

## RESEARCH ARTICLE

# Chicken colour discrimination depends on background colour

Peter Olsson<sup>1,‡</sup>, Robin D. Johnsson<sup>1,\*</sup>, James J. Foster<sup>1</sup>, John D. Kirwan<sup>1</sup>, Olle Lind<sup>2</sup> and Almut Kelber<sup>1</sup>

## ABSTRACT

How well can a bird discriminate between two red berries on a green background? The absolute threshold of colour discrimination is set by photoreceptor noise, but animals do not perform at this threshold; their performance can depend on additional factors. In humans and zebra finches, discrimination thresholds for colour stimuli depend on background colour, and thus the adaptive state of the visual system. We have tested how well chickens can discriminate shades of orange or green presented on orange or green backgrounds. Chickens discriminated slightly smaller colour differences between two stimuli presented on a similarly coloured background, compared with a background of very different colour. The slope of the psychometric function was steeper when stimulus and background colours were similar but shallower when they differed markedly, indicating that background colour affects the certainty with which the animals discriminate the colours. The effect we find for chickens is smaller than that shown for zebra finches. We modelled the response to stimuli using Bayesian and maximum likelihood estimation and implemented the psychometric function to estimate the effect size. We found that the result is independent of the psychophysical method used to evaluate the effect of experimental conditions on choice performance.

**KEY WORDS:** Adaptation, Bird vision, Colour vision, Psychometric function, Visual ecology

## INTRODUCTION

For many animals, including primates, birds and insects, colour is an important cue to identify, recognise and evaluate objects, such as mates (e.g. birds: Hill, 1991; Hunt et al., 1999) or food (e.g. primates: Osorio and Vorobyev, 1996; bees: Hempel de Ibarra et al., 2001, 2002; birds: Schaefer et al., 2006). Accordingly, colour vision and its relationship to colouration has been investigated extensively (e.g. Kemp et al., 2015; Renoult et al., 2017; Lind et al., 2017; Cuthill et al., 2017), and ecological studies have frequently used colour vision models (e.g. Vorobyev and Osorio, 1998) to predict discrimination thresholds.

Behavioural studies have tested how well animals can use colour to detect objects against a background (e.g. bees: Hempel de Ibarra et al., 2001; crows: Schaefer et al., 2006) and to discriminate between objects (e.g. bees: Hempel de Ibarra et al., 2002; chickens: Olsson et al., 2015). Colour vision models have also been used to

understand both the discrimination of an object colour from a background colour (e.g. bees: Hempel de Ibarra et al., 2001; primates: Sumner and Mollon, 2000) and discrimination between object colours (e.g. bees: Hempel de Ibarra et al., 2002; chickens: Olsson et al., 2015). The receptor noise limited (RNL) model of colour discrimination builds on the assumption that discrimination thresholds are ultimately limited by photoreceptor noise (Vorobyev and Osorio, 1998) and has been shown to accurately predict absolute colour discrimination thresholds of many species under laboratory conditions (see Renoult et al., 2017; Olsson et al., 2018). The RNL model has also been widely applied to predict colour discrimination in natural contexts, but under these conditions, additional parameters may influence discrimination performance (Olsson et al., 2018). One such parameter is the adaptive state of the visual system, which depends on the colour and intensity of the illumination and the reflectance of the background.

Changes in colour discrimination thresholds caused by changes in illumination colour in bright light can be predicted satisfactorily by the RNL model (e.g. Olsson et al., 2016; Olsson and Kelber, 2017), whereas additional assumptions on photon shot noise, dark noise and spatial pooling have to be considered to predict the performance of the visual system in dim light (e.g. Olsson et al., 2015, 2017).

However, studies on the discrimination of object colours have rarely considered the colour of the adaptive background, often chosen as neutral grey. In natural visual scenes, objects, such as conspecifics and food items, are seen against differently coloured backgrounds, such as green vegetation, the blue sky or sandy ground. Moreover, some animals position themselves against specific backgrounds and under specific illumination during courtship displays, indicating that background colour may influence the perception of plumage presented by a prospective mate (e.g. Endler, 1993; Endler and Mielke, 2005).

For humans, the colour of the background, and thus the adaptive state of the visual system, affects colour discrimination (Krauskopf and Gegenfurtner, 1992; Smith et al., 2000). Humans are better at discriminating red colours presented against a reddish background than against a greenish background, and vice versa. Lind (2016) found a similar effect of background colour on colour discrimination in a bird, the zebra finch (*Taeniopygia guttata*).

This study aimed to determine the effect of the background on colour discrimination in a widely used bird model system, the chicken (*Gallus gallus*). We trained chickens to discriminate green or orange stimulus colours on green or orange backgrounds and compared their discrimination performance. We applied different psychophysical methods to describe both the discrimination threshold and the slope of psychometric curves.

## MATERIALS AND METHODS

### Animals

Thirty-two Lohmann white chickens (*Gallus gallus* L.) of both sexes (Gimranäs AB, Herrljunga, Sweden) were used. We hatched eggs in a commercial incubator (Covatutto 24, Högberga AB,

<sup>1</sup>Department of Biology, Lund University, 223 62 Lund, Sweden. <sup>2</sup>Department of Philosophy, Lund University, 223 62 Lund, Sweden.

\*Present address: Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, VIC 3086, Australia.

‡Author for correspondence (polsn84@gmail.com)

ORCID iD: P.O., 0000-0001-9026-3274; R.D.J., 0000-0002-6526-677X; J.J.F., 0000-0002-4444-2375; J.D.K., 0000-0001-5537-3574; O.L., 0000-0002-5490-4705; A.K., 0000-0003-3937-2808

**Table 1. Batches and experimental groups of chickens used in the experiments**

Batch	Background colour	Rewarded stimulus colour and group		Sex ratios (F:M)
		Group 1	Group 2	
a	green	Group 1	Group 2	Group 1 (5:3) Group 2 (5:3)
b		G+	O+	
c	orange	Group 3	Group 4	Group 3 (4:4) Group 4 (4:4)
d		G+	O+	

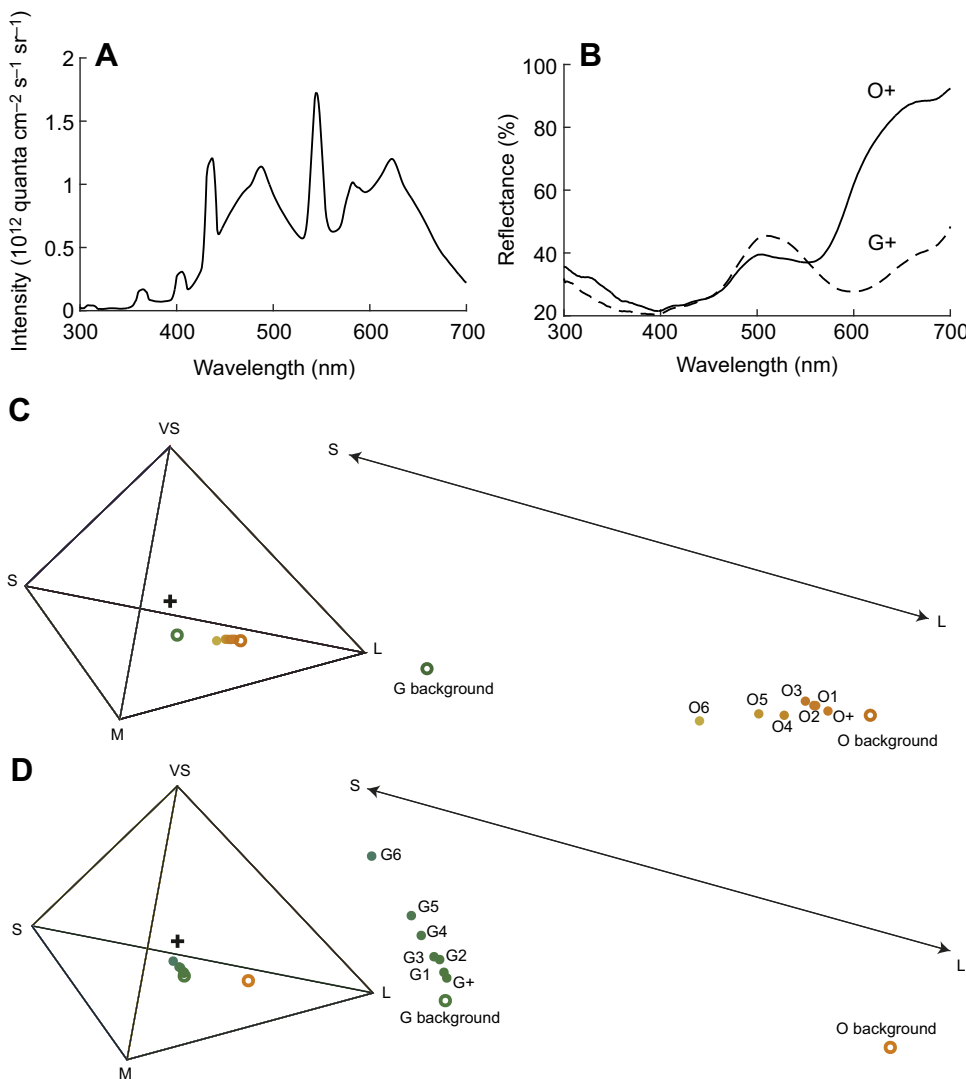
Matfors, Sweden) and housed the chickens following regulations (permit no. M111-14, Swedish Board of Agriculture) in a 1×1 m wooden box with 1.5 m high walls, a mesh on top, a perch and water available *ad libitum*. Food (chick crumbs, Fågel Start, Svenska Foder AB, Staffanstorp, Sweden) was available during experimental sessions and after the last experimental session each day. On days without sessions, food was available *ad libitum*. The experiment ended when the chickens were between 4 and 5 weeks old.

We raised, trained and tested four batches of eight chickens each. Animals of two batches were trained to discriminate colours presented on an orange background, whereas the other two batches were trained with a green background. On the third day after hatching, four animals were randomly assigned to training with

green stimuli, and four animals to training with orange stimuli. This resulted in four experimental groups with eight animals each: group 1 discriminated green stimulus colours on a green background, group 2 discriminated orange stimulus colours on a green background, group 3 discriminated green stimulus colours on an orange background and group 4 discriminated orange stimulus colours on an orange background (Table 1).

**Experimental setup and stimuli**

We conducted all experiments in a wooden arena (0.7×0.4 m) with matte grey walls and floor, illuminated from above by four fluorescent tubes (Biolum 18 W, Osram, Munich, Germany; for the spectrum, see Fig. 1A). We presented stimuli on a uniformly coloured orange or green background that covered part of the floor (25 cm wide×15 cm deep) and the wall (25 cm wide×15 cm high) (Fig. S1). At 30 cm distance (the starting condition), the background subtended 34 deg horizontally and 26 deg vertically. As stimuli, we used conical food containers, folded from paper uniformly printed with green or orange colours, similar to those in previous studies with chickens (Olsson and Kelber, 2017; Olsson et al., 2015, 2016, 2017; Osorio et al., 1999). All colours were created in Adobe Illustrator, Creative Suite package version 5, using CMYK colour coding and printed with a Canon Pro 9000 MkII.



**Fig. 1. Colours used in the study.** (A) The radiance of a white standard illuminated by the experimental illumination. (B) The reflectance of the two rewarded colour stimuli. (C) Chromatic loci of all colour stimuli in the orange stimulus series and both backgrounds. (D) Chromatic loci of all colour stimuli in the green stimulus series and both backgrounds. An enlarged version of all stimuli and backgrounds for easier visualisation is located to the right. The letters at each corner (VS, S, M, L) refer the different cone types and the letters within the space (O, G) refer to individual stimuli.

We used one rewarded (O+) and six unrewarded orange stimuli (O1–6), as well as one rewarded (G+) and six unrewarded green stimuli (G1–6). The achromatic contrasts between all unrewarded stimulus colours and the rewarded stimulus colour were below the achromatic discrimination threshold of chickens (ca. 7%; see Jarvis et al., 2009; Gover et al., 2009). Spectra of backgrounds and colour loci in bird colour space are given in Fig. 1B–D. All spectra are given in Table S3.

### Visual modelling

The colour differences between stimuli, and between stimuli and backgrounds (Table 2), were calculated using the RNL model (Vorobyev and Osorio, 1998) with units of just noticeable difference (JND), and the colour discrimination threshold set as 1 JND. We have previously calibrated the noise assumptions of this model to the behavioural colour discrimination threshold of chickens (Olsson et al., 2015). The calculations are based on these established methods (Vorobyev et al., 1998; Olsson et al., 2015).

Colour differences,  $\Delta S$ , were calculated as:

$$\Delta S^2 = (\omega_1 \omega_2)^2 (\Delta f_4 - \Delta f_3)^2 + (\omega_1 \omega_3)^2 (\Delta f_4 - \Delta f_2)^2 + (\omega_1 \omega_4)^2 (\Delta f_3 - \Delta f_2)^2 + \frac{(\omega_2 \omega_3)^2 (\Delta f_4 - \Delta f_1)^2 + (\omega_2 \omega_4)^2 (\Delta f_3 - \Delta f_1)^2 + (\omega_3 \omega_4)^2 (\Delta f_2 - \Delta f_1)^2}{(\omega_1 \omega_2 \omega_3)^2 + (\omega_1 \omega_2 \omega_4)^2 + (\omega_1 \omega_3 \omega_4)^2 + (\omega_2 \omega_3 \omega_4)^2}, \quad (1)$$

where  $\omega$  refers to noise in the visual channel of cone type  $i$  (VS:S:M:L), set to  $0.06 \omega_L$ , and is calculated as:

$$\omega_i = \frac{\sigma}{\sqrt{\eta_i}}, \quad (2)$$

where  $\sigma$  is the coefficient of variation of noise and  $\eta_i$  is the relative abundance of cone type  $i$  in the retina. In chickens, the relative abundance of VS:S:M:L cone types has been estimated as 1:1.5:2.5:2 (Kram et al., 2010). The contrast,  $\Delta f_i$ , in receptor channel  $i$ , is calculated as:

$$\Delta f_i = \ln \left( \frac{Q_{i,\text{stimulus1}}}{Q_{i,\text{stimulus2}}} \right), \quad (3)$$

**Table 2. Colour distances between stimuli and backgrounds, and between the unrewarded stimuli and the rewarded stimulus**

	Colour difference to green background (JND)	Colour difference to orange background (JND)	Distance to G+ and O+ (JND)
G+	1.0	13.9	0
G1	1.2	14.0	0.3
G2	1.7	14.1	0.6
G3	1.7	14.3	1.0
G4	2.5	14.8	1.8
G5	3.3	15.2	2.5
G6	5.6	16.8	5.0
O+	12.5	1.3	0
O1	12.1	1.7	0.3
O2	12.1	1.7	0.7
O3	11.8	2.0	1
O4	11.1	2.7	1.6
O5	10.4	3.5	2.4
O6	8.6	5.4	4.3

JND, just noticeable difference.

where  $Q$  refers to the quantum catch of the receptor type  $i$  and is calculated as:

$$Q_i = \int_{300}^{700} R_i(\lambda) S(\lambda) d\lambda. \quad (4)$$

Here,  $R$  refers to the spectral sensitivity of cone type  $i$  and  $S$  refers to the radiance of the stimulus. The radiance of the stimuli was measured with a spectroradiometer (RSP 700, International Light, Peabody, MA, USA) under the experimental illumination.

Colour loci as represented in Fig. 1C,D were calculated using the functions given in Kelber et al. (2003), for the colour tetrahedron of tetrachromats.

The achromatic differences between colours were calculated as Michelson's contrast ( $C$ ) for the double cone, which is assumed to mediate the achromatic vision of birds in bright light (Osorio and Vorobyev, 2005; Lind et al., 2013), as:

$$C = \frac{Q_{i,\text{stimulus1}} - Q_{i,\text{stimulus2}}}{Q_{i,\text{stimulus1}} + Q_{i,\text{stimulus2}}}. \quad (5)$$

### Training and testing procedure

Training of the chickens started on the third day after hatching. Chickens were assigned to one of four groups (see Table 1) and trained in two sessions each day, one before and one following noon, with at least 1 h between the two sessions. On the first training day, four chickens were trained together, with several food containers of the rewarded stimulus colour (orange O+ or green G+) at the same time, each filled with several chicken crumbs and presented on the green or orange background, depending on the group. On the second day, chickens were trained in pairs with only one filled food container of the rewarded stimulus colour available at any time. On the third day, each pair of chickens was initially placed behind a grey cardboard wall and given access to the food container only after the wall was removed. On the fourth day, each chicken was trained individually, as per day 3. A second chicken was placed in an adjacent cage separated from the experimental cage by mesh, allowing for visual and audio contact with the experimental chicken to reduce stress. On the fifth day, an empty food container of a distinctly different unrewarded stimulus colour (either O6 or G6) was introduced. From this day on, each session consisted of 20 trials. Testing started when a chicken had reached the learning criterion of 0.75 correct choices in two consecutive sessions.

In tests, we presented chickens with one filled food container of the rewarded stimulus colour (G+ or O+) and one food container of an unrewarded stimulus colour on a green background (groups 1 and 2) or an orange background (groups 3 and 4; Table 2). If the chicken pecked the rewarded container, it was allowed to feed on any spilled food, and the unrewarded container was removed. By rewarding the chickens continuously for correct choices, we ensured high motivation throughout the experiment. If the chicken pecked the unrewarded container, both containers were removed. Testing started with the largest colour difference, O+ against O6 and G+ against G6, and continued with smaller colour differences between stimuli, ending with O1 or G1. We presented each colour difference in four sessions of 20 trials each, two training sessions, and two test sessions, during which we analysed the choices ( $N=40$  choices for each chicken and stimulus pair). A proportion of 0.65 correct choices differs significantly from chance in binomial tests with  $\alpha=0.05$  ( $N=40$ , one-tailed binomial test). Chickens that failed to reach 0.65 correct choices with a specific unrewarded stimulus

colour were excluded from subsequent tests with unrewarded stimulus colours that were more similar to the rewarded stimulus colour; thus, a chicken made a maximum of 240 choices, if tested with all six unrewarded stimuli.

### Data analysis

We analysed data from each bird separately by calculating the correct choice frequency for the last 40 choices with each stimulus pair (one rewarded and one unrewarded stimulus colour) as a function of the colour difference between the two colours. We used three different methods to analyse the results: (i) generalised linear modelling using maximum likelihood estimation (MLE), incorporating the psychometric function (psychometric MLE), (ii) generalised linear mixed modelling using Bayesian estimation, incorporating the psychometric function (psychometric Bayesian), and (iii) generalised linear mixed modelling using MLE, excluding the psychometric function (non-psychometric MLE). The last two methods apply mixed models by estimating random as well as fixed effects. Discrimination thresholds were estimated from the fitted psychometric functions. We used two thresholds: (i) the proportion of correct choices that differs significantly from chance according to binomial statistics (0.65), to compare with previous studies; and (ii) the proportion of correct choices at the inflection point of the psychometric function.

For the psychometric MLE method (method i), the dependent variable was choice performance, the proportion of correct choices,

and the independent variables were colour difference (between rewarded and unrewarded stimuli), background colour, sex, batch, and individual identity. We fitted models of the psychometric functions using the *glm.WH* function in the *psyphy* package (<https://cran.r-project.org/web/packages/psyphy/>) in R ([www.r-project.org](http://www.r-project.org)). These used MLE to fit a logistic regression model, via the probit link, for two alternative forced choice (2AFC) experiments; an upper asymptote was estimated and a lower asymptote at 0.5 was specified. For model selection, for both methods i (psychometric MLE) and iii (non-psychometric MLE), we used likelihood ratio tests on nested models via the *anova* command in R, starting from a model including a 3-way interaction between colour difference, background colour and sex with additional fixed effects of batch and individual identity, and looked for significant differences in deviance when dropping variables and their higher order interactions. We preferred models with significantly lower deviance and lower AIC. Additionally, models with reduced complexity that did not significantly reduce deviance were also preferred (Tables 3–5; Tables S1 and S2). Whenever a variable or an interaction was removed in this way, we updated the model so that further comparisons were always between the current, updated, model and the proposed model with further variable or interaction reduction. In essence, we performed a backwards model selection procedure. When we arrived at a final model, with only variables that were preferred to remain in the model, we successively added each dropped variable and interaction, individually, back to new

**Table 3. Model selection results for the psychometric MLE method (method i) comparing models explaining the performance of chickens discriminating green colours**

Term(s)	AIC	$\Delta$ Deviance	$\Delta$ d.f.	P-value	Rank status
a. Comparison with a null model					
Discrimination performance~ Colour.difference:background colour: sex +individual identity +batch	442.35				Deficient 2×Ind Batch B–D Background:sex
Discrimination performance~1	1090.69	686.34	19	<2.2e–16	Full
b. Model selection					
Colour.difference:background colour: sex	442.72	2.37	1	0.123	Deficient 2×Ind Background:sex
Background colour:sex	442.72	0	0		Deficient 2×Ind Batch B–D
Background colour:colour difference	441.59	0.86	1	0.354	Deficient 2×Ind Batch B–D
Colour.difference:sex	439.73	0.14	1	0.71	Deficient 2×Ind Batch B–D
Batch	439.73	0	0	0	Deficient 2×Ind
Sex	439.73	0	0	0	Deficient 1×Ind
Background colour	439.73	0	0	0	Full
<b>Individual identity</b> (retained)	515.87	106.14	15	8.84e–16	Full
<b>Colour.difference</b> (retained)	1029.19	712.55	1	<2.2e–16	Full
c. Final model					
Discrimination performance~Colour.difference +individual identity	439.73				

The psychometric MLE method (method i), compares models explaining the performance of chickens discriminating green colours. (a) Likelihood ratio test comparing the full model with a null model. (b) Model selection, using likelihood ratio tests to compare models with and without each of the fixed effects terms specified. (c) Formula for the model chosen, including only the terms that were retained. Deviance and d.f. refer to residual deviance and residual degrees of freedom. Terms that were retained in the final model are indicated in bold. Rank status reports whether the model was rank deficient or not and specifies term(s) that were not estimated. Note that in some cases the removal of a variable did not change degrees of freedom, which are related to changes in rank status.



proposed models and compared those with the final model. The changes in deviance reported in Tables 3–5 are from the final comparison. However, for the psychometric MLE method, the models where the interactions colour difference:background, and the 3-way interaction in the green colour series (Table 3) and the 3-way interaction in the orange colour series (Table 4) were added back in could not be fitted. The reported deviance change for those comparisons are from the backwards selection procedure. To estimate thresholds, we fitted the final models to each group and used the *predict* function to generate psychometric functions along with confidence intervals and used the *approx* command to find the threshold colour difference.

We calculated the slope of each psychometric function at threshold (0.65) by fitting a linear function to  $x$  and  $y$  values just above and below threshold (by a proportion of 0.01) and looked for significant interaction effects, between stimulus colour difference (continuous variable) and background colour (categorical variable), indicating differences in the slopes of the psychometric functions.

It should be noted that psychometric MLE models including batch and individual identity as random effects did not converge on appropriate parameter values (possibly because of the large number of higher order interactions). We therefore chose to model batch and individual identity as fixed effects (i.e. categorical variables with intercepts and slope coefficients) rather than random effects (i.e. correlated group-level intercepts and slopes scaled by a population-level standard deviation). For categorical variables with many levels, these two methods may yield similar results. One crucial difference is that some predictive power could be gained from assigning random effects to variables for which it was impossible to reasonably observe all possible levels in a single experiment, such

as individual identity and batch. This reduces their influence on the estimates of other fixed effects, improving predictions for future experiments (Henderson, 1982) with different individuals and batches.

For the psychometric Bayesian method (method ii), we applied a probabilistic model to estimate discrimination thresholds, using the Stan language (Carpenter et al., 2017; <http://mc-stan.org/>) via the *brms* package (v2.6.0; Bürkner, 2018; <https://cran.r-project.org/web/packages/brms/>) within R. Here, we used a logistic regression model incorporating the psychometric function (see Kirwan and Nilsson, 2019), with success rate for random guessing (0.5) as the estimated lower asymptote and the lapse rate, found in tests with the unrewarded colours O6 and G6 in each experiment, as the upper asymptote. We reparameterised the psychometric function to directly assess the effects of conditions on the curve's inflection point (threshold  $T_{ip}$ ) and the range of colour differences that account for 80% of the response range [threshold ( $m$ ), width ( $w$ ): the ' $m,w$ ' parameterisation (Kuss et al., 2005; Houpt and Bittner, 2018; also known as 'threshold, support': Alcalá-Quintana and García-Pérez, 2004).

The fixed (population-level) effects accounted for in this model were colour difference, background colour, whether background and stimulus were of the same colour type (e.g. green stimuli on a green background) or different, sex of the individual tested, and their higher order interactions. Individual identity and batch were included as random (group-level) effects, permitting unique thresholds, threshold-width and lapse rates for each individual and batch.

We applied informative priors for threshold, width and lapse rate (to restrict estimates to the range of colour differences sampled) but applied a specific informative prior to keep the lower asymptote near 0.5. Using the fitted psychometric function, an additional threshold

**Table 4. Model selection results for the psychometric MLE method (method i) comparing models explaining the performance of chickens discriminating orange colours**

Term(s)	AIC	$\Delta$ Deviance	$\Delta$ d.f.	P-value	Rank status
a. Comparison with a null model					
<i>Fixed effects model</i>	381.21				Deficient
Discrimination performance~Colour.difference:background colour:sex +individual identity +batch					2×Ind batch b background:_sex
<i>Null model</i>	972.52	629.33	19	<2.2e−16	Full
Discrimination performance~1					
b. Model selection					
Colour.difference:background colour:sex	379.34	0.14	1	0.71	Deficient 2×Ind background:_sex
Background colour:sex	379.34	0	0		Deficient 2×Ind
Colour.difference:sex	377.42	0.077	1	0.78	Deficient 2×Ind
<b>Colour.difference:background</b> (retained)	410.24	34.82	1	3.61e−09	Deficient 2×Ind
Batch	377.42	0	0		Deficient 2×Ind
Sex	377.42	0	0		Deficient 1×Ind
<b>Individual identity</b> (retained)	393.17	41.83	13	0.0001	Full
c. Final model					
Discrimination performance~Colour.difference:background +individual identity	377.42				

(a) Likelihood ratio test comparing the full model with a null model. (b) Model selection, using likelihood ratio tests to compare models with and without each of the fixed effects terms specified. (c) Formula for the model chosen, including only the terms that were retained. Deviance and d.f. refer to residual deviance and residual degrees of freedom. Terms that were retained in the final model are indicated in bold. Rank status reports whether the model was rank deficient or not and specifies term(s) that were not estimated. Note that in some cases the removal of a variable did not change degrees of freedom, which are related to changes in rank status.

**Table 5. Model selection for the non-psychometric MLE model (method iii)**

Term(s)	$\Delta$ Deviance	$\Delta$ d.f.	P-value
<b>a. Comparison with a null model</b>			
<i>Fixed effects model</i>			
Discrimination performance ~ colour difference +background +same +sex +colour difference:background +colour difference: same +background: same +colour difference:sex +background:sex +same:sex +colour difference:background: same +colour difference:background:sex +colour difference: same:sex +background: same:sex +colour difference:background:same:sex +(1+colour difference   individual) +(1+colour difference * background   batch)			
<i>Null model</i>	991.61	28	<2.2e-16
Discrimination performance~1 +(1   individual) +(1   batch)			
<b>b. Model selection</b>			
Colour.difference:background:same:sex	10.8325	13	0.6248
Colour.difference:same:sex	5.9578	5	0.3103
Colour.difference:background:sex	1.9502	6	0.9242
Colour.difference:background:same	6.2955	5	0.2785
Background:same:sex	6.0654	6	0.4159
Colour.difference:sex	0.4841	2	0.7850
Background:sex	1.0469	3	0.7899
Colour.difference:background	0.6674	2	0.7163
Background:same	1.9928	2	0.3692
Colour.difference:same	3.4083	1	0.06487
Same:sex	2.3527	2	0.3084
Background	0.0292	1	0.8644
Sex	0.3568	1	0.5503
<b>Same</b> (retained)	14.768	1	0.0001
<b>Colour.difference</b> (retained)	961.17	1	<0.0001
<b>c. Final model</b>			
Discrimination performance~Colour.difference +same +(1+colour difference   individual) +(1+colour difference * background   batch)			

(a) Likelihood ratio test comparing the fixed effects model with the null model (assuming choice performance was based on random individual and batch variation). (b) Model selection, using likelihood ratio tests to compare models with and without each of the fixed effects terms specified. (c) Formula for the model chosen, including only the terms that were retained. NB Because different individuals did not experience different backgrounds, random effects of individual on the 'background' coefficient were not modelled. Terms that were retained in the final model are indicated in bold.

was calculated from the point at which correct-choice rates were modelled at 0.65, as for the psychometric MLE method (method i).

The 'threshold, width' model estimated had the form:

$$\psi = \gamma + (1 - \lambda - \gamma) \times \text{logistic}\left(\frac{2 \ln((1/\alpha) - 1)(x - m)}{w}\right), \quad (6)$$

where  $\psi$  is the predicted proportion of correct choices for colour difference  $x$ , given a threshold at colour difference  $m$  with a width of  $w$ , also in units of colour difference. Changes in the proportion of correct choices occur in the range between the proportion  $\lambda$ , expected for random guessing, and  $(1-\lambda)$ , the maximum possible proportion of correct choices given a lapse rate  $\lambda$ . The auxiliary parameter  $\alpha$  scales threshold-width  $w$ , so that it covers  $1-2\alpha$  of the range of  $x$  for which

the response curve is non-asymptotic (Kuss et al., 2005). Following Houpt and Bittner (2018), we chose  $\alpha=0.1$ , so that  $x \pm (w/2)$  gives the range of colour differences for which predicted success rate is between the lower 10% and top 10% of the guess rate–maximum success rate interval  $[\lambda, (1-\lambda)]$ , across which 80% of changes in the proportion of correct responses would be predicted.

It should be noted that this is equivalent to the more commonly used parameterisation:

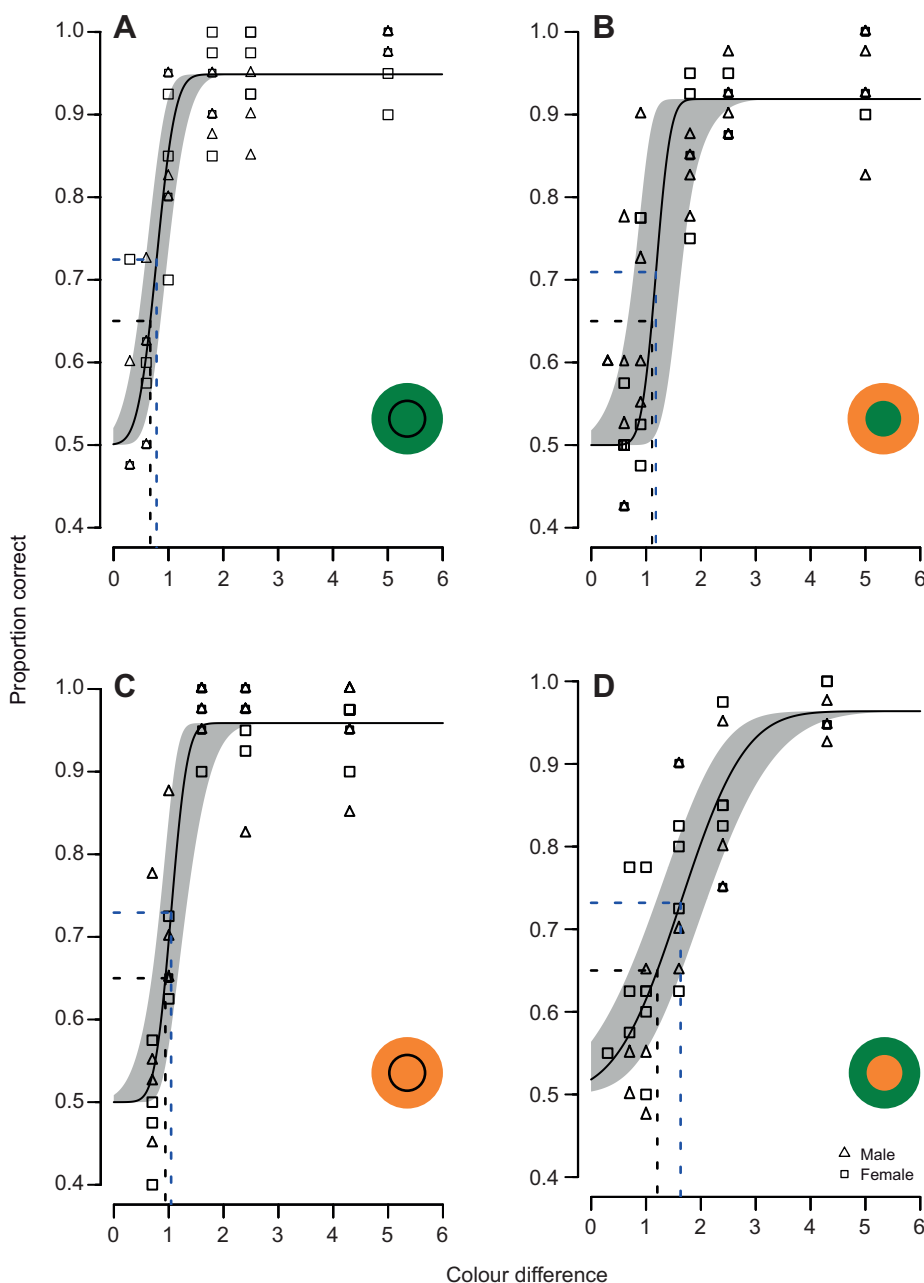
$$\begin{aligned} \psi &= \gamma + (1 - \lambda - \gamma) \times \text{logistic}(bx + c) \\ &= \gamma + \frac{(1 - \lambda - \gamma)}{1 + e^{-(bx+c)}}, \end{aligned} \quad (7)$$

in which  $b$  and  $c$  are the logistic curve's slope and intercept in logit

space, respectively. An advantage of fitting the model in the form of Eqn 6, as opposed to Eqn 7, is that the parameters can be directly interpreted as properties of the observer's response threshold, and are comparable with the parameters of thresholds obtained with other commonly used response functions (e.g. probit, Gumbel and Weibull: Kuss et al., 2005).

We chose to account for individual differences in psychometric curves by allowing for different threshold, width and lapse rate estimates for each individual and batch, producing a mixed-effects model. As Eqns 6 and 7 include the terms both within and outside of the logistic transform, neither  $\psi$  nor  $\text{logit}(\psi)$  change linearly as a function of  $x$ . Instead, changes in the rate of correct choices are non-linear with respect to  $x$ . We therefore used the non-linear modelling function in the *brms* package (Bürkner, 2018) in R ([https://cran.r-project.org/web/packages/brms/vignettes/brms\\_nonlinear.html](https://cran.r-project.org/web/packages/brms/vignettes/brms_nonlinear.html)) to fit mixed-effects psychometric models.

In order to restrict estimates of threshold and width to positive numbers (colour differences  $>0$ ) and produce estimates of lapse rate between zero and one, these parameters were estimated as the natural logarithms of threshold and width, and as the logit transform  $\{\ln[x/(1-x)]\}$  of lapse rate. An informative prior distribution of  $\text{Normal}(0,1)$  was chosen for  $\ln(\text{threshold})$ ,  $\ln(\text{width})$  and their fixed effects coefficients, maintaining 95% of prior probability density on estimates between 0.14 and 7.10. A bounded prior distribution of  $\text{Normal}(-3,10)$ , with an upper bound at  $-1$ , was chosen for  $\text{logit}(\text{lapse rate})$ , to exclude only lapse rates smaller than  $10^{-10}$  and greater than 0.27. Such large lapse rates would suggest a maximum proportion correct of 0.73, below the learning criterion required for an animal to qualify for these experiments. Indeed, maximum proportion correct was above 0.80 for all test animals (Fig. 2). While such restrictions may not be necessary in cases where lapse rates can credibly reach 0.27 and above, we recommend prior distributions for



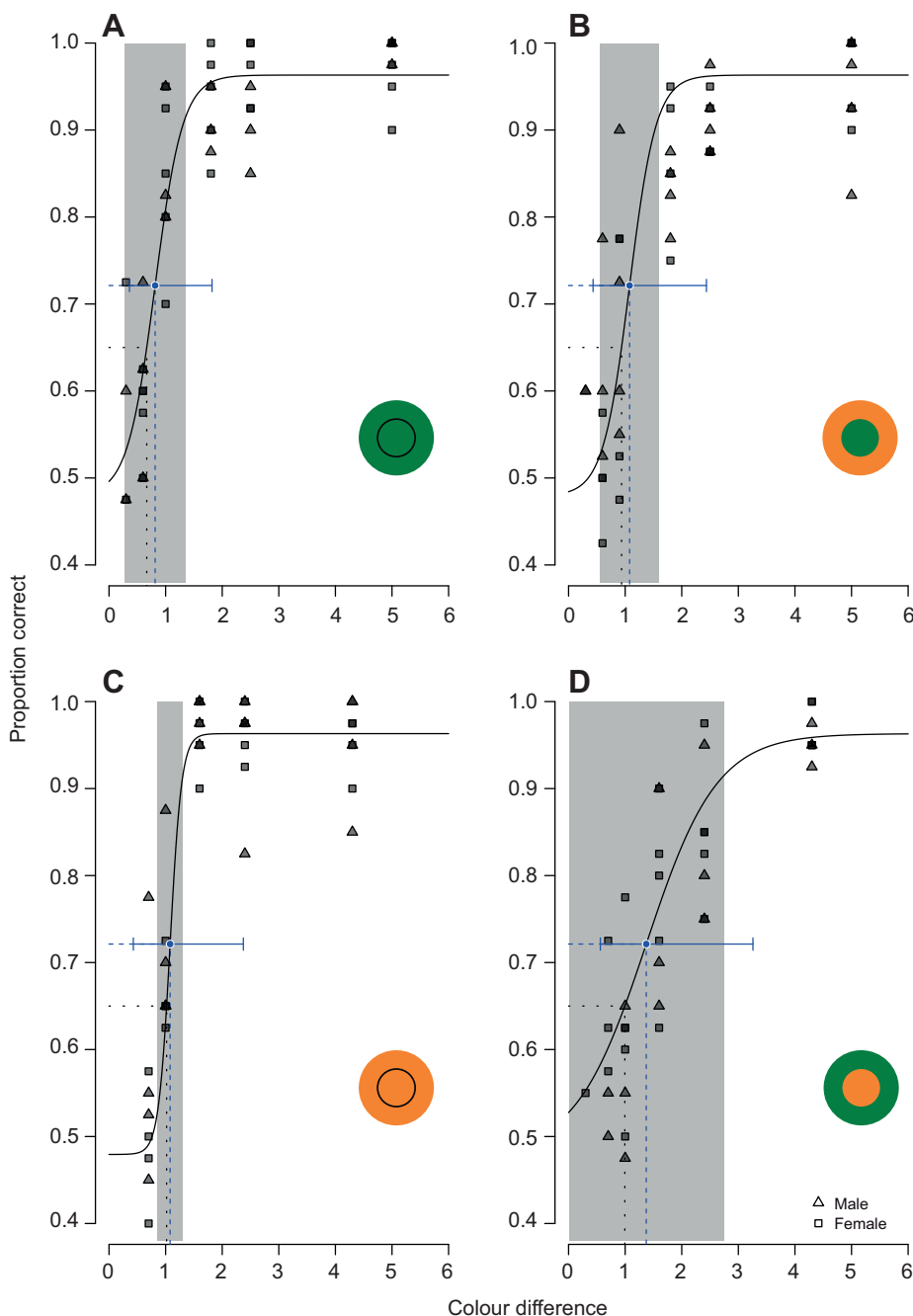
**Fig. 2. Colour discrimination performance as a function of colour difference to the rewarded colour stimulus.** The solid line in each panel shows the maximum likelihood estimation (MLE) psychometric model (method i), and the shaded area gives the confidence interval of the fit. The dotted lines give the estimated threshold (0.65 correct choices). The colour inset in each figure shows the experimental condition: the border is the background and the centre is the stimulus series colour. (A) Group 1, discriminating green colours on the green background. (B) Group 3, green colours on the orange background. (C) Group 4, orange colours on the orange background. (D) Group 2, orange colours on the green background. Each group had 8 chickens; each circle or square represents the proportion of correct choices ( $N=40$ ) of one chicken; note that some points lie behind the others.

logit(lapse rate) that are bounded near the limit for credible lapse rates, where known. This helps to avoid positive feedback loops of increasing lapse rates and widths, approaching a flat curve, or a reversal of the positions of the lapse rate and guess rate (when lapse rate > 1 – guess rate) during estimation. The default prior distribution for the standard deviation of random effects in *brms* (Bürkner, 2018) was used for effects of individual identity on ln(threshold), ln(width) and logit(lapse): a half Student *t* distribution with 3 degrees of freedom, a mean at 0 and a standard deviation of 10. A specific bounded prior distribution of Beta( $\alpha=250$ ,  $\beta=250$ ) was applied to guess rate, with a lower bound at 0.25 and an upper bound at 0.75. This generally excluded guess rate estimates outside of the range from 0.45 to 0.55, and avoided a reversal of the positions of the lapse rate and guess rate during estimation, while permitting slightly higher rates of correct guessing in the control condition than

might be expected. In cases where higher rates of correct guesses are more credible, a less specific, unbounded prior distribution could be applied.

Parameter estimation used four Markov–Chain Monte Carlo (MCMC) chains, each including 5000 ‘warmup’ iterations and generating 5000 ‘post-warmup’ samples. The chains converged well for all parameters (potential scale reduction statistic: Rhat=1.00), producing effective sample sizes that represented more than 20% of ‘post-warmup’ samples in all cases. For further predictive diagnostic checks and cross-validation see Fig. S4.

For the non-psychometric MLE method (method iii), which estimates random effects, we applied the methods used by Olsson et al. (2016), and Olsson and Kelber (2017), using a mixed logistic regression model, without incorporating the psychometric function. This model was chosen specifically to allow for direct comparison



**Fig. 3. Bayesian model predictions for performance as a function of colour difference.**

The solid line in each panel shows the estimated psychometric model, the blue points indicate the estimated thresholds and the shaded area gives the ‘width’ for each threshold. The black dotted lines give the estimated threshold for 0.65 correct choices, whereas the blue dashed lines give the inflection-point threshold modelled. The blue error bars indicate the 95% credible interval for each estimate of threshold, representing the bounds containing 95% of model estimates generated during Bayesian estimation. The colour inset in each figure shows the experimental condition as in Fig. 2. (A) Group 1, green stimuli, green background. (B) Group 3, green stimuli, orange background. (C) Group 4, orange stimuli, orange background. (D) Group 2, orange stimuli, green background. Each group had 8 chickens; each circle or square represents the proportion of correct choices ( $N=40$ ) of one chicken; note that some points lie behind the others.



with the previous studies (Olsson et al., 2016; Olsson and Kelber, 2018). We used the *lme4* package (<https://cran.r-project.org/web/packages/lme4/>) in R (see Olsson et al., 2016, 2017). These models tested choice performance as a function of the colour difference between stimuli. As per the psychometric Bayesian method (method ii), fixed effects of colour difference, background colour, similarity of background and stimulus colour type, sex and their higher order interactions were accounted for, and individuals and batches were included as random effects, with different intercepts and slopes.

## RESULTS

All chickens were highly motivated throughout the experiment and learned to discriminate the rewarded stimulus colour (O+ or G+) from the most different unrewarded stimulus colour (G6 or O6) on both backgrounds. In all experimental groups, choice performance differed depending on the colour difference between the unrewarded and the rewarded stimulus colour (Figs 2 and 3).

### Psychometric MLE method

For this method, model selection, comparing the deviance and AIC of fitted models, suggested a model which included the variables colour difference between stimuli, background colour, individual identity and an interaction between colour difference and background colour, for the orange stimuli, and a model including the variables colour difference and individual identity for the green stimuli (Tables 3 and 4). The inclusion of an interaction between colour difference and background indicates that the slope of the psychometric function depends on the background colour, with steeper slopes when stimulus and background colours were similar (orange stimuli on orange background) and shallower slopes when stimuli and background were very different (orange stimuli on green background). However, some models converged poorly with this method, which hampered accurate estimation of the effect of some variables, and several models were rank deficient, mainly concerning estimates for batch and a few individual chickens (Tables 3 and 4). Model summaries can be found in Tables S1 and S2.

The predictions of the selected models for the four individual groups are compared in Fig. 2. These models included the variables colour difference between stimuli and individual identity. The thresholds (set at 0.65 correct choices) of chickens discriminating orange stimuli were 0.94 JND (credible interval, CI 0.69–1.16 JND) on the orange background, and 1.21 JND (CI 0.70–1.65 JND) on the green background, and the slopes of the psychometric functions at threshold were 0.78 and 0.20, with the orange and green background, respectively (Fig. 2A,B, Table 6). For chickens discriminating green stimuli, we found thresholds of 1.11 JND (0.66–1.54 JND) with the orange background and 0.67 JND (0.45–0.86 JND) with the green background (Fig. 2C,D, Table 6) and the respective slopes were 0.10 and 0.66.

When using the inflection point as threshold estimate, the following thresholds (and error estimates) were obtained; 1.04 (0.83–1.31) JND for chickens discriminating orange stimuli on orange background and 1.63 (1.17–2.09) JND on green background, and 1.18 (0.78–1.64) JND for chickens discriminating green stimuli on orange background and 0.79 (0.59–0.99) JND on green background (Table 6).

### Psychometric Bayesian method

The results using the Psychometric Bayesian method (Fig. 3; Fig. S2) were similar but expressed greater uncertainty regarding the shallow slopes of the psychometric functions observed in both experiments, in which stimulus colours and background colour were very different. For this method, we used the inflection point of the psychometric function as a threshold estimate. As these values are more strongly influenced by the function's intercept, the variations in performance for below-threshold conditions have a strong influence on the fitted intercept and hence the estimated threshold. Grey shaded regions indicate estimated threshold-widths (80% of the rising region of the curve), the region within which the threshold would be expected to occur. The widths of these regions are inversely proportional to the slope.

The CIs shown as blue error bars in Fig. 3 indicate the distributions of model estimates generated during Bayesian estimation, illustrating the robustness of the estimate. The narrower CIs for the experiments in which stimulus and background colours were of the same type (in comparison to those with differing type) indicate lower uncertainty in these threshold estimates. Thresholds, estimated at the inflection point, were 1.07 JND (CI 0.84–1.37 JND) for orange stimuli against an orange background, 1.36 JND (CI 1.04–1.76 JND) for orange stimuli on a green background, 1.09 JND (CI 0.83–1.43 JND) for green stimuli against an orange background and 0.81 JND (CI 0.62–1.05 JND) for green stimuli against a green background. Posterior distributions for thresholds and threshold-widths in each condition and fitted curves and thresholds for each individual are available in Figs S2 and S3. Posterior predictive checks, which compared the observed data with simulated data (derived from model predictions) accurately recovered the count (trial success and failure) proportions, supporting the validity of this model (Fig. S4).

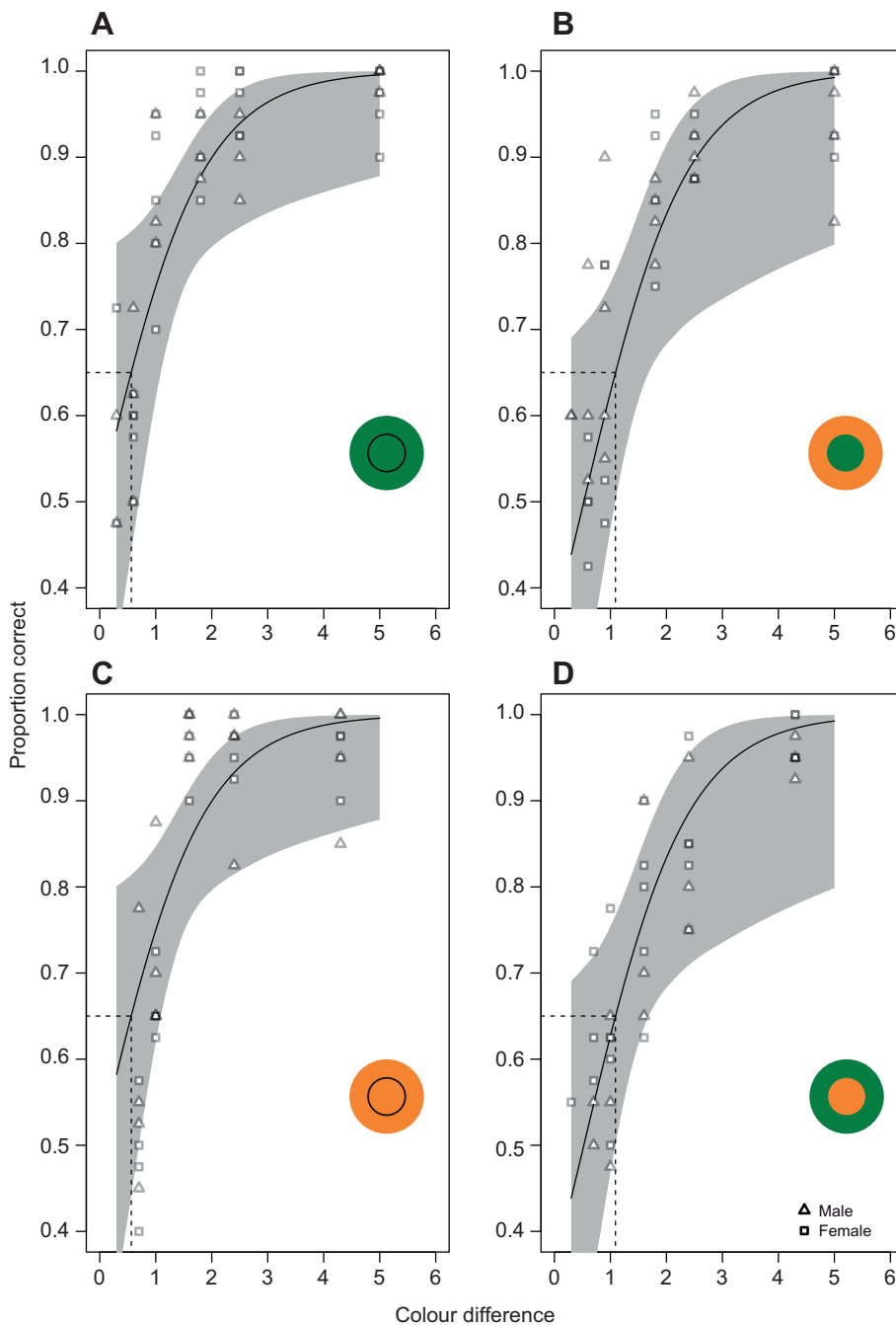
### Non-psychometric MLE method

The non-psychometric MLE model found significant effects of the difference between stimulus colours (colour difference) and the match between stimulus and background colour (same), but no independent effects of background colour or sex on correct choice rates, for both green and orange stimuli (Table 5). The model including colour difference as a fixed effect had lower AIC scores and significantly lower deviance than the null models, which

**Table 6. The estimated discrimination thresholds**

Method	Threshold estimate (JND)			
	Group 1 (Green on green)	Group 2 (Green on orange)	Group 3 (Orange on green)	Group 4 (Orange on orange)
Psychometric MLE (i) Binomial (65%)	0.67 (0.45–0.86)	1.11 (0.66–1.54)	1.21 (0.70–1.65)	0.94 (0.69–1.16)
Psychometric Bayesian (ii) Binomial (65%)	0.66 (0.53–0.81)	0.96 (0.77–1.15)	0.98 (0.80–1.17)	1.01 (0.80–1.26)
Non-psychometric MLE (iii) Binomial (65%)	0.57 (0.30–0.88)	1.09 (0.74–1.36)	1.09 (0.74–1.36)	0.57 (0.30–0.88)
Psychometric MLE (i) Inflection point	0.79 (0.59–0.99)	1.18 (0.78–1.64)	1.63 (1.17–2.09)	1.04 (0.83–1.31)
Psychometric Bayesian (ii) Inflection point	0.81 (0.62–1.05)	1.09 (0.83–1.43)	1.36 (1.04–1.76)	1.07 (0.84–1.37)

Discrimination thresholds estimated for each experiment and for each method, both at the level just above chance according to binomial statistics as well as the inflection point, in the different experimental conditions from different psychometric methods. Error estimates correspond to confidence intervals.



**Fig. 4. Results with non-psychometric MLE models.** The models include individual and batch variation as random effects and colour difference, background, stimulus similarity, sex and their interactions as fixed effects as predictors of choice performance. Lines present 0.65 correct choices, the shaded area is the confidence interval of the estimate, and dots present average choices for the chickens ( $n=8$ , making 40 choices each). For details of the methods used, please see Olsson et al. (2016, 2017).

assumes that choice performance was an effect of only individual and batch variance (Table 5). The resulting model fits can be found in Fig. 4.

## DISCUSSION

In our experiments, we found that the colour discrimination performance of chickens depends on background colour. Chickens had higher discrimination thresholds when the background colour differed strongly from the stimulus colours, in keeping with earlier results from zebra finches (Lind, 2016) and humans (Krauskopf and Gegenfurtner, 1992; Smith et al., 2000). Discriminability differed less in our experiments than for zebra finches (Lind, 2016); additionally, one analysis method (non-psychometric MLE), out of three, suggested that chickens discriminating green colours were not affected by the background

colour. Our finding that the slopes of the psychometric functions differed depending on background colour (Figs 2 and 3) agrees with the data in fig. 5 of Lind (2016). The difference in the size of the effect between the two studies may be due to species differences, but different experimental conditions may also have contributed. In the previous study, the zebra finches saw stimuli and backgrounds on a computer screen, while in the current study, chickens saw three-dimensional stimuli. Moreover, fewer animals were tested in the zebra finch study, and the psychometric functions fitted to the choice data had shallower slopes even in the control conditions.

All statistical models used to analyse behavioural performance agreed that the similarity of stimulus colours to the background colour affected choice performance, with one exception, which suggests that this conclusion is robust. The psychometric MLE method has the benefit of being easy to use. However, the method

had problems estimating effects of all variables and properly identifying some models. In addition, fixed effects models, such as this one, treat errors within subjects equally to errors between subjects and may produce invalid standard errors of parameter estimates (Moscatelli et al., 2012). The non-psychometric MLE method includes random effects, which may correct for this problem. However, functions fitted for the population do not fit the data very well and are poor predictors for future performance, as they do not estimate the necessary parameters to fit a psychometric function, although we do not exclude this possibility. Therefore, it is also difficult to estimate the threshold of performance from the fit. The psychometric Bayesian approach resolves these problems. It considers individual identity and models a psychometric function from which thresholds can be estimated for the population. In addition, Bayesian statistics treat probability in a more intuitive manner than frequentist statistics and provide a better framework to test hypotheses (O'Hagan, 2004). Furthermore, the convergence, model identification and rank deficiency problems of the psychometric MLE method were not found in the Bayesian approach. This may indicate overfitting in psychometric MLE models, in which individual variation was treated as a fixed effect rather than a random effect, so it is possible that psychometric Bayesian models provide a better representation of the predictions that can be drawn from the data, given the spread of individual data points. The Bayesian approach requires a prior – a probabilistic specification of the parameter, aside from the data – that can influence the (posterior) estimate of that parameter. We propose that in comparable cases, the psychometric MLE method could obtain a first estimate for prior specification on thresholds across the population, to apply the psychometric Bayesian method to perform the analysis, as it is shown to better estimate variables when data are limited.

The differences in the slopes of the psychometric functions are striking. In human psychophysics, the slope of a psychometric function measures the certainty of the subject (e.g. Olkkonen and Allred, 2014), and a shallower slope reflects greater uncertainty. Krauskopf and Gegenfurtner (1992) and Smith et al. (2000) did not report individual psychometric functions, but a similar effect is known from achromatic vision. In tests with high achromatic contrasts between stimuli and background, psychometric functions have shallower slopes than in tests with low contrasts (e.g. Wallis et al., 2013). To our knowledge, the slopes of psychometric curves have rarely been analysed from other species. We suggest that, as for humans, they may indicate the certainty of the subject and impact visually guided behaviours, such as detecting a food item or assessing mate quality.

The dependence of the slope on background colour could be a direct result of the sigmoid photoreceptor response properties ( $V$ -log  $I$ -curves; Lind, 2016). Photoreceptors adapt to background intensity (Boynton and Whitten, 1970; Normann and Werblin, 1974), and their contrast sensitivity – and thus, certainty – is greatest in a narrow range centred on background intensity. High contrast against the background reduces the contrast sensitivity of each receptor channel, such that uncertainty is higher for the detection of small contrasts between stimuli. An additional mechanism that could explain the observed differences is simultaneous colour contrast, which builds on lateral interactions in the retina such that the colour of the background changes the appearance of the stimulus colours (e.g. Neumeyer, 1980; Dörr and Neumeyer, 1996; Lotto and Purves, 2000).

The fact that thresholds differed little with background colour indicates that the RNL model (Vorobyev and Osorio, 1998) is a

good predictor of absolute thresholds, even under these conditions. However, the model does not account for uncertainty resulting from large colour differences to the background and, thus, different adaptive states of visual system. As in the experiments with zebra finches (Lind, 2016), we found large inter-individual variation in discrimination thresholds and the slopes of the psychometric functions (see Fig. S3). These differences, and potential differences between batches of chickens, caution against over-interpretation of small variations among colour discrimination data. Importantly, when slopes differ, threshold estimates also depend on the definition of the threshold, i.e. whether we use the proportion of 0.65 correct choices as threshold, as appropriate for a binomial distribution with  $n=40$ , or as a more conservative estimate, the inflection point of the psychometric curve.

The influence of background colours on colour discrimination is not only interesting from a psychophysics perspective but also directly relates to ecology: colours of potential mates or ripe fruit are often seen against complex, strongly contrasting backgrounds. Background colour may thus influence the certainty with which animals interpret the colour and, thus, the quality of such objects. Indeed, sexual traits that have strong contrast against background colours are more variable in birds, presumably to overcome the reduced visual discriminability (Delhey et al., 2017). Our study, together with Lind's (2016) study on zebra finches and studies on humans (e.g. Brown and Macleod, 1997), suggests that ignoring background colours in the estimation of colour discrimination performance may sometimes predict colour perception in a natural environment inaccurately. We hope that our results can contribute to a future extension of the RNL model to also include such realistic conditions in visual ecology.

#### Acknowledgements

We thank Marie Dacke for her support of the project. We are grateful to two anonymous reviewers for their constructive criticism.

#### Competing interests

The authors declare no competing or financial interests.

#### Author contributions

Conceptualization: P.O., A.K.; Methodology: P.O., R.D.J.; Software: J.J.F., J.D.K.; Validation: P.O., J.J.F., J.D.K.; Formal analysis: P.O., J.J.F., J.D.K., A.K.; Investigation: P.O., R.D.J., J.J.F., J.D.K., A.K.; Data curation: P.O., J.J.F., J.D.K.; Writing - original draft: P.O., R.D.J., J.J.F., J.D.K., A.K.; Writing - review & editing: P.O., J.J.F., J.D.K., O.L., A.K.; Visualization: P.O., J.J.F., J.D.K.; Supervision: O.L., A.K.

#### Funding

This work was supported by Human Frontier Science Program (grant no. RGP0017/2011), the Swedish Research Council (Vetenskapsrådet; 2012–2212 to A.K. and 637-2013-388 to O.L.) and the Knut och Alice Wallenbergs Stiftelse (Ultimate Vision).

#### Data availability

The R code used for the statistical analysis of the behavioural data, the response data derived from the behavioural experiments, the HTML outputs of the R markdown versions of each script, supplementary figures, and spreadsheets of light measurements used in these experiments are available from GitHub ([https://github.com/JohnKirwan/Olsson\\_colour\\_discrimination/](https://github.com/JohnKirwan/Olsson_colour_discrimination/)) and from figshare ([https://figshare.com/articles/figure/Olsson2020\\_bonus\\_material\\_zip/13233812](https://figshare.com/articles/figure/Olsson2020_bonus_material_zip/13233812)).

#### Supplementary information

Supplementary information available online at <https://jeb.biologists.org/lookup/doi/10.1242/jeb.209429.supplemental>

#### References

Alcalá-Quintana, R. and García-Pérez, M. A. (2004). The role of parametric assumptions in adaptive Bayesian estimation. *Psychol. Methods* **9**, 250-271. doi:10.1037/1082-989X.9.2.250

- Boynton, R. M. and Whitten, D. N.** (1970). Visual adaptation in monkey cones: recordings of late receptor potentials. *Science* **170**, 1423-1426. doi:10.1126/science.170.3965.1423
- Brown, R. O. and MacLeod, D. I. A.** (1997). Color appearance depends on the variance of surround colors. *Curr. Biol.* **7**, 844-849. doi:10.1016/S0960-9822(06)00372-1
- Bürkner, P.-C.** (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *R Journal* **10**, 395-411. doi:10.32614/RJ-2018-017
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A.** (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1-32. doi:10.18637/jss.v076.i01
- Cuthill, I. C., Allen, W. L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M. E., Hill, G. E., Jablonski, N. G., Jiggins, C. D., Kelber, A. et al.** (2017). The biology of color. *Science* **357**, eaan0221. doi:10.1126/science.aan0221
- Delhey, K., Szecsenyi, B., Nakagawa, S. Peters, A.** (2017). Conspicuous plumage colours are highly variable. *Proc. R. Soc. B* **284**, 20162593. doi:10.1098/rspb.2016.2593
- Dörr, S. and Neumeyer, C.** (1996). Simultaneous color contrast in goldfish – a quantitative study. *Vision Res.* **37**, 1581-1593. doi:10.1016/S0042-6989(96)00320-3
- Endler, J. A.** (1993). The color of light in forests and its implications. *Ecol. Monogr.* **63**, 1-27. doi:10.2307/2937121
- Endler, J. A. and Mielke, P. W.** (2005). Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc.* **86**, 405-431. doi:10.1111/j.1095-8312.2005.00540.x
- Gover, N., Jarvis, J. R., Abeyesinghe, S. M. and Wathes, C. M.** (2009). Stimulus luminance and the spatial acuity of domestic fowl (*Gallus g. domesticus*). *Vision Res.* **49**, 2747-2753. doi:10.1016/j.visres.2009.08.011
- Hempel De Ibarra, N., Giurfa, M. and Vorobyev, M.** (2001). Detection of coloured patterns by honeybees through chromatic and achromatic cues. *J. Comp. Physiol. A* **187**, 215-224. doi:10.1007/s003590100192
- Hempel De Ibarra, N., Giurfa, M. and Vorobyev, M.** (2002). Discrimination of coloured patterns by honeybees through chromatic and achromatic cues. *J. Comp. Physiol. A* **188**, 503-512. doi:10.1007/s00359-002-0322-x
- Henderson, C. R.** (1982). Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics* **38**, 623. doi:10.2307/2530044
- Hill, G. E.** (1991). Plumage coloration is a sexually selected indicator of male quality. *Nature* **350**, 337-339. doi:10.1038/350337a0
- Houpt, J. W. and Bittner, J. L.** (2018). Analyzing thresholds and efficiency with hierarchical Bayesian logistic regression. *Vision Res.* **148**, 49-58. doi:10.1016/j.visres.2018.04.004
- Hunt, S., Cuthill, I. C., Bennett, A. T. D. and Griffiths, R.** (1999). Preferences for ultraviolet partners in the blue tit. *Anim. Behav.* **58**, 809-815. doi:10.1006/anbe.1999.1214
- Jarvis, J. R., Abeyesinghe, S. M., McMahon, C. E. and Wathes, C. M.** (2009). Measuring and modelling the spatial contrast sensitivity of the chicken (*Gallus g. domesticus*). *Vision Res.* **49**, 1448-1454. doi:10.1016/j.visres.2009.02.019
- Kelber, A., Vorobyev, M. and Osorio, D.** (2003). Colour vision in animals - behavioural tests and physiological concepts. *Biol. Rev.* **78**, 81-118. doi:10.1017/s1464793102005985
- Kemp, D. J., Herberstein, M. E., Fleishman, L. J., Endler, J. A., Bennett, A. T. D., Dyer, A. G., Hart, N. S., Marshall, J. and Whiting, M. J.** (2015). An integrative framework for the appraisal of coloration in nature. *Am. Nat.* **185**, 705-724. doi:10.1086/681021
- Kirwan, J. D. and Nilsson, D.-E.** (2019). A millipede compound eye mediation low-resolution vision. *Vision Res.* **165**, 36-44. doi:10.1016/j.visres.2019.09.003
- Kram, Y. A., Mantey, S. and Corbo, J. C.** (2010). Avian cone photoreceptors tile the retina as five independent, self-organizing mosaics. *PLoS ONE* **5**, e8992. doi:10.1371/journal.pone.0008992
- Krauskopf, J. and Gegenfurtner, K.** (1992). Color discrimination and adaptation. *Vision Res.* **32**, 2165-2175. doi:10.1016/0042-6989(92)90077-V
- Kuss, M., Jäkel, F. and Wichmann, F. A.** (2005). Bayesian inference for psychometric functions. *J. Vis.* **5**, 478-492. doi:10.1167/5.5.8
- Lind, O.** (2016). Colour vision and background adaptation in a passerine bird, the zebra finch (*Taenopygia guttata*). *R. Soc. Open Sci.* **3**, 160383. doi:10.1098/rsos.160383
- Lind, O., Karlsson, S. and Kelber, A.** (2013). Brightness discrimination in budgerigars (*Melopsittacus undulatus*). *PLoS ONE* **8**, 8-10. doi:10.1371/journal.pone.0054650
- Lind, O., Henze, M. J., Kelber, A. and Osorio, D.** (2017). Coevolution of coloration and colour vision? *Philos. Trans. R. Soc. B* **372**, 20160338. doi:10.1098/rstb.2016.0338
- Lotto, R. B. and Purves, D.** (2000). An empirical explanation of color contrast. *Proc. Natl. Acad. Sci. USA* **97**, 12834-12839. doi:10.1073/pnas.210369597
- Moscatelli, A., Mezzetti, M. and Lacquaniti, F.** (2012). Modeling psychophysical data at the population-level: the generalized linear mixed model. *J. Vis.* **12**, 1-17. doi:10.1167/12.11.26
- Neumeyer, C.** (1980). Simultaneous color contrast in the honeybee. *J. Comp. Physiol. A* **139**, 165-176. doi:10.1007/BF00657079
- Normann, R. A. and Werblin, F. S.** (1974). Control of retinal sensitivity: I. Light and dark adaptation of vertebrate rods and cones. *J. Gen. Physiol.* **63**, 37-61. doi:10.1085/jgp.63.1.37
- O'Hagan, A.** (2004). Bayesian statistics: principles and benefits. In *Bayesian Statistics and Quality Modelling in the Agro-Food Production Chain* (ed. M. A. J. S. van Boekel, A. Stein and A. H. C. van Bruggen), pp. 12-13. Netherlands: Springer.
- Olkkonen, M. and Allred, S. R.** (2014). Short-term memory affects color perception in context. *PLoS ONE* **9**, 1-11. doi:10.1371/journal.pone.0086488
- Olsson, P. and Kelber, A.** (2017). Relative colour cues improve colour constancy in birds. *J. Exp. Biol.* **220**, 1797-1802. doi:10.1242/jeb.155424
- Olsson, P., Lind, O. and Kelber, A.** (2015). Bird colour vision: behavioural thresholds reveal receptor noise. *J. Exp. Biol.* **218**, 184-193. doi:10.1242/jeb.111187
- Olsson, P., Wilby, D. and Kelber, A.** (2016). Quantitative studies of animal colour constancy: using the chicken as model. *Proc. R. Soc. B* **283**, 20160411. doi:10.1098/rspb.2016.0411
- Olsson, P., Wilby, D. and Kelber, A.** (2017). Spatial summation improves bird colour vision in low light intensities. *Vision Res.* **130**, 1-8. doi:10.1016/j.visres.2016.10.009
- Olsson, P., Lind, O. and Kelber, A.** (2018). Chromatic and achromatic vision: parameter choice and limitations for reliable model predictions. *Behav. Ecol.* **29**, 273-282. doi:10.1093/beheco/arx133
- Osorio, D. and Vorobyev, M.** (1996). Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. B* **263**, 593-599. doi:10.1098/rspb.1996.0089
- Osorio, D. and Vorobyev, M.** (2005). Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proc. R. Soc. B Biol. Sci.* **272**, 1745-1752. doi:10.1098/rspb.2005.3156
- Osorio, D., Vorobyev, M. and Jones, C. D.** (1999). Colour vision of domestic chicks. *J. Exp. Biol.* **202**, 2951-2959.
- Renoult, J. P., Kelber, A. and Schaefer, H. M.** (2017). Colour spaces in ecology and evolutionary biology. *Biol. Rev.* **92**, 292-315. doi:10.1111/brv.12230
- Schaefer, H. M., Levey, D. J., Schaefer, V. and Avery, M. L.** (2006). The role of chromatic and achromatic signals for fruit detection by birds. *Behav. Ecol.* **17**, 784-789. doi:10.1093/beheco/ar1011
- Smith, V. C., Pokorny, J. and Sun, H.** (2000). Chromatic contrast discrimination: Data and prediction for stimuli varying in L and M cone excitation. *Color Res. Appl.* **25**, 105-115. doi:10.1002/(SICI)1520-6378(200004)25:2<105::AID-COL5>3.0.CO;2-G
- Sumner, P. and Mollon, J. D.** (2000). Chromaticity as a signal of ripeness in fruits taken by primates. *J. Exp. Biol.* **203**, 1987-2000.
- Vorobyev, M. and Osorio, D.** (1998). Receptor noise as a determinant of colour thresholds. *Proc. R. Soc. B* **265**, 351-358. doi:10.1098/rspb.1998.0302
- Vorobyev, M., Osorio, D., Bennett, A. T. D., Marshall, N. J. and Cuthill, I. C.** (1998). Tetrachromacy, oil droplets and bird plumage colours. *J. Comp. Physiol. A* **183**, 621-633. doi:10.1007/s003590050286
- Wallis, S. A., Baker, D. H., Meese, T. S. and Georgeson, M. A.** (2013). The slope of the psychometric function and non-stationarity of thresholds in spatiotemporal contrast vision. *Vision Res.* **76**, 1-10. doi:10.1016/j.visres.2012.09.019