined with additional data. Furthermore, to establish if the observed differences in performance are caused by factors such as social system, feeding ecology, or other traits, we need to compare closely related species (e.g., sister taxa) or compare among populations within species, that vary in these traits.

Here, we tested spontaneous quantity discrimination in a social lizard, the Australian gidgee skink (*Egernia stokesii*). Gidgee skinks live in stable family groups of up to 17 individuals comprising of one monogamous pair and their offspring from multiple years (Gardner et al. 2001; Chapple 2003; Cogger 2014) to which they give live birth (viviparity; Cogger 2014). Our aim was to test quantity discrimination in a group-living lizard species with kin-based sociality. Gidgee skinks could benefit from a number sense in social situations, such as when deciding to disperse or when making judgments about social status (Shettleworth 2010).

We based our methodology on previous work in reptiles (Miletto Petrazzini et al. 2017; Gazzola et al. 2018) in which individuals were presented with a simultaneous choice between two food quantities in a number of unique combinations. We replicated previous work in reptiles as best as possible to produce data that are comparable to those collected previously. We presented up to four identical carrot strips in a discrimination based on number, while, in the experiment based on size, we used single strips differing in length from 1 to 4 cm. Our study extends previous work (Miletto Petrazzini et al. 2017; Gazzola et al. 2018) by 1) testing a lizard species with a family-based social system and 2) including two additional ratios: 0.33 (one vs. three) and 0.5 (one vs. two) to the four ratios (0.25: one vs. four; 0.5: two vs. four; 0.67: two vs. three; and 0.75: three vs. four) tested previously (Miletto Petrazzini et al. 2017; Gazzola et al. 2018).

METHODS

Animal capture and husbandry

We used 12 adult gidgee skinks of undetermined sex (Supplementary Table S1), wild-caught at Fowlers Gap Arid Zone Research Station (−31.086972 S, 141.704836 E), New South Wales, Australia, during March/November 2018. Lizards were captured by hand or Elliott traps and transported back to Macquarie University by car in cloth bags a maximum of 1 week after capture and set up individually in plastic tub enclosures (683 L × 447 W × 385 H mm) in a temperature-controlled room. Lizards were not individually marked but identified using individual markings and coloration. Room temperature was set to 24 °C with a light cycle of 12 h (6:00 AM to 6:00 PM) and a heat cord under one side of the enclosure increased the temperature to 33 °C (±2 °C standard deviation [SD]) to allow lizards to thermoregulate. To ensure optimal temperatures for lizards, temperature was monitored hourly from within enclosures using Thermochron iButtons (model DS1921). Enclosures contained a refuge (upside down brown plant saucer 200 mm in diameter), a water bowl (poly resin reptile water bowls made to look like rock, 130 L × 110 W × 40 H mm), and, as enrichment, one or two wooden ramps, a stone, some bark, leaves, and a 150-mm long PVC tube. Lizards were kept on butcher's paper (taped down during trials to prevent lizards from crawling underneath and out of sight), which was replaced once a week on Fridays to give lizards ample time to spread their scent over the weekend before the next test day.

When not in experiments, lizards were fed three times per week (Monday, Wednesday, and Friday, between 08:00 and 12:00 AM) with small cut fruits and vegetables presented in green food dishes (60 mm diameter) adding two or three crickets (powdered with aristopet Repti-vite and URS Ultimate Calcium) on Fridays. During the time of experimental testing, lizards were fed carrots during trials from Monday to Friday (6–12 times 0.065 g ± 0.021 SD each day), adding their regular diet of fruits, vegetables, and crickets on Fridays. Lizards had ad libitum access to water.

Setup

All tests were carried out in the animal's home enclosure to avoid the negative effects of handling stress (Langkilde and Shine 2006) and lizards were tested either between 07:00 and 11:00 AM or between 12:00 and 4:00 PM. The area in which lizards were tested

was located within the same room they were housed in but on a separate shelf about 3 m away from the animals. The area was surrounded by a gray curtain to reduce distractions and obscure the researcher during trials, and a heat cord was placed on the shelf (similar to their housing conditions) to elevate the temperature allowing lizards to thermoregulate during trials.

We used carrot as a reward and test stimuli throughout the experiment (carrot is a highly preferred food item). Carrot strips were prepared fresh each day, but the same strips (stimuli) were used for all lizards tested within a day. We created carrot strips using a grater to achieve equal width and cut the strips into the appropriate length for each experiment using a ruler.

General procedure

We conducted three tests: first, a side preference and, thereafter, two quantity discrimination tests. All lizards were tested in the side preference test and we used a between-individuals test design in which six lizards were tested on a number discrimination and the other six on a size discrimination. The number discrimination test was run in August 2019 and the size discrimination test was run in September to October 2019 except for lizard ID 4, which was tested on the number discrimination during September to October 2019.

First, a lizard was gently carried in its enclosure to the test area. Lizards were given 3 min to acclimate from being carried over to the test area. In this time, the lizard was covered with the hide if not already under it and moved backward to the starting position at the furthest end of the enclosure from the experimenter (Figure 1d) and all enrichment items and the water bowl were removed from the enclosure. Next, a wooden divider (Figure 1c) was placed inside the enclosure at the end closest to the experimenter (Figure 1d). After the 3-min acclimation period was over, dishes were introduced at each side of the divider and then the hide was removed to expose lizards to the experimental setup. The lizard was left to choose a dish (a choice was defined as the first dish touched by the lizard). Once a choice was made, the hide was replaced and the lizard was gently moved back to the starting position (by hand or under the hide) in preparation of the next trial for an intertrial interval of 30 s after which another set of dishes was introduced. The order in which lizards were tested each day was randomized. All trials were recorded automatically from above using a CCTV camera setup (H.264 Digital Video Recorder, 3-Axis Day & Night Dome Cameras) and the behavior of lizards was observed live on a computer screen.

All lizards had previously participated in cognitive assays studying their visual discrimination and reversal learning ability (Szabo et al. 2021b), as well as their ability to exercise motor response inhibition (Szabo et al. 2020). This pre-experience with other cognitive tests might have influenced lizard behavior and made them better habituated to testing procedures.

Side preference test

Two green food dishes (60 mm diameter) were placed on either side of the wooden divider, each including one strip of carrot (0.065 g ± 0.021 SD). Lizards were familiar with these dishes as they were used for feeding during regular husbandry. After the hide was removed, the trial ran until a lizard chose a food dish. Once a dish was chosen, the lizard was allowed to eat the carrot within the chosen dish, while the second dish was removed. We conducted 14 trials with an intertrial interval of 30 s and recorded the side of the first dish chosen in each trial. Each lizard received all 14 trials within a single day. At the beginning of the first out of 14 trials, each lizard was fed an additional carrot strip (to the reward placed within food dishes) presented with forceps to prime them and initiate food searching behavior. Between trials, the enclosure was gently rotated 180° (after which lizards were given at least 10 s before the next trial started) to increase the likelihood that lizards chose based on an egocentric strategy rather than external cues.

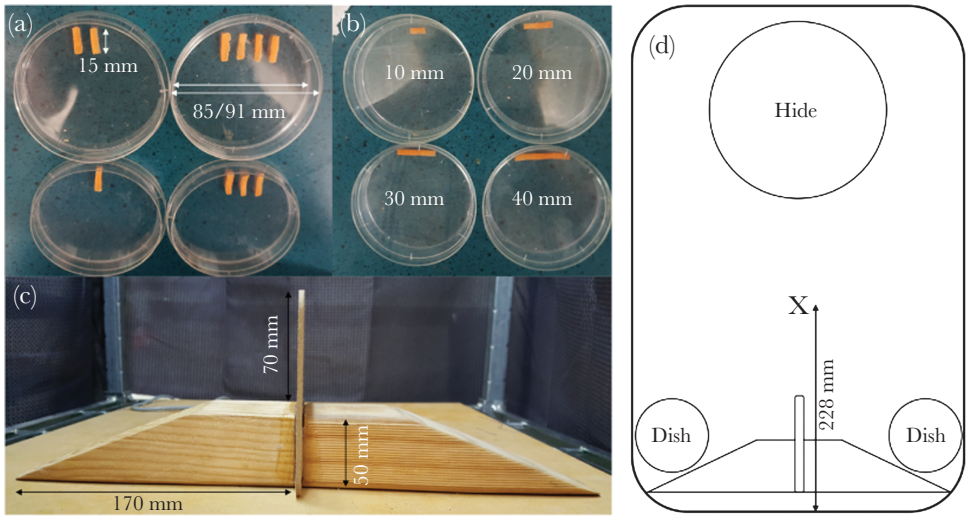

Figure 1

(a) Petri dish setup used during the number discrimination test. Strips were always approximately 5 mm apart. (b) Petri dish setup used in the size discrimination test. (c) Picture of the wooden divider (height 93 mm) used in all tests. Two wooden ramps were glued together using nontoxic silicone with a wooden coaster in between. (d) Schematic setup. The wooden divider was placed on one end of the enclosure, while the lizard was moved to the other end of the enclosure underneath the hide. Next, dishes were placed at either side of the divider. The X indicates the starting position from which each lizard made its choice (lured there by presenting a strip of carrot with forceps), equidistant from both dishes, 228 mm away from the wall behind the divider.

Quantity discrimination

We presented lizards with six combinations of quantities: one versus four, one versus three, one versus two, two versus four, two versus three, and three versus four (ratio: 0.25, 0.33, 0.50, 0.50, 0.67, and 0.75, respectively). Lizards performed six trials a day from Monday to Friday. Within a day, each combination was only presented once and, every day, the order of presentation followed a different, predetermined random order. This order was different for lizards tested on number and size. However, by accident, one out of the six lizards in each test group received the order of the other group. The side the larger quantity was presented was randomized across trials. Each lizard received a minimum of 90 trials. Trials were continued until we had collected data from 15 trials of each combination per lizard. Both quantity discrimination tests were conducted within 4 weeks and immediately followed the preference test.

Four petri dishes were prepared for each lizard each day. We used cleaned petri dishes for each lizard each day, and all used petri dishes were cleaned using 70% ethanol to remove previously deposited olfactory traces that could impact choice. In the number discrimination, equal-sized strips (15 mm long, 0.065 g ± 0.021 SD each) were arranged perpendicular and 5 mm from one another (Figure 1a) and, in the size discrimination one, 10 (0.053 g ± 0.01 SD), 20 (0.109 g ± 0.03 SD), 30 (0.151 g ± 0.03 SD), or 40 mm (0.210 g ± 0.02 SD) strips were placed longitudinally to the edge of the dishes (Figure 1b). On the first test day, an additional trial was conducted in which lizards were presented with a choice between an empty dish and a dish including one carrot strip (15 mm long). Behavior during test trials indicated that all lizards perceived the carrots in the dishes as food because they tried to reach and eat them. To control for the use of food odor during trials, we hid four carrot strips (15 mm long) behind the divider. To induce food searching behavior in test trials and keep motivation high, a lizard was lured to a starting position (marked by an X on the enclosure floor; Figure 1d) by presenting a carrot strip with forceps at the start of each trial after the hide was removed and before the lizard was allowed to choose a dish. This was done to ensure that lizards always started each trial equidistant from both choices. The lizard was allowed to eat the carrot while the experimenter moved behind the curtain. During trials, petri dishes, including stimuli, were angled about 45° to ensure that lizards had an unobstructed view of all stimuli from the starting position. Petri dishes were not elevated because lizards were able to view the stimuli within the dishes unobstructed from the staring position when dishes were placed on the enclosure floor.

We recorded the food quantity within the first dish touched. No blind coding of choice behavior was possible due to the nature of the experiment. Lizards were not allowed to eat the stimuli to prevent learning. If the lizard did not make a choice (had not moved from the staring position) by 2 min after trial start, an additional carrot piece was given to increase motivation. By 5 min, if still no choice (no movement) had been made, the lizard was placed back under the hide in preparation for the next trial. These trials were recorded as failed and repeated on additional test days following the completion of all sessions at the end of the 3-week period to ensure that each lizard finished 15 trials per combination. One lizard (ID 2) lost motivation during testing only completing part of the 90 test trials by the end of the 3-week period. Consequently, we replaced this lizard with another animal (ID 4) to prevent the effects of motivation influencing our results. Due to an error, one lizard (ID 14) received 16 trials in combination two versus three and 14 trials in combination one versus three in the number discrimination test, which we accounted for in the statistical analysis.

Statistical analyses

We used Bayesian generalized linear mixed models (GLMMs; Hadfield 2010) to investigate if lizards were more likely to choose the larger quantity (number and size) of food. Choosing the larger quantity (larger = 1 and smaller = 0, Bernoulli variable) was included as the response variable and combination (contrasts: one vs. four, one vs. three, one vs. two, two vs. four, two vs. three, and three vs. four) as the fixed effect. The model included a random intercept (animal identity) and random slope (session) to account for repeated measures and autocorrelation between successive choices across days. We used the resulting posterior of these models to calculate the difference of each combination from chance and differences between combinations (results shown in Table 1). We assumed statistical significance ($P < 0.05$) if the 95% confidence interval (CI) did not cross 0. Furthermore, we compared the proportion of choices toward the larger quantity between tests using a Mann–Whitney *U*-test.

To investigate if and how time of day affected lizards choice performance, we ran a Bayesian GLMM with the choice behavior of each lizard (1—choosing the larger quantity, 0—choosing the smaller quantity, Bernoulli variable) as the response variable and time tested as the only fixed effect. Before analysis, time of day was converted into a numeric variable. The lowest values reflected the earliest time any of the 12 lizards that were tested during the whole experiment, while the largest number reflected the latest time a lizard was tested. We ran a separate model for the number and size discrimination test. The models included a random intercept of animal identity and a random slope of session accounting for repeated measures and autocorrelation between successive choices across days. For all Bayesian models, we confirmed that no autocorrelation (correlation between lags <0.1; Hadfield 2010) occurred, that sufficient mixing (by visually inspecting plots of Markow-Chain-Monte-Carlo chains; Hadfield 2010) took place, and that the Markov chain was run for long enough (Heidelberg and Welch diagnostic tests; Hadfield 2010).

We also investigated if lizards showed a side bias in the side preference test using a binomial test. Additionally, we wanted to know if motivation caused by differences in hunger level led to our test groups difference in performance. We calculated the scaled mass index (SMI; Peig and Green 2009) and compared SMI between groups using a Mann–Whitney *U*-test. All analyses were conducted in R version 4.0.3 (R Core Team 2020).

ETHICAL NOTE

We followed the guidelines laid out by the Association for the Study of Animal Behaviour/Animal Behaviour Society for the treatment of animals in behavioral research and teaching (2018). All test were noninvasive behavioral observations approved by the Macquarie University Animal Ethics Committee (ARA # 2013/031). The collection of lizards was approved by the New South Wales National Parks and Wildlife Service, Office of Environment and Heritage (License # SL101972). Animals were captured by hand or Elliott traps and transported back to Macquarie University, NSW, by car in cloth bags a maximum of 1 week after capture.

Table 1

Estimates (logit scale, bold) and highest posterior density intervals (95% CIs) produced by the Bayesian mixed effects models looking at choice performance (1—larger, 0—smaller; response variable) in the number discrimination (left) and size discrimination (right). The horizontal contrasts (one vs. four, one vs. three, two vs. four, one vs. two, two vs. three, and three vs. four) are the contrasts to which every other contrast (vertical row) was compared (e.g., in the number test, one vs. four was performed less well compared to one vs. three: estimate = −0.248; but better than two vs. four: estimate = 0.118). Values within bold boxes indicate if the larger amount was chosen more often. The remaining boxes represent the difference between contrasts within the same test (number or size). The models included a random effect allowing for a random intercept (animal identity) and a random slope (session) to account for repeated measures and autocorrelation between successive choices across days. The effective sample size for each model was, therefore, six individuals. We assumed statistical significance ($P < 0.05$) if CIs were not crossing 0. Significant results are highlighted in gray

	Number discrimination						Size discrimination					
Estimate												
Lower CI												
Upper CI	1 vs. 4	1 vs. 3	2 vs. 4	1 vs. 2	2 vs. 3	3 vs. 4	1 vs. 4	1 vs. 3	2 vs. 4	1 vs. 2	2 vs. 3	3 vs. 4
1 vs. 4	**1.075** 0.451 1.687	**0.248** −0.543 1.019	**−0.118** −0.889 0.641	**−0.635** −1.387 0.103	**−0.721** −1.451 0.037	**−0.348** −1.100 0.396	**0.455** −0.130 1.017	**−0.060** −0.782 0.659	**−0.222** −0.939 0.498	**−0.3887** −1.143 0.301	**0.058** −0.664 0.778	**−0.555** −1.263 0.168
1 vs. 3	**−0.248** −1.019 0.543	**1.323** 0.687 1.978	**−0.366** −1.135 0.422	**−0.883** −1.655 −0.136	**−0.969** −1.717 −0.209	**−0.596** −1.364 0.159	**0.060** −0.659 0.782	**0.394** −0.188 0.968	**−0.162** −0.885 0.556	**−0.327** −1.069 0.368	**0.118** −0.621 0.834	**−0.494** −1.201 0.235
2 vs. 4	**0.118** −0.641 0.889	**0.366** −0.422 1.135	**0.957** 0.347 1.583	**−0.517** −1.268 0.215	**−0.603** −1.370 0.112	**−0.230** −0.974 0.514	**0.222** −0.498 0.939	**0.162** −0.556 0.885	**0.232** −0.355 0.793	**−0.165** −0.876 0.551	**0.280** −0.431 1.018	**−0.333** −1.040 0.390
1 vs. 2	**0.635** −0.103 1.387	**0.883** 0.136 1.655	**0.517** −0.215 1.268	**0.440** −0.163 1.029	**−0.086** −0.802 0.646	**0.287** −0.447 1.010	**0.387** −0.301 1.143	**0.327** −0.368 1.069	**0.165** −0.551 0.876	**0.067** −0.509 0.639	**0.445** −0.277 1.168	**−0.167** −0.889 0.537
2 vs. 3	**0.721** −0.037 1.451	**0.969** 0.209 1.717	**0.603** −0.112 1.370	**0.086** −0.646 0.802	**0.354** −0.227 0.954	**0.373** −0.360 1.090	**−0.058** −0.778 0.664	**−0.118** −0.834 0.621	**−0.280** −1.018 0.431	**−0.445** −1.168 0.277	**0.512** −0.054 1.101	**−0.613** −1.307 0.133
3 vs. 4	**0.348** −0.396 1.100	**0.596** −0.159 1.364	**0.230** −0.514 0.974	**−0.287** −1.010 0.447	**−0.373** −1.090 0.360	**0.727** 0.106 1.316	**0.555** −0.168 1.263	**0.494** −0.235 1.201	**0.333** −0.390 1.040	**0.167** −0.537 0.889	**0.613** −0.133 1.307	**−0.100** −0.664 0.479

RESULTS

In the number discrimination test, lizards chose the larger number of carrots significantly more often when presented with one versus four (GLMM, median = 0.73, estimate = 1.075, CI = 0.451–1.687), one versus three (GLMM, median = 0.73, estimate = 1.323, CI = 0.687–1.978), two versus four (GLMM, median = 0.70, estimate = 0.957, CI = 0.347–1.583), and three versus four strips (GLMM, median = 0.63, estimate = 0.727, CI = 0.106–1.316; Figure 2 and Tables 1 and 2). Furthermore, lizards chose the larger amount significantly more often when presented with one versus three compared to one versus two (GLMM, 1v3 = 0.883, CI = 0.136–1.655) and two versus three (GLMM, 1v3 = 0.969, CI = 0.209–1.717). In the size discrimination test, lizards did not choose the longer carrot strip more often in any of the combinations, and the combinations did not differ from each other (GLMM, CIs crossing 0 for all estimates; Figure 2 and Tables 1 and 2).

The comparison between the two tests revealed that lizards chose 3 over 1 (Mann–Whitney *U*-test, $W = 33$, estimate = 0.13, CI = 0.07–0.27, $P = 0.018$) and 4 over 3 (Mann–Whitney *U*-test, $W = 31$, estimate = 0.12, CI = 0.00005–0.27, $P = 0.039$) significantly more often in the number discrimination compared to the size discrimination (Figure 2). Overall, Table 2 shows that individual lizards performed better at combinations with a larger distance or smaller ratio in the number test but no such pattern is apparent in the size test. We found no correlation between the probability of choosing the larger quantity and the time of day a lizard was tested in the experiment conducted to investigate spontaneous quantity discrimination based on the number of food items (GLMM, estimate = −0.001, CI = −0.01 to 0.01, $P = 0.868$). In the size discrimination test, however, lizards chose the larger quantity more often later in the day (GLMM, estimate = 0.010, CI = −0.0002 to 0.020, $P = 0.044$). None of the lizards showed a significant preference for left or right during the side preference test (binomial test, $P > 0.05$; Supplementary Table S1); all lizards chose one strip over zero strips 100% of the time and body condition did not differ between test groups (Mann–Whitney *U*-test, $W = 9$, estimate = −19.52, CI = −48.00 to 14.59, $P = 0.180$).

DISCUSSION

Our lizards were able to discriminate between quantities differing in number but failed to do so when quantities differed in size (i.e., length). As a group, lizards chose the larger quantity (i.e., number) above chance in combinations one versus four, one versus three, two versus four, and three versus four carrot strips but not in combinations one versus two and two versus three, although individual lizards did choose the larger number of carrot strips at a high frequency in these combinations. Furthermore, all lizards chose one

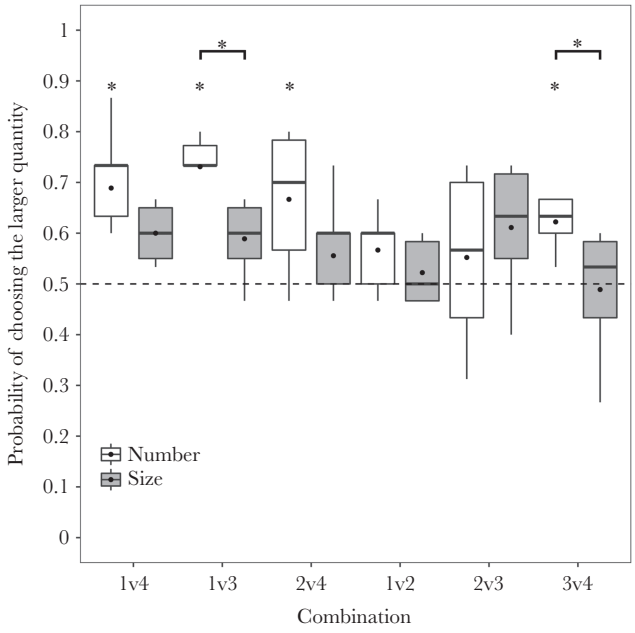

Figure 2
Boxplot summarizing performance (the proportion of choices toward the larger quantity of food) of the six lizards performing a spontaneous discrimination based on food item number (white) and the six lizards performing a discrimination based on food size (gray). The bold line within boxes represents the median, edges of boxes represent the 25% (lower edge) and 75% (upper edge) quartiles, and vertical lines signify the data range excluding outliers. Points represent mean group values. The dashed line represents equal probability of choosing the larger or smaller amount (0.5). Asterisk (*) above single boxes indicate significant difference from chance, asterisk above solid line segments indicate a significant difference between tests. *$P < 0.05$.

over zero carrot strips and none of our lizards showed a side bias during the side preference test. Importantly, we can exclude the possibility that our lizards learnt to discriminate between quantities because they were not allowed to eat the stimuli (carrot pieces). Therefore, our results show spontaneous quantity discrimination in gidgee skinks without prior training.

Our experiment aimed to exploit the tendency of animals to maximize the food amount obtainable at a single point in time (Krebs et al. 1974; Agrillo 2015; Giaquinto 2018). Gidgee skinks only reliably went for more food when quantities differed in number but not when quantities differed in size. In the number test, lizards could rely both on discrete as well as continuous quantity information to choose more food. In the size test, discrete information was missing (we presented only single strips of carrots), and lizards could only rely on continuous information to choose more food. In fishes and primates, processing of discrete numbers is as cognitively demanding as processing continuous quantities (Cantlon and Brannon 2007; Agrillo et al. 2011), but fishes are more proficient when both kinds of information are presented simultaneously (Agrillo et al. 2011). The presence of both continuous and discrete quantity information in the number test might have been an advantage, resulting in above chance performance in those six lizards that had both kinds of information available to make a choice. These results also contrast with other previous findings: dogs (*Canis lupus familiaris*), for example, preferentially use food amount over number (Miletto Petrazzini and Wynne 2016). Guppies (*Poecilia reticulata*) also chose the larger quantity of food

Table 2
Individual performance in percentage of trials in which the larger quantity was chosen for each single lizards tested on each of the six combinations (smallest to largest ratio—top to bottom) presented in the number as well as the size discrimination test. Numbers in brackets indicate the number of choices toward the larger quantity out of all trials tested in the respective combination.

	Number test						Size test					
Combination	ID14	ID15	ID19	ID4	ID5	ID9	ID10	ID11	ID16	ID6	ID7	ID8
1 vs. 4	60% (9/15)	73% (11/15)	47% (7/15)	73% (11/15)	73% (11/15)	87% (13/15)	67% (10/15)	53% (8/15)	53% (8/15)	60% (9/15)	67% (10/15)	60% (9/15)
1 vs. 3	79% (11/14)	73% (11/15)	73% (11/15)	73% (11/15)	60% (9/15)	80% (12/15)	67% (10/15)	60% (9/15)	60% (9/15)	47% (7/15)	53% (8/15)	67% (10/15)
2 vs. 4	47% (7/15)	67% (10/15)	80% (12/15)	80% (12/15)	53% (8/15)	73% (11/15)	73% (11/15)	47% (7/15)	60% (9/15)	60% (9/15)	33% (5/15)	60% (9/15)
1 vs. 2	47% (7/15)	60% (9/15)	60% (9/15)	60% (9/15)	47% (7/15)	67% (10/15)	47% (7/15)	47% (7/15)	60% (9/15)	53% (8/15)	80% (12/15)	27% (4/15)
2 vs. 3	31% (5/16)	60% (9/15)	73% (11/15)	53% (8/15)	40% (6/15)	73% (11/15)	67% (10/15)	40% (6/15)	73% (11/15)	53% (8/15)	73% (11/15)	60% (9/15)
3 vs. 4	60% (9/15)	67% (10/15)	67% (10/15)	67% (10/15)	60% (9/15)	53% (8/15)	53% (8/15)	60% (9/15)	53% (8/15)	27% (4/15)	40% (6/15)	60% (9/15)

even if it was associated with the smaller set of food items (Lucon-Xiccato et al. 2015) and Italian wall lizards (*Podarcis sicula*) also performed better when items differed in size compared to number (Miletto Petrazzini et al. 2017). Apart from differences in information content, the number test also presented lizards with more food (larger amount in each dish) than the size test. This discrepancy could have affected lizards' motivation to obtain food in general and could, therefore, be partly responsible for the difference in performance across test groups. Gidgee skinks are, however, highly motivated to retrieve carrots. They prefer carrots to other vegetables and will eat them even after being fed their regular diet. We, therefore, believe that motivation was unlikely to have had any influence. The test groups also did not differ in body condition, suggesting that differences in motivation based on hunger level did not cause our result. We found, however, that lizards tested on the size discrimination chose the larger quantity of food more often later in the day. This result might still point toward an effect of motivation to retrieve the larger food quantity. In future tests, the size of the carrot strips used in the number experiment could be adjusted to occupy the same area as the pieces used in the size test, which would control for differences in continuous quantities across tests. It is also possible that discriminating quantities relying on the length of items is especially challenging for gidgee skinks. These lizards eat mainly fruit and other plant material (Duffield and Bull 1998) and it would be interesting to see how these lizards discriminate between large and small clusters of food items. Item size might be a less relevant stimulus when food plants produce large numbers of very small fruits, such as weeds and shrubs. Importantly, all food items presented in this study were well within the size range of food items consumed by our lizards during regular feeding.

We did not control for continuous quantities correlating with discrete numbers in our number discrimination test. Gidgee skink quantity discrimination ability might, therefore, not necessarily involve any true numerical abilities. However, our results allow some inferences. First, by using carrot strips, we can exclude movement as a cue, which might bias test subjects to the larger quantity. For example, prey movement has been shown to affect choice in mosquito fish (Agrillo et al. 2008) and salamanders (Krusche et al. 2010). Second, we also controlled for food odor by hiding carrot strips within the enclosure during trials. Lizards were unlikely to be using food chemicals alone because, otherwise, they would have chosen the larger quantity in both experiments. If we compare the two experiments, we might infer that food weight, volume, and area occupied were not solely used to discriminate quantities. In both tests, four (strips or centimeters) was always twice as much (weight, volume, and area) as two and four times as much as one. Further tests are, however, necessary to elucidate what information gidgee skinks rely on when making quantity judgments. For example, the length of carrot strips could be adjusted to equalize continuous quantities between the different combinations (e.g., four 5-mm strips vs one 20-mm strip) in the number test. If lizards fail discriminations in which the length of the carrot strips were made equal it would show stronger reliance on continuous quantities rather than discrete information.

The results from the number test suggest that our lizards relied on a parallel individuation system (i.e., object file system) to pick the larger of two food quantities in the small number range rather than the approximate number system. For one thing, our lizards' performance only weakly conformed to Weber's law. Their accuracy was highest discriminating between one versus four (ratio of 0.25) items and decreased from one versus three (ratio 0.33) to two versus four (ratio 0.5) and two versus three (ratio 0.67). The difference was, however, only significant between the combinations one versus three and one versus two, as well as one versus three and two versus three items. Our lizards failed at discriminating between one versus two with a ratio of 0.5 and a distance of 1, while they were able to discriminate between two versus four, also with a ratio of 0.5 but a distance of 2. Under the parallel individuation system, quantities of similar ratio are of similar discriminability (Agrillo 2015). Our lizards' performance also lacked a distance effect; performance was similar when item numbers differed by three (one vs. four) and two (one vs. three and two vs. four) and lizards even discriminated one out of three combinations with a distance of one (three vs. four). The latter result is important because combination three versus four represents the largest ratio (with a distance of one) and should have been the most difficult to discriminate if lizards relied on the approximate number system. Other species that are capable of this discrimination are, for example, rhesus macaques (Hauser et al. 2000), Asian elephants (Irie-Sugimoto et al. 2009), North Island robins (Hunt et al. 2008), Hermann's tortoises (Gazzola et al. 2018), and honey bees (Howard et al. 2020). Conversely, human infants (Feigenson et al. 2002), dogs (Ward and Smuts 2007; Miletto Petrazzini and Wynne 2016), African gray parrots (Aïn et al. 2009), guppies (Lucon-Xiccato et al. 2015), Italian wall lizards (Miletto Petrazzini et al. 2017), red-backed salamanders (Uller et al. 2003), and fire-bellied toads (Stancher et al. 2015) fail at discriminating three versus four discrete food items. However, comparisons of species performance across studies must be done with caution because methodological differences, as well as differences in motivation or background of subjects, amongst others, can factor into such differences (Shettleworth 2010). Interestingly, our lizards showed no significant difference in performance on the discrimination between three versus four items and one versus four (the easiest based on the approximate number system), further indicating that they used the parallel individuation system. To confirm our conclusions, however, lizards should be tested using larger quantities (more than four items) in future experiments because only large quantities are sufficiently variable to cause a detectable ratio effect (Nieder 2020), making it possible to establish if lizards do use the approximate number system for some quantity discriminations. Furthermore, additional tests investigating the point at which discriminations of small quantities exceed the maximum trackable item number would be informative. Rhesus monkeys (*Macaca mulatta*), for example, spontaneously choose the larger amount when discriminating between one versus two, two versus three, three versus four, and three versus five apple slices but not between four versus five, four versus six, four versus eight, or even three versus eight, indicating that the maximum number of apple slices they can successfully track is about three or four (Hauser et al. 2000).

Our results support the use of the parallel individuation system (rather than the approximate number system) in the small number range, although we are unable to provide data on gidgee skinks performance in the large number range. Agrillo et al. (2012) compared the spontaneous nonverbal quantity judgment of small and larger numbers between fishes and humans and found that, in the large number range, performance followed a ratio effect, while it did not in the small number range. Discriminating between one versus four was as easy as discriminating three versus four, demonstrating the use of two separate systems in both humans and fishes. Similarly, some studies in humans and nonhuman primates support the existence of two separate systems (Cantlon and Brannon 2006;

vanMarle and Wynn 2009), while others report the use of the approximate number system also in the small number range (Cordes et al. 2001; Beran et al. 2006). Such contradicting findings fuel a continuing discussion about the existence of either one numerical processing system (the approximate number system) that is used to judge quantities of any size or two numerical systems. In the latter case, one system would apply to small quantities (parallel individuation system), while the other would apply to judging large quantities (the approximate number system). The study of more basal vertebrates (e.g., amphibians and reptiles) can give some indication of the evolutionary basis of the two cognitive processing systems. The parallel individuation system might be a more recent evolutionary invention to cope with the need to be more precise when judging small quantities. Therefore, we would expect to only find it in species more closely related to humans and nonhuman primates and not in species that are more distantly related. Conversely, the parallel individuation system might have been present in an ancestor but has been lost in some lineages over evolutionary time (Agrillo 2015). Importantly, species might also differ in when they use and how strongly they rely on the different processing systems depending on how selective pressures act on these cognitive traits. With so few data available in reptiles, however, we are unable to make any distinct conclusions regarding the evolutionary history of the processing systems or how selection might have acted on numerical abilities across species, but further studies in a diverse range of reptiles will likely provide more clarity.

In summary, our results show that gidgee skinks likely used a parallel individuation system to judge quantities and spontaneously select for more in small sets of food items (four or fewer items). Further studies including numbers greater than four will give better insight into if and when lizards use the approximate number system. Implementing additional controls for continuous quantities would clarify if lizards rely on item numbers to discriminate between discrete quantities or if they need information from multiple sources (discrete and continuous) to reliably select the larger quantity of food. Our results also markedly differ from those found in other reptiles (Italian wall lizards, Miletto Petrazzini et al. 2017; and Hermann's tortoises Gazzola et al. 2018) in that our gidgee skinks excelled in the number discrimination tests while failing at the size discrimination test as well as the lack of a clear ratio effect in the number test. These marked differences indicate that species might possess specific quantitative skills related to their ecology or sociality. More experiments with a greater phylogenetic range within reptiles are, however, needed to test this hypothesis.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *Behavioral Ecology* online.

FUNDING

This project was funded by the Australian Society of Herpetologists (Student research grant given to B.S.), Macquarie University (Rice Memorial Field Research Award received by B.S.), and the Australian National University (ANU Futures Grant to D.W.A.N.).

We thank Isabel Damas, Bruno Pleno, Faustine Degottex, and Victor Frichot for their help catching skinks and Catarina Vila Pouca and Amanda Ridley, as well as three anonymous reviewers, for their helpful comments on this manuscript.

Data availability: Analyses reported in this article can be reproduced using the data provided by Szabo et al. (2021a).

Handling editor: Amanda Ridley

REFERENCES

Agrillo C. 2015. Numerical and arithmetic abilities in non-primate species. In: Kadosh R, Dowker A, editors. The Oxford handbook of numerical cognition. Oxford: Oxford University Press. p. 214–236.

Agrillo C, Bisazza A. 2017. Understanding the origin of number sense: a review of fish studies. Philos Trans R Soc Lond B Biol Sci. 373:20160511.

Agrillo C, Dadda M, Serena G, Bisazza A. 2008. Do fish count? Spontaneous discrimination of quantity in female mosquitofish. Anim Cogn. 11:495–503.

Agrillo C, Piffer L, Bisazza A. 2011. Number versus continuous quantity in numerosity judgements by fish. Cognition. 119:281–287.

Agrillo C, Piffer L, Bisazza A, Butterworth B. 2012. Evidence for two numerical systems that are similar in humans and guppies. PLoS One. 7:e31923.

Aïn SA, Giret N, Grand M, Kreutzer M, Bovet D. 2009. The discrimination of discrete and continuous amounts in African grey parrots (*Psittacus erithacus*). Anim Cogn. 12:145–154.

Beran MJ, Taglialatela LA, Flemming TM, James FM, Washburn DA. 2006. Nonverbal estimation during numerosity judgements by adults humans. Q J Exp Physiol. 59:2065–2082.

Cantlon JF, Brannon EM. 2006. Shared system for ordering small and large numbers in monkeys and humans. Psychol Sci. 17:401–406.

Cantlon JF, Brannon EM. 2007. How much does number matter to a monkey (*Macaca mulatta*)? J Exp Psychol Anim Behav Process. 33:32–41.

Chapple DG. 2003. Ecology, life-history, and behavior in the Australian scincid genus Egernia, with comments on the evolution of complex sociality in lizards. Herpetol Monogr. 17:145–180.

Cogger HG. 2014. Reptiles & amphibians of Australia. 7th ed. Collingwood (Australia): CSIRO Publishing.

Cordes S, Gelman R, Gallistel CR, Whalen J. 2001. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. Psychon Bull Rev. 8:698–707.

Duffield GA, Bull CM. 1998. Seasonal and ontogenetic changes in the diet of the Australian skink *Egernia stokesii*. Herpetologica. 54:414–419.

Feigenson L, Carey S, Hauser M. 2002. The representations underlying infants' choice of more: object files versus analog magnitudes. Psychol Sci. 13:150–156.

Feigenson L, Dehaene S, Spelke E. 2004. Core systems of number. Trends Cogn Sci. 8:307–314.

Gardner MG, Bull CM, Cooper SJ, Duffield GA. 2001. Genetic evidence for a family structure in stable social aggregations of the Australian lizard *Egernia stokesii*. Mol Ecol. 10:175–183.

Gazzola A, Vallortigara G, Pellitteri-Rosa D. 2018. Continuous and discrete quantity discrimination in tortoises. Biol Lett. 14:20180649.

Giaquinto M. 2018. Cognitive access to numbers: the philosophical significance of empirical findings about basic number abilities. Philos Trans R Soc Lond B Biol Sci. 373:20160520.

Guidelines for the treatment of animals in behavioural research and teaching. 2018. Anim Behav. 135:I– X doi:10.1016/j.anbehav.2017.10.001.

Hadfield JD. 2010. MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. J Stat Softw. 33:1–22.

Harper DGC. 1982. Competitive foraging in mallards: "ideal free" ducks. Anim Behav. 30:575–584.

Hauser MD, Carey S, Hauser LB. 2000. Spontaneous number representation in semi-free-ranging rhesus monkeys. Proc Biol Sci. 267:829–833.

Howard SR, Schramme J, Garcia JE, Ng L, Avargues-Weber A, Greentree AD, Dyer AG. 2020. Spontaneous quantity discrimination of artificial flowers by foraging honeybees. J Exp Biol. 223:jeb223610

Hunt S, Low J, Burns KC. 2008. Adaptive numerical competency in a food-hoarding songbird. Proc Biol Sci. 275:2373–2379.

Hyde DC. 2011. Two systems of non-symbolic numerical cognition. Front Hum Neurosci. 5:150.

Irie-Sugimoto N, Kobayashi T, Sato T, Hasegawa T. 2009. Relative quantity judgment by Asian elephants (*Elephas maximus*). Anim Cogn. 12:193–199.

Kennedy M, Gray RD. 1993. Can ecological theory predict the distribution of foraging animals? A critical analysis of experiments on the ideal free distribution. Oikos. 68:158–166.

Krebs JR, Ryan JC, Charnov EL. 1974. Hunting by expectation or optimal foraging? A study of patch use by chickadees. Anim Behav. 22:953–964.
Krusche P, Uller C, Dicke U. 2010. Quantity discrimination in salamanders. J Exp Biol. 213:1822–1828.
Langkilde T, Shine R. 2006. How much stress do researchers inflict on their study animals? A case study using a scincid lizard, *Eulamprus heatwolei*. J Exp Biol. 209:1035–1043.
Lucon-Xiccato T, Miletto Petrazzini ME, Agrillo C, Bisazza A. 2015. Guppies discriminate between two quantities of food items but prioritize item size over total amount. Anim Behav. 107:183–191.
McComb K, Packer C, Pusey A. 1994. Roaring and numerical assessment in the contests between groups of female lions, *Panther leo*. Anim Behav. 47:379–387.
Miletto Petrazzini ME, Fraccaroli I, Gariboldi F, Agrillo C, Bisazza A, Bertolucci C, Foa A. 2017. Quantitative abilities in a reptile (*Podarcis sicula*). Biol Lett. 13:20160899.
Miletto Petrazzini ME, Wynne CD. 2016. What counts for dogs (*Canis lupus familiaris*) in a quantity discrimination task? Behav Processes. 122:90–97.
Nieder A. 2020. The adaptive value of numerical competence. Trends Ecol Evol. 35:605–617.
Peig J, Green AJ. 2009. New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. Oikos. 118:1883–1891.
R Core Team. 2020. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Shettleworth SJ. 2010. Numerical competence. In: Shettleworth SJ, editor. Cognition, evolution, and behaviour. New York: Oxford University Press. p. 340–370.
Stancher G, Rugani R, Regolin L, Vallortigara G. 2015. Numerical discrimination by frogs (*Bombina orientalis*). Anim Cogn. 18:219–229.
Szabo B, Hoefer S, Whiting MJ. 2020. Are lizards capable of inhibitory control? Performance on a semi-transparent version of the cylinder task in five species of Australian skinks. Behav Ecol Sociobiol. 74:118.
Szabo B, Noble DWA, McCloghry KJ, Monteiro MES, Whiting MJ. 2021a. Data from: spontaneous quantity discrimination in a family-living lizard. Behav Ecol. doi:10.5061/dryad.h70rxwdht.
Szabo B, Noble DWA, Whiting MJ. 2021b. Learning simple and compound stimuli in a social lizard (*Egernia stokesii*). J Comp Psychol. doi:10.1037/com0000260.
Tanner CJ. 2006. Numerical assessment affects aggression and competitive ability: a team-fighting strategy for the ant *Formica xerophila*. Proc Biol Sci. 273:2737–2742.
Uller C, Jaeger R, Guidry G, Martin C. 2003. Salamanders (*Plethodon cinereus*) go for more: rudiments of number in an amphibian. Anim Cogn. 6:105–112.
vanMarle K, Wynn K. 2009. Infants' auditory enumeration: evidence for analog magnitudes in the small number range. Cognition. 111:302–316.
Ward C, Smuts BB. 2007. Quantity-based judgments in the domestic dog (*Canis lupus familiaris*). Anim Cogn. 10:71–80.
Wilson ML, Hauser MD, Wrangham RW. 2001. Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? Anim Behav. 61:1203–1216.

Spontaneous quantity discrimination in a family-living lizard

Birgit Szabo*, Daniel W. A. Noble, Kaitlin J. McCloghry, Marco E. S. Monteiro and Martin J. Whiting

Behavioral Ecology

*Correspond to: Birgit Szabo, PhD, Division of Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland; email: birgit.szabo@gmx.at, ORCID: 0000-0002-3226-8621

Supplementary statistical analyses

We investigated repeatability across combinations (1 versus 4, 1 versus 3, 2 versus 4, 1 versus 2, 2 versus 3 and 3 versus 4) with all data pooled from both tests and separated into the number and size discrimination test. Furthermore, we calculated repeatability within combinations in both tests. Across combinations we calculated adjustment repeatability for pooled data and agreement repeatability as well as adjustment repeatability for data separated into the two tests. We calculated agreement repeatability from the intercept model using choice (1 – choosing the larger quantity, 0 – choosing the smaller quantity) as the response variable, no fixed effects and session nested in animal identity as the random effect. From the posterior of this intercept model we extracted the between-individual variance (σ_α^2), the residual variance (σ_ℓ^2) and the link scale intercept (β_0) to calculate repeatability on the original proportion scale (R_{propA}) following Nakagawa and Schielzeth (2010). Because MCMCglmm (Hadfield, 2010) uses additive overdispersion GLMMs, we used the equations provided for such models by Nakagawa and Schielzeth (2010) to calculate repeatability:

Inverse logit transformation $$P = \frac{\exp(\beta_0)}{1+\exp(\beta_0)}$$

Repeatability on the original scale $$R_{propA} = \frac{\sigma_\alpha^2 * P^2/(1+\exp(\beta_0))^2}{(\sigma_\alpha^2+\sigma_\ell^2) * P^2/(1+\exp(\beta_0))^2 + P*(1-P)}$$

We calculated adjustment repeatability from models with choice (1 – choosing the larger quantity, 0 – choosing the smaller quantity) as the response variable, combination as the only

fixed effect and session nested in animal identity as the random effect. Again, we extracted the between-individual variance, the residual variance and the link scale intercept to calculate repeatability on the original scale using the same equations.

To be able to look at individual repeatability within combinations we ran one model each on a subset of the whole data set which only included the data from a single combination (e.g. 1 versus 4 only). We used choice (1 – choosing the larger quantity, 0 – choosing the smaller quantity) as the response variable, session as the only fixed effect and session nested in animal identity as the random effect. Therefore, we ran six models, one for each combination for both tests (12 models total), and calculated adjustment repeatability following the same procedure as described above.

Supplementary results

Across combinations, repeatability was low when data was pooled (Adjustment repeatability: mean = 0.010, median = 0.004, mode = 0.001, $CI_{low,\ HPD}$ = 0.00003, $CI_{up,\ HPD}$ = 0.042) and when repeatability was calculated separately for both the number (Agreement repeatability: mean = 0.016, median = 0.005, mode = 0.001, $CI_{low,\ HPD}$ = 0.00002, $CI_{up,\ HPD}$ = 0.071; Adjustment repeatability: mean = 0.015, median = 0.004, mode = 0.001, $CI_{low,\ HPD}$ = 0.00003, $CI_{up,\ HPD}$ = 0.062) and the size test (Agreement repeatability: mean = 0.018, median = 0.005, mode = 0.001, $CI_{low,\ HPD}$ = 0.00004, $CI_{up,\ HPD}$ = 0.079; Adjustment repeatability: mean = 0.017, median = 0.005, mode = 0.001, $CI_{low,\ HPD}$ = 0.00002, $CI_{up,\ HPD}$ = 0.076). Within combinations individual repeatability was low but larger than across combination in both the number as well as the size test (Table S2).

Table S1. Summary table including the identity of each animal tested (ID), their age class, origin, body length (SVL – snout-vent-length) and weight. The table also includes information on which experiment an animal participated in (number or size), how many times it chose the right food dish (out of 14 trials) during the side preference test and the results of the binomial test (*p*-values) analyzing choices to the right side during the side preference test. * Animal lost motivation during the experiment only completing part of the test trials; it was replaced with another individual (ID 4).

Experiment	ID	Age class	Origin	SVL (mm)	Weight (g)	Preference test (right choices)	*p*-value (Binomial test)
Number	2*	Adult	Wild	205	288	11	-
Number	4	Adult	Wild	165	153	10	0.180
Number	5	Adult	Wild	190	179	6	0.791
Number	9	Adult	Wild	210	222	9	0.424
Number	14	Adult	Wild	185	154	6	0.791
Number	15	Adult	Wild	195	173	5	0.424
Number	19	Adult	Wild	193	199	10	0.180
Size	6	Adult	Wild	183	192	4	0.180
Size	7	Adult	Wild	175	160	8	0.791
Size	8	Adult	Wild	175	197	5	0.424
Size	10	Adult	Wild	164	114	5	0.424
Size	11	Adult	Wild	180	145	5	0.424
Size	16	Adult	Wild	157	117	7	1.000

Table S2. Summary table presenting individual adjustment repeatability within combinations for both the number as well as the size discrimination test.

Combination	Mean	Median	Mode	$CI_{low, HPD}$	$CI_{up, HPD}$
Number discrimination test					
1 versus 4	0.067	0.016	0.002	0.00003	0.324
1 versus 3	0.061	0.011	0.0006	0.000009	0.307
2 versus 4	0.048	0.010	0.0009	0.0000000003	0.236
1 versus 2	0.343	0.327	0.002	0.00002	0.769
2 versus 3	0.083	0.022	0.002	0.00004	0.385
3 versus 4	0.058	0.013	0.0006	0.00003	0.283
Size discrimination test					
1 versus 4	0.078	0.018	0.002	0.00003	0.376
1 versus 3	0.185	0.091	0.003	0.00004	0.651
2 versus 4	0.053	0.013	0.0006	0.00004	0.256
1 versus 2	0.085	0.022	0.002	0.00002	0.395
2 versus 3	0.061	0.015	0.0004	0.00003	0.293
3 versus 4	0.135	0.045	0.001	0.000008	0.547

References

Hadfield JD. 2010. MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package. J Stat Softw. 33:1-22.

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol Rev Camb Philos Soc. 85:935-956. doi:10.1111/j.1469-185X.2010.00141.x