Linköping University | Department of Physics, Chemistry and Biology
Master thesis, 60 hp | Applied Ethology and Animal Biology
Spring term 2021 | LITH-IFM-A-EX—21/3961--SE

Exploring a simplified affective state test in the red junglefowl

Alessandra Munari

Examiner: Per Jensen, IFM Biology, Linköping University
Supervisors: Hanne Løvlie & Laura Garnham, IFM Biology, Linköping University

Avdelning, institution
Division, Department

Department of Physics, Chemistry and Biology
Linköping University

Datum
Date

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English

☐ _______________

Rapporttyp
Report category

☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport

☐ ____________

ISBN

ISRN: LITH-IFM-x-EX--yy/xxxx--SE

Serietitel och serienummer
Title of series, numbering

ISSN
____

URL för elektronisk version

Titel
Title
Exploring a simplified affective state test in the red junglefowl

Författare
Author
Alessandra Munari

Sammanfattning
Abstract

Affective states of animals are emotions with positive or negative valance. Positive and negative affective states affect animal welfare, and can bias interpretation of information positively or negatively, respectively. Judgement bias tests measure affective states based on responses to ambiguous cues, intermediate to cues with learnt positive and negative outcomes. Responses closer to those of positive cues indicate positive affective state. However, animals need extensive training to learn initial associations to reference cues. Therefore, I here aimed to validate an alternative affective state test based on instinctive avoidance of patterns resembling eyespots. Responses to ambiguous eyespot cues similar to responses to full eyespot cues could indicate negative affective state. To test this, behavioural responses of red junglefowl (Gallus gallus) chicks to ambiguous cues from a validated judgement bias test were compared to responses to cues resembling eyespots. In a second cohort of birds, I developed simplified tests with only one ambiguous cue in each tests. I predicted that responses in both tests would correlate positively. In the original tests, shorter distance and latency to approach ambiguous cues correlated positively with latency to approach one of the eyespot cues, a full eyespot cue. This pattern was only observed in females. In the simplified tests, at 4 weeks of age, shortest latency to approach cues correlated among tests. This pattern was not observed when chicks were tested at 2 weeks of age. Overall, the eyespot test is a promising alternative affective state test, but further studies exploring sex- and age-effects, are needed.

Nyckelord
Keyword
Affective state, animal welfare, chicken, eyespot, *Gallus gallus*, judgement bias, optimism, pessimism

Contents

1. Abstract

Affective states of animals are emotions with positive or negative valance. Positive and negative affective states affect animal welfare, and can bias interpretation of information positively or negatively, respectively. Judgement bias tests measure affective states based on responses to ambiguous cues, intermediate to cues with learnt positive and negative outcomes. Responses closer to those of positive cues indicate positive affective state. However, animals need extensive training to learn initial associations to reference cues. Therefore, I here aimed to validate an alternative affective state test based on instinctive avoidance of patterns resembling eyespots. Responses to ambiguous eyespot cues similar to responses to full eyespot cues could indicate negative affective state. To test this, behavioural responses of red junglefowl (Gallus gallus) chicks to ambiguous cues from a validated judgement bias test were compared to responses to cues resembling eyespots. In a second cohort of birds, I developed simplified tests with only one ambiguous cue in each tests. I predicted that responses in both tests would correlate positively. In the original tests, shorter distance and latency to approach ambiguous cues correlated positively with latency to approach one of the eyespot cues, a full eyespot cue. This pattern was only observed in females. In the simplified tests, at 4 weeks of age, shortest latency to approach cues correlated among tests. This pattern was not observed when chicks were tested at 2 weeks of age. Overall, the eyespot test is a promising alternative affective state test, but further studies exploring sex- and age-effects, are needed.

Keywords: affective state, animal welfare, chicken, eyespot, *Gallus gallus*, judgement bias, optimism, pessimism.

2. Introduction

Cognition is the mechanism by which animals acquire, process, store, and act on information from the environment (Shettleworth, 2010). Emotions are stimulus-directed affective states consisting of behavioural, physiological, and cognitive components (Crump et al., 2018). Emotions can modulate cognition, but can also be modulated by cognitive processes (Mendl et al., 2009). This means that an emotion, or affective state, can affect how individuals respond to an external cue (Mendl et al., 2009). Affective states have valence, where an affective state that is pleasurable is defined as positive, while an affective state that is unpleasurable is defined as negative (Mendl et al., 2009). Due to increased interest in reducing negative affective state and increasing positive affective state to improve animal welfare, estimating the affective state of animals has been a main research focus in recent decades. Nowadays, animal welfare is still one of the biggest challenges for scientists, mainly due to difficulty in interpreting indicators of animal affective states (Mendl et al., 2009). Therefore, affective state tests need to be developed, made more efficient and simpler to implement, to allow for easier and large-scale application.

Affective state can be measured behaviourally using a judgement bias test (reviewed by Lagisz et al., 2020). Humans show cognitive bias, as they interpret ambiguous cues in a more positive way if they are in a less negative state (e.g. less anxious), and in a more negative way if they are in a more negative state (e.g. more anxious, or depressed; Paul et al., 2005). Optimism is defined as expectation of a positive outcome, or overestimation of the chances that there will be a benefit when the presented cue is novel and ambiguous (Mendl et al., 2009). Pessimism is defined as expectation of a negative outcome or overestimation of the chances that there will be a negative consequence when the presented cue is novel and ambiguous (Mendl et al., 2009). Such cognitive bias is present in non-human animals (Harding et al., 2004; Paul et al., 2005). For animals, the judgement bias test consists of training animals to perform one response to a positive cue (i.e. a rewarded cue) or unrewarded cue, and a different response to an unrewarded cue or to a negative cue (e.g. one that is punished; Mendl et al., 2009). The value of learnt cues can vary in combination, as long as the value of the two reference cues are not the same (e.g. one rewarded and one unrewarded, one rewarded and one punished, or one unrewarded and one punished; Mendl et al., 2009). This training phase is based on discrimination learning. Discrimination learning is successful if the animal performs the correct response for a cue for a specific number of times (i.e. having reached a set learning criterion). After fulfilment of the learning criterion, behavioural responses towards neither rewarded nor punished ambiguous cues whose properties are intermediate of the cues with learned values, are measured (Mendl et al., 2009). These responses are used to measure optimism or pessimism based on the hypothesis that individuals in a relatively more positive affective state will react more optimistically, or less pessimistically, to an ambiguous cue (e.g. approach it faster; Mendl et al., 2009). On the other hand, individuals in a relatively more negative affective state are predicted to react more pessimistically, or less optimistically, to the same cue (e.g. approach it slower; Mendl et al., 2009). Optimism is measured by responses to the novel, intermediate cues that are more similar to responses to the cue associated with a positive experience than to the cue associated with a negative experience (Mendl et al., 2009; Figure 1A). A gradient of optimism is measured if none of the learnt cues are associated with negative experiences (Mendl et al., 2009; Figure 1B). Pessimism is measured by responses to the novel, intermediate cues that are more similar to responses to the cue associated with a negative experience than to the cue associated with a positive experience (Mendl et al., 2009; Figure 1A). A gradient of pessimism is measured if none of the learnt cues are associated with a positive experience (Mendl et al., 2009; Figure 1C). Studies altering affective states after making housing conditions unpredictable (Harding et al., 2004) or after providing enrichment (Burman et al., 2008, Matheson et al., 2008; reviewed in Lagisz et al., 2020; Mendl et al., 2009; Neville et al., 2020; Zidar et al., 2018a) showed that cognitive judgement bias test could detect such changes in affective state. This makes the judgement bias test currently the most accurate affective state test for animals (Paul et al., 2005; reviewed in Mendl et al., 2009). This test has quickly attracted research interest and has been applied in a variety of versions, making it a relatively well explored method (Lagisz et al., 2020; Mendl et al., 2009; Neville et al., 2020). On the other hand, the judgement bias test has some drawbacks: the test requires extensive training and assumes that animals are not anxious or neophobic (Mendl et al., 2009). These aspects can result in time-consuming habituation and training prior to testing or in the use of only a subset of the available population (e.g. only the set of animals that perform in the test setup). Moreover, animals can learn that the ambiguous cue is unrewarded, resulting in loss of motivation and slower responses to ambiguous cues, which can be easily misinterpreted as indicators of negative affective state. This phenomenon, known as ‘loss of ambiguity’, can cause biased results (Bethell, 2015; Košťál et al., 2020). Despite these drawbacks, the number of studies investigating simpler alternatives to the judgement bias test to detect affective states in animals, are still limited.

More recently developed methods for detecting affective states of animals are based on the interpretation of innate responses to naturally (i.e. ecologically relevant) aversive, or appetitive cues (Bateson, 2016; Mendl et al., 2009). Reactions to such cues are unlearned, meaning that training to establish an association between a cue and a specific outcome is not needed prior to testing (Bateson, 2016; Mendl et al., 2009).

Among these methods is a test based on initial responses to cues resembling “eyespots” (i.e. circular marking resembling the eye of a vertebrate; Brilot et al., 2009). Patterns resembling eyespots are often used as antipredator strategy (Ruxton et al., 2004) and can be naturally aversive for various bird species (European starling, *Sturnus vulgaris*, Brilot et al., 2009; domestic chicken, Olofsson et al., 2012; Olofsson et al., 2015; European robin, *Erithacus rubecula*, Chaffinch, *Fringilla coelebs*, Great tit, *Parus major*, House sparrow, *Passer domesticus*, Stevens et al., 2007; Stevens et al., 2008). However, the reason for the innate aversive reaction of studied birds to eyespot cues is still unclear. There are, to date, two main hypotheses that aim to explain the innate response of avoidance: the conspicuousness hypothesis and the predator’s eye mimicry hypothesis (Stevens and Ruxton 2014). Studies supporting the first hypothesis suggest that the appearance of an eyespot overstimulates the visual system of vertebrate predators (Stevens et al., 2007; Stevens et al., 2008). For example, a bird predating a butterfly might be overstimulated by eyespot patterns on the butterfly wings. Studies supporting the second hypothesis propose that the eyespot patterns suggest to the predator the presence of its own predator (Olofsson et al., 2012; Olofsson et al., 2015). For example, the eyespots on a butterfly's wing could look like owl or mammalian eyes for the bird that predates the butterfly. If the reason for eyespots’ aversiveness is the predator’s eye mimicry, exposure to eyespot patterns would be likely to be perceived as a negative experience. The negativity of exposure is less clear if the conspicuousness hypothesis explains the observed aversiveness. Regardless of the reason for their aversiveness, eyespot patterns have been used as naturally negative cues when measuring affective state in starlings (Brilot et al., 2009). First, to make sure the animal was motivated enough to be tested, each animal was habituated to the position of a rewarded control cue (i.e. a background lacking any eyespot pattern). Then, the reaction of test animals to rewarded full eyespot cues and ambiguous eyespot cues in the same position as the control cue, was recorded (Mendl et al., 2009). Researchers predicted to observe the strongest reaction (e.g. longest latency) to the full eyespot cues, no reaction to the control cue, and an intermediate reaction to intermediate, ambiguous eyespot cues (Brilot et al., 2009). Here, it is not possible to define optimism and pessimism based on how individuals respond to learnt rewarded or punished cues, because there is no associated learning of rewards or punishment, but merely an instinct based response to avoid (Bateson, 2016; Stevens and Ruxton, 2014). A new definition of pessimistic bias, specific for the eyespot test, can be established with the following two assumptions. First, the predator's eye mimicry hypothesis explains the aversiveness of eyespot patterns. Alternatively, eyespots' conspicuousness is perceived as a negative experience. Second, as all cues are rewarded, the observed reaction to full and ambiguous eyespot cues is not affected by the cue being rewarded. More specifically, the measured reaction to full and ambiguous eyespot cues automatically controls for effects from positive experiences, linked to the presence of the reward. This means that responses in this test are assumed to directly reflect a gradient of negative experiences when the bird is exposed to the full and the ambiguous eyespot cues. With these assumptions, the eyespot test can measure affective states only in their negative spectrum. Therefore, only a definition of pessimism is applicable for tests using eyespot patterns. The definition of pessimism is presented in detail below, and summarised and compared with the definition of pessimism and optimism from the judgement bias test in Appendix, Table A1. In the eyespot test, pessimism can be defined as unlearnt perceived predation risk when exposed to an ambiguous aversive cue containing eyespot patterns (Stevens and Ruxton, 2014). Pessimism is measured by latency to approach an ambiguous adverse cue, expected to always be longer than the latency to approach a control cue (Brilot et al., 2009; Figure 2A). European starlings showed longer latencies to approach cues with eyespot patterns, compared to latencies to approach cues without eyespot patterns (Brilot et al., 2009). However, no differences were found between responses to ambiguous eyespot cues and control cue (Brilot et al., 2009). From these results, it is unclear if starlings perceived the ambiguous eyespot cue as mildly aversive (i.e. intermediate between the control cue and the eyespot cue). In case starlings perceived the ambiguous eyespot cue as ambiguous (i.e. mildly aversive), the fact that they showed similar responses to the ambiguous cue as to the control cue would indicate a consistently positive affective state. A comparison of responses to eyespot cues with responses to cues from already established affective state tests (e.g. judgement bias test), could have explored if birds were perceiving ambiguous eyespot cues as mildly aversive and full eyespot cues as aversive. In that case, an eyespot test with the use of both full eyespot cues and ambiguous eyespot cues could have been validated to measure pessimism (Figure 2A). The logic behind the eyespot test and the potential way to validate the aversiveness of eyespot cues, are encouraging for using it as an alternative to classical judgement bias test, but needs to be tested.

Here, I therefore aim to perform a validation of the eyespot test by comparing results from it directly with results from a judgement bias test, to explore if the eyespot test can be used as a simpler method to measure affective state of animals. I did this by comparing the behavioural responses of red junglefowl chicks to ambiguous cues used in a judgement bias test to eyespot cues including two full and two ambiguous eyespot cues from tests run prior to my work. I analysed this data aiming to validate a complete eyespot test with the use of both full eyespot cues and ambiguous eyespot cues (Figure 2A). Further, in a new set of birds, I performed a simplified version of both tests, in which only one ambiguous cue for the judgement bias test and one full eyespot cue for the eyespot test were used. This comparison aimed to validate a simplified eyespot test with the use of just one full eyespot cue as ambiguous near aversive cue (i.e. cue displaying eyespot patterns functionally perceived as ambiguous; Figure 2B). In the red junglefowl, judgement bias tests have been routinely used to measure variation in optimism (e.g. Boddington et al., 2020; Garnham et al., 2019; Sorato et al., 2018; Zidar et al., 2017a; Zidar et al., 2017b; Zidar et al., 2018b; Zidar et al., 2019). This offers relatively standardised experimental procedure. Domestic chickens have been used for studying responses to eyespot patterns (e.g. Olofsson et al., 2012; Olofsson et al., 2015). Thus, it is established that such patterns are perceived as aversive in the fowl. Based on the previous work, I hypothesised that responses of fowl to eyespots would correlate with their responses to ambiguous, intermediate cues of the judgement bias test. In particular, I predicted that responses to full eyespot cues and ambiguous eyespot cues would correlate with responses to ambiguous cues from the judgement bias test, with the ambiguous eyespot cues showing weaker, but still correlations with ambiguous cues from the judgement bias test. If so, an eyespot test could be used to detect variation in pessimistic bias, and be used as a simpler alternative to a judgement bias test.

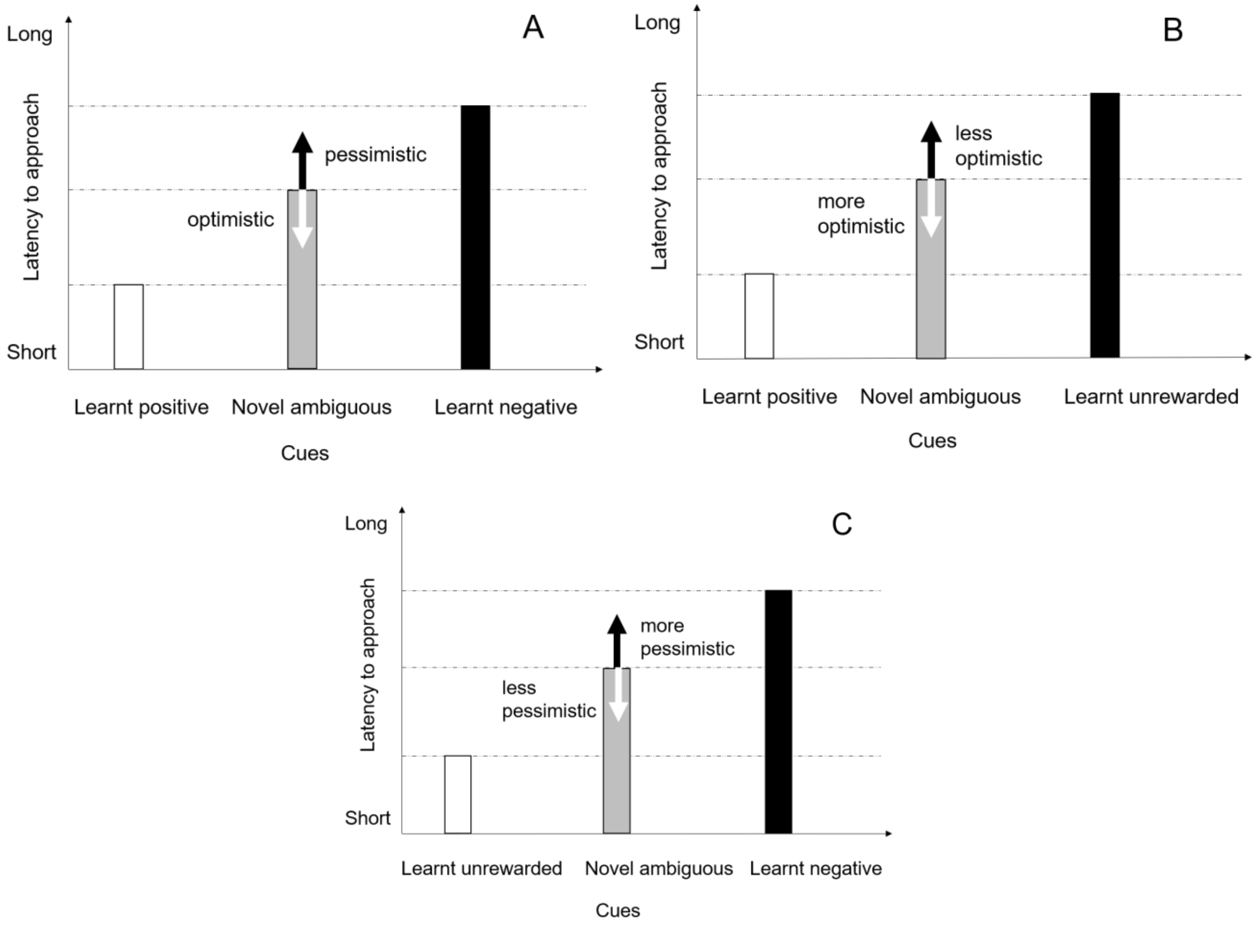

Figure 1: Predicted responses to cues from a judgement bias test. The value of learnt cues determines if optimism or pessimism is being measured. Predicted responses differ if learnt cues are A) positively reinforced (i.e. rewarded) vs. negatively reinforced (i.e. punished); B) positively reinforced vs. unrewarded; C) unrewarded vs. negatively reinforced (i.e. punished).
To measure optimism or pessimism the ambiguous cue, intermediate between learnt cues, is used (Mendl et al., 2009; Lagisz et al., 2020). Top horizontal dotted lines indicate maximum observed response. Lower horizontal dotted lines indicate minimum observed response. Middle horizontal dotted lines indicate mean of responses to learnt cues. White columns represent the response to the relatively more positive cue. Black columns represent the response to the relatively more negative cue. Grey columns represent the response to the intermediate, novel, ambiguous cue. Upwards black arrows indicate responses to the ambiguous cue biased towards responses to the more negative cue. Downwards white arrows indicate responses to the ambiguous cue biased towards responses to the more positive cue. Latency to approach cue is commonly used as an example of response measured.

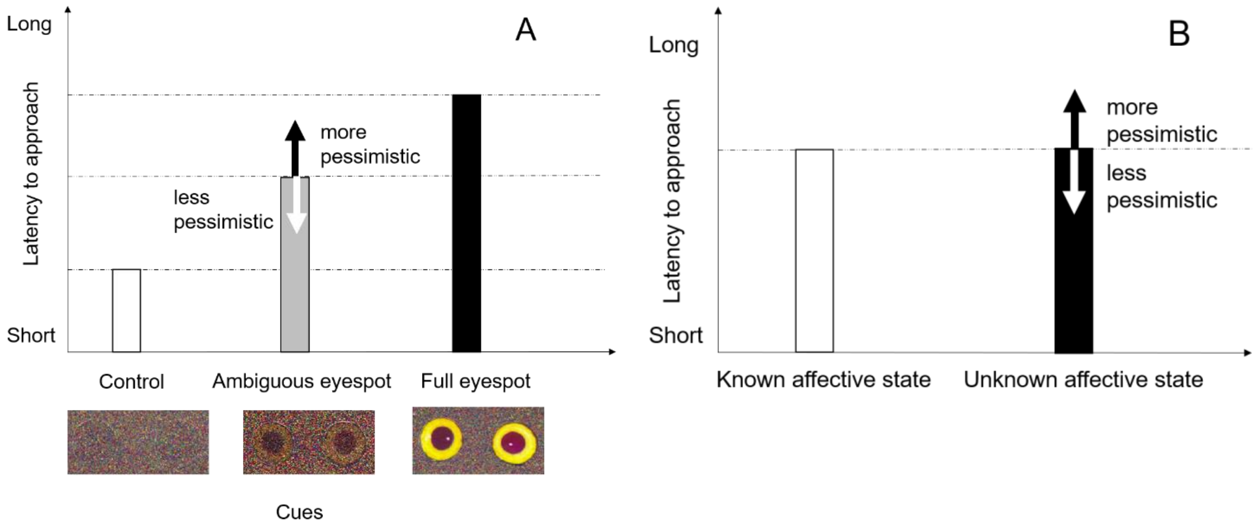

Figure 2: Predicted responses to cues from an eyespot test. To measure pessimism, instinctive aversion to eyespot patterns is used. Pessimism is measured A) by comparing reactions to ambiguous eyespot cues to reactions of the same individuals to reference cues (Brilot et al., 2009). White column represents the response to the control cue. Black column represents the response to the putatively aversive cue. Grey column represents the response to the intermediate, novel, ambiguous cue. Upwards black arrows indicate responses to the ambiguous cue biased towards responses to the aversive cue. Downwards white arrows indicate responses to the ambiguous cue biased towards responses to the control cue; B) by comparing reactions to an ambiguous, near aversive eyespot cue of individuals in unknown affective state to reactions to the same cue of individuals in a known affective state. White column represents the response to the ambiguous near aversive eyespot cue from animals with known affective state. Black column represents the response to the same cue from animals with unknown affective state.

Top horizontal dotted line indicates maximum observed response. Lower horizontal dotted line indicates minimum observed response. Middle horizontal dotted line indicates mean of responses to reference unlearnt cues. A single horizontal dotted line indicates mean of responses to ambiguous near aversive cues. Latency to approach cue is commonly used as an example of response measured.

3. Methods

3.1 Animals and housing

A total of 148 red junglefowl from a captive, pedigree-bred population kept at Linköping University, Sweden, in two cohorts, one hatched in 2017 (n = 49) and one in 2020 (n = 99), were used. Birds were wing tagged soon after hatching to facilitate individual recognition. Hatching and housing were controlled and standardised among individuals to ensure naivety to training and test cues (i.e. black, grey, white and eyespot cues; see below), before training and testing. Birds were housed in a laboratory in several groups, each containing a mix of test and control individuals (2017: three groups, three treatments; 2020: six groups, three treatments; see section 3.2). Each group (n = 10-25), was housed in a pen (72 × 71 × 53 cm, L × W × H). Birds were kept on a 12:12hr light:dark cycle, at around 24 ºC, and had *ad libitum* access to food, water, perches, sawdust and heating plates. When the birds were three weeks old, heating plates were turned off but kept in the pens as shelter and extra floor level. Two room heaters were active throughout the birds' stay in the laboratory. During experiments, the sex of birds was unknown, since chicks are monomorphic until the age of six-seven weeks. At age eight weeks in 2017 and five weeks in 2020, birds were moved to the chicken facility at Vreta Gymnasiet, Linköping. At the age of three months, birds were sexed and sorted according to their sex. The study followed ethical requirements in Sweden and was approved by Linköping ethical committee (288-2019).

3.2 Experimental setup

Prior to training and testing, to reduce stress, all birds were habituated to human handling, to test arenas and to be left alone (sensu Boddington et al., 2020). Birds were tested singly, and between 8 AM and 6 PM local time. The food reward (a third of a mealworm) was the same for all training and tests. The behavioural test battery differed in the two cohorts (Table 1). In 2017, prior to starting my project, birds were tested with two affective state tests: a classical judgement bias test and an eyespot test (hereafter referred to as the 'original' tests). In 2020, I tested birds with a simplified version of these two tests. The simplification of both tests consisted mainly of reducing the number of test cues to only one test cue per test. In 2017, as part of another study, before both affective state tests, birds were divided into three treatments (control: n = 34; treatment one: n = 65; treatment two: n = 34). The aim of treatments was to provide cognitive or environmental enrichment, to study effects on impulsivity tests. As only birds in treatment one were tested in both affective state tests, I used data only from birds from this treatment in my study. In 2020, as part of another study, after the first set of affective state tests (hereafter referred to as the 'simplified 2-week' tests), birds were divided into three treatments (control: n = 22, n_{males} = 12, $n_{females}$ = 10; treatment one: n = 31, n_{males} = 13, $n_{females}$ = 18; treatment two: n = 28, n_{males} = 15, $n_{females}$ = 13). Birds in treatment one were tested for reversal learning (i.e. training for learning the association opposite to that in discriminative learning). Birds in treatment two were given fluoxetine, a serotonin reuptake inhibitor. Following these treatments, to study effects of enrichment on affective state tests, control birds and those in treatment one were tested with a second set of both affective state tests (hereafter referred to as the 'simplified 4-week' tests). To study effects of drugs on affective state tests, birds in treatment two were tested with a second judgement bias test. As these aims were not the focus of the present study, the result of these tests will not be discussed. To avoid experimenter bias, testing after treatment followed a blinded procedure (i.e. experimenters did not know which treatment the bird being tested belonged to).

Table 1: testing history of red junglefowl chicks in the two cohorts used in the current tesis

2017

2-5 d	2 w	2 w	8 w
Habituation	Discriminative learning	Refresh sessions and original CJB	Habituation and testing original ES

2020

2-5 d	5-9 d	9-13 d	13-15 d	25-30 d	30-33 d	34-35 d
Habituation	Discriminative learning	Refresh sessions and 2-week CJB	Habituation and testing 2-week ES	Reversal learning for treatment group one	Refresh sessions and 4-week CJB	Habituation and testing 4-week ES

Given is age of animals in days (d) or weeks (w) after hatch. Discriminative learning = learning an association of cue-reward for judgement bias test. CJB = cognitive judgement bias test. ES = eyespot test. Refresh sessions = training sessions carried out just before testing. See main text for further details.

3.3 Previous experiments with original affective state tests in 2017

3.3.1 Original judgement bias test

In 2017, at the age of two weeks (Table 1), birds (n = 49) were trained to associate a white bowl with a reward and a black bowl with the absence of the reward. Each bowl (5 × 3 cm, Ø × H) was associated with a laminated card (9 cm^2) matching its colour. For the rewarded cue, a reward was provided, which was out of sight for the bird, until it had made its choice. A choice was measured as when a chick had its head within 2 cm from the bowl. The two cues were presented simultaneously to the bird. The arena used for learning the association was the same arena used later for the judgement bias test (47 × 38 × 18 cm, L × W × H; Figure 3). Previous research showed that six correct choices in a row is a sufficiently reliable criterion to avoid having false learners in associative learning (i.e. individuals reaching the learning criterion by chance; Sorato et al., 2018). Therefore, when a bird successfully approached the rewarded cue and ate the reward six trials in a row, the bird was assumed to have learnt the association between the positive cue and the reward. After this learning criterion was met, the bird was tested. Each bird was allowed 15 minutes per training session for a maximum of four sessions, separated by at least one hour of rest. Prior to testing, to make sure that birds remembered the association, birds underwent a maximum of four refresh sessions (i.e. training sessions carried out just before testing; Table 1). During testing, to measure positive affective state (i.e. optimism), each bird was exposed 12 times to each of the original training cues (i.e. black, and white; Figure 4), and three times each to three unrewarded, novel, intermediate ambiguous cues (‘light grey’ 75% white / 25% black; ‘middle grey’ 50 %white / 50% black; ‘dark grey’ 25% white / 75% black; Figure 4). There was a total of 33 trials per test, organised in a pseudo-random order (Sorato et al., 2018; Zidar et al., 2018a). If a bird showed lack of motivation by flying out of the arena on consecutive times, the session was put on hold for at least one hour and the bird was given the possibility to resume the test after that time. This procedure was carried out for a maximum of one time for each bird, resulting in a maximum of two test sessions per bids. During the test, only the positive cue was rewarded. A trial lasted maximum 30 s, it started when the bird was placed in the starting position in the test arena (Figure 3), and ended when it reached the cue and ate the reward, when present. Latency (s) to approach the cue was measured for each trial with a chronometer. A bird’s latency was reported as 30 s if it did not approach within the maximum allocated time. Latency values to approach the middle grey cue that were closer to latency values to approach the rewarded cue, than the unrewarded cue were considered as corresponding to a more optimistic response. In the current study, to keep the test simple, I did not use measures from the other ambiguous cues. A bird was considered to fail the test if it showed no motivation towards approaching the rewarded cue in both test sessions. Nine birds failed the test and were therefore excluded from further analyses.

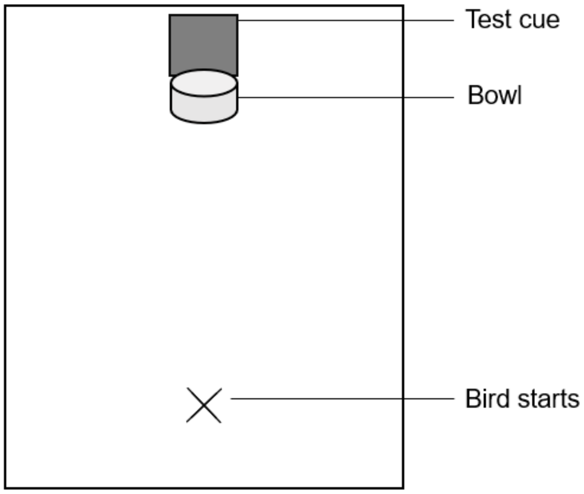

Figure 3: Overview of test arena used for judgement bias test measuring latency to approach learnt cues and ambiguous cues in red junglefowl chicks. The test arena was the same in original and simplified judgement bias tests (47 × 38 × 18 cm, L × W × H).

Figure 4: Cues used in the judgement bias test measuring latency to approach learnt cues (black and white) and test cues in red junglefowl chicks. From left to right: black, dark grey (25% white / 75% black), middle grey (50 %white / 50% black), light grey (75% white / 25% black), white. Each cue consists of a colour printed on one side of a paper sheet that was then matt laminated. All cues were used in the original judgement bias test and only the black, middle grey and white cue were used in both simplified judgement bias tests. The middle grey cue was the only ambiguous cue used in the simplified version, as it was the only one causing responses that correlated to responses to a cue from the eyespot test in the original tests.

3.3.2 Original eyespot test

In 2017, at the age of eight weeks (Table 1), birds (n = 49) were habituated to the test arena used for the eyespot test (76 × 114 cm; Figure 5A). This was done by placing a reward on top, and in the centre, of a control cue (i.e. a background pattern lacking any eyespot patterns; cue 0; Figure 6), which was placed flat on the floor 60 cm from the starting position. The floor of the arena was brown, corrugated cardboard. One 15-minutes habituation session was given. When a bird successfully approached the cue and ate the reward three times in a row, the bird was assumed to be sufficiently motivated and to know where the reward was provided, to be tested. During the test, birds were split into four groups. Each group consisted of 12-13 birds with an even sex ratio, balanced so that each group saw one of the four eyespot cues (1, 2, 3, 4, Figure 6) first, regardless of the cue being a full eyespot cue or an ambiguous eyespot cue. This subdivision was applied to account for any effects of previous experience of a cue on the response to the current cue. During the test, the arena allowed for two birds to be tested simultaneously, but visually separated from conspecifics (Figure 5A). After three trials of exposure to the control cue, each bird experienced all the four eyespot cues at least three times each (four for the first cue they experienced). Exposure to cues followed a semi-randomised order that depended on the test group of the bird. All trials were rewarded. A trial lasted maximum 60 s, it started when the bird was placed in the starting position in the test arena, and ended when it reached the cue and ate the reward. If the bird did not approach the cue or did not eat the reward within this time, latency was recoded as 60 s. To allow any needed recovery from any effect of previous cue, each exposure to an eyespot cue was followed by three trials of exposure to the control cue. As a result, a test session consisted of 52 trials. For each trial, two measurements were recorded. First, the closest distance of each chick to the cue (measured to the closest cm from the bird’s eye until point A; Figure 5A) was measured through a scale marked on the arena wall. Second, the chick’s latency (s) to approach the cue was recorded. If, during a control trial, the bird did not eat the reward, but reached to the cue, the trial was not considered completed, because of lack of motivation. In this case, to motivate the bird, the uncompleted trial was followed by another control trial with a reward presented halfway to the cue and another reward on the cue. The bird was allowed to resume the test when this simplified control trial was completed. One test session was given. One bird failed the test and was therefore excluded from the analyses.

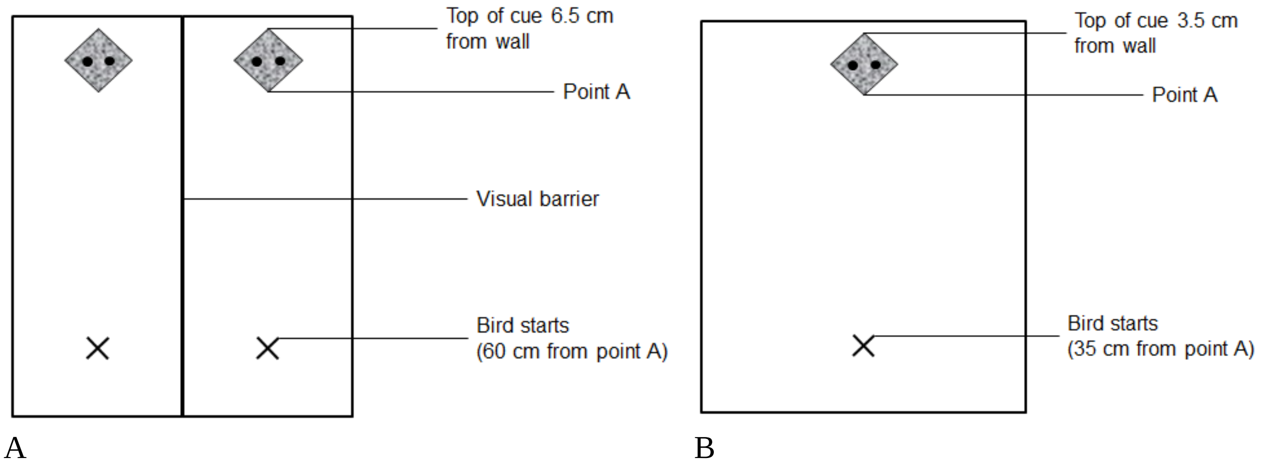

Figure 5: Overview of test arenas used for eyespot test measuring latency to approach and distance approached from eyespot cues, in red junglefowl chicks. The orientation of the cue is shown. ‘Point A’ indicates the tip of the cue which is closer to the bird. A) Arena used in the original eyespot test (76 × 114 cm, L × W). B) Arena used in both simplified eyespot tests (59 × 39 ×14 cm, L × W × H).

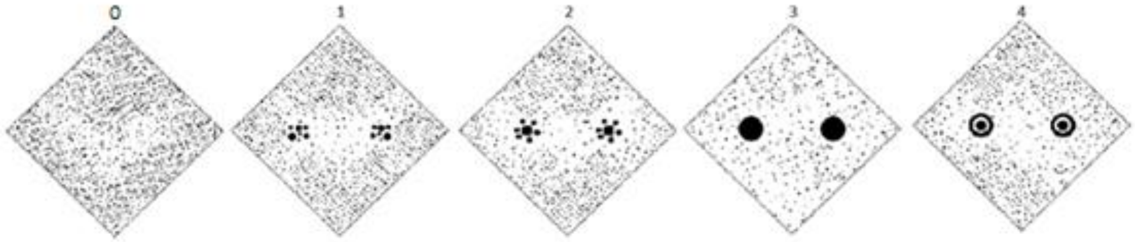

Figure 6: Cues used in an eyespot test measuring latency to approach and distance approached from eyespot cues, in red junglefowl chicks. 0 = control cue; 1 and 2 = ambiguous eyespot cues; 3 and 4 = full eyespot cues (i.e. circular marking resembling the eye of a vertebrate). Each cue consists of a black pattern printed on 7.5 × 7.5 cm transparent OHP paper. All cues were used in the original eyespot test and only cue 0 and 3 were used in both simplified eyespot tests. Cue 3 was the only eyespot cue used in the simplified version, as it was the only one causing responses that correlated to responses to a cue from the judgement bias test, in the original tests.

3.4 Current experiments with simplified affective state tests

3.4.1 Simplified judgement bias test

In 2020, birds ($n = 99$, $n_{males} = 43$, $n_{females} = 56$) took part in a simplified version of the original judgement bias test. At the age of five to nine days (Table 1), I trained half of the birds to associate a white laminated card with a reward and a black laminated card with the absence of the reward. To account for potential effects of the rewarded cue colour, I trained the other half to associate the black laminated card with the reward and the white laminated card with the absence of the reward. For training, I used the same modalities as those described in section 3.3.1. At the age of nine to 13 days, to make sure that all chicks remembered the association, birds that learnt the association ($n = 89$, $n_{males} = 42$, $n_{females} = 47$) underwent a maximum of four refresh sessions, and were then tested with a judgement bias test (i.e. simplified 2-week judgement bias test, Table 1). To measure positive affective state (i.e. optimism), each bird was exposed seven times to the positive cue, six times to the unrewarded middle grey ambiguous cue (50% white / 50% black), and four times to the negative cue. There was a total of 17 trials per session, organised in a pseudo-random order. Number of test sessions, trial time, test arena dimensions, measured responses and rules for test failure were the same as those described in section 3.3.1. Eight birds did not complete the simplified 2-week test. I therefore excluded these birds from the analyses and from further training or testing for the 4-week simplified judgement bias test. As part of another study, between the simplified 2-week tests and the simplified 4-week tests, birds in treatment one ($n = 31$) were trained to reverse the association they learned before the 2-week simplified judgement bias test (i.e. reversal learning; Table 1). At the age of 30-33 days (Table 1), to make sure that birds remembered the association after this treatment, birds from the control group, and treatment one that successfully reversed the learnt association ($n = 51$, $n_{males} = 24$, $n_{females} = 27$), underwent a maximum of three refresh sessions and were then tested a second time (i.e. simplified 4-week judgement bias test; Table 1). I adapted the simplified 4-week judgement bias test for treatment one by inverting rewarded and unrewarded colour so that each bird was rewarded on the cue it had learnt previously that was rewarded. Four birds did not complete the simplified 4-week test, so I excluded them from the analyses.

3.4.2 Simplified eyespot test

In 2020, at the age of 13 to 15 days for the first test (i.e. simplified 2-week eyespot test; Table 1) and at the age of 34 to 35 days for the second test (i.e. simplified 4-week eyespot test; Table 1), I habituated birds (2-weeks: n = 89, n_{males} = 42, $n_{females}$ = 47; 4-weeks; n = 53, n_{males} = 24, $n_{females}$ = 29) to the test arena (59 × 39 ×14 cm, L × W × H; Figure 5B). I followed the same procedures described in section 3.3.2, except for two aspects. First, the distance of the reward from the starting position was 35 cm. The smaller dimensions of the arena and the shorter distance of the reward from the starting position were adaptations to the younger age of birds in the simplified eyespot tests, than in the original eyespot test. Second, I gave birds two training sessions to maximize the number of birds that learnt the position of the reward. During the test, after three trials of exposure to the control cue, birds were exposed three times to cue 3 (Figure 6), as cue 3 was the only eyespot cue to which responses showed significant correlation with responses to the middle grey cue when comparing the two original tests (Table 3). For testing, I followed the same procedure as in section 3.3.2, except for the following aspects. First, the maximum trial time was 30 s. Second, as each exposure to cue 3 was followed by three trials of exposure to the control cue, there was a total of 12 trials per session. Third, I gave birds two test sessions. One bird did not pass the simplified 4-week test, so I excluded the bird from analyses.

4. Data analyses

4.1 General approach

Despite not personally collecting data in 2017 from the original tests, I analysed data and here report results from both cohorts. I used R version 3.6.1 for statistical analyses. Data distributions of variables obtained were not normal (Shapiro tests, p-value < 0.05), thus, as the assumptions for parametric statistics were not met, non-parametric statistical tests were used throughout.

I controlled for potential differences in locomotor activity between birds. This was done in both the judgement bias test and the eyespot test by using a subtraction of baseline activity approach. In the judgement bias test, the individual shortest latency value to the rewarded cue out of all rewarded trials (i.e. 12 trials in the original test; seven trials in the simplified tests) was subtracted from each individual mean latency to each other cue, separately. The measure that resulted from this subtraction was defined as ‘relative mean latency’ to each cue from the judgement bias test. In the eyespot test, this was done for both distance and latency measures. For distance measures, the lowest measure of closest distance from cue 0 was subtracted from the mean measure of closest distance from the eyespot cues (i.e. cue 1-4, separately in the original test; cue 3 in the simplified tests). For latency measures, each individual's lowest measure of shortest latency to cue 0 was subtracted from the individual's mean measure of shortest latency to the eyespot cues (i.e. cue 1-4, separately in the original test; cue 3 in the simplified tests). The measures that resulted from these subtractions were defined respectively as 'relative mean distance' and 'relative mean latency' to approach each eyespot cue in the eyespot test.

4.2 Comparisons between affective state tests

In the original tests, to explore how relative mean latency to approach cues in the judgement bias test related to relative mean latency to approach cues and the relative mean distance the cues were approached to in the eyespot test, I ran Spearman rank correlations using the cor.test function from the basic R package 'stats'. Correlations were run using data from birds that completed both affective state tests (n = 39, males = 21, females = 18). To test for potential sex effects, correlations were initially run with males and females separately. Data were not pooled if Spearman's Rho (R_S) was > +0.3 or < –0.3 in just one sex, since that would indicate a correlation only in one sex (Cohen, 1988; Table 2). As all birds tested in the original tests belonged to the same treatment, I did not test for any potential treatment effects. As none of the correlations between the middle grey cue from the judgement bias test and the eyespot cues 1, 2, and 4 were strong (Table 3), these correlations were not analysed further.

In the simplified 4-week eyespot test, all birds scored a value of 0 in relative mean distance from cue 3 (i.e. all birds approached this cue completely in all trials). Therefore, this data was not used for further analyses due to lack of variation. To explore how relative mean latency to approach cues in the simplified judgement bias test related to relative mean latency to approach cues and relative mean distance approached in the simplified eyespot test, I ran Spearman rank correlations. Correlations were run using data from birds that successfully completed both affective state tests (simplified 2-week tests: n = 81, males = 39, females = 42; simplified 4-week tests: n = 46, males = 23, females = 23). To test for any sex effects, correlations were initially run with males and females separately. To test for any treatment effects when exploring the relationships for the 4-week tests, birds from the control treatment and from treatment one were analysed separately and, for each treatment, the two sexes were analysed separately by Spearman rank correlations. I used the same rule as above regarding pooling data from sexes and, when applicable, from different treatments (Table 2).

4.3 Comparisons accounting for loss of ambiguity

In the simplified judgement bias test, some birds showed maximum latency to approach the middle grey cue after several exposures to this cue. The number of total exposures to the middle grey cue in a complete test session was equal to the number of trials in a row required for reaching the learning criterion in the training prior to the test (i.e. six). Therefore, it was not possible to establish if birds showing repeated maximum trial time during testing, did so due to a bias induced by a negative affective state, or to learning that the ambiguous cue was unrewarded (a phenomenon called ‘loss of ambiguity’; Bethell, 2015; Košťál et al., 2020). To avoid considering values from trials in which birds could have potentially learnt that the middle grey cue was unrewarded, a further set of Spearman rank correlations was run. In this set of correlations, data were analysed using only values from the trial of each bird with the shortest latency to the middle grey cue. I created the measures used in these correlations by subtracting the individual’s shortest latency value to the rewarded cue out of all rewarded trials from the shortest latency to the middle grey cue. I called this measure ‘relative shortest latency’ to the middle grey cue. This adjustment (i.e. using the relative shortest latency instead of the relative mean latency) was not carried out for data from the simplified eyespot test for the following reasons. First, I assumed that birds could not learn to ignore the aversiveness of cue 3 by associating cue 3 with the reward, since the total number of exposures to cue 3 (i.e. six considering both the simplified 2-week and the simplified 4-week tests) was never higher than the learning criterion in training prior to judgement bias test (i.e. six). The learning criterion from the judgement bias test was here used as a reference for the number of exposures to cue 3, instead of using the habituation criterion from the eyespot test. The reason for this is that while habituation to the eyespot test arena just aims to teach the association between the location and the reward to keep motivation, the judgement bias training aims to teach the association between the cue itself and the reward. Second, the smallest values for latency to approach cue 3 were very similar among all birds and the shortest distance from cue 3 was always 0, resulting respectively in almost absent, or totally absence of variation. To do a between-cohort comparison of detected relationships between responses to tests, the same adjusted analysis was run for the original version of the tests. In the original tests, the adjustment applied when using shortest values resulted in some slight negative values for latencies to the middle grey cue (range = - 0.10 s to 27.82 s). In the simplified version, the adjustment resulted in some negative values in two cases. The first case was the subtraction of latencies to the middle grey cue in the 2-week simplified judgement bias test (range = – 0.58 s to 2.07 s). The second case was the subtraction of latencies from the middle grey cue in the simplified 4-week judgement bias test (range = –0.66 s to 2.97 s). Bonferroni correction was not applied to any comparison.

4.4 Outliers

To detect potential error outliers (‘PEO’; i.e. candidates for being error extreme values; Aguinis et al., 2013), both variables in each relationship of a correlation were visually inspected by the use of a boxplot, which did not assume a normal distribution (Aguinis et al., 2013; Dovoedo and Chakraborti, 2014). Values falling outside the whiskers (i.e. values < 1st quartile - 1.5 x Interquartile range or values > 3rd quartile + 1.5 x Interquartile range) were considered PEO. Each detected PEO could potentially be an error outlier (‘EO’; i.e. data points that lie at distance from other data points because they are the result of inaccuracies in a test, e.g. an error in observation or recording; Aguinis et al., 2013). In that case, to see how such values could affect results, a parallel set of Spearman correlations was run with excluding EO from the involved test and from the test whose results were used in the comparison with the involved test (Table 2; Appendix; Table A2). To analyse to what extent results depended on extreme values not due to experimental errors (i.e. PEO), a second parallel set of Spearman correlations was run with excluding all detected PEO (Table 2; Appendix, Table A2). Results are presented in section 5.1, 5.2 and 5.3 without excluding outliers. A description of how results were affected by these two parallel analyses is provided in section 5.4.

Table 2: Summary of data pooling and outliers removal in comparisons between responses of red junglefowl chicks to the judgement bias test and the eyespot test.

Comparison	Data pooling	Removal of outliers
Original		
Mean MG – Cue 1, distance	M/F	NA
Mean MG – Cue 1, latency	Pooled	NA

Mean MG – Cue 2, distance	Pooled	NA
Mean MG – Cue 2, latency	Pooled	NA
Mean MG – Cue 3, distance	M/F	PEO (Males)
Shortest MG – Cue 3, distance	M/F	PEO (Figure A1)
Mean MG – Cue 3, latency	M/F	PEO (Males)
Shortest MG – Cue 3, latency	M/F	PEO (Figure A2)
Mean MG – Cue 4, distance	Pooled	NA
Mean MG – Cue 4, latency	Pooled	NA
Simplified 2-week		
Mean MG – Cue 3, distance	Pooled	PEO
Shortest MG – Cue 3, distance	Pooled	PEO
Mean MG – Cue 3, latency	Pooled	PEO (Figure A3B); EO (Figure A3A)
Shortest MG – Cue 3, latency	Pooled	PEO (Figure A4B); EO (Figure A4A)
Simplified 4-week		
Mean MG – Cue 3, distance	Pooled	NA
Shortest MG – Cue 3, distance	Pooled	NA
Mean MG – Cue 3, latency	C/T	PEO (Figure A6)
Shortest MG – Cue 3, latency	Pooled	PEO (Figure A5)

‘Comparison’ indicates the pair of measurements that are compared. For each pair of measurements, ‘Mean’ or ‘Shortest’ refer to the middle grey cue from the judgement bias test (‘MG’), and ‘distance’ or ‘latency’ refer to the mean value of each measure for cue 3 from the eyespot test. In bold is the version of judgement bias test and eyespot test to which the pairs of measures refer. ‘Data pooling’ indicates if data from the two sexes (and/or the two treatments, in the simplified version) were pooled or not, where ‘M/F’ means that sexes were not pooled, and ‘C/T’ means that treatments were not pooled. ‘Removal of outliers’ indicates if any potential error outliers (‘PEO’) and error outliers (‘EO’) were detected (Appendix; Table A2), and the corresponding figure in Appendix, when applicable. ‘NA’ indicates that the outlier analysis was not operated on that comparison.

5. Results

5.1 Eyespot cue and approach

In the original eyespot test, 13 out of 39 birds did not approach eyespot cue 3, while only four birds at two weeks did not when using simplified eyespot tests.

In the original eyespot test, five and four birds out of 48 did not approach cue 3 in the first and last exposure to cue 3, respectively. This means that the number of birds not approaching the eyespot cue did not seem to decrease from the first to the last trial in which the presented cue was eyespot cue 3. In the simplified test at two weeks, zero and five birds out of 89 did not approach cue 3 respectively in the first and last exposure to cue 3. This means that the number of birds not approaching from the first to the last trial of exposure to cue 3 again did not differ much.

5.2 Comparisons between responses from original tests

In the original tests, relative mean measure of shortest latency to approach the middle grey cue in the judgement bias test tended to positively correlate with relative mean measure of closest distance from, and latency to, cue 3 in the eyespot test, in females (Table 3; Figure 7A-B). Females' relative shortest measure of shortest latency to approach the middle grey cue in the judgement bias test was positively correlated with relative mean measure of closest distance from cue 3 in the eyespot test and tended to positively correlate with relative mean measure of shortest latency to approach cue 3 in the eyespot test (Table 3). No comparison was significant in males in the original tests (Table 3).

5.3 Comparisons between responses from simplified tests

In the simplified tests, there was no significant relationships between responses to the middle grey cue in the judgement bias test and cue 3 in the eyespot test when birds were two-week-old (Table 3). Relative mean measure of shortest latency to approach the middle grey cue in the simplified judgement bias test was positively correlated with relative mean measure of shortest latency to approach cue 3 in the simplified eyespot test when birds were four-week-old (Table 3; Figure 7C). There was no significant relationship between responses to the middle grey cue in the simplified judgement bias test and to cue 3 in the eyespot test after taking loss of ambiguity into account (Table 3).

5.4 Comparisons between responses from affective state tests with removal of outliers

One error outlier was detected in both simplified 2-week tests. However, inclusion or removal of this did not affect any relationships related to those tests.

In the original tests, removing two potential error outliers that were not error outliers, affected two comparisons. First, the comparison between the relative shortest measure of shortest latency to approach the middle grey cue and the relative mean measure of closest distance from cue 3 shifted to non-significant (from showing positive tendency). Second, the comparison between the relative shortest measure of shortest latency to approach the middle grey cue and the relative mean measure of shortest latency to cue 3 shifted to non-significant (from showing positive tendency).

In the simplified 2-week tests, removing 13 potential error outliers that were not error outliers, resulted in that the comparison between relative shortest measure of shortest latency to middle grey cue and relative mean measure of shortest latency to cue 3 was stronger and shifted from non-significant to showing a positive tendency. Removing one potential error outlier in the simplified 4-week tests, resulted in that the comparison between relative mean measure of shortest latency to middle grey cue and relative mean measure of shortest latency to cue 3 reduced from significantly positive to showing positive tendency.

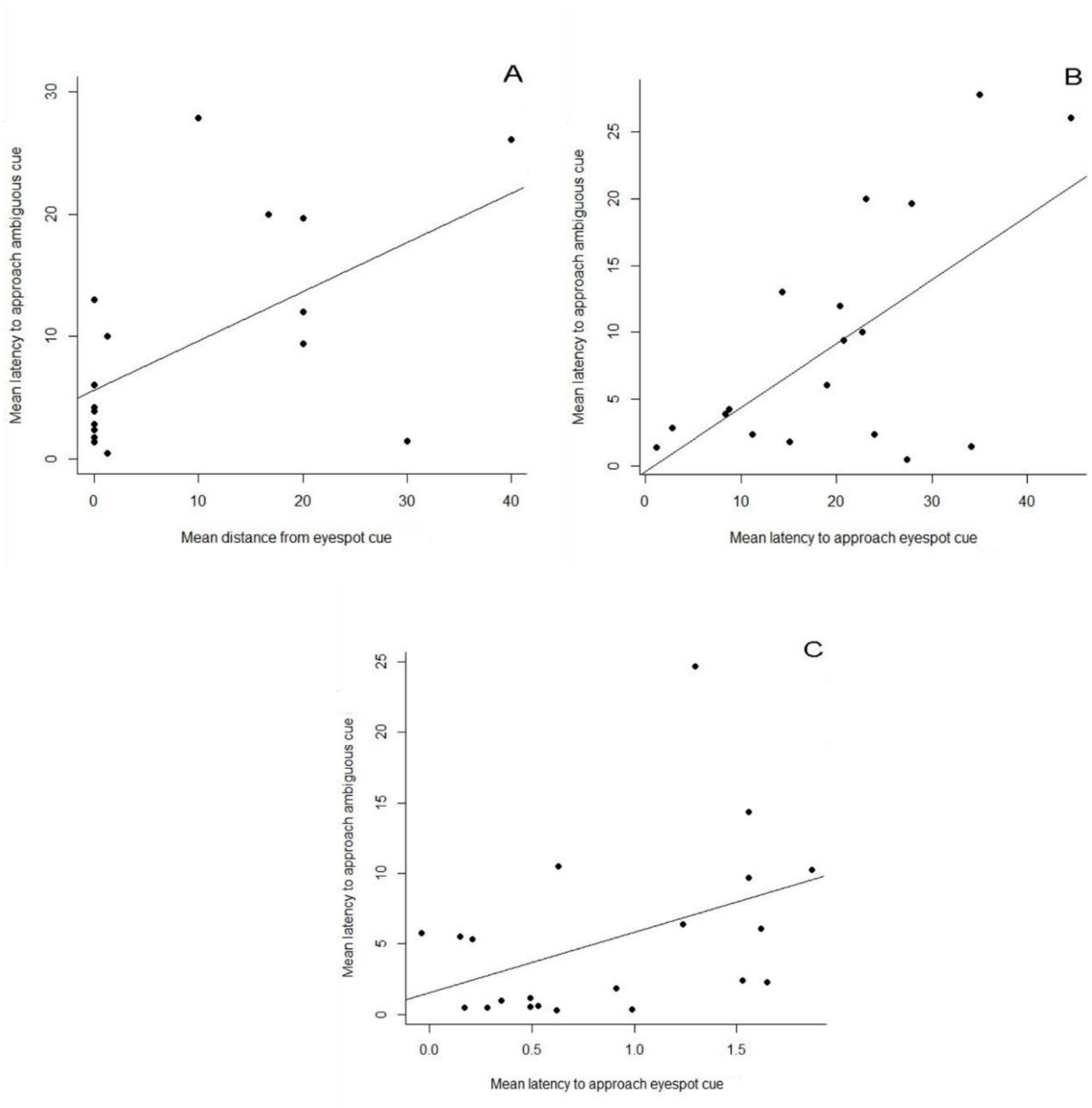

Figure 7: Comparison of behavioural responses of red junglefowl to a judgement bias test and eyespot test. A) The original version with distance from the eyespot cue 3 in females only; B) the original version with latency to approach the eyespot cue 3 in females only; C) The simplified version of tests at four weeks of both sexes. 'Mean distance from eyespot cue' is the relative mean measure of closest distance (cm) reached by the chick from the eyespot cue 3 (i.e. a full eyespot cue) in the eyespot test. 'Mean latency to approach ambiguous cue' is the relative mean measure of shortest latency (s) to approach the middle grey cue (i.e. novel, ambiguous cue) in the judgement bias test. 'Mean latency to approach eyespot cue' is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 in the eyespot test. Each point represents an individual bird. The trend line is marked in black.

Table 3: Comparisons between behavioural responses of red junglefowl chicks to the judgement bias test and the eyespot test.

	MG: Mean latency	MG: Shortest latency
Original tests		
Cue 1: Mean distance	♂♂: $R_S = -0.22$, $n = 21$, $p = 0.34$	NA
	♀♀: $R_S = 0.33$, $n= 18$, $p = 0.19$	
Cue 1: Mean latency	$R_S = 0.12$, $n = 39$, $p = 0.47$	NA
Cue 2: Mean distance	$R_S = 0.20$, $n = 39$, $p = 0.22$	NA
Cue 2: Mean latency	$R_S < 0.01$, $n = 39$, $p = 0.99$	NA
Cue 3: Mean distance	♂♂: $R_S = 0.22$, $n = 21$, $p = 0.33$	♂♂: $R_S = 0.16$, $n = 21$, $p = 0.49$
	♀♀: $R_S = 0.45$, $n = 18$, $p = 0.06$*	♀♀: $R_S = 0.47$, $n = 18$, $p = 0.05$**
Cue 3: Mean latency	♂♂: $R_S = 0.16$, $n = 21$, $p = 0.49$	♂♂: $R_S = 0.20$, $n = 21$, $p = 0.38$
	♀♀: $R_S = 0.39$, $n = 18$, $p = 0.11$	♀♀: $R_S = 0.41$, $n = 18$, $p = 0.09$*
Cue 4: Mean distance	$R_S = 0.11$, $n = 39$, $p = 0.50$	NA
Cue 4: Mean latency	$R_S = 0.11$, $n = 39$, $p = 0.51$	NA
Simplified tests, 2-week		
Cue 3: Mean distance	$R_S = -0.12$, $n = 81$, $p = 0.27$	$R_S = -0.01$, $n = 81$, $p = 0.94$
Cue 3: Mean latency	$R_S = 0.14$, $n = 81$, $p = 0.22$	$R_S = 0.12$, $n = 81$, $p = 0.27$
Simplified tests, 4-week		
Cue 3: Mean distance	NA	NA
Cue 3: Mean latency	$R_S = 0.46$, $n = 21$, $p = 0.04$**	$R_S = 0.02$, $n = 46$, $p = 0.87$

**significance ($R_S > 0.40$; $p < 0.05$); *tendency ($R_S > 0.40$; $p < 0.10$)

Cue 1, cue 2, cue 3 and cue 4 are eyespot cues from the eyespot test. MG is the ambiguous, middle grey cue from the judgement bias tests. 'Cue 3: Mean distance' is the relative mean measure of closest distance (cm) of a bird from cue 3, the full eyespot cue whose responses correlated with responses to one ambiguous cue from the judgement bias test. 'Cue 3: Mean latency' is the relative mean measure of shortest latency (s) of a bird to approach cue 3. For the original tests, these measurements are available also for cue 1, 2 and 4. 'MG: Mean latency' is the relative mean measure of shortest latency (s) of a bird to approach the middle grey cue, the ambiguous cue whose responses correlated with responses to one eyespot cue from the eyespot test. 'MG: Shortest latency' is the relative shortest measure of shortest latency (s) of a bird to approach the middle grey cue, when accounted for possible loss of ambiguity. '♂♂' refers to the male subset. '♀♀' refers to the female subset. 'NA' indicates that the analysis was not run for that comparison.

6. Discussion

I here explored relationships between behavioural responses obtained in two tests aiming to measure affective state in animals, namely a judgement bias test (Mendl et al., 2009) and an eyespot test (Mendl et al., 2009). I did so in the original version of the tests and in the here developed simplified versions. I used two cohorts of red junglefowl chicks for this work. Responses to ambiguous cues from the judgement bias test did not correlate with responses to more ambiguous eyespot cues. Females' responses correlated between tests from the original version when using one of the two full eyespot cues. Males' responses did not correlate in the original version. Both sexes' responses correlated in the simplified tests at four weeks of age when using the same full eyespot cue. However, the relationship disappeared if considering the possibility that birds learnt that the ambiguous cue was unrewarded in the judgement bias test. This phenomenon, known as loss of ambiguity, is observable as a decline of responses to unreinforced cues, regardless of the affective state (Bethell, 2015; Košťál et al., 2020). These results are discussed below, in turn.

6.1 Differences between original and simplified tests

The original affective state tests, and the simplified affective state tests carried out at two and four weeks could detect the relationship between the judgment bias test and the eyespot test in different experimental conditions. These experimental conditions, referred to as 'conditions' from now on, include the use of latency to or distance from the eyespot cue, the use of one or both sexes, including or excluding potential loss of ambiguity, and including or excluding extreme values (Appendix; Table A4). The increased number of individuals used in the simplified tests did not result in a clearer relationship between tests in any condition, when compared to the original tests. Many factors could have affected the ability of each version of the tests to detect a relationship. These factors include differences in experimental setup, in age and in naivety of birds to eyespot cues (Appendix; Table A4). Below, I will discuss how these factors could have influenced each of the mentioned conditions in the different tests.

6.1.1 Ability to detect a relationship between affective state tests using distance to approach

The relationship between affective state tests with the use of closest distance to a cue in the eyespot test was detected by neither of the simplified tests. The original and simplified versions of the eyespot test differed in two main methodological aspects: number of trials (original: 52 trials; simplified: 12 trials) and runway length (i.e. distance from starting point; original: 60 cm; simplified: 35 cm). The number of birds approaching the eyespot cue in the last trial did not differ much from the number of birds approaching the same cue in the first trial in both versions of the eyespot test. The number of cues is therefore likely not relevant, as birds' likelihood to approach the eyespot cue did not seem to decrease when more cues were presented. Concerning runway length, longer runways have been used to account for size differences due to age (Favati et al., 2016; Hedlund et al., 2021; Olofsson et al., 2015). For this reason, as birds were 8 weeks old when tested with the original eyespot test, I used shorter runway lengths for the simplified eyespot tests. This adaptation could have made the runway too short in the simplified version, causing distance measures to be more clustered towards lower values and more measures to stand out as potential error outliers, if compared to distance measures from the original version. This could have prevented us from considering possible effects of age and naivety to the eyespot cue on distance approached. Future studies aiming to clarify the role of age and naivety to cues when validating responses to the eyespot pattern by distance measures should repeat affective state tests at different ages with small experimental adaptations to account for size.

6.1.2 Ability to detect a relationship between affective state tests with latency measures

As a relationship between affective state tests was detectable when measuring latency to approach eyespot cues by both the original tests and the simplified test at four weeks, which differed in both experimental setup and birds' naivety to the eyespot cue, these two factors are likely not strongly affecting it. That the only test in which the relationship between tests was not detectable without removal of extreme values is the simplified test at two weeks, suggests that age might be a relevant factor for the presence of a relationship between tests. More specifically, age could play a role in how birds react to eyespot cues. In my data, 10 out of 81 birds had much higher latency to approach the full eyespot cue in the test at two weeks while just two out of 46 in the test at four weeks and none out of 18 in the original test. This seemingly lower number in individuals supposedly showing stronger aversion to eyespot patterns might suggest that a bird's perception of eyespot patterns or reaction to aversive situations changes with age (e.g. between age two weeks and age four weeks). However, this speculation needs further investigation. The shift would in so case be from particularly strong reactions to weaker reactions possibly linked to pessimism, if observed in an eyespot test. This could explain why some birds were particularly slow when approaching the full eyespot cue in the test at two weeks (10 birds out of 81) while other birds already showed weaker reactions (i.e. by not showing extreme values). This could also explain why all individuals that were particularly slow when approaching the full eyespot cue in the test at two weeks, scored non-extreme values in latency measurements for full the eyespot cue in the test at four weeks. In accordance with this explanation, four-week-old chickens reacted to eyespot patterns stronger than older birds (Olofsson et al., 2015). Studies investigating reactions of chickens to eyespot patterns and chicken's retinal ontogeny focused on birds which are either younger than one week or older than three weeks (Olofsson et al., 2012; Olofsson et al., 2015; Rejdak et al., 2002). Future studies should explore age and ontogeny effects on responses to eyespot patterns.

6.1.3 Alternative ways to detect pessimism using an eyespot test

I used distance and latency to approach an eyespot cue to measure pessimism in the eyespot test as these were easy to record. However, reactions of chickens to eyespot patterns have also been recorded after birds reached or stopped walking towards the cue (Olofsson et al., 2012; Olofsson et al., 2015). Recorded reactions were alarm calls, attacks to the cue and latency to start foraging or to walk back to the area surrounding the cue (Olofsson et al., 2012; Olofsson et al., 2015). Additionally, reactions of starlings to eyespot cues were analysed through percentage of time spent in areas that differed for their distance from the eyespot cue (Brilot et al., 2009). To understand if any of these measurements could be a better tool to detect pessimism with the eyespot test, future work on validation of the eyespot test as a measure of affective state is encouraged to include them.

6.1.4 Effects of sex on detection of relationships between tests

In the original test, only females showed a relationship between tests. By contrast, in both simplified tests, there were no sex differences for detection of a relationship between tests or not. Therefore, as neither birds here naïve to the full eyespot cue or birds already exposed to the full eyespot cue showed sex effects, naivety to the full eyespot cue can likely be excluded as a factor affecting the relationship between tests. Differences in experimental setup cannot be excluded a priori as a factor, but age is a more likely explanation for differences between the simplified and original tests in detecting the relationship between tests depending on sex. In effect, adult female fowls are known to be more motivated to work for food rewards than males, and studies often exclude males from the sample when performing judgement bias tests or training for associating a cue to a reward (Zidar et al., 2017 b; Zidar et al., 2018 b). Therefore, tests performed when birds are closer to sexual maturity might lead to sex differences in latencies to approach cues due to differences in food motivation.

6.1.5 Effects of loss of ambiguity on detection of relationships between tests

In the original tests and the simplified tests that were carried out at two weeks, the presence or absence of a relationship between tests was not affected by the consideration of loss of ambiguity (i.e. learning that the ambiguous cue was unrewarded). More specifically, loss of ambiguity did not affect the possibility to detect relationships between tests, despite birds being tested at different ages and with different experimental setups. However, the tests at two weeks were followed by the tests at four weeks, while no affective state tests followed the original tests. Therefore, repeated exposures to the ambiguous cue can be a cause for loss of ambiguity affecting the tests at four weeks, as four-week-old birds had more experience with the judgement bias test (Bethell, 2015; Doyle et al., 2010; Košťál et al., 2020; Starling et al., 2014). The problem of repeated exposures will be further discussed in section 6.2.2.

6.1.6 Effects of extreme values on detection of relationships between tests

Extremely long latencies to approach either the ambiguous cues in the judgement bias tests or the eyespot cues in the eyespot tests, or both (Appendix section; Table A2; Figure A1-6) could have influenced results both when accounting, and when not accounting, for loss of ambiguity.

When accounting for loss of ambiguity, in the original tests, birds that scored extremely long latencies did so exclusively when approaching the ambiguous cue from the judgement bias test. Removing these extreme values caused the relationship between the judgement bias test and the eyespot test to disappear. When accounting for loss of ambiguity, in the simplified tests, birds that scored extremely long latencies did so more often when approaching the full eyespot cue than when approaching the ambiguous cue. Removing these extreme values caused the relationship between the judgement bias test and the eyespot test at two weeks to be revealed. Thus, when accounting for loss of ambiguity, the original tests were more likely to produce extreme values in the judgement bias test than in the eyespot test, while the opposite was true for the simplified tests. Consequently, the relationship between original tests seemed more dependent on extreme values scored in the judgement bias test, while the relationship between simplified tests seemed more dependent on extreme values scored in the eyespot test. Experimental setup differences alone cannot explain why removing extreme values affected the presence of the relationship between behavioural responses in the two tests in different ways when birds were tested in the original and simplified test at two weeks. Both the number of birds scoring extremely long latency to approach the eyespot cue and the ability of extreme values to disrupt the relationship between tests seemed to decrease with age. Previous research showed that young chickens were more intimidated by eyespot patterns than older chickens (Olofsson et al., 2015). Age effects on perception of eyespot patterns could therefore explain why extremely long latencies to approach the eyespot cue affect the presence of the relationship between tests in different ways depending on the used affective test version. Future studies exploring relationships between the judgement bias test and the eyespot test should therefore test birds that are older than two weeks.

When not accounting for loss of ambiguity, birds seemed to score extremely long latencies to approach the ambiguous cue fewer times than when accounting for loss of ambiguity. Moreover, when accounting for loss of ambiguity, the relationship between responses of females in the two original tests disappeared completely. On the contrary, when not considering loss of ambiguity in the tests at four weeks, only small changes occurred in the detection of the relationship between responses to the two tests. These patterns may indicate that results from both the original and simplified version are less affected by extreme values when including potential effects from loss of ambiguity. I recommend the use of an alternative method to control for loss of ambiguity than the one used here, as the method used in the present study (see section 4.2) caused more extreme values, which, in turn, seemed to affect the relationships on these. A method to exclude birds that experienced loss of ambiguity could be used directly in the testing phase. This could be done by exposing animals to a longer version of the judgement bias test to record responses to the ambiguous cue a higher number of times than the learning criterion in the training phase. Animals showing maximum latency to approach the ambiguous cue for the same number of trials as the learning criterion would be considered experiencing loss of ambiguity. To prevent the described higher number of needed test trial with the ambiguous cue to cause loss of ambiguity, the number of test trials with the ambiguous cue should not increase a lot. To limit this increase to the minimum, the number of trials in a row for the learning criterion could be reduced to five, which is still supported as a valid criterion (Sorato et al., 2018). In this way, a judgement bias test with five or more exposures to the ambiguous cue would be likely to detect if any individuals experienced loss of ambiguity directly in the testing phase.

6.2 Relationships between responses to the judgement bias test and the eyespot test

6.2.1 Original tests

Relationships between behavioural responses to the ambiguous middle grey cue in the original judgement bias test and to one of the full eyespot cues in the original eyespot test were present and positive even when accounting for the loss of ambiguity of the ambiguous cue in the judgement bias test. This confirms my hypothesis that a relationship between behavioural responses observed in the two tests can be present. I here assume that individuals with much longer latencies to the eyespot cue correspond to birds that found that cue particularly aversive. I make this assumption as I exclude that birds can ignore the instinctive aversion to eyespots. The only way for birds to overcome their instinct to avoid, could be by learning that the eyespot cue is rewarded or not dangerous. However, I assume this to be unlikely, given such low number of exposures to the cue (three in the original, and four in the simplified eyespot test, at two weeks of age). Individuals having much longer latencies to the ambiguous cue (i.e. that are particularly low in optimism, Mendl et al., 2009) also had longer latencies to the full eyespot cue (i.e. perceived the full eyespot cue as particularly aversive, possibly being particularly pessimistic, Bateson, 2016). These individuals led the relationships between responses to the two original tests in a positive direction, supporting my hypothesis.

6.2.2 Simplified tests

Contrary to my hypothesis, no relationships were detected between the judgement bias test and the eyespot test run when birds were only two weeks. The finding that relationships between behavioural responses to the two tests showed similar strength and directions independent of accounting or not for possible loss of ambiguity, suggests that loss of ambiguity did not occur. However, removal of individuals showing extreme values in either test, revealed a relationship between tests only when accounting for loss of ambiguity. This dependence on extreme values shows that individuals that are particularly low in optimism and that perceived the full eyespot cue as particularly aversive when accounting for loss of ambiguity, were disrupting the relationship between tests, instead of leading it. More specifically, some individuals had extremely long latencies to the full eyespot cue without showing long latencies to the ambiguous cue, or the opposite. Therefore, not all individuals that showed extreme values were both less optimistic and more pessimistic. The mismatch between extreme values in behavioural responses from the two tests is probably the cause for disrupting the relationship between responses observed in the two tests.

The predicted relationship was detected between responses from tests run when birds were four weeks. However, this relationship was absent when accounting for possible loss of ambiguity, which indicates that birds might have learnt that the ambiguous cue was unrewarded. Loss of ambiguity is a common methodological problem when repeatedly exposing animals to ambiguous cues in judgement bias tests (Doyle et al., 2010; Starling et al., 2014; reviewed in Bethell 2015; Košťál et al., 2020). Loss of ambiguity has been observed through decreased likelihood and increased latency of approaching ambiguous cues (Doyle et al., 2010; Starling et al., 2014). Here in my work, the upper limit of the range of mean latencies to approach the ambiguous cue seemed to shift towards longer latencies when birds were tested the second time at four weeks, which may support that birds experienced loss of ambiguity (test at two weeks: range = 0.39 - 16.01 s; test at four weeks: range = 0.29 - 24.70 s). It is unlikely that this shift in the range of mean latency is due to birds having a worse affective state, since relative mean latency measurements were originated from birds in the control group, whose affective state was not experimentally manipulated, and since housing conditions did not change. The observation that some birds repeatedly showed maximum trial time over the last trials in the judgement bias test at four weeks (Appendix; Table A3) also supports the occurrence of loss of ambiguity. As a result, the simplified test cannot reliably detect a relationship between tests at four weeks, as it shows dependency on whether the possibility of loss of ambiguity is considered or not. Additionally, it is not possible to exclude that the presence of the relationship between behavioural responses from the two tests was partially affected by age. The impossibility to disentangle effects of age and loss of ambiguity and to exclude these effects leaves open the question on how to interpret the relationship between the tests at four weeks that was detected when excluding loss of ambiguity. A possible explanation is that affective state of birds changed between testing instances due to other factors than housing conditions (e.g. personality, social interactions; Košťál et al., 2020). Alternatively, contrary to my initial assumptions, birds could have overcome aversion towards eyespot patterns by learning the association between the full eyespot cue and the reward (e.g. after further exposures to the full eyespot cue at four weeks). This learning process would explain shorter latencies to approach the full eyespot cue at four weeks than at two weeks. Latencies to approach the full eyespot cue seemed shorter also when removing individuals that found the eyespot cue particularly aversive in the tests at two weeks (Appendix; Figure A3). Outlier removal in the test at two weeks and further exposure to full eyespot cues experienced by all birds tested at four weeks had one more common point. Both after removing outliers in the 2-week tests and after birds were exposed again to eyespot cues in the 4-week test, the relationship between tests appeared with tendency. This suggests that, similarly to when removing those individuals that were putatively most scared by the full eyespot cue in the test at two weeks, further exposures to the full eyespot cue could have reduced or eliminated the instinctive aversion to eyespot patterns in the test performed at four weeks. In other words, birds might have overcome the instinctive aversion to eyespot cues. Therefore, there is a possibility that the relationship between responses to the two tests that was observed at four weeks is merely a side effect of learning the association between the eyespot cue and the reward. However, these speculations need further investigation of the possibility of learning that an eyespot pattern is not aversive and the consequences of this phenomenon on limited repeatability of the eyespot test and reliability of test results.

6.3 Possibility of using the full eyespot cue as an ambiguous cue for future tests

Overall, only one full eyespot cue presented the predicted correlation with responses to one ambiguous cue from the judgement bias test. This correlation was present under specific conditions of age, sex, and loss of ambiguity. Moreover, the ambiguous eyespot cues seemed not to be functionally perceived as ambiguous, and one full eyespot cue still needs to be fully validated as ambiguous. Therefore, more work is needed before implementing an eyespot test based on observations of reactions to both full and ambiguous eyespot cues (Brilot et al., 2009; Figure 2A). However, further studies controlling for effects linked to age, sex and loss of ambiguity can potentially validate the full eyespot cue as an ambiguous near aversive cue (i.e. functionally perceived as ambiguous, despite displaying aversive eyespot patterns). This would result in the possibility to implement an eyespot test based on just one ambiguous near aversive cue (Figure 2B). In such a test, reactions of birds with known affective state to the ambiguous near aversive cue would be compared to reactions of birds with unknown affective state to the same cue. This would allow to indirectly determine affective state using the pessimism bias in a way that is easy to apply.

7. Conclusion

I have here shown that red junglefowl chicks had functionally similar behavioural responses to two different affective state tests: a judgement bias test and an eyespot test. The relationship was present despite potential loss of ambiguity when the original tests were used and for females. The relationship was not present for males tested with the original tests. The relationship was present, but likely affected by loss of ambiguity and age for young chicks tested with the simplified tests. To my knowledge, this is the first study evaluating the use of eyespot patterns for detecting affective states by comparing it to a judgement bias test. My work should encourage more studies exploring the possibility of the eyespot test as a simpler alternative to the classical judgement bias test. Future studies should further explore effects of age and loss of ambiguity and compare the two affective state tests at different ages to improve our understanding of how to implement the eyespot test.

8. Societal and ethical aspects of my work

During this study, animals were exposed to some mildly stressful procedures that were necessary to carry out all experiments according to the established setup. Most of these procedures, such as vaccinations, wing tagging, and exposure to putatively aversive eyespot cues, were short-term stressors. Birds were then allowed to be in their home cages with conspecifics and were not trained or tested in the same day. Other procedures might have been perceived as relatively longer stressors than the procedures described above. These procedures included the process of transporting chicks from the hatchery to the lab and temporary separation from conspecifics, occasionally in novel environments (i.e. during testing). For the first procedure, time spent outdoors during transportation was minimised to avoid heat loss and thermal stress. For carrying out transportation in the most efficient way, chicks were divided into large groups which they were transported together in until the laboratory was reached. For the second procedure, birds were habituated to being alone by gradually reducing the number of conspecifics. After transportation or habituation, birds were allowed to rest and were not trained or tested in the same day. All procedures followed the Swedish national guidelines for handling captive research animals and were approved by the Swedish Regional Ethical Committee.

This study contributes to the field of animal welfare research, as it is the first to find encouraging results for validating the use of a simpler test using eyespot cues to measure affective states in the spectrum of pessimism. More specifically, I showed that responses to a full eyespot cue are comparable to responses to ambiguous cues from an already validated test (the judgement bias test). By doing so, I made research move the first steps towards validating the use of full and ambiguous eyespot cues to detect variation in pessimism in avian models. However, more studies are needed to confirm my findings. In particular, more results in this direction could lead to acceptance of the eyespot test as a tool for affective state assessment used for research aimed to measure affective state. In this framework, the eyespot test would be a more easily implemented alternative to the judgement bias test. The advantages of a simplified affective state test are many for both researchers and study animals. Firstly, researchers would need less time to train animals and animals would spend shorter time away from their conspecifics, which would likely decrease their stress level. Secondly, animals would be less frustrated than in a judgement bias test, as the eyespot test has always rewarded cues, regardless of the cue being aversive or not. Finally, researchers would benefit from excluding all these effects from their data. As a result, research in animal welfare focusing on measuring affective state of animals would be easier, faster, and more efficient.

Additionally, as the ancestor of domestic chicken was used, this study offers applications for animal welfare assessment in poultry industry, where birds typically suffer poor welfare. Chickens in industrial environments are highly stocked and exposed to extreme levels of ammonia, which can cause problems to visual and respiratory systems (Robins & Phillips, 2011). The use of a simple affective state test, such as the eyespot test, for industrial chickens, would provide further evidence of their poor welfare, and contribute to the drive to improve their living conditions. In a long-term perspective, the eyespot test could be also used to monitor how efficient interventions are in improving chickens' welfare.

9. Acknowledgements

I would like to thank my supervisors Hanne Løvlie and Laura Garnham, for their guidance and feedback during my master project. Thanks to Kristoffer Lundgren, Lelde Šuksta, Charlie Clarke, Ina Eriksson and Clara Farnebäck for their support during experimental planning, data collection in the chick lab, and advice for statistics. I would also like to thank Veronica Samelius for taking care of our chicks once they were moved to Vreta, and my examiner Per Jensen and my fellow students for reviewing my thesis draft.

10. References

Aguinis, H., Gottfredson, R.K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16 (2), 270-301.

Bateson, M. (2016). Optimistic and pessimistic biases: a primer for behavioural ecologists. *Science Direct*, 12, 115–121.

Bethell, J., E. (2015). A "how-to" guide for designing judgment bias studies to assess captive animal welfare. *Journal of Applied Animal Welfare Science*, 18, S18-S42.

Boddington, R., Gómez Dunlop, C.A, Garnham, L.C., Ryding, S., Abbey-Lee, R.N., Kreshchenko, A., & Løvlie, H. (2020). The relationship between monoaminergic gene expression, learning, and optimism in red junglefowl chicks. *Animal Cognition*, 23, 901-911.

Brilot, B. O., Normandale, C. L., Parkin, A., & Bateson, M. (2009). Can we use starlings' aversion to eyespots as the basis for a novel 'cognitive bias' task? *Applied Animal Behaviour Science*, 118, 182–190.

Burman, O.H.P., Parker, R., Paul, E.S., & Mendl, M. (2008). A spatial judgement task to determine background emotional state in laboratory rats, Rattus norvegicus. *Animal Behaviour*, 76, 801–809.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd Edition, Lawrence Erlbaum Associates Publishers.

Crump, A., Arnott, G., & Bethell, E.J. (2018). Affect-driven attention biases as animal welfare. Indicators: review and methods. *Animals*, 8, 136.

Dovoedo, Y.H., & Chakraborti, S. (2015). Boxplot-based outlier detection for the location-scale family. *Communications in Statistics - Simulation and Computation*, 44, 1492-1513.

Doyle, R.E., Vidala, S., Hinch, G.N., Fisher, A.D., Boissy, A., & Lee, C. (2010). The effect of repeated resting on judgement biases in sheep. *Behavioural Processes*, 83, 349-352.

Favati, A., Zidar, J., Thorpe, H., Jensen, P., Løvlie, H. (2016). The ontogeny of personality traits in the red junglefowl, *Gallus gallus*. *Behavioural ecology*, 27, 484-493.

Garnham, L.C., Porthén, S.A., Child, S., Forslind, S., & Løvlie, H. (2019). The role of personality, cognition, and affective state in same-sex contests in the red junglefowl. *Behavioral Ecology and Sociobiology*, 73, 149-161.

Harding, E.J., Paul, E.S., & Mendl, M. (2004). Animal behaviour: cognitive bias and affective state. *Nature*, 427, 312.

Hedlund, L., Palazon, T., Jensen, P. (2021). Stress during Commercial Hatchery Processing Induces Long-Time Negative Cognitive Judgement Bias in Chickens. *Animals*, 11, 1083.

Košťál, L., Skalná, Z., & Pichová, K. (2020). Use of cognitive bias as a welfare tool in poultry. *Journal of Animal Science*, 98, S63-S79.

Lagisz, M., Zidar, J., Nakagawa, S., Neville, V., Sorato, E., Paul, E.S., Bateson, M., Mendl, M., & Løvlie, H. (2020). Optimism, pessimism and judgement bias in animals: a systematic review and meta-analysis. *Neuroscience and Biobehavioural Reviews*, 118, 3-17.

Matheson, S.M., Asher, L., & Bateson, M. (2008). Larger, enriched cages are associated with 'optimistic' response biases in captive European starlings (Sturnus vulgaris). *Applied Animal Behaviour Science*, 109, 374–383.

Mendl, M., Burman, O. H., Parker, R. M., & Paul, E. S. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, 118, 161–181.

Neville, V., Nakagawa, S., Zidar, J., Paul, E.S., Lagisz, M., Bateson, M., Løvlie, H., & Mendl, M. (2020). Pharmacological manipulations of judgement bias: a systematic review and meta-analysis. *Neuroscience and Biobehavioural Reviews*, 108, 269–286.

Olofsson M., Løvlie H., Tibblin J., Jakobsson S., & Wiklund C. (2012). Eyespot display in the peacock butterfly triggers antipredator behaviours in naive adult fowl. *Behavioral Ecology*, 24, 305-310.

Olofsson, M., Wiklund, C., & Favati, A. (2015). On the deterring effect of a butterfly's eyespot in juvenile and sub-adult chickens. *Current Zoology*, 61 (4), 749–757.

Paul, E.S., Harding, E.J., & Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*, 29, 469–491.

Robins, A. & Phillips, C., J., C. (2011). International approaches to the welfare of meat chickens. *World's Poultry Science Journal*, 67 (2), 351-369.

Rejdak, R., Zarnowski, T., Turski, W.A., Kocki, T., Zagorski, Z., Gwenther, E., Kohler, K., & Zrenner, E. (2002). Changes of kynurenic acid content in the rat and chicken retina during ontogeny. *Graefe's Archive for Clinical and Experimental Opthalmology*, 240, 687-691.

Ruxton, G.D., Sherratt, T.N., & Speed, M.P. (2004). *Avoiding Attack*. 2nd Edition, Oxford University Press.

Shettleworth, S.J. (2010). *Cognition, Evolution, and Behaviour*. 2nd Edition, Oxford University Press.

Sorato, E., Zidar, J., Garnham, L., Wilson, A., & Løvlie, H. (2018). Heritabilities and co-variation among cognitive traits in red junglefowl. *Philosophical Transactions Royal Society B*, 373, 20170285.

Starling, M.J., Branson, N., Cody, D., Starling, T.R., & McGreevy, P.D. (2014). Canine Sense and sensibility: tipping points and response latency variability as an optimism index in a canine judgement bias assessment. *PLOS ONE*, 9, 1-15.

Stevens, M., Hopkins, E., Hinde, W., Adcock, A., Connolly, Y., Troscianko, T., & Cuthill, I.C. (2007). Field experiments on the effectiveness of 'eyespots' as predator deterrents. *Animal Behaviour*, 74, 1215-1227.

Stevens, M., Hardman, C.J., & Stubbins, C.L. (2008). Conspicuousness, not eye mimicry, makes "eyespots" effective antipredator signals. *Behavioral Ecology*, 19 (3), 525–531.

Stevens, M., & Ruxton, G.D. (2014). Do animal eyespots really mimic eyes? *Current Zoology*, 60 (1), 26–36.

Zidar, J., Sorato, E., Malmqvist, A.M., Jansson, E., Rosher, C., Jensen, P., Favati, A., & Løvlie, H. (2017 a). Early experience affects adult personality in the red junglefowl: a role for cognitive stimulation? *Behavioural Processes*, 134, 78–86.

Zidar, J., Balogh, A., Favati, A., Jensen, P., Leimar, O., & Løvlie, H. (2017 b). A comparison of animal personality and coping styles in the red junglefowl. *Animal Behaviour*, 130, 209-220.

Zidar, J., Campderrich, I., Jansson, E., Wichman, A., Winberg, S., Keeling, L., & Løvlie, H. (2018 a). Environmental complexity buffers against stress-induced negative judgment bias in female chickens. *Scientific Reports*, 8, 5404.

Zidar, J., Balogh, A., Favati, A., Jensen, P., Leimar, O., Sorato, E., & Løvlie, H. (2018 b). The relationship between learning speed and personality is age- and task-dependent in red junglefowl. *Behavioral Ecology and Sociobiology*, 72, 168.

Zidar, J., Balogh, A.C.V., Leimar, O., & Løvlie, H. (2019). Generalization of learned preferences covaries with behavioral flexibility in red junglefowl chicks. *Behavioral Ecology*, 30, 1375-1381.

11. Appendix

Table A1: Overview of two behavioural tests used to measure affective state in animals; judgement bias test and eyespot test.

	Judgement bias	**Eyespot**
Training	Positive reinforcement associated with a cue is used to measure optimism. Negative reinforcement associated with a cue is used to measure pessimism.	Habituation to reward associated with a position (control cue).
Optimism	Expectation of a positive outcome (e.g. reward) when the test cue is novel and ambiguous[1,2]. Latencies to the **ambiguous test cue** closer to the + cue indicate +/– ➡ optimism +/N➡ increased optimism Latencies to the **ambiguous test cue** closer to the other learnt cue indicate +/– ➡ pessimism +/N➡ decreased optimism	Not applicable (see below).
Pessimism	Expectation of a negative outcome (e.g. punishment) when the presented cue is novel and ambiguous[1,2]. Latencies to the **ambiguous test cue** closer to the – cue indicate +/– ➡ pessimism –/N➡ increased pessimism Latencies to the **ambiguous test cue** closer to the other learnt cue indicate +/– ➡ optimism –/N➡ decreased pessimism	Unlearnt perceived predation risk when exposed to an ambiguous aversive cue containing eyespot patterns. Latencies to the **ambiguous eyespot cue**[3] closer to the control cue indicate decreased pessimism, and latencies to the **ambiguous eyespot cue**[3] closer to the full eyespot cue indicate increased pessimism.

Definitions of optimism and pessimism are given, together with description of methods for judgement bias test (Harding et al., 2004), and eyespot test (Brilot et al., 2009). ‘Training’ indicates if training is needed prior to testing, with further details if needed. ‘Optimism’ refers to how optimism manifests in a test and to the measured variable. ‘Pessimism’ refers to how pessimism manifests in a test and to the measured variable. In bold is the used test cue. +/–; +/N; –/N indicate combinations of learnt cues, where + indicates that one learnt cue is rewarded, – that one learnt cue is punished, N that one learnt cue is neither rewarded nor punished. In all tests, increase or decrease in both optimism and pessimism levels is referred to baseline measurements from the same individual or the same group. References: 1. Mendl et al., 2009; 2. Lagisz et al., 2020; 3. Brilot et al., 2009.

Table A2: Comparisons between behavioural responses of red junglefowl chicks to the judgement bias test and the eyespot test with removal of outliers detected with boxplot.

	MG: Mean latency	MG: Shortest latency
Original		
Cue 3: Mean distance	♂♂:	♂♂:
	Removal of PEO (n = 1)	Removal of PEO
	R_S = 0.28, n = 20, p = 0.23	No variation in data (mean distance to cue 3 = 0)
		♀♀:
		Removal of PEO (n = 2)
		R_S = 0.32, n = 16, p = 0.23; Figure A1
Cue 3: Mean latency	♂♂:	♂♂:
	Removal of PEO (n = 1)	Removal of PEO (n = 3)
	R_S = 0.26, n = 20, p = 0.27	R_S = 0.21, n = 18, p = 0.40
		♀♀:
		Removal of PEO (n = 2)
		R_S = 0.16, n = 16, p = 0.56; Figure A2
Simplified 2-weeks		
Cue 3: Mean distance	Removal of PEO	
	No variation in data (mean distance to cue 3 = 0)	
Cue 3: Mean latency	Removal of EO (n = 1)	Removal of EO (n = 1)
	R_S = 0.11, n = 80, p = 0.33; Figure A3A	R_S = 0.13, n = 80, p = 0.26; Figure A4A
	Removal of PEO (n = 11)	*Removal of PEO (n = 13)*
	R_S = 0.09, n = 70, p = 0.44; Figure A3B	_R_S = 0.22, n = 68, p = 0.08*; Figure A4B_

Simplified 4-weeks		
Cue 3: Mean distance	NA	NA
Cue 3: Mean latency	_Removal of PEO (n = 1)_ _$R_S = 0.42$, n = 20, p = 0.06*; Figure A6_	Removal of PEO (n = 7) $R_S = 0.03$, n = 39, p = 0.87; Figure A5

*tendency ($R_S > 0.40$; $p < 0.10$)

Cue 3 is the full eyespot cue whose responses correlated with responses to one ambiguous cue from the judgement bias test. ‘Cue 3: Mean distance’ is the relative mean measure of closest distance (cm) of a bird from cue 3, one of the two full eyespot cues used in the eyespot test. ‘Cue 3: Mean latency’ is the relative mean measure of shortest latency (s) of a bird to approach cue 3. MG is the ambiguous cue whose responses correlated with responses to one eyespot cue from the eyespot test. ‘MG: Mean latency’ is the relative mean measure of shortest latency (s) of a bird to approach the middle grey cue. ‘MG: Shortest latency’ is the relative shortest measure of shortest latency (s) of a bird to approach the middle grey cue, which refers to correlations that account for possible loss of ambiguity. ‘♂♂’ refers to the male subset. ‘♀♀’ refers to the female subset. ‘NA’ indicates that the analysis was not run for that comparison. ‘Removal of PEO’ indicates the results of correlations run when potential error outliers were removed. ‘Removal of EO’ indicates the results of correlations run when error outliers were removed. In italics are relationships that were influenced by eliminating potential error outliers.

Table A3: Summary of maximum latency in trials where red junglefowl chicks approached the ambiguous middle grey cue in the simplified judgement bias tests.

Test	**Number of individuals**
Simplified 2-week test	
1, last	19
2, last	2
3, last	2
Simplified 4-week test	
1, last	2
2, last	2
3, last	1
4, last	2
3, middle	3

The numbers in the left column refer to the number of trials in a row where a bird had maximum latency to approach the ambiguous middle grey cue. ‘last’ indicates that the last trial in the sequence of maximum-time-trials is also the last trial of the test session. ‘middle’ indicates that the last trial in the sequence of maximum-time-trials is not the last trial of the test session. The right column indicates the number of individuals that had maximum latency for the corresponding number or trials of exposure to the ambiguous middle grey cue.

Table A4: Summary of relevant factors influencing the presence and strength of the relationship between responses of red junglefowl chicks to judgement bias tests and eyespot tests, in different versions and experimental conditions.

	Distance	Latency	Sex	Loss of ambiguity	PEO
Original test	N	N	Y	N	Y
Simplified test at 2-week	Y	Y	N	N	Y
Simplified test at 4-week	Y	N	N	Y	N
Possible explanation	Runway length	Age	Age	Number of exposures to middle grey cue	Experimental setup, Age

Each column represents a single factor considered when exploring a relationship between the two affective state tests. ‘Distance’ indicates if the relationship is detectable when using closest distance approached from cue 3, the full eyespot cues whose responses correlated with responses to one ambiguous cue from the judgement bias test. ‘Latency’ indicates if the relationship is detectable when using shortest latency to approach cue 3. ‘Sex’ indicates if the relationship differs between sexes. ‘Loss of ambiguity’ indicates if the relationship is altered when using shortest measure of shortest latency instead of mean measure of shortest latency to approach the middle grey cue, the ambiguous cue whose responses correlated with responses to one full eyespot cue from the eyespot test. ‘PEO’ (i.e. potential error outlier) indicates if the relationship is altered when removing potential error outliers and assuming the possibility of loss of ambiguity (i.e. Learning that the ambiguous cue is unrewarded). ‘Y’ and ‘N’ stand respectively for yes and no. ‘Possible explanation’ refers to factor(s) most likely to explain the ability of the different test versions to detect the relationship between affective state measurements in the specified experimental condition (i.e. with the use of distance, latency, one sex, considering loss of ambiguity and PEO).

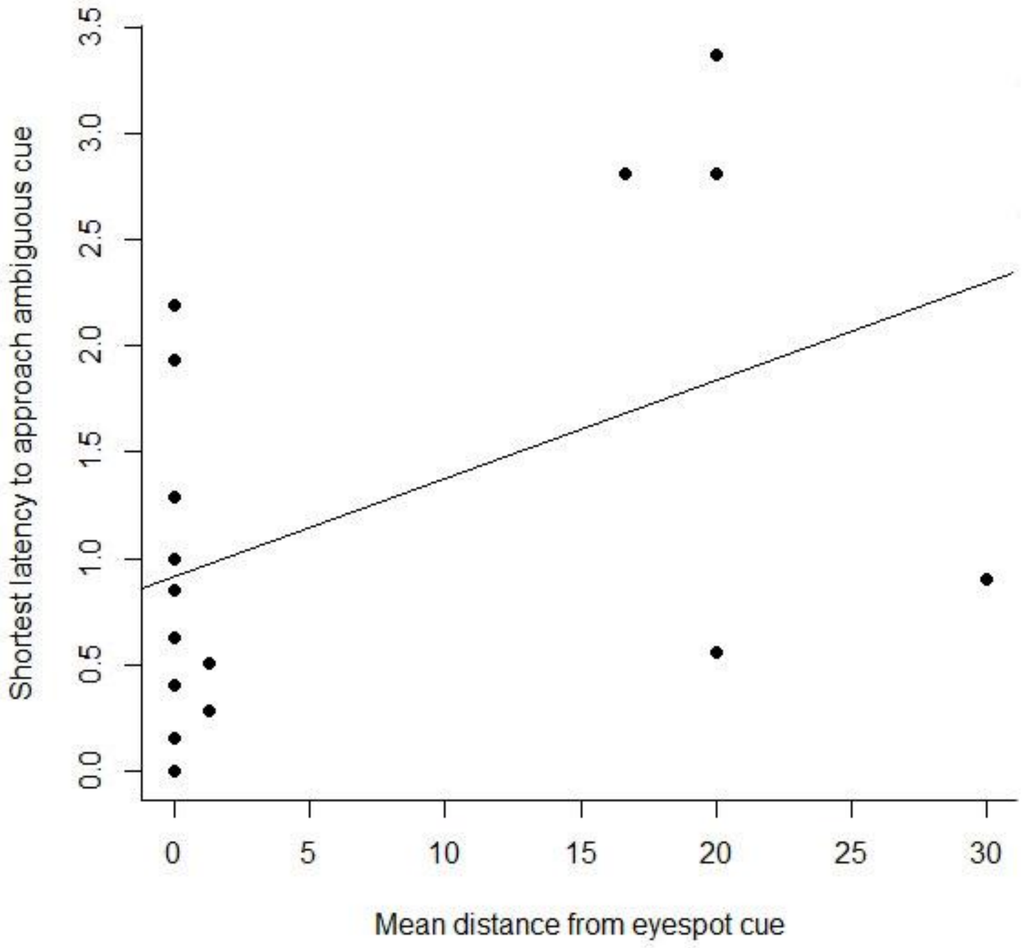

Figure A1: Comparison of behavioural responses to the original judgement bias test and eyespot test in female red junglefowl chicks, with mean measure of closest distance to the eyespot cue 3 (i.e. a full eyespot cue) and exclusion of potential error outliers. ‘Shortest latency to approach ambiguous cue’ is the relative shortest measure of shortest latency (s) to approach the middle grey cue (i.e. novel, ambiguous cue) in the judgement bias test. ‘Mean distance from eyespot cue’ is the relative mean measure of shortest distance (cm) reached from the eyespot cue 3 in the eyespot test. Each point represents an individual bird. The trend line is correlation trend line. Two potential error outliers for extremely high values in relative shortest measure of shortest latency to approach the middle grey cue are removed.

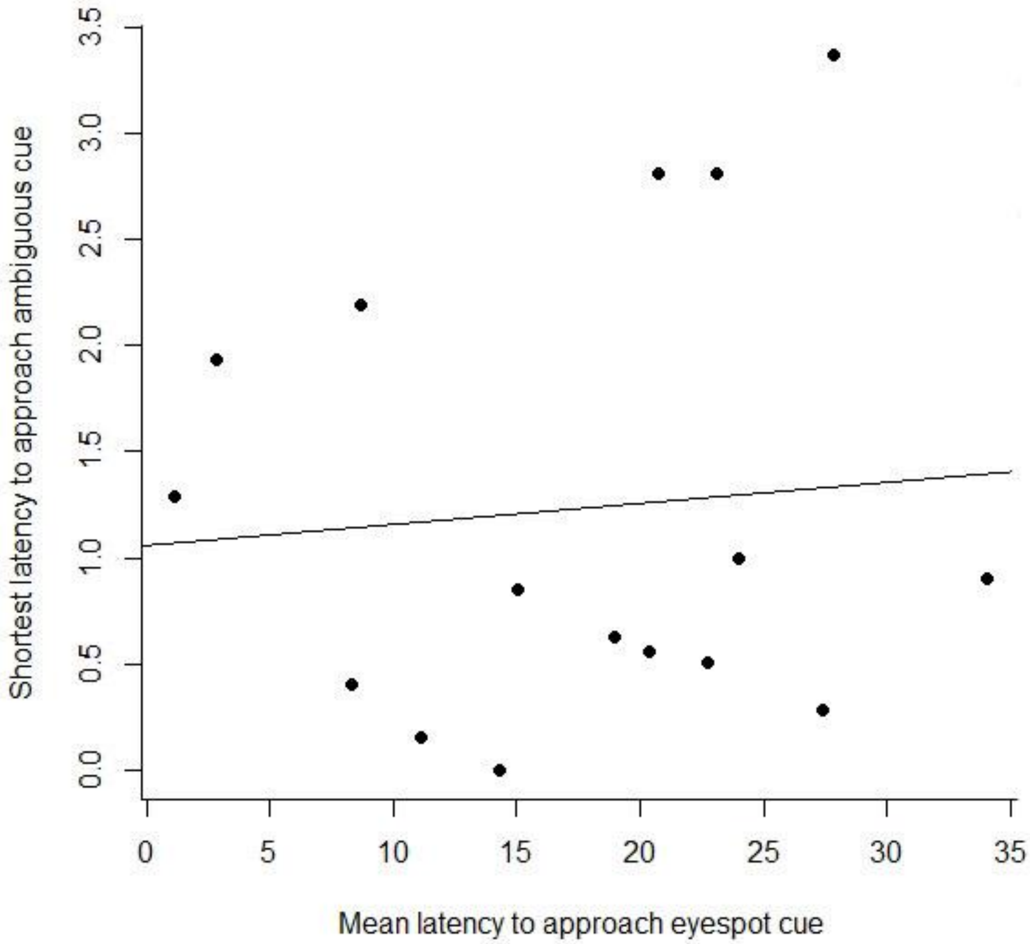

Figure A2: Comparison of behavioural responses to the original judgement bias test and eyespot test in female red junglefowl chicks, with mean measure of shortest latency to approach the eyespot cue 3 (i.e. a full eyespot cue) and exclusion of potential error outliers. ‘Shortest latency to approach ambiguous cue’ is the relative shortest measure of shortest latency (s) to approach the middle grey cue (i.e. novel, ambiguous cue) in the judgement bias test. ‘Mean latency to approach eyespot cue’ is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 in the eyespot test. Each point represents an individual bird. The trend line is correlation trend line. Two potential error outliers for extremely high values in relative shortest measure of shortest latency to approach the middle grey cue are removed.

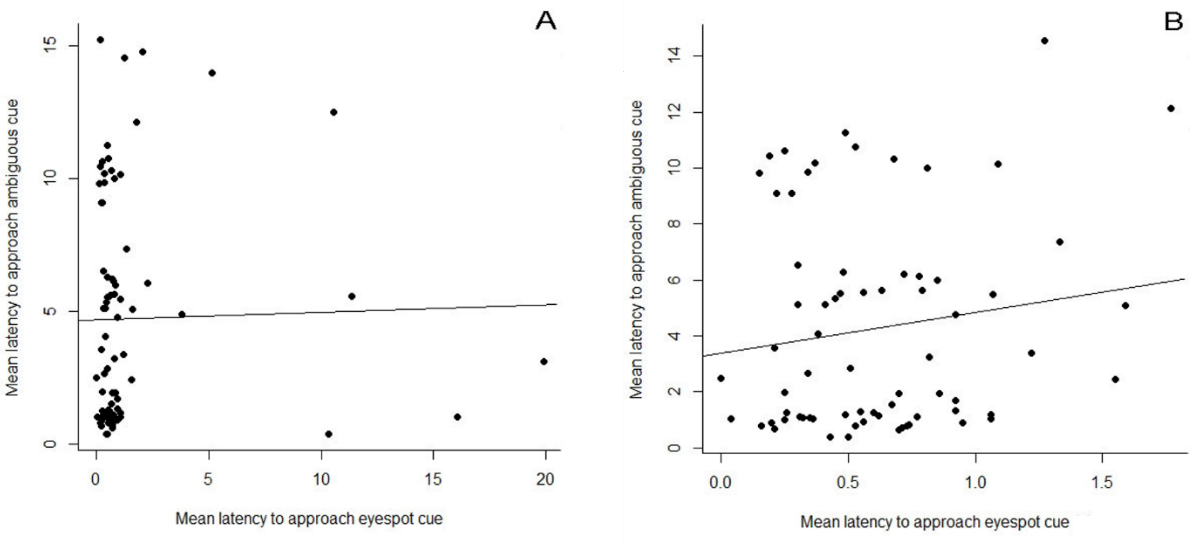

Figure A3: Comparison of behavioural responses to the judgement bias test and eyespot test in red junglefowl chicks at two weeks, without accounting for possible loss of ambiguity and with exclusion of outliers. A) One error outlier is removed for incorrect completion of data collection sheet in the judgement bias test and inability to start trials unless facing cue in the eyespot test. B) Ten potential error outliers for extremely high values in mean measure of shortest latency to approach cue 3 (i.e. a full eyespot cue) and three potential error outliers for extremely high values in mean measure of shortest latency to approach the middle grey cue (i.e. novel, ambiguous cue) are removed. 'Mean latency to approach ambiguous cue' is the relative mean measure of shortest latency (s) to approach the middle grey cue in the judgement bias test. By using the mean measure of shortest latency instead of the shortest measure of shortest latency, potential effects of loss of ambiguity are considered. 'Mean latency to approach eyespot cue' is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 in the eyespot test. Each point represents an individual bird. The trend line is correlation line.

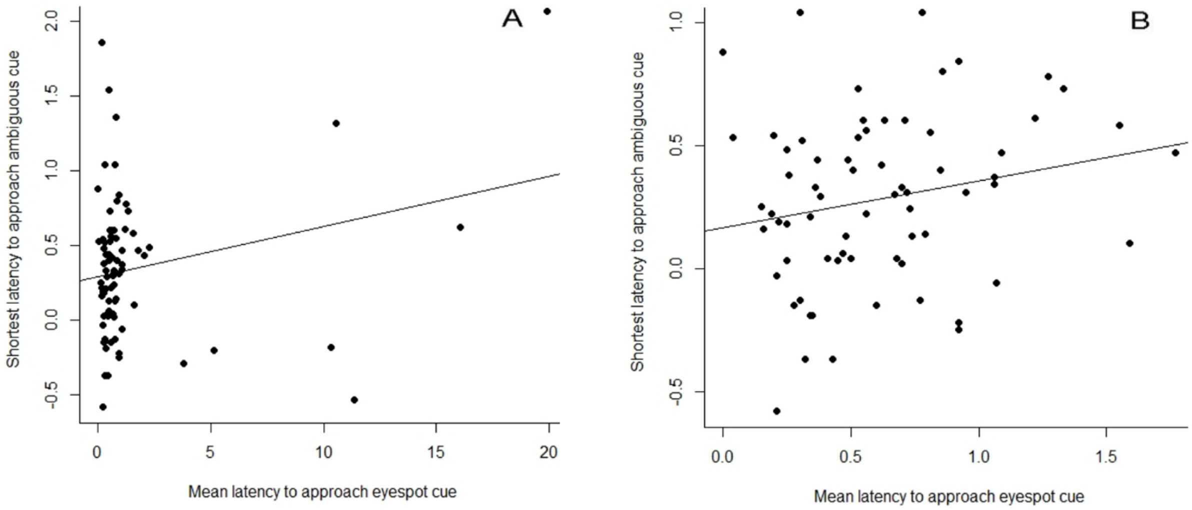

Figure A4: Comparison of behavioural responses to the judgement bias test and eyespot test in red junglefowl chicks at two weeks, accounting for possible loss of ambiguity and with exclusion of outliers. A) One error outlier is removed for incorrect completion of data collection sheet in the judgement bias test and inability to start trials unless facing cue in the eyespot test. B) 10 potential error outliers for extremely high values in mean measure of shortest latency to approach cue 3 (i.e. a full eyespot cue) and five potential error outliers for extremely high values in mean measure of shortest latency to approach the middle grey cue (i.e. novel, ambiguous cue) are removed. 'Shortest latency to approach ambiguous cue' is the relative shortest measure of shortest latency (s) to approach the middle grey cue in the judgement bias test. By using the shortest measure of shortest latency instead of the mean measure of shortest latency, potential effects of loss of ambiguity are considered. 'Mean latency to approach eyespot cue' is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 in the eyespot test. Each point represents an individual bird. The trend line is correlation line.

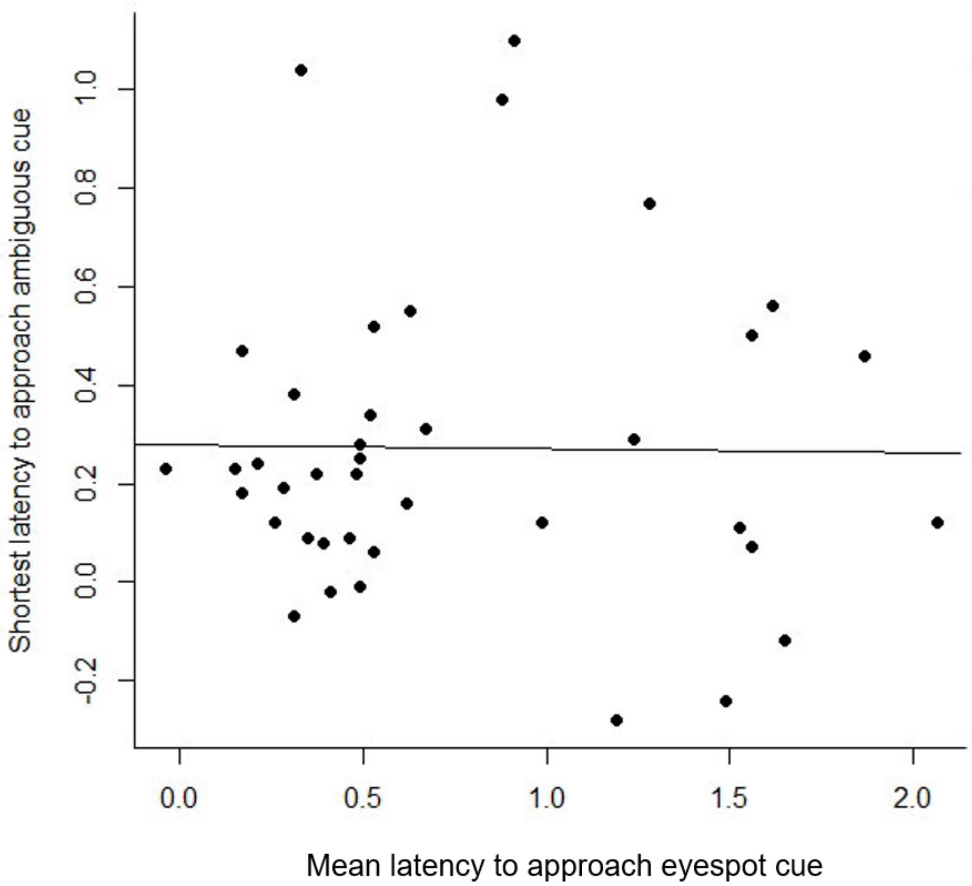

Figure A5: Comparison of behavioural responses to the judgement bias test and eyespot test in red junglefowl chicks at four weeks, accounting for possible loss of ambiguity and with exclusion of potential error outliers. ‘Shortest latency to approach ambiguous cue’ is the relative shortest measure of shortest latency (s) to approach the middle grey cue (i.e. novel, ambiguous cue) in the judgement bias test. By using the shortest measure of shortest latency instead of the mean measure of shortest latency, potential effects of loss of ambiguity are considered. ‘Mean latency to approach eyespot cue’ is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 (i.e. a full eyespot cue) in the eyespot test. Each point represents an individual bird. The trend line is correlation line. Two potential error outliers for extremely high values in relative shortest measure of shortest latency to approach cue 3, one potential error outlier for extremely low and five for extremely high values in mean measure of shortest latency to approach the middle grey cue are removed.

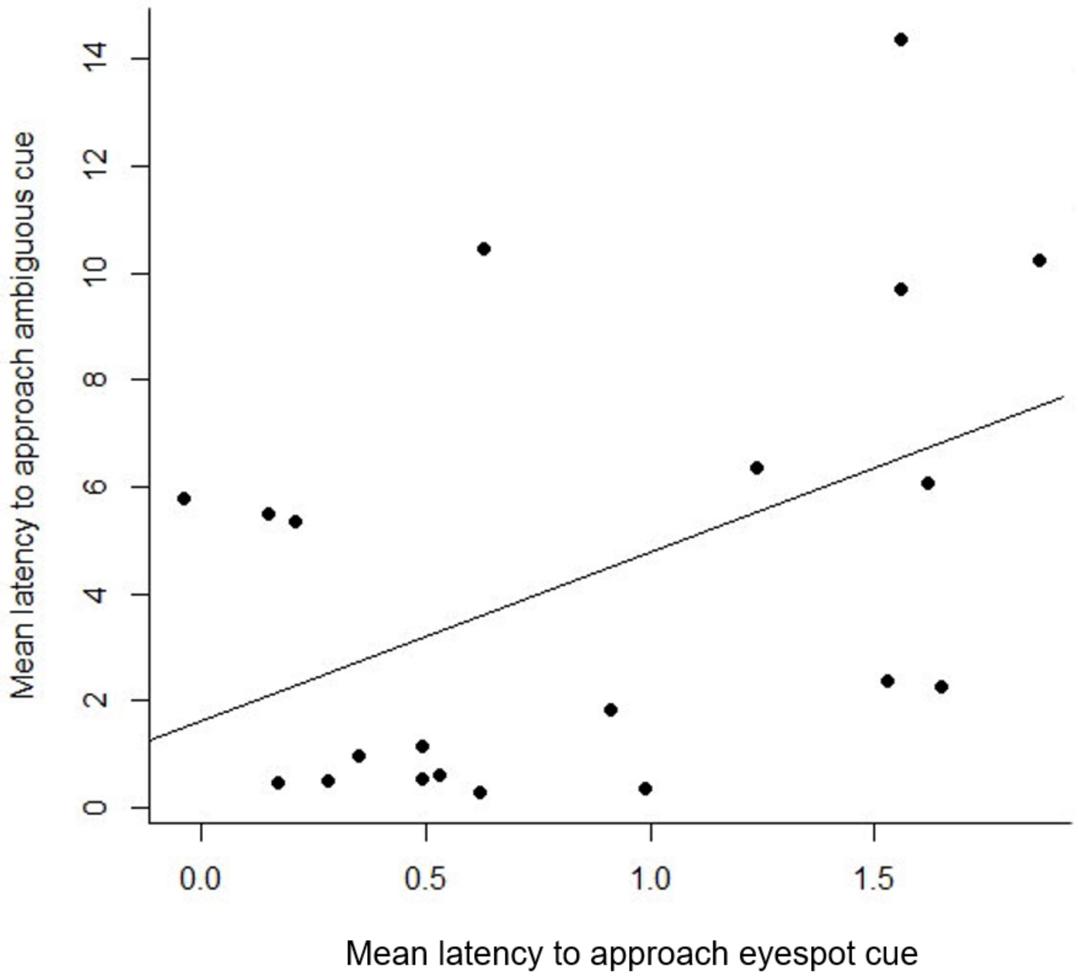

Figure A6: Comparison of behavioural responses to the judgement bias test and eyespot test in red junglefowl chicks at four weeks, without accounting for possible loss of ambiguity and with exclusion of potential error outliers. 'Mean latency to approach ambiguous cue' is the relative mean measure of shortest latency (s) to approach the middle grey cue (i.e. novel, ambiguous cue) in the judgement bias test. By using the mean measure of shortest latency instead of the shortest measure of shortest latency, potential effects of loss of ambiguity are not considered. 'Mean latency to approach eyespot cue' is the relative mean measure of shortest latency (s) to approach the eyespot cue 3 (i.e. a full eyespot cue) in the eyespot test. Each point represents an individual bird. The trend line is correlation trend line. One potential error outlier for extremely high values in relative mean measure of shortest latency to approach the middle grey cue is removed.