Nectar-feeding bats learn the rule behind serial reversals

Shambhavi Chidambaram1,2, Sabine Wintergerst3, Alex Kacelnik4, York Winter1,2\*, Vladislav Nachev1

1 Institute of Biology, Humboldt University, Berlin, Germany

2 Berlin School of Mind and Brain, Humboldt University, Berlin, Germany

3 Fairchild Tropical Botanic Garden, Miami, USA

4 Department of Zoology, University of Oxford

**\*For correspondence:** [york.winter@hu-berlin.de](mailto:york.winter@hu-berlin.de)

**Present Address:** Institute of Biology, Humboldt University, Philippstr. 13, 10115 Berlin, Germany

# Abstract

Flowering plants are static, as is the resource they provide – nectar. Nectar levels are variable in time depending on when the flower empties, replenishes itself or dies. Nectar-feeding animals must trade off exploring new flowers that are potentially rewarding and exploiting flowers known to be rewarding. In such a foraging environment discrimination between the rewarding properties of flowers and flexible responses to changes in those properties are abilities with strong advantages. Commissaris’s long-tongued bat is known to have the former; we aimed to explore the latter through a spatial serial reversal learning task. Wild bats were trained to obtain rewards from two artificial flowers. At any given time only one of the flowers provided a reward. After the bats had experienced the rewarding properties of the flowers for some time, a reversal happened: the rewarding flower became non-rewarding and vice versa. These reversals of rewarding properties occurred repeatedly.

We found strong evidence that the bats learn the rule underlying reversing: when a food location suddenly dried up they quickly abandoned visiting it and switched to the alternate option. This shows that they can ‘learn to learn.’ After experiencing multiple such serial reversals, bats switched more rapidly to the rewarding option after a reversal and chose the rewarding flower more frequently overall. In fact, this increase in the number of choices for the rewarding flower became high enough that it could not increase much further: a ceiling effect was seen in the animals’ preferences. In the trade-off between exploration and exploitation the animals’ behaviour approached a ‘best of both worlds’ strategy. The bats showed near-exclusive preference for the rewarding option, which they rapidly changed to after a reversal as they had learned the rule behind the environmental change and relied on it. The question of whether nectar-feeding bats can understand other predictable but more complex environmental patterns produced by the deterministic process of floral nectar secretion, or even potentially anticipate and predict them, is exciting grounds for further research.

# Introduction

Nectar-feeding bats face the challenge of exploiting resources that continually change with time, but are relatively predictable in space. Floral nectar content varies within any given night, as well from night to night; flowering plants are stationary, however, they bloom seasonally and single flowers on plants themselves wither and die every day or every few days. Bats need to detect the changing reward contingencies in their environment and adjust their behaviour accordingly. This ability of animals to recognize and respond to changes is crucial for behavioural efficiency and includes cognitive or behavioural flexibility (Tello-Ramos et al. 2019). One experimental protocol that has been widely used to demonstrate cognitive/behavioural flexibility is reversal learning, a task that is a simple experimental analogue to some of the decision-making bats do in their natural environment.

Reversal learning essentially involves two separate stages of learning. An animal must first learn about multiple stimulus response associations. These can be two spatial locations that potentially provide a food reward. The animal is then confronted with the next step which is a situation where in fact only one of the two options is profitable. In this phase of the experiment, it is optimal to only pay attention to this one option and ignore other possibilities. However, after a longer series of such repetitive stimulus response actions, reinforcement stops. The optimal strategy for the animal is to then immediately abolish its recent behaviour and instead perform another stimulus response action that in the past had proven profitable. It is this ability to abruptly terminate an ongoing behaviour in favour of switching to another behaviour that the animal should learn for efficient exploitation. In a serial reversal learning procedure the reward contingencies reverse repeatedly. An animal that responds to the rewarding stimulus more frequently than to the non-rewarded stimulus receives food more often and performs more efficiently on the task. Choosing the non-rewarded stimulus is, in the context of the task, an ‘error.’ This performance critically depends on the animal ‘s ability to switch abruptly after a reversal.

The serial reversal learning protocol can be adapted to the behaviour and sensory physiology of many different species, thus allowing comparative research. It has been done using visual stimuli in bumblebees (Strang and Sherry 2014) and guppies (Boussard et al. 2020); visual and spatial stimuli in both corvids (Bond, Kamil, and Balda 2007) and cowbirds (Lois-Milevicich et al. 2021); spatial stimuli in rats (Boulougouris, Dalley, and Robbins 2007), great tits (Hermer et al. 2018a) and grey squirrels (Chow et al. 2015); and olfactory stimuli in rats (Kinoshita et al. 2008). Reversal learning, specifically serial reversal, has been used as an explicit comparative measure of animal ‘intelligence’ (Bitterman 1964): animals considered then to be ‘higher,’ like pigeons, rats and monkeys showed a progressive improvement on the task while species considered to be ‘lower,’ like turtles and fish, did not. Though the idea of such a hierarchy is outdated, comparative research using reversal learning can reveal important differences in behaviour and learning that have evolved under the selection pressures faced by different species.

Improvement in the reversal learning task is clearly demonstrable and is therefore a meaningful criterion when comparing the performance of different species. First-order learning happens when an animal perceives a stimulus and learns which behavioural response leads to a reward - the stimulus-reward association - and changes its behaviour according to the strength of this reinforcement. Higher-order or second-order learning is the learning of rules or strategies. This is the ability to actually choose efficiently among potential behavioural responses in any given situation. In serial reversal learning the same stimuli are successively paired with a reward and then not paired with a reward. Under such conditions animals must learn the second-order rule to quickly abandon a suddenly unsuccessful response that previously resulted in reward. A rule that can maximize reward in a deterministic, reversal learning schedule is ‘win-stay; lose-shift’: repeat the behaviour which produces a reward until it stops being rewarding, and then immediately switch to another behavioural response. Such a rule means in practice that the animal makes exactly one ‘error’ per reversal. After learning the task, the perfectly optimal animal will first exclusively respond to the rewarded stimulus. After the first experience of a lack of reward (the error), the animal will change its actions and exclusively respond to the other stimulus which is now paired with a reward.

Animals are rarely, if ever, so optimal in real life. Large and deterministic changes are uncommon in nature and more gradual shifts in behaviour may be better suited to natural environmental changes. Errors in reversal learning tasks are common and can occur in different ways. An animal can perseverate: continue to choose a non-rewarding cue after the reversal has occurred or they can make anticipatory errors: changing their behaviour and choosing the other cue before a reversal has occurred. Progressive ‘improvement’ in this task, where an animal makes fewer and fewer errors per reversal is indication that the animal is learning the rule of reversal, or ‘learning to learn’(Shettleworth 2010).

What performance on the serial reversal task says about the deeper cognitive mechanisms at work, and whether the task is a measure of cognitive or behavioural flexibility, are not completely settled questions. Cognitive flexibility cannot be directly observed; it is inferred to have occurred through changes in behaviour, or behavioural flexibility (Tait et al. 2018). However, behavioural flexibility does not necessarily indicate cognitive flexibility (Dhawan, Tait, and Brown 2019). The term ‘behavioural flexibility’ itself has been used widely but inconsistently, applied to many traits that have different underlying neural mechanisms or do not co-vary (Audet and Lefebvre 2017). Behavioural flexibility in animals has evolved in response to selection pressures from different foraging environments: the flexibility required to deal with seasonal changes in fruit availability is not the same kind of flexibility required to deal with capturing a prey animal intent on escaping.

There is a sense in which the foraging ecology of some nectar-feeding animals is a natural analogue to the serial reversal learning task. The Neotropical bat species *Glossophaga commissarisi* relies primarily on flower nectar for energy. These bats have remarkably high metabolic rates for their body mass (Voigt and Winter 1999; Voigt, Kelm, and Visser 2006), due to the energetic demands of flight (Winter and von Helversen 1998; von Helversen and Reyer 1984). Many plants visited by bats put out only a few flowers every night, but their flowering seasons that can last for several months. Bat-pollinated flowers can secrete up to 1-2 mL of nectar a night (von Helversen 1993) but as only small droplets are available each time they are visited (Voigt, Kelm, and Visser 2006), the bats make several hundred flower visits per night. Indeed, a certain time after a flower is emptied the nectar-levels are replenished, so bats can visit the same flower multiple times. The bats relocate the flowering plant primarily through their excellent spatial memory in addition to local echo-acoustic cues to find the individual flowers (Thiele and Winter 2005; von Helversen and von Helversen 1999). The longer the bat waits, the more the flower refills but the higher the likelihood that a competitor could exploit the flower first. To make repeated, profitable visits to a flower, a bat must remember both the location of the flower and, ideally, estimate the flower’s expected reward value. The serial reversal learning task requires an animal to respond to a change in the profitability of available options and remember all potentially rewarding options. Though this is a simplified situation with no inter-individual competition, the behavioural flexibility required by the experimental protocol is the kind of flexibility required in the natural foraging circumstances of a nectar-feeding bat.

To investigate the flexibility of responding to depleting food locations we carried out a serial reversal learning task with wild *G. commissarisi* individuals. The bats were given two potentially rewarding options to choose between. At the start of the night, only one of the options was rewarding, the ‘S+’ option and the other was not rewarding, the ‘S-’ option. After a certain number of visits had been made by the bats, the reward contingencies reversed without any signal or cue to the bats: the previously rewarding option was now unrewarding and the previously unrewarding option was rewarding, and this reversal happened five times in a night.

Our aims with this experiment were as follows. First, we wanted to test how the bats would alter their preference between the two reward sources according to their transient rewarding properties. Second, if the bats demonstrated flexibility in responding to the reversals, we wanted to explore how this was reflected in their decision-making. What would be the relative number of visits made to the rewarding and the non-rewarding options, and how swiftly would behavioural allocation between the two options change after a reversal? Third, we wanted to see if the bats were capable of second-order learning, or ‘learning to learn.’ Could the bats learn the rule behind the change in their environment and converge to the maximal payoff of one error per reversal achievable through win-stay, lose-shift?

After the analyses described above were done and the data and results examined, we performed further analyses to explore the conclusions of our confirmatory analyses. The difference between these results must be clearly noted. We reasoned that there is a difference between the first visits of a night, before any experience of a change in contingencies, and all the subsequent visits after at least one reversal had occurred. We statistically tested for this difference in the bats’ choice behaviour and compared the results of the confirmatory and exploratory analyses.

# Methods

## Study site and subjects

The experiment took place at La Selva Biological Field Station, Province Heredia, Costa Rica. Bats of the species Glossophaga commissarisi were captured from the wild and retained in a flight cage through the experiment. The bats were attracted to a particular location in the forest using chicken-feeders filled with sugar-water (see Reward below) as bait. The feeders had cotton swabs soaked in dimethyl disulphide on them, a chemical attractant produced by many bat-pollinated flowers [@von\_helversen\_sulphur-containing\_2000]. and then caught in mist-nets. The bats were sexed on capture and housed in two outdoor, meshed flight-cages (4 x 6 m) under ambient light conditions. All individuals were weighed and marked with radio frequency identification (RFID) tags placed as collars around their necks.

A total of 16 bats participated in the main experiment and the first stage of the experiment began on the same night the bats entered the cages. A group of four experimental bats of the same sex were placed in a flight cage together. Two such groups were run in parallel, one in each flight-cage so the data were collected simultaneously. Two groups were run in parallel, one in each flight cage, and the data were collected simultaneously. At the end of the experiment, the RFID collars were removed and the bats were released back into the wild. All the data collection was completely automatized. Two of the bats did not drink a sufficient amount of sugar-water to meet minimum energy requirements; they were released before the end of the experiment and not replaced, and the data from these two individuals were not analyzed. Thus, 14 bats (7 males and 7 females) completed the experiment. Permission for this research was granted by Sistema Nacional de Areas de Conservación (SINAC) at the Ministerio de Ambiente y Energía (MINAE), Costa Rica.

## Experimental Setup

### Reward

The reward received by the bats during the experiment was also their main source of food. The reward was a 17% by weight solution of sugar dissolved in water, hereafter referred to as ‘nectar.’ The sugar consisted of a 1:1:1 mass-mixture of sucrose, fructose and glucose. The nectar was thus similar in composition and concentration to the nectar produced by wild chiropterophilous plants [@baker\_sugar\_1998]. Every night, the bats were also given ad-libitum access to 10 mL of supplemental food: 3.5 g of hummingbird food (NektarPlus, Nekton) and 3.5 g of milk powder (Nido 1+, Nestle) in 10 mL of water. They were also given a small bowl of locally-sourced bee pollen.

### Flower and pump setup

Each flight cage had a square plastic frame in the center (2 × 2 × 1.5 m). Eight reward-dispensing devices - hereafter referred to as ‘flowers’ - were fixed in a radial pattern on this frame, two on each side of the square (see Figure 1) with a distance of 40 cm between adjacent flowers. At this distance, bats can easily discriminate neighbouring flowers [@thiele\_hierarchical\_2005]. Each flower had the following parts: an RFID reader mounted on a plastic cylinder around the head of the flower; an infra-red photo gate; and an electronic pinch valve through which a silicon tube was placed and fixed to the head of the flower.

A stepper-motor pump was placed in the center of the plastic frame in each cage. The pumps contained a 25 mL Hamilton glass syringe (Sigma Aldrich). The step volume of the two pumps differed slightly: the pump in Cage 1 delivered 2.11 L per step of the stepper-motor, whereas the pump in Cage 2 delivered 3.33 L per step. The glass syringe was connected to the tubing system of the flowers through five pinch valves [@nachev\_psychophysics\_2012]. The pinch valves controlled the flow of liquid from the pump to the system and from a reservoir of liquid to the pump. The reservoir (500 mL thread bottle, Roth, Germany) was filled with fresh nectar every day and was connected to the syringe through the valves.

Every day at around 1000 h, the old nectar was emptied from the system, which was rinsed and filled with plain water until 1500 h, when it was filled again with fresh nectar. Twice a week, the system was filled with 70% ethanol for an hour to prevent microbial growth, then repeatedly rinsed with water.

When a tagged bat approached a flower, the individual RFID number was read. If the bat then poked its nose into the flower and interrupted the light beam, it triggered the release of a reward. The pinch valve opened and the pump moved the correct number of pre-programmed steps to dispense nectar to the head of the flower. The bat could easily hover in front of the flower and lick up the nectar. A reward was given only when both events occurred, i.e., the RFID reader identified a bat and the photo gate was triggered. The flowers and the pump were connected to a Windows PC, which ran the experimental programs and the program used to automatically flush, clean and fill the pump and tubing system (PhenoSoft Control, PhenoSys).

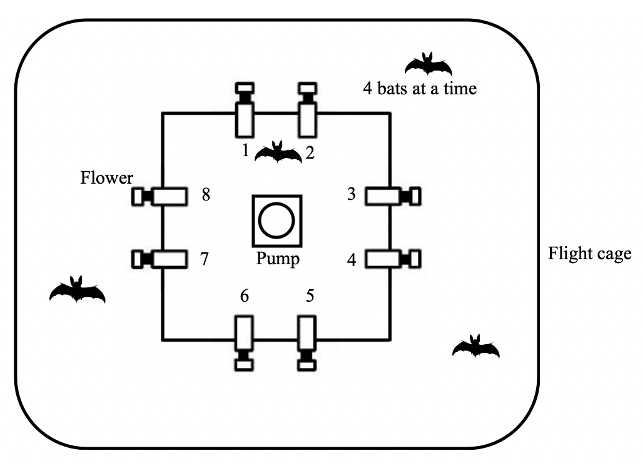


Figure 1: Schematic of the cage and flower set-up

## Experimental procedure

Out of the array of eight flowers, each bat was uniquely assigned two adjacent flowers on the same side of the square frame, programmed to reward only one of the four bats in the cage. After the system was filled with fresh nectar at approximately 1700 h, the program was left running for data-collection till the next morning. Thus, the bats could begin visiting the flowers to collect a reward whenever they chose, which was at nightfall at approximately 1800 h every night. During the main experiment, the bats could make a maximum of 300 visits a night.

During the course of the night, when the syringe of the pump had been emptied, the pump re-filled automatically. This event happened only once every night. On the main experimental days, this process took 4.5 minutes (SD = ±0.18) for the horizontal pump, and 2.4 minutes (SD = ±0.04) for the vertical pump.

About 1% (SD = ±0.74) of all visits made by the bats over all three experimental nights were wrongly unrewarded: the bats did not receive a reward during these visits even when they were made to a flower assigned to them that was rewarding at the time. This happened either during the pump refill times or when the pump was moving to reward a visit made by another bat that happened almost at the same time. Such events did not count towards the total of 300.

## Experimental design

The experiment proceeded through the following stages.

### Training

On the night the naïve bats were captured and placed into the flight cages, they could receive a reward from any of the flowers whenever they visited them throughout the night. To enable the bats to find the flowers, a small cotton pad soaked in dimethyl disulphide was placed on each flower. A small drop of honey was applied to the inside of the flowers to encourage the bats to place their heads inside, interrupt the photo gate, and trigger a nectar reward. By the end of the night, all the bats had found the flowers and learned to trigger rewards.

The next stage of training involved assigning the bats uniquely to two out of the eight flowers in the array. For an individual animal, only the two flowers assigned to it would elicit rewards from this stage of training until the end of the experiment. This stage was similar to the previous one, except the bats could only trigger a reward at their assigned flowers, and the chemical attractant and honey were not used.

To ensure that the bats were familiar with both flowers assigned to them, they went through one final stage of training: forced alternation after every single rewarded visit between the two assigned flowers for the duration of one night.

### Serial Reversal Learning Task

In the serial reversal learning task, the bats had continuous access to two flowers: one that gave a 40-L nectar reward, and one that remained empty. The location of the rewarding flower was not cued; however, after completing the alternation training phase, each bat knew the locations of both flowers that were potentially rewarding to it. After a bat had made 50 visits in total to the two flowers (regardless of relative allocation), a reversal occurred: the previously rewarding flower became the non-rewarding flower and *vice versa*. Importantly, only visits to the two flowers assigned to a bat counted towards the visit tally. Each set of 50 visits to the assigned two flowers, either at the start of each night or between reversals, was termed a ‘block’. There were six blocks and five reversals per night, unless the bat ceased visiting earlier. This was repeated for three consecutive nights. The same flower started the sequence every night. Consequently, the last flower to be rewarding one night was non-rewarding at the start of the next.

## Data analysis

The raw data collected during this study were the computer logged events of feeder visits recorded in comma-separated value (CSV) files. Each event included the time stamp, animal ID, photo gate interruption duration and the volume of nectar dispensed. The CSV files were read into R, which was used for all statistical analyses and creation of plots.

All the statistical models were fitted in a Bayesian framework using Hamiltonian Monte Carlo in the R package brms [@burkner\_brms\_2017], which is a front-end for rstan (Stan Development Team, 2020). The details of these models are provided in the **Supplementary Material**.

Except when stated otherwise, all the visits made by the bats to their two assigned flowers - up to 300 - during a night were included in the analyses (some of the bats did not complete all 300 visits on every night). The bats made some visits and approaches to the flowers that were not assigned to them; however, these were the minority, and were not considered for the analysis (see **Supplementary Material** for details). At the end of each of the first five blocks, a reversal occurred; the end of the last block was the end of data-collection for the night. Each block was further divided into five bins of ten visits, in order to examine the bats’ behaviour within each block. Generalized linear mixed models were used for the analyses (see **Supplementary Material** for details of the model fitting). We report here the mean as a measure of central tendency and the 89% quantile-based credible intervals for the parameters. (89% boundaries are the default for reporting credible intervals - [@mcelreath\_statistical\_2020]).

We investigated the effect of experimental night and block on the number of perseverative visits, i.e., consecutive visits to the previously-rewarding option just after the occurrence of a reversal; and the effect of experimental night, block and bin on the proportion of visits to the rewarding flower. The proportion of visits to the rewarding flower was calculated as the number of visits to the S+ divided by total number of visits to both the S+ and S-, and we denote this as the Proprew:

After examining these results, we conducted further post-hoc exploratory analyses. We investigated the effect of the first night and first block of a night as predictor variables separate from the later two nights and the later five blocks on Proprew. This was because the animals had had no prior experience of any reversals on the first night, and during the first block of every night, they had not yet experienced any reversals. Thus night and block were taken as categorical variables with levels ‘first night’ and ‘later nights’; and ‘first block’ and ‘later blocks’ respectively.

A few of the results report 95% confidence intervals, as opposed to credible intervals, and these are noted specifically. The confidence intervals were calculated by non-parametric bootstrapping without assuming a normal distribution of the data, using the Hmisc package, which is part of the tidyverse system of R packages [@wickham\_welcome\_2019].

All statistical analyses and creation of plots were done in R.

## Data availability

All data and analysis code are available online at …..

# Results

## Confirmatory analyses

### Bats made a very high proportion of their visits to the rewarding option

The bats made a very high number of visits to the rewarding option, quickly abandoning an option when it stopped being rewarding after a reversal and switching to the other (Figure 2). A consistent pattern emerged over all three nights: a sharp decrease in the proportion of visits to the formerly rewarding option immediately following a reversal, concomitant with a rapid increase in visits to the newly rewarding option. There are two particularly interesting points about this overall pattern: the behaviour during the first bins of each block; and the changes between the first night and the following two nights.

At the start of the first night, in the very first bin of ten visits when the bats did not yet have any information about the available options and had never experienced a reversal, the Proprew (the proportion of visits to the rewarding option) averaged across individuals was close to chance: 54.5% [95% CI 46.8, 62.3]. Within the next ten visits however, Proprew increased to 92.1% [95% CI 87.1, 96.4] and by the last bin of this first block was 99.3% [95% CI 97.9, 100]. Immediately after the first experience of a reversal, the Proprew dropped down to 13.6% [95% CI 8.4, 18.8] in the first ten visits, but came back up to 96.4% [95% CI 92.9, 99.3] by the last bin of this block.

At the very start of the second and third nights, in the first bin of visits before any experience of a reversal on that night, the average Proprew of all the bats was 69.8% [95% CI 64.6, 74.7]. This was significantly higher than random choice and higher than the Proprew in the corresponding bin of the first night. For the remainder of these nights the Proprew showed a similar pattern to the first night: a decrease immediately after the reversal and then an increase to near 100%: 94.8% [95% CI 94.1, 95.5], comparable to the 93% [95% CI 91.8, 94.1] on the first night.

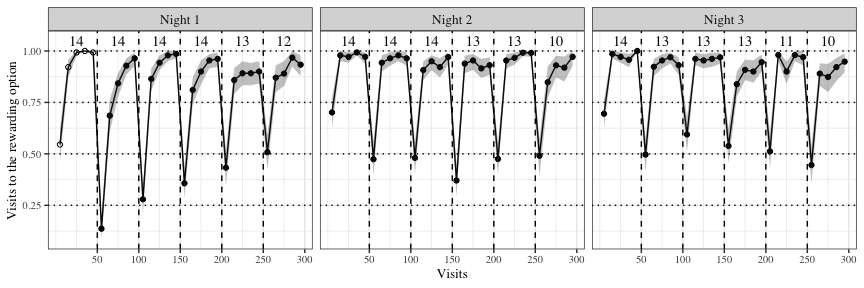


Figure 2: Visits to the rewarding of two options across the three experimental nights. Data are average proportions for bins of ten visits averaged over all 14 individuals. Data are indicated by white points in the first block on the first night before the bats had experienced any reversals; the bin averages of the other blocks are indicated by black points. Numbers indicate the bats that participated in a block. Shading shows 95% confidence intervals. Dashed lines show reversals

### Bats switched to the rewarding option faster as they experience more reversals

As the bats experienced more reversals, they made their first visit to the rewarding flower faster and faster. Immediately after a reversal, the number of visits to the previously-rewarding flower (perseverative visits) decreased as the animals experienced more reversals (Figure 3). Visually examining the data however indicates that this change does not seem to occur on the second and third nights. It would appear therefore, that the bats get faster at responding to reversals only on the first night until they reach a plateau, which persists over the second and third nights.

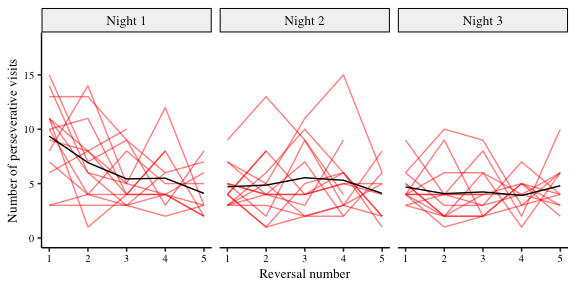
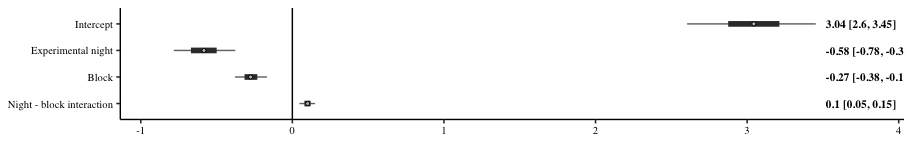
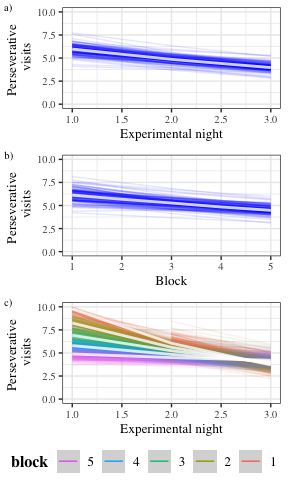


Figure 3: Number of perseverative visits made by the bats after each reversal across all three nights. A perseverative visit is a visit to the previously-rewarding option after a reversal and before any visit to the newly-rewarding option. By definition, there were no perseverative visits in the first block of a night. The red lines show individual data and the black lines the group average (N = 14). Reversals were between the two flowers assigned to an individual bat

The results of the statistical analysis are summarized in Figure 4 and Figure 5. The bats make fewer perseverative visits as they experience more reversals and more experimental nights, but this effect on perseverative visits decreases until a plateau is reached by the second experimental night. Overall, though the bats ‘improved’, there was no evidence that the number of perseverative visits decreased to one per reversal, the optimum performance.



*Figure 4: Forest plot of the estimates of the effect of night and block on perseverative visits. Circles represent the means of the posterior distributions of the slope coefficients, thick horizontal lines represent 50% credible intervals, and thin horizontal lines 89% credible intervals. The numbers in bold are the means of the posterior distributions and 89% credible intervals*

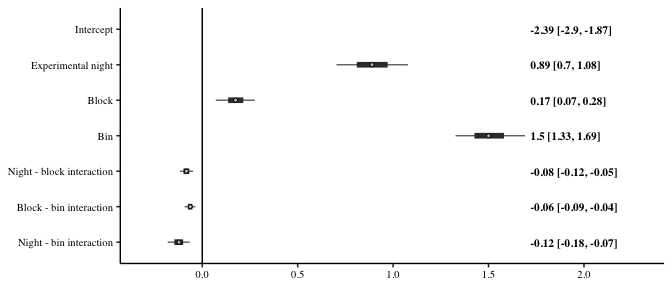


*Figure 5: Conditional effects plots from the model of the perseverative visits showing the effect of the predictors and interaction conditional on the value of the other predictors being equal to their mean, sampling from the posterior distribution. a) Conditional effects of experimental night on the number of perseverative visits b) Conditional effects of block on the number of perseverative visits c) Conditional effects of the interaction of experimental night and block*

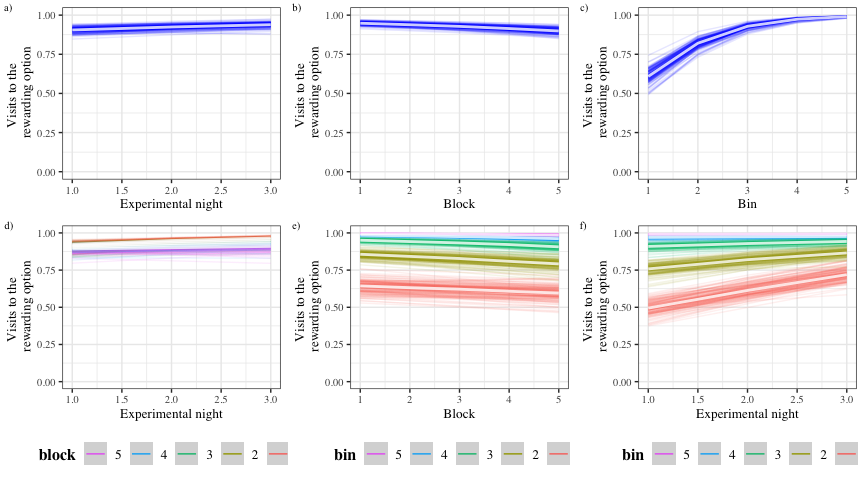
### Bats make more visits to the rewarding option overall as they experience more reversals

As the experiment proceeded and the bats experienced more reversals on more nights the overall proportion of the animals’ choices for the rewarding flower changed (see Figure 2). Within each block the change in the Proprew was very large, showing a rapid increase from less than 0.7 at the start of a block, when the animals had either experienced the start of the experimental night or a reversal, to more than 0.9 by the second bin of 10 visits. Across the three nights the animals’ choices for the rewarding flower also increased a little. However, there was a decrease within each night in the Proprew: the bats made slightly fewer visits to the rewarding flower at the end of the night compared to the start of the night.

As the results of the statistical analysis are summarized in Figure 6 and Figure 7. These results show the bats made an increasing number of visits to the rewarding option as each block progressed from its start to the next reversal, and as they experienced more experimental nights. However, this increase was dampened over the course of each individual night.



*Figure 6: Forest plot of the estimates of the effect of night, block, bin and their two-way interactions on the Proprew. Circles represent the means of the posterior distributions of the slope coefficients, thick horizontal lines represent 50% credible intervals, and thin horizontal lines 89% credible intervals. The numbers in bold are the means of the posterior distributions and 89% credible intervals*



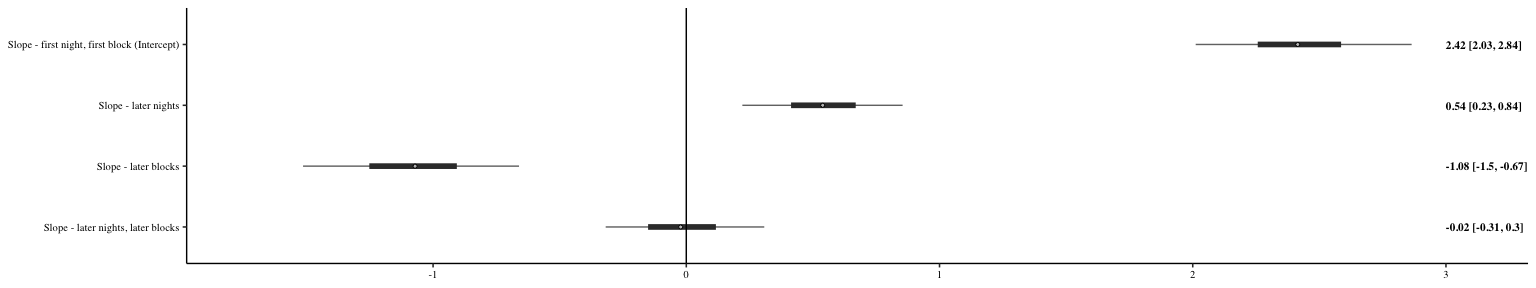
*Figure 7: Conditional effects plots from the model of the Proprew showing the effect of the predictors and interactions conditional on the value of the other predictors being equal to their mean, sampling from the posterior distribution. a) Conditional effects of experimental night on the Proprew b) Conditional effects of block on the Proprew c) Conditional effects of bin on the Proprew d) Conditional effects of the interaction of night and block on the Proprew e) Conditional effects of the interaction of block and bin on the Proprew f) Conditional effects of the interaction of night and bin on the Proprew*

## Exploratory analyses

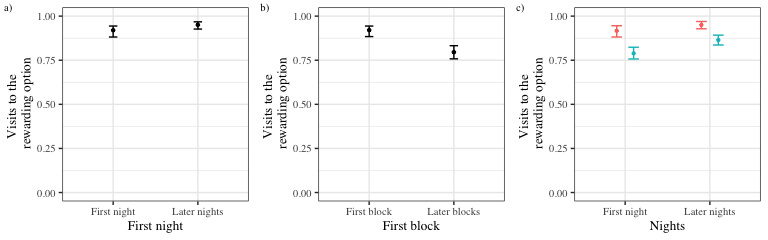
### The effect of the first experimental night and the first block of each night

The first block of an experimental night was qualitatively different from the other blocks, as this was the only part of the night when the bats had not yet experienced a reversal. A similar argument can be made about the very first experimental night: before this night the bats had never experienced a reversal at all. Therefore, after examining the results of the analyses described above, we did a further analysis to specifically explore the effects of the first block of a night and the first night.

The results of this analysis were consistent with the results of the confirmatory analysis, summarized in Figure 8 and Figure 9. The first block of a night had an effect on the Proprew: the bats made more visits to the rewarding option in the very first block of a night compared to all the later blocks. This effect was greater on the very first night compared to the later two nights, i.e., on the very first block of the very first night when the bats had experienced reward at only one option, and no reversals at all.



*Figure 8: Forest plot of the estimates of the effect of night, block, bin and their two-way interactions on the Proprew. Circles represent the means of the posterior distributions of the slope coefficients, thick horizontal lines represent 50% credible intervals, and thin horizontal lines 89% credible intervals. The numbers in bold are the means of the posterior distributions and 89% credible intervals*



Diagram

Description automatically generated

Figure 9: Conditional effects plots show the effect of a particular predictor or interaction conditional on the value of the reference category of the predictor variables, sampling from the posterior distribution a) Conditional effects of first and later experimental nights on the Proprew b) Conditional effects of first and later blocks on the Proprew c) Conditional effects of the interaction of the first and later nights and first and later blocks on the Proprew

# Discussion

Primates show a very interesting difference in between first- and second-order learning. In one study, thirteen different species of primates were compared on a visual reversal learning task (Rumbaugh, Savage-Rumbaugh, and Washburn 1996), where they were trained to discriminate a pair of stimuli up to either a 67% or 84% level of correct choices for the rewarded option. At these two different levels of training they were then given a single reversal of reward contingencies. If the primates’ behaviour was driven mainly by first-order learning, they would be expected to switch to the newly-rewarding stimulus more quickly when their training had not yet reached a high level of performance, i.e., 67%. The opposite would be true if second-order learning or rule-learning was occurring: knowing the reversal rule, the primates would be expected to more quickly adapt to the reversed option when they had reached a high level of performance in their training, i.e., 84%. The results showed that Prosimian species tended to perform better when trained to 67% and apes when trained to 84% level of correct choices before the reversal.

Progressive improvement on the serial reversal task as more reversals are experienced, which is evidence of rule-learning, has been shown in many different species: bumblebees (Strang and Sherry 2014); great tits from different environments (Hermer et al. 2018a); three different species of corvids (Bond, Kamil, and Balda 2007); rats (Castañé, Theobald, and Robbins 2010); and marmosets (Clarke, Robbins, and Roberts 2008). A decrease in the number of errors to reach a criterion (a pre-decided number of correct choices) with multiple reversals is seen in mice (Caglayan, Stumpenhorst, and Winter 2021), even when each reversal is done with a different pair of stimuli, indicating the formation of a learning set.

In our experiment wild nectar-feeding bats participated in a spatial 2-AFC serial reversal learning task where the locations of rewards changed after blocks of 50 trials, five times a night. The task was relatively simple: there were only two available options that could be potentially rewarding, and only one of these was rewarding at any given time.

Bats detected and responded to the changing reward contingencies, evidencing first-order learning. As the animals experienced more reversals, their performance improved in two significant ways.

## A faster switch

With each successive reversal the bats were quicker to switch from the previously-rewarding option to the newly-rewarding option. The number of perseverative visits decreased, meaning that the first visit to the newly-rewarding option happened earlier as more reversals were experienced, up until the fifth reversal. After this point there was no further change in the number of perseverative visits: a plateau in performance was maintained over the second and third experimental nights (Figure 3 and Figure 4).

## A change in the choices for the rewarding option

As the results summarized in Figure 2, Figure 6 and Figure 7 show, the choices made to the rewarding option increased within the course of every single block from the start of the block to the end when the reversal occurred. A similar increase occurred as the bats experienced more experimental nights. However, within each experimental night, the largest number of visits to the rewarding option were made at the very beginning of the night when no reversals had been experienced at all, and then decreased over the course of the night as the bats experienced the reversals.

## The effect of experience at different time-scales on the exploration-exploitation trade-off

When foraging efficiently in a natural context, such as the one we simulated in our experiment, animals face the classic trade-off between exploration and exploitation. The greater the behavioural allocation to the currently better alternative, the lower the ability to immediately detect an increase in reward elsewhere. The balance between these alternatives may be affected by experience. To explore this possibility, we examined whether experiencing repeated alternations had an impact on this trade-off at two different time scales: within a single night and repeatedly across nights.

At the start of a night, especially the first night, the trade-off is completely in favour of exploitation because only one option is experienced to be rewarding. When a reversal occurs, the trade-off is entirely in favour of exploration until the reward is experienced at the other location, and then once again exploitation is the only profitable option. Until the first reversal occurs the bats have no information on how the environment might change; the possibility that exploration could be profitable arose when the first reversal happened and the hitherto rewarding flower was suddenly empty. Across the three nights the bats make an increasing number of visits to the rewarding option; by the second bin of 10 visits after a reversal the proportion of choices to the rewarding option is as high as it was just before the reversal and close to 1: it is evident that the bats learned to exploit the environment of the serial reversal learning task to near optimum levels. However, the bats balance this exploitation against exploration in the following way: as they experience more reversals within a night, they make a small but increasing number of visits to the non-rewarding location. Though the number of visits to the previously rewarding option when a reversal has just been experienced never falls to one per reversal, it nevertheless reaches a plateau of approximately five such visits per reversal on average – it appears exploitative behaviour persists until the bats are ‘sure’ that an option is exhausted before they initiate exploratory behaviour. In a nutshell, the bats exploit their environment more over the three nights, but they increase their exploratory behaviour within the course of each night.

The rewards at one spatial location and the lack of rewards at the alternate location are strong reinforcers of the current behaviour. After a short sequence of unrewarded visits, or even a single unrewarded visit in many individual cases, the bats abandon a location that had until recently been rewarding in favour of the newly-rewarding location. Notably this near-optimal response to the environmental change occurred after an ‘improvement’ in performance over about five blocks, approximately one experimental night, and did not improve much further. At this point the bats had reached a plateau of near-optimal performance and did not ever show exclusive preference for the rewarding location as they continued to make a small number of exploratory visits in every block. We consider this strong evidence of second order learning. Further, we suggest that the time taken to reach a plateau in choice behaviour is a parameter that might vary among different species of animals that are capable of such learning.

The fact that experience has an effect on the trade-off between exploration and exploitation is underlined by the results of the exploratory analysis. The bats make more visits to the rewarding location on the first block of every night compared to all the others, and this difference between the first block and the blocks after the first reversal is largest on the first night. Second-order learning was only possible after at least one reversal had occurred; before that the bats were unaware of the possibility of a reversal and first-order learning led to almost the highest possible amount of exploitation.

## Learning mechanisms in a reversal learning task

Previous work has shown that the bats’ behaviour can be described well by a choice-history dependent behavioural model (Nachev et al. 2017). In such a model, an animal has an estimate of the profitability of the available option, based on its memory of past reward at these options. The past experiences are weighted by recency: the most recent experience affects the estimate of the option more than those in the past, and the estimates are updated according to the animal’s learning rate. Our results, supported by both the confirmatory and the exploratory analyses, with bats improving performance within blocks of trials are in line with the expectations of choice-history dependent reinforcement learning (Worthy and Maddox 2014).

As bats were continually confronted with reversals, however, they learned a second reinforced response that was conditioned to a novel cue. This novel cue was the single or repeated absence of expected reinforcement. This absence of reinforcement became a conditioned stimulus that was coupled to the novel conditioned response of switching to the other reward location. This conditioned response of switching location was reinforced. The two conditioned responses acquired through reinforcement learning were therefore: the win-stay response and the lose-shift response. It was the ability of abruptly switching from one type of response, the win-stay response, to another type of response, the lose-shift response, that required behavioural flexibility. This flexibility allowed the action selection of both conditioned responses in alternation. The result of this could be observed as the win-stay lose-shift behaviour of the bats.

The difference between a large reward and no reward, (‘something’ versus ‘nothing’) is the easiest possible discrimination. The change from one option to another by the animals is less remarkable than the fact that this change happened more and more rapidly: more experience of the reversal led to better exploitation of this particular type of environmental change. A pure win-stay, lose-shift behaviour would result in optimum performance. The bats’ behaviour became a close approximation of win-stay, lose-shift already after the fifth reversal by the end of the first night of the experiment.

## How do other animal species compare with the bats?

There are several key points of similarity between the bats’ performance on the serial reversal task and that of other animals. Bumblebees show improvement primarily through a reduction in the perseverative errors (Strang and Sherry 2014; Chittka 1998) on a colour reversal task. Notably, this ability to improve at the task seems to be achieved through the large number of trials, just as we had in our experiment. When the task was done with a small number of trials between the reversals, both bumblebees (Couvillon and Bitterman 1986) and honeybees (Mota and Giurfa 2010) stopped discriminating and began responding to both the rewarding and non-rewarding stimuli at chance levels.

Several different species of birds also showed performance on this task that was similar to the bats. Corvids (Bond, Kamil, and Balda 2007) show both a decrease in perseverative errors as well as an increase in preference for the rewarding stimulus as they experienced successive reversals. The improvement in performance however is seen only in a colour reversal task and not a spatial reversal task. Great tits, sampled from two different locations, do even better on a spatial reversal task than the Corvids: both trials within a block and reversal number had a positive effect on the proportion of visits to the rewarding option (Hermer et al. 2018b). A similar performance is seen in pigeons on a colour reversal task (Diekamp, Prior, and Güntürkün 1999). Among mammals, a decrease in perseverative errors is seen both in marmosets on a visual reversal task (Clarke, Robbins, and Roberts 2008) and in rats on a spatial reversal task (Castañé, Theobald, and Robbins 2010). The bats’ improvement on the serial reversal learning task thus seems to follow a similar pattern to the improvement of several other animal species, potentially indicating similar learning mechanisms.

The role of the sensory modality of the experimental stimuli must not be overlooked: we suggest that animals are likely to perform best on a version of the task that uses a sensory modality relevant to their natural foraging ecology and this is consistent with the results of serial reversal experiments with multiple species. Indeed, the transfer of improved performance across stimuli (as seen in the Corvids), is extremely strong evidence of rule-learning, and suggests an important follow-up experiment to the one we have carried out.

Our results are consistent with previous work of the same species of bat under natural conditions in the same environment (La Selva Biological Field Station, Costa Rica). This study, using the serial reversal task, evaluated the behavioural flexibility of nectar-feeding bats to fluctuations in food resource availability. Free-flying, ID-tagged wild bats interacted with 50 ID-sensor equipped artificial flowers placed over a 100 x 100 m area in the open forest that varied in their rate of nectar production. The allocation of flower types to spatial locations changed with the same pattern every night (Thiele 2006). Within each night bats adapted to the changes in resource availability. However, they needed four nights before they had adapted to the underlying recurring, predictable pattern of resource variability.

Under natural conditions bats make many small hovering visits every night to many flowers. These flowers are from many different species of plants that vary in their flowering season, flowering duration, the number of flowers that bloom per night, and the quantity of nectar they provide. Bats can distinguish flowers that differ in nectar secretion rates and return to more profitable flowers sooner (Tölch 2006), utilizing their excellent spatial memory to find the flower again and adjusting the time interval between successive visits based on the secretion rate of the experimental flowers. Such more frequent visitation of more profitable flowers may not necessarily mean the bats have the ability to estimate the temporal component of secretion rates but could be interpreted as matching visitation rate to different fixed ratio reward schedules.

In most cases under natural foraging conditions flowers are emptied in a single visit, so win-stay is not often a profitable strategy. There are however certain plants such as species of *Agave* that hold large amounts of nectar, or *Vriesea*, which if undetected all night may hold so much nectar that it requires multiple hovering visits to deplete. The perception of flower nectar volume is subject to Weber’s Law (Tölch and Winter 2007). Due to these factors, a foraging bat would need to make more than one visit to such a flower (as the optimal ‘win-stay, lose-shift’ strategy requires) to perceive that nectar levels have been depleted so much that future visits will not be profitable. Situations where lose-shift is the optimum response might be more common, but win-stay situations do occur in the bats’ foraging environments, and thus the ability to adopt a ‘win-stay, lose-shift’ strategy is likely part of the bats’ natural foraging ecology.

## Conclusions

[to be written after some discussion]

## Acknowledgements

We thank Alexej Schatz for the programming of the PhenoSoft Control software. We thank the members of the Winter lab, \_\_\_, \_\_ and \_\_\_ for many useful discussions and our colleagues at La Selva Biological Field Station for all their support. We also thank \_\_\_ for their comments and suggestions for the improvement of the manuscript.

## Author Contributions

**SC**: data collection, formal analysis, data curation, writing - original draft, writing - review and editing. **SW**: conceptualization, experimental methodology, data collection. **AK**: formal analysis, writing - review and editing, supervision. **YW**: conceptualization, experimental methodology, resources, formal analysis, writing - review and editing, supervision. **VN**: formal analysis, data curation, writing - review and editing, supervision.

## Funding

Open Access funding enabled by …

## Availability of data and code

All data and code are available in the Zenodo repository …

# Declarations

## Funding

This work was funded partly by a scholarship from the Deutscher Akademischer Austauschdienst (DAAD) to SC. Support was provided by EXC ….. and EXC ….. [need to look up the two project numbers]

## Conflict of interest

YW owns PhenoSys equity

## Code availability

All data and code are available in the Zenodo repository …

## Open Access

Licenses

## Electronic Supplementary Material

## Visits and approaches to the unassigned flowers

Only two out of the array of eight flowers were assigned uniquely to each bat but all the flowers were accessible to all the animals. The number of approaches to and attempts to get a reward from all the flowers, both assigned and not assigned, is shown in Figure 10.

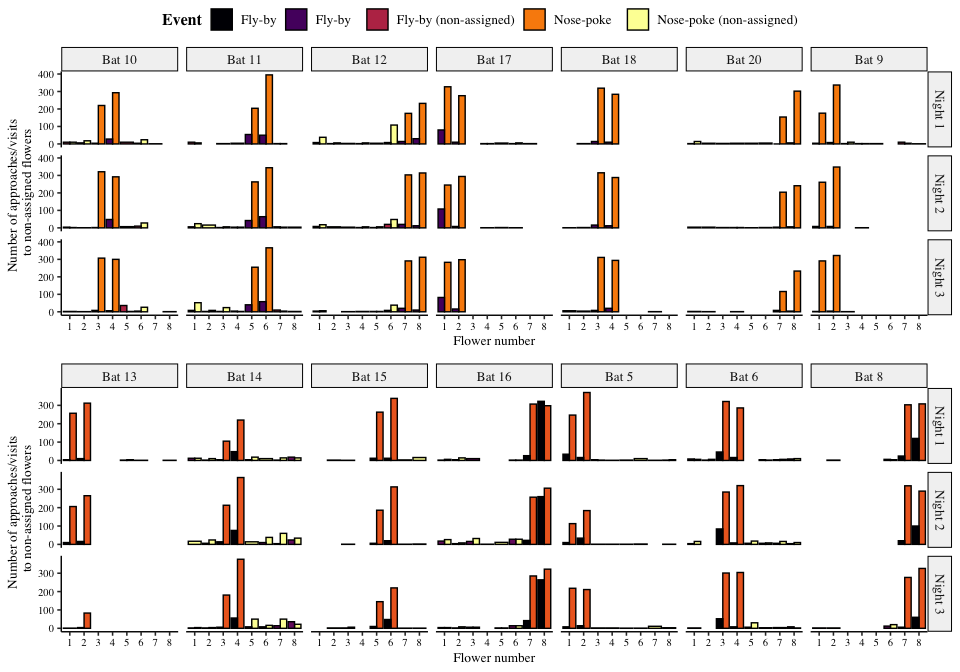


Figure 10: Visits made by the bats to all the flowers, including the ones that were not assigned to them. Yellow bars are nose-pokes at the assigned flowers, where the bats attempted to get a reward by breaking the light-barrier. Purple bars are ‘fly-by’ events near the assigned flowers where the bat flew near the flower but did not attempt to get a reward. Orange bars are nose-pokes at the non-assigned flowers and black bars are fly-bys at the non-assigned flowers.

The number of approaches or attempts to get a reward at the non-assigned flowers was a small proportion of the overall number of approaches and reward-attempts at the flowers, less than 10% every night on average as Figure 11 shows.

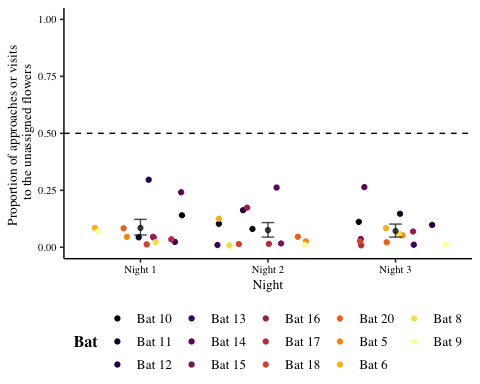


Figure 11: Proportion of visits or approaches to the un-assigned flowers out of the total number of visits or approaches to flowers. Coloured points are data from individual bats. Black points are the mean proportion per night and the error bars are 89% CIs

## Details of the statistical analyses

Weakly informative priors were used for the generalized linear mixed-models in brms. The random intercepts and slopes were given a Normal distribution with a mean of 0, and a standard deviation drawn from a Cauchy distribution with a mean of 0 and a standard deviation of 1. All the models were estimated using 4 chains with a thinning interval of 3, with 1200 warm-up samples and 1200 post-warm-up samples for the model with the first experimental night and block treated differently; 2000 warm-up samples and 2000 post-warm-up samples for the model of the first bin of 10 visits after a reversal; and 1000 warm-up samples and 1000 post-warm-up samples for the others.

For the model investigating the effect of experimental night and block on the number of perseverative visits, a negative-binomial likelihood function was used. Experimental night, block and their interaction were fixed effects and random slopes and intercepts were used to fit regression lines for each individual animal.

The model investigating the effect of experimental night, block and bin on the Proprew (calculated only over the two flowers assigned to a bat) used a binomial likelihood function with experimental night, block, bin and their interactions as fixed effects; random slopes and intercepts were used to fit regression lines for the individuals. The model of the change in Proprew in the exploratory analysis also used a binomial likelihood function with only night and block and their interaction as fixed effects; random slopes and intercepts were used to fit regression lines for the individuals. The first night and the first block of every night was treated as one level of the categorical variables and the other nights and other blocks of each night as the other level. The first night and first block were the reference categories.

Visual inspection of the trace plots, the number of effective samples, the Gelman-Rubin convergence diagnostic () and the calculation of posterior predictions for the same clusters were all used to assess the fit of the models. In all of the models the was equal to 1 for all the chains.

## Posterior predictive checks of the model of the Proprew

We carried out a posterior predictive check of the confirmatory model of the Proprew in order to check the models’ descriptive adequacy - the agreement of the models with the empirical data, which is a basic goal of modeling [@shiffrin\_survey\_2008]. Posterior samples from the posterior predictive distribution were computed for the data used to fit the models and plotted in Figure 12. A visual comparison of the two reveals a close correspondence between the data and the posterior samples of the model.

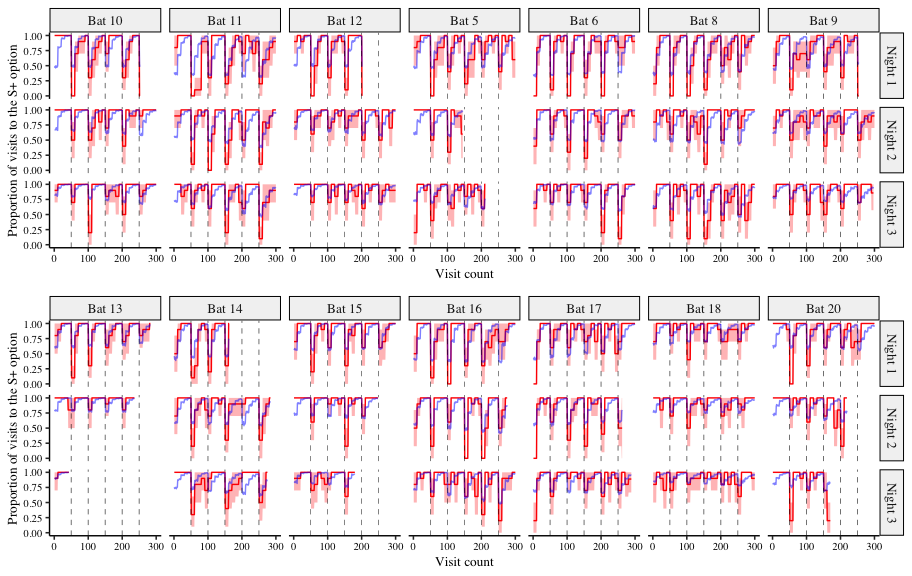


Figure 12: A comparison of the posterior predictions of the generalized linear mixed-effects model of the visits to the rewarding option and the empirical data from the bats. The red line indicates the average proportion of visits to the rewarding option per bin made by the individual bats, with the red shading indicating 95% confidence intervals; the blue line indicates the corresponding posterior prediction of the model