RESEARCH ARTICLE

Open Access

Open Peer-Review

Do Ecological Valid Stop Signals Aid Detour Performance? A Comparison of Four Bird Species.

Cite as:

Correspondence:
Anneleen.Dewulf@UGent.be

Anneleen Dewulf [1,2], Clara Garcia-Co [1,3], Wendt Müller [3], Joah R. Madden [4], An Martel [1,5], Luc Lens [1,6] & Frederick Verbruggen[1,2,6]

Recommender:

Reviewers:

[1]Centre for Research on Ecology, Cognition and Behaviour of Birds, Ghent University, BE [2]Department of Experimental Psychology, Ghent University, BE [3]Department of Biology, Behavioural Ecology and Eco-Physiology Group, University of Antwerp, BE [4]Centre for Research in Animal Behaviour, Psychology, University of Exeter, UK [5]Department of Pathobiology, Farmacology and Zoological Medicine, Wildlife Health Ghent, Ghent University, BE [6]Department of Biology, Terrestrial Ecology Unit, Ghent University, BE

This version of the article has been peer-reviewed and recommended by (https://doi.org/)

Abstract

Response inhibition, or the stopping of actions, is considered a key component of flexible and adaptive behaviour. Across fields, response inhibition is often treated as a unitary cognitive mechanism. However, we propose that response inhibition consists of a chain of cognitive processes, including the detection of a stimulus, the selection of an appropriate behaviour (go or stop), and the implementation of it (execution or inhibition of a motor response). From this, we propose that individual variation in response inhibition can arise at the early signal detection stage. This idea was tested in a detour barrier task, which is one of the most popular tools to study response inhibition in non-human animals. The role of signal detection in detour tasks has been largely neglected, with a few notable exceptions. We therefore partially replicated two previous studies that manipulated the perceptual characteristics of the barrier, while addressing some conceptual and methodological shortcomings of the original work. Specifically, we compared how detour performance of four bird species (i.e., white leghorn chickens, Japanese quails, herring gulls and domestic canaries) is differently influenced by vertical- and horizontal-barred barriers. In contrast to the previous work, performance on the detour task did not improve when the perceptual characteristics of the barrier matched the ecological niche of the species. However, all species, showed some level of learning, as evidenced by shorter detour latencies (except in herring gulls) and fewer persisting attempts. These findings highlight the need for replication studies and emphasize on the importance of improving methodological and conceptual design factors to further investigate the underlying mechanisms of response inhibition in animals. Preregistered Stage 1 protocol: https://osf.io/qvxgh (date of in-principle acceptance: 20/03/2023).

Keywords: Response Inhibition; Stop-Signal Detection; Comparative Approach; Birds; Detour

Introduction

Response inhibition (RI) refers to stopping or cancelling actions that are no-longer relevant, inappropriate, or overly risky (Verbruggen and Logan, 2008b, 2017). It is often regarded as a critical component of flexible and adaptive behaviour (Verbruggen and Logan, 2008b). For example, animals living in urban environments must often inhibit no-longer relevant behaviours when confronted with environmental conditions that differ significantly from their ancestral ones (Lee and Thornton, 2021); lower-ranked animals need to inhibit inappropriate, disobedient behaviour in the presence of dominant animals (Johnson-Ulrich and Holekamp, 2020); and foraging animals must refrain from approaching a food source when this action becomes overly risky due to the emergence of a predator (Tvardíková and Fuchs, 2012). These examples demonstrate that RI (or a lack thereof) can have important fitness consequences (e.g., the animals may be predated if they fail to stop foraging when the predator emerges).

One of the most popular tasks to study RI in animals is the detour task (Kabadayi, Jacobs, et al., 2017; Miller and Tallarico, 1974; Regolin, Giorgio, et al., 1995; Regolin, Vallortigara, et al., 1994; Van Horik, Beardsworth, Laker, Whiteside, et al., 2020; Van Horik, Langley, et al., 2018; Zucca, Antonelli, et al., 2005; Zucca and Sovrano, 2008). In this task, the direct path to a motivationally-salient stimulus (e.g., food or a social companion) is blocked by a barrier or cylinder. Animals have to inhibit their prepotent response to go directly for the reward (as the direct path is blocked), and instead make a detour around the barrier or cylinder to obtain the reward. Detour tasks have been used in non-human animals, such as birds, to study e.g., how the social or physical environment shapes RI. For example, wild Australian magpies (*Gymnorhina tibicen*) demonstrated superior detour performance when reared in large compared with small social groups (Ashton et al., 2018). Another study found that pheasants (*Phasianus colchicus)* showed superior detour performance when reared in spatially unpredictable compared with predictable environments (Van Horik, Beardsworth, Laker, Langley, et al., 2019). Combined, these findings suggest that RI development is facilitated in e.g., environments with high social demands or environments that promote the expression of diverse foraging strategies.

Typically, performance in the detour task has been linked to the variation in the effectiveness of a single cognitive control function, 'response inhibition', or more generally, 'inhibitory control' (which is an umbrella term for various types of inhibition, which may or may not be related to each other; Bari and Robbins, 2013). However, by referring to general ill-defined cognitive constructs such as RI (or even worse, a general umbrella term such as 'inhibitory control'), we do not explain the underlying cognitive mechanisms or building blocks of RI (Verbruggen, McLaren, et al., 2014), as the explanation is 'just as mysterious as the thing it is supposed to explain' (Press et al., 2022). To address this issue in the human RI literature, a theoretical framework of RI has been proposed (Verbruggen, McLaren, et al., 2014). Based on empirical work in humans, primates, and rodents, the authors of the framework proposed that RI involves a chain of processes. More specifically, RI would involve at least three basic processes: the detection of a 'stop signal' (detection), the stochastic accumulation of information (selection), and suppression of the motoric output (implementation). Furthermore, these core processes can be modulated by a set of processes that take place on shorter (seconds, minutes, hours or days) and longer (months or years) timescales. Depending on the species, this can involve, e.g., outcome monitoring, anticipatory adjustments, and both short-term and long-term learning. Here we argue that some of these cognitive processes play a role in RI across species (without assuming a one-to-one mapping for the full processing chain). In particular, in the present study we propose that one of these core processes, namely stop-signal detection, is a crucial (but largely ignored) building block of RI across species, including avian species.

The Crucial Role of Stop-Signal Detection

Several lines of evidence indicate that signal detection may play a critical role in RI, particularly in humans and non-human primates. ~~(at least in humans)~~. For example, several behavioural studies reveal that RI is impaired when visual distractors occur in the environment (Verbruggen, Stevens, et al., 2014), or when stop signals are hard to perceive (Van Der Schoot et al., 2005). Neurophysiological and computational work also demonstrated that early perceptual processing of potential stop signals (which could be e.g., an obstacle, or in case of humans, a red light) determines to a large extent whether individuals can successfully inhibit a response or not (Bekker et al., 2005; Boehler et al., 2009; Elchlepp and Verbruggen, 2017; Pani et al., 2018; Salinas and Stanford, 2013).

Thus, it appears that RI may largely depend on the outcome of perceptual processes. However, the crucial contribution of these processes to successful RI is rarely acknowledged or studied in the non-human animal cognition domain, with a few notable exceptions. For example, researchers found that avian RI was improved when the visibility of a stop signal (i.e., a predator) was improved (e.g., when the predator occurred against a white background, in bright light, or in short grass) (Devereux et al., 2006; Nebel et al., 2019). Other studies suggested that RI in the detour barrier task is affected by the perceptual characteristics of the barrier (i.e., the type of stop signal). For example, Regolin and colleagues (Exp 1) (1994) exposed 2-day old white leghorn chickens *(Gallus gallus domesticus)* to a variety of barrier types. These included a barrier with vertical bars or stripes, and a barrier with horizontal bars. The authors found that RI performance was impaired (i.e., the time required to successfully detour around the barrier) when faced with vertical-bar barriers compared with horizontal-bar barriers. Both vertical- and horizontal-bar barriers occluded the reward behind the barrier to a similar degree (i.e., 20% compared with a fully transparent barrier). Thus, the differences between these two barrier types cannot be attributed to differences in physical reward occlusion. Instead, the authors suggested two alternative potential explanations for this asymmetrical effect, namely (1) the degree of subjective occlusion and (2) the ecological validity of stop signals.

First, despite equal reward occlusion by each barrier type, the behavioural repertoire of ground-moving animals consists primarily of horizontal movements (e.g., walking, running). Consequentially, these animals can 'subjectively' perceive a reward as less occlusive (i.e., more visible) with vertical- than horizontal-bar barriers (making it harder to inhibit the response to go directly for the reward) (Kabadayi, Bobrowicz, et al., 2018). However, follow-up experiments in which the occlusion of the reward was directly manipulated, were inconsistent with this 'subjective occlusion' account (Regolin, Vallortigara, et al., 1994). Second, differential performance between vertical- and horizontal-bar barriers might be due to the ecological niche of the species. Gallinaceous birds such as chickens are mainly terrestrial animals that have occupied niches that consist of penetrable long grass and twigs. Regolin et al. (1994) therefore argued that it might be harder for gallinaceous birds to detour around vertical- than horizontal-bar barriers, as the former would mimic the penetrable vertical vegetation of their ecological niche (whereas in the detour task, the vertical-bar barrier is of course, not penetrable).

Zucca and colleagues (2005) further investigated this ecological-niche hypothesis by comparing detour performance in another gallinaceous bird species, hybrid (Japanese) quails *(Coturnix coturnix x C. japonica)*, with performance in two species with a substantially different ecological niche, namely yellow-legged gulls *(Larus michahellis)* [1] and domestic canaries *(Serinus canaria)*. They used a variant of the detour task with multiple compartments and again, vertical- and horizontal-bar barriers (14% reward occlusion compared with a transparent barrier[2]). In this study, the authors considered both probability of a correct response (i.e., going to the

---

[1] Zucca and colleagues (2005) mention in their paper that they tested herring gulls, but used the scientific name of the Caspian gull *Larus cachinnans*. The taxonomy of the genus Larus has been updated over the years. Given that the authors mentioned that the gulls were from a breeding colony in Trieste (Italy), we assume they tested yellow-legged gulls, which were formerly treated as a subspecies of the herring gull (as was the Caspian gull).

[2] Zucca and colleagues (2005) mention that they used a 23 x 26 cm barrier, barrier lines/rods with a width of 0.2 cm, and in-between

correct compartment during their first attempt) and the latency to detour as measures of RI. They found that the detour accuracy for quails was lower (i.e., RI performance was impaired) for vertical- than for horizontal-bar barriers. This seems consistent with the findings of Regolin and colleagues (1994), although it should be noted that Zucca et al. (2005) only found a significant effect for detour accuracy but not for the latency to detour, the measure of RI in the study of Regolin et al. (1994). For yellow-legged gulls, detour accuracy was not influenced by barrier type, but detour latency was. Specifically, the latency to detour was longer (i.e., RI was impaired) for horizontal- than vertical-bar barriers. Again, the authors attributed this to the species' ecological niche. Specifically, Zucca et al. (2005) argued that in the (original) ecological niche of young, yellow-legged gulls, chicks are accustomed to consider the vertical ground vegetation of sand dunes as largely impenetrable. According to the authors, it might therefore be harder for (young) gulls to detour around horizontal- than vertical barriers, as the latter would be perceived as less penetrable (note that Zucca et al. (2005) tested juvenile gulls that could not fly yet). Lastly, canaries were unable to detour around the barrier, although they made several attempts to fly over the barrier demonstrating that they were sufficiently motivated. Zucca et al. (2005) therefore hypothesized that the detour task is not considered to be a real problem for the two-month old canaries. The authors suggested that, after all, canaries are aerial birds, allowing them to tackle obstacle problems by simply flying over them (but which was not possible in the detour task due to the dimensions of the used apparatuses).

In sum, the results of Regolin et al. (1994) and Zucca et al. (2005) indicate that the characteristics of the 'stop signal' matter in the detour task, potentially shedding new light on RI in avian species. However, some concerns can be raised about certain features of the previous studies (which are summarized in Table 1). First, the sample size was low (at least for some species), the studies (inconsistently) used within- and between-species designs, the number of trials per barrier type differed within and between species and the number of sessions per barrier type fluctuated between species (e.g., yellow-legged gulls received three sessions per barrier type spread over three days, while hybrid quails received one session per barrier type). The latter two issues are also problematic from a conceptual point of view, as previous work indicates that learning will influence RI (Verbruggen, Best, et al., 2014; Verbruggen and Logan, 2008a), including in the detour task (Van Horik, Beardsworth, Laker, Whiteside, et al., 2020). Second, both studies used less-common variants of the detour task, which complicates comparisons with the wider literature. Additionally, they used less common, hard to standardize (social) rewards, which complicates between-species comparisons of RI behaviour. Similarly, the large differences in developmental trajectories and the lack of adapting the test apparatus to the morphological differences between species, also complicated the between-species comparisons.

A Partial Replication of Regolin et al. (1994) and Zucca et al. (2005)

To prove the significance of the previous study, our study investigated the role of stop-signal detection in avian RI by partially replicating the studies of Regolin et al. (1994) and Zucca et al. (2005). The importance of the original studies is indubitable, as they are one of the few studies that aimed to 'deconstruct' avian RI performance by focusing on the underlying cognitive processes (in this case, stop-signal detection). Additionally, Zucca et al. (2005) implemented a comparative approach to investigate whether differences in how the stop signal might be perceived by different species could contribute to variation in RI.

In our partial replication, we made several changes to address commonly raised concerns in the detour literature (including the concerns raised in the previous section, see Table 1). First, we directly compared four species (white leghorn chickens, Japanese quails, herring gulls (*Larus argentatus* [3]) and domestic canaries), in a well-powered mixed design analysis with *Species* as between-species factor, and *Barrier* (vertical-bar vs.

---

line/rod gaps of 1.25 cm. Given these measurements, only a combination of 18 rods/lines and 18 in-between gaps is possible, which is the equivalent of a vertical- or horizontal-bar barrier with a 14% reward occlusion.

[3] Given the high availability of the herring gull at the North Sea coast (Belgium), herring gulls were selected to represent the gull species in this comparative study. But see also footnote 1.

Table 1. Methodological features of the Regolin et al. (1994), the Zucca et al. (2005) and the present study.

Source	Regolin (1994)	Zucca (2005)			Current Study
1. Methodology					
Species	**White leghorn chicken**	**Hybrid quail**	**Yellow-legged gull**	**Canary**	**All four species**
Design	Between	Between	Within	Within	Mixed
Total Sample	750 (250) [1]	90	5	26	240 (60/species)
Sample per Barrier Type	102 (34) [1]	18	5	26	60/species
Trials per Barrier Type	1	10	10	1 or 10 [2]	3
Sessions per Barrier Type	1	1	3	1	1
2. Detour Task	Two compart.	Four compart.	Four compart.	Four compart.	Simple
3. Reward	Cagemates	Reflection	Reflection	Reflection	Food
4. Baseline covariate	No	No	No	No	Yes
5. DV's	Latency	Latency Accuracy	Latency Accuracy	Latency Accuracy	Latency Persistence
6. Enclosure					
Social density	3	1	5	5	± 10
Fence	NA	Vertical	Bricks	Vertical	Mesh netting
7. Test Age	2 days	1 M	10-25 days	4-6 M	Species specific
8. Apparatus					
Test box: L x W	120 x 35	150 x 75	150 x 75	150 x 75	Scaled/species
Test box: H	60	40	40	40	Barrier H
Barrier-Entry Distance	27	27	27	27	Scaled/species
Barrier: L x H	10 x 20	23 x 26	23 x 26	23 x 26	Scaled/species
Barrier line: W	0.3	0.2	0.2	0.2	Scaled/species
Gap between barrier lines	1.2	1.25	1.25	1.25	Scaled/species

Note. DV's = Dependent variables, [1] *= Animals were reared and tested in groups of three. The means of each trio was used as individual data for the subsequent analysis;* [2] *= 6/26 canaries received 1 trial/barrier, 20/26 canaries received 10 trials/barrier. Measurements are in cm.*

horizontal-bar barrier) as within-species factor. Hereafter, each species will be referred to by its common name for clarity: chicken (excluding 'White Leghorn'), quail (excluding 'Japanese'), gull (excluding 'herring'), and canary (excluding 'domestic'). All species were given an equal amount of trials and sessions per barrier type (see below). Second, the perceptual characteristics of the barrier (i.e., vertical-bar vs. horizontal-bar barriers) were manipulated in a simple detour barrier task (which is the most common variant of the detour problem; (Kabadayi, Bobrowicz, et al., 2018), rather than a four- (Zucca, Antonelli, et al., 2005) or two-compartment (Regolin, Vallortigara, et al., 1994) detour task). See Figure 1, for an overview of the designs. Third, the unconditional reward was food instead of a social stimulus (as in Regolin, Vallortigara, et al., 1994 and Zucca, Antonelli, et al., 2005). Food is a common reward in laboratory tests and has a high incentive value across species and individuals. Furthermore, it's subjective value can be better standardised both within and between species compared with social rewards. Fourth, non-cognitive, motivational states can influence detour performance (Kabadayi, Krasheninnikova, et al., 2017; Van Horik, Langley, et al., 2018). Therefore, we collected for each individual a 'multi-baseline' measure of their general motivational state (which could be a combination of, e.g., non-transparent obstacle neophobia, test box neophobia, food motivation, or motivation to explore). This 'multi-baseline' measure was obtained with an opaque barrier during habituation (see below). We included this as a covariate in our statistical models to increase the likelihood of detecting barrier type effects within species conditional on/adjusted for the 'multi-baseline' measure of an individual's general motivational state[4]. Fifth, our study considered two measures of interest, namely the latency to detour (Regolin, Vallortigara, et al., 1994; Zucca, Antonelli, et al., 2005) and the time spent in proximity to the barrier (persistence). The last variable was not included in the original studies, but adds substantial information about variation in (un)successful inhibition following the different barrier types. Note that this measure also captures 'accuracy', as all birds that did not peck at the barrier (i.e., an accurate response) got a score of 0, whereas all birds that pecked, got a score > 0. Sixth, for all species, (fledged) chicks were raised in groups of approximately 10 individuals, as variations in detour performance had already been reported with fluctuating social group sizes (Ashton et al., 2018). Furthermore, mesh netting was used for the enclosures of all species (preventing variation in experiences with (im)prenetrable vertical- or horizontal-bar objects in the enclosures). Seventh, detour performance of the different species was compared when they are on similar levels in their developmental trajectory (see e.g., Kabadayi, Jacobs, et al. (2017), Kabadayi, Krasheninnikova, et al. (2017) and Verbruggen, McLaren, et al. (2014) for the influence of cognitive maturation on RI), and again, with similar experiences in the enclosure, keeping in mind the precocial-altricial spectrum (see below). Finally, the size of the test apparatuses and barriers (including the width and the in-between line gaps) was scaled based on the morphological characteristics of each species (see below). For example, Zucca et al. (2005) used the same test box for all three species, and argued that this was appropriate because they tested species at different ages. Nevertheless, one could still expect substantial differences in body size (e.g., a one-month old quail is substantially larger than a two-month old canary). As (relative) distance towards the reward influences detour performance (Kabadayi, Bobrowicz, et al., 2018), it was therefore important to control for this as well.

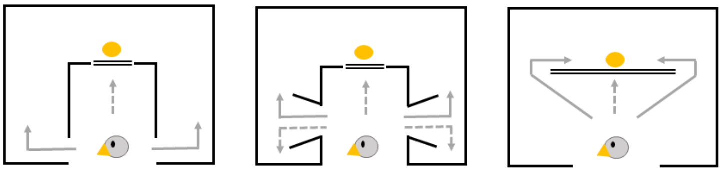

Figure 1. A display of the three detour task variants: the two-compartment detour task (left, Regolin, Vallortigara, et al., 1994) the four-compartment detour task (middle, Zucca, Antonelli, et al., 2005) and the simple detour barrier task (right, current study). Double line: barrier; full arrows: correct responses; dashed arrows: incorrect responses.

Predictions

First, we predicted better RI performance for ecologically valid compared with non-valid stop signals, as should be reflected in shorter latencies to detour and less time spent persisting. As the (original) ecological niche of our species substantially differs (chicken and quails: penetrable vertical terrestrial vegetation; gulls: impenetrable vertical vegetation of sand dunes; canaries: aerial environment), the ecological validity of stop signals would be species-specific. Specifically, for chickens and quails, we expected better detour performance for horizontal- (HB) compared with vertical-bar barriers (VB; thus, detour performance HB > VB). We expected the opposite pattern for gulls (i.e., detour performance HB < VB). Finally, based on the findings of Zucca et al. (2005), we did not expect differences between vertical- and horizontal-bar barriers for canaries (i.e., detour performance HB = VB). Overall, this pattern should result in a statistical interaction between *Barrier* and *Species* (**Prediction 1**).

---

[4] The original studies administered a variety of barrier types, including a transparent barrier, which was then also used to interpret performance with the other barrier types. Yet, we opted to exclude transparent barriers from our design for three reasons. First, the comparison of transparent and barred barriers tests another hypothesis, namely the role of reward occlusion, rather than the ecological validity of the barrier type (which can only be tested by comparing vertical- and horizontal-bar barriers). Second, performance in the detour task can be influenced by differential individual experiences with transparent obstacles (Kabadayi, Krasheninnikova, et al., 2017; Van Horik, Langley, et al., 2018). Last, omitting the transparent barrier reduces the contact of wild gulls to humans, which is important as gulls are wild animals and will be raised as such, prior to their post-test release back to nature.

Second, as each session consisted of three trials (of the same barrier type), we also looked at how detour performance improved within each session. Based on previous studies, we predicted that detour performance should improve across trials within a session (**Prediction 2**).

Furthermore, we explored if the learning effect (i.e., improved detour performance across trials) interacts with the ecological validity of the stop signals. There are two possible patterns that would result in a three-way interaction between *Species*, *Barrier* (horizontal- vs. vertical-bar barriers), and *Trial* (1-3) (**Explorative Prediction 3**). First, detour performance might be better for ecologically valid compared with non-valid stop signals at the beginning, but this pattern might diminish over time as individuals learned to stop (i.e., the differences between barrier types would decrease). Second, detour performance might be poor at the beginning for both barrier types, but learning to stop might be easier for ecologically valid signals compared with non-valid stop signals (i.e., the differences between barrier types would increase). Both patterns would be theoretically meaningful, but we did not have *a-priori* predictions about the direction of the three-way interaction.

Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Subjects and Housing

Quails, gulls and chickens were raised and tested at the avian research facilities of Ghent University (Lab number LA1400452), located at the Wildlife Rescue Center (WRC) in Ostend (Belgium). The canaries were raised and tested at the avian research facilities of the University of Antwerp (Lab number LA1100161) in Wilrijk (Belgium).

Sample size

We originally registered to test 60 individuals per species. *A-priori* power sensitivity analyses done in G*Power (Faul et al., 2009) indicated that this was sufficient to detect small effects; it was also the largest number that was practically feasible[5]. For the sensitivity analysis, we used a mixed ANOVA model with one between-subjects factor (4 levels; corresponding to our *Species* factor) and two within-subjects factor (one with 2 levels – *Barrier* – and one with 3 levels – *Trial*). This indicated that our sample size of 60 animals per species (240 in total) was sufficient to detect a *Species* x *Barrier* interaction effect (**Prediction 1**) with a small effect size (Cohen's *f* effect size of 0.12; Cohen, 1988; Power = .80; cor. among RM = 0.5; we used an alpha of .025 to correct for the fact that we had two dependent variables measuring (slightly) different aspects of detour performance). Second, our sample size was sufficient to detect a small effect of *Trial* (**Prediction 2**; Cohen's *f* effect size of 0.09; Cohen, 1988). Third, our sample size was sufficient to *explore* a small effect (Cohen's *f* effect size of 0.09; Cohen, 1988) for the *Species* x *Barrier* x *Trial* interaction effect (**Explorative Prediction 3**). Due to higher than expected post-hatch mortality, the quail sample size was slightly reduced to 58. This reduction did not result in meaningful differences in effect size for our three predictions [6].

Our sensitivity analyses were based on mixed ANOVAs (fixed-effects models with between- and within-species factors). However, as discussed below, we analysed our data with (G)LMMs, which are currently not covered by G*Power or most other power-estimation tools. These mixed-effect models are more flexible in assigning variance as they allow for the specification of both fixed and random effects. However, by accounting for unexplained variance (see below), our proposed mixed effect models are more powerful than the fixed-effect model ANOVAs used in our sensitivity analyses (and than the models used in the studies of

---

[5] Farrar et al. (2020) mention in their paper on replications in comparative cognition that power analyses are not the golden standard in this research domain, and 'in many cases comparative cognition researchers could be better off performing design or sensitivity analyses based on their resource constraints.'

[6] This adjustment was made after data collection but prior to any data analysis. These changes are still considered part of Stage 1 IPA and were pre-approved.

Regolin, Vallortigara, et al., 1994, and Zucca, Antonelli, et al., 2005). Thus, the sensitivity analyses discussed here are a conservative estimate.

White leghorn chickens and Japanese quails

Chicken and quail eggs were obtained from local breeders in Belgium. At the WRC, the eggs were incubated in Brinsea Ova-Easy incubators (temperature = 37.5°C; humidity = 45% for first 15 [quail] or 17 [chicken] days, after which humidity = 70% till hatching). Once hatched, chicks received a unique colour ring combination prior to being housed in groups of ± 10 chicks[7] per indoor enclosure (size = 1m x 1m x 2m; L x W x H; ambient temperature = 15-25°C; humidity = 40%-80%; photoperiod = 14:10 L:D; type of wire fencing = mesh netting). Birds were *ad libitum* provided with a chicken meal mixture (Aveve Chicken Start Mash) and water. Shelter, additional heating panels (30 x 30 cm; till Day 7), and pecking objects (pine cones) were available. The (precocial) chickens and quails were tested at ±3 weeks (see below for justification of species-specific test age). After testing, the individuals were euthanized by certified staff.

Herring Gulls

Gull eggs were collected by the 'Agentschap voor Natuur en Bos' and the gull patrol team in Ostend (https://www.oostende.be/meeuwen) who are authorized to remove gull eggs along the Belgium coast for various reasons. The eggs were collected prior to pipping, and were safely transported to the WRC. At the WRC, the eggs were further incubated in Brinsea Ova-Easy incubators (temperature = 37.5°C; humidity = 45%) and checked twice per day for signs of pipping. When gull embryos reached the pipping stage, they were placed in a hatchery (temperature = 37.2°C; humidity = 50%). Once hatched, the semi-precocial gull chicks received a unique colour ring combination prior to being placed in boxes with netting bottoms (size = 1.20m x 0.60m x 0.60m; L x W x H) within heated rooms (ambient temperature = 15-25°C; humidity = 40%-80%; typical photoperiod = the natural photoperiod at the latitude of Belgium; type of netting = grid) for 5 days (and till their body mass exceeded 60 grams). During this period, the gulls were hand-fed small pieces of fish and dog pellets (soaked in water), supplemented with Akwavit (Kiezebrink Focus on Food, The Netherlands). We also provided one heating panel per box. After this initial indoor period in the boxes, the gull chicks were housed in groups of ± 10 individuals[7] per outdoor enclosure (size = 5 m x 1,95 m x 2,65; L x W x H), type of wire fencing = mesh netting), with an extra heating panel for the first couple of days (note that the exact number of days depended on the weather conditions). Food (a mixture of 75% dog food soaked in water and 25% defrosted fish, supplemented with Akwavit) was provided four times per day (the default policy at the WRC); water was provided *ad libitum*. The (semi-precocial) gulls were tested when they were approximately 3 weeks old (see below). After testing, gulls were moved to large flight cages to dehabituate them from human handling (and hence improve their survival rates). They were released into the wild when they were approximately 8-10 weeks old.

Domestic canaries

*D*omestic canaries (of the Fife Fancy type) were obtained from long-term, breeding populations at the Department of Biology ('Behavioural Ecology and Ecophysiology' research group) of Antwerp University. Canaries are altricial species, and nestlings are thus highly dependent of their parents for food. Therefore, chicks were only separated from their parents at the end of the nestling period (i.e., when they were ± 25 days old) (Garcia-Co et al., 2024)[8]. At this point, the canaries were moved in groups of ± 10 individuals[7] to indoor aviaries of

---

[7] This adjustment was made after data collection but prior to data analysis. These changes are still considered part of Stage 1 IPA and were pre-approved. Additionally, an analysis incorporating group size as an extra random variable in the statistical models is provided in the supplementary materials (Dewulf, Garcia-Co, et al., 2023).

[8] During the nestling period, birds were kept in cages with vertical wires (the default policy at the University of Antwerp). However, we believe that this does not affect the interpretation of our results. First the birds did not interact with the cage during the nestling period, so they could not learn whether the bars or strips were impenetrable or not. Second, Zucca et al. (2005) argued that detour performance is improved when the perceptual characteristics of the stop signal match the ecological niche (rather than the prior experience) of the

Antwerp University (size : 1m x 2m x 2m; L x W x H; ambient temperature = 15-25°C; humidity = 40% - 80%, photoperiod = the natural photoperiod at the latitude of Belgium; type of wire fencing = mesh netting). The canaries were marked with a permanent marker for individual recognition at hatching, and ringed with a closed metal ring when their body mass exceeded the predetermined threshold of 7 g. Upon arrival at the indoor aviaries, canaries received a unique number-color ring combination (the default policy at the University of Antwerp). In the aviaries, canaries were provided with canary seed mixture and egg food (van Camp, Belgium), water, shell grit, and cuttlefish bone *ad libitum*. They were tested at 7 weeks (approximately three weeks after fledging; see below). After testing, canaries returned to their local breeding population in the University of Antwerp.

A Comparative Testing Age

Our previous work (Troisi et al., 2024) indicates that 3 weeks is an ideal testing age for large gulls (incl. herring gulls) in detour tasks and other related cognitive tests. Gulls are semi-precocial, but only require hand-feeding for the first couple of days (and most start eating independently after two/three days). Furthermore, the gull chicks can already move around (and explore their environment) from Day 1. Chickens and quails are precocial, which implies that they can feed independently and explore their environment from Day 1. Given the overall similarities, we therefore tested chickens, quails and gulls when they were approximately 3 weeks old (i.e., habituation happened $\approx$ on day 16-18; testing happened on $\approx$ day 19-20). By contrast, canaries only become independent when they are approximately 25 days old (see previous subsection). At this point they were moved to larger enclosures and housed in groups. To ensure that the altricial canaries had a similar (15-day) experience with their enclosure and their cagemates as the (semi-)precocial species, habituation and testing of canaries was delayed with 25 days (i.e., habituation happened $\approx$ on day 41-43; testing on $\approx$ day 44-45)

Apparatus

For each species, the test apparatus consisted of a two door start box, a test box, a barrier, and a feeding bowl. Performance of the birds was monitored using a camera placed centrally at the top of the testing arena (Sony Action Cam HDR-AS50). In the test box, a vertical- or horizontal-bar barrier blocked the direct path to the unconditional reward (i.e., the food in the bowl) that was immediately placed behind the barrier. The species-specific unconditional food reward (chickens and quails: chicken meal, gull: dog pellets and fish, canaries: canary seed mixture and egg food) consisted of clearly visible food, placed in a coloured bowl. For chickens and quails, these were coloured green and yellow (brand: Junai, The Netherlands); for gulls and canaries, these were coloured orange-brown (brand: Elho, Belgium) [9]. To avoid satiation after the first trial on test trials (see next section), the pile of food was largely covered with a transparent perspex cover, with only a small bit of accessible food placed on top of the cover. The vertical- and horizontal-bar barriers were made of transparent perspex on which 18 vertical and horizontal lines, respectively, were painted per species (see below). To prevent canaries from flying over the barrier (as as alternative way of avoiding the barrier), floor-to-ceiling barriers were used for all species.

The size of the test apparatus was adjusted per species. In a recent study from our lab, we tested gulls in a detour task (akin to the task proposed here, but with transparent and non-transparent barriers). In this study, the starting box was 35 x 20 x 26 cm (L x W x H), and the test box was 145 x 88 x 132 cm (L x W x H). The barrier was 40 x 40 cm (L x H), and was placed 50 cm from the start box entrance (with approximately 24 cm between the edges of the barrier and the sides of the test box). In the present study we used the

---

species. Notably, although the canaries in Zucca et al. (2005) were reared in cages with vertical bars from two months until tested at four to six months, performance did not differ between barrier types.

[9] Potential variations in colour perceptions and sensitivities across species are mitigated by the developed preference for their respective food bowls prior to the start of the experiment through repeated exposure (i.e., 10-day habituation to the food bowl in the enclosure) and learning (pairing of the coloured bowl with food during these 10 days).

same set-up for the gulls, and re-scaled all values based on tarsus length at testing age (see Table 2 for the values for each species). For the chickens, quails and gulls we used the growth curves (Figure 2) for tarsus length reported in previous studies (Dudusola and Bashiru, 2020; Troisi et al., 2024; Yeasmin and Howlider, 2013). For canaries, no such growth curves were available. However, in a recent study, Garcia-Co et al. (2024) measured tarsus length at day 25. Given that morphological traits (incl. tarsus) seem to plateau at a similar moment in the lifespan of a canary, we used this tarsus measure at day 25 as our measure for the tarsus length at testing age. In addition, the black painted barrier lines (18 in total per species) occluded the food reward by approximately 14% (Zucca, Antonelli, et al., 2005). As a consequence, the width of the barrier lines (and of the in-between gaps) as described in the study Zucca and colleagues (2005) was adjusted to the re-scaled barrier size dimensions per species (see Table 2).

Table 2. The upper table shows the tarsus length and species-specific re-scaled test apparatuses based on the herring gull detour set-up of Troisi et al. (2024). The lower table shows the re-scaled width of the barrier lines (and of the in-between gaps) for each species based on the detour set-up of Zucca et al. (2005)

	White leghorn chicken	**Japanese quail**	**Herring gull**	***Fife Fancy* Canary**
Source	Yeasmin (2013)	Dudusola (2020)	Troisi (2024)	Garcia-Co (2024)
Test age (days)	19	19	19	44
Sample size	130	2591	42	69
Mean Tarsus (mm)	30.71	25.32	55.56	17.81
Ratio	1.81	2.19	1	3.12
Apparatus (rescaled)				
Start box: L x W	19.35 x 11.05	15.95 x 9.11	35 x 20	11.22 x 6.41
Test box: L x W	80.15 x 48.64	66.08 x 40.10	145 x 88	46.48 x 28.21
Barrier-Entry Distance	27.64	22.79	50	16.03
Barrier: L x H	22.11 x 22.11	18.23 x 18.23	40 x 40	12.82 x 12.82

	White leghorn chicken	**Japanese quail**	**Herring gull**	***Fife Fancy* Canary**
Barrier line: W	0.17	0.14	0.31	0.10
Gap between barrier lines	1.06	0.87	1.91	0.61

Note. The estimated tarsus length at testing age for (semi-)precocial species, is derived from a linear equation using the two nearest measuring points for a mixed-sex sample (except for white leghorn chickens, where an additional average was calculated over pullets and cockerels). For re-scaling the test apparatuses, the unrounded factor per species was utilized. Unless specified otherwise, measurements are in cm.

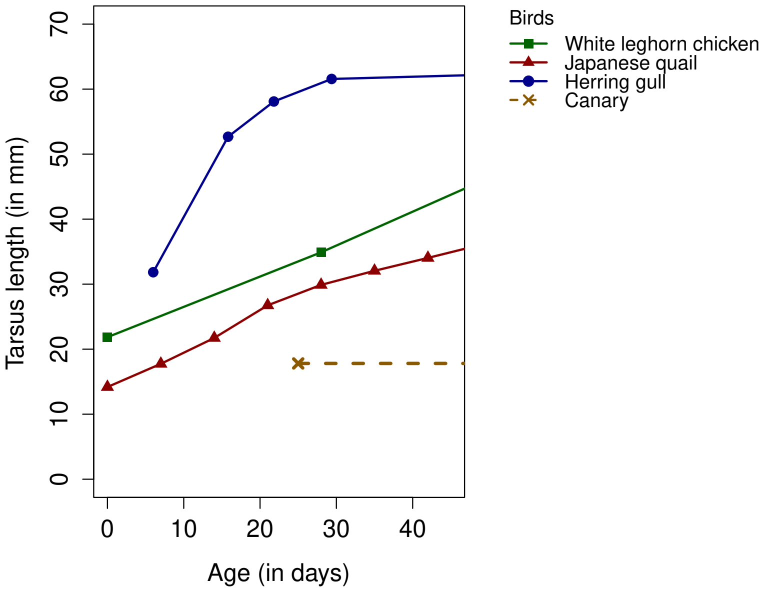

Figure 2. Full lines: reported tarsus growth during early life for white leghorn chickens (Yeasmin and Howlider, 2013), Japanese quails (Dudusola and Bashiru, 2020) and herring gulls (Troisi et al., 2024). Dashed lines: hypothetical tarsus length for canaries based the assumption that tarsus length does not change (much) after fledging (Garcia-Co et al., 2024)

Procedure

Prior to the start of the experiment, birds were habituated for 10 days in their enclosure to feed from a coloured food bowl, which was identical to the bowl used during both habituation and testing in the test box. For all species, the feeder was placed on the ground, to simulate ground feeding during the test. When they reached the appropriate age (see above), animals were tested for five consecutive days (i.e., three habituation and two testing days). Food in the enclosures was provided *ad libitum*, but the evening before an individual's habituation or testing day, the feeders were removed from the enclosures at 6PM (after the last feeding time). This created a non-feeding period during the night (which is normal and also happens in non-experimental conditions), followed by (shortly) delayed feeding in the morning to prevent birds from overindulging prior to habituation or testing. This is in line with other studies using the same species (chicken: e.g., Bollweg and Sparber, 1998; quail: e.g., Ueno and Suzuki, 2014 and unpublished data from our lab; gulls: e.g., Dewulf, Knoch, et al., 2025; canaries: e.g., Müller et al., 2008). After all individuals of one enclosure completed the habituation or testing trials for the day, food was again provided *ad libitum*.

On the three habituation days (08:00 AM - 10:30 AM), each bird received 1 trial per day where it could freely explore the test box and feed from a centrally placed coloured food bowl. During the second and third habituation day, an opaque barrier was placed just behind the coloured food bowl. This allowed us to obtain a 'multi-baseline' measure of an individual's general motivational state (which could be a combination of e.g., non-transparent obstacle neophobia, test box neophobia, food motivation, motivation to explore; see below). The current habituation set-up (i.e., the food bowl *in front* of the barrier) was designed in such a way that acquiring a motor routine during habituation was unnecessary and could not confound subsequent detour performance with the barred barriers (Van Horik, Beardsworth, Laker, Whiteside, et al., 2020).

Each bird participated in one session per day on the two testing days (10:30AM - 02:30 PM), ~~each bird performed one session~~. Each session consisted ~~ing~~ of 3 trials with one barrier type. The order of barrier type (i.e., horizontal-bar or vertical-bar barrier) was pseudo-randomized within and between species, across the two testing days.

Due to the natural breeding season of the wild gull and the canary, birds hatched non-simultaneously. In order to guarantee an appropriate test age (see above), we grouped individuals of a similar age per enclosure; and then habituated or tested birds per enclosure (by taking into account the average age of the enclosure). Although there was no fixed breeding season for quails and chickens, incubation happened in 'batches' (due to reduced egg production/supply). As a result, an identical grouping procedure within these species was applied.

At the beginning of each trial, each bird was gently placed in the dark two-door start box. The trial started when the researcher opened the first non-transparent cardboard door of the start box. This permited the bird to see the test arena but not access it. After 15 seconds, the second, transparent door of the start box was opened and the bird could enter the test box. If the bird did not exit the start box within 30 seconds, it was gently pushed forward (by sliding the back of the starting box forward; Troisi et al., 2024). The habituation trials ended when the individual ate from the food bowl for 30 seconds or when the maximum trial time had been reached (i.e., 5 min 15 seconds). The test trials ended immediately when the individual ate from the food bowl (to avoid food satiation on subsequent trials) or when the maximum trial time had been reached (i.e., 2 min 15 seconds). Maximum trial times during habituation were longer than during testing, as the main goal of the habituation was to familiarize each bird with the test material (and obtain a 'multi-baseline' measure of an individual's general motivational state). The maximum duration of a test trial was 2 minutes (after an additional 15 seconds inside the start box with the second, transparent door), which is in line with other studies (e.g., Kabadayi, Krasheninnikova, et al. 2017 and Vernouillet at al. 2016). Two minutes should be sufficient, especially because our barriers were not entirely transparent (hence, partially occluded the food reward), making it easier to execute a detour response (Kabadayi, Bobrowicz, et al., 2018).

Gulls were tested during the second half of June in 2023 and 2024 (restricted to breeding season), quails in November 2023 (autumn), chickens in February 2024 (late winter), and canaries in May 2024 (late spring, again, restricted to breeding season).

Data Processing and Analysis

Video Recording and Analysis

The videos of the second and third habituation trial and the three test trials per test session were coded using the free, open-source 'Behavioural Observation Research Interactive' Software (BORIS, v.7.13.6) (Friard and Gamba, 2016). We coded five (types of) events (see Table 3 and Figure 3): latency to leave the start box (for habituation trials 2 and 3, as well as the six test trials), persisting (test trials only), moment of detouring the barrier (test trials only), interacting with the food bowl (for habituation trials 2 and 3, as well as the six test trials) and leaving the species-specific 'test box zone of interest' (test trials only). All videos were coded by the first author. A second person, blind to the hypotheses coded 10$\%$ of the videos per species. An average Cohen's Kappa (McHugh, 2012) was calculated for these videos to provide a measure of inter-rater, cross-species reliability. We had registered that, in the case no perfect inter-rater, cross species agreement (0.81 $\leq$ Cohen's Kappa $\leq$ 1) had been reached, discrepancies in inter-rater reliability would be investigated by calculating the average Cohen's kappa (McHugh, 2012) for each species, separately. By doing so, a species-specific or overall low Cohen's Kappa would reveal whether the videos have to be recoded for one or all four species. However this was not needed, as the average cross-species Cohen's kappa value indicated a strong level of inter-rater, cross species agreement (k = 0.927; McHugh, 2012).

To compare detour performance between species, we extracted our two response variables from the behavioural events coded in BORIS. First, the latency to detour (in seconds) was determined as the time between leaving the start box and the moment the individuals detoured the barrier. A maximum trial duration of 135 seconds for detour latency was assigned to the 20 trials (1.40% of the data set) in which birds did not ~~failed to~~ detour but entered the species-specific ‘barrier zone of interest’. Second, the time spent persisting (in seconds) was calculated as the cumulative time that the individual spent in the species-specific ’barrier zone of interest’ (size = Barrier L x 25% of the Barrier-Entry Distance; L x W, see Table 2 for the species-specific dimensions). A minimum trial duration of 0 seconds for persisting was assigned to the 483 trials (33.82% of the data set) in which birds detoured without entering ~~did not enter~~ the species-specific ’barrier zone of interest’ first. Third, a ’multi-baseline’ measure of an individual’s general motivational state (in seconds) was calculated, by *averaging* the time between leaving the start box and touching the food (bowl) placed in front of the opaque barrier on habituation trial 2 and 3. Note that if a bird was unsuccessful on trial 2, a non-averaged ’multi-baseline’ score was calculated based on habituation trial 3 only.

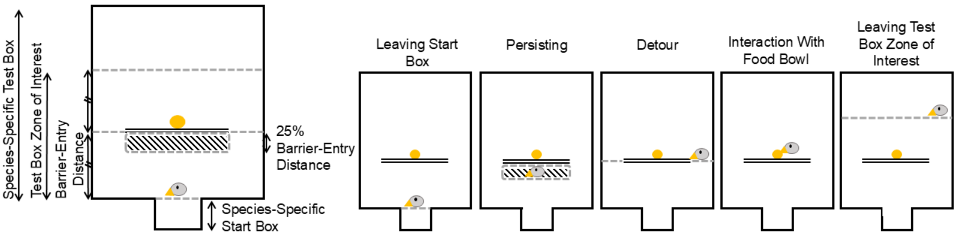

Figure 3. Visualisation of the species-specific dimensions (left) and the behaviours (right) that were coded in BORIS. Double line: Barrier; hatched area: species-specific ’barrier zone of interest’; dashed lines: (fictitious) lines that needed to be crossed by the bird, see Table 3.

Table 3. The description of the behaviours that were coded in BORIS.

Behaviour	Description
Leaving start box	- When the bird voluntarily leaves the start box: when both feet of the bird are visibly inside the test box, or (when the feet are not visible) when the front body half of the bird is inside the test box. - When the bird needs to be pushed: When the bird’s entire body is inside the test box [1] .
Persisting	At least the bird’s whole head crosses the (fictitious) lines of the rectangular-shaped, species-specific ’barrier zone of interest’ [2].
Detour	At least the bird’s whole head crosses the (fictitious) straight line from the barrier to the side of the test box (with a modifier whether they detour on the right or left side)
Interaction with food bowl	Bird touches the food or food bowl with its beak.
Leaving the ’test box zone of interest’	At least the bird’s whole head crosses the (fictitious) straight line at $\approx$ 2/3 of the test box length [2].

Note. [1] *= When a bird needed to be gently pushed, it was most likely that the individual was lying down in the start box. As a result, a gentle push put the bird forward and resulted in the bird ending up in a standing position in the test box. As a result, we coded whether these individuals had left the start box when the bird’s entire body (vs. front half body) was inside the test box.* [2] *= The fictitious lines that marked a zone of interest were defined by two wooden sticks attached to each side of the test box.*

Data Exclusion Criteria

Individuals that failed to visit the food bowl at the third habituation day, were excluded from subsequent test trials (pre-test criterion). This exclusion criterion guaranteed a similar within- and between-species proficiency with the basic task demands (e.g., the perceptual, motoric and motivational requirements to retrieve a food reward; for a similar pre-test exclusion criterion see, MacLean et al., 2014).

Birds that did not detour around the barrier nor entered the species-specific 'barrier zone of interest' in a test trial, were excluded from subsequent test trials (and data of that individual was excluded from all statistical analyses). This mid-test exclusion criterion 1 was applied for two reasons. First, birds that did not obtain a measure for one of the two dependent variables within 2 minutes were likely to be unmotivated or be in distress. Furthermore, observations from similar RI test paradigms in our lab demonstrate that such individuals are unlikely to eat at all with a prolonged test time or on subsequent test trials (within the same day).[10] In addition, removing birds from subsequent trials (rather than assigning a maximum trial limit for both dependent variables) reduced the risk of data skewing.

Individuals that left the species-specific 'test box zone of interest' (size = 2 times the Barrier-Entry Distance, see Table 2 for the species-specific dimensions) without touching the food (bowl) were also excluded from further testing and all analyses. This mid-test exclusion criterion 2 assured that we avoid confusing general exploration behaviour (without initial interest in the food) with successful detour performance (which assumes interest in the food). Thus, by excluding birds with differential trial experiences (due to e.g., demotivation, distress, distraction or exploration; for a similar mid-test exclusion criterion see, Van Horik, Beardsworth, Laker, Langley, et al., 2019), we aimed to ensure that each barrier orientation is standardized within- and between species. We registered that we would test all individuals of each species in a single 'season', as we incubated per season 20% more eggs than the number of individuals required for the testing; we expected that this surplus would allow us to replace all excluded individuals. For an overview of the birds excluded per criterion, see Table 4. However, due to the fearful and stress responses of gulls during testing (we come back to this in the general discussion), the exclusion rate was higher than expected so we had to include a second breeding season.

Table 4. Number (and %) of birds excluded per criterion.

Species	**White leghorn chicken**	**Japanese quail**	**Herring gull**	**Fife fancy Canary**	**Total**
Initial sample	90	85	147	170	492
Excluded for:					
Pre-test criterion	10 (11%)	2 (2%)	35 (24%)	1 (1%)	48 (10%)
Mid-test criterion 1	2 (2%)	4 (5%)	39 (26%)	14 (8%)	59 (12%)
Mid-test criterion 2	7 (8%)	21 (25%)	0 (0%)	31 (18%)	59 (12%)
Technical issues/Sick birds	2 (2%)	0 (0%)	0 (0%)	9 (5%)	11 (2%)
Remaining sample	69 (77%)	58 (68%)	73 (50%)	115 (68%)	315 (64%)

Note. All raised birds were subjected to habituation and (part of) testing. As can be seen, the total number of birds tested was higher than registered for all species (apart from the quails). This was due to the fact that these individuals were ~~also~~ reused for other studies, with different sample size requirements. Reusing individuals in other behavioral studies is possible when they share similar prior experiences (Van Horik, Langley, et al., 2018), and facilitates future analyses, such as exploring correlations between different tasks and making comparisons across studies. The first 60 individuals (58 for quails) that did not fail any exclusion criteria were selected for this study, ensuring a balanced design and minimizing group variation.

---

[10] In a continuous RI task with a sample size of 80 herring gulls, birds that failed on the first trial, were likely to fail again on the second trial of the same test day (Dewulf, Knoch, et al., 2025)

Statistical Analysis

Statistical analyses were performed using R. v. 4.2.2 (R Core Team, 2021). Models were fitted by means of the *lme4* package (D Bates, Mächler, et al., 2015) and parameter estimation and p-values for the generated models were provided by means of the ~~lmerTest package~~ (Kuznetsova et al., 2017) ~~via the Satterthwaite's degrees of freedom method (linear mixed model, LMM) or via~~ *carData* (Fox, Weisberg, and Price, 2022) and *car* (Fox and Weisberg, 2019) packages , which are suited for both linear mixed models (LMM) with temporal correlation structures and generalized linear mixed models (GLMM). For the ~~(G)~~LMM, we used partial eta-squared (η_p^2) as effect sizes for the relevant statistical models (linear mixed model, LMM) and they were calculated by means of the *effectsize* (Ben-Shachar et al., 2020) package.

Registered model

We registered that we would perform a (G)LMM with Type III sum of squares on the latency to detour and the cumulative time spent in the species-specific 'barrier zone of interest' (persisting). These registered ~~Both~~ models (model specification 1) included the between-species factor: *Species* (i.e., chickens, quails, gulls and canaries) and both within-species factors: *Barrier* (i.e., vertical- and horizontal-bar) and *Trial* (i.e., 1-3), and their interactions. In addition, they included ~~we add~~ two extra explanatory variables ~~to the model~~: a *'multi-baseline'* measure of an individual's general motivational state (and its interaction with *Species*, as we mean-centered this *'multi-baseline'* measure within *Species*, see Chen et al., 2014 for an example of within-group centering); and *Barrier Order* (with two levels: the individual received the horizontal-bar barrier on the first test day 1 and the vertical-bar barrier on the second test day; or vice versa), as species might demonstrate superior performance with the last encountered barrier, irrespective of its type and ecological validity. Bird identity and enclosure (social group) were included as a random intercept in the models, with bird identity nested in enclosures. In addition, we included by-individual (nested in enclosures) random slopes that varied for the levels of *Species* (corresponding with species-specific intercepts). The registered model is presented in model specification 1.

$$\text{Log(Outcome, s)} \sim \text{Species} \times (\text{Barrier} \times \text{Trial} + \text{Baseline}) + \text{BarrierOrder} + (\text{Species} | \text{Id:Enclosure})) \tag{1}$$

We registered that we would generate plots by means of the package *performance* (Lüdecke et al., 2021) to inspect for violations of the model assumptions: 1) heteroscedasticity (plotting the square root of the residuals (y-axis) and fitted values (x-axis)), 2) non-normality of residuals (plotting the sample quantiles (y-axis) on the standard normal distribution quantiles) and 3) outliers (plotting standard residuals (y-axis) and leverage). Additionally, the multicollinearity between fixed main factors (via the variance inflation factor, VIF) and the autocorrelation between residuals (via a Durbin-Watson-Test) would be calculated via functions provided by the *performance* (Lüdecke et al., 2021) package. Potential violations of model assumptions would be addressed by transforming the (in)dependent variables (i.e., via log-transformation) or by changing the error distribution (family) or the link function of the model (switching a default LMM that will be fitted to a GLMM). Fixed main effects with a VIF of >5 were planned to be removed and logical outliers (i.e., recording/entry errors) would be inspected and corrected (if possible). In the case that the outlier could not be corrected, all data of that individual was planned to be excluded from all statistical analyses.

Applied model

Following the registered inspections and analyses, changes were made to the registered models to address model complexity, violations of certain assumptions, and issues with model convergence.

For detour latency, the registered model (model specification 1) was simplified by removing the random slope for Species in order to reduce the model's complexity. This decision was based on the presence of perfect or near-perfect correlations among random effects, indicating redundancy and boundary singularity. In order to address violations of the model assumptions (i.e., heteroscedasticity and non-normality of residuals), the dependent variable was log-transformed (see supplementary Figure 1, Dewulf, Garcia-Co, et al., 2023). To address autocorrelation in the residuals [11], the model was further extended with a temporal correlation structure using the *nlme* package (Pinheiro and DM Bates, 2000). This temporal correlation structure accounts for the correlation in residuals from repeated measurements across Time (i.e., 1-6 trials; for each bird, nested within enclosures). Specifically, each bird participated in two sessions, with one session per barrier type and three trials per session, resulting in six interdependent trials. The autocorrelation parameter (ϕ), estimated by the model at lag 1, was 0.319. Explicitly modeling this autocorrelation properly accounts for the residuals' temporal dependencies (see supplementary Figure 2, Dewulf, Garcia-Co, et al., 2023), leading to an improved model fit (AIC = 4063.716 with the correlation structure vs. AIC 4275.634 without) and more accurate parameter estimates. Adding this correlation structure was required as adjusting the error distribution (e.g., gamma or inverse Gaussian) did not resolve the autocorrelation issue, as the models with adjusted error distributions encountered convergence problems. VIF scores were all below 5, and no logical outliers were detected, so we did not have to remove any outliers. The applied model for detour latency is presented in model specification 2, for an overview of the evolution of the model structure, see supplementary materials (Dewulf, Garcia-Co, et al., 2023).

$$\text{Log(Detour Latency, s)} \sim \text{Species} \times (\text{Barrier} \times \text{Trial} + \text{Baseline}) + \text{BarrierOrder} + (1|\text{Id:Enclosure}) + \text{corAR1}(\text{Time}|\text{Id:Enclosure}) \tag{2}$$

For persisting, the registered model (model specification 1) was simplified by removing the random slope for Species, for the same reasons as in the statistical model for detour latency. The simplified model demonstrated violations of model assumptions (i.e., heteroscedasticity and non-normality of residuals; see supplementary Figure 3 Dewulf, Garcia-Co, et al., 2023), which could not be addressed by log-transforming the dependent variable due to the presence of zeros in the data. To meet model assumptions, various models with different error distributions were explored, including Poisson, negative binomial (ZI) and zero-inflated negative binomial (ZINB), with the selection guided by the data characteristics. The dependent variable, persisting, was therefore also converted to integer counts by scaling the original data to frames (30 frames per second), which was necessary to meet the model's requirements while preserving the precision of short latencies that would otherwise be rounded to zero. The selected models are designed to account for overdispersion and excess zeros, reducing the need for explicit tests of heteroscedasticity and non-normality of residuals. However, additional diagnostics were conducted using the *DHARMa* package (Hartig, 2022), assessing: 1) residual uniformity (Kolmogorov-Smirnov test), 2) over/under-dispersion, 3) outliers, 4) zero-inflation, and 5) autocorrelation (via residual plots [11]). Ultimately, the ZINB model, implemented via the *glmmTMB* package [12] satisfied the final model assumptions (see supplementary Figure 4 Dewulf, Garcia-Co, et al., 2023). The ZINB model included a negative binomial component to capture variability in persisting and a zero-inflated part to account

---

[11] The Stage 1 registered Durbin-Watson statistic (using *performance* Lüdecke et al., 2021 package) and the ACF plot (using the base R package) are not suitable for (G)LMM (due to the random effect structure) an issue identified during the data analysis. Consequently, we opted for alternative methods to assess autocorrelation in the residuals. Specifically, we employed ACF plots (for LMM; using the package *nlme*, Pinheiro and DM Bates, 2000) or simulation-based residual plots (GLMM; using the package *DHARMa*, Hartig, 2022) for inspecting autocorrelation between residuals.

[12] The *glmmTMB* (Brooks et al., 2017) package depends on both the *Matrix* (D Bates, Maechler, et al., 2023) and *TMB* (Thygesen et al., 2017) packages. To fit a (ZI)NB model with random effects, it is essential to reinstall the *Matrix* (D Bates, Maechler, et al., 2023) package from source (version 1.6.2) followed by reinstalling *TMB* (Thygesen et al., 2017) and *glmmTMB* (Brooks et al., 2017) from source.

for the excess of zeros in persisting. The best-fitting zero-inflation component was determined using AIC comparisons across models with different combinations of main and interaction effects in the zero-inflated model's formula, which resulted in the inclusion of *Barrier*, *Baseline*, *Species*, *Trial* and the interaction between *Species* and *Trial* as zero-inflated effects. VIF scores were all below 5, and no logical outliers were detected. The applied model for persisting is presented in model specification 3, see supplementary materials (Dewulf, Garcia-Co, et al., 2023).

$$\mathrm{NB2(Persisting,frames)} \sim \mathrm{Species} \times (\mathrm{Barrier} \times \mathrm{Trial} + \mathrm{Baseline}) + \mathrm{BarrierOrder} + (1|\mathrm{Id{:}Enclosure}) + \mathrm{zi} \sim \mathrm{Barrier} + \mathrm{Baseline} + \mathrm{Species} + \mathrm{Trial} + \mathrm{Species{:}Trial}) \tag{3}$$

We had preregistered that, in case of (a) significant *Barrier* x *Species* interaction effect(s) (**Prediction 1**), ~~and/or~~ (b) a main effect of *Trial* (**Prediction 2**), ~~further post-hoc Bonferroni-holm corrected linear contrasts upon the model will be performed to compare respectively, the performance with different ecological (non-) valid barrier per Species and performance over Trials (within one session). In case we find~~ or (c) a significant three-way *Species* x *Barrier* x *Trial* interaction effect(s) (**Explorative Prediction 3**), ~~further exploratory Bonferroni-Holm corrected linear contrasts upon the model will be performed to compare Trial performance of Species on different types of the Barrier~~ follow-up tests would be performed. We only found a significant effect of *Trial* (**Prediction 2**). In line with the preregistration, we performed post-hoc Bonferroni-Holm corrected (Holm, 1979) linear contrasts upon the model to compare performance over trials (within one session). Follow-up linear contrasts along with the corresponding ~~and Cohen's *d*~~ effect sizes (observed Cohen's *d* for LMMs, Incidence Rate Ratio (IRR) for the negative binomial part of the ZINB model and Odds Ratios (OR) for the zero-inflated part of the ZINB model) were calculated by means of the *emmeans* (Lenth, 2023) and *lsr* (Cohen's *d*, Navarro, 2015) packages.

Results

Detour Latency

Registered comparisons ~~Main Analyses~~

Descriptive statistics appear in Figure 4 and Tables 6, 7 and 8; inferential statistics appear in Table 5 and supplementary Tables 1, 2 and 3 (Dewulf, Garcia-Co, et al., 2023). The Species x Barrier (**Prediction 1**) and the Species x Barrier x Trial (**Prediction 3**) interaction effects were not significant. However, there was a significant main effect of Trial (**Prediction 2**), as shown in Table 5.

Table 5. Output: LMM with temporal correlation structure on detour latency (s)

Parameter	$\tilde{X}^2$	Df	P	np^2
(Intercept)	1322.498	1	**<0.001**	
Species	77.015	3	**<0.001**	0.249
Barrier	1.343	1	0.246	0.001
Trial	64.249	2	**<0.001**	0.051
Baseline_centered	0.568	1	0.451	0.000
BarrierOrder	2.852	1	0.091	0.013
Barrier:Trial	0.151	2	0.927	0.000
Species:Barrier	0.147	3	0.986	0.000
Species:Trial	56.035	6	**<0.001**	0.045
Species:baseline_Centered	5.452	3	0.142	0.023
Species:Barrier:Trial	8.228	6	0.222	0.007

Note. Significant effects are indicated with bold p-values

Follow-up contrasts upon the model for the main effect of Trial showed that performance improved over trials, with significantly slower detour latencies on Trial 1 compared to Trial 2 and Trial 3. There was no significant difference in detour latencies on Trial 2 compared to Trial 3 (see Table 6 and Figure 4 panel A). Further inferential statistics are provided in the supplementary Table 1.

Table 6. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in seconds) and the observed means (also on the original scale, in seconds) for detour latency across different Trial levels.

	Model		exp(model)		observed
Trial	**Mean(SE)**	**CI**	**Mean**	**CI**	**Mean (SD)**
Trial1	2.005 (0.057)	1.892 - 2.118	7.428	6.635 - 8.316	15.276 (22.114)
Trial2	1.664 (0.058)	1.550 - 1.777	5.279	4.712 - 5.914	11.636 (21.879)
Trial3	1.617 (0.057)	1.504 - 1.730	5.038	4.500 - 5.640	10.418 (18.523)

Additional Exploratory Analyses

Further examination of the model revealed an unexpected significant main effect of Species (see Table 5). All pairwise comparisons were statistically significant (see Table 7 and Figure 4 panel B). Further inferential statistics are provided in the supplementary Table 2).

Table 7. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in seconds) and the observed means (also on the original scale, in seconds) for detour latency across different Species levels.

	Model		exp(model)		Observed
Species	**Mean(SE)**	**CI**	**Mean**	**CI**	**Mean (SD)**
Canary	2.002 (0.097)	1.810 - 2.194	7.404	6.110 - 8.972	14.591 (22.203)
Chicken	1.531 (0.097)	1.339 - 1.724	4.625	3.817 - 5.604	8.245 (13.867)
Gull	2.313 (0.097)	2.121 - 2.505	10.100	8.335 - 12.239	18.517 (27.911)
Quail	1.202 (0.099)	1.007 - 1.397	3.327	2.736 - 4.044	8.280 (14.771)

There was also an unexpected interaction between Species x Trial (see Table 5). Post-hoc linear contrasts showed that, performance improved for Canaries, Chickens and Quails, but not for Gulls. Specifically, Quails exhibited slower detour latencies on Trial 1 compared to Trial 2 and Trial 3. Quails were also significantly faster on Trial 3 than Trial 2. Canaries and Chickens exhibited slower detour latencies on Trial 1 compared to Trial 2 and Trial 3. There was no significant effect between Trial 2 and Trial 3 for either species. For Gulls detour latencies were significantly faster on Trial 1 compared to Trial 3. Similarly, detour latencies on Trial 2 were significantly faster than on Trial 3. No significant effect was observed between Trial 1 and Trial 2 (see Table 8 and Figure 4 panel C). Further inferential statistics are provided in the supplementary Table 3.

Table 8. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in seconds) and the observed means (also on the original scale, in seconds) for detour latency across different Trial nested within Species levels.

Condition	Model Mean(SE)	CI	exp(model) Mean	CI	Observed Mean (SD)
Canary					
Trial1	2.417 (0.114)	2.192 - 2.642	11.209	8.952 - 14.035	20.890 (26.208)
Trial2	1.874 (0.115)	1.647 - 2.100	6.511	5.193 - 8.165	13.992 (24.149)
Trial3	1.716 (0.114)	1.491 - 1.941	5.561	4.441 - 6.963	8.892 (12.004)
Chicken					
Trial1	1.888 (0.114)	1.664 - 2.113	6.608	5.278 - 8.274	11.423 (17.141)
Trial2	1.395 (0.115)	1.169 - 1.621	4.035	3.218 - 5.060	6.510 (9.563)
Trial3	1.311 (0.114)	1.086 - 1.536	3.709	2.963 - 4.644	6.801 (13.399)
Gull					
Trial1	2.220 (0.114)	1.995 - 2.445	9.206	7.352 - 11.527	16.243 (22.341)
Trial2	2.213 (0.115)	1.986 - 2.439	9.140	7.289 - 11.461	18.424 (31.530)
Trial3	2.505 (0.114)	2.280 - 2.730	12.247	9.781 - 15.335	20.885 (29.086)
Quail					
Trial1	1.496 (0.116)	1.268 - 1.725	4.465	3.552 - 5.612	12.452 (20.766)
Trial2	1.174 (0.117)	0.944 - 1.404	3.234	2.569 - 4.071	7.479 (12.313)
Trial3	0.936 (0.116)	0.707 - 1.164	2.549	2.028 - 3.204	4.909 (6.770)

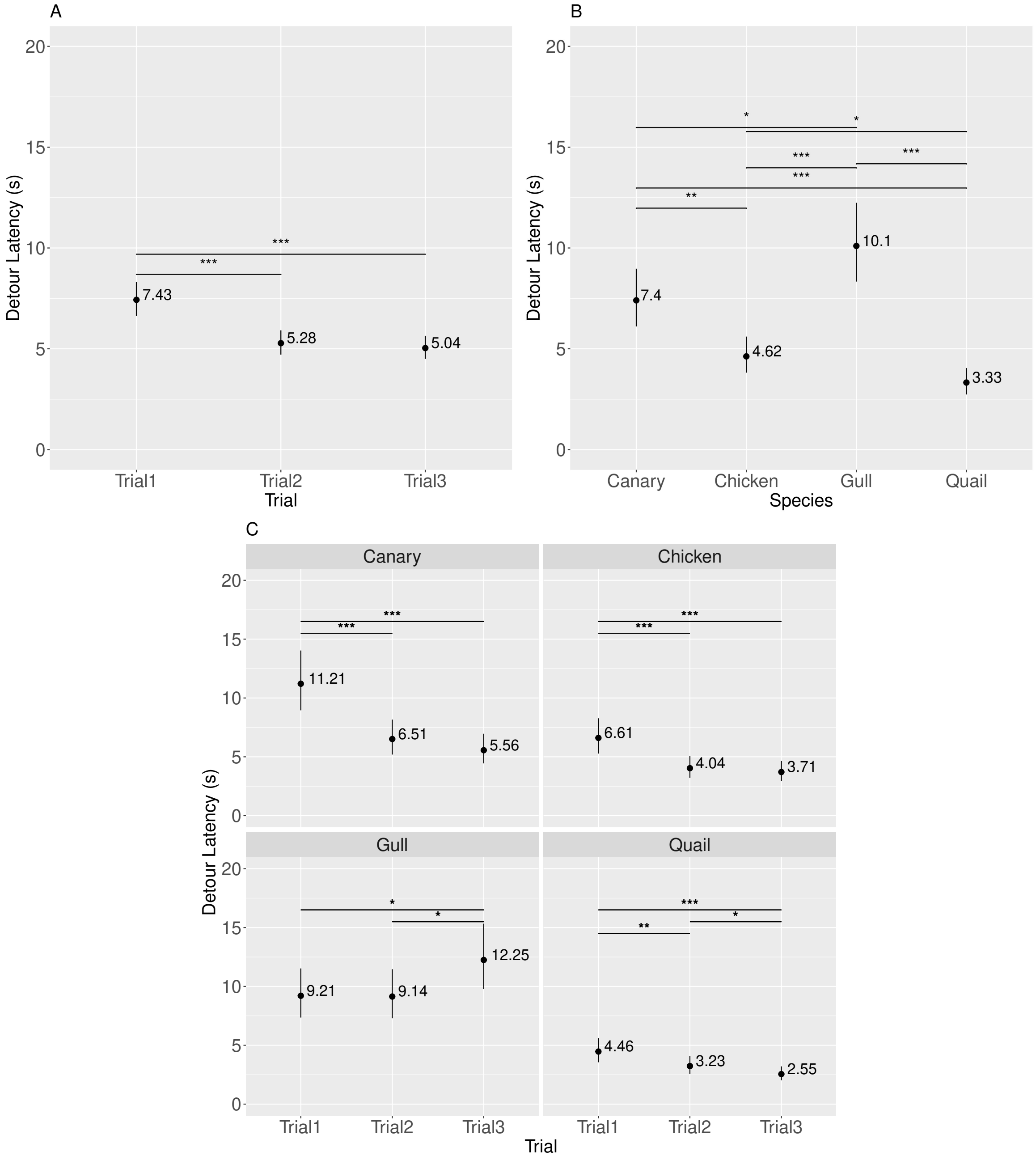

Figure 4. Visualization of model-predicted means (back-transformed to detour latency in seconds) along with their 95% CI across different Trial levels (Panel A), Species (Panel B), and Species by Trial interactions (Panel C). Significant effects are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Persisting

Registered comparisons ~~Main Analyses: Negative Binomial Part~~

Descriptive statistics for the negative binomial part of the model appear in Figure 5 and Tables 10, 11 and 12; inferential statistics appear in Table 9 and supplementary Tables 4, 5 and 6 (Dewulf, Garcia-Co, et al., 2023).

The Species x Barrier (**Prediction 1**) and the Species x Barrier x Trial (**Prediction 3**) interaction effects were not significant. However, there was a significant main effect of Trial (**Prediction 2**), as shown in Table 9.

Table 9. Output: GLMM on persisting (frames)

Parameter	$\tilde{X}^2$	Df	p
	Negative binomial part		
(Intercept)	9266.6~~59~~ 80	1	**<0.001**
Species	24.031	3	**<0.001**
Barrier	0.005	1	0.942
Trial	97.22~~5~~ 2	2	**<0.001**
baseline_centered	1.506	1	0.220
BarrierOrder	2.30~~1~~ 2	1	0.129
Barrier:Trial	8.514	2	**0.014**
Species:Barrier	5.292	3	0.152
Species:Trial	3.949	6	0.684
Species:baseline_centered	4.20~~1~~ 0	3	0.241
Species:Barrier:Trial	4.150	6	0.656
	Zero-inflated part		
(Intercept)	153.73~~2~~ 1	1	**<0.001**
Barrier	11.758	1	**< 0.001**
baseline_centered	12.73~~0~~ 3	1	**<0.001**
Species	174.5~~49~~ 52	3	**<0.001**
Trial	35.177	2	**<0.001**
Species:Trial	12.57~~2~~ 3	6	0.050

Note. Significant effects are indicated with bold p-values

Follow-up contrasts upon the negative binomial part of the model to further investigate the main effect of Trial showed that performance improved over trials, with significantly more persisting on Trial 1 compared to Trial 2 and Trial 3. There was also significantly more persisting on Trial 2 compared to Trial 3 (see Table 10 and Figure 5 panel A). Further inferential statistics are provided in the supplementary Table 4.

Table 10. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in frames) and the observed means (also on the original scale, in frames) for persisting across different Trial levels.

	Model		**exp(model)**		**Observed**
Trial	**Mean(SE)**	**CI**	**Mean**	**CI**	**Mean (SD)**
Trial1	5.594 (0.070)	5.456 - 5.732	268.7~~72~~ 69	234.14~~7~~ 4 - 308.51~~8~~ 4	295.861 (389.194)
Trial2	5.025 (0.077)	4.875 - 5.176	152.233	130.941 - 176.987	204.441 (396.460)
Trial3	4.664 (0.083)	4.501 - 4.827	106.03~~6~~ 7	90.09~~4~~ 5 - 124.~~799~~ 800	148.066 (288.298)

Additional Exploratory Analyses ~~Negative Binomial Part~~

Further examination of the negative binomial part of the model revealed an unexpected significant main effect of Species (see Table 9). Post-hoc linear contrasts showed that Canaries and Chickens persisted less compared to Gulls and Quails. All pairwise comparisons were statistically significant, except for the comparisons between Canaries and Chickens, and between Gulls and Quails (see Table 11 and Figure 5 panel B). Further inferential statistics are provided in the supplementary Table 5. The main effect of Species will be addressed further in the discussion.

Table 11. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in frames) and the observed means (also on the original scale, in frames) for persisting across different Species levels.

	Model		exp(model)		observed
Species	**Mean(SE)**	**CI**	**Mean**	**CI**	**Mean (SD)**
Canary	4.827 (0.114)	4.604 - 5.051	124.88~~5~~ 6	99.898 - 156.123	190.718 (337.446)
Chicken	4.878 (0.092)	4.698 - 5.059	131.41~~7~~ 8	109.747 - 157.36~~7~~ 8	152.094 (171.974)
Gull	5.396 (0.093)	5.214 - 5.579	220.62~~5~~ 7	183.83~~6~~ 8 - 264.77~~5~~ 8	305.363 (505.505)
Quail	5.275 (0.120)	5.039 - 5.511	195.42~~4~~ 0	154.33~~4~~ 1~~X~~ X - 247.4~~54~~ 48	238.181 (345.849)

There was also an unexpected interaction between Barrier x Trial (see Table 9). Post-hoc linear contrasts showed that overall performance improved with both types of barriers. For the Horizontal-Bar Barrier, individuals persisted significantly more on Trial 1 compared to Trial 2 and Trial 3. There was no significant difference between Trail 2 and Trial 3. For the Vertical-Bar Barrier, individuals persisted significantly more on Trial 1 compared to Trial 2 and Trial 3. In addition, there was significantly more persisting on Trial 2 compared to Trial 3 (see Table 12 and Figure 5 panel C). Further inferential statistics are provided in the supplementary Table 6).

Table 12. The model predicted means (on the log scale), the back-transformed model-predicted means (on the original scale, in frames) and the observed means (also on the original scale, in frames) for persisting across Trial nested in Barrier levels.

	Model		exp(model)		observed
Condition	**Mean(SE)**	**CI**	**Mean**	**CI**	**Mean (SD)**
Horizontal					
Trial1	5.635 (0.102)	5.435 - 5.835	~~280.005~~ 279.997	229.29~~8~~ 3 - 341.9~~24~~ 14	18.442 (23.536)
Trial2	4.862 (0.107)	4.653 - 5.072	129.346	104.915 - 159.466	14.458 (24.518)
Trial3	4.776 (0.119)	4.543 - 5.009	118.6~~09~~ 10	93.96~~6~~ 8 - 149.71~~3~~ 5	16.337 (26.444)
Vertical					
Trial1	5.553 (0.092)	5.372 - 5.734	257.99~~0~~ 1	215.250 - 309.218	19.378 (24.638)
Trial2	5.188 (0.102)	4.989 - 5.388	179.170	146.77~~3~~ 4 - 218.71~~6~~ 7	16.757 (26.925)
Trial3	4.552 (0.108)	4.340 - 4.763	94.79~~6~~ 7	76.74~~3~~ 4 - 117.097	12.214 (17.545)

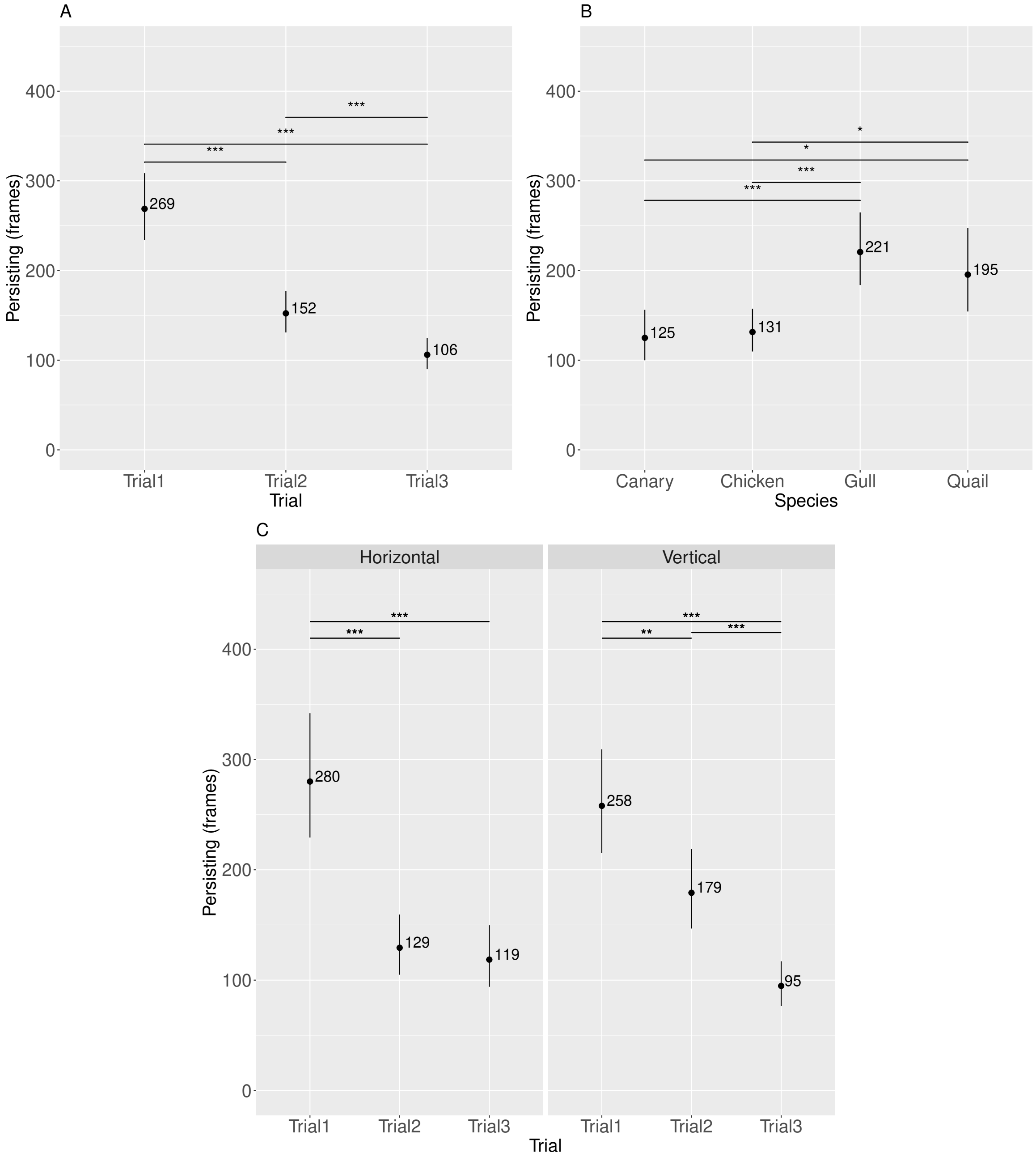

Figure 5. Visualisation of the negative binomial model-predicted means (back-transformed log estimates to the original scale, representing, persisting in frames) along with their asymptotic CI across different Trial levels (Panel A), Species (Panel B) and Trial by Barrier interactions (Panel C). Significant effects are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

~~Additional Analyses Zero-Inflated Part~~

In addition to the negative binomial component, the statistical model for persisting also included a zero-inflated component that accounted for excess zeros in persisting (i.e., capturing all birds that did not persist). Descriptive statistics for the zero-inflated part of the model appear in Figure 6 and Tables 13, 14, 15 and 16; inferential statistics appear in Table 9 and supplementary Tables 7, 8 and 9 (Dewulf, Garcia-Co, et al., 2023).

Examination of the zero-inflation part of the model revealed a significant main effect of Trial (see Table 9). Post-hoc linear contrasts showed that overall there was an increase in the probability of zeros for persisting in later trials, indicating improved accuracy. Specifically, Trial 3 showed a significant higher probability of zeros for persisting compared to Trial 2 and Trial 1. Trial 2 also had a significantly higher probability of zeros for persisting compared to Trial 1 (see Table 13 and Figure 6 panel A). Further inferential statistics are provided in the supplementary Table 7.

Table 13. The model predicted means (on the log odds ratio scale), the back-transformed zero probabilities (on the original scale, in frames) and the observed zero probabilities (also on the original scale, in frames) for persisting across Trial levels.

	Model		Prob(model)		Observed
Trial	**Mean(SE)**	**CI**	**prob**	**CI**	**prob**
Trial1	-1.438 (0.143)	-1.718 - -1.158	0.192	0.153 - 0.239	0.244
Trial2	-0.832 (0.116)	-1.058 - -0.605	0.303	0.258 - 0.353	0.342
Trial3	-0.388 (0.107)	-0.597 - -0.180	0.404	0.355 - 0.455	0.429

Again, we had not predicted a main effect of Species (see Table 9). However, Canaries and Quails had a higher probability of zeros for persisting (indicating higher accuracy) compared to Chickens and Gulls. All pair-wise comparisons were statistically significant, exception for the comparisons between Canaries and Quails, and between Chickens and Gulls (see Table 14 and Figure 6 panel B). Further inferential statistics are provided in the supplementary Table 8). The main effect of Species will be addressed further in the discussion.

Table 14. The model predicted means (on the log odds ratio scale), the back-transformed zero probabilities (on the original scale, in frames) and the observed zero probabilities (also on the original scale, in frames) for persisting across Species levels.

	Model		Prob(model)		Observed
Species	**Mean(SE)**	**CI**	**prob**	**CI**	**prob**
Canary	-0.139 (0.114)	-0.362 - 0.085	0.466	0.410 - 0.521	0.478
Chicken	-1.831 (0.175)	-2.173 - -1.489	0.138	0.102 - 0.184	0.169
Gull	-1.785 (0.157)	-2.093 - -1.477	0.144	0.110 - 0.186	0.158
Quail	0.211 (0.111)	-0.006 - 0.428	0.553	0.499 - 0.605	0.555

There was also an unexpected significant main effect of Barrier (see Table 9). Post-hoc linear contrasts showed a significant higher probability of zeros for persisting (indicating higher accuracy) for the Horizontal-Bar Barrier than Vertical-bar Barrier (see Table 15 and Figure 6 panel C). Further inferential statistics are provided in the supplementary Table 9)

Table 15. The model predicted means (on the log odds ratio scale), the back-transformed zero probabilities (on the original scale, in frames) and the observed zero probabilities (also on the original scale, in frames) for persisting across Barrier levels.

	Model		Prob(model)		Observed
Barrier	**Mean(SE)**	**CI**	**Prob**	**CI**	**Prob**
Horizontal	-0.665 (0.092)	-0.846 - -0.485	0.339	0.300 - 0.381	0.377
Vertical	-1.106 (0.100)	-1.302 - -0.910	0.249	0.214 - 0.287	0.300

An unexpected main effect of Baseline was also observed (see Table 9). Follow-up analyses indicated that the probability of zeros for persisting was estimated at 0.292, when birds had an average motivation score

(Baseline at zero, due to within-species mean-centering). Descriptive statistics appear in Table 16; inferential statistics appear in Table 9.

Table 16. The model predicted means (on the log odds ratio scale), the back-transformed zero probabilities (on the original scale, in frames) and the observed zero probabilities (also on the original scale, in frames) for the average value of Baseline.

	Model		**Prob(model)**		**Observed**
average baseline value	**Mean(SE)**	**CI**	**prob**	**CI**	**prob**
0	-0.886 (0.071)	-1.026 - -0.746	0.292	0.264 - 0.322	0

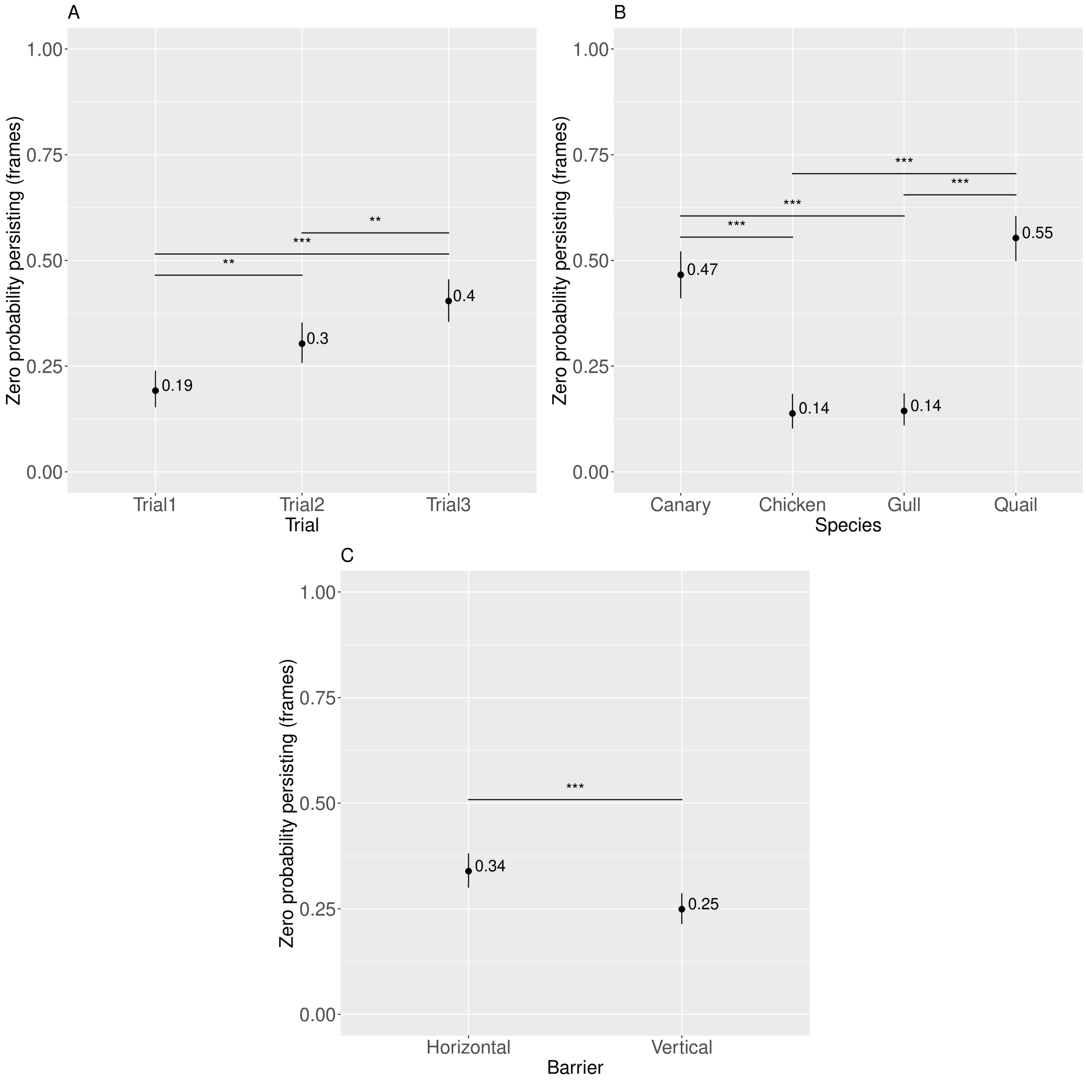

Figure 6. Visualisation of the zero-inflated model-predicted means (back-transformed to the probability zeros for persisting in frames) along with their asymptotic CI across different Trial (Panel A), Species (Panel B) and Barrier levels (Panel C). Significant effects are indicated as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Additional Analysis: Group Size as a Random effect

Due to post-hatch mortality (in canaries, gulls, and quails), group size deviated slightly from the intended 10 individuals per group (Table 17) as batch incubation limited the ability to replace these losses with age-matched individuals. Note that group sizes greater than 10 resulted from the merging of two high mortality enclosures where birds were of the same age. For the canaries, this was further complicated by the need to introduce 'tutors' for the juveniles (i.e. adult demonstrators to teach independent feeding), which meant that each group of juveniles (about 10) had at least one extra adult for a few days. To examine the possible impact of variation in group size on performance, additional analyses were conducted using 'Group Size' as a random variable. For both detour latency and persistence, the extended model did not outperform the above-mentioned reported models. As a result, these additional analyses are discussed in the supplementary materials (Table 10, Table 11).

Table 17. Visualization of the number of individuals that met our exclusion criteria, in relation to the enclosure ~~raised in each~~ group size and the ~~across~~ species.

	Group Size								
Species	**#6**	**#8**	**#9**	**#10**	**#11**	**#12**	**#14**	**Mean**	**Range**
Canary	0	0	15	20	14	0	11	10.717	9 - 14
Chicken	0	0	0	60	0	0	0	10	10 - 10
Gull	0	22	3	35	0	0	0	9.217	8 - 10
Quail	3	2	7	39	0	7	0	9.845	6 - 12

Discussion

We argue that stop-signal detection is a critical cognitive component of response inhibition (RI) across species, including birds. This study explored this idea further by investigating whether RI is improved when the perceptual characteristics of the stop signal (i.e., barrier) in the detour barrier task correspond to the species' ecological niche, as shown by Regolin et al. (1994) and Zucca et al. (2005). However, we failed to replicate this earlier work, as RI was not significantly improved when the barrier type supposedly matched the ecological niche of the species (**Prediction 1)**. However, we did find that performance generally improved over trials (**Prediction 2**) for both detour latency and persisting, but again, this did not interact with the species-specific ecological validity of the stop signal ~~vary between species~~ (**Prediction 3**). ~~The latter indicates that learning occurs independently of the ecological validity of the stop signal.~~

Most importantly, we were unable to replicate the finding that barrier type had a species-specific influence on detour performance, even though our study has several methodological and conceptual strengths, including a well-powered design, standardised experimental procedures, controlled prior experience (through pre- and mid-test exclusion criteria), and baseline measures (to minimise confounding by non-cognitive, motivational traits). Thus, our findings do not support the ecological-niche hypothesis as proposed by Regolin et al. (1994) and Zucca et al. (2005), suggesting that the adaptation to a specific ecological niche cannot account for variation in stop-signal detection (at least, not in the detour task). This does not necessarily imply that stop-signal detection is not important at all for RI, but it does indicate that differences between the four bird species tested here are not caused by variation in how they perceive or interpret vertical and horizontal-bar barriers.

As trials progressed, most individuals became faster in detouring (except for gulls) and made fewer attempts to persist at interacting with the barrier, regardless of the barrier type (confirming Prediction 2 but disconfirming Prediction 3). In the habituation set-up (or training phase) of our study, the food bowl was placed *in front* of the opaque barrier; this ensured that birds had no prior experience of retrieving food from behind a barrier (which standardised baseline performance). But without this experience, in the test phase, individuals had to learn both to inhibit their prepotent response to go directly for the reward (as the direct path is blocked) ~~their initial behaviour~~ and to navigate around the barrier (Van Horik, Beardsworth, Laker, Whiteside, et al., 2020), explaining the observed improvements over trials.

The learning pattern observed for the gulls was unexpected, as it appears that gulls learned to inhibit interacting with the barrier itself but without an overall improvement in detour latency, whereas the other species became faster at detouring and interacted less over time with the barrier. At present, we have no explanation as to why, for gulls, learning was only observed for the persistence measure and not detour latency as was seen in the other three species. However, this pattern demonstrates the value of looking at detour latency and time spent interacting with the barrier. One might assume that lower persistence scores should automatically result in shorter detour latencies but for gulls, this was not the case. This indicates that overall task performance (i.e., detour latency) captures additional behaviours, potentially unrelated to response inhibition (e.g., the time taken to approach the barrier, time spent not interacting, time needed to navigate the barrier, etc.). The observed differences in learning also highlight two further issues. First, the fact that gulls showed evidence of learning in measures of persistence but not in the measure of detour latency suggests that, at least for some species, ~~tasks include several subcomponents and that some of these are not equally influenced by learning across species~~ certain task components are more influenced by learning (inhibiting an unrewarded repetitive response) than others (inhibiting the response to go straight for the food or navigating around a barrier, which are both captured by detour latency). Speculatively, this could be related to the ecological niche adaptations of the species as well. Certain behaviours, such as inhibition of unrewarded responses, may be more critical than others in certain ecological niches, making them easier to learn. In contrast, other behaviours, such as navigating obstacles, may be more influenced by context-specific factors, and therefore, harder to learn for certain species (although follow-up work is required to test this idea). Second, the learning differences stress the need to take the role of learning in RI (and cognition in general) into account when aiming to interpret the variation in RI between species. For example, while canaries and gulls were initially slow at detouring (compared to chickens and quails; Figure 4), detour latencies of canaries gradually decreased, while those of gulls did not. This suggests that the differences between these two species in a putative test of response inhibition could at least partly reflect variations in learning rather than inhibition, with interaction effects between species and trial potentially explaining these findings (Willcox et al., 2024).

Alongside the effects of trial, we also identified general latency differences between species. Even though such differences are hard to interpret, one notable finding stands out, namely that gulls appeared to 'underperform' compared to the other species, as they were generally slower (compared to the three other species; Figure 4), more likely to peck (compared to canaries and quails; Figure 6), and when they did peck, they pecked for longer (compared to canaries and chickens; Figure 5). As noted above, the gulls' detour latencies also did not decrease over trials. We consider two (not mutually exclusive) hypotheses. First, the gulls are a wild species, whereas the three other species are domesticated. Domesticated species are generally less fearful and stressed than wild species (Kaiser et al., 2015). For example, Gjøen and colleagues (2023) compared the behavioural responses of white leghorn chickens with their wild counterparts, red junglefowls (*Gallus gallus*), in risk-taking situations, such as the encounter of a novel object during food retrieval. They found that red junglefowls were more stressed and fearful of the object and reached the food later than white leghorns. If gulls were indeed more fearful and stressed than the other species, this could have influenced their detour performance. Consistent with this idea, there was a high number of drop-outs among gulls (compared to the three other species; Table 4) due to the pre-test (i.e., a failure to interact with the food bowl in the presence of a novel barrier in a new test environment) or mid-test 1 (i.e., a failure to obtain a measure for one of the two dependent variables during a test trial, indicating little interaction with the experimental task) exclusion criteria. Second, even though we standardised the testing age in terms of number of days, the developmental trajectories of RI (and cognition in general) may have differed between species. Gulls have a much longer maximum lifespan (49 years) compared to canaries (24 years), chickens (15-20 years), and quails (6 years; Tacutu et al., 2018); based on life history theories, one could speculate that neuro-cognitive development would be protracted in the semi-precocial and long-living gulls compared with e.g. the precocial and shorter-living quails and chickens (Bunge, 2012). However, this idea should be further tested.

Finally, it is noteworthy that canaries successfully solved (and learned) the detour problem (irrespective of barrier type). In contrast,~~while~~ in the study of Zucca et al. (2005), canaries were unable to solve the detour problem and repeatedly attempted to fly over the barrier (again, irrespective of barrier type). The authors attributed this inability to the canaries' adaptation to an aerial environment, which enables them to navigate obstacles by simply flying over them in natural environments. However, several other studies have already shown that species, adapted to an aerial lifestyle, such as ravens (Kabadayi, Jacobs, et al., 2017), ring doves (*Streptopelia risoria*, Miller, 1974; Miller and Tallarico, 1974) and pigeons (*Columbia livia*, Miller, 1974; Miller and Tallarico, 1974), are capable of solving the detour barrier task as well. We speculate that canaries were able to solve the detour problem in our study, but not in the original work, due to the exclusion criteria we implemented, which ensured proficiency with the basic task demands (e.g., the perceptual, motoric, and motivational requirements for retrieving a food reward; MacLean et al., 2014). Specifically, our pre-test exclusion criterion ensured that all included birds visited and ate from a food bowl placed in front of a barrier (novel object) in the habituation phase before access to the food bowl was restricted by moving the barrier in front of it in the test phase. We believe that experience with retrieving the reward may be critical for measuring detour performance, potentially more so in aerially adapted birds. After all, Zucca et al. (2005) found that, even after prolonged exposure to the test situation, a large proportion of canaries were unable to solve the detour problem. This suggest, that the problem was not a lack of familiarity with the test itself, but rather a lack of experience with retrieving the reward. However, this explanation is speculative and requires further investigation.

In summary, we failed to provide support for the 'ecological niche hypothesis', as proposed by (Regolin, Vallortigara, et al., 1994; Zucca, Antonelli, et al., 2005). Our study adds to the growing body of evidence for the critical need for replication studies (Farrar et al., 2020), and highlights the need to consider methodological and conceptual design factors, as these can significantly impact results. Although our study did not provide strong evidence for the idea that interspecies differences in the perception of barrier types influence detour performance (and cause species differences), this does not negate the need for further research into the influence of the characteristics of the stop signal or other underlying mechanisms of RI. More generally, future research should focus on the cognitive mechanisms underlying RI. Understanding these mechanisms will help explain inter-individual variation such as in decision-making in dynamic environments (Johnson-Ulrich and Holekamp, 2020), predator avoidance and foraging optimization (Tvardíková and Fuchs, 2012), as well as responses to broader ecological pressures (Lee and Thornton, 2021).

Acknowledgements

We thank Dries Debeer for statistical support, and two anonymous reviewers and the specialist recommender for their helpful comments on the paper. We also thank Sophia Knoch for assisting with data validation, as well as to the staff of the ECoBird Centre, the Department of Biology at the University of Antwerp and the Wildlife Rescue Centre Ostend for general project support. Preprint Stage 1 version 2 of this article has been peer-reviewed and recommended by Peer Community In Registered Reports (https://osf.io/qvxgh).

Fundings and Ethics

This work is supported by a FWO (Flemish Research Foundation) PhD Fellowship grant (No. 11F0823N, awarded to Anneleen Dewulf), an ERC Consolidator grant (European Union's Horizon 2020 research and innovation programme, grant agreement No. 769595, awarded to Frederick Verbruggen), and a Methusalem grant (Methusalem Project 01M00221, Ghent University, awarded to Frederick Verbruggen, Luc Lens and An Martel). This experiment has been approved by the Ethical Committee for Animal Experimentation, Faculty of Sciences, Flemish Institute for Biotechnology (VIB) of Ghent University (EC2022-091, EC2024-023).

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, script, code, and supplementary information availability

Data, script, codes and supplementary information is available at the OSF repository (Dewulf, Garcia-Co, et al., 2023, https://osf.io/u63w2/?view_only=db4c1a620de841c28d7fc9a52e326cfd)

Authors' Contributions

'Conceptualization' – AD, CG, WM, JM, AM, LL, FV ; 'Data Curation' – AD; 'Formal Analysis' – AD; 'Funding Acquisition' – AD, AM, LL, FV; 'Investigation' - AD, CG; 'Methodology' - AD, CG, WM, JM, AM, LL, FV; 'Project Administration' – AD, FV; 'Software' – AD; 'Resources' – AD, WM, FV; 'Supervision' – JM, LL, FV ; 'Validation' – AD, JM, LL, FV; 'Visualization' – AD; 'Writing – Original Draft' – AD; 'Writing – Review & Editing' – AD, CG, WM, JM, AM, LL, FV.

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
1 Does detour performance improve when the perceptual characteristics of the barrier (in a detour task) match the ecological niche of a bird species?	**As we will test four bird species with substantially different ecological niches, the perception of a match between the perceptual characteristics and the ecological niche will be species-specific (see below). We therefore predict an interaction between barrier type and species.** **1:1** For white leghorn chickens *(Gallus gallus domesticus)* and Japanese quails *(Coturnix japonica)*: better detour performance for horizontal- (HB) compared with vertical-bar barriers (VB; thus, detour performance HB > VB). **1:2** For herring gulls *(Larus argentatus)*: better detour performance for vertical- compared with	We will test 60 individuals per species (total N = 240). *A-priori* power sensitivity analyses done in G*Power (Faul et al., 2009) indicate that this is sufficient to detect small effects; it is also the largest number that is practically feasible given our resource constraints and study design (see Farrar et al., (2020)) For the sensitivity analysis, we used a mixed ANOVA model with one between-subjects factor (4 levels; corresponding to our *Species* factor) and two within-subjects factor (one with 2 levels – *Barrier* – and one with 3 levels – *Trial*). This indicated that our sample size of 60 animals per species (240 in total) is sufficient to detect a *Species* x *Barrier* interaction effect (**Question 1**) with a small effect size (Cohen's *f* effect size of 0.12; Cohen, 1988; Power = .80; cor. among RM = 0.5; we used an alpha of .025 to correct for the fact that we will have two dependent variables measuring (slightly) different aspects of detour	A (G)LMM with type III sum of squares will be used to analyse our two dependent variables, namely 1) the latency to detour and the 2) cumulative time spent in the species-specific 'barrier zone of interest'(persisting). Both models will include the between-species factor: *Species* (i.e., 4 levels) and both within-species factors: *Barrier* (i.e., 2 levels) and *Trial* (i.e., 3 levels), and the two control variables (as covariates), namely (a) a *'muti-baseline'* measure of an individual's motivational state (and its interaction with *Species*, as we will mean-center this *'multi-baseline'* measure within *Species*), and (b) *Barrier Order* (i.e., 2 levels). Individual birds and enclosure (social group) will be included as random effects in the models, with individual	We performed a-priori power sensitivity analyses (alpha corrected = .025, Power = .80, cor. among RM = 0.5) with G*Power (Faul et al., 2009) for the different effects of interest and confirmed that our sample size is sufficient to detect small effects (with a Cohen's *f* effect size varying between 0.09 -0.12 ; Cohen, 1988). See sampling plan for more information.	This study is a partial replication of the study of Regolin et al., (1994) and the comparative study of Zucca et al., (2005). Across research fields, there is a clear need for these replication studies (see Farrar et al., 2020, for the comparative cognition field), as they are "a central part of the iterative maturing cycle of description, prediction, and explanation", and as such, play an "important, exciting, and generative role in scientific progress" (Nosek & Errington, 2020). As a consequence, the absence of a Barrier x Species interaction effect (which we predict on the basis of the previous studies) informs us about the	We propose that stop-signal detection (hence, barrier detection) is a crucial, cognitive building block of RI across species (Verbruggen et al., 2014), including birds. Here we will take this idea one step further and propose that ecologically valid signals are easier to detect (or to perceive as a stop signal), and this will enhance stopping. The role of stop-signal detection in avian response inhibition, and in particular, the interaction with the ecological niche of the species, should be revised if we cannot replicate the previous work (Regolin et al., 1994; Zucca et al., 2005).

	horizontal-bar barriers (thus, detour performance HB < VB). **1:3** For domestic canaries we do not expect differences between vertical- and horizontal-bar barriers (i.e., detour performance HB = VB)	performance). Second, our sample size is sufficient to detect a small effect of *Trial* (**Question 2**; Cohen's *f* effect size of 0.09; Cohen, 1988). Third, our sample size is sufficient to explore a small effect (Cohen's *f* effect size of 0.09; Cohen, 1988) for the *Species* x *Barrier* x *Trial* interaction effect (**Question 3: Explorative**). <u>Remark 1:</u> Our sensitivity analyses are based on mixed ANOVAs (fixed-effects models with between- and within-species factors). However, as discussed below, we will analyse our data with (G)LMMs, which are currently not covered by G*Power or most other power-estimation tools. These mixed-effect models are more flexible in assigning variance as they allow for the specification of both fixed and random effects. However, by accounting for unexplained variance (see below), our proposed mixed effect models are more powerful than the fixed-effect model ANOVAs used in our sensitivity analyses (and than the models used in the studies of Regolin,	birds nested in enclosures. In addition, we will include by-individual (nested in enclosures) random slopes that can vary for the levels of *Species* (corresponding with species-specific intercepts). Model plots will be generated by means of the package *performance* (Lüdecke et al., 2021) to inspect for violations of the model assumptions: 1) heteroscedasticity (plotting the square root of the residuals (y-axis) and fitted values (x-axis)), 2) non-normality of residuals (plotting the sample quantiles (y-axis) on the standard normal distribution quantiles), and 3) outliers (plotting standard residuals (y-axis) and leverage). Additionally, the multicollinearity between fixed main factors (via the variance inflation factor, VIF) and the autocorrelation between residuals (via a Durbin-Watson-Test) will be calculated via functions provided by the *performance* package (Lüdecke et al., 2021).		validity of the original findings (Regolin et al., 1994; Zucca et al., 2005) and can further emphasize on the need of replication in general.	

		Vallortigara, et al., 1994, and Zucca, Antonelli, et al., 2005). Thus, the sensitivity analyses discussed here are a conservative estimate. Remark 2: We will incubate 20% more eggs than the number of individuals required for testing (to account for possible drop outs and guarantee statistical power during the whole study)	Potential violations of model assumptions will be addressed by transforming the (in)dependent variables (i.e., via log-transformation) or by changing the error distribution (family) or the link function of the model (switching a default LMM that will be fitted to a GLMM). Fixed main effects with a VIF of >5 will be removed and logical outliers (i.e., recording/entry errors) will be inspected and corrected (if possible). In the case that the outlier cannot be corrected, all data of that individual will be excluded from all statistical analyses. In case we find (a) significant *Barrier* x *Species* interaction **-** effect(s) **(Question 1)** further post-hoc Bonferroni-Holm corrected linear contrasts upon the model will be performed to compare performance with different (ecological valid) barriers per species **(1:1, 1:2, 1:3)**.			

2 Does detour performance improve over trials?	**2:1** Detour performance will gradually improve over trials (thus, detour performance trial 1 < trial 2 < trial 3).		In case we find (a) significant main effect of *Trial* **(Question 2)** further post-hoc Bonferroni-Holm corrected linear contrasts upon the model will be performed to compare performance over trials **(2:1).**		The absence of a main effect of trial would demonstrate that detour performance is consistent over trials.	We propose that detour performance improves over trials. Extensive work on skill acquisition in humans has shown that performance generally improves rapidly at first and then more slowly over time (see e.g., Logan, 1988, Thorndike, 1913). If we do not find a difference between trials, this would indicate that detouring cannot be learned easily by avian species.
3 (explorative) Does the learning effect (i.e., improved detour performance across trials) depend on the ecological validity of the barrier?	**3:1 (explorative)** We will *explore* if the learning effect (i.e., improved detour performance across trials) interacts with the ecological validity of the stop signals. There are two possible patterns that would result in a three-way interaction between *Species*, *Barrier* (horizontal- vs. vertical-bar barriers), and *Trial* (1-3) (**Explorative Prediction 3**). First, detour performance might be better for ecologically valid compared with non-		In case we find (a) significant three-way *Species* x *Barrier* x *Trial* interaction effect(s) **(Question 3 explorative)**, further exploratory Bonferroni-Holm corrected linear contrasts upon the model will be performed to compare *Trial* performance of *Species* on different types of the *Barrier* **(3:1 explorative).**		Our design enables us to investigate potential interaction effects with Trial (e.g., the three-way *Species* x *Barrier* x *Trial* interaction effect, **Question 3 explorative**), with which we can further **explore** whether learning effects are completely absent in our design or are influenced by the (ecological validity) of the different barrier types.	We will *explore* if the learning effect (i.e., improved detour performance across trials) will depend on the ecological validity of signals. If we do not find such a three-way interaction effect, we can conclude (a) that superior detour performance with ecological valid than non-valid trials is independent of trial number (in case we do find a *Species* x *Barrier* interaction

	valid stop signals at the beginning, but this pattern might diminish over time as individuals learn to stop (i.e., the differences between barrier types would decrease). Second, detour performance might be poor at the beginning for both barrier types, but learning to stop might be easier for ecologically valid signals compared with non-valid stop signals (i.e., the differences between barrier types would increase). Both patterns would be theoretically meaningful, but we do not have *a-priori* predictions about the direction of the three-way interaction.					effect) or (b) that the interaction between the stop signal and the ecological niche of the species should be revised (in case we do not find a *Species* x *Barrier* interaction effect).

Guidance Notes

- **Question**: articulate each research question being addressed in one sentence.
- **Hypothesis**: where applicable, a prediction arising from the research question, stated in terms of specific variables rather than concepts. Where the testability of one or more hypotheses depends on the verification of auxiliary assumptions (such as positive controls, tests of intervention fidelity, manipulation checks, or any other quality checks), any tests of such assumptions should be listed as hypotheses. Stage 1 proposals that do not seek to test hypotheses can ignore or delete this column.
- **Sampling plan**: For proposals using inferential statistics, the details of the statistical sampling plan for the specific hypothesis (e.g power analysis, Bayes Factor Design Analysis, ROPE etc). For proposals that do not use inferential statistics, include a description and justification of the sample size.

- **Analysis plan**: For hypothesis-driven studies, the specific test(s) that will confirm or disconfirm the hypothesis. For non-hypothesis-driven studies, the test(s) that will answer the research question.
- **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis**: For hypothesis-driven studies that employ inferential statistics, an explanation of how the authors determined a relevant effect size for statistical power analysis, equivalence testing, Bayes factors, or other approach.
- **Interpretation given different outcomes**: A prospective interpretation of different potential outcomes, making clear which outcomes would confirm or disconfirm the hypothesis.
- **Theory that could be shown wrong by the outcomes**: Where the proposal is testing a theory, make clear what theory could be shown to be wrong, incomplete, or otherwise inadequate by the outcomes of the research.

References

Ashton BJ, AR Ridley, EK Edwards, and A Thornton (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. en. *Nature* 554.7692, 364–367. ISSN: 0028-0836, 1476-4687. https://doi.org/10.1038/nature25503.

Bari A and TW Robbins (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. en. *Progress in Neurobiology* 108, 44–79. ISSN: 03010082. https://doi.org/10.1016/j.pneurobio.2013.06.005.

Bates D, M Mächler, B Bolker, and S Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.1, 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates D, M Maechler, and M Jagan (2023). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.6-2.

Bekker EM, JL Kenemans, MR Hoeksma, D Talsma, and MN Verbaten (2005). The pure electrophysiology of stopping. en. *International Journal of Psychophysiology* 55.2, 191–198. ISSN: 01678760. https://doi.org/10.1016/j.ijpsycho.2004.07.005.

Ben-Shachar MS, D Lüdecke, and D Makowski (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software* 5.56. Publisher: The Open Journal, 2815. https://doi.org/10.21105/joss.02815.

Boehler CN, TF Munte, RM Krebs, H Heinze, MA Schoenfeld, and J Hopf (2009). Sensory MEG Responses Predict Successful and Failed Inhibition in a Stop-Signal Task. en. *Cerebral Cortex* 19.1, 134–145. ISSN: 1047-3211, 1460-2199. https://doi.org/10.1093/cercor/bhn063.

Bollweg G and S Sparber (1998). Relationships Between Midembryonic 5-HT2 Agonist and/or Antagonist Exposure and Detour Learning by Chickens. en. *Pharmacology Biochemistry and Behavior* 60.1, 47–53. ISSN: 00913057. https://doi.org/10.1016/S0091-3057(97)00555-8.

Brooks ME, K Kristensen, KJ van Benthem, A Magnusson, CW Berg, A Nielsen, HJ Skaug, M Maechler, and BM Bolker (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* 9.2, 378–400. https://doi.org/10.32614/RJ-2017-066.

Bunge SA (2012). The Developing Human Brain: A Frontiers Research Topic. en. *Frontiers in Human Neuroscience* 6. ISSN: 1662-5161. https://doi.org/10.3389/fnhum.2012.00047.

Chen G, NE Adleman, ZS Saad, E Leibenluft, and RW Cox (2014). Applications of multivariate modeling to neuroimaging group analysis: a comprehensive alternative to univariate general linear model. eng. *NeuroImage* 99. Place: United States, 571–588. ISSN: 1095-9572 1053-8119. https://doi.org/10.1016/j.neuroimage.2014.06.027.

Cohen J (1988). *Statistical power analysis for the behavioral sciences*. en. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates. ISBN: 978-0-8058-0283-2.

Devereux CL, MJ Whittingham, E Fernández-Juricic, JA Vickery, and JR Krebs (2006). Predator detection and avoidance by starlings under differing scenarios of predation risk. en. *Behavioral Ecology* 17.2, 303–309. ISSN: 1465-7279, 1045-2249. https://doi.org/10.1093/beheco/arj032.

Dewulf A, C Garcia-Co, W Müller, JR Madden, A Martel, L Lens, and F Verbruggen (2023). Do Ecological Valid Stop Signals Aid Detour Performance? A Comparison of Four Bird Species. Available at: https://osf.io/u63w2/.

Dewulf A, S Knoch, JR Madden, A Martel, L Lens, and F Verbruggen (2025). Unpublished manuscript.

Dudusola I and H Bashiru (2020). Repeatability estimates of body weight and shank length in Japanese quail (Coturnix coturnix Japonica). *Nigerian Journal of Animal Science* 22.1, 27–32.

Elchlepp H and F Verbruggen (2017). How to withhold or replace a prepotent response: An analysis of the underlying control processes and their temporal dynamics. en. *Biological Psychology* 123, 250–268. ISSN: 03010511. https://doi.org/10.1016/j.biopsycho.2016.10.005.

Farrar B, M Boeckle, and N Clayton (2020). Replications in Comparative Cognition: What Should We Expect and How Can We Improve? en. *Animal Behavior and Cognition* 7.1, 1–22. ISSN: 23725052, 23724323. https://doi.org/10.26451/abc.07.01.02.2020.

Faul F, E Erdfelder, A Buchner, and AG Lang (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. en. *Behavior Research Methods* 41.4, 1149–1160. ISSN: 1554-351X, 1554-3528. https://doi.org/10.3758/BRM.41.4.1149.

Fox J and S Weisberg (2019). *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage.

Fox J, S Weisberg, and B Price (2022). *carData: Companion to Applied Regression Data Sets*.

Friard O and M Gamba (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in ecology and evolution* 7.11. Publisher: Wiley Online Library, 1325–1330.

Garcia-Co C, J Morales, F Verbruggen, and W Müller (2024). Unpublished manuscript. Department of Biology, Behavioural Ecology and Eco-Physiology Group, University of Antwerp.

Gjøen J, H Jean-Joseph, K Kotrschal, and P Jensen (2023). Domestication and social environment modulate fear responses in young chickens. en. *Behavioural Processes* 210, 104906. ISSN: 03766357. https://doi.org/10.1016/j.beproc.2023.104906.

Hartig F (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.6.

Holm S (1979). A Simple Sequentially Rejective Multiple Test Procedure. en. *Scand J Statist* 6.2, 65–70.

Johnson-Ulrich L and KE Holekamp (2020). Group size and social rank predict inhibitory control in spotted hyaenas. en. *Animal Behaviour* 160, 157–168. ISSN: 00033472. https://doi.org/10.1016/j.anbehav.2019.11.020.

Kabadayi C, K Bobrowicz, and M Osvath (2018). The detour paradigm in animal cognition. en. *Animal Cognition* 21.1, 21–35. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-017-1152-0.

Kabadayi C, I Jacobs, and M Osvath (2017). The Development of Motor Self-Regulation in Ravens. en. *Frontiers in Psychology* 8, 2100. ISSN: 1664-1078. https://doi.org/10.3389/fpsyg.2017.02100.

Kabadayi C, A Krasheninnikova, L O'Neill, J van de Weijer, M Osvath, and AMP von Bayern (2017). Are parrots poor at motor self-regulation or is the cylinder task poor at measuring it? en. *Animal Cognition* 20.6, 1137–1146. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-017-1131-5.

Kaiser S, MB Hennessy, and N Sachser (2015). Domestication affects the structure, development and stability of biobehavioural profiles. en. *Frontiers in Zoology* 12.Suppl 1, S19. ISSN: 1742-9994. https://doi.org/10.1186/1742-9994-12-S1-S19.

Kuznetsova A, PB Brockhoff, and RHB Christensen (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82.13, 1–26. https://doi.org/10.18637/jss.v082.i13.

Lee VE and A Thornton (2021). Animal Cognition in an Urbanised World. en. *Frontiers in Ecology and Evolution* 9, 633947. ISSN: 2296-701X. https://doi.org/10.3389/fevo.2021.633947.

Lenth RV (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*.

Lüdecke D, MS Ben-Shachar, I Patil, P Waggoner, and D Makowski (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software* 6.60, 3139. https://doi.org/10.21105/joss.03139.

MacLean EL, B Hare, CL Nunn, E Addessi, F Amici, RC Anderson, F Aureli, JM Baker, AE Bania, AM Barnard, NJ Boogert, EM Brannon, EE Bray, J Bray, LJN Brent, JM Burkart, J Call, JF Cantlon, LG Cheke, NS Clayton, MM Delgado, LJ DiVincenti, K Fujita, E Herrmann, C Hiramatsu, LF Jacobs, KE Jordan, JR Laude, KL Leimgruber, EJE Messer, AC de A. Moura, L Ostoji, A Picard, ML Platt, JM Plotnik, F Range, SM Reader, RB Reddy, AA Sandel, LR Santos, K Schumann, AM Seed, KB Sewall, RC Shaw, KE Slocombe, Y Su, A Takimoto, J Tan, R Tao, CP van Schaik, Z Viranyi, E Visalberghi, JC Wade, A Watanabe, J Widness, JK Young, TR Zentall, and Y Zhao (2014). The evolution of self-control. en. *Proceedings of the National Academy of Sciences* 111.20, E2140–E2148. ISSN: 0027-8424, 1091-6490. https://doi.org/10.1073/pnas.1323533111.

McHugh ML (2012). Interrater reliability: the kappa statistic. en. *Biochemia Medica*, 276–282. ISSN: 18467482. https://doi.org/10.11613/BM.2012.031.

Miller DB (1974). A comparative analysis of detour behavior in ring doves (Streptopelia risoria) and rock doves (Columbia livia) as a function of barrier density. PhD Thesis. University of Miami.

Miller DB and RB Tallarico (1974). On the correlation of brain size and problem-solving behavior of ring doves and pigeons. *Brain, Behavior and Evolution* 10, 265–273. https://doi.org/https://doi.org/10.1159/000124318.

Müller W, J Vergauwen, and M Eens (2008). Yolk testosterone, postnatal growth and song in male canaries. en. *Hormones and Behavior* 54.1, 125–133. ISSN: 0018506X. https://doi.org/10.1016/j.yhbeh.2008.02.005.

Navarro D (2015). *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6)*. R package version 0.5.1. University of New South Wales. Sydney, Australia.

Nebel C, P Sumasgutner, A Pajot, and A Amar (2019). Response time of an avian prey to a simulated hawk attack is slower in darker conditions, but is independent of hawk colour morph. en. *Royal Society Open Science* 6.8, 190677. ISSN: 2054-5703. https://doi.org/10.1098/rsos.190677.

Pani P, F Giarrocco, M Giamundo, R Montanari, E Brunamonti, and S Ferraina (2018). Visual salience of the stop signal affects the neuronal dynamics of controlled inhibition. en. *Scientific Reports* 8.1, 14265. ISSN: 2045-2322. https://doi.org/10.1038/s41598-018-32669-8.

Pinheiro JC and DM Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer. https://doi.org/10.1007/b98882.

Press C, D Yon, and C Heyes (2022). Building better theories. en. *Current Biology* 32.1, R13–R17. ISSN: 09609822. https://doi.org/10.1016/j.cub.2021.11.027.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Regolin L, V Giorgio, and Z Mario (1995). Object and spatial representations in detour problems by chicks. en. *Animal Behaviour* 49.1, 195–199. ISSN: 00033472. https://doi.org/10.1016/0003-3472(95)80167-7.

Regolin L, G Vallortigara, and M Zanforlin (1994). Perceptual and motivational aspects of detour behaviour in young chicks. *Animal Behaviour* 47, 123–131. https://doi.org/https://doi.org/10.1006/anbe.1994.1014.

Salinas E and TR Stanford (2013). The Countermanding Task Revisited: Fast Stimulus Detection Is a Key Determinant of Psychophysical Performance. en. *Journal of Neuroscience* 33.13, 5668–5685. ISSN: 0270-6474, 1529-2401. https://doi.org/10.1523/JNEUROSCI.3977-12.2013.

Tacutu R, D Thornton, E Johnson, A Budovsky, D Barardo, T Craig, E Diana, G Lehmann, D Toren, J Wang, VE Fraifeld, and JP de Magalhães (2018). Human Ageing Genomic Resources: new and updated databases. en. *Nucleic Acids Research* 46.D1, D1083–D1090. ISSN: 0305-1048, 1362-4962. https://doi.org/10.1093/nar/gkx1042.

Thygesen UH, CM Albertsen, CW Berg, K Kristensen, and A Nielsen (2017). Validation of ecological state space models using the Laplace approximation. *Environmental and Ecological Statistics*, 1–23. https://doi.org/10.1007/s10651-017-0372-4.

Troisi CA, A Vernouillet, R Allaert, S Knoch, A Martel, L Lens, and F Verbruggen (2024). Manuscript in preparation. in prep.

Tvardíková K and R Fuchs (2012). Tits recognize the potential dangers of predators and harmless birds in feeder experiments. en. *Journal of Ethology* 30.1, 157–165. ISSN: 0289-0771, 1439-5444. https://doi.org/10.1007/s10164-011-0310-0.

Ueno A and K Suzuki (2014). Comparison of learning ability and memory retention in altricial (Bengalese finch, *Lonchura striata var. domestica*) and precocial (blue-breasted quail, *Coturnix chinensis*) birds using a color discrimination task: Visual Learning and Retention in Birds. en. *Animal Science Journal* 85.2, 186–192. ISSN: 13443941. https://doi.org/10.1111/asj.12092.

Van Der Schoot M, R Licht, TM Horsley, and JA Sergeant (2005). Effects of stop signal modality, stop signal intensity and tracking method on inhibitory performance as determined by use of the stop signal paradigm: Factors affecting inhibitory performance in the stop task. en. *Scandinavian Journal of Psychology* 46.4, 331–341. ISSN: 00365564. https://doi.org/10.1111/j.1467-9450.2005.00463.x.

Van Horik J, CE Beardsworth, PR Laker, EJG Langley, MA Whiteside, and JR Madden (2019). Unpredictable environments enhance inhibitory control in pheasants. en. *Animal Cognition* 22.6, 1105–1114. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-019-01302-0.

Van Horik J, CE Beardsworth, PR Laker, MA Whiteside, and JR Madden (2020). Response learning confounds assays of inhibitory control on detour tasks. en. *Animal Cognition* 23.1, 215–225. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-019-01330-w.

Van Horik J, EJG Langley, MA Whiteside, PR Laker, CE Beardsworth, and JR Madden (2018). Do detour tasks provide accurate assays of inhibitory control? en. *Proceedings of the Royal Society B: Biological Sciences* 285.1875, 20180150. ISSN: 0962-8452, 1471-2954. https://doi.org/10.1098/rspb.2018.0150.

Verbruggen F, M Best, WA Bowditch, T Stevens, and IPL McLaren (2014). The inhibitory control reflex. *Neuropsychologia* 65, 263–278. https://doi.org/10.1016/j.neuropsychologia.2014.08.014.

Verbruggen F and GD Logan (2008a). Automatic and controlled response inhibition: associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General* 137.4, 649–72. ISSN: 0096-3445. https://doi.org/10.1037/a0013170.

— (2008b). Response inhibition in the stop-signal paradigm. en. *Trends in Cognitive Sciences* 12.11, 418–424. ISSN: 13646613. https://doi.org/10.1016/j.tics.2008.07.005.

— (2017). Control in Response Inhibition. en. In: *The Wiley Handbook of Cognitive Control*. Ed. by Egner T. Chichester, UK: John Wiley & Sons, Ltd, pp. 97–110. ISBN: 978-1-118-92049-7 978-1-118-92054-1. https://doi.org/10.1002/9781118920497.ch6.

Verbruggen F, IPL McLaren, and CD Chambers (2014). Banishing the Control Homunculi in Studies of Action Control and Behavior Change. en. *Perspectives on Psychological Science* 9.5, 497–524. ISSN: 1745-6916, 1745-6924. https://doi.org/10.1177/1745691614526414.

Verbruggen F, T Stevens, and CD Chambers (2014). Proactive and reactive stopping when distracted: An attentional account. en. *Journal of Experimental Psychology: Human Perception and Performance* 40.4, 1295–1300. ISSN: 1939-1277, 0096-1523. https://doi.org/10.1037/a0036542.

Vernouillet A, J Anderson, D Clary, and DM Kelly (2016). Inhibition in Clark's nutcrackers (Nucifraga columbiana): results of a detour-reaching test. en. *Animal Cognition* 19.3, 661–665. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-016-0952-y.

Willcox K, A Vernouillet, L Lens, and F Verbruggen (2024). Early-Life Group Size Does Not Influence Japanese Quails' Learning in a Response Inhibition Task. *bioRxiv*. https://doi.org/10.1101/2024.06.19.599677. eprint: https://www.biorxiv.org/content/early/2024/06/19/2024.06.19.599677.full.pdf.

Yeasmin T and M Howlider (2013). Effects of autosomal dwarf gene on growth and shank length of chicken. en. *Bangladesh Veterinarian* 30.1, 25–32. ISSN: 1012-5949. https://doi.org/10.3329/bvet.v30i1.16282.

Zucca P, F Antonelli, and G Vallortigara (2005). Detour behaviour in three species of birds: quails (Coturnix sp.), herring gulls (Larus cachinnans) and canaries (Serinus canaria). en. *Animal Cognition* 8.2, 122–128. ISSN: 1435-9448, 1435-9456. https://doi.org/10.1007/s10071-004-0243-x.

Zucca P and VA Sovrano (2008). Animal lateralization and social recognition: Quails use their left visual hemifield when approaching a companion and their right visual hemifield when approaching a stranger. en. *Cortex* 44.1, 13–20. ISSN: 00109452. https://doi.org/10.1016/j.cortex.2006.01.002.