# Discussion

Results and their interpretation

* Opening statement summarizing all the results

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 straightforward: 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.

* Result 1: Perseverative visits and an interpretation of it

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 on the first night. After this 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).

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 a faster response to 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.

The rewards at one spatial location and the lack of rewards at the alternate location strongly reinforce choice for the former spatial location. 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.

As bats were continually confronted with reversals they learned a second reinforced response that was conditioned to a novel cue. This novel cue was the 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.

* Result 2: Choice for the rewarding option and an interpretation of it – exploration and exploitation

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.

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 strategy. 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 made 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 was 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 balanced this exploitation against exploration in the following way: as they experienced more reversals within a night, they made 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 had just been experienced never fell to one per reversal, it nevertheless reached a plateau of approximately five such visits per reversal on average – it appears exploitative behaviour persisted until the bats were ‘sure’ that an option was exhausted before they initiated exploratory behaviour. In a nutshell, the bats increased their exploitation of options known to be rewarding over the three nights, but they increased their exploratory behaviour within the course of each night.

As nectar-feeding bats repeatedly experience a change in the rewarding properties of spatial locations they respond to the change faster and show an increasing overall preference for the rewarding option over multiple experimental nights until they reach a plateau. We suggest that these changes in behaviour are strong evidence of second-order learning. If the bats showed only first-order learning they would switch from the non-rewarding to the rewarding option after a reversal, but repeated experience would not result in a faster shift or higher preference for the rewarding option. 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.

Reinforcement-based learning models have been widely applied in understanding animals’ behaviour in reversal learning tasks. Reinforcement learning is learning what actions to perform to maximize a reward in a situation or context where those actions lead from one state to another (Dayan and Niv, 2008; Sutton and Barto, 2016). Such learning can be either model-based, where experience constructs an internal model of the world to make predictions of future value (Dhawan et al., 2019), or model-free where the long-run values of actions are learned, or experience is stored in cached values of actions without any model (Dayan and Niv, 2008). Rule-learning, as we have in the bats’ behaviour, is a type of model-based learning. Evidence from rats shows that when the animals are trained past a criterion on a reversal task, they require fewer trials to reverse than animals trained to the criterion (Reid, 1953; Racht-Delatour and Massioui, 2000; Dhawan et al., 2019); if reversal learning were driven by model-free learning then increased association between stimulus and reward through over-training should result in more and not fewer trials to reverse choice behaviour.

Weighing up the evidence

* The effect of block: evidence from the exploratory analysis and interpreting through probability matching

The fact that experience has an effect on the trade-off between exploration and exploitation is underlined by the fact that there is a difference in preference for the rewarding option in the first block of a night compared to the other blocks, and this difference is highest on the first night. The bats made 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 was 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.

When reward was not experienced at an option that had been rewarding until that point, the perceived probability of reward at that option would drop below 100%. One way of interpreting the decrease in choice for the rewarding option as multiple reversals are experienced in a night is that the bats is probability matching: the probability of their choosing an option is proportionate to the probability that they will be rewarded at that option (Houston, Trimmer and McNamara, 2021, Ellerby and Tunney, 2019). If the animals experience no reward at an option at multiple points during a night, it is consistent with probability matching that they correspondingly decrease the proportion of visits they make to that option. In the context of the 2-AFC serial reversal task, as there are only two potentially rewarding options, this translates to an increase in exploratory behaviour to the option previously experienced as non-rewarding, which is what we see in the bats’ behaviour.

* Strength of evidence for perseverative visits

We note however that the effect of experience more reversals on the perseverative visits is limited. The effect seemed to be present, but only on the first night, and there seemed to be a great deal of variation between the individual animals. This might partly be due to the small sample size – only 14 animals participated in the experiment and not all of those animals completed all six blocks on any of the three nights. Another possibility has to do with the technical aspects of the pump and flower system. Once every night the pump refilled the glass syringe, taking about 2-4 minutes. During this time the flowers could not deliver a reward even if the bats visited them when they were supposed to be rewarding. Furthermore, the pump could only deliver a reward to one flower at a time. As the bats were housed in groups in four, a small number of their visits to the rewarding flower occurred when the pump was already delivering a reward to another flower. Therefore, the bats had a small number of experiences when they made a visit to the rewarding flower and were improperly unrewarded, followed by the proper rewarding schedule. This may have contributed to the animals’ persistence at a non-rewarding option immediately after a reversal, and possibly to their reaching a plateau.

Strengths and weaknesses of the study

* These are wild bats and we struck a good balance between getting enough data and being as close as possible to the natural environment.

One of the main strengths of our study is that it was done in an environment that was part of the animals’ natural with wild-caught individuals that knew how to forage in such an environment. Free-flying bats provide the most realistic data of foraging behaviour, but sample size and the quantity of usable data are then very difficult to influence. Placing wild bats in flight cages in their natural habitat was our compromise solution.

One key difference between our experiment and the natural environment is that bats rarely encounter such rich options in the latter. During the training days the bats not only learned the spatial location of the flowers and how to trigger rewards at them but also that a flower that gives a reward can continue to do so indefinitely, which is impossible for flowers in the wild.

* Effect size, sample size and follow ups

Though our sample size was small, we still had several hundred data points per night, and the statistical methods we chose meant that we did not have to exclude any animal that did not complete the experiment on any night. There was a large and obvious effect on choice behaviour within each block, increasing from 50-70% to approximately 80-90% and both our exploratory and confirmatory analyses showed the effect of night on the animals’ choice behaviour. The effect of reversals within a night, i.e., the effect of block, was smaller, and a larger experiment would provide a rigorous test of our conclusion that the choice for the rewarding option declines slightly from the beginning to the end of an experimental night. Additionally, such an experiment could study of the effect of age, sex and other such characteristics which have ignored in our study.

The choice for the rewarding option compared to the non-rewarding option (first-order learning) showed an enormous effect size: the bats made approximately 90% of their visits to the rewarding option, compared to the 50%. This is what one would expect, given the ease of discrimination between something and nothing, but the role of the sensory modality of the experimental stimuli must not be overlooked. Spatial location was the only difference between the two options, which an environmental characteristic that is highly salient to nectar-feeding bats. Animals are likely to show the strongest first-order, and probably second-order learning in a reversal learning task sensory modality relevant to their natural foraging ecology. However, the transfer of improved performance across sensory modality is the most rigorous test to show the occurrence of rule-learning (Bond, Kamil, and Balda 2007), which is a cognitive and not sensory phenomenon. This suggests an important follow-up experiment to the one we have carried out: if nectar-feeding bats can indeed learn the rule behind environmental change they should be able to do so when faced with more than one sensory modality.

Bigger picture

* Bat behaviour and cognition

Our results are also 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. In addition, as mentioned earlier, this experiment produced data only from four individuals, a sample size much smaller than ours.

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 “jackpot” 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. Win-stay is a very profitable strategy when a jackpot is encountered as competitors might find and empty it at any time (Ohashi and Thomson, 2005). 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.

* Other animals

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.