Lizards lack speed-accuracy trade-offs in a quantitative foraging task when unable to sample the reward

Lisa M. Greis [a], Eva Ringler [b,1], Martin J. Whiting [c,2], Birgit Szabo [b,3,*]

[a] *Faculty of Life Sciences, University of Vienna, Vienna, Austria*
[b] *Division of Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland*
[c] *School of Natural Sciences, Macquarie University, Sydney, Australia*

ARTICLE INFO

ABSTRACT

To make decisions, animals gather information from the environment in order to avoid costs (e.g., reduced survival) and increase benefits (e.g., foraging success). When time is limited or information is insufficient, most animals face a speed-accuracy trade-off (SAT) – they have to balance the benefits of making quick decisions against the costs of inaccurate decisions. Here, we investigated the relationship between decision accuracy and decision speed in gidgee skinks (*Egernia stokesii*) performing a food-based spontaneous quantity discrimination task. Rather than a SAT we found a speed-accuracy alignment; lizards made decisions that were fast and accurate, rather than inaccurate. Furthermore, we found only within-, but no between-individual differences in decision making indicating behavioural plasticity in the absence of individual decision styles. Finally, latency to choice was highly repeatable, more so than choice accuracy. Previous work has shown that learning, the costs of a bad decision and task difficulty frequently result in SATs. The lack of a SAT in our lizards might be a direct consequence of our simple testing methodology which prevented learning by not allowing lizards to consume the chosen quantity. To fully understand how SATs develop, different methodologies that control the costs and benefits of decisions should be compared.

1. Introduction

To appropriately respond to a given situation and avoid the costs of a bad decision, animals need to constantly gather information about their environment (de Froment et al., 2014). The more time an individual invests in gathering information the more likely it is that the resulting decision is accurate. However, since time is limited, the time spent performing essential tasks such as resting, foraging, scanning for predators or engaging in social interactions needs to be carefully balanced. Therefore, spending more time gathering information in one context (e. g. foraging) takes away time from other activities (e.g. vigilance; Verdolin, 2006). When time becomes such a limiting factor, individuals face a speed-accuracy trade-off (SAT). On the one hand, SATs are more likely when problems are complex and more information is needed, increasing the time required to maintain accuracy. On the other hand, errors increase when time is limited and information gathering is constrained (Chittka et al., 2009; Franks et al., 2003).

Studies investigating cognition largely focus on the accuracy of a decision without considering the time needed to make a choice (Chittka et al., 2009; Jones et al., 2020). Often, accuracy is used to determine the limit of an ability (Chittka et al., 2009) and ignoring the time needed to make a decision can lead to false interpretations, especially when the study design places a limit on trial length. SAT is considered a primary property of cognitive performance (Heitz, 2014; Jones et al., 2020) and it occurs across many contexts including foraging and predator detection (Chittka et al., 2009; Dyer and Chittka, 2004; Hemingway et al., 2019; Ings and Chittka, 2008), nest site selection (Marshall et al., 2009) and when choosing a mate (Hemingway et al., 2019). Not all studies have, however, demonstrated the expected trade-off between decision speed and accuracy. When presented with an acoustic playback of a territory intruder (via a loudspeaker) or a visible intruder using a frog model, a male poison frog's (*Allobates femoralis*) decision whether or not to attack
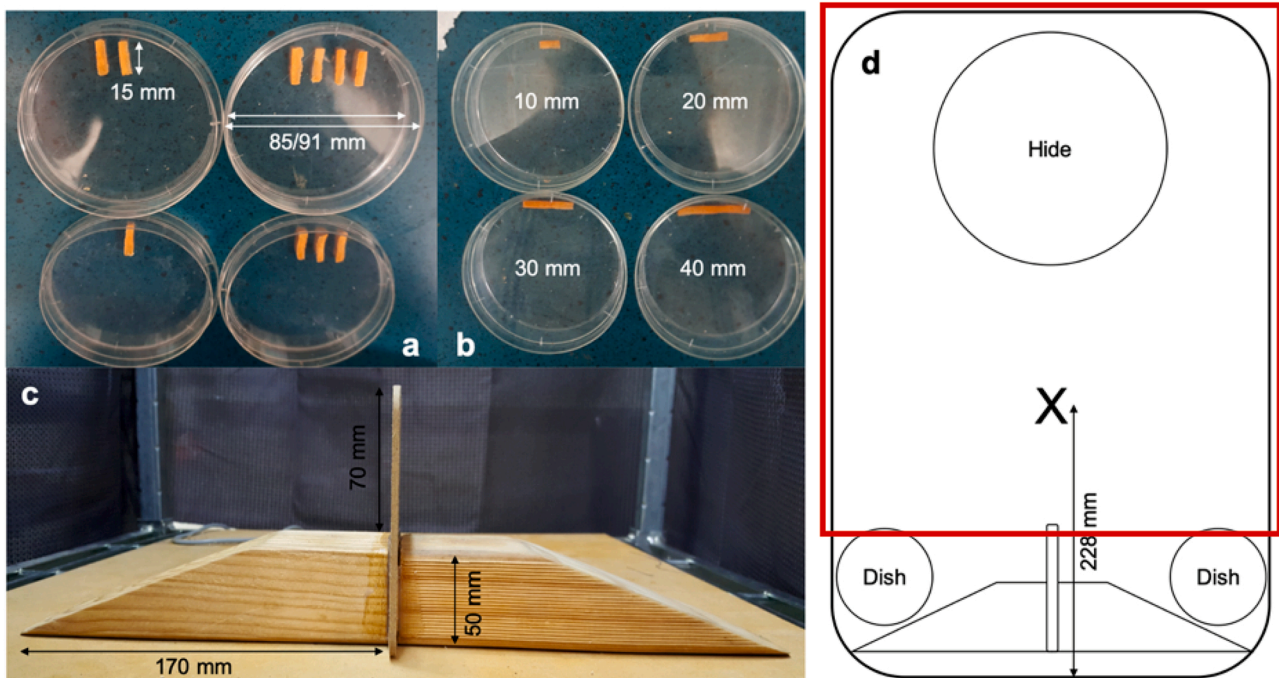
Fig. 1. (a) Number discrimination test setup for petri dishes. The gap between strips was about 5 mm. (b) Size discrimination test setup for petri dishes. (c) Wooden divider (height 93 mm) made up of a wooden coaster glued between two wooden ramps. (d) Schematic of the experimental setup: The skink was put under the hide at the far end while the wooden divider was placed at the opposite side. Petri dishes were placed to the left and right of the divider. The lizard was lured to the starting position (X) with a strip of carrot presented with pincers to make sure the animal started from the same distance from both dishes.
The red frame marks the area visible in the videos provided for the speed-accuracy trade-off analysis (modified from Szabo et al., 2021).

(accuracy) does not relate to their reaction time or approach speed (Sonnleitner et al., 2020).

Individuals can vary greatly in their cognitive ability (Boogert et al., 2018) and this may translate into distinct decision styles. Response patterns have been classified as either reactive or proactive coping strategies, which are consistent over time and in different situations (Koolhaas et al., 1999). Some individuals are more prone to making rapid decisions that are more likely to produce errors (proactive) while others are slower but more accurate (reactive) (Chittka et al., 2009). Therefore, there is a clear difference between the two strategies in the amount of external stimuli necessary to influence behaviour (Benus et al., 1987; van Oortmerssen et al., 1990; Koolhaas et al., 1999). Importantly, such individual differences in decision style are usually only detectable when making an accurate decision is difficult (Chittka et al., 2009). For example, archerfish (*Toxotes chatareus*) trained to shoot at artificial targets that provided different amounts of food depending on their colour learnt to shoot at the more highly rewarded targets but at a cost of increased decision latency. Consistently fast individuals initially made more errors than slow individuals and this difference was only detectable in the more difficult task providing three targets (Jones et al., 2020). Similarly, great tits (*Parus major*) exhibit between-individual differences in SAT but not within individuals. Slower deciding tits made more accurate decisions than fast deciding birds, but individual decision speed did not relate to accuracy (Moiron et al., 2016). Bumblebees also express inter-individual variation in SAT (Chittka et al., 2003). Such individual differences are even evident in slime moulds (*Physarum polycephalum*) where faster choosing slime moulds are more likely to make errors (Latty and Beekman, 2011).

We recently published a study that investigated spontaneous quantity discrimination in a foraging context in the Australian gidgee skink (*Egernia stokesii*; Szabo et al., 2021), an omnivorous, gregarious, social lizard species (Cogger, 2014; Chapple, 2003). Previously, we found that gidgee skinks were more accurate at choosing the larger amount of individual carrot pieces compared to single carrot pieces differing in size.

Lizards also showed reduced accuracy when presented with more similar quantities. These differences in accuracy suggest that information is processed about which quantity to choose, potentially translating into SATs. This is especially likely to be the case when choosing between similar quantities or when less information is available (discrete and continuous information in the number discrimination versus continuous information only in the size discrimination).

In the present study, we scored decision latency from the videos produced during our previous study (Szabo et al., 2021) to investigate if gidgee skinks show a SAT in a spontaneous quantity discrimination task. Based on previous findings in other species (Chittka et al., 2009; Dyer and Chittka, 2004; Pachella, 1974) we expected that choice latency would correlate positively with choice accuracy and predicted that lizards would show longer latencies when making a correct decision (choosing the larger quantity). We also expected that latencies would increase with the difficulty of the decision (smaller distance between quantities). Lastly, we also tested for individual differences in the expression of the SAT, possibly revealing individual decision styles (slow-accurate versus fast-inaccurate), especially when choosing between similar quantities.

2. Methods

Here, we analysed choice latency in gidgee skinks (*Egernia stokesii*) performing a spontaneous quantity discrimination based on food. The methods have been described in detail elsewhere (Szabo et al., 2021) but are given in brief below.

2.1. Animals, captive conditions and husbandry

Twelve adult gidgee skins of undetermined sex were tested; six in a number-based spontaneous quantity discrimination task and six in a size-based task (analysis showed no difference in body condition between test groups, see Szabo et al., 2021). All skinks were captured

during November 2018 at Fowlers Gap Arid Zone Research Station (−31.086972 S, 141.704836 E), New South Wales, Australia and kept at Macquarie University in a temperature controlled room set to 24 °C and a light: dark cycle of 12 h: 12 h. Each lizard was kept in a plastic tub (683 L x 447 W x 385 H mm) lined with butcher's paper and provided with a plastic hide, a 150 mm long PVC tube and some natural items such as leaves and bark for enrichment purposes. To enable lizards to thermoregulate, the tubs were also equipped with a heat coil under one third of the enclosure, locally raising the temperature to 33 °C (± 2 °C SD). Enclosure temperatures were monitored on an hourly basis using Thermochron iButtons (model DS1921). A bowl of water within the enclosure ensured unlimited access to water for the lizards. Feeding occurred on Monday, Wednesday and Friday with a diet of small cut fruit and vegetables as well as crickets. During tests the lizards were fed carrots (0.065 g ± 0.021 g SD) 6–12 times daily (as rewards in the experiment) Monday to Friday. On Fridays after trials were finished, lizards were fed their regular diet including 2–3 crickets dusted in aristopet Repti-vite and URS Ultimate Calcium in addition to the carrots fed during testing. The butcher's paper was replaced every Friday to allow the lizard's scent to spread for two days before the next day of testing, thereby avoiding the development of scent trails.

2.2. Set-up and testing procedure

Three tests were conducted: one side preference test and two spontaneous quantity discrimination tests (number and size discrimination). Quantity discrimination tests immediately followed the side preference test and were concluded within four weeks. Only the quantity discrimination tests were used for analysis of SATs.

2.2.1. General testing procedure

Lizards were tested within their familiar home enclosure to minimise stress (Langkilde and Shine, 2006) and carried within their enclosure to a curtained-off shelf (testing area) 3 m from the other lizards within the temperature-controlled room before the first trial each test day. The curtain minimized distractions and also hid the researcher during trials. A heat cord underneath a third of the enclosure ensured that lizards were able to thermoregulate during test trials. Next, the lizard was carefully covered with its hide and left undisturbed for 3 min to recover from being moved (in its enclosure) to the trial area. In the meantime, we removed all items from the enclosure except for the hide and we inserted a wooden divider inside the enclosure (Fig. 1c) close to the experimenter. If necessary, after the recovery period, the lizard was moved to the far end of the enclosure (Fig. 1d). After the recovery period, two dishes with stimuli were placed simultaneously to the left and right of the divider and the hide was removed. Freshly prepared carrot strips (a preferred food item) were used as rewards and test stimuli during the trials. To ensure that all lizards started each trial from the same position they were lured to a starting position at the same distance from both food dishes (marked with an X, Fig. 1d) using a carrot strip presented in forceps; this also initiated foraging behaviour. After eating this carrot, the lizard was left to make a choice. If two minutes passed without the animal making a choice, motivation was increased by presenting the lizard with another strip of carrot. After five minutes (300 s), the trial was aborted, the animal placed back under the hide and the next trial prepared. These failed trials were repeated after all other trials had finished ensuring that all individuals performed 90 successful trials.

A choice was scored when the skink first touched a dish or the animal moved into close proximity to one of the dishes and made a snapping motion. After a choice was made, the animal was returned to the far side of the enclosure (by hand or by moving the hide with the lizard) in preparation for the next trial. The animals were not allowed to eat the food within the chosen dish, to avoid learning related to quantity based on the food quantity consumed (i.e., degree of satiation). A camera setup (H.264 Digital Video Recorder, 3-Axis Day & Night Dome Cameras) was installed above the testing area to automatically record the lizards'

behaviour. Individuals were tested in a random order each day.

2.2.2. Test stimuli

Six combinations of quantities were presented to the animals (treatments): 1 versus 4, 1 versus 3, 1 versus 2, 2 versus 4, 2 versus 3 and 3 versus 4 (15 trials each = 90 trials per lizard). All combinations were tested each day (Monday to Friday) but only once. The presented stimuli and the side the larger quantity was presented on were also randomised but counter-balanced (showing the larger quantity no more than twice on the same side). Each test day, four petri dishes with carrots (preferred food) were prepared. The carrots were first grated (equal width) and then cut to size with the help of a ruler. Possible olfactory cues were removed after use by cleaning the dishes with 70% ethanol solution, so choice would not be affected. Additionally, four carrot strips (15 mm long) were placed behind the divider to control for food odour.

2.2.3. Number discrimination test

For the number discrimination trials, carrot strips of the same size (15 mm long, 0.065 g ± 0.021 g SD) were placed in the two dishes. Strips were parallel and 5 mm apart from each other, close to the far edge of the dish (Fig. 1a). Petri dishes were placed at an angle of approximately 45 degrees. This provided the lizards with an unobstructed view of the stimuli from the starting position.

2.2.4. Size discrimination test

For the size discrimination, only one carrot strip of different size was presented, parallel to the edge of the dish (Fig. 1b). Lengths were 10 mm (0.053 g ± 0.01 g SD), 20 mm (0.109 g ± 0.03 g SD), 30 mm (0.151 g ± 0.03 g SD), or 40 mm (0.210 g ± 0.02 g SD) respectively. Petri dishes were also placed at an angle of approximately 45 degrees to provide an unobstructed view.

2.3. Data collection

We analysed the existing video data for the 12 adult gidgee skinks to evaluate if the animals showed a SAT that reflects the difficulty of numerical discrimination by collecting and analysing the time (latency, measured in seconds) it took individual lizards to choose a quantity. Scoring of videos was done blind. All videos were edited prior to analysis so the screen section visible did not include the dishes with the stimuli (Fig. 1d). Latency was measured from the time an animal consumed the motivational carrot until a decision was clearly made, i.e., when a food dish was touched or the animal moved into close proximity to one of the dishes and made a snapping motion (de Froment et al., 2014).

2.4. Statistical analyses

First, we analysed if latency differed between correct (larger quantity) and incorrect (smaller quantity) choices using a Bayesian generalised linear mixed model (MCMCglmm package; Hadfield, 2010) with Gaussian family for the number and size group separately (treatment pooled). We used the log transformed latency as the response variable and accuracy as the fixed effect. Latency was log transformed to better conform to a normal distribution. Additionally, we were interested in if latency changed across the experiment, if lizards showed a change in motivation across trials within a day and if lizards took longer to make a choice when choosing between more similar quantities. Date, trial and treatment were, therefore, also included as fixed effects. To compare treatments we performed a post hoc least square means (LSM) test using the package emmeans (Lenth, 2021).

Next, we looked at the speed-accuracy trade-off. We were interested if a SAT occurred in general (treatment pooled) and, following the approach by Moiron et al. (2016), if individuals differed from each other and across treatments. We used within-subject centering (van de Pol and Wright, 2009) and derived a representation of within-subject variation by subtracting the average latency from each individual latency ($x_{ij} -$

Table 1
Individual first trial choice (quantity chosen) and choice latency (in seconds) for each treatment separated into number and size group. The group average accuracy and latency for the first trial across all individuals are also given. ID – animal identity.

Number group					Size group				
Treatment	ID	Choice	Latency (sec)	Group average	Treatment	ID	Choice	Latency (sec)	Group average
1 versus 4	4	4	8	Accuracy = 0.5	**1 versus 4**	6	4	31	Accuracy = 0.67
	5	4	24	Latency = 83.83 s		7	4	19	Latency = 33.00 s
	9	4	198			8	4	91	
	14	1	219			10	1	10	
	15	1	32			11	4	9	
	19	1	22			16	1	38	
1 versus 3	4	3	7	Accuracy = 0.83	**1 versus 3**	6	1	36	Accuracy = 0.83
	5	3	29	Latency = 33.50 s		7	3	13	Latency = 18.00 s
	9	3	23			8	3	29	
	14	3	81			10	3	11	
	15	1	37			11	3	8	
	19	3	24			16	3	11	
2 versus 4	4	4	10	Accuracy = 0.67	**2 versus 4**	6	4	40	Accuracy = 0.83
	5	2	28	Latency = 26.33 s		7	2	9	Latency = 16.50 s
	9	4	15			8	4	19	
	14	2	48			10	4	12	
	15	4	20			11	4	7	
	19	4	37			16	4	12	
1 versus 2	4	2	13	Accuracy = 0.5	**1 versus 2**	6	2	28	Accuracy = 0.5
	5	2	34	Latency = 35.00 s		7	2	32	Latency = 22.33 s
	9	1	21			8	1	19	
	14	1	93			10	2	10	
	15	1	32			11	1	16	
	19	2	17			16	1	29	Accuracy = 0.67
2 versus 3	4	2	9	Accuracy = 0.5	**2 versus 3**	6	2	30	Latency = 18.17 s
	5	2	28	Latency = 25.17 s		7	3	16	
	9	3	19			8	3	30	
	14	2	39			10	3	14	
	15	3	32			11	3	5	
	19	3	24			16	2	14	
3 versus 4	4	3	8	Accuracy = 0.67	**3 versus 4**	6	3	11	Accuracy = 0.33
	5	4	36	Latency = 39.00 s		7	3	12	Latency = 12.00 s
	9	3	23			8	3	10	
	14	4	105			10	4	12	
	15	4	20			11	4	11	
	19	4	42			16	3	16	

$\overline{x}_j$). Between subject variation is represented by the mean alone ($\overline{x}_j$). We then built two Bayesian models with Binomial error distribution for each test group (number and size). The first model included accuracy as the response variable and both within ($x_{ij} - \overline{x}_j$) and between subject variation ($\overline{x}_j$) as the fixed effects. In the second model we also used accuracy as the response variable but individual latency (x_{ij}) and the mean latency ($\overline{x}_j$) as the fixed effects. While the first model estimates within- and between-subject effects separately, the second model estimates the correlation of accuracy and latency (SAT) and the difference between within- and between-subject variation. Additionally, we were interested if choice accuracy changed across trials (possible effect of motivation) and included trial as a third fixed effect in the second model.

Because we used different treatments by presenting different combinations of food quantities, we also analysed within and between subject effects within each treatment for both test groups separately. We calculated an average latency within each treatment to derive a within and between subject effect for each treatment. We then ran a single Bayesian model with Binomial error distribution for each treatment using the accuracy as the response variable and the within- and between-subject variation as the fixed effects. Finally, we calculated adjusted repeatability accounting for treatment (package rptR, Stoffel et al., 2017) to look at within and between individual consistency in latency measures to compare to the repeatability calculated for choice in our previous analysis (Szabo et al., 2021).

Finally, we investigated how motivation (size of the chosen quantity and the total presented quantity of both stimuli combined) affected our lizards' choice latency. To this end, we ran separate Bayesian GLMMs (one for the number group and one for the size group) with the log

transformed latency as the response variable and included the chosen quantity as well as the total quantity presented in a given trial as the fixed effects. We were forced to build a separate model instead of including these fixed effects in the model looking at how latency differed between correct and incorrect choices, because both fixed effects are derived from treatment and treatment was included as a fixed effect in the first model.

All GLMMs included a random intercept of animal identity to account for repeated measures. We ensured that lags were not correlated (< 0.1; Hadfield, 2010), that the MCMC chain mixed sufficiently (by visually inspecting plots; Hadfield, 2010) and we confirmed that the MCMC chain was run for long enough by using a Heidelberg and Welch diagnostic test (Hadfield, 2010). All analyses were run in R version 4.0.3 (R Core Team, 2021) and we report our results based on Muff et al. (2022): $p > 0.1$ no evidence, $0.1 < p < 0.05$ weak evidence, $0.05 < p < 0.01$ moderate evidence, $0.01 < p < 0.001$ strong evidence, $p < 0.001$ very strong evidence. All raw data generated for this study and the code used for analysis are available on the Open Science Framework (OSF, doi: 10.17605/OSF.IO/WB8JC).

2.5. Ethical note

We followed the guidelines laid out by the Association for the Study of Animal Behaviour/Animal Behaviour Society for the ethical treatment of animals in behavioural research and teaching (ASAB Ethical Committee, ABS Animal Care Committee, 2022). All tests were non-invasive behavioural 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
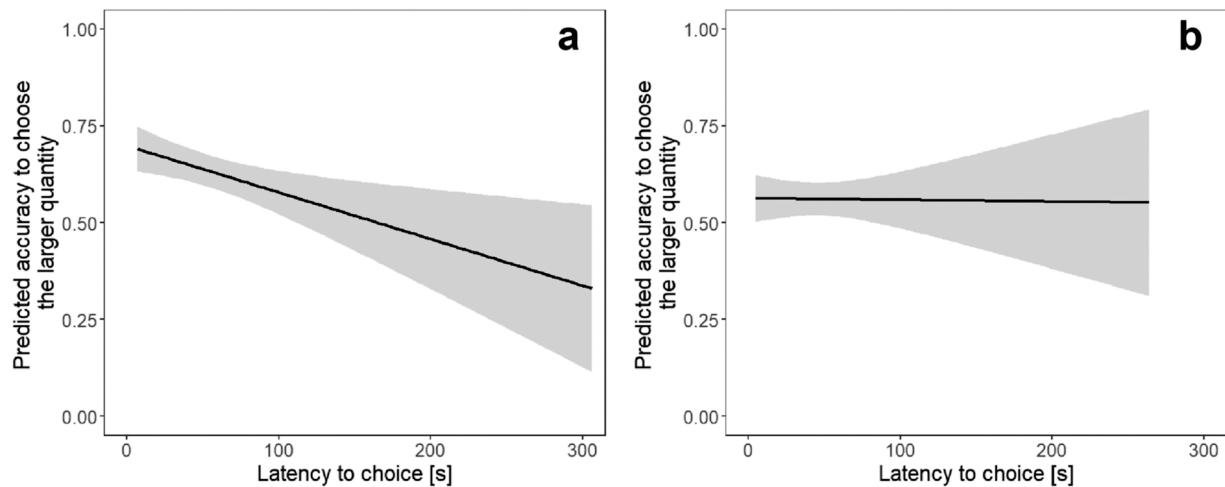
Fig. 2. Relationship between accuracy and latency to choice. (a) Significant negative correlation between the accuracy to choose the larger quantity and the latency to make a choice in the number group (treatment pooled). (b) No correlation between the accuracy to choose the larger quantity and the latency to make a choice in the size group (treatment pooled). The grey area represents the 95% confidence interval.
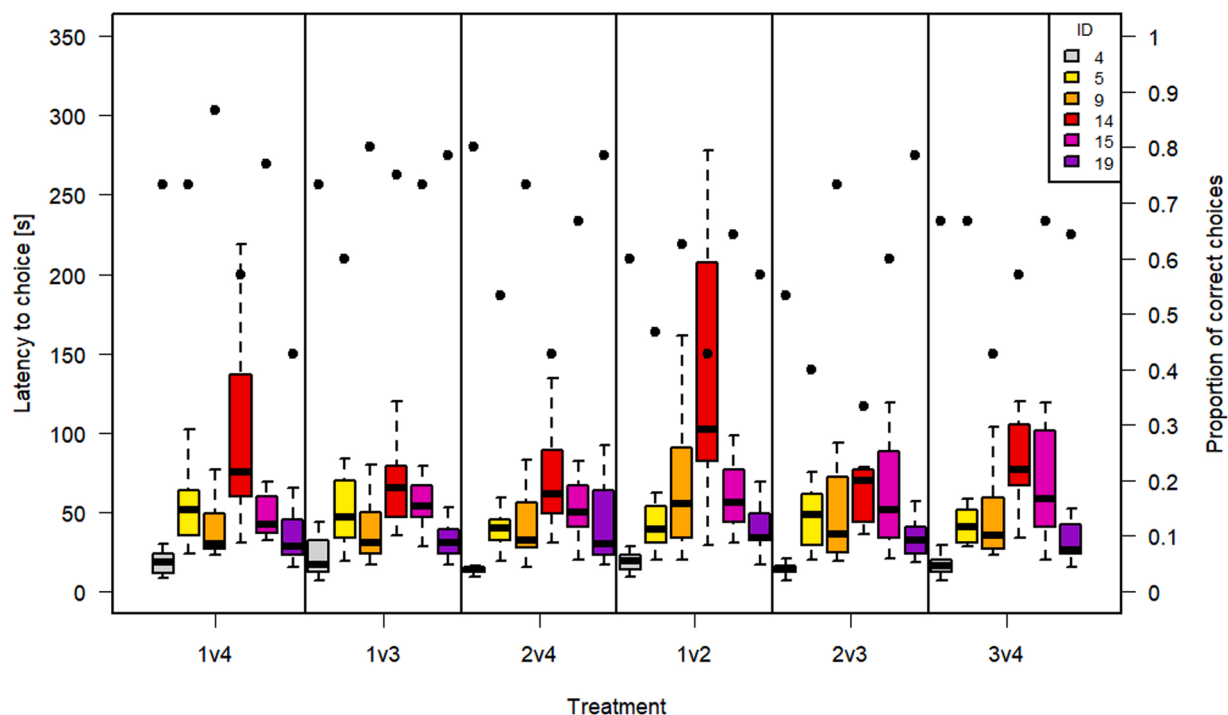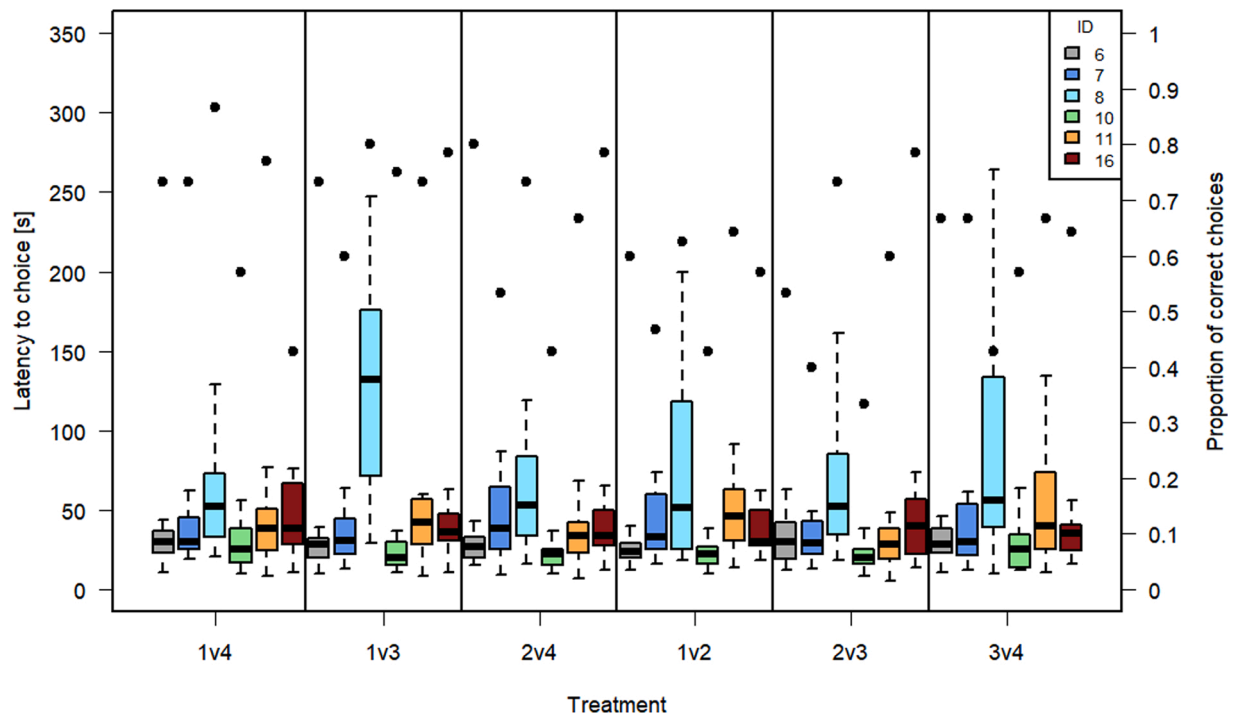

Fig. 3. Box plot of latency to make a choice in seconds for each individual in each treatment (left y-axis) for the *number* group. The dots represent the accuracy for each individual as the proportion of correct choices (right y-axis). The bold line indicates the median, the upper edge of the box represents the upper quartile, the lower edge the lower quartile, the top edge of the whisker the maximum and the bottom edge of the whisker the minimum (outliers are not shown). Each colour represents a different individual.

3. Results

We found weak evidence that number group lizards made faster accurate, compared to inaccurate, choices (GLMM, estimate $= -0.111$, $CI_{low} = -0.211$, $CI_{up} = -0.015$, $p = 0.028$; Appendix Table A1) while latency did not change across test days (GLMM, estimate $= 0.004$, $CI_{low} = -0.005$, $CI_{up} = 0.014$, $p = 0.389$; Appendix Table A1) or across trials

(GLMM, estimate $= -0.011$, $CI_{low} = -0.039$, $CI_{up} = 0.016$, $p = 0.436$; Appendix Table A1). Furthermore, we found weak evidence that lizards chose more slowly in treatment 1 versus 2 compared to 1 versus 4 (GLMM, estimate $= 0.147$, $CI_{low} = -0.006$, $CI_{up} = 0.312$, $p = 0.068$; Appendix Table A1) and every other treatment (LSM, confidence interval not crossing 0; Appendix Table A1). Conversely, we found no evidence that latency differed between correct and incorrect choices in the size group (GLMM, estimate $= -0.021$, $CI_{low} = -0.116$, $CI_{up} = 0.073$, $p = 0.655$; Appendix Table A2), while we found very strong evidence that latency increased over test days (GLMM, estimate $= 0.042$, $CI_{low} = 0.032$, $CI_{up} = 0.053$, $p < 0.00002$; Appendix Table A2) but no correlation with trial (GLMM, estimate $= 0.002$, $CI_{low} = -0.026$, $CI_{up} = 0.031$,

Fig. 4. Box plot of latency to make a choice in seconds for each individual in each treatment (left y-axis) for the *size* group. The dots represent the accuracy for each individual as the proportion of correct choices (right y-axis). The bold line indicates the median, the upper edge of the box represents the upper quartile, the lower edge the lower quartile, the top edge of the whisker the maximum and the bottom edge of the whisker the minimum (outliers are not shown). Each colour represents a different individual.

Table A1

Estimates and test statistics for the Bayesian model looking at the effect of accuracy, date, treatment and trial on latency in the number group only. The response variable was log transformed. Included are also estimates and test statistics for the PostHoc comparison of treatment based on least square means (LSM). Results with a *p*-value below 0.1 are highlighted in bold. We assumed results to be significant if the confidence interval did not cross 0. CI – Confidence interval, n.s. – not significant, s. – significant.

Parameter	Estimate	Lower 95% CI	Upper 95% CI	*p*-value
Intercept	-69.30	-251.9	96.96	0.414
Accuracy 1	**-0.111**	**-0.211**	**-0.015**	**0.028**
Date	0.004	-0.005	0.014	0.389
1 versus 3	-0.041	-0.196	0.129	0.618
2 versus 4	-0.081	-0.251	0.068	0.327
1 versus 2	**0.147**	**-0.006**	**0.312**	**0.068**
2 versus 3	-0.075	-0.232	0.087	0.356
3 versus 4	-0.037	-0.201	0.122	0.657
Trial	-0.011	-0.039	0.016	0.436
PostHoc comparison LSM				
1v4 – 1v3	0.039	-0.116	0.207	n.s.
1v4 – 2v4	0.081	-0.085	0.241	n.s.
1v4 – 1v2	-0.146	-0.305	0.018	n.s.
1v4 – 2v3	0.075	-0.078	0.244	n.s.
1v4 – 3v4	0.035	-0.127	0.201	n.s.
1v3 – 2v4	0.041	-0.114	0.205	n.s.
1v3 – 1v2	**-0.185**	**-0.352**	**-0.033**	**s.**
1v3 – 2v3	0.034	-0.122	0.202	n.s.
1v3 – 3v4	-0.005	-0.167	0.159	n.s.
2v4 – 1v2	**-0.226**	**-0.385**	**-0.068**	**s.**
2v4 – 2v3	-0.008	-0.166	0.154	n.s.
2v4 – 3v4	-0.046	-0.214	0.111	n.s.
1v2 – 2v3	**0.220**	**0.061**	**0.376**	**s.**
1v2 – 3v4	**0.180**	**0.024**	**0.344**	**s.**
2v3 – 3v4	-0.039	-0.201	0.123	n.s.

Table A2

Estimates and test statistics for the Bayesian model looking at the effect of accuracy, date, treatment and trial on latency in the size only. The response variable was log transformed. Included are also estimates and test statistics for the PostHoc comparison of treatment based on least square means (LSM). We assumed results to be significant if the confidence interval did not cross 0. CI – Confidence interval, n.s. – not significant, s. – significant.

Parameter	Estimate	Lower 95% CI	Upper 95% CI	*p*-value
Intercept	**-766.9**	**-954.2**	**-579.8**	**0.00003**
Accuracy 1	-0.022	-0.116	0.072	0.657
Date	**0.042**	**0.032**	**0.053**	**0.00003**
1 versus 3	0.069	-0.098	0.224	0.406
2 versus 4	-0.005	-0.215	0.109	0.541
1 versus 2	-0.009	-0.175	0.152	0.910
2 versus 3	-0.086	-0.249	0.080	0.307
3 versus 4	0.014	-0.149	0.176	0.867
Trial	0.002	-0.026	0.031	0.861
PostHoc comparison LSM				
1v4 – 1v3	-0.069	-0.232	0.090	n.s.
1v4 – 2v4	0.050	-0.115	0.207	n.s.
1v4 – 1v2	0.009	-0.153	0.174	n.s.
1v4 – 2v3	0.085	-0.078	0.251	n.s.
1v4 – 3v4	-0.014	-0.177	0.148	n.s.
1v3 – 2v4	0.119	-0.043	0.279	n.s.
1v3 – 1v2	0.079	-0.085	0.241	n.s.
1v3 – 2v3	0.155	-0.017	0.313	n.s.
1v3 – 3v4	0.056	-0.106	0.218	n.s.
2v4 – 1v2	-0.041	-0.204	0.124	n.s.
2v4 – 2v3	0.036	-0.125	0.204	n.s.
2v4 – 3v4	-0.064	-0.225	0.098	n.s.
1v2 – 2v3	0.076	-0.084	0.241	n.s.
1v2 – 3v4	-0.023	-0.189	0.138	n.s.
2v3 – 3v4	-0.100	-0.265	0.066	n.s.

p = 0.861; Appendix Table A2). Lizards from the size group did not change latency across treatments (GLMM, *p* > 0.1; Appendix Table A2).

When looking at the first trial only, lizards did not take more time to make an accurate decision. This was true for the number as well as the size group. First trial data reflect a similar trend as what we found

Table A3

Estimates and test statistics for the Bayesian models looking at within and between individual SAT variation within each treatment, number group only. Results with a *p*-value below 0.1 are highlighted in bold. CI – Confidence interval.

Parameter	Estimate	Lower 95% CI	Upper 95% CI	*p*-value
1 versus 4				
Intercept	1.111	-1.0002	3.467	0.238
Between individual variation	-0.002	-0.039	0.035	0.886
Within individual variation	0.0002	-0.014	0.0152	0.997
1 versus 3				
Intercept	**2.731**	**-0.166**	**3.714**	**0.070**
Between individual variation	-0.006	-0.023	0.012	0.523
Within individual variation	-0.004	-0.047	0.036	0.829
2 versus 4				
Intercept	**2.313**	**0.696**	**4.070**	**0.011**
Between individual variation	**-0.027**	**-0.054**	**0.001**	**0.051**
Within individual variation	-0.006	-0.019	0.008	0.359
1 versus 2				
Intercept	0.671	-0.826	2.121	0.357
Between individual variation	-0.007	-0.032	0.019	0.572
Within individual variation	-0.0007	-0.011	0.009	0.900
2 versus 3				
Intercept	1.205	-1.126	3.588	0.219
Between individual variation	-0.018	-0.058	0.022	0.286
Within individual variation	**-0.019**	**-0.036**	**-0.003**	**0.010**
3 versus 4				
Intercept	0.750	-0.745	2.307	0.313
Between individual variation	-0.004	-0.030	0.022	0.768
Within individual variation	-0.002	-0.017	0.013	0.733

Table A4

Estimates and test statistics for the Bayesian models looking at within and between individual SAT variation within each treatment, size group only. Results with a *p*-value below 0.1 are highlighted in bold. CI – Confidence interval.

Parameter	Estimate	Lower 95% CI	Upper 95% CI	*p*-value
1 versus 4				
Intercept	0.831	-0.778	2.462	0.302
Between individual variation	-0.008	-0.042	0.027	0.633
Within individual variation	-0.010	-0.033	0.012	0.369
1 versus 3				
Intercept	-0.306	-1.985	1.445	0.723
Between individual variation	0.018	-0.017	0.058	0.322
Within individual variation	-0.009	-0.025	0.007	0.258
2 versus 4				
Intercept	0.499	-1.380	2.315	0.557
Between individual variation	-0.005	-0.043	0.034	0.755
Within individual variation	-0.007	-0.224	0.009	0.397
1 versus 2				
Intercept	1.290	-0.718	3.334	0.166
Between individual variation	-0.027	-0.071	0.016	0.180
Within individual variation	**0.014**	**-0.002**	**0.031**	**0.080**
2 versus 3				
Intercept	0.630	-1.217	2.424	0.452
Between individual variation	-0.001	-0.039	0.038	0.961
Within individual variation	0.005	-0.012	0.023	0.573
3 versus 4				
Intercept	-0.855	-2.566	0.824	0.295
Between individual variation	0.018	-0.018	0.053	0.290
Within individual variation	0.001	-0.013	0.014	0.926

looking at repeated measures: lizards made fast and accurate and slow and inaccurate decisions (Table 1).

In the number group, we found weak evidence for within-subject variation (GLMM, estimate $= -0.005$, $CI_{low} = -0.010$, $CI_{up} = -0.0002$, $p = 0.032$) but found no evidence for between-subject variation (GLMM, estimate $= -0.009$, $CI_{low} = -0.023$, $CI_{up} = 0.005$, $p = 0.155$). Furthermore, we found weak evidence for a negative correlation between accuracy and latency (GLMM, estimate $= -0.005$, $CI_{low} = -0.010$, $CI_{up} = -0.0004$, $p = 0.0341$; Fig. 2a) and no evidence for a difference between within- and between-subject variation (GLMM, estimate $= -0.004$, $CI_{low} = -0.019$, $CI_{up} = 0.011$, $p = 0.576$) or an effect of trial on accuracy (GLMM, estimate $= -0.042$, $CI_{low} = -0.168$, $CI_{up} = 0.082$, $p = 0.522$). In the size group, we found no evidence for either a within- (GLMM, estimate $= 0.0002$, $CI_{low} = -0.006$, $CI_{up} = 0.006$, $p = 0.958$) or between-subject effect (GLMM, estimate $= -0.002$, $CI_{low} = -0.017$, $CI_{up} = 0.014$, $p = 0.840$). Furthermore, we found no evidence for a correlation between accuracy and latency (GLMM, estimate $= 0.0002$, $CI_{low} = -0.006$, $CI_{up} = 0.006$, $p = 0.961$), that within- and between-subject variation differed from each other (GLMM, estimate $= -0.002$, $CI_{low} = -0.018$, $CI_{up} = 0.015$, $p = 0.823$) or an effect of trial on accuracy (GLMM, estimate $= 0.011$, $CI_{low} = -0.109$, $CI_{up} = 0.133$, $p = 0.855$).

We found no evidence for within- or between-individual effects when analysing each treatment separately (Appendix Tables A3 and A4) except for weak evidence of a within-individual effect in the number group treatment 2 versus 3 (GLMM, estimate $= -0.019$, $CI_{low} = -0.036$, $CI_{up} = -0.003$, $p = 0.010$), a weak between-individual effect in the number group treatment 2 versus 4 (GLMM, estimate $= -0.027$, $CI_{low} = -0.054$, $CI_{up} = -0.001$, $p = 0.051$) and a weak within-individual effect in the size group treatment 1 versus 2 (GLMM, estimate $= 0.014$, $CI_{low} = -0.002$, $CI_{up} = 0.031$, $p = 0.080$). While we did not find evidence that choice performance was repeatable in our previous analysis (Szabo et al., 2021), we found very strong evidence that latency was repeatable in the number (R $= 0.493$, $CI_{low} = 0.139$, $CI_{up} = 0.724$, $p < 0.0001$) as well as the size group (R $= 0.228$, $CI_{low} = 0.046$, $CI_{up} = 0.419$, $p < 0.0001$) (Figs. 3 and 4).

We found moderate evidence for a negative correlation between latency and the chosen quantity in the number group (GLMM, estimate $= -0.059$, $CI_{low} = -0.114$, $CI_{up} = -0.006$, $p = 0.036$) but no evidence for such a correlation in the size group (GLMM, estimate $= -0.028$, $CI_{low} = -0.087$, $CI_{up} = 0.029$, $p = 0.335$). Furthermore, we found no evidence that the total amount of presented food within a trial did affect latency to choice in the number (GLMM, estimate $= -0.014$, $CI_{low} = -0.058$, $CI_{up} = -0.031$, $p = 0.560$) or the size group (GLMM, estimate $= 0.005$, $CI_{low} = -0.045$, $CI_{up} = 0.053$, $p = 0.846$).

4. Discussion

Contrary to our expectations, we found that gidgee skinks were more likely to make fast decisions that were correct rather than incorrect, at least in the number-based quantity discrimination test. Even when looking at the choice within the first trial only, we found that lizards made either fast and accurate or slow and inaccurate choices. The number group also demonstrated a negative correlation between decision time and accuracy which is opposite to what would be expected

from a speed-accuracy trade-off. Our analysis revealed within- but no between-individual differences when treatment was pooled indicating that our lizards did not exhibit decision styles. Additionally, we found greater latency to choice (i.e. slower) in treatment 1 versus 2 compared every other treatment, which is consistent with our expectation that more similar quantities are harder to distinguish. In the size group, we found no evidence for a correlation between decision time and accuracy or within- and between-individual differences. When we analysed SAT within each treatment we only found weak evidence for within- and between-individual variation in some treatments. We found some indication, albeit only in the number group, that choosing larger quantities (e.g. 4 and 3) was done faster than choosing smaller quantities (2 and 1) indicative of some influence of motivation on choice latency. Finally, latency to choice was repeatable, more so in the number group than the size group.

Our analyses generally show a lack of SAT; the pattern we found might be better described as a speed-accuracy alignment. We found a negative correlation between accuracy and latency; more accurate decisions were made faster than more inaccurate decisions albeit only in the number group. In the original study (Szabo et al., 2021), the size group did not show any evidence that they could discriminate between the presented quantities and here, they did not show any evidence of a correlation between accuracy and decision time. Furthermore, contrary to the number group, neither the chosen quantity nor the total amount presented affected latency in the size group. We might, therefore, conclude that lizards from the size group randomly chose one of the presented quantities which might be interpreted as them being unable to discriminate single food items differing in size. In our original study, we found that discriminating between 1 versus 2 and 2 versus 3 were most difficult for the lizards in the number group (lowest accuracy). SAT predicts increased decision time in difficult tasks to maintain accuracy. Here, we found increased latency when choosing between 1 versus 2 compared all other treatments which did, however, not improve accuracy as would be predicted by a SAT. Overall, an increase in decision time did, therefore, not improve our lizards' choice accuracy.

During our study, skinks were not exposed to stressors or costs, such as starvation or predation risk. The animals were also not fed the quantity of carrots they chose, to avoid learning based on food consumed (reward), satiation level or time to consume the reward. This also means that they did not pay a cost for choosing the smaller quantity or failing to choose any food at all. Additionally, lizards were given only two choice which might have been easier compared to, for example, three choices (Chittka et al., 2009; Jones et al., 2020). The test methodology might, therefore, have prevented SATs, because although time to make a choice was limited, the fact that stimuli could not be consumed after a choice meant that all choices had the exact same cost. Stress has been shown to lead to faster, more inaccurate decisions in different animals or an increase in decision time to maintain accuracy (e. g. Chittka et al., 2003; Latty and Beekman, 2011). For example, bumblebees took longer to make a decision when flower colour was more similar. They also took longer to improve their accuracy when a wrong or inaccurate decision led to a negative effect, for example if cryptic predators are abundant or if the foraging reward is not favourable, such as salt water (Chittka et al., 2003; Ings and Chittka, 2008). Previous studies that demonstrated SATs in a range of animal species all tested learned discriminations or presented more than two choices, increasing task difficulty (Chittka et al., 2003; Dyer and Chittka, 2004; Ducatez et al., 2015; Ings and Chittka, 2008; Jones et al., 2020). This suggests that learning and task difficulty might play a role in the formation of SATs. Changing our methodology by increasing the costs of inaccurate choices, rewarding lizards with the chosen quantity to enable them to learn the pay-offs of their choices or adding a third choice could lead to the expected expression of SATs.

We did not find evidence for decision styles (fast but inaccurate or slow but accurate) in our test subjects, although we found weak evidence for such a style in the number group treatment 2 versus 4. This is unexpected, because from the existing literature (Chittka et al., 2003, 2009; Jones et al., 2020; Latty and Beekman, 2011; Moiron et al., 2016) we would expect such styles to become evident only when the task is difficult. Although 2 versus 4 was not the easiest quantity discrimination, it is also not the hardest among those tested. On the one hand, it is possible that 2 versus 4 was the hardest discrimination that lizards were able to solve while 1 versus 2 and 2 versus 3 were unsolvable. This would make 2 versus 4 the condition in which we would be able to detect decision styles. On the other hand, without further testing we cannot be sure that this result is not another artefact of the testing methodology.

When looking at repeatability we only found an average low repeatability $(0.1 < R < 0.2)$ in accuracy (Szabo et al., 2021). Conversely, we found quite high repeatability in latency (R = 0.493) at least in the number group; the size group showed lower repeatability (R = 0.228). A meta-analysis of repeatability in behaviour reported an average of 0.37 (Bell et al., 2009). On average, we found similar repeatability in latency to choice. Repeatability reflects both between-individual compared to within- individual consistency across repeated measures. When looking at our lizards' choice latency (Figs. 3 and 4) we see that number group animals show higher among individual compared to within-individual variance. This explains why the number group's repeatability is higher than the repeatability of the size group.

In conclusion, we found a lack of SAT in our lizards when choosing different food quantities. We even found a negative correlation between decision accuracy and speed. Lizards made fast and accurate or slow and inaccurate decisions. We also found some evidence that lizards were more motivated to approach larger numbers of food items. Learning might be an important component when gathering information about what choice to make (Chittka et al., 2003, 2009; Jones et al., 2020; Latty and Beekman 2010; Moiron et al., 2016). Learning through consumption of the presented stimuli (i.e., reward) was prevented in our spontaneous quantity discrimination test, so further testing is required to reveal how increased costs (including task difficulty) and learning influence the relationship between the speed of decision-making and choice accuracy.

CRediT authorship contribution statement

Birgit Szabo: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Lisa M. Greis:** Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Eva Ringler:** Supervision, Writing – original draft, Writing – review & editing. **Martin J. Whitingc:** Resources, Writing – original draft, Writing – review & editing.

Acknowledgements

Appendix

See Appendix Table A1, Table A2, Table A3, Table A4.

References

ASAB Ethical Committee, ABS Animal Care Committee, 2022. Guidelines for the treatment of animals in behavioural research and teaching. Anim. Behav. 183, I–XI. https://doi.org/10.1016/S0003-3472(21)00389-4.

Bell, A.M., Hankison, J., Laskowski, K.L., 2009. The repeatability of behaviour: a meta-analysis. Anim. Behav. 77, 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022.

Benus, R.F., Koolhaas, J.M., Van Oortmerssen, G.A., 1987. Individual differences in behavioural reaction to a changing environment in mice and rats. Behaviour 100, 105–122.

Boogert, N.J., Madden, J.R., Morand-Ferron, J., Thornton, A., 2018. Measuring and understanding individual differences in cognition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373 https://doi.org/10.1098/rstb.2017.0280.

Chapple, D.G., 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.

Chittka, L., Dyer, A.G., Bock, F., Dornhaus, A., 2003. Bees trade off foraging speed for accuracy. Nature 424, 388. https://doi.org/10.1038/424388a.

Chittka, L., Skorupski, P., Raine, N.E., 2009. Speed–accuracy tradeoffs in animal decision making. Trends Ecol. Evol. 24, 400–407. https://doi.org/10.1016/j.tree.2009.02.010.

Cogger, H.G., 2014. Reptiles & amphibians of Australia, Seventh ed. CSIRO Publishing, Collingwood (Australia).

van de Pol, M., Wright, J., 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. Anim. Behav. 77, 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006.

Ducatez, S., Audet, J.N., Lefebvre, L., 2015. Problem-solving and learning in Carib grackles: individuals show a consistent speed-accuracy trade-off. Anim. Cogn. 18, 485–496. https://doi.org/10.1007/s10071-014-0817-1.

Dyer, A.G., Chittka, L., 2004. Bumblebees (*Bombus terrestris*) sacrifice foraging speed to solve difficult colour discrimination tasks. J. Comp. Physiol. A 190, 759–763. https://doi.org/10.1007/s00359-004-0547-y.

Franks, N.R., Dornhaus, A., Fitzsimmons, J.P., Stevens, M., 2003. Speed versus accuracy in collective decision making. Proc. R. Soc. B Biol. Sci. 270, 2457–2463. https://doi.org/10.1098/rspb.2003.2527.

de Froment, A.J., Rubenstein, D.I., Levin, S.A., 2014. An extra dimension to decision-making in animals: the three-way trade-off between speed, effort per-unit-time and accuracy. PLoS Comput. Biol. 10, e1003937 https://doi.org/10.1371/journal.pcbi.1003937.

Hadfield, J.D., 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33, 1–22. https://doi.org/10.18637/jss.v033.i02.

Heitz, R.P., 2014. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front. Neurosci. 8, 1–19. https://doi.org/10.3389/fnins.2014.00150.

Hemingway, C.T., Lea, A.M., Page, R.A., Ryan, M.J., 2019. Effects of information load on response times in frogs and bats: mate choice vs. prey choice. Behav. Ecol. Sociobiol. 73, 1–7. https://doi.org/10.1007/s00265-019-2726-4.

Ings, T.C., Chittka, L., 2008. Speed-accuracy tradeoffs and false alarms in bee response to cryptic predators. Curr. Biol. 18, 1520–1524. https://doi.org/10.1016/j.cub.2008.07.074.

Jones, N.A.R., Webster, M., Newport, C., Templeton, C.N., Schuster, S., Rendell, L., 2020. Cognitive styles: speed-accuracy trade-offs underlie individual differences in archerfish. Anim. Behav. 160, 1–14. https://doi.org/10.1016/j.anbehav.2019.11.019.

Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A., Blokhuis, H.J., 1999. Coping styles in animals: current status in behaviour and stress-physiology. Neurosci. Biobehav. Rev. 23, 925–935. https://doi.org/10.1016/s0149-7634(99)00026-3.

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. https://doi.org/10.1242/jeb.02112.

Latty, T., Beekman, M., 2011. Speed–accuracy trade-offs during foraging decisions in the acellular slime mould *Physarum polycephalum*. Proc. R. Soc. B: Biol. Sci. 278, 539–545. https://doi.org/10.1098/rspb.2010.1624.

Lenth, R.V., 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.0. ⟨https://CRAN.R-project.org/package=emmeans⟩.

Marshall, J.A., Bogacz, R., Dornhaus, A., Planqu, R., Kovacs, T., Franks, N.R., 2009. On optimal decision-making in brains and social insect colonies. J. R. Soc. Interface 6, 1065–1074. https://doi.org/10.1098/rsif.2008.0511.

Moiron, M., Mathot, K.J., Dingemanse, N.J., 2016. A multi-level approach to quantify speed-accuracy trade-offs in great tits (*Parus major*). Behav. Ecol. 5, 1539–1546. https://doi.org/10.1093/beheco/arw077.

Muff, S., Nilsen, E.B., O'Hara, R.B., Nater, C.R., 2022. Rewriting results sections in the language of evidence. Trends Ecol. Evol. 37, 203–210. https://doi.org/10.1016/j.tree.2021.10.009.

van Oortmerssen, G.A., Den Daas, S., Koolhaas, J.M., Benus, R.F., 1990. Routine formation and flexibility in social and non-social behaviour of aggressive and non-aggressive male mice. Behaviour 112, 176–193. https://doi.org/10.1163/156853990X00185.

Pachella, R.G., 1974. The interpretation of reaction time in information-processing research. In: Kantowitz, B. (Ed.), Human Information Processing: Tutotial in Performance and Recognition. Erlbaum, Hillsdale, NJ, pp. 41–82.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (URL). ⟨https://www.R-project.org/⟩.

Sonnleitner, R., Ringler, M., Loretto, M.C., Ringler, E., 2020. Experience shapes accuracy in territorial decision-making in a poison frog. Biol. Lett. 16 https://doi.org/10.1098/rsbl.2020.0094.

Stoffel, M.A., Nakagawa, S., Schielzeth, H., Goslee, S., 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.

Szabo, B., Noble, D.W., McCloghry, K.J., Monteiro, M.E., Whiting, M.J., 2021. Spontaneous quantity discrimination in a family-living lizard. Behav. Ecol. 32, 686–694. https://doi.org/10.1093/beheco/arab019.

Verdolin, J.L., 2006. Meta-analysis of foraging and predation risk trade-offs in terrestrial systems. Behav. Ecol. Sociobiol. 60, 457–464. https://doi.org/10.1007/s00265-006-0172-6.