Quantifying camouflage and conspicuousness using visual salience

Thomas W. Pike*

School of Life Sciences, University of Lincoln, Lincoln, LN2 2UU

*Correspondence to: tpike@lincoln.ac.uk

1. Being able to quantify the conspicuousness of animal and plant colouration is key to understanding its evolutionary and adaptive significance. Camouflaged animals, for example, are under strong selection pressure to minimise their conspicuousness to potential predators. However, successful camouflage is not an intrinsic characteristic of an animal, but rather an interaction between that animal's phenotype and the visual environment that it is viewed against. Moreover, the efficacy of any given camouflage strategy is determined not by the signaller's phenotype per se, but by the perceptual and cognitive capabilities of potential predators. Any attempts to quantify camouflage must therefore take both predator perception and the visual background into account.

2. Here I describe the use of species-relevant saliency maps, which combine the different visual features that contribute to selective attention (in this case the luminance, colour and orientation contrasts of features in the visual environment) into a single holistic measure of target conspicuousness. These can be tuned to the specific perceptual capabilities of the receiver, and used to derive a quantitative measure of target conspicuousness. Furthermore, I provide experimental evidence that these computed measures of conspicuousness significantly predict the performance of both captive and wild birds when searching for camouflaged artificial prey.

3. By allowing the quantification of prey conspicuousness, saliency maps provide a useful tool for understanding the evolution of animal signals. However, this is not limited to inconspicuous visual signals, and the same approach could be readily used for quantifying conspicuous visual signals in a wide variety of contexts, including, for example, signals involved in mate choice and warning colouration.

Keywords: Selective attention, signal evolution, crypsis, visual salience, conspicuousness

Introduction

Being able to attend to relevant objects in a cluttered visual scene has considerable evolutionary significance because it allows an animal to rapidly identify potential food, mates and predators. Indeed, some stimuli are intrinsically conspicuous, or salient, in a given context; for example, in humans a ripe red fruit among green leaves automatically and involuntarily attracts attention (Frey *et al.* 2011). Saliency is independent of the nature of the particular task, operates very rapidly, and is primarily driven in a bottom-up manner that reflexively directs visual focus based on certain low-level visual features (e.g. colour, orientation and/or brightness contrasts). If a stimulus is sufficiently salient, it will therefore 'pop out' of a visual scene (Itti & Koch 2001). As a result, the concept of visual salience has clear implications for understanding the evolution of animal signals which, broadly speaking, either aim to maximise saliency (as in the case of animals producing conspicuous mating signals) or minimise it (as in animals that rely on camouflage to avoid detection by potential predators) (Ruxton, Sherratt & Speed 2004; Endler & Mielke 2005; Stevens & Merilaita 2009).

However, despite its importance, predicting an animal's salience from its visual appearance remains a major challenge. This is in part because saliency is not an intrinsic characteristic of an animal, but rather an interaction between that animal's phenotype and the visual environment that it is viewed against which, in nature, is likely to be heterogeneous and visually cluttered (Godfrey, Lythgoe & Rumball 1987; Merilaita 2003; Dimitrova & Merilaita 2014). Because of this, an animal that is well camouflaged against one background may be highly salient against another; any useful measure of saliency must therefore take into account the relative characteristics of both the target and its background (Xiao & Cuthill 2016). Moreover, conspicuousness is determined not by the signaller's visual phenotype per se, but is a function of the perceptual and cognitive capabilities of potential receivers (Thery & Casas 2002; Stevens & Cuthill 2006; Osorio & Vorobyev 2008; Chiao *et al.* 2009). Different species vary in their perceptual abilities (e.g. in the spectral sensitivity of their retinal photoreceptors) and in the cognitive mechanisms underpinning how perceptual information is processed and integrated (Kesner & Olton 2014), and this will necessarily impact on how salient prey with particular phenotypic characteristics appear. Animals which appear highly salient to one receiver may completely lack salience for another, even against the same

visual background, emphasising the importance of incorporating species-relevant perceptual and cognitive information into estimates of salience wherever possible (Xiao & Cuthill 2016; Troscianko, Skelhorn & Stevens 2017).

To address the challenge of quantifying an animal's visual salience a wide variety of metrics have been suggested, some of which are inspired by known features of animals' visual and cognitive systems. These include metrics for quantifying internal and external edges (Stevens & Cuthill 2006; Lovell *et al.* 2013; Webster *et al.* 2013; Kang *et al.* 2015; Troscianko, Skelhorn & Stevens 2017), the orientation of which often contrasts with those in the background or with edges intrinsic to the prey itself; those involving pattern detection or the identification of pattern contrasts (Spottiswoode & Stevens 2010; Stoddard, Kilner & Town 2014; Troscianko *et al.* 2016; Troscianko, Skelhorn & Stevens 2017); those which calculate chromatic (Kang *et al.* 2015) or luminance (i.e. perceived brightness) differences or contrasts between a prey and its background (Troscianko *et al.* 2016); and those that quantify the complexity of the visual scene against which the prey is viewed (Xiao & Cuthill 2016). Many of these are supported by empirical evidence demonstrating their efficacy in quantifying predation risk. However, while the application of these various metrics has made significant contributions to our understanding of the visual features that influence prey conspicuousness (Troscianko, Skelhorn & Stevens 2017), they tend to be employed independently, despite the fact that the visual features they encapsulate are typically available simultaneously to any animal viewing a scene. This limits our understanding of how these different visual features may be differentially weighted by a predator's visual system. Moreover, differences in the way these various metrics are implemented and the different assumptions they make (Troscianko, Skelhorn & Stevens 2017) means they are not easily combined into a holistic measure of signal conspicuousness, which is ultimately what choice is based on (Stevens & Merilaita 2009). One recent exception to this is the study by Xiao and Cuthill (2016), which used various measures of 'visual clutter' in the background against which prey were viewed to estimate detectability. Their approach allowed the relative efficacy of chromatic, achromatic and textural (i.e. orientation-based) clutter to be explored independently, but could also be combined into a composite measure that simultaneously considered clutter across all three feature types.

Here I describe an alternative approach, based on the neurophysiologically-inspired model of bottom-up visual attention described by Itti, Koch and Niebur (1998). The adaptation of this model described here allows the computation of species-relevant 'saliency maps', which topographically encode conspicuity over an entire visual scene and hence intrinsically incorporate the relative salience of both the target and its (heterogeneous) background. I then demonstrate that relative target salience is a good predictor of the performance of avian predators searching for camouflaged artificial prey both under constrained conditions in the lab, using Japanese quail (*Coturnix japonica*) searching for computer-generated targets on a computer screen, and in the field, using the predation of artificial moth-like targets by wild birds. In order to provide a comparison with other approaches that have successfully been used to characterise prey conspicuousness in comparable experiments, I also compare the performance of the saliency model described here with the best-performing models identified by Troscianko, Skelhorn and Stevens (2017) in their comprehensive comparison of models available at the time, and those used previously by Xiao and Cuthill (2016).

Methods

Modelling visual salience

In order to model the salience of features within a heterogeneous visual scene I adapt the model of bottom-up visual attention described by Itti, Koch and Niebur (1998). This model, and adaptations of it, are widely used within computer vision, neuroscience and human cognition (Itti & Koch 2001; Borji & Itti 2013), and have also been used to address questions in animal signalling (Peters 2010). Although many extensions to the model have been proposed in order to improve the fit to psychophysical data on human saliency perception (Borji & Itti 2013), the original version of the model still provides an excellent base from which to adapt the concept for non-human animals. For a full description of the underlying rationale and computation details readers are referred to Koch and Ullman (1985), Itti, Koch and Niebur (1998) and Walther and Koch (2006); here I provide an overview of the model architecture (Fig. 1), noting in particular where adaptations have been made to improve

generality and to address the specific question of target detection. Wherever possible, notation follows that used in Itti, Koch and Niebur (1998) for consistency.

The original model was inspired by the neurophysiological characteristics of human (and other trichromatic primate) visual systems, and so includes some assumptions that may not be appropriate for modelling saliency in other species. In particular, most applications of the model use an RGB image as input, in which each of the three colour channels (R, G and B) is assumed to broadly correspond to the response of one of the three cone classes in the human retina (Mollon 1989), and luminance is estimated as the mean of these three colour channels (Walther & Koch 2006). However, it is unlikely that these assumptions are appropriate for the majority of animal species (Osorio & Vorobyev 2005). In order to increase the generality of the model, I therefore adapted it to accept an arbitrary number of $n \times m$ grayscale 'images' I as inputs, each of which is assumed to provide a topographical representation of the quantum catch of one of the viewing animal's cone classes. In this paper I use birds as model predators (see below), and so the model was explicitly adapted for a tetrachromatic visual system in which four classes of single cone (long wavelength-, medium wavelength-, short wavelength- and ultraviolet/violet-sensitive, denoted L, M, S and U, respectively) are assumed to contribute to colour perception, and double cones (D) are assumed to contribute to luminance perception (Osorio, Miklosi & Gonda 1999; Jones & Osorio 2004; Osorio & Vorobyev 2005), although it would be straightforward to modify this to cope with variable numbers of cone classes (e.g. to represent dichromatic or pentachromatic visual systems) and different luminance perception mechanisms (e.g. those based on the summed input from two or more cone classes; Endler and Mielke (2005)). These input images are denoted I_{L}, I_{M}, I_{S}, I_{U} and I_{D}, respectively.

For each of these input images, a Gaussian pyramid is then constructed by iteratively low-pass filtering and subsampling the image to produce a sequence of reduced-resolution images (Walther & Koch 2006). At each successive iteration, the next levels $\sigma = [0, \ldots, 7]$ of the pyramid are obtained, such that the resolution of level σ is $1/2^{\sigma}$ times the original image resolution; i.e., the seventh level has a resolution of 1/128th that of the input image. Each level of the pyramid is then further decomposed into a series of 'maps', corresponding to the early visual features of luminance, colour and orientation. For luminance, the local map at level σ, $M_L(\sigma)$, is simply

$$M_L(\sigma) = I_{\mathrm{D}}(\sigma). \quad (1)$$

The original model assumes that colour can be encoded using four broadly tuned colour channels – namely red, green, blue and yellow (i.e. a linear combination of the red and green channels) – and that local colour maps can be constructed on the basis of red–green/green–red and blue–yellow/yellow–blue double opponent interactions (Livingstone & Hubel 1984; Itti & Koch 2001; Walther & Koch 2006). However, while this may be appropriate for the visual system of trichromatic primates on which the original model was based, it is unlikely that these particular opponent interactions are appropriate for the overwhelming majority of species (Kelber, Vorobyev & Osorio 2003). In this adaptation of the model, rather than assume that colour perception results from specific opponent interactions, I adopt a more general approach in which all possible pairwise colour opponent interactions between the cone classes putatively contributing to colour perception are considered (*sensu* Vorobyev and Osorio (1998)). Because birds are used here, I therefore considered six putative opponent interactions: LM, LS, LU, MS, MU and SU, although it would be straightforward to incorporate or restrict this to specific known or hypothesised opponent interactions, if this information was available for the species under study (e.g. Osorio, Miklosi and Gonda (1999)). Local colour maps are computed following Walther and Koch (2006): for the putative LM opponent mechanism, for example, the corresponding colour map $M_{\mathrm{LM}}(\sigma)$ at level σ is calculated as

$$M_{\mathrm{LM}}(\sigma) = \frac{|I_{\mathrm{L}}(\sigma) - I_{\mathrm{M}}(\sigma)|}{I_{\mathrm{D}}(\sigma)}. \quad (2)$$

Maps encoding for the putative LS, LU, MS, MU and SU mechanisms are created in a similar way.

In the model it is assumed that textural (i.e. orientation-based) features are detected using achromatic information (Itti & Koch 2001; Walther & Koch 2006). Local orientation maps $M_\theta(\sigma)$ are therefore computed by convolving (Russ & Neal 2016, p. 352) the levels of the I_{D} pyramid with Gabor filters, such that

$$M_\theta(\sigma) = |I_{\mathrm{D}}(\sigma) * G_0(\theta)| + |I_{\mathrm{D}}(\sigma) * G_{\pi/2}(\theta)|, \quad (3)$$

where $G_{\psi}(\theta)$ is a Gabor filter with a standard deviation of 7/3 pixels, a phase of $\psi \in \{0, \pi/2\}$ and an orientation of $\theta \in \{0°, 45°, 90°, 135°\}$ (following Walther and Koch (2006)), and $*$ denotes convolution.

The next step is to construct a number of 'feature maps' that encode local luminance, colour and orientation contrasts, using a set of linear 'centre-surround' operations analogous to visual receptive fields (Hubel & Wiesel 1959). Typical visual neurons are most sensitive to a small region of visual space (the centre), while stimuli presented in a broader antagonistic region around the centre (the surround) inhibit the neural response. This increases sensitivity to local spatial discontinuities, and so is particularly well-suited to detecting regions of space which locally stand out from their surround (i.e. which are salient). Centre-surround operations are implemented in the model as differences between a 'centre' fine scale c and a 'surround' coarser scale s. Specifically, the centre is a pixel at scale $c \in \{2,3,4\}$ and the surround is the corresponding pixel at scale $s = c + \delta$, where $\delta \in \{3,4\}$. Such across-scale differences, denoted '$\ominus$' below, are obtained by interpolation to the finer scale followed by point-by-point subtraction (Itti & Koch 2001). A feature map F for a particular centre and surround is therefore calculated as

$$F_k(c, s) = N(|M_k(c) \ominus M_k(s)|), \quad (4)$$

where $\forall k\ K \in \{L\} \cup \{\mathrm{LM}, \mathrm{LS}, \mathrm{LU}, \mathrm{MS}, \mathrm{MU}, \mathrm{SU}\} \cup \{0°, 45°, 90°, 135°\}$, and $N(\cdot)$ is an iterative, nonlinear normalisation operator, simulating local competition between neighbouring salient locations (Itti & Koch, 2001). The normalisation process is fully described elsewhere (Itti, Koch & Niebur 1998; Walther & Koch 2006), but in brief each feature map is normalised to the range [0,1] and then iteratively convolved by a two-dimensional difference-of-Gaussian filter. Between iterations, the original image is added to the new one and negative values set to zero. The effect of this is (i) to eliminate feature-dependent differences caused by different feature extraction mechanisms, and (ii) to promote regions of the map which differ most from the average (i.e. which are likely to be the most salient), while suppressing homogenous or repetitive regions.

These feature maps are then combined into three 'conspicuity maps', for luminance C_L, colour C_C, and orientation C_O (Fig. 2), using across-scale addition (denoted '$\oplus$' below),

which consists of reduction of each map to scale $\sigma = 4$ and point-by-point addition (Itti, Koch & Niebur 1998), to give

$$C_L = N\left(\bigoplus_{c=2}^{4}\bigoplus_{s=c+3}^{c+4} N\big(F_L(c,s)\big)\right), \tag{5}$$

$$C_C = N\left(\bigoplus_{c=2}^{4}\bigoplus_{s=c+3}^{c+4} \sum_{\varphi\in\{\mathrm{LM,LS,LU,MS,MU,SU}\}} N\left(F_\varphi(c,s)\right)\right), \tag{6}$$

$$C_O = N\left(\sum_{\theta\in\{0°,45°,90°,135°\}} N\left(\bigoplus_{c=2}^{4}\bigoplus_{s=c+3}^{c+4} N\big(F_\theta(c,s)\big)\right)\right). \tag{7}$$

Finally, these three conspicuity maps are linearly combined to produce a single overall saliency map S (Fig. 2), such that

$$S = \omega_L C_L + \omega_C C_C + \omega_O C_O, \tag{8}$$

where ω_L, ω_C and ω_O are weighting factors in the range $[0,1]$, that allow the three feature types to contribute differentially to saliency. The resulting saliency map topographically encodes conspicuity over the entire visual scene, and therefore provides a continuous measure of salience at any given location.

Computing target salience

In order to identify the location of a relevant target (e.g. a prey item) in a visual scene, a viewing animal must be able to distinguish the region containing the target of interest from other (possibly equally) salient regions of the background (i.e. the signal must be sufficiently large relative to the prevailing noise; Navalpakkam & Itti 2006). The more salient the elements of the background (or the less salient the elements of the target) are, on average, the harder this task will be. Quantifying the relative salience of a target therefore requires calculating an appropriate measure of distance between the value of pixels within the target and the value of pixels in the background (see Fig. 3). Because these pixel values can follow any arbitrary distribution (and so metrics based on mean pixel values are not always appropriate; Navalpakkam & Itti 2006), here I used a histogram-based method which compares the empirical cumulative histograms of pixel saliency values for the target and its

background; a common technique in image analysis (Pal & Peters 2010) which is insensitive to the specific distributions of the data. Specifically, relative target salience S_t is found by taking the sum of differences between the cumulative histograms of pixel salience values for the background H_b and the target H_t in a given saliency map as

$$S_t = \frac{1}{N}\sum_{j=1}^{N} H_b(j) - H_t(j), \quad (9)$$

where each cumulative histogram is divided into N bins, where $j \in \{1, 2, 3, \ldots, N\}$. Here N was set to 100. If pixel values within one or more visual features of the target (e.g. in colour, luminance and/or orientation) are high compared to the background, then S_t will be high (>> 0) and locating the target is predicted to be relatively easy (e.g. Fig. 3a,b); if the target and background share many visual features in common, or if the pixel values of the background are high compared to the target, then S_t will be low (≈ 0) and locating the target is predicted to be hard (e.g. Fig. 3c,d). This metric therefore defines a holistic measure of 'target salience', which takes into account the salience of both the target itself and the salience of features within the background it is viewed against.

The implementation of the saliency model used here is based on a Matlab (MathWorks, Natick, MA) version of Itti, Koch and Niebur (1998)'s original model (Harel, Koch & Perona 2006), adapted as described above, and available from github.com/thomaswpike/salience.

Predation experiments

In order to test whether the model is able to predict the behaviour of real animals searching for targets that varied in their relative salience, I conducted two experiments in which avian predators were tasked with searching for and locating camouflaged artificial prey. Experiment 1 was run under controlled conditions in the lab, using Japanese quail (*Coturnix japonica*) searching for computer-generated targets on a computer screen. Experiment 2 was conducted in the field, employing a widely used approach (Cuthill *et al.* 2005) to quantify the detection of artificial targets by wild birds. In both cases, targets consisted of moth-like patterned triangles, viewed against a bark background. The visual scene in which each of the prey targets was viewed (either the computer screen, or calibrated photographs of the targets in situ in the field) was then used to construct the five quantum catch 'images'

needed for the computation of target salience. Full methodological details of these experiments are given in the supplementary material.

Comparison with alternative metrics

A large number of metrics have been developed to characterise prey conspicuousness, many of which have been very successful in predicting predator performance across a range of species and contexts. Here I provide a qualitative comparison of the performance of the saliency model described in this paper with some of these other approaches. However, rather than exhaustively test every available metric (not least because such a comparison has recently been conducted; Troscianko, Skelhorn and Stevens (2017)), here I focus specifically on the best-performing metrics identified by Troscianko, Skelhorn and Stevens (2017) that take into account both the characteristics of the target and the characteristics of its background (and so provide a meaningful comparison with the saliency model described here), along with the visual 'clutter' metrics used in the recent paper by Xiao and Cuthill (2016). These metrics are listed in Table 1, summarised in the supplementary material and described in detail in the original publications (Rosenholtz *et al.* 2005; Stevens & Cuthill 2006; Stoddard, Kilner & Town 2014; Xiao & Cuthill 2016; Troscianko, Skelhorn & Stevens 2017).

Statistical analysis

To test whether target salience predicted predator success in the two experiments, I used (generalised) linear mixed-effect models, fitted using the 'lmer' and 'glmer' functions in the 'lme4' package (Bates *et al.* 2015) for R version 3.3.1. Full details are given in the supplementary material. In each case significance was determined by comparing a full model to models lacking the effect of interest using likelihood ratio tests (Crawley 2005), and assumptions validated following Zuur, Ieno and Elphick (2010).

Because the relative contribution of the different feature types (luminance, colour and orientation) to the perception of overall salience is unknown for birds (Xiao & Cuthill 2016), target salience was initially calculated from saliency maps in which each conspicuity map was weighted equally (i.e. $\omega_L = \omega_C = \omega_O = 1$ in Eq. 8). However, it is unlikely that animals

do in fact weight these different features equally (Rosenholtz *et al.* 2005), and so it is useful to explore which set of feature weights provides the best predictive power. To do this, I systematically varied the values of ω_L, ω_C and ω_O in the computation of the final saliency map, and then reran the analyses for each combination of weights. In each case, the quality of the model fit was quantified using its AIC score, with the 'optimal' combination of weights defined as those which resulted in the lowest AIC (Burnham & Anderson 2002). For ease of comparison, AIC scores are presented as differences from this smallest AIC (i.e. in terms of their ΔAIC; Burnham & Anderson 2002).

In order to compare the relative performance of the saliency model described here with the various alternative metrics, each of the analyses was rerun, but substituting 'target salience' for each of the alternative metrics in turn. The quality of the model fit in each case was quantified using its ΔAIC score, as above, allowing qualitative comparison between the metrics. Models were considered equally well-fitting if ΔAIC < 2 (Burnham & Anderson 2002).

Results

Experiment 1

For quail predating virtual moths the time taken to catch camouflaged prey was significantly predicted by the salience of the target ($\chi^2(1) = 19.77$, $p < 0.001$), with time taken decreasing as the target became increasingly salient (Fig. 4a). There was no evidence of predator learning over successive trials ($\chi^2(1) = 0.17$, $p = 0.680$), or any evidence that prey nearer the centre of the screen were quicker to catch ($\chi^2(1) = 0.11$, $p = 0.740$). However, assuming that the visual features contributing to target salience were equally weighted (i.e. $\omega_L = \omega_C = \omega_O = 1$) did not produce the best-fitting model (Fig. 4b-d); instead, model fit increased roughly linearly as the relative luminance (ω_L) and orientation (ω_O) weights increased (Fig. 4b), with a moderate contribution from colour (ω_C) (Fig. 4c,d). The best-fitting model had the following feature weights: $\omega_L = 1.0$, $\omega_C = 0.5$ and $\omega_O = 0.7$ ($\chi^2(1) = 20.38$, $p < 0.001$).

Comparing the alternative camouflage metrics, the best-fitting models were the 'optimally'-weighted (ΔAIC = 0.0) and equally-weighted (ΔAIC = 0.9) saliency models, both of which

provided a substantially better fit than the next best metric, the Gabor Edge Disruption Ratio (ΔAIC = 10.6) (Table 1).

Experiment 2

Target salience significantly predicted the survival of moth-like targets deployed in the field ($\chi^2(1) = 8.93$, $p = 0.003$), such that those surviving predation by birds had a significantly lower target salience than those that were predated (Fig. 5a). As for Experiment 1, model fit varied considerably with feature weight, although the overall pattern was somewhat different. Specifically, the best-fitting model occurred when orientation was weighted high ($\omega_O = 0.9$), luminance was weighted relatively low ($\omega_L = 0.3$), and colour did not contribute at all to target salience ($\omega_C = 0.0$) ($\chi^2(1) = 18.56$, $p < 0.001$; Fig. 5b,c,d).

When comparing between the different metrics, the best-fitting models were those including Sub-band Entropy (ΔAIC = 0.0) and the 'optimally'-weighted saliency model (ΔAIC = 1.5). The next-best fitting models included Luminance Feature Congestion (ΔAIC = 4.8), the equally-weighted saliency model (ΔAIC = 11.2), Overall Feature Congestion (ΔAIC = 11.4) and Orientation Congestion (ΔAIC = 11.8) (Table 1).

Discussion

This study explored the efficacy of species-relevant saliency maps as predictors of predator performance in two tasks involving locating cryptic targets against noisy backgrounds. The results clearly demonstrate that across both laboratory and field contexts target salience is a good predictor of predator performance, with laboratory quail locating salient virtual moths quicker than those that were estimated to appear less salient (Experiment 1), and wild birds most successfully predating artificial moths that were deemed the most salient (Experiment 2). Moreover, it allowed information on the possible weighting of the different feature types contributing to a predator's perception of target salience to be inferred. Interestingly, these weightings differed between the two experiments. In Experiment 1, birds appeared to be using a combination of luminance, colour and orientation features to inform their behaviour, although the highest weightings came from luminance and orientation. In

Experiment 2, the birds appeared to be predominantly using orientation features, with a lesser reliance on luminance and no contribution at all from colour. While this provides some evidence that the relative efficacy of luminance-based cues may be greater than chromatic cues (Stevens and Cuthill (2006); cf. Schaefer and Stobbe (2006)), it is impossible to know whether the difference in the relative weightings of the three feature types between the two experiments was driven by differences in vision or cognition between the species involved, or by differences between the experimental setups. For example, in Experiment 2 the distance at which prey were viewed was likely to be both initially greater and considerably more variable than in Experiment 1, which would be important if the weighting of the different feature types depended on distance or perceived prey size. Moreover, the search space in Experiment 2 included three-dimensional information (providing a possible explanation for the reduced reliance of luminance cues, as these may be less useful when searching in a three-dimensional environment; Zhang *et al.* (2010)), and would have included elevated (but unmeasured) noise in luminance and colour due to short-term illumination changes, possibly rendering colour and luminance cues less reliable. However, despite these differences the findings of both experiments are broadly consistent with previous studies, in which the textural (i.e. orientation-based) complexity of the background (Xiao & Cuthill 2016) and the conspicuousness of the prey's outline (Stevens & Cuthill 2006; Lovell *et al.* 2013; Webster *et al.* 2013; Kang *et al.* 2015; Troscianko *et al.* 2016; Troscianko, Skelhorn & Stevens 2017) have been identified as important determinants of predator success. Orientation features therefore appear to be a key component in the detection of camouflaged prey across a range of species and contexts, although the results of this study emphasise the need to also consider the relative contribution of other feature types if we are to fully understand the mechanisms predators use to detect prey.

As well as predicting predator performance in the two experiments reported here, the performance of the saliency model also compared very favourably with a number of alternative metrics that have been proposed to quantify prey conspicuousness in analogous situations (Xiao & Cuthill 2016; Troscianko, Skelhorn & Stevens 2017). In Experiment 1 it performed substantially better than all the other metrics tested, possibly because the birds appeared to be using a combination of luminance, colour and orientation features to inform their behaviour; something that is not encapsulated in metrics that focus on a single feature

type. For example, the next best performing metric, the Gabor Edge Disruption Ratio, was found to perform extremely well in Troscianko, Skelhorn and Stevens (2017)'s human-based study, possibly because achromatic stimuli were used. While focussing on achromatic stimuli was entirely reasonable, given that the luminance channel in primates has numerous oriented edge detectors suitable for shape processing (Hesse & Georgeson 2005), colour is also known to contribute to target detection by facilitating the segregation of surfaces that differ in chromaticity (Gegenfurtner & Rieger 2000). It is possible, therefore, that had chromatic information also been unavailable in the present study, birds may have weighted orientation-based features more heavily. In Experiment 2, the best-performing metrics were Sub-band Entropy (cf. Xiao and Cuthill (2016)) and the 'optimally'-weighted saliency model, with Luminance Congestion also performing well. Such variation in model fit between the various metrics is likely to stem, at least in part, from what they are actually quantifying, as well as the characteristics of the specific prey and backgrounds used. For example, the 'congestion' and 'clutter' metrics (which include Sub-band Entropy and Luminance Congestion) are global measures of the background against which the prey is viewed, and do not explicitly compare features of the prey with those of its background (Xiao & Cuthill 2016). As such, a plain prey item against a congested background could actually appear highly salient. Similarly, other metrics focussing specifically on the outline of the prey, such as the number of true edges detected by the Hough transform (Stevens & Cuthill 2006) and the Gabor Edge Disruption Ratio (Troscianko, Skelhorn & Stevens 2017), ignore at least some of the prey's internal features, which may themselves be highly salient. Further work is therefore needed to identify the strengths and weaknesses of these various approaches across different contexts, particularly with regard to the alternative mechanisms of camouflage (Stevens & Merilaita 2009). For example, the relative performance of the edge-based metrics may well have been improved if the targets explicitly incorporated disruptive patterns rather than simply representing samples of the background. It should also be noted that, while the saliency model performed well, measures of overall salience per se tell us little about the mechanisms underpinning successful camouflage. To address this, we still need to consider the various component parts separately.

The model of visual salience used in this paper is primarily concerned with 'bottom-up' salience, which reflexively directs visual focus based on certain low-level visual features (Itti

& Koch 2001). This mimics the case where a predator has no a priori knowledge of the prey or its background, and is not an unrealistic assumption for the experiments described here given that each of the moths and background combinations was unique. However, given repeated exposure to a prey item with particular identifying characteristics, a predator may have the opportunity to learn about the statistical properties of both the prey and its background and use this to optimise its search (Navalpakkam & Itti 2006; Borji & Itti 2013). A camouflaged prey item that lacks bottom-up salience could therefore still be effectively detected through 'top-down', or knowledge-based, guidance to known prey locations and features. In terms of the model used here, this could be implemented by optimising the weighting given to each of the bottom-up feature and conspicuity maps when computing the final saliency map, with the aim of giving high weighting to features predominantly found in prey and low weighting to features that predominate in the background. This would be akin to a sensory system enhancing neurons tuned to properties of the prey and/or supressing neurons tuned to properties of the background, thereby maximising target detection speed (Navalpakkam & Itti 2006).

The focus of this paper has been on using salience to describe the efficacy of camouflage in animals. However, the general approach would apply equally well to the assessment of conspicuity in animals or plants that have evolved to maximise their probability of detection, including those displaying conspicuous signals within a mate choice context. Because it is possible to use the feature and (colour, luminance and orientation) conspicuity maps to make inferences about which feature channel most contributes to the saliency of a target, it may allow us to better understand both signal design and receiver cognition. For example, Fig. 6 shows the three peafowl-specific conspicuity maps derived from a calibrated colour image of a displaying peacock (*Pavo cristatus*), in which the relative contribution of colour, luminance and orientation features are presented. There is little evidence of luminance salience in the elements of the peacock's colouration compared to the background they are viewed against (although this may be because saliency was derived from a static photograph; due to the iridescence of the peacock's eyespots [Loyau *et al.* 2007], there is likely to be large modulations of luminance with movement, creating salience through signal change). However, the eyespots on the tail feathers, which have been repeatedly implicated as a target for female choice (Petrie, Halliday & Sanders 1991; Petrie

& Halliday 1994; Loyau *et al.* 2007; Dakin & Montgomerie 2011), exhibit clear colour salience when viewed against their local background, with the different colour elements of the eyespots clearly delineated. Furthermore, the radial changes in the tail feather orientation around the train result in local orientation contrasts and hence regions of high orientation salience. Combined, these make elements of the peacock's train highly salient against their local background. This is, of course, simply an illustrative example; however, an approach like this could allow studies to identify and focus on particular aspects of a signal that contribute disproportionately to its conspicuousness, while avoiding aspects that may be poorly perceived.

In this study the focus was necessarily on avian visual systems under fairly constrained experimental conditions, although the visual salience of a given target may in fact differ considerably between receivers of different species and in response to variation in the physical and biological environment. In particular, the spectrum, intensity and orientation of illuminating light, as well as the presence of features such as shadows, will likely play a significant role in determining how salient a target appears. This has been widely explored in terms of chromatic contrasts (Uy & Endler 2004), although less so in terms of luminance and orientation (Troscianko *et al.* 2016). Moreover, several features which have been linked to salience in humans remain largely unexplored in animals, including contrasts arising from variation in depth (Zhang *et al.* 2012; Ma & Hang 2015) and motion (Belardinelli, Pirri & Carbone 2009; Peters 2010); the approach used here provides the flexibility needed to incorporate these different visual features (Walther & Koch 2006) to explore how salient targets appear to a variety of different visual systems, across a range of different biological and physical contexts.

Acknowledgements

I thank Oliver Burman and Charles Deeming for valuable discussions, and Innes Cuthill, Martin Stevens and an anonymous reviewer for comments that greatly improved the manuscript. This work was supported in part by a Natural Environment Research Council fellowship (NE/F016514/1).

Data Accessibility

Data associated with this paper is available from eprints.lincoln.ac.uk/31393.

References

Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software,* **67,** 1-48.

Belardinelli, A., Pirri, F. & Carbone, A. (2009) Motion Saliency Maps from Spatiotemporal Filtering. *Attention in Cognitive Systems,* **5395,** 112-123.

Borji, A. & Itti, L. (2013) State-of-the-Art in Visual Attention Modeling. *Ieee Transactions on Pattern Analysis and Machine Intelligence,* **35,** 185-207.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach,* 2nd ed. edn. Springer-Verlag.

Chiao, C.C., Chubb, C., Buresch, K., Siemann, L. & Hanlon, R.T. (2009) The scaling effects of substrate texture on camouflage patterning in cuttlefish. *Vision Research,* **49,** 1647-1656.

Crawley, M.J. (2005) *Statistics: an introduction using R.* John Wiley, Chichester, UK.

Cuthill, I.C., Stevens, M., Sheppard, J., Maddocks, T., Parraga, C.A. & Troscianko, T.S. (2005) Disruptive coloration and background pattern matching. *Nature,* **434,** 72-74.

Dakin, R. & Montgomerie, R. (2011) Peahens prefer peacocks displaying more eyespots, but rarely. *Animal Behaviour,* **82,** 21-28.

Dimitrova, M. & Merilaita, S. (2014) Hide and seek: properties of prey and background patterns affect prey detection by blue tits. *Behavioral Ecology,* **25,** 402-408.

Endler, J.A. & Mielke, P.W. (2005) Comparing entire colour patterns as birds see them. *Biological Journal of the Linnean Society,* **86,** 405-431.

Frey, H.P., Wirz, K., Willenbockel, V., Betz, T., Schreiber, C., Troscianko, T. & Konig, P. (2011) Beyond correlation: do color features influence attention in rainforest? *Frontiers in Human Neuroscience,* **5**.

Gegenfurtner, K.R. & Rieger, J. (2000) Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology,* **10,** 805-808.

Godfrey, D., Lythgoe, J.N. & Rumball, D.A. (1987) Zebra Stripes and Tiger Stripes - the Spatial-Frequency Distribution of the Pattern Compared to That of the Background Is Significant in Display and Crypsis. *Biological Journal of the Linnean Society,* **32,** 427-433.

Harel, J., Koch, C. & Perona, P. (2006) Graph-based visual saliency. *Advances in neural information processing systems*, pp. 545-552.

Hart, N.S. (2002) Vision in the peafowl (Aves: *Pavo cristatus*). *Journal of Experimental Biology,* **205,** 3925-3935.

Hesse, G.S. & Georgeson, M.A. (2005) Edges and bars: where do people see features in 1-D images? *Vision Research,* **45,** 507-525.

Hubel, D.H. & Wiesel, T.N. (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology,* **124,** 574–591.

Itti, L. & Koch, C. (2001) Computational modelling of visual attention. *Nature Reviews Neuroscience,* **2,** 194-203.

506 Itti, L., Koch, C. & Niebur, E. (1998) A model of saliency-based visual attention for rapid
507 scene analysis. *Ieee Transactions on Pattern Analysis and Machine Intelligence,* **20,**
508 1254-1259.
509 Jones, C.D. & Osorio, D. (2004) Discrimination of oriented visual textures by poultry chicks.
510 *Vision Research,* **44,** 83-89.
511 Kang, C., Stevens, M., Moon, J.Y., Lee, S.I. & Jablonski, P.G. (2015) Camouflage through
512 behavior in moths: the role of background matching and disruptive coloration.
513 *Behavioral Ecology,* **26,** 45-54.
514 Kelber, A., Vorobyev, M. & Osorio, D. (2003) Animal colour vision - behavioural tests and
515 physiological concepts. *Biological Reviews,* **78,** 81-118.
516 Kesner, R.P. & Olton, D.S. (2014) *Neurobiology of Comparative Cognition*. Psychology Press,
517 New York, NY.
518 Koch, C. & Ullman, S. (1985) Shifts in Selective Visual-Attention - Towards the Underlying
519 Neural Circuitry. *Human Neurobiology,* **4,** 219-227.
520 Livingstone, M.S. & Hubel, D.H. (1984) Anatomy and Physiology of a Color System in the
521 Primate Visual-Cortex. *Journal of Neuroscience,* **4,** 309-356.
522 Lovell, P.G., Ruxton, G.D., Langridge, K.V. & Spencer, K.A. (2013) Egg-Laying Substrate
523 Selection for Optimal Camouflage by Quail. *Current Biology,* **23,** 260-264.
524 Loyau, A., Gomez, D., Moureau, B.T., Thery, M., Hart, N.S., Saint Jalme, M., Bennett, A.T.D. &
525 Sorci, G. (2007) Iridescent structurally based coloration of eyespots correlates with
526 mating success in the peacock. *Behavioral Ecology,* **18,** 1123-1131.
527 Ma, C.Y. & Hang, H.M. (2015) Learning-based saliency model with depth information.
528 *Journal of Vision,* **15**.
529 Merilaita, S. (2003) Visual background complexity facilitates the evolution of camouflage.
530 *Evolution,* **57,** 1248-1254.
531 Navalpakkam, V. & Itti, L. (2006) An integrated model of top-down and bottom-up attention
532 for optimizing detection speed. *2006 IEEE Computer Society Conference on Computer*
533 *Vision and Pattern Recognition (CVPR'06)*, pp. 2049-2056.
534 Osorio, D., Miklosi, A. & Gonda, Z. (1999) Visual ecology and perception of coloration
535 patterns by domestic chicks. *Evolutionary Ecology,* **13,** 673-689.
536 Osorio, D. & Vorobyev, M. (2005) Photoreceptor spectral sensitivities in terrestrial animals:
537 adaptations for luminance and colour vision. *Proceedings of the Royal Society B-*
538 *Biological Sciences,* **272,** 1745-1752.
539 Osorio, D. & Vorobyev, M. (2008) A review of the evolution of animal colour vision and
540 visual communication signals. *Vision Research,* **48,** 2042-2051.
541 Pal, S.K. & Peters, J.F. (2010) *Rough Fuzzy Image Analysis: Foundations and Methodologies*.
542 CRC Press, Boca Raton, FL.
543 Peters, R.A. (2010) Movement-based signalling and the physical world: modelling the
544 changing perceptual task for receivers. *Modelling perception with artifical neural*
545 *networks* (eds C.R. Tosh & G.D. Ruxton), pp. 269-292. Cambridge University Press,
546 Cambridge.
547 Petrie, M. & Halliday, T. (1994) Experimental and Natural Changes in the Peacocks (Pave
548 Cristatus) Train Can Affect Mating Success. *Behavioral Ecology and Sociobiology,* **35,**
549 213-217.
550 Petrie, M., Halliday, T. & Sanders, C. (1991) Peahens Prefer Peacocks with Elaborate Trains.
551 *Animal Behaviour,* **41,** 323-331.

552 Rosenholtz, R., Li, Y., Mansfield, J. & Jin, Z. (2005) Feature congestion: a measure of display
553 clutter. *Proceedings of the SIGCHI conference on human factors in computing*
554 *systems, ACM*, pp. 761-770.
555 Russ, J.C. & Neal, F.B. (2016) *The Image Processing Handbook,* 7th edn. Taylor & Francis,
556 Boca Raton, FL.
557 Ruxton, G.D., Sherratt, T.N. & Speed, M.P. (2004) *Avoiding Attack*. Oxford University Press,
558 Oxford, UK.
559 Schaefer, H.M. & Stobbe, N. (2006) Disruptive coloration provides camouflage independent
560 of background matching. *Proceedings of the Royal Society B-Biological Sciences,* **273,**
561 2427-2432.
562 Spottiswoode, C.N. & Stevens, M. (2010) Visual modeling shows that avian host parents use
563 multiple visual cues in rejecting parasitic eggs. *Proceedings of the National Academy*
564 *of Sciences of the United States of America,* **107,** 8672-8676.
565 Stevens, M. & Cuthill, I.C. (2006) Disruptive coloration, crypsis and edge detection in early
566 visual processing. *Proceedings of the Royal Society B,* **273,** 2141-2147.
567 Stevens, M. & Merilaita, S. (2009) Animal camouflage: current issues and new perspectives.
568 *Philosophical Transactions of the Royal Society B-Biological Sciences,* **364,** 423-427.
569 Stoddard, M.C., Kilner, R.M. & Town, C. (2014) Pattern recognition algorithm reveals how
570 birds evolve individual egg pattern signatures. *Nature Communications,* **5,** 4117.
571 Thery, M. & Casas, J. (2002) Predator and prey views of spider camouflage - Both hunter and
572 hunted fail to notice crab-spiders blending with coloured petals. *Nature,* **415,** 133-
573 133.
574 Troscianko, J., Skelhorn, J. & Stevens, M. (2017) Quantifying camouflage: how to predict
575 detectability from appearance. *BMC Evolutionary Biology,* **17,** 7.
576 Troscianko, J., Wilson-Aggarwal, J., Stevens, M. & Spottiswoode, C.N. (2016) Camouflage
577 predicts survival in ground-nesting birds. *Scientific Reports,* **6**.
578 Uy, J.A.C. & Endler, J.A. (2004) Modification of the visual background increases the
579 conspicuousness of golden-collared manakin displays. *Behavioral Ecology,* **15,** 1003-
580 1010.
581 Vorobyev, M. & Osorio, D. (1998) Receptor noise as a determinant of colour thresholds.
582 *Proceedings of the Royal Society B,* **265,** 351-358.
583 Walther, D. & Koch, C. (2006) Modeling attention to salient proto-objects. *Neural Networks,*
584 **19,** 1395-1407.
585 Webster, R.J., Hassall, C., Herdman, C.M., Godin, J.G.J. & Sherratt, T.N. (2013) Disruptive
586 camouflage impairs object recognition. *Biology Letters,* **9**.
587 Xiao, F. & Cuthill, I.C. (2016) Background complexity and the detectability of camouflaged
588 targets by birds and humans. *Proceedings of the Royal Society B,* **283,** 20161527.
589 Zhang, H.L., Lei, J.J., Fan, X.H., Wu, M.M., Zhang, P. & Bu, S.P. (2012) Depth Combined
590 Saliency Detection Based on Region Contrast Model. *Proceedings of 2012 7th*
591 *International Conference on Computer Science & Education, Vols I-Vi***,** 763-766.
592 Zhang, Y., Jiang, G.Y., Yu, M. & Chen, K. (2010) Stereoscopic Visual Attention Model for 3D
593 Video. *Advances in Multimedia Modeling, Proceedings,* **5916,** 314-324.
594 Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010) A protocol for data exploration to avoid common
595 statistical problems. *Methods in Ecology and Evolution,* **1,** 3-14.

596

597

598 Table 1. Relative performance of the various metrics used to quantify target
599 conspicuousness, in terms of their ΔAIC. Please refer to the supplementary material for a
600 full description of these metrics and details of the analysis. For each experiment, the best-
601 fitting model is denoted by an asterisk (*).

Predictor	ΔAIC (Experiment 1)	ΔAIC (Experiment 2)
This Model (equal weighting of feature types)	0.9	11.2
This Model ('optimal' weighting of feature types)	0.0*	1.5
Gabor Edge Disruption Ratio	10.6	20.0
Number of SIFT Feature Correspondences	27.4	18.3
Colour Congestion	17.2	18.9
Luminance Congestion	16.7	4.8
Orientation Congestion	19.0	11.8
Overall Feature Congestion	19.0	11.4
Sub-band Entropy	20.3	0.0*
Number of Hough Edges	23.6	18.9

602
603
604
605
606
607
608
609

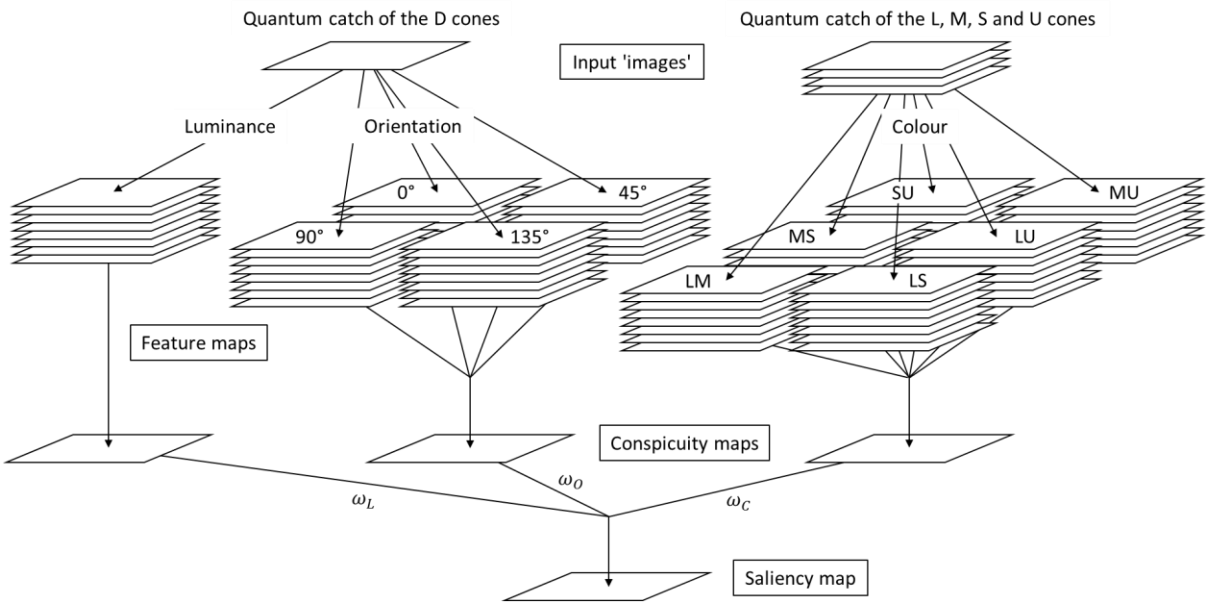


Figure 1. Schematic illustration of the saliency model used here, adapted from Itti, Koch and Niebur (1998). Input to the model is a series of grayscale 'images', each representing topographical variation in the estimated quantum catch of one of the viewing bird's cone classes; one for each of the four single cones (L, M, S and U, which are assumed to contribute to the perception of colour) and one for the double cones (D, which are assumed to encode luminance). These are used to construct feature maps that encode local colour, luminance and orientation contrasts, before being aggregated hierarchically, first by grouping features by type into conspicuity maps, then by combining these conspicuity maps (using the weights ω_L, ω_C and ω_O) to compute the final saliency map. Please refer to the text for full details.

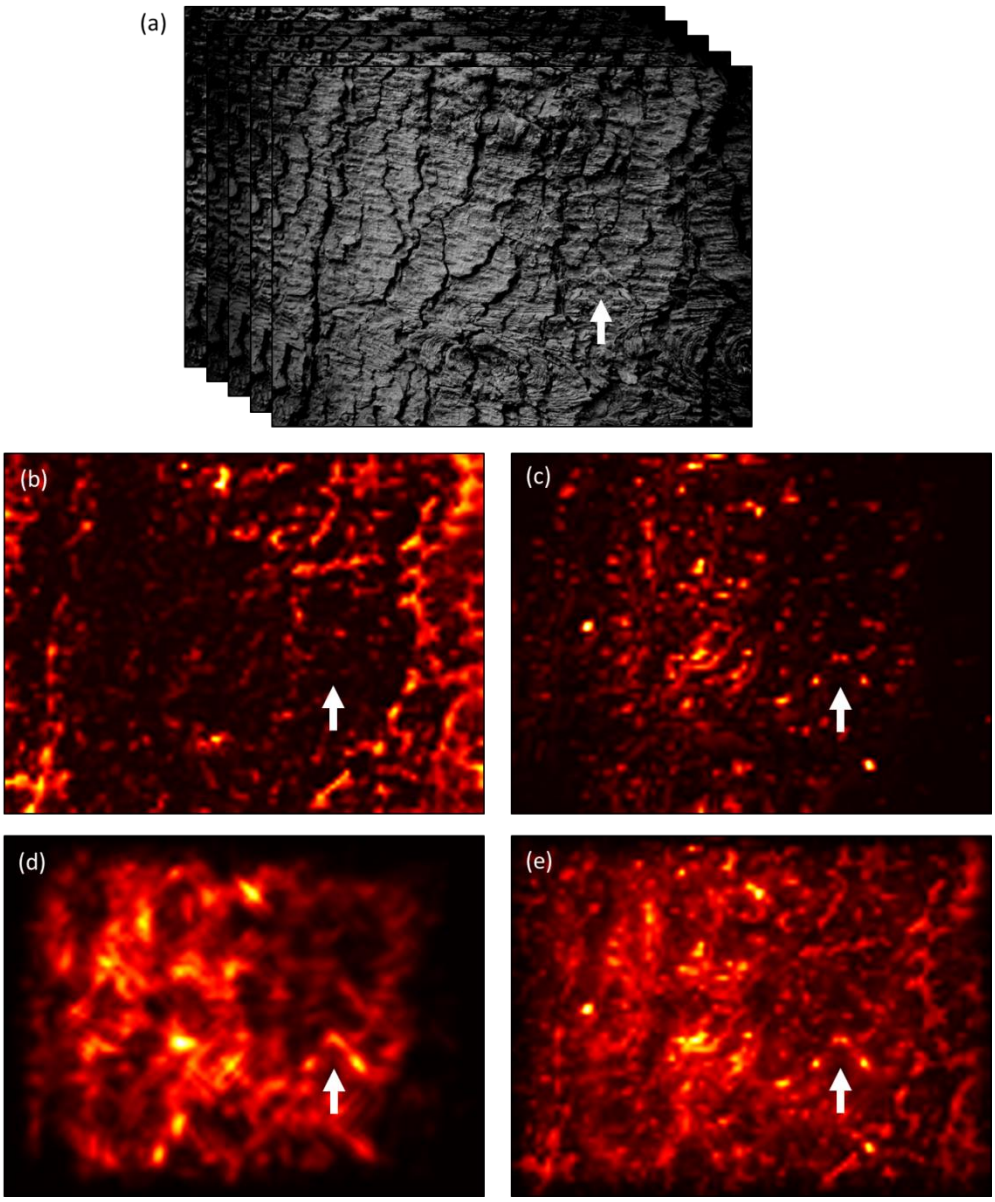
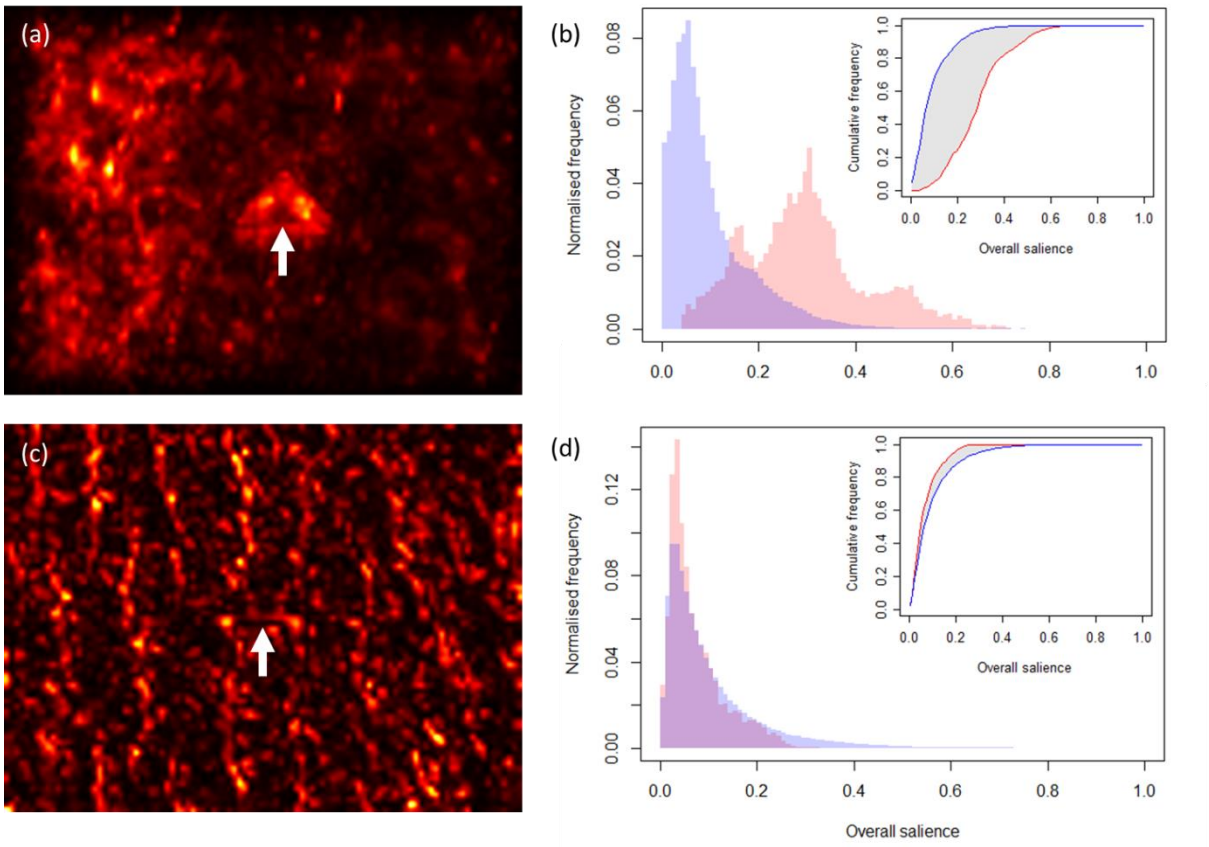


Figure 2. (a) Representative input 'images' for one of the stimuli used in Experiment 1. These were used to compute conspicuity maps for (b) colour, (c) luminance and (d) orientation, which were then combined (in this case using equal-weighting, i.e. $\omega_L = \omega_C = \omega_O = 1$) to produce the final saliency map (e). Please refer to the text for full details. In each map, colour is proportional to salience, with lighter colours denoting regions of relatively high salience and darker colours regions of relatively low salience. The camouflaged virtual moth is shown by the white arrow, and is at the same corresponding position in each map. In this example there is little evidence of colour salience in the target compared to its background. However, some elements of the target's pattern are relatively salient in the luminance channel (seen as blobs of high salience corresponding to the positions of brighter regions on the outer edge of the wings), and edges that differ in direction from the

surrounding background are clearly evident in the orientation channel. Combined, these features contribute to the overall salience of the target. Note that in each map the target is not necessarily the only (or most) salient region, but its salience is sufficiently high to likely make it conspicuous against this particular background to this particular predator.

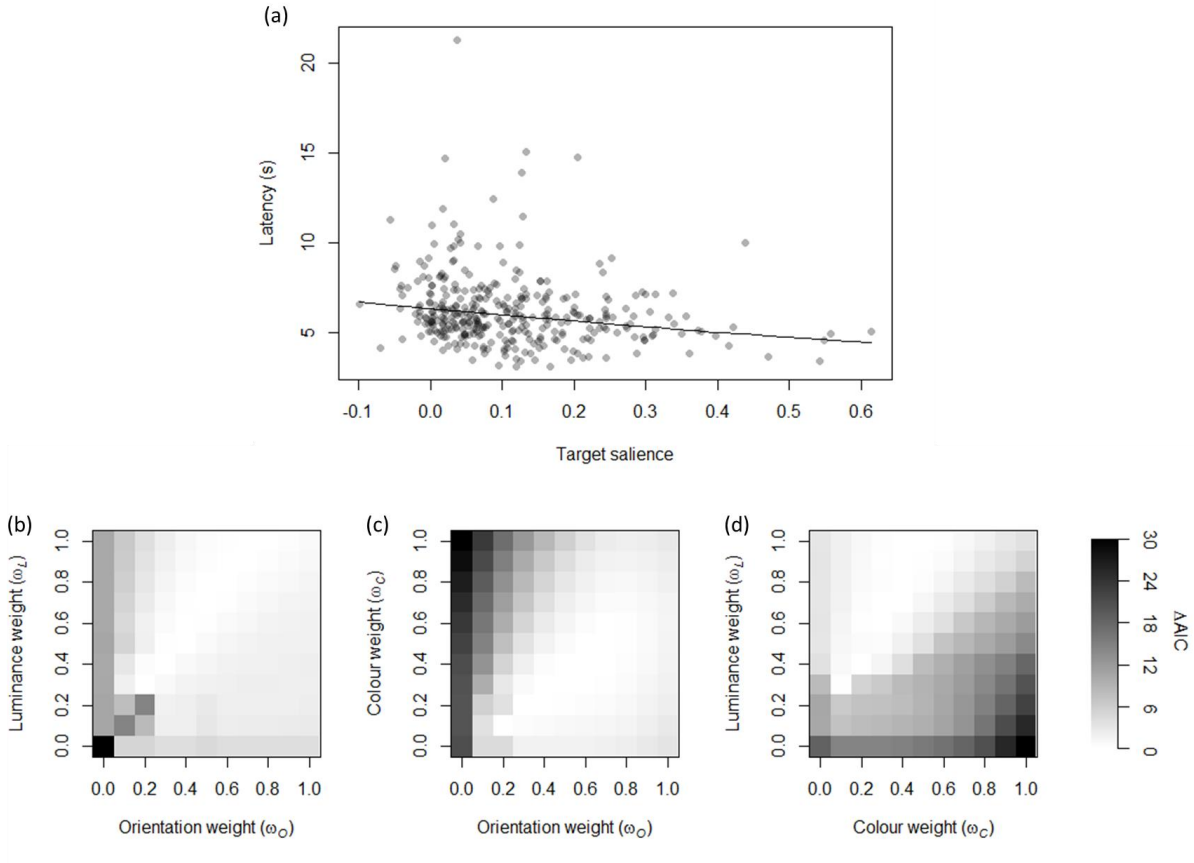


667

668 Figure 3. Calculation of target salience, illustrated using representative stimuli from
669 Experiment 2. (a) Overall saliency map, which includes a relatively salient moth-like target
670 (indicated by the white arrow). Colour is proportional to salience, with lighter colours
671 denoting regions of relatively high salience and darker colours regions of relatively low
672 salience. (b) Frequency histogram of pixel salience values for the background (blue) and
673 target (red) of the scene shown in (a), with the region of overlap shown in purple. Both
674 histograms have been normalised to aid comparison. The inset shows the cumulative
675 histogram of these data, with the grey shaded region indicating the difference between
676 histograms from which relative target salience was calculated. Please refer to the text for
677 full details. (c) Overall saliency map including a relatively unsalient moth-like target
678 (indicated by the white arrow), along with (d) the corresponding frequency and cumulative
679 histograms of pixel salience values for the background (blue) and target (red).

680

681



682

683 Figure 4. (a) Time taken for Japanese quail to successfully predate virtual moths, as a
684 function of target salience. Each data point represents one moth, and data from all birds
685 have been combined for clarity. The solid line denotes the estimated fit from the linear
686 mixed-effects model. For simplicity, the three feature types (luminance, colour and
687 orientation) were assumed to contribute equally to the computation of target salience (i.e.
688 $\omega_L = \omega_C = \omega_O = 1$). (b-d) Variation in model ΔAIC as the relative weight of the luminance,
689 colour and orientation features types was systematically changed. Grey values denote the
690 minimum ΔAIC score for the given combination of weights, with lighter shades indicative of
691 better fitting models.

692

693

694

695

696

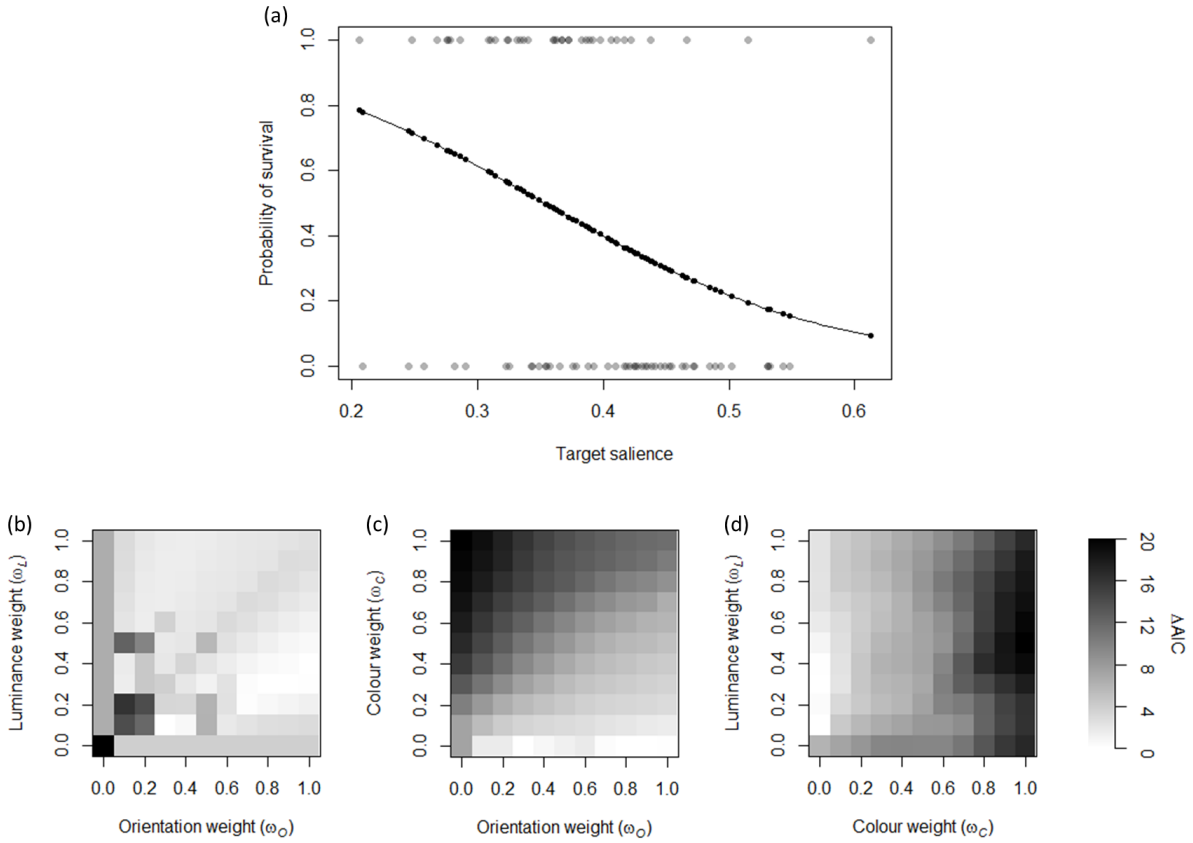


Figure 5. (a) The probability that moth-like targets survived predation by wild birds over a 24 hour period as a function of their salience. Individual data points represent a single target, and the curve represents the fit of the binomial generalized linear mixed model used to model the data. For simplicity, the three feature types (luminance, colour and orientation) were assumed to contribute equally to the computation of target salience (i.e. $\omega_L = \omega_C = \omega_O = 1$). (b-d) Variation in model ΔAIC as the relative weight of the luminance, colour and orientation features types was systematically changed. Grey values denote the minimum ΔAIC score for the given combination of weights, with lighter shades indicative of better fitting models.

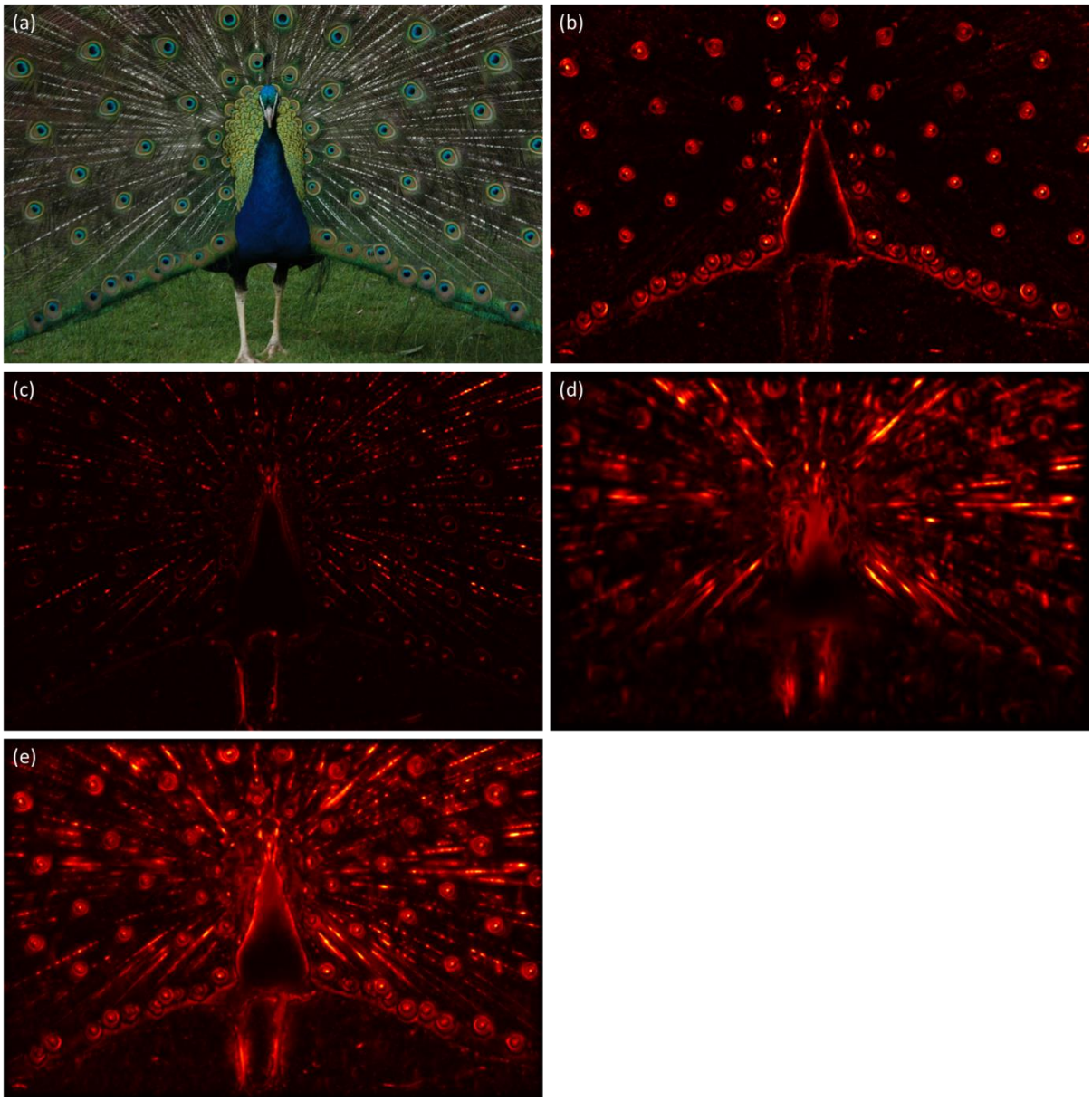


Figure 6. (a) Calibrated colour image of a displaying peacock (*Pavo cristatus*), and the conspicuity maps for (b) colour, (c) luminance and (d) orientation that result from applying the model of visual salience used in this paper. (e) The final overall saliency map. In each map, colour is proportional to salience, with lighter colours denoting regions of relatively high salience and darker colours regions of relatively low salience. The procedure used was as described for experiment 2, but using data on the peafowl's photoreceptor spectral sensitivity from Hart (2002).