Central Lancashire Online Knowledge (CLoK)

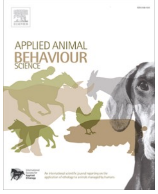
Title	Whoa, No-Go: Evidence consistent with model-based strategy use in horses during an inhibitory task
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/52113/
DOI	https://doi.org/10.1016/j.applanim.2024.106339
Date	2024
Citation	Evans, Louise, Cameron-Whytock, Heather and Ijichi, Carrie (2024) Whoa, No-Go: Evidence consistent with model-based strategy use in horses during an inhibitory task. Applied Animal Behaviour Science, 277. ISSN 0168-1591
Creators	Evans, Louise, Cameron-Whytock, Heather and Ijichi, Carrie

Applied Animal Behaviour Science 277 (2024) 106339

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim



Whoa, No-Go: Evidence consistent with model-based strategy use in horses during an inhibitory task

Louise Evans[a,*,1], Heather Cameron-Whytock[b,2], Carrie Ijichi[a,3]

[a] *School of Animal, Rural & Environmental Science, Nottingham Trent University, Southwell, Nottingham NG25 0QF, United Kingdom*
[b] *School of Veterinary Medicine, University of Central Lancashire, Fylde Rd, Preston PR1 2HE, United Kingdom*

ARTICLE INFO

Keywords:
Model-based learning
Strategy
Cognition
Stop-signal
Equine
Behavioural inhibition

ABSTRACT

It is thought horses lack the prospection and brain architecture required for goal-directed, reflective model-based learning which considers future states. We investigate whether horses can use model-based strategy in an inhibitory task. Twenty subjects were trained for three sessions of a Stop-Signal paradigm using positive reinforcement (+R) for correct responding. All twenty failed to reach learning criterion. Subjects continued to touch in "Stop" contingencies indicating horses either: cannot complete Stop-Signal tasks; need further consolidation; or are utilising model-based cost-benefit analysis. Subjects underwent a further three sessions with the addition of negative punishment (+R/-P) as a cost for errors of emission (EE). If horses lack the ability to complete Stop-Signal tasks, EE would remain high across both treatments. If horses found Stop-Signal difficult but the introduction of cost aided their learning, EE would gradually decrease throughout the +R/-P condition. If horses built a cognitive model of the task but developed a strategy of indiscriminate responding in the +R condition, EE would suddenly decrease with the introduction of cost. A significant, immediate reduction in EE was observed when cost was introduced ($p=0.02$) that remained stable throughout the +R/-P condition providing evidence consistent with model-based cost-benefit analysis in horses.

1. Introduction

Instrumental learning can be acquired through both model-free and model-based learning. Model-free learning is a relatively simple computation based on accrued trial-and-error learning that forms habits. As such, changes in contingencies and environmental conditions can only be responded to slowly, as a new "cache" of trials must be built (Dayan and Berridge, 2014a). The accruing cache adjusts the reward prediction error which signals the value achieved by a given action to maximise future expected rewards (Sutton and Barto, 2018). This is achieved by comparing differences between rewards actually received and those expected, based on previous experience (Schultz, 2016). Simple tasks can be reliably learnt, given sufficient opportunities, by building associations between stimuli and their outcomes. For example, an animal may reflexively touch a target because they have been conditioned through repeated exposure to associate it with a positive outcome such as food. The target elicits an urge to make contact with it, and no more complex processing of what the target signifies. Therefore, there is no need for higher order executive function. By contrast, a model-based strategy involves prospective cognition (Dayan and Berridge, 2014a) which refers to the ability to think about possible future states and understand the likely outcomes of various responses so that optimal responses can be selected. It therefore allows an individual to select the response most suitable to the current situation out of all potential options. The model refers to this cognitive "map" of possible outcomes and is built using the state prediction error (SPE) (Gläscher et al., 2010). SPE registers violated expectancies by comparing discrepancies between the current cognitive model (cognitive map) and the current observed state (reality) (Gläscher et al., 2010). Whilst model-based is goal-directed and reflective, model-free cognition is habitual and reflexive. The former is considered more computationally demanding than the latter which is counterbalanced by its value in promoting flexible adaptive responses in dynamic situations (Huang et al., 2020a).Inhibitory control is the ability to actively ignore

---

* Corresponding author.
E-mail address: louise.evans2020@my.ntu.ac.uk (L. Evans).
[1] Orcid id 0000-0002-8718-1128
[2] Orcid id 0000-0003-0760-2584
[3] Orcid id 0000-0003-1271-8813

https://doi.org/10.1016/j.applanim.2024.106339
Received 25 February 2024; Received in revised form 27 June 2024; Accepted 29 June 2024

distracting stimuli (attentional inhibition) and/or suppress a behavioural response (response inhibition) (Tiego et al., 2018). It is used as a marker for intelligence in non-human animals, as it can reveal whether an animal is giving a habitual behavioural response or using context to flexibly alter responses. Inhibitory control indicates how impulsive an individual is, making it a clinically relevant executive function for individuals with ADHD (Ma et al., 2016; Schachar et al., 2000). The stop-signal task is a paradigm for measuring response inhibition (Raud et al., 2020). In the stop-signal task, an individual must inhibit an already initiated response. The stop-signal task challenges stimulus control, an important executive function for human behaviour (Dinsmoor, 1995), and is a marker of cognitive flexibility (Flagel et al., 2011).

Higher impulsivity score (on the Barratt scale; Barratt, 1975) has been associated with increased physiological arousal in humans during a stop-signal task (Zhang et al., 2015). Inhibitory control is a function of cognitive control, defined as the action of ignoring distracting or emotion-inducing stimuli, to maintain task focus and increase the likelihood of successful outcomes (Neill et al., 1995). Inhibitory control has been studied using Go/No-Go and stop-signal learning paradigms (Gomez et al., 2007; Rubia et al., 2001). Success in a Stop-Signal or Go/No-Go task relies on effective cognitive control to maintain task focus, even where there is competition between possible responses, in order to attain a specific goal or reward (Gunther and Pérez-Edgar, 2021).

Here, subjects had been positively reinforced with food rewards to touch a target (Evans et al., 2024). In a subsequent Stop-Signal inhibitory control test (Evans et al., In Prep), subjects were rewarded for touching the target unless a new light cue was on, in which it was a "Stop" contingency, and no reward was offered for touching. The aim of that study was originally to measure individual differences in inhibitory control and cue dependency. However, after three training sessions, all subjects still failed to reach criterion in this task which had previously been sufficient for the cohort to learn a two-choice discrimination reversal task (Evans et al., 2024). In addition, this species should be able to inhibit previously learned responses (Brucks et al., 2022a). Instead, subjects responded indiscriminately and touched the target, regardless of the light cue and reward delivery. This is known as error of emission where a response is given but is incorrect (as opposed to error of omission, where no response is given). There are three explanations for this poor performance in order of increasing cognitive complexity following Lloyd Morgan's Canon (Dwyer and Burgess, 2011). First, horses lack the ability to complete Stop-Signal tasks due to poor behavioural inhibition or executive function limitations. However, horses have evidenced Stop-Signal ability, albeit in a spatial task (Hintze et al., 2018), and demonstrate inhibitory control in a delayed-gratification task (Brucks et al., 2022). Second, it may be that horses can learn Stop-Signal paradigms, but the stimulus used was not available to them. This was unlikely as the No-Go stimulus was a light presented within the right field of vision (Murphy et al., 2009) and was distinctly off and on. In addition, they may be able to learn this paradigm but needed longer to do so. Finally, subjects developed a model of the task and were using a cost-benefit strategy. Touching the target often results in reward and there is minimal energetic cost to indiscriminate responding, whilst contingent responding requires higher computational cost (Falkenstein et al., 1999). Therefore, it may be beneficial to touch the target regardless of the contingency. However, prospective cognition and executive function required for this strategy are not thought to be well developed in horses due to their limited prefrontal cortex (de Lahunta et al., 2015).

To determine whether subjects might be using a cost-benefit analysis, an additional experiment was conceptualised. The aim was to determine whether horses had developed a model of the possible outcomes, allowing them to use strategy to respond optimally, using the least effort to gain the most reward. A cost was introduced for errors of emission and these errors monitored and compared to predictions (Fig. 1). If subjects had no model of the task, the addition of the cost may have helped them understand the task and thus improve performance gradually, by using model-free operant conditioning. This is because model-free learning requires building a new cache of experiences of the cost, to modify behaviour. If so, a steady reduction in errors would be expected following introduction of the cost (Fig. 1a) (Dayan, 2009). If the Stop-Signal stimulus was not available or salient to the subjects, no reduction in errors would be expected (Fig. 1b). This is because adding the cost would not help them perceive the light cue, thus behaviour would not be modified. If horses cannot inhibit responses in a Stop-Signal test, again, no reduction in errors would occurs after introducing the cost (Fig. 1b). If subjects developed a model of the task and were making errors of emission due to using strategy based on the low cost of the error, a sudden reduction in errors following the introduction of a cost would be observed (Fig. 1c). This is because the cost chosen removed the opportunity to gain reward, therefore the most efficient strategy to gain reward was to touch in Go contingencies and inhibit under No-Go. To achieve this flexible response to a dynamic situation, subjects would need to adaptively compute the ideal response by searching simulated outcomes (Huang et al., 2020a). This would be expected to be relatively stable over the subsequent three sessions as a result of latent learning (Tolman, 1948). Therefore, errors would remain low across sessions 4, 5 and 6.

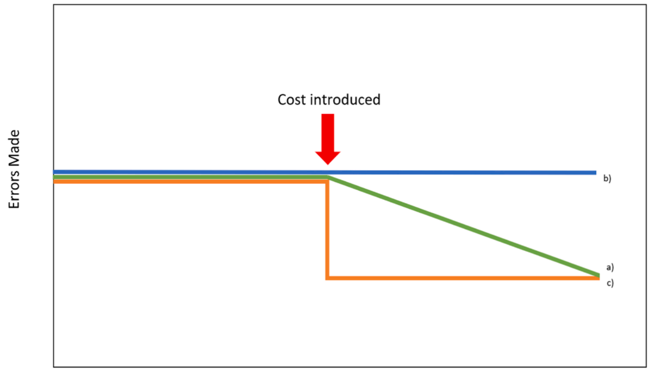


Fig. 1. Expected patterns of responding if subjects: a) did not understand the task but the addition of negative punishment improved learning; b) did not understand the task, even after negative punishment was introduced *or* the no-go signal was not available; c) were using a cost/benefit strategy and had successfully built a model of the task.

2. Methods

2.1. Ethics

This study was conducted in compliance with UK and EU law relating to the use of animals in research. Ethical approval was granted by the Nottingham Trent University Ethical Review Committee (approval code: ARE202129). A key ethical consideration for this study was allowing subjects the choice to participate or not in the training paradigm. This was a free-choice paradigm in which horses could choose not to engage with the training.

2.2. Subjects

Subjects were twenty horses aged 11–22 years (mean=16.05 years ±2.98) including 7 females and 13 castrated males of mixed breeds housed and managed at Nottingham Trent University in their usual facilities. Subject's general routines were unchanged; their regular exercise and turnout times were unaffected and their usual night-time sleeping and resting hours were uninterrupted. Subjects received *ad libitum* access to forage and water with concentrate feeds appropriate for their individual needs. Subjects have regular access to free outdoor movement and socialisation and a workload tailored to their age and physical ability.

2.3. *Pre-conditioning*

Stop-Signal paradigms test inhibition of a previously conditioned response, rather than the ability to learn a cue dependent response from the outset to determine when a target should be touched. To ensure that the task tested inhibition of a conditioned response, subjects needed to be pre-conditioned through target training to touch a target. The target was an A3 sized laminated card, either black or white (pseudo-randomised and counterbalanced across the sample). Pre-conditioning was applied to shape a nose touch response to the cue card target as per Sections 2.3.1 and 2.3.2 below. The motor requirements of this touch response were to touch any part of the target with the nose or muzzle. The aim of pre-conditioning was to operantly condition an association between target presentation and a touch response. Pre-conditioning was completed as part of a previous study (Evans et al., 2024). Training and testing sessions took place in the horses' usual stables at Nottingham Trent University (NTU) Brackenhurst Equestrian Centre. Sessions were completed between 10:00 and 15:00. The same trainer (LE) trained all 20 horses, using the same method and training schedule.

2.3.1. *Conditioning the secondary reinforcer*

Subjects completed two 15-minute whistle conditioning sessions, each on consecutive days, as per clicker training in horses (Ellis, Greening, 2016). The whistle sound was conditioned using classical, or Pavlovian, conditioning. A whistle, rather than a clicker, was used as this sample had been previously clicker trained by students at the university, whereas the whistle was a previously unconditioned stimulus. However, the principle of whistle training is identical to that of clicker training. To condition the whistle as a secondary reinforcer, the trainer (LE) stood directly in front of the horse in their usual stable and blew the whistle. Every time the whistle sounded the horse was given a small handful of palatable Ulsakind cubes (approximately five to six small pellets). Ulsakind was part of subjects' usual diets and was approved by their key caregiver (AG). This conditioning of the whistle sound was repeated continuously for three minutes, followed by a two-minute break, and then repeated for a further three minutes. This training schedule was repeated for a total of 15 minutes. The role of the whistle was to provide an instant signal that the response was correct, bridging any time delay between the horse touching the target and the food delivery. This avoids timing-related errors and is considered gold-standard protocol for operant conditioning (Skinner, 1963).

2.3.2. *Conditioning the target touch response*

Once all subjects had received two sessions of secondary reinforcer conditioning, they were conditioned to touch a target. Subjects were trained to touch one of two A3 size, laminated cue cards (one black, one white) as part of a previous study (Evans et al., 2024). The touch response was a nose or muzzle contact with any part of the target. The targeted cue card was pseudo-randomised using a random number generator for each horse and counterbalanced overall. The side the target was presented on was also pseudo-randomised for each attempt and counterbalanced. Cue cards were black and white because these are two shades which are clearly visible to horses (Macuda and Timney, 1999). The touch response was achieved using positive reinforcement. Positive reinforcement refers to the addition (+) of something pleasant when a desired response is given to increase the likelihood of that being offered again (reinforcement). The whistle sound acted as a secondary reinforcer and a food reward (Ulsakind cubes) as a primary reinforcer. The use of a secondary reinforcer bridges any gap between the desired response and delivery of the reward, thus aiding learning (Skinner, 1963). The cue card was introduced to the horses by the same trainer as the whistle conditioning training. The target was held in front of the horse, at the level of the horse's chest. During the initial shaping phase, any movement towards the cue card was rewarded with a whistle sound and food reward. As the horse began to give the nose touch response more readily, only complete touches were rewarded. The cue card response was considered successful when active and immediate nose touch responses were offered upon presentation of the cue card.

Each horse completed a minimum of two shaping sessions. Each session lasted one hour, with regular breaks throughout. Horses were given 5 minutes to rest for every 15 minutes of training completed. The criteria for completing the touch response training was at least 80 % correct behavioural responses to the cue card in two consecutive training sessions (Neave et al., 2013). As the horse's touch response improved, the cue card was held in different positions and at different angles, encouraging the horse to actively seek the cue card for reward. As the cards were laminated, light would reflect in different ways so moving the cue card into different positions meant that the horses learned to generalise their response to the cue card even when it looked slightly different depending on position and lighting. All subjects reached learning criteria, indicating that they were successfully conditioned to touch a particular target and not simply any item placed in front of them.

2.4. *Stop-signal paradigm*

Only the conditioned target for each subject from the previous study (Evans et al., 2024) was presented in Stop-Signal testing as the intent was to measure inhibitory control of a conditioned response, rather than discrimination between two targets. Since only a single target was now being presented that target was now positioned centrally in front of the horse by the same trainer (LE). Sessions lasted approximately five minutes, consisting of 28 Go contingencies and 12 Stop contingencies. Within each contingency subjects could offer multiple responses, limited only by the length of time the contingency lasted. Therefore, although a total of 40 continencies were offered, error rates could be much higher than this if multiple errors of emission were made within the timeframe. All subjects completed the same number of sessions (see Section 2.5).

Stop was signalled by a light cue, consisting of a battery-powered LED cyclists' helmet torch light. The light was attached to an adjustable, elasticated helmet band, worn by the trainer in the central upper abdomen area. The trainer was 151 cm in height, so the position of the light was such that it was visible to the horses, but not shining directly into their eyes. The light was a bright white LED, as this is a shade known to be visible to horses (Roth et al., 2007). In Go, the light was switched off, as this was consistent with pre-conditioning of the touch response, where no light cue was present. The light was manually operated using an on/off switch on the top edge of the LED casing.

The Stop load was 30 % which is sufficient to maintain motivation (Neave et al., 2013). After two to three Go contingencies Stop would be initiated and maintained for five, ten, fifteen or twenty seconds (*a priori* selected using a random number generator and balanced across subjects, such that each subject had a total of 150 seconds in Stop), preventing subjects predicting when the Go contingency would restart and ensured that it was the light itself that signalled the No-Go contingency and not predictable intervals (Kononowicz et al., 2022). Touches of the target under Go resulted in a whistle sound and reward (+R). Responses under Stop depended on the experimental condition (see Section 2.5).

2.5. *Experimental design*

The experiment was a within-individual design, with all subjects completing both treatments and acting as their own control. Subjects completed three sessions (one session per week) with only positive reinforcement (+R). Under Stop, the target was presented and maintained until the condition ended. Should a subject touch the target in this time this was not rewarded or punished and the target remained in position. Each touch was an error of emission. Following a fallow period of three weeks, subjects completed three further sessions (one session per week) with positive reinforcement plus the introduction of negative punishment which represented a "cost" (+R/-P). Negative punishment is the removal (-) of a desirable stimulus to reduce the likelihood of the animal offering that response again (Punishment). Errors of emission during the +R/-P phase (sessions 4–6) resulted in a 10-second "time-out" where no rewards could be earned in addition to the already scheduled five to twenty second Stop contingency. Time-out periods were standardised to 10 seconds to minimise the application of punishment as much as possible. The trainer stepped back out of the reach of subjects, lowered the target and looked down. At the end of the time-out period, they stepped back in range and re-presented the target to complete that current Stop contingency. Thus time-outs extended the time subjects must wait for the next opportunity to earn reward but did not limit how many errors could be made within the Stop contingency. For example, upon completion of the 10-second time-out, the handler may present the target again only for the subject to immediately make another error of emission and trigger a further time out period. Time-outs extended total session length, dependent on the performance of the individual, but 28 Go and 12 Stop contingencies were always available. Any reduction in errors observed during sessions 4–6 was therefore due to improved behavioural inhibition and not time-outs reducing the opportunity to make errors. Correct Stop responses were not rewarded in either phase (+R or +R/-P), to avoid simply conditioning horses to stand still during Stop trials; it was important that horses were actively inhibiting an impulse rather than performing a learned behaviour.

2.6. *Learning performance*

A Canon Legria video camera (Canon Inc., Tokyo, Japan) and tripod were used to record sessions for retrospective analysis of errors. Touches of the target under Stop contingencies were recorded as errors of emission, defined as a given response which is incorrect. These were scored by the trainer (LE) and validated by a second hypothesis blind rater with excellent agreement (weighted Cohen's kappa = 0.97).

2.7. *Statistics*

Data was statistically analysed in R version 4.2.1 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria). Data were assessed for normality using a Shapiro-Wilks test. Wilcoxon tests were used throughout to determine differences in Errors of Emission between consecutive sessions. P-values reported are adjusted using Bonferroni correction.

3. Results

No significant differences were seen between sessions 1–3 (+R) and sessions 4–6 (+R/-P) (Table 1). A significant reduction in errors of emission was observed between treatments (sessions 3 and 4) (Table 1; Fig. 2).

Table 1
Tests of differences for errors of emission made between consecutive sessions (n=20). Treatments were positive reinforcement (+R) and positive reinforcement plus negative punishment (+R/-P).

Sessions	Treatment	Median (IQR)	Wilcoxon V	p value
1	+R	12.5 (11.75)	72.5	0.86
2	+R	14.5 (20.5)		
2	+R	14.5 (20.5)	52.5	0.86
3	+R	16 (16)		
3	+R	16 (16)	206.5	**0.02**
4	+R/-P	5.5 (8.25)		
4	+R/-P	5.5 (8.25)	95.5	1
5	+R/-P	7 (6.5)		
5	+R/-P	7 (6.5)	72	1
6	+R/-P	6.5 (7.5)		

4. Discussion

The results of this study are consistent with model-based learning. It appears that in sessions one to three, horses developed a model of the task based on a cost-benefit analysis of outcomes. Indiscriminate touch responses did not result in loss and sometimes resulted in reward. The level of reward was therefore the same regardless of whether horses chose to follow the rules, requiring mental energy (Falkenstein et al., 1999), or respond indiscriminately to the target. The introduction of a time-out period for errors of emission at session four may outweigh the mental cost of following the rules, altering the horses' strategy in favour of a cost-avoidant model-based strategy. The evidence consistent with this is that from session three (+R) to session four (+R/-P), there was a significant, immediate reduction in errors. By contrast, model-free learning requires building a new cache of experiences of the cost to modify behaviour which would result in a steady reduction in errors. Therefore, the response observed here does not appear to be a subsequent learning effect, as the reduction in errors was immediate without the opportunity to cache new information (Dayan, 2009). According to Lloyd Morgan's Canon (Morgan, 1903), animal behaviour should only be considered evidence of higher cognitive processing if it cannot be explained by simpler mechanisms. If subjects had no model of the task, no reduction in errors would have been observed in session four but a slow reduction may have been observed over three sessions if the cost improved their understanding of the task. If horses cannot complete Stop-Signal tasks, or the light was not salient, no reduction would be seen in any session. Therefore, we propose that the explanation most consistent with the observed reduction in errors was that horses used a model-based cost-benefit strategy.

There are other suggested reasons why performance improved at session four, and it is important to acknowledge these. First, seasonal effects might have an influence on motivation in animals. There was a three-week period between sessions three and four due to the logistical challenge of conceptualising, planning and data collection for the new treatment (+R/-P), which was not originally part of this study. As such, it is possible that within that three-week period, horses experienced motivational changes. However, this is unlikely given that the data collection period still took place in summer and management was not changed during this time. Additionally, it would be expected that any changes would occur gradually, whereas there was an immediate, highly significant decrease in errors between sessions three and four. To the authors' knowledge seasonal changes in motivation have not been evidenced in this species. In addition, there was a two-week gap between sessions one and three, and no improvement in performance occurred in that time period when there was ample opportunity to rest and consolidate learning.

A second, important point to acknowledge is the Clever Hans effect. It is possible that the subjects used subtle, unintended cues from the trainer which influenced their responses. However, though it is never possible to rule out a Clever Hans effect when a human is present, such unintended cues would have been present in sessions one to three, where subjects did not demonstrate that they understood the task. Therefore, Clever Hans is unlikely to explain the sudden improvement in performance between sessions three and four. The only cue that changed was intentional, in that the trainer stepped back and gave a time out. However, if horses found the task difficult, but the time-out cost aided their learning, a gradual improvement in performance over the course of sessions four, five and six would be expected, as this would reflect learning and caching responses. Horses may have used Clever Hans in sessions one to three but chose a low-cost indiscriminate responding strategy. When this strategy was no longer effective due to cost introduced, they appear to have switched to cue-dependent discriminate responding.

It is not possible to entirely rule out other mechanisms involved in

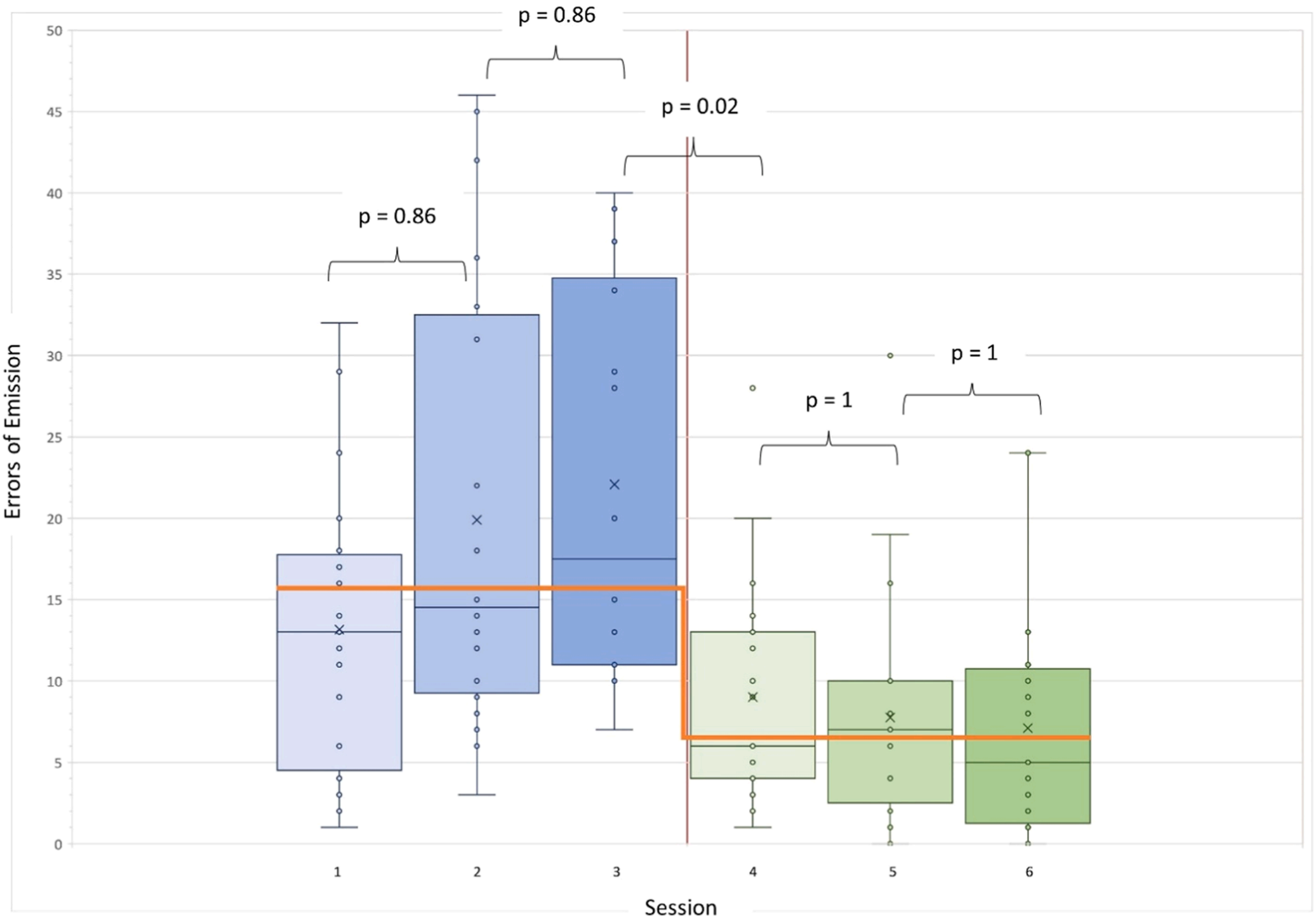


Fig. 2. Errors of emission for each session. In sessions 1–3 (blue) only positive reinforcement (+R) is used. In sessions 4–6 (green) a "cost" for errors is suddenly introduced using negative punishment (+R/-P), resulting in an immediate, significant reduction in errors (p=0.02, n=20). The predicted pattern of errors for model-based strategy (Fig. 1) is superimposed for comparison. The boxes represent the first to the third quartile of data, with a bold line at the median value. The maximum and minimum values are indicated by the upper and lower whiskers, respectively.

learning this task based on the results of the current study alone. This study was conceptualised based on an interesting observation about horses' responses during a separate investigation. As such, this was an opportunistic approach to investigating model-based learning in horses. More detailed investigation involving a larger sample and more sophisticated methods (such as EMG; Sitole and Sup, 2023) is required to rule out other learning mechanisms. However, the current study provides preliminary evidence to suggest that horses may possess model-based learning ability. This is the first evidence consistent with this ability in this species and should be explored further.

Basic cost-benefit analysis has been observed in a range of species (Burtsev and Turchin, 2006; Georgiev et al., 2013), including snails (Gillette et al., 2000), bats (Wilkinson, 1992) and starlings (Wiebe, 2003). As a simple strategy, starlings use cost-benefit analysis to decide where to spend the most time and energy on foraging, based on predicted yield (Dall and Cuthill, 1997). More complex strategies, such as creating shortcuts, have been observed in dogs (Chapuis and Varlet, 1987). Macaque monkeys have demonstrated highly complex strategy use in a three-armed bandit task, where the reward values of three different behavioural options fluctuated (Walton et al., 2010). Even basic model-based learning involves consideration of the consequences of choosing one action over another, based on the learner's mental model of the environment (Huang et al., 2020b; Huys et al., 2014; Wunderlich et al., 2012). This is a relatively complex cognitive skill requiring good executive function, particularly for an animal with an underdeveloped prefrontal cortex (PFC), such as the horse (Hausberger et al., 2019; Schmidt et al., 2019). The PFC is thought to be instrumental in model-based learning (Bunge et al., 2003; Mcdannald et al., 2012; Tsujimoto et al., 2011) due to its role in processing cause-and-effect relations between choices and outcomes (Tsujimoto et al., 2011). However, other brain structures implicated in model-based learning are available to the horse, including the hippocampus (Jacobs and Schenk, 2003; Schmidt et al., 2019; Vikbladh et al., 2019). The hippocampus creates a cognitive map (Jacobs and Schenk, 2003; O'Keefe and Nadel, 1979), which allows an animal to form a model of their environment. In model-based learning, a map of events and environmental stimuli is created, allowing the learner to prospectively evaluate the consequences of their actions (Dayan and Berridge, 2014b). The horse has a particularly well-developed hippocampus (Schmidt et al., 2019), contributing to their adeptness for spatial learning (Baragli et al., 2011; Hanggi, 2010; McLean, 2004; Murphy, 2009). The horse's demonstrable ability to form spatial models (Baragli et al., 2011; Hanggi, 2010; McLean, 2004; Murphy, 2009) may have relevance to the apparent model-based strategy observed in the current study. However, the mechanisms underlying the use of model-based strategy seen here are unclear without the use of in vivo brain imaging, not yet developed in horses (Schmidt et al., 2019; Tsujimoto et al., 2011; Walton et al., 2010). Based on the preliminary evidence we present here, this is an interesting area that should be investigated further using more sophisticated methods.

The results presented here suggest that horses may have higher-order cognitive capabilities than have been previously demonstrated (Hanggi, 2005). Higher order functions, also referred to as executive functions, allow animals to optimise their response to novel circumstances. The horse's ability for conditioned stimulus-response learning (Brubaker and Udell, 2016; Hanggi, 2005; Murphy and Arkins, 2009) and spatial memory (Baragli et al., 2011; Hanggi, 2010; McLean, 2004; Murphy, 2009) has been well evidenced. However, the results of the current study suggest that horses may be capable of certain higher-order processes. Concept learning appears to have been observed in a small number of equine studies to date (Gabor and Gerken, 2012; Hanggi, 2003; Uller and Lewis, 2009). Horses were able to select novel, two-and-three-dimensional objects based on the concept of relative size difference (Hanggi, 2003). In a separate study, horses repeatedly demonstrated an ability to select the greater of two quantities of apples, indicating a basic understanding of relative quantity difference (Uller and Lewis, 2009). If horses can form concepts, it is possible that conceptualisation helps them to develop models during learning, as the results of the current study suggest. A recent experiment found that horses were able to use self-control to wait for higher value rewards (Brucks et al., 2022b). This suggests that horses may consider consequences, including costs and benefits of their behaviour. Recently, it has even been suggested that horses may have the ability to use tools to help them access food, aid social interactions and for comfort (Krueger et al., 2022). Though this requires further investigation, it indicates a possible sense of cause-and-effect, and even prospective planning (Krueger et al., 2022; Osman et al., 2014). The research to date has identified behaviours in horses which point towards some higher-order cognition (Brucks et al., 2022b; Hanggi, 2003; Krueger et al., 2022; Uller and Lewis, 2009). The current study has experimentally tested this proposed potential for horses to use prospective planning to achieve a goal (which we term strategy).

It is worth noting that considerable individual variation in errors was observed in all sessions indicating strategy was not consistently demonstrated by all subjects. It is possible that the number of time-out periods may have influenced this variation. Clearly, the number of time-out periods applied directly related to the number of errors made, as time-out was a cost for making errors. As such, horses with more errors spent more time in time-out. However, it is possible that time-out periods early in the session influenced the subsequent number of errors made. Time-out periods may have induced fatigue or frustration-related arousal which in turn may have affected horses' ability to learn the task. Future research should investigate the association between arousal and inhibitory control performance in horses. Further, it would be interesting to compare physiological responses during both +R and +R/-P conditions, to assess the effect of time-out on arousal levels, and how this may relate to performance on the task. This was beyond the scope of the current study. However, though some individuals reduced their errors from session three to four more than others, all but one individual reduced errors in response to the introduction of the cost suggesting a remarkably prevalent ability within the group to do so. Variation in strategy use across individuals and the reasons for this are intriguing but beyond the scope of the current paper. However, the data does suggest a species-level capability for model-based learning.

5. Conclusions

Here we provide novel, preliminary evidence consistent with model-based strategic decision making in horses. Subjects rapidly altered their responses in a Stop-Signal paradigm, shifting from indiscriminate responding with high rates of errors of emission to conservative responding with few errors when a cost was introduced. To the authors' knowledge, this is the first evidence suggesting higher order, executive function in horses, a species thought to possess only moderate cognitive capabilities.

Funding

This project was funded by Nottingham Trent University Vice-Chancellors Award (2021) and The Horse Trust (Grant no. G3020).

CRediT authorship contribution statement

Louise Kate Evans: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Heather Cameron-Whytock:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition. **Carrie Ijichi:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to acknowledge the equine research technicians at the Brackenhurst Equine Centre, in particular Cath Hake and Anna Gregory, for facilitating access to participants and research equipment.

References

Baragli, P., et al., 2011. Encoding the object position for assessment of short term spatial memory in horses (Equus caballus). Int. J. Comp. Psychol. 24 (3) https://doi.org/10.46867/IJCP.2011.24.03.02.

Barratt, E.S., 1975. *Barratt impulsiveness scale*. ETS m 1975.

Brubaker, L., Udell, M.A.R., 2016. Cognition and learning in horses (Equus caballus): what we know and why we should ask more. Behav. Process. 126, 121–131. https://doi.org/10.1016/J.BEPROC.2016.03.017.

Brucks, D., Härterich, A., König von Borstel, U., 2022. Horses wait for more and better rewards in a delay of gratification paradigm. Front. Psychol. 13 https://doi.org/10.3389/FPSYG.2022.954472/PDF.

Brucks, D., Härterich, A., von Borstel, U., 2022a. Horses wait for more and better rewards in a delay of gratification paradigm. Front. Psychol. 13, 954472 https://doi.org/10.3389/fpsyg.2022.954472.

Brucks, D., Härterich, A., von Borstel, U., 2022b. Horses wait for more and better rewards in a delay of gratification paradigm. Front. Psychol. 13, 954472 https://doi.org/10.3389/fpsyg.2022.954472.

Bunge, S.A., et al., 2003. Neural circuits subserving the retrieval and maintenance of abstract rules. J. Neurophysiol. 90 (5), 3419–3428. https://doi.org/10.1152/jn.00910.2002.

Burtsev, M., Turchin, P., 2006. Evolution of cooperative strategies from first principles. Nature 440 (7087), 1041–1044. https://doi.org/10.1038/nature04470.

Chapuis, N., Varlet, C., 1987. Short cuts by dogs in natural surroundings. Q. J. Exp. Psychol. Sect. B 39 (1), 49–64. https://doi.org/10.1080/14640748708402251.

Dall, S.R.X., Cuthill, I.C., 1997. Searching in patches by European starlings, Sturnus vulgaris. Behav. Process. 39 (2), 149–159. https://doi.org/10.1016/S0376-6357(96)00053-8.

Dayan, P., 2009. Goal-directed control and its antipodes. Neural Netw. 22 (3), 213–219. https://doi.org/10.1016/J.NEUNET.2009.03.004.

Dayan, P., Berridge, K.C., 2014a. Model-based and model-free pavlovian reward learning: revaluation, revision and revelation [online]. Cogn., Affect. Behav. Neurosci. 14 (2), 473. Available at: /pmc/articles/PMC4074442/ [Accessed 13 January 2023].

Dayan, P., Berridge, K.C., 2014b. Model-based and model-free pavlovian reward learning: revaluation, revision and revelation. Cogn., Affect. Behav. Neurosci. 14 (2), 473. https://doi.org/10.3758/S13415-014-0277-8.

de Lahunta, A., Glass, E.N., Kent, M., 2015. Veterinary Neuroanatomy and clinical neurology Fourth. Elsevier Saunders, St Louis.

Dinsmoor, J.A., 1995. Stimulus control: part I. The Behavior Analyst 18, 51–68.

Dwyer, D.M., Burgess, K. v, 2011. Rational accounts of animal behaviour? Lessons from C. Lloyd Morgan's Canon [online]. Int. J. Comp. Psychol. 24, 349–364. Available at: http://www.comparativepsychology.org/2011-ijcp-vol24-4/03.Dwyer_Burgess_PDF.pdf [Accessed 19 January 2023].

Ellis, S., Greening, L., 2016. Positively reinforcing an operant task using tactile stimulation and food – a comparison in horses using clicker training. J. Vet. Behav. 15, 78. https://doi.org/10.1016/j.jveb.2016.08.008.

Evans, L., Cameron-Whytock, H., Ijichi, C., 2024. Eye understand: physiological measures as novel predictors of adaptive learning in horses. Appl. Anim. Behav. Sci. 271, 106152 https://doi.org/10.1016/J.APPLANIM.2023.106152.

Falkenstein, M., Hoormann, J., Hohnsbein, J., 1999. ERP components in Go/Nogo tasks and their relation to inhibition. Acta Psychol. 101 (2–3), 267–291. https://doi.org/10.1016/S0001-6918(99)00008-6.

Gabor, V., Gerken, M., 2012. Cognitive testing in horses using a computer based apparatus. Appl. Anim. Behav. Sci. 139 (3–4), 242–250. https://doi.org/10.1016/J.APPLANIM.2012.04.010.

Georgiev, A.V., et al., 2013. When violence pays: a cost-benefit analysis of aggressive behavior in animals and humans. Evolut. Psychol. 11 (3), 147. https://doi.org/10.1177/147470491301100313.

Gillette, R., et al., 2000. Cost-benefit analysis potential in feeding behavior of a predatory snail by integration of hunger, taste, and pain. Proc. Natl. Acad. Sci. 97 (7), 3585–3590. https://doi.org/10.1073/pnas.97.7.3585.

Gläscher, J., et al., 2010. States versus Rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66 (4), 585–595. https://doi.org/10.1016/J.NEURON.2010.04.016.

Gomez, P., Ratcliff, R., Perea, M., 2007. A model of the go/no-go task. Journal of Experimental Psychology: General 136 (3), 389.

Gunther, K.E., Pérez-Edgar, K., 2021. Dopaminergic associations between behavioral inhibition, executive functioning, and anxiety in development. Developmental Review 60, 100966.

Hanggi, E.B., 2003. Discrimination learning based on relative size concepts in horses (Equus caballus). Appl. Anim. Behav. Sci. 83 (3), 201–213. https://doi.org/10.1016/S0168-1591(03)00136-9.

Hanggi, E.B., 2005. The thinking horse: cognition and perception reviewed. AAEP Proc. 51, 246–255.

Hanggi, E.B., 2010. Short-term memory testing in domestic horses: experimental design plays a role. J. Equine Vet. Sci. 30 (11), 617–623. https://doi.org/10.1016/J.JEVS.2010.10.004.

Hausberger, M., et al., 2019. Mutual interactions between cognition and welfare: the horse as an animal model. Neurosci. Biobehav. Rev. 107, 540–559. https://doi.org/10.1016/J.NEUBIOREV.2019.08.022.

Hintze, S., et al., 2018. A cross-species judgement bias task: integrating active trial initiation into a spatial Go/No-go task. Sci. Rep. 8 (1), 5104. https://doi.org/10.1038/s41598-018-23459-3.

Huang, Y., Yaple, Z.A., Yu, R., 2020a. Goal-oriented and habitual decisions: neural signatures of model-based and model-free learning. NeuroImage 215. https://doi.org/10.1016/J.NEUROIMAGE.2020.116834.

Huang, Y., Yaple, Z.A., Yu, R., 2020b. Goal-oriented and habitual decisions: neural signatures of model-based and model-free learning. NeuroImage 215. https://doi.org/10.1016/J.NEUROIMAGE.2020.116834.

Huys, Q.J.M., Cruickshank, A., Seriès, P., 2014. Reward-based learning, model-based and model-free. Encycl. Comput. Neurosci. 1–10. https://doi.org/10.1007/978-1-4614-7320-6_674-1.

Jacobs, L.F., Schenk, F., 2003. Unpacking the cognitive map: the parallel map theory of hippocampal function. Psychol. Rev. 110 (2), 285–315. https://doi.org/10.1037/0033-295X.110.2.285.

Krueger, K., et al., 2022. Tool use in horses. Animals 12 (15), 1876. https://doi.org/10.3390/ani12151876.

Ma, I., van Duijvenvoorde, A., Scheres, A., 2016. The interaction between reinforcement and inhibitory control in ADHD: A review and research guidelines. Clinical psychology review 44, 94–111.

Macuda, T., Timney, B., 1999. Luminance and chromatic discrimination in the horse (Equus caballus) [online]. Behav. Process. 44 (3), 301–307. Available at: https://www.sciencedirect.com/science/article/pii/S0376635798000394.

Mcdannald, M.A., et al., 2012. Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur. J. Neurosci. 35 (7), 991–996. https://doi.org/10.1111/J.1460-9568.2011.07982.X.

McLean, A.N., 2004. Short-term spatial memory in the domestic horse. Appl. Anim. Behav. Sci. 85 (1–2), 93–105. https://doi.org/10.1016/J.APPLANIM.2003.09.009.

Morgan, C.L., 1903. An introduction to comparative psychology, New ed., rev. Walter Scott Publishing Co, London, England, 10.1037/13701-000.

Murphy, J., 2009. Assessing equine prospective memory in a Y-maze apparatus. Vet. J. 181 (1), 24–28. https://doi.org/10.1016/J.TVJL.2009.03.028.

Murphy, J., Hall, C., Arkins, S., 2009. What horses and humans see: a comparative review. Int. J. Zool. 14. https://doi.org/10.1155/2009/721798.

Neave, H.W., et al., 2013. Pain and pessimism: dairy calves exhibit negative judgement bias following hot-iron disbudding. [online]. PloS One 8 (12), e80556. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0080556 [Accessed 17 November 2015].

Neill, W.T., Valdes, L.A., Terry, K.M., 1995. Selective attention and the inhibitory control of cognition. Interference and inhibition in cognition 207–261.

O'Keefe, J., Nadel, L., 1979. Précis of O'Keefe & Nadel's The hippocampus as a cognitive map. Behav. Brain Sci. 2 (4), 487–494. DOI: 10.1017/S0140525X00063949.

Osman, M., Meiser, T., Krummenacher, J., 2014. What are the essential cognitive requirements for prospection (thinking about the future)? Front. Psychol. 5, 626. https://doi.org/10.3389/fpsyg.2014.00626.

Raud, L., Westerhausen, R., Dooley, N., Huster, R.J., 2020. Differences in unity: The go/no-go and stop signal tasks rely on different mechanisms. NeuroImage 210, 116582.

Roth, L.S.V., Balkenius, A., Kelber, A., 2007. Colour perception in a dichromat. J. Exp. Biol. 210 (16), 2795–2800. https://doi.org/10.1242/jeb.007377.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M.J., Bullmore, E.T., Sharma, T., Simmons, A., Williams, S.C., Giampietro, V., Andrew, C.M., Taylor, E., 2001. Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. Neuroimage 13 (2), 250–261.

Schachar, R., Mota, V.L., Logan, G.D., Tannock, R., Klim, P., 2000. Confirmation of an inhibitory control deficit in attention-deficit/hyperactivity disorder. Journal of abnormal child psychology 28, 227–235.

Schmidt, M.J., Knemeyer, C., Heinsen, H., 2019. Neuroanatomy of the equine brain as revealed by high-field (3Tesla) magnetic-resonance-imaging. PLOS ONE 14 (4), 213814.

Schultz, W., 2016. Dopamine reward prediction error coding [online]. Dialog-. Clin. Neurosci. 18 (1), 23. Available at: /pmc/articles/PMC4826767/ [Accessed 17 January 2023].

Sitole, S.P., Sup, F.C., 2023. Continuous prediction of human joint mechanics using EMG signals: a review of model-based and model-free approaches. IEEE Trans. Med. Robot. Bionics 5 (3), 528–546. https://doi.org/10.1109/TMRB.2023.3292451.

Skinner, B.F., 1963. Operant behavior. Am. Psychol. 18 (8), 503.

Sutton, R., Barto, A., 2018. Reinforcement learning: An introduction Second. The MIT Press, Cambridge, Massachusetts.

Tiego, J., Testa, R., Bellgrove, M.A., Pantelis, C., Whittle, S., 2018. A hierarchical model of inhibitory control. Frontiers in psychology 9, 1339.

Tolman, E.C., 1948. Cognitive maps in rats and men. [online]. Psychol. Rev. 55 (4), 189–208. Available at: https://psycnet.apa.org/fulltext/1949-00103-001.pdf?auth_token=b6edfaaf11968200612fb9bd81436b291eed4243 [Accessed 17 January 2023].

Tsujimoto, S., Genovesio, A., Wise, S.P., 2011. Comparison of Strategy Signals in the Dorsolateral and Orbital Prefrontal Cortex. J. Neurosci. 31 (12), 4583–4592. https://doi.org/10.1523/JNEUROSCI.5816-10.2011.

Uller, C., Lewis, J., 2009. Horses (Equus caballus) select the greater of two quantities in small numerical contrasts. Anim. Cogn. 12 (5), 733–738. https://doi.org/10.1007/s10071-009-0225-0.

Vikbladh, O.M., et al., 2019. Hippocampal contributions to model-based planning and spatial memory. Neuron 102 (3), 683–693.e4. https://doi.org/10.1016/j.neuron.2019.02.014.

Walton, M.E., et al., 2010. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65 (6), 927–939. https://doi.org/10.1016/J.NEURON.2010.02.027.

Wiebe, K.L., 2003. Delayed timing as a strategy to avoid nest-site competition: testing a model using data from starlings and flickers. Oikos 100 (2), 291–298. https://doi.org/10.1034/j.1600-0706.2003.12046.x.

Wilkinson, G.S., 1992. Information transfer at evening bat colonies. Anim. Behav. 44 (PART 3), 501–518. https://doi.org/10.1016/0003-3472(92)90059-I.

Wunderlich, K., Smittenaar, P., Dolan, R.J., 2012. Dopamine enhances model-based over model-free choice behavior. Neuron 75 (3), 418–424. https://doi.org/10.1016/J.NEURON.2012.03.042.

Zhang, S., Hu, S., Hu, J., Wu, P.L., Chao, H.H., Li, C.S.R., 2015. Barratt impulsivity and neural regulation of physiological arousal. PLoS One 10 (6), e0129139.