,

Neophobia across social contexts in juvenile herring gulls

Reinoud Allaert[1,2], Sophia Knoch[2,3], Simon Braem[1,2], Dries Debeer[4], An Martel[2,5], Wendt Müller[6], Eric Stienen[7], Luc Lens[1,2], & Frederick Verbruggen[1,2,3]

[1] Dept. Biology, Ghent University
[2] Centre for Research on Ecology, Cognition and Behaviour of Birds
[3] Dept. Experimental Psychology, Ghent University
[4] Research Support Office, Faculty of Psychology and Educational Sciences,
[5] Dept. Pathobiology, Pharmacology and Zoological Medicine, Ghent University
[6] Dept. Biology, Antwerp University
[7] Research Institute for Nature and Forest

Correspondence: reinoud.allaert@ugent.be

Abstract

Neophobia, the fear or avoidance of the unfamiliar, can have significant fitness consequences. It is typically assessed by exposing individuals to unfamiliar objects when they are alone, but in social species the presence of conspecifics can influence neophobia. However, previous research on the effect of group dynamics on neophobic responses has produced mixed results. Here, we explored the degree of neophobia of an individual in different social contexts in a highly social species, the herring gull. To this end, we exposed juvenile herring gulls to novel objects in both individual and group settings, repeating each condition twice to establish reproducibility. Individuals in groups were quicker to eat, and spent more time near a novel object than individuals tested alone. The results of our study suggest that individuals mitigate risk by distributing it among group members. Preregistered Stage 1 protocol: https://osf.io/qvxgh (date of in-principle acceptance: 17/05/2023)

Keywords: Animal Behaviour, Behavioural Inhibition, Neophobia, Social Behaviour, Herring Gull, Animal Personality

Introduction

Neophobia is the fear or reluctance to engage with new or unfamiliar objects, places or scenarios. It is often considered to be a consistent personality trait across species, affecting an individual's survival and adaptation (Both et al., 2005; Greggor et al., 2015; Kimball and Lattin, 2023; Vrublevska et al., 2015). Research into animal behaviour is increasingly focusing on neophobia because of its significance in the context of rapid environmental change. The world is rapidly urbanising, with the footprint of urban land cover expected to at least double by the end of the century (Gao and O'Neill, 2020). Many species must therefore adapt to human-induced changes in their environment, and hence, to unfamiliar scenarios (Lee and Thornton, 2021; McKinney, 2002). In such situations, neophobia can, on the one hand, serve as a survival mechanism, allowing individuals to avoid potential threats and increase their chance of survival (Greenberg and Mettke-Hofmann, 2001). On the other hand, excessive aversion to novelty can restrict exploratory behaviour, limiting an individual's ability to locate and exploit novel resources, learn from its novel environment and adapt to environmental changes (Biondi et al., 2010; Greenberg, 2003).

To assess neophobia, individuals are typically exposed to novel food, objects, or spaces (Greggor et al., 2015; Mettke-Hofmann, 2017). For example, in the 'novel object task', which we ~~use~~ used in the present study, an individual encounters an unfamiliar object, often placed next to a food reward, in a familiar environment. The latency to approach the food (in the presence of the novel object) or to interact with the novel object itself, is then used as a measure of neophobia (Greggor et al., 2015; Miller, Lambert, et al., 2022; Vernouillet and DM Kelly, 2020). These measures have been used in cross-species comparisons to investigate, for example, the socio-ecological drivers of neophobia (Mettke-Hofmann et al., 2002; Miller, Lambert, et al., 2022), or within species, to investigate both the causes and consequences of individual differences in neophobia (Greenberg and Mettke-Hofmann, 2001).

Most research on neophobia has focused on individual animals, both in laboratory and field settings. However, it is important to consider that many species are to various extents reliant on social information, so individuals can influence each other's behaviour. This is also true in the context of adapting to environmental changes and urbanisation (Lee and Thornton, 2021). For instance, when individuals encounter a new environment, they may learn from others about appropriate roosting or nesting sites, food sources, or unfamiliar predators (Harel et al., 2017; Keen et al., 2020; Loukola et al., 2012). In this context, several studies suggest that the presence of conspecifics also influences neophobia. However, the mechanisms behind this social phenomenon are still a topic of debate due to the various patterns that have been observed.

First, some studies have found that individuals in groups are generally less neophobic than when tested alone. For example, Coleman and Mellgren presented zebra finches (*Taeniopygia guttata*) with novel feeders and decorated the feeders with novel objects (Coleman and Mellgren, 1994). Individuals in a group approached and started using the new and decorated feeders more quickly than when tested alone. Other studies reported similar patterns in different species for some (but not necessarily all) measures of neophobia (Benson-Amram and Holekamp, 2012; Kareklas et al., 2018; Moretti et al., 2015; Soma and Hasegawa, 2004). Such mitigating effects of social context on neophobia may be attributed to 'risk dilution' (Krause and Ruxton, 2002) or 'social buffering' (Kikusui et al., 2006). These theories predict that neophobia, or fear responses in general, are reduced in the presence of others, as individuals in a group collectively share the potential risks associated with novel situations or threats, causing them to behave more similarly.

Second, some studies found the opposite pattern. For example, common ravens (*Corvus corax*) and carrion x hooded crows (hybrid; *C. corone, C. cornix*) approached novel objects faster when alone than when accompanied by a conspecific (Miller, Bugnyar, et al., 2015; Stöwe, Bugnyar, Heinrich, et al., 2006; Stöwe, Bugnyar, Loretto, et al., 2006). Other studies have observed similar patterns in other species, including Indian mynahs, *Acridotheres tristis* (Griffin et al., 2013), house sparrows, *Passer domesticus* (TR Kelly et al., 2020), and even zebra finches (Kerman et al., 2018; St. Lawrence et al., 2021), thus failing to replicate the findings of the aforementioned study by Coleman and Mellgren (1994). Interestingly, however, some of these studies found that once

individuals reached the novel object, they spent more time interacting with it when in the presence of others (either in pairs or in groups) than when isolated (Miller, Bugnyar, et al., 2015; St. Lawrence et al., 2021; Stöwe, Bugnyar, Heinrich, et al., 2006). It has therefore been suggested that the slower approach latencies may be due to conspecifics 'negotiating', by using behavioural cues to coordinate their actions and deciding who will approach the novel object first. Consequently, this may lead to a convergence of individual behaviours, as group members align their actions based on these cues.

Third, some studies failed to find effects of social context on average neophobic responses altogether (e.g Apfelbeck and Raess, 2008). While ~~,~~it is of course possible that social context does not matter for some species, it is also possible that the presence of conspecifics alters behaviour of individuals without changing the mean response. Specifically, in environments where conspecifics' behaviour serves as an indicator of appropriate responses, individuals may adjust their own behaviour to match that of others (Herbert-Read et al., 2013). This synchronisation of behaviours within the group, or 'social conformity', enhances cohesion and helps the group to adapt to their environment. ~~For example, observations~~ Observations in a variety of species, such as zebra finches (Schuett and Dall, 2009) and gouldian finches, *Erythrura gouldiae* (King et al., 2015), show how individuals adapt their behaviour and mirror their partners' character traits. For instance, if a gouldian finch exhibited bold behaviour, the observing individual tended to become bolder as well, while if the partner displayed shyness, the observing individual mirrored this trait (King et al., 2015). Thus, this study found that the neophobic response was similar on average for individuals tested alone or in pairs, but there was less variation between individuals in the paired condition compared to the alone condition.

Current study The aim of this study is to investigate if and how the social context affects neophobia in the herring gull (*Larus argentatus*). Gulls' natural coastal habitat is rapidly disappearing, forcing them to live closer to humans in urban environments and to rely more on anthropogenic food sources (Coulson, 2015; Nager and O'Hanlon, 2016). Although reports in popular media may suggest that herring gulls are generally not neophobic due to their approach towards humans or stealing food, such anecdotes do not necessarily reflect the species' behaviour at a population level (Inzani et al., 2023). In fact, ~~significant~~ widely differing levels of neophobia as well as individual differences therein exist within populations (Inzani et al., 2023). The latter finding suggests that for some individuals, it might be easier to adapt to environmental change and urbanisation than for others. Indeed there is considerable intraspecific variation in how herring gulls utilise urbanised areas, ranging from minimally to almost complete dependence (O'Hanlon et al., 2017; Pavlova and Wronski, 2020).

Herring gulls are a highly social species, utilising cues not only from conspecifics, but even from other species, including humans. This suggests that social learning is a key aspect of gull behaviour (Feist et al., 2023; Frings et al., 1955; Gandolfi, 2009; Goumas et al., 2020). Thus, when assessing their neophobia, it is important to do this not only in an individual context, but also in a social (group) context.

Based on previous findings (as reviewed above), we predict that the distribution of neophobic responses will depend on the social context. However, the direction of the effects will depend on the social mechanisms at play. In Figure 1, we provide a template for testing the three different hypotheses of group effects, taking into account two measures, namely the average neophobic response and the variance between individuals.

Overall, we predict that there will be lower variance between individuals when they are tested in a group, compared to when they are tested alone. After all, all of the major hypotheses discussed above assume that individuals become more similar to each other by spreading risk, jointly buffering stress, negotiating with each other, or simply through social conformity. However, there are three possible scenarios regarding the average neophobic response. First, the 'risk dilution' hypothesis predicts that herring gulls will be *less* neophobic on average when in a group compared to when they are alone (scenario A in Figure 1). Second, the 'negotiation' hypothesis predicts that individuals will ~~approach novel objects~~ be ~~*slower*~~*more* neophobic when in group (scenario B in Figure 1). Third, according to the 'social conformity' hypothesis, individuals will tend to mimic one another's behaviours—those who are neophobic will show a decrease in their fear of novel objects when surrounded by others who are less neophobic, and vice versa (scenario C in Figure 1). Thus, in this third

scenario, there is a reduction of variance but no change in the average neophobic response. These three predictions are contrasted with the null hypothesis that social context does not modulate variance, or group means ('Null Hypothesis', Figure 1).

To test these predictions, juvenile herring gulls ~~will be~~ were subjected to four distinct conditions: individual or group tests paired with a control or novel object. Each condition ~~will be~~ was repeated twice. The guidelines for designing neophobia tests of Greggor et al. (2015) were followed, and a within-subject design with a relatively large sample size (N = ~~80~~67 individuals) was chosen to further increase the statistical power of the study. One additional reason for the inconsistent previous findings is that sample size was relatively low in many studies (see also Farrar et al., 2020). In addition, the herring gulls used in this study ~~will be~~ were raised by hand from the egg to control for sampling bias, a recurring issue when testing wild animals. After testing, they ~~will be~~ were released in the wild.

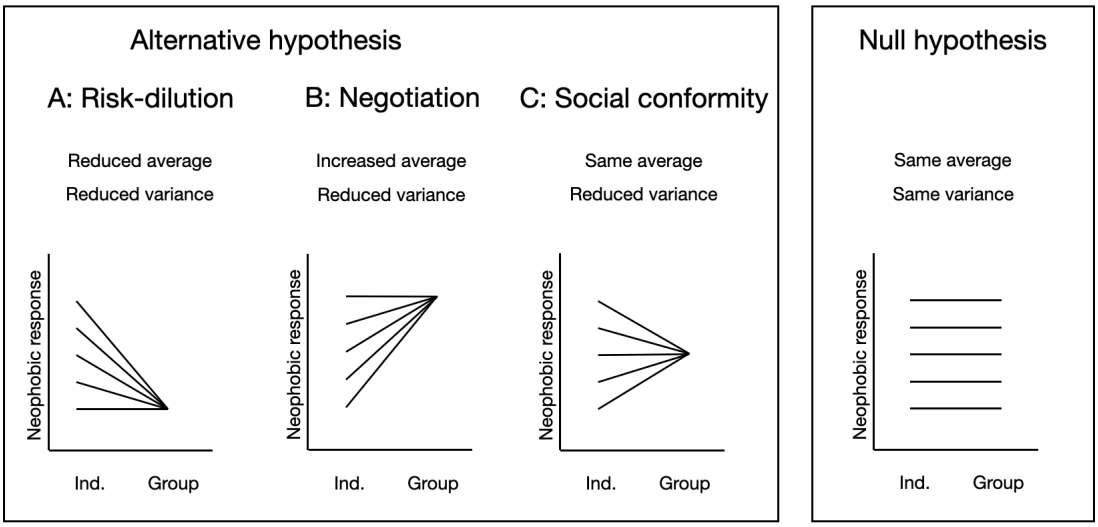

Figure 1. Overview of hypotheses

Material and methods

Sample size

We ~~will test 80*~~ originally planned to test 80 herring gulls twice across a 2x2 design (thus eight tests per individual; see above). We performed an *a-priori* power sensitivity ~~analyses~~ analysis using G*Power (Erdfelder et al., 2009), for a repeated measures MANOVA with three within-subject factors: ~~*Context*~~*context* (with levels ~~Group and Individual),~~ group and individual), ~~*Object*~~*object* (with levels ~~Control and Novel Object~~control and novel object), and ~~*Trial*~~*trial* (with levels 1 and 2). Our ~~sample size is~~ initial analysis indicated that a sample size of 80 would be sufficient to detect small main effects of ~~*Context*~~*context*, ~~*Object*~~*object*, *and* ~~*Trial*~~*trial* (Cohen's f effect size of 0.11 (Cohen, 2013); Power = 0.80 ; cor. among RM = 0.5), as well as an interaction between ~~Context and Object~~ context and object with small effect size (0.11; Power = 0.80 ; cor. among RM = 0.5). ~~Our sensitivity analyses are based on MANOVAs (repeated-measures, within-species~~ We reared the gulls from the eggs (see the 'Subject section' below) and we anticipated that in some cases herring gull eggs would be mistaken for those of the phylogenetically and ecologically related lesser black-backed gull (LBBG) during egg collection. For logistical reasons, the chicks could only be identified to the species level after testing by visual inspection of plumage differences. To mitigate the potential reduction in sample size (due to the exclusion of LBBGs), we conducted a second *a-priori* power analysis accounting for a potential 10% dropout rate. This a-priori analysis revealed that even with a 10% reduction, our study would still have sufficient statistical power (Cohen's f effect size of 0.17) to detect significant effects.

Due to unanticipated mortality, we were only able to test 67 birds (instead of the registered 80). Of these 67 birds, 13 individuals were later identified as LBBG (a higher percentage than we had anticipated) and were excluded from further analysis in accordance with the registered protocol. This further reduced our final sample size (N = 54). Although this is a significant reduction from our planned sample size (N = 72 after exclusion of LBBG), it is important to note that our sensitivity analyses were based on repeated measures MANOVAs (within species factors). ~~However, as discussed below, we will analyse our data with~~ This type of analysis does not take into account the additional flexibility offered by (G)LMMs, which are ~~currently not~~ not currently 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~~ power estimation tools. The mixed effects models used in this study (in line with the registered protocol) are more robust and better equipped to deal with unexplained variance than the fixed effects MANOVAs used in our sensitivity analysis. Thus, despite the reduction in sample size, our proposed mixed-effects models are expected to retain sufficient power to detect the effects of interest. ~~By accounting for unexplained variance, our proposed mixed-effect models are more powerful than the fixed-effect MANOVAs used in our sensitivity analyses.~~

~~*As gulls are reared from the egg, in a small number of cases (typically less than 10%), herring gull eggs are mistaken for those of the phylogenetically and ecologically related lesser black-backed gull. The species can only be determined after testing (when the individuals are older). Test data from lesser black-backed gulls (if any) will be excluded from subsequent analysis. We conducted a power analysis that accounts for a potential 10% drop-out to ensure that even with this potential reduction, our study would still have sufficient statistical power (Cohen's f effect size of 0.17) to detect significant effects.~~

Subjects

Egg Collection and Incubation

The herring gulls used in this study are part of a larger research project and ~~are~~ were raised and tested at the avian research facilities of Ghent University (Lab number LA1400452), located at the Wildlife Rescue Centre (WRC) in Ostend, Belgium. Eggs ~~are~~ were collected in May and June 2024, from nests of roof-breeding parents, by the Research Institute for Nature and Forest (INBO) under the license of the Agentschap voor Natuur en Bos (ANB) and the ~~gull patrol~~ 'gull patrol' team, authorised to remove eggs along the Belgian coasts for nuisance prevention. Collected before the pipping stage, the eggs ~~are~~ were transported to the WRC under stable conditions for further incubation, using Brinsea Ova-Easy incubators (temperature = 37.5°C; humidity = 45%). Upon arrival eggs ~~are~~ were marked with a unique nest identifier and the two largest eggs~~are~~, which are typically the first laid eggs of a clutch (Parsons, 1972), were incubated. They ~~are~~ were checked twice daily for small cracks, indicating pipping. Eggs showing signs of pipping, ~~are~~ were moved to a MS700U Hatchery (temperature = 37.2°C; humidity = 50%).

Chick Rearing

Once hatched and fully dried, the chicks ~~receive~~ received a unique combination of colour rings for identification. The chicks ~~are~~ were then housed in groups of 10 in boxes with netting bottoms (size = 120 x 60 x 60cm, LWH) within heated rooms (ambient temperature= 15-25°C; humidity=40%-80%; under natural light conditions). Each box ~~contains~~ contained a heating plate (30 x 30cm). The semi-precocial chicks ~~are~~ were hand-fed small pieces of fish and dog pellets soaked in water, supplemented with Akwavit, a complementary feed specially developed for fish eating animals (Kasper Faunafood, The Netherlands). Food ~~is~~ was available *ad libitum*. Once the chicks ~~are~~ were at least 5 days old and their weight ~~exceeds~~ exceeded 60 grams, they ~~are~~ were moved to outside enclosures (size = 500 x 205 x 265cm, LWH), housed in stable groups of 10 individuals. Outside, heating plates ~~are~~ were provided during the first few days ~~when~~ if night-time temperatures ~~are forecast~~ were forecasted to drop below 5°C, or in the event of adverse weather conditions such as heavy rain or storms. Food ~~consists~~ consisted of a mixture of dog pellets soaked in water and fish, provided 4 times per

day, following the default policy at the WRC. Water ~~is~~ was provided *ad libitum*. Individuals ~~are~~ were tested when they ~~are~~ were approximately 30 days old, shortly before they ~~reach~~ reached fledging age. After testing, the birds ~~are~~ were moved to a large flight cage (approximately 180m^2) for dehabituation from handling. Once they ~~are~~ were 8-10 weeks old, birds ~~are~~ were released in the wild, and a subset (n = ~~50) receives a GPS-tracker~~23) received a GPS-device.

Behavioural Test: Novel Object Task

Task Design: For testing purposes, each home enclosure containing ~~ten birds is~~ eight to ten birds was pseudo-randomly divided into two stable testing groups of ~~five known individuals . This division ensures nestmates are~~ four to five individuals that were familiar with each other. Within these subgroups, we ensured that nestmates were not placed in the same testing group. This arrangement ~~allows~~ allowed to maintain consistent housing conditions when not testing, while ~~facilitating specific configurations during testing sessions .~~ ensuring that testing sessions consistently involved the same subgroups of four to five individuals.

In the 'novel object' condition, birds ~~are~~ were exposed to a pseudo-randomly selected novel object~~()~~. Conversely, in the 'control object' condition, birds were exposed to a familiar object~~is placed in the home enclosure for six days prior to testing~~. By placing a familiar object behind the food plate ~~prior to testing, we can observe responses during testing that are~~ in the control condition, we ensured that responses in the 'novel object' condition were elicited by the novelty of the object and not just the presence of the object itself ~~(see e. g. for justification).~~ (see e.g. Greggor et al., 2015, for justification). The familiar object ~~remains in place ,~~ remained in place throughout the testing and habituation period to avoid dishabituation from the familiar object. It ~~is~~ was replaced by the novel object only during the novel object testing sessions. To preserve the integrity of the experimental design, the novel object introduced in each of the four sessions ~~is~~ was unique, thus each bird's interaction with it ~~marks~~ marked their first encounter. The experimental timeline ~~spans~~ spanned from late June to mid-July, ~~lasting for~~ and lasted 8 consecutive days.

~~**Objects:** We will use five~~ We used five objects (Figure 2) of similar size (approximately the same size as a four weeks old gull), but of different colour, form and texture.

Prior to the Task: In preparation of the novel object task, and following a series of cognitive tests as part of another study (three tests in total), the ~~will be~~ test setup (Figure 3) was introduced into the birds' home enclosure when the birds ~~are~~ were not present. This setup ~~includes~~ included the pre- and post-testing pens, the start area, and one of our five pseudo-randomly selected ~~, which will later act~~ objects, which later acted as the control object in the neophobia assessments. After having introduced the test setup, birds ~~are~~ were allowed to accustom to the presence of the test apparatus for a period of six days. This habituation period ~~minimises~~ minimised any potential stress towards a new environment, which may influence the behavioural outcome of the test trials.

In order to distinguish the birds when they ~~are~~ were being tested in a group, each individual ~~is~~ was given a unique ~~marker~~ marking (marker pen, Raidex) a few days before the test, which ~~can~~ could be easily detected by a roof-mounted camera, as ~~the colour rings are~~ colour rings were not visible in the video recordings.

Testing Protocol: The testing ~~commences~~ commenced after the six-day habituation period. Order of conditions ~~is~~ was counterbalanced to incorporate control and novel object conditions, as well as individual versus group settings, with the entire sequence being repeated twice. The animals ~~are~~ were food deprived since their last feeding moment the evening before each test at 5:30 PM, to reduce motivational differences before testing. Testing ~~begins at 8~~began around 7:30 AM and ~~is expected to be~~ was completed around 11 AM. In both group and individual trials, individuals ~~will have~~ were given a maximum of 10 minutes for entering the test arena, and an additional 10 minutes to feed, which is consistent with previous novel object studies (Brown and Nemes, 2008; Bruijn and Romero, 2021; Lecuelle et al., 2011). All tests ~~will be~~ were recorded with roof-mounted cameras.

Prior to testing, all the birds ~~will be~~ were moved to the pre-testing holding pen. Next, a stacked plate of fish and an object (novel or control, depending on the condition) ~~will be~~ was placed at the back of the enclosure,

Figure 2. ~~Test setup in home enclosure~~Novel or control objects.

with the food plate placed in front of the object to rule out directional preference. A single bird, or group of birds, depending on the social context, ~~will be~~ was placed in the start area. The tester ~~will lift~~ lifted the door of the start area after 15 seconds and ~~leave~~left, giving the bird(s) access to their home enclosure (Figure 3). The first ~~10 minutes start~~ 10-minute period started when the door ~~begins~~ began to move, the second ~~10 minutes start~~ 10-minute period started for each bird individually when it ~~leaves~~ left the start area. The test session ~~ends~~ ended 10 minutes after the bird ~~has~~ had left the start area in individual trials, or after all birds had left the start area in group trials. Next, the tester ~~moves~~ moved the tested bird(s) to the post-testing holding pen and ~~starts~~ started a new test with a new (group of) bird(s).

Data processing and analysis

Video coding. We ~~will code~~ coded all videos using the free, open-source software BORIS (Behavioural Observation Research Interactive Software) (Friard and Gamba, 2016). ~~We will code four events~~ Four events were coded, namely 'start of trial', 'test arena entry', 'eating', and 'zone of interest' (see Table 1 for full descriptions). Based on the coded events, we ~~will determine~~ determined latencies and cumulative times. By extracting the time difference between 'start of trial' and 'test arena entry', we ~~will determine~~ determined the latency to leave the start area (Figure 3). In order to determine the latency to approach the food, we ~~will extract~~ extracted the time difference between 'test arena entry' and 'eating'. Time spent in the zone of interest (i.e. in proximity to the food reward and/or novel object, see ~~) is~~ Figure 3) was calculated as the cumulative time over the length of the trial. If an individual ~~does not perform a specific behaviour, we will assign~~ did not perform one of the target behaviours, we assigned the maximum latency, ~~meaning~~ representing the full task duration (in seconds), to that behaviour. For example, the behaviour ~~"~~'test arena entry~~' will have '~~ has a latency of 600 seconds if an individual ~~does~~ did not enter the test arena. This maximum latency applies only to latency measures; for time spent in the zone of interest (ZOI), a value of 0 was recorded if a bird did not enter the ZOI. For the group tests,

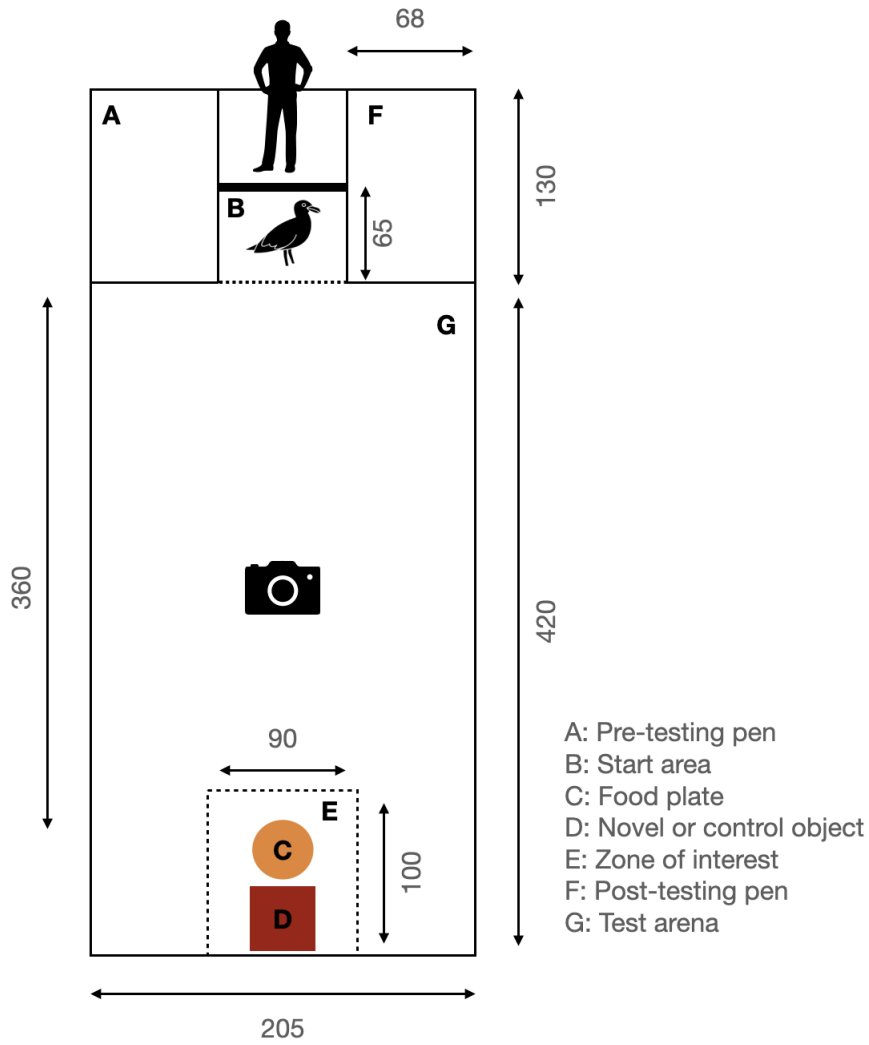

Figure 3. ~~Novel or control objects~~Test setup in home enclosure.

we ~~will follow~~ followed each bird individually to code their behaviours.

Video coding ~~will be a shared task between~~ was conducted collaboratively by multiple experimenters, with 20 percent of all videos being double-coded by a third experimenter to assess inter-rater-reliability (IRR) using Cohen's Kappa. ~~We aim for 0.81 ≤~~ Our analysis resulted in a Cohen's Kappa ~~≤ 1.0~~of 0.89, which indicates strong ~~to almost perfect~~ agreement between coders (McHugh, 2012). ~~If we will have a Cohen's Kappa below this value, we will assess each behaviour individually to determine which behaviours need to be recoded for all videos.~~

Statistical analysis ~~Statistical analyses will be~~

Statistical analyses were conducted using R, version ~~4.3.X~~ 4.4.1 (R Core Team, 2021). All package version numbers are documented and managed using the `renv` package (Ushey and Wickham, 2024). Mixed-Effects Models (~~MMs), either linear MMs (LMMs) or generalised LMMs (GLMMs), will be~~ LMMs) were fitted using the `lme4` package (Bates et al., 2015)~~. For LLMs, parameter estimation and~~ , and parameter estimation along with p-values ~~for the estimated models will be calculated by means of~~ were calculated using the `lmerTest` package (Kuznetsova et al., 2017)~~via the the~~ , via Satterthwaite's degrees of freedom method~~; for GLMMs, the~~ . Model assumptions, including normality and heteroscedasticity, were assessed using the ~~`caror` `carData` `package` `will be used. For the GLMM, we will use partial` η`-squared (`η_p^2`) as effect sizes, and they will be calculated by means of the` `r2glmm` `package. Models will be fitted to the different~~

Table 1. ~~**Ethogram of behaviours that will be coded in BORIS.**~~ **Ethogram of behaviours that were coded in BORIS.** The '~~Zone~~ zone of interest' ~~is~~ was defined as a fixed rectangle that ~~includes~~ included the object and the food bowl. To ensure comprehensive observation coverage, this area ~~is~~ was expanded by the approximate body length of a 4-week-old gull (30 cm). This ~~ensures~~ ensured that all relevant activities within and around the novel object ~~are~~ were captured.

Action	Definition
Start of trial (Point event)	Moment the door starts moving.
Test arena entry (Point event)	When the entire bird is outside the start area.
Eating (Point event)	When the beak touches the food.
Zone of interest (State event)	When the front half of the bird crosses the (notional) line.

~~latency measures separately, as well as combined. For the combined analysis, the approach proposed by Snijders and Bosker, 2012 will be used, which allows for the simultaneous analysis of multiple dependent variablesin the case of nested data structures, thereby considering within-group and between-group variance in latency measures.~~ `performance` package (Lüdecke et al., 2021), and transformations (log or Box-Cox) were applied where necessary.

~~As we aim~~ Initial diagnostic plots indicated non-normality of residuals and heteroscedasticity in the models. To address these violations, we followed a structured approach. Each dependent variable was first fitted using the raw data. When this did not meet assumptions, a log transformation was applied. This approach sufficiently improved model fit for the ZOI duration model, and the model was refitted accordingly. For the latency to enter and latency to eat models, however, the log transformation did not resolve assumption violations. In these cases, a Box-Cox transformation was implemented, with optimal lambda values determined using maximum likelihood estimation. The estimated lambda values were $\lambda = -0.869$ for latency to enter and $\lambda = -0.828$ for latency to eat. The models were refitted using the Box-Cox transformed dependent variables, leading to improved model assumptions.

The primary objective of the analysis was to determine whether ~~the average neophobic response differs~~ neophobic responses differed between individual and group trials~~, a (G)LMM~~. LMMs were fitted to different latency measures (latency to enter, latency to eat, and ZOI duration) under appropriate transformations (log or Box-Cox). Models were selected based on the best fit and diagnostics, with Type III sum of squares ~~will be performed on the latency measures (). This analysis will include both fixed and random effectsto explore the impact of different experimental conditions. The model will incorporate *Object*, *Context*, and their interaction as key fixed effects to explore how the type of object and the social setting (alone vs. in a group)interactively affect latency responses. Additionally,~~ used to ensure appropriate partitioning of variance for the fixed effects. Key fixed effects included *object* (control vs. novel objects) and *context* (individual vs. group). We added ~~*Trial*~~*trial* ~~will be included~~ as a fixed effect to ~~control for the impact of trial repeat. To specifically assess the variability in latency across individual~~ account for repeated testing. Initially, we fitted a full random effects structure (in line with the preregistered report) accounting for variability at the *NestID*, *GroupID*, and *BirdID* levels, with specific terms for individual (*indiv_dummy*) and group (*group_dummy*) conditions to capture within-subject and ~~group trials, we will compare the estimated variance components within our mixed-effects model. Variance for individual trials will be estimated from the *Indiv_Dummy* effect at the BirdID level. For group trials, the combined estimated variances of the *Group_Dummy* effect at both the BirdID and GroupID levels will be evaluated. This comparison aims to determine whether individual differences are more pronounced in solitary compared to group settings, with an expectation that individual variances and~~ within-group variation. The full random effects model was:

$$\begin{aligned}\text{Latency} \sim & \text{Object} \times \text{Context} + \text{Trial} \\ & + (1|\text{NestID}) \\ & + (-1 + \text{Group\_dummy}|\text{GroupID}) \\ & + (-1 + \text{Indiv\_dummy} + \text{Group\_dummy}|\text{BirdID})\end{aligned} \tag{1}$$

However, the ~~total variance might be higher in individual trials. Additionally, an analysis at the BirdID level between the estimated variances of the *Indiv_Dummy* and *Group_Dummy* effects will further elucidate how individual differences manifest under different trial conditions , potentially highlighting the influence of group dynamics on individual behaviour.~~ full random effects structure, outlined in the preregistered report, led to over-parameterisation. Consequently, non-significant interactions were dropped to simplify the model. In addition, the final models were simplified by including only random intercepts for *BirdID*, while retaining the dummy variables where the model allowed it. This approach effectively captured individual-level variability in both individual and group conditions, avoiding over-fitting. The final model structure for each latency measure is as follows:

$$\begin{aligned}\text{Latency} \sim & \quad , \text{Object} \times \text{Context} + \text{Trial} \\ & + (1|\text{NestID}) \\ & + (-1 + \text{Group\_Dummy}|\text{GroupID}) \\ & + (-1 + \text{Indiv\_Dummy} + \text{Group\_Dummy}|\text{BirdID})\end{aligned}$$

$$\begin{aligned}\text{Box-Cox(Latency to enter)} \sim & \text{Object} + \text{Context} + \text{Trial} \\ & + (-1 + \text{Indiv\_dummy} + \text{Group\_dummy}|\text{BirdID})\end{aligned} \tag{2}$$

~~In the model, *Object* refers to the stimulus presented, distinguishing between control and novel objects. *Trial* captures the two testing sessions conducted, and~~

$$\begin{aligned}\text{Box-Cox(Latency to eat)} \sim & \text{Object} \times \text{Context} + \text{Trial} \\ & + (1|\text{BirdID})\end{aligned} \tag{3}$$

$$\begin{aligned}\text{Log(ZOI duration)} \sim & \text{Object} \times \text{Context} + \text{Trial} \\ & + (-1 + \text{Indiv\_dummy} + \text{Group\_dummy}|\text{BirdID})\end{aligned} \tag{4}$$

For models fitted on Box-Cox transformed latency data, transformation parameters were estimated using the `MASS` package (Venables and Ripley, 2002). Marginal means for the fixed effects (context and object) were computed and back-transformed to the original scale (seconds) using the Box-Cox inverse transformation, or an inverse log transformation for eating latency, with the `emmeans` package (Lenth, 2024). Random effect variances for individual (*indiv_dummy*) and group (*group_dummy*) contexts were extracted from the model outputs using the `lme4` package (Bates et al., 2015). To aid interpretation, we back-transformed the random effects by simulating random effects for 1000 individuals under both conditions (individual and group), while accounting for the covariance between *indiv_dummy* and *group_dummy* using the `mvtnorm` package (Genz et al.,

2020). These simulated random effects were then combined with the predicted fixed effects for each condition (individual vs. group, control vs. novel object) and back-transformed to the original scale. The inverse Box-Cox transformation was applied to latency to enter and ZOI duration, while the inverse log transformation was applied to latency to eat.

We also fitted a multivariate model on the combined dataset for *latency to enter* and ~~*Contextindicates the social environment, differentiating between individual and group settings. Random effects structures are tailored to accurately reflect the individual and group-level variability in responses. Specifically, NestID is included to control for similarities within nests, Group_Dummy identifies trials conducted in group setting, effectively marking the presence of social interactions during the test. Conversely, Indiv_Dummy indicates the absence of such group dynamics, highlighting trials where subjects are tested alone.~~ *latency to eat*, using a contrast for behaviour type (*eat_vs_leave_contrast*) to account for potential correlations between the two outcomes. The multivariate model confirmed the findings from the univariate analyses, indicating consistent effects across both latency measures. However, for ease of interpretation, the results of the univariate models are presented in the main manuscript, as they allow a more straightforward interpretation of the individual effects of each predictor on the dependent variables. Although the multivariate model is not discussed in detail, a full walk-through of all intermediate models, including the preregistered version of the statistical analysis and the multivariate model results, is provided in the supplementary material.

~~In all instances, model plots will be generated using the `performance`~~ Post-hoc analyses of significant interactions were performed using estimated marginal means via the `emmeans` package ~~to inspect violations of model assumptions, such as heteroscedasticity, non-normality of residuals, and the presence of outliers. Multicollinearity and autocorrelation will be evaluated, with potential model adjustments including transformation of variablesor modification of the model structure (e. g., switching from LMM to GLMM). In terms of model design, binary predictors will be encoded using contrast coding~~ (Lenth, 2024), with appropriate back-transformations applied for models with transformed dependent variables. Random effect variances were compared between individual and group trials using likelihood ratio tests to assess whether separate variance components were warranted for each condition. Binary predictors were contrast-coded as (-0.5 vs .~~0~~0.5) ~~, optimizing the interpretability and efficiency of our analyses in the context of our perfectly balanced predictor variables. Post-hoc analyses, following significant findings, will be performed with Bonferroni-Holm corrected contrasts to further explore the data. Given the~~ to optimise interpretability. Multicollinearity concerns were minimal due to the balanced nature of ~~our model predictors, concerns related to multicollinearity are minimised, negating the need for the predictors, so~~ variance inflation factor (VIF) assessments ~~traditionally used to identify redundancy among predictors~~were not required. Finally, model assumptions were verified through diagnostic plots, and pairwise comparisons for significant findings were adjusted using Bonferroni-Holm corrections.