# Deep Neural Network and field experiments reveal how transparent wing windows reduce detectability in butterflies

Mónica Arias*[1,2], Cynthia Tedore*[1,3], Marianne Elias[2], Lucie Leroy[1], Clément Madec[1], Louane Matos[1], Julien P. Renoult[1], Doris Gomez[1, 4]

Affiliations:

[1] CEFE, CNRS, Univ. Montpellier, Univ. Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

[2] ISYEB, CNRS, MNHN, Sorbonne Univ., EPHE, Univ. Antilles, 45 rue Buffon CP50, Paris, France

[3] Univ. Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Institute of Zoology, Hamburg, Germany

[4] INSP, CNRS, Sorbonne Univ., Paris, France

*first co-authors

Corresponding author email: moarias@gmail.com

**Abstract**

Lepidoptera – a group of insects in which wing transparency has arisen multiple times - exhibit much variation in the size and position of transparent wing zones. However, little is known as to how this variability affects detectability. Here, we test how the size and position of transparent elements affect predation of artificial butterflies by wild birds in the field. We also test whether deep neural networks (DNNs) might be a reasonable proxy for live predators, as this would enable one to rapidly test a larger range of hypotheses than is possible with live animals. We compare our field results with results from six different DNN architectures (AlexNet, VGG-16, VGG-19, ResNet-18, SqueezeNet, and GoogLeNet). Our field experiment demonstrated the effectiveness of transparent elements touching wing borders at reducing detectability, but showed no effect of transparent element size. DNN simulations only partly matched field results, as larger transparent elements were also

harder for DNNs to detect. The lack of consistency between wild predators' and DNNs' responses raises questions about what both experiments were effectively testing, what is perceived by each predator type, and whether DNNs can be considered to be effective models for testing hypotheses about animal perception and cognition.

**Keywords**

background matching, disruptive coloration, transparency, wild predators, birds, artificial prey, deep learning

## Introduction

The coevolutionary arms race between prey and predators has generated some of the most striking adaptations in the living world, including lures, mimicry and camouflage in prey, and highly sophisticated detection systems in predators (Ruxton et al., 2004). Transparency, by definition, constitutes the perfect background matching against virtually all types of backgrounds. Transparency is common in pelagic environments where there is no place to hide (Johnsen, 2001), and often concerns the entire body because: 1) small differences in refractive index between water and biological tissues limit reflections that would betray the presence of a prey, 2) ultraviolet-protective pigments are not necessary in the depths of oceans because UV radiation is filtered out by water and, 3) water provides support, limiting the need for thick, sustaining (often opaque) tissues. Contrary to aquatic species, transparency on land has only evolved as "elements" (i.e. parts of their body). It is currently unknown which visual configurations of transparent and opaque elements are the best at reducing prey detectability on land.

Unlike in aquatic taxa, transparency has evolved in few terrestrial groups, and is mostly restricted to insect wings. Within the Lepidoptera, however, transparency has evolved several times independently in typically opaque-winged butterflies and moths. According to recent experimental evidence, the presence of well-defined transparent windows reduces detectability in both conspicuous (Arias et al., 2019) and cryptic (Arias et al., 2020) terrestrial

prey. However, wing transparency can occur in different forms (Fig. 1): as small windows (e.g., as in the moths *Attacus atlas*, or *Carriola ecnomoda*) or as large windows (e.g., as in the sphinx *Hemaris fuciformis*) delimited by brownish elements or conspicuous wing borders (e.g., as in the Ithomiine tribe). Wing transparency can also be associated with discontinuous borders, touching prey edges (e.g., as in the elytra of the tortoise beetle *Aspidomorpha miliaris* or in moths of the genus *Bertholdia*). Whether the size of transparent windows or their disrupting contour have any effect on prey detectability remains to be studied.

This diversity of forms suggests that including transparent elements can reduce detectability by different mechanisms. For example, a transparent window surrounded by an opaque border could reduce detectability by increasing background resemblance. The larger the transparent area, the larger proportion of the surface area of the prey that matches its background ("background matching hypothesis"). On the other hand, transparent elements breaking the prey outline can hamper detection as a form of disruptive coloration (Cuthill et al., 2005; Stevens and Cuthill, 2006). In disruptively coloured prey, colours that imitate the background are combined with contrasting elements that are located on the prey border, breaking its visual edges and making it difficult to detect ("disruptive coloration hypothesis"). Broken edges have been shown to work better than background matching (Cuthill et al., 2005), especially when low and intermediate colour contrast elements are included (Stobbe and Schaefer, 2008). Additionally, transparent elements touching prey borders can make the prey appear to be a smaller size than it actually is, reducing its perceived profitability and its risk of being attacked, as predation rate is directly correlated with prey size (Berger et al., 2006). To understand the relative importance of these different mechanisms in partially transparent terrestrial prey, we need to investigate the effect of size and position of transparent elements.

Field experiments with artificial prey models are a common way to explore the behaviour of natural predators in response to the visual characteristics of prey and thus to infer the perceptual mechanisms of predation avoidance. This approach has been used, for instance, to reveal the defence advantages of disruptive coloration (Cuthill et al., 2005),

which is more efficient when less symmetrical (Cuthill et al., 2006) and in the absence of marks attracting predators' attention to non-vital body parts (Stevens et al., 2013). However, field experiments are time-consuming and labor-intensive (from 11 to 20 weeks per experiment in the field to obtain attack rates ranging from 11.3% to 18% (Arias et al., 2016; Nordberg and Schwarzkopf, 2019; Rönkä et al., 2020)), which reduces the combination of conditions that can be tested simultaneously.

Artificial intelligence algorithms, and deep neural networks (DNNs) in particular, are becoming common tools for tackling questions in ecology and evolution while reducing human labour (Christin et al., 2019; Norouzzadeh et al., 2018). DNNs are able to represent complex stimuli, such as animal phenotypes and their colour patterns, in a high-dimensional space in which Euclidean distances can be calculated (Charpentier et al., 2020; Cuthill et al., 2019; Wham et al., 2019). As in a colour space, the distance between phenotypes or patterns in the representational space of a DNN have been shown to be correlated with their perceived similarity (Zhang et al., 2018).

DNNs are made up of multiple layers of neurons that lack interconnections within layers but are connected to neurons in adjacent layers. The firing intensity of a neuron depends on the strength of signal received by connected upstream neurons and the neuron's nonlinear activation function. Learning comes about by iteratively tweaking the weight of each neuron's output to downstream neurons until a more accurate final output is attained. Although the architecture and computational mechanics of artificial neural networks were originally inspired by biological neural networks, the primary goal of most DNNs today is to produce accurate outputs (in our case, categorizations) rather than to mimic the mechanics of biological neural networks. Indeed, this prioritization of functionality over mechanistic mimicry is true for the DNNs used in the present study. Interestingly, however, the spatial distribution of different classes of objects in the representational space of DNNs has been shown in various studies to be correlated with that of the primate visual cortex (Cadieu et al., 2014; Güçlü and van Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014; Mély and Serre, 2017; Renoult et al., 2019; Yamins et al., 2014). This is suggestive of some

degree of mechanistic similarity in the way that artificial and biological neural networks encode and categorize objects. Beside primates, the ability of DNNs to predict the responses of animals remains almost untested (but see (Fennell et al., 2019)).

Here, we explored the perceptual mechanisms by which transparent wing windows decrease predation, using wild avian predators and DNNs as observers of artificial moths. As both types of observers were naïve to transparent butterflies, our study also shed light on the selective advantages that could favour the evolution of these mechanisms. First, we carried out field experiments to analyze the effect of different visual characteristics of transparent windows and adjacent elements on the attack of prey by avian predators. We predicted that larger transparent windows touching prey outlines would make prey more difficult to detect, presumably *via* both background matching and disruptive coloration. Second, we studied the ability of DNNs to detect moths with varying sizes and configurations of transparent elements, aiming to compare DNN responses to the behaviour of real predators. Finally, using DNNs, we explored which characteristics of transparent elements might promote the establishment of new cryptic morphs in a given population. Our study sheds light on how well predator behaviour can be predicted by artificial intelligence.

## Material and methods

### Field experiments

We followed an experimental design similar to that described in (Arias et al., 2020). We performed predation experiments in April and May 2019 in two localities in southern France: La Rouvière forest (43.65°N, 3.64°E) and the Montpellier zoo (43.64°N, 3.87°E), for three 1-week sessions at each place. We monitored artificial prey survival from predation by bird communities that include great tits (*Parus major*), blue tits (*Cyanistes caeruleus*), Eurasian jays (*Garrulus glandarius*), common chaffinches (*Fringilla coelebs*), golden orioles (*Oriolus oriolus*) and European robins (*Erithacus rubecula*). Artificial prey (body and wings) were pinned on green oak (*Quercus ilex*) tree trunks (>10cm in diameter, with little or no moss

cover) every 10m. To deter ant attacks, we applied Vaseline and double-sided transparent tape between the wings and the trunk. Prey were placed perpendicularly to the ground and mostly on the north-facing side of tree trunks to reduce direct sunlight reflection that could facilitate their detection. We monitored prey survival once per day for the following four consecutive days after placing them on trunks, and removed them afterwards.

Artificial moths

As described by Arias *et al* (2020) and similarly to other experiments, artificial moths consisted of paper wings and an edible body (Cuthill et al., 2005; Stevens and Cuthill, 2006). Wings consisted of right triangles resembling resting generic moths (i.e., not representing any real species). Triangle dimensions were 25mm height by 36mm width, thus a surface area of 450mm². To determine wing colouration, we performed reflectance measurements of green oak trunk colouration (120 measurements on 6 trunks) and laminated grey paper using a deuterium halogen lamp (Avalight DHS) emitting over 300-700 nm range including UV to which birds are sensitive (Chen and Goldsmith, 1986), a spectrophotometer (Starline Avaspec-2048 L), an optic probe with tip cut at 45° (FC-UV200-2-1.5 x 100, Avantes), and a white diffuse reference (spectralon, WS2). We calculated colour and brightness contrasts between paper and trunk as seen by birds by applying Vorobyev and Osorio discriminability model (Vorobyev and Osorio, 1998) using *pavo* package with R software (Maia et al., 2013). We modeled the discriminality for both UVS vision (blue tit, (Hart et al., 2000)) and VS vision (shearwater, (Hart, 2001)). We used a forest shade light environment (Gomez and Théry, 2007), a Weber fraction of 0.1 for the chromatic response (as reported for Pekin robin *Leiothrix lutea* in (Gomez and Théry, 2007; Maier and Bowmaker, 1993) and 0.2 for the brightness response (as the average reported values for known bird species (Lind et al., 2013)). We found that colour Grey155 (R=G=B=155), printed on Canson® sketch paper with a HP officejet pro 6230 printer was chromatically indistinguishable from trunk coloration (chromatic contrast < 1JND), and marginally lighter than oak trunks (achromatic contrast ≤

1.64 JND). This colour thus enabled us to explore the effect of transparent windows and their spatial configuration on reducing detectability of cryptic prey.

We analysed five types of artificial grey moths with different wing characteristics (Fig. 1): opaque (O morph), with small transparent windows within the wing (SW morph), with large transparent windows within the wing (LW morph), with large transparent windows touching the bottom edge of the wing (BE morph) and with large transparent windows touching all three wing edges (as each window is touching two edges of the three butterfly edges, B3E morph). Morphs that included transparent elements were built by cutting two right triangular windows out of the laminated grey triangle. Dimensions for each triangular window were 7 mm width by 10 mm height (transparency occupying thus 15% of original grey surface) for the SW morph, and 14mm width by 18mm height for the LW, BE and B3E morphs (transparency occupies thus 46.6% of originally grey surface). To simulate transparent wing surfaces, we then added a transparent film (3M for inkjet, chosen for its high transparency even in the UV range see ESM, Fig S1) underneath the remaining parts. On top of the moth wings, we added an artificial body made from pastry dough (428g flour, 250g lard, and 36g water, following (Carroll and Sherratt, 2013)), dyed grey by mixing yellow, red and blue food dyes (spectrum in Fig. S1). This malleable mixture allowed us to make small bodies on which we could record marks made by bird beaks and distinguish them from those made by insects.

Data collection and analysis

During monitoring, we considered artificial moths as attacked by birds when their body showed V-shaped or U-shaped marks, or when the body was missing without signs of invertebrate attacks (i.e. no body scraps left on wings or around the moth on the trunk). We removed all remains of artificial moths attacked by birds, and replaced them when attacked by invertebrates or when the entire artificial prey (wings, body and pin) were missing, as we could not exclude that the prey item fell down or was blown away by the wind. Non-attacked prey were treated as censored data in the analyses (i.e. prey that survived at least until the

end of the experiment). We analysed prey survival, fitting three independent Cox proportional hazard regressions (Cox, 1972), using as explanatory variables one of the three different ways to describe the visual characteristics of morphs: 1) per morph (O, SW, LW, BE, B3E), 2) per border characteristics (having a full wing border (O, SW, LW) or a border broken by transparent windows in one (BE) or three edges (B3E)), and 3) per transparent surface (relative size of both transparent windows (0 for O, 0.14 for SW and 0.57 for LW, BE and B3E)). We also included week as an explanatory variable in order to control for time in the season as well as place, as the experiment was never run simultaneously at La Rouviére forest and at the zoo during the same week. Interactions between morph or its characteristics and week were also included as explanatory variables in order to explore whether differences in locality and/or time in the breeding season might affect predation on different morphs. Overall significance was measured using a Wald test. Statistical analyses were performed in R (R Foundation for Statistical Computing, 2014) using the *survival* package (Therneau and Lumley, 2009). Using the Cox survival model results, we performed pairwise post-hoc comparisons between survival of the different morphs and their characteristics. Additionally, using the Cox survival model coefficients, we produced “distance matrices” of pairwise differences of survival between morphs (survival coefficient for morph A - survival coefficient for morph B): 1) per morph, 2) per transparent surface size and 3) per border characteristics (Tables S4). Differences in attack rates between places and/or weeks could be produced by differences in predation pressure, that can be higher when more nests are occupied and/or when more chicks on their developmental peak are present in the population. Therefore, we explored whether differences in attack rate per week/locality were associated with great and blue tit chick hatching, the number of chicks between 9 and 15 days old (age when the food demand by great tit and blue tit chicks is the highest (A. Charmantier pers. comm.)), or the number of active nests or number of nests having chicks between 9 and 15 days old. We fitted linear models with attacks per day as the response variable and including all, some or none of these demographic variables as explanatory variables. We selected the best model using the AIC criterion.

Deep Neural Networks

We used deep neural networks (DNNs) for the second and third aims of this study. We trained DNNs to detect moths on tree trunks and scored their performance at detecting different moth forms. We assessed the performance of different DNNs at detecting moths, and compared the responses of DNNs and avian predators when detecting the same prey. Additionally, using DNNs, we predicted the characteristics that will render a prey morph highly cryptic and thus potentially able to establish in a population.

Prey used

We took pictures of artificial grey moths identical to those used in the field experiments. We used a smartphone mounted on a tripod to take pictures of each of five artificial moth morphs, each pinned to the same position on a tree trunk, as well as the tree trunk alone with no moth present. This controlled setup ensured that the DNNs used distinguishing characteristics of the different morphs, rather than finding unintended correlations between each morph and other characteristics of the visual background. Each morph was photographed against 222 unique visual backgrounds.

Selected DNNs and general training

Rather than training DNNs from scratch (i.e. initializing all neuronal outputs with random weights), we used DNNs pre-trained to correctly identify 1000 object categories from more than a million images from the ImageNet database. Pre-trained networks have already learned to detect image features common to many classes of objects, and have pre-defined neuronal weights that are useful for learning new combinations of features. We took such DNNs as a starting point, and then used transfer learning to train them on the new task of detecting artificial moths. Transfer learning replaces the last few layers of an existing DNN with new learning and classification layers that rapidly adapt to a new classification task after training on only a limited dataset. To test the robustness of results obtained from DNNs, we used six different DNNs that were similarly trained: AlexNet, VGG-16, VGG-19, ResNet-18, SqueezeNet, and GoogLeNet. Weights were frozen in the first ten layers to preserve simple

feature detectors already learned from the ImageNet database. The initial learning rate in subsequent layers was set to $2*10^{-4}$, and was increased by a factor of ten in the final, fully-connected layer. The $L_2$ regularization term was set to 0.0001. Different solvers, which identify optimal weights, were used for different DNNs to maximize the learning capability (measured as the test set accuracy) of the different DNN architectures. The adaptive moment estimation (Adam) solver was used for AlexNet, ResNet-18, SqueezeNet, and GoogLeNet, and the stochastic gradient descent with momentum (sgdm) solver was used for VGG-16 and VGG-19. To avoid overfitting DNNs to the training set, training was halted when the loss on the validation set failed to decrease three times. At this point, training was considered complete. After training and testing, each DNN returned a score between 0 and 1 indicating its level of confidence that there was a moth in each picture. This score was determined by a sigmoid binary classifier (softmax layer in the DNN). Images with scores greater than 0.5 were classified as displaying a moth. To verify that DNNs were using features of the wings, and not the body alone, to detect butterflies, we generated class activation maps of the final layer of the DNNs. These maps indicated which parts of the image contributed the most towards determining that the image contained a butterfly. SqueezeNet, ResNet-18, and GoogLeNet can produce such maps, but AlexNet, VGG-16, and VGG-19 cannot due to the fact that their final layer consists of multiple interconnected layers. All training and testing of DNNs and generation of class activation maps was conducted using MATLAB (Tha MathWorks In., 2019) (see Figshare for the MATLAB code and images used).

We used two training methods to explore two different questions: 1) in a community of different moth morphs, all of which are familiar to predators, which morphs will be the most successful at evading predation? and, 2) if predators are only experienced with opaque morphs, which novel morph(s) with transparent elements will most easily establish themselves in the population? (Fig. 3).

*Detection of familiar morphs by DNNs*

We trained DNNs to detect moths using all morphs at once to test the effect of transparent elements at reducing detectability among familiar prey. Out of the 222 images of each morph, the training set included 144 randomly-selected images of each of the five morphs and each corresponding background image in quintuplicate. We included each background in quintuplicate so that there would be the same number of positive (with a moth) and negative stimuli (without a moth) in the training set. Thus, in total, 1,440 images were used in training. Thirty-nine randomly-selected images of each morph and their corresponding background images in quintuplicate were used in the validation set (for a total of 390 images), and the remaining 39 images (out of 222) of each morph and the background were used in the test set (for a total of 234 images). We repeated this training and testing procedure 100 times, each time retraining the same initial DNN (pretrained on the ImageNet database) and randomly selecting which images (out of the full 222 images available per morph) were to be allocated for training, for validation, and for testing. Minibatch size (the subset of images used during one iteration of training) was set to 80 for AlexNet, ResNet-18, SqueezeNet, and GoogLeNet, and to 70 for VGG-16 and VGG-19 (lower in these latter two due to computer memory limitations). Each minibatch contained eight (for AlexNet, ResNet-18, SqueezeNet, and GoogLeNet) or seven (for VGG-16 and VGG-19) sets of ten images. Each set of ten images had the same background, five of which contained one each of the O, SW, LW, BE, and B3E morphs. The remaining five in each set of ten was the same background image in quintuplicate. Minibatch sizes during training were larger for this than for the training method described below to keep the number of different backgrounds in each minibatch the same under each training scenario.

*Which new morph would establish in a population?*

In the second experiment, we trained DNNs to discriminate between pictures with a moth present on a trunk (O morph only) and pictures of the background only. The aim was to simulate the reaction of predators familiar with opaque prey and naive to prey exhibiting transparent elements. This approach allowed us to test which new visual configuration is less often detected, and thus potentially more likely to establish in a population. Each time,

the training set of images consisted of 144 randomly-selected images of the O morph and 144 images of the corresponding backgrounds (with no moth present), the validation set included 39 randomly-selected images of the O morph and 39 images of the corresponding background, and the test set consisted of the remaining 39 (out of 222) images of the O morph, 39 images of the corresponding background and 222 images of each of the other morphs. As before, in each of these sets of images, the positive and negative stimuli had the same visual backgrounds; only the morph type varied. The minibatch size was set to 16, and each minibatch contained eight O morph images and eight background images that had the same backgrounds as the O morph images. All other training parameters were identical to those above.

Analyses of DNNs results

First we compared the DNNs' performances at moth detection under different training scenarios by comparing their false positive rates (proportion of background images with scores higher than 0.5). Additionally, and for each training method, we computed summary statistics (mean and standard deviation) of DNN scores per morph and DNN architecture. As the DNNs' performances were rather similar (Table S2), we pooled them together. We created a binomial response variable that was set to one for DNN scores higher than 0.5 and set to zero if it did not. We fitted a Generalised Linear Mixed model assuming a binomial distribution, using a binary response variable describing whether a moth was detected (1) or not (0) in each DNN replicate. The explanatory variable was the moth morph and had five levels: O, SW, LW, BE and B3E. Independent models were fitted for the two different experiments testing the contrast: a) O was more detected than other morphs, b) SW was more detected than LW, c) BE was more detected than B3E and d) morphs with complete windows (SW and LW) were more detected than morphs with broken edges (BE and B3E). Random effects included image ID, DNN architecture and replicate of each DNN architecture (100 replicates of six architectures). Finally, and similar to the field experiment, we produced "distance matrices" for detection differences between prey types: 1) per morph, 2) per

transparent surface size and 3) per border (Tables S4). Distances were calculated as the difference between DNNs scores per morph. Applying a Mantel test, we compared results from the field and from both DNNs experiments.

## Results

*Effect of window size and position on prey survival in the field*

555 artificial moths were attacked out of the 1583 used in the field experiment (35.1% of attack rate). We detected no significant differences in attacks between morphs when separating all the morphs (Wald test = 3.92, df = 4 $p$ = 0.4, Fig. 2). We did, however, find differences when grouping morphs according to the characteristics of their transparent windows. When transparent windows touched the border (morphs BE and B3E), prey were less attacked ($z$ = -2.03, $p$ = 0.04), suggesting that this characteristic renders morphs less detectable. However, predation of morphs sporting only different transparent window sizes was rather similar (morph SW vs morphs LW, BE, B3E: $z$ = -1.49, $p$ = 0.13). No difference in attacks was registered between the two locations (z = -2.03, $p$ = 0.73). However, more attacks were registered at the zoo at the beginning of the experiment (attacks: week 2 > week 5, $z$ = 4.41 , $p$ <0.001), and at La Rouviére forest at the end of the experiment (attacks: week 6 > week 4, $z$ = 2.75, $p$ = 0.006). We did not find any correlation between attacks per day or per week and predator demographic dynamics that could either increase or decrease predation pressure (the number of active nests, the number of chicks between 9 and 15 years old or the number of hatchings for great and blue tits (Fig. S2, best model includes none of the demographic variables to explain the variation in attack rate).

*Image elements used by DNNs to detect butterflies*

Class activation maps (CAMs) showed that DNNs used features of both the wings and body to recognize butterflies. SqueezeNet produces the highest-resolution CAMs, so we show examples of CAMs generated by SqueezeNet for each of five morphs against two different backgrounds in Fig. 4.

*DNNs' detection of moth morphs used in the field experiment*

When trained only on the “O” form, AlexNet produced the highest false positive rate (0.19). The lowest was produced by ResNet (0.01). When trained on all morphs, AlexNet again produced the highest false positive rate (0.31), while GoogLeNet produced the lowest (0.09, Table S1).

Different DNN architectures and training methods produced similar results (Fig. 5 & Table S2). When pooling together all DNNs, we found that “O” was the easiest morph to detect in both experiments (detecting familiar prey experiment: $z = 9.8$, $p<0.001$; detecting novel prey experiment: $z = 19.06$, $p<0.001$). Having large windows reduced prey detectability more than having small windows (detecting familiar prey experiment: $z = 4.66$, $p<0.001$; detecting novel prey experiment: $z = 11$, $p<0.001$, Table 1). Morphs having windows touching three edges were more difficult to detect than morphs touching only one edge especially when detecting familiar prey (detecting familiar prey experiment: $z = 5.67$, $p<0.001$; detecing novel pre experiment: $z = 2.37$, $p=0.02$). Finally, morphs having complete windows (pooling together SW and LW) were slighlty more detected than those with the broken edges (pooling together BE and B3E, detecting familiar prey experiment: $z = 2.96$, $p=0.003$, 83% detected for complete windows vs 80% for broken edges; detecting novel prey experiment: $z = 8.71$, $p<0.001$, 62.3% detected for complete windows vs 48.3% detected for broken edges).

*Comparisons between birds and DNNs*

For both types of observers, the O form was among the easiest forms to detect, while forms touching 3 edges were among the more difficult morphs to detect. However, differences in the detectability of different morphs were not similar between field experiments and DNNs detecting either familiar or novel prey (Table S3 for Mantel tests comparing field experiments and all DNNs mean score per morph). When prey forms are pooled according to window surface area or the number of broken edges, field work experiments and DNNs also produce different results (Tables S3 and S4).

## Discussion

<u>Important features of transparent elements</u>

By studying prey detection by natural avian predators, we found that in addition to their presence, the position of transparent elements is key at reducing prey detectability. Moths with large transparent elements touching wing edges were more efficient at reducing detectability in our field experiment. Our results are consistent with the findings of (Schaefer and Stobbe, 2006) that breaking prey contour is more efficient at reducing prey detectability than 1) having a uniform colour that matches one of the several colours present in the background, and 2) sporting elements that represent a random sample of the background, which is similar to the visual effect created by large transparent elements. Surprisingly, however, the disruption of wing outline does not seem to be the dominant strategy used by butterflies and moths exhibiting transparent wing elements. Several clearwing Lepidoptera have well defined wing borders, suggesting that factors other than reducing detection by predators are also shaping transparency evolution. Transparency in wings is sometimes restricted to small areas, and/or is combined with highly visible elements, suggesting an important role in visual communication, beyond crypsis, as in several unpalatable comimics of the Ithomiinae tribe (Pinna et al., 2020). It is also possible that in Lepidoptera, evolving large transparent areas touching wing edges is constrained by factors linked to wing strength or flight dynamics. Whether transparent zones without borders are indeed more fragile or sensitive to wind and/or water remains an open question.

<u>Can DNNs predict avian predation?</u>

Both natural predators and DNNs detected “O” the easiest, whereas “B3E” was among the hardest morphs to detect for both observers. This suggests that transparent elements breaking several prey edges should indeed be favoured as they are more difficult for predators to detect, a benefit that could ease their establishment in a cryptic opaque population. Our results also suggest that DNNs can be used to predict predators’ behaviour to a certain extent. However, their behaviours toward the other morphs were rather different during our study. Differences between morph detection is small for both types of obsevers, although theese differences are always significant for DNNs. Grouping morphs by their

characteristics also did not produce similar results between wild predators and DNNs. While broken edges reduced predation rates in the field experiment, DNNs detected a stronger effect of size rather than position of the window on decreasing detection. This difference was especially high when DNNs detected novel prey, case expected to be closer to field experiments as the local community of predators are expected to be familiar with opaque and not with transparent prey.

One potential explanation for the lack of agreement between the field and DNN results could be related to what we were actually testing in our experiments. The DNN experiment only explored the detectability of different prey items. By contrast, the field experiment not only tested the detectability of different prey items, but also the decision of the predator to attack, without the possibility to disentangle which of these factors was more important. Both detection and motivation to attack in biological neural networks are likely mediated by both top-down (e.g. neophobia, hunger, attention) and bottom-up (feature-driven) cognitive processes, whereas our DNNs make use of bottom-up processes only. This may partly explain why DNNs and birds did not show the same relative ability to detect the different morphs.

Another possibility is that the attack rate in the field experiment was too high to see differences in predation rates among morphs. The attack rate in this experiment was over double that of a similar study carried out under similar conditions (35.1% in the current study vs 14.08% in (Arias et al., 2020)). Under strong predation pressure, once the most detectable prey morphs in a given area have already been attacked, less detectable prey are more likely to be attacked. This could potentially have led to a homogenisation of attack rates among forms. Another possibility is that there was some kind of unintentional bias in the placement of butterflies in the field experiment. While in the DNN experiments all DNNs had access to images in which we were sure the butterflies were visible, this may not have been the case for all prey items in the field experiment. Natural predators may have mostly attacked the moths placed in the most visible locations. If visible locations unintentionally

contained more of certain morphs, then these morphs may have been more heavily predated upon than would otherwise be expected.

Other potential explanations for the difference between the field and DNN experiments could be related to differences in the color and achromatic contrasts seen by birds and DNNs. DNNs were fed standard RGB photos designed to mimic the colors experienced by human observers. Avian observers possess a UV cone and visible-light cones that are differently tuned from those of humans. Also unlike humans, birds experience much higher receptor noise in their achromatic double cone channel than in their single cones responsible for color vision (Olsson et al., 2018). Automatic image processing software on board the smartphone that took the pictures can be expected to have artificially enhanced achromatic contrasts beyond what would be seen by either human or avian observers.

The spatial configuration of elements seen by birds and DNNs may have also differed. Birds may search for butterflies from a different viewing angle than the butterflies were photographed from. All photographs were taken from a viewing angle perpendicular to the surface of the butterfly wings. However, birds may search while flying or walking on the trunk or the ground. Different viewing angles can be expected to impact not only the apparent configuration of opaque elements, but also the specular reflection (i.e. glare) by transparent elements. In the first set of images in Figure 4, some specular reflection can be seen from the transparent windows of the BE morph. Real butterflies with transparent elements show comparatively less specular reflection; thus, it would be interesting to run a similar set of experiments with artificial butterflies containing empty windows rather than transparent windows made from synthetic materials.

It is, of course, possible that the lack of agreement between live predators and DNNs could be due to limitations of the DNNs themselves. Although DNNs have made incredible strides in mimicking biological perception in the last decade, they are still susceptible to error in ways that computer scientists are still discovering (Serre, 2019). It was not practical to analyze all 363,000 class activation maps (CAMs) produced by our dataset, but we did

notice a strong bias towards the highest activations overlapping the anterior wing edges, with a slight bias toward the left anterior edge, regardless of whether the DNN was trained on the opaque or on all morphs, and regardless of whether the images were mirrored such that the left side became the right side and vice versa. The importance of diagonally-oriented wing edges is particularly evident in the bottom row of images in Figure 4, where moderate activations were produced in response to a diagonally-oriented tree branch viewed against the sky. It is interesting to note that the anterior wing edges of the BE morph were thicker than those of the LW morph. If the anterior wing edge was more important than the posterior wing edge for correct categorization by DNNs, but not for live predators, this may explain why DNNs detected the BE morph more easily than the LW morph, but the opposite was true of live birds.

Recent studies have shown that DNNs do not encode global object shape, but rather local object features, and perform poorly when asked to classify images composed of silhouettes without any internal object texture (Kubilius et al., 2016). Indeed, Baker et al. (2018) found VGG-19's performance to be abysmal when objects had clearly-defined shapes against a white background but incongruous internal textures. This represents a key difference between human and DNN perception and is perhaps the most likely explanation for why the DNNs' and avian predators' responses did not mirror each other. Moving forward, we would advise caution in assuming that DNNs will mimic the finer points of biological perception in simulations like ours without validating that this is true in live animals. However, we remain optimistic that studies like this one may lend some insight to computer scientists attempting to develop DNN algorithms that model biological perception more accurately.

**Conclusions**

According to both the DNN and fieldwork experiments' results, having transparent elements breaking several prey borders seems to be the most effective mechanism for reducing detectability of prey and subsequent predation. However, to be able to directly compare

DNNs' and wild predators' reactions, careful experimental design, paying special attention to what may affect their output, is needed.

## Acknowledgements

This work was funded by Clearwing ANR project (ANR-16-CE02-0012), HFSP project on transparency (RGP0014/2016) and a France-Berkeley fund grant (FBF #2015--58).

## References

**Arias, M., le Poul, Y., Chouteau, M., Boisseau, R., Rosser, N., Théry, M. and Llaurens, V.** (2016). Crossing fitness valleys: empirical estimation of a fitness landscape associated with polymorphic mimicry. *Proc. R. Soc. B* **283**, 20160391.

**Arias, M., Mappes, J., Desbois, C., Gordon, S., McClure, M., Elias, M., Nokelainen, O. and Gomez, D.** (2019). Transparency reduces predator detection in mimetic clearwing butterflies. *Functional Ecology* **33**, 1110–1119.

**Arias, M., Elias, M., Andraud, C., Berthier, S. and Gomez, D.** (2020). Transparency improves concealment in cryptically coloured moths. *Journal of Evolutionary Biology* **33**, 247–252.

**Baker, N., Lu, H., Erlikhman, G. and Kellman, P. J.** (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology* **14**, e1006613.

**Berger, D., Walters, R. and Gotthard, K.** (2006). What keeps insects small?—Size dependent predation on two species of butterfly larvae. *Evolutionary Ecology* **20**, 575.

**Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J. and DiCarlo, J. J.** (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput Biol* **10**, e1003963.

**Carroll, J. and Sherratt, T.** (2013). A direct comparison of the effectiveness of two anti predator strategies under field conditions. *Journal of Zoology* **291**, 279–285.

**Charpentier, M., Harté, M., Poirotte, C., de Bellefon, J. M., Laubi, B., Kappeler, P. and Renoult, J.** (2020). Same father, same face: Deep learning reveals selection for signaling kinship in a wild primate. *Science Advances* **6**, eaba3274.

**Chen, D.-M. and Goldsmith, T. H.** (1986). Four spectral classes of cone in the retinas of birds. *Journal of Comparative Physiology A* **159**, 473–479.

**Christin, S., Hervet, É. and Lecomte, N.** (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution* **10**, 1632–1644.

**Cox, D. R.** (1972). Models and life-tables regression. *JR Stat. Soc. Ser. B* **34**, 187–220.

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

**Cuthill, I. C., Stevens, M., Windsor, A. M. and Walker, H. J.** (2006). The effects of pattern symmetry on detection of disruptive and background-matching coloration. *Behavioral Ecology* **17**, 828–832.

**Cuthill, J. F. H., Guttenberg, N., Ledger, S., Crowther, R. and Huertas, B.** (2019). Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. *Science advances* **5**, eaaw4967.

**Fennell, J., Talas, L., Baddeley, R., Cuthill, I. and Scott-Samuel, N.** (2019). Optimizing colour for camouflage and visibility using deep learning: the effects of the environment and the observer's visual system. *Journal of The Royal Society Interface* **16**, 20190183.

**Gomez, D. and Théry, M.** (2007). Simultaneous crypsis and conspicuousness in color patterns: comparative analysis of a neotropical rainforest bird community. *the american naturalist* **169**, S42–S61.

**Güçlü, U. and van Gerven, M. A.** (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience* **35**, 10005–10014.

**Hart, N. S.** (2001). Variations in cone photoreceptor abundance and the visual ecology of birds. *Journal of comparative physiology. A, Sensory, neural, and behavioral physiology* **187**, 685–697.

**Hart, N., Partridge, J., Cuthill, I. and Bennett, A. T.** (2000). Visual pigments, oil droplets, ocular media and cone photoreceptor distribution in two species of passerine bird: the blue tit (Parus caeruleus L.) and the blackbird (Turdus merula L.). *Journal of Comparative Physiology A* **186**, 375–387.

**Johnsen, S.** (2001). Hidden in Plain Sight: The Ecology and Physiology of Organismal Transparency. *The Biological Bulletin* **201**, 301–318.

**Khaligh-Razavi, S.-M. and Kriegeskorte, N.** (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology* **10**, e1003915.

**Kubilius, J., Bracci, S. and Op de Beeck, H. P.** (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS computational biology* **12**, e1004896.

**Lind, O., Karlsson, S. and Kelber, A.** (2013). Brightness discrimination in budgerigars (Melopsittacus undulatus). *PLoS One* **8**, e54650.

**Maia, R., Eliason, C. M., Bitton, P., Doucet, S. M. and Shawkey, M. D.** (2013). pavo: an R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution* **4**, 906–913.

**Maier, E. and Bowmaker, J.** (1993). Colour vision in the passeriform bird, Leiothrix lutea: correlation of visual pigment absorbance and oil droplet transmission with spectral sensitivity. *Journal of Comparative Physiology A* **172**, 295–301.

**Mély, D. A. and Serre, T.** (2017). Towards a theory of computation in the visual cortex. In *Computational and cognitive neuroscience of vision*, pp. 59–84. Springer.

**Nordberg, E. J. and Schwarzkopf, L.** (2019). Predation risk is a function of alternative prey availability rather than predator abundance in a tropical savanna woodland ecosystem. *Scientific reports* **9**, 1–11.

**Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C. and Clune, J.** (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* **115**, E5716–E5725.

**Olsson, P., Lind, O. and Kelber, A.** (2018). Chromatic and achromatic vision: parameter choice and limitations for reliable model predictions. *Behavioral Ecology* **29**, 273–282.

**Pinna, C., Vilbert, M., Borenztajn, S., de Marcillac, W. D., Piron-Prunier, F., Pomerantz, A., Patel, N., Berthier, S., Andraud, C., Gomez, D., et al.** (2020). Convergence in light transmission properties of transparent wing areas in clearwing mimetic butterflies. *bioRxiv* 2020.06.30.180612.

**R Foundation for Statistical Computing, R. C.** (2014). *R: A language and environment for statistical computing.* Vienna, Austria.

**Renoult, J. P., Guyl, B., Mendelson, T. C., Percher, A., Dorignac, J., Geniet, F. and François, M.** (2019). Modelling the Perception of Colour Patterns in Vertebrates with HMAX. *bioRxiv* 552307.

**Rönkä, K., Valkonen, J. K., Nokelainen, O., Rojas, B., Gordon, S., Burdfield Steel, E. and Mappes, J.** (2020). Geographic mosaic of selection by avian predators on hindwing warning colour in a polymorphic aposematic moth. *Ecology Letters* **23**, 1654–1663.

**Ruxton, G. D., Sherratt, T. N. and Speed, M. P.** (2004). *Avoiding attack*. Oxford University Press.

**Schaefer, H. M. and Stobbe, N.** (2006). Disruptive coloration provides camouflage independent of background matching. *Proceedings of the Royal Society of London B: Biological Sciences* **273**, 2427–2432.

**Serre, T.** (2019). Deep learning: the good, the bad, and the ugly. *Annual Review of Vision Science* **5**, 399–426.

**Stevens, M. and Cuthill, I. C.** (2006). Disruptive coloration, crypsis and edge detection in early visual processing. *Proc Biol Sci* **273**, 2141.

**Stevens, M., Marshall, K. L., Troscianko, J., Finlay, S., Burnand, D. and Chadwick, S. L.** (2013). Revealed by conspicuousness: distractive markings reduce camouflage. *Behavioral Ecology* **24**, 213–222.

**Stobbe, N. and Schaefer, H. M.** (2008). Enhancement of chromatic contrast increases predation risk for striped butterflies. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1535–1541.

**Tha MathWorks In.** (2019). *MATLAB*. Natick, Massachusetts.

**Therneau, T. and Lumley, T.** (2009). Survival: Survival analysis, including penalised likelihood. R package version 2.35-7. *R foundation for Statistical Computing2011*.

**Vorobyev, M. and Osorio, D.** (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society B-Biological Sciences* **265**, 351–8.

**Wham, D. C., Ezray, B. and Hines, H. M.** (2019). Measuring Perceptual Distance of Organismal Color Pattern using the Features of Deep Neural Networks. *bioRxiv* 736306.

**Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D. and DiCarlo, J. J.** (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111**, 8619–8624.

**Zhang, R., Isola, P., Efros, A. A., Shechtman, E. and Wang, O.** (2018). The unreasonable effectiveness of deep features as a perceptual metric.pp. 586–595.

**Figures and Tables**

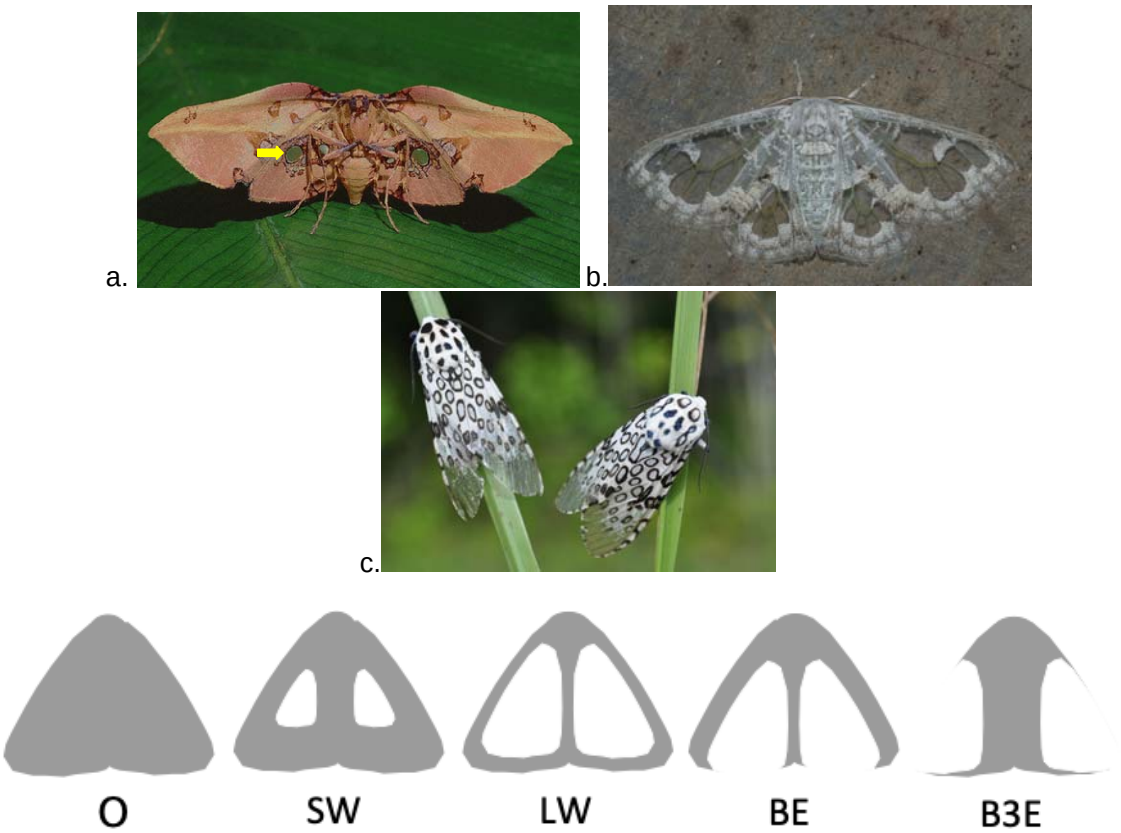


Figure 1. Examples of diversity of position and size of transparent elements in butterflies. a. *Siculodes aurorula* shows small windows (indicated by the yellow arrow, photo: © Adrian Hoskins), b. *Carriola thyridophora* shows large windows (photo: © Vijay Anand Ismavel), c. *Hypercompe scribonia* shows windows breaking the bottom edge (photo: © Green Futures). In the third row, 5 morphs of artificial moths used in this study.

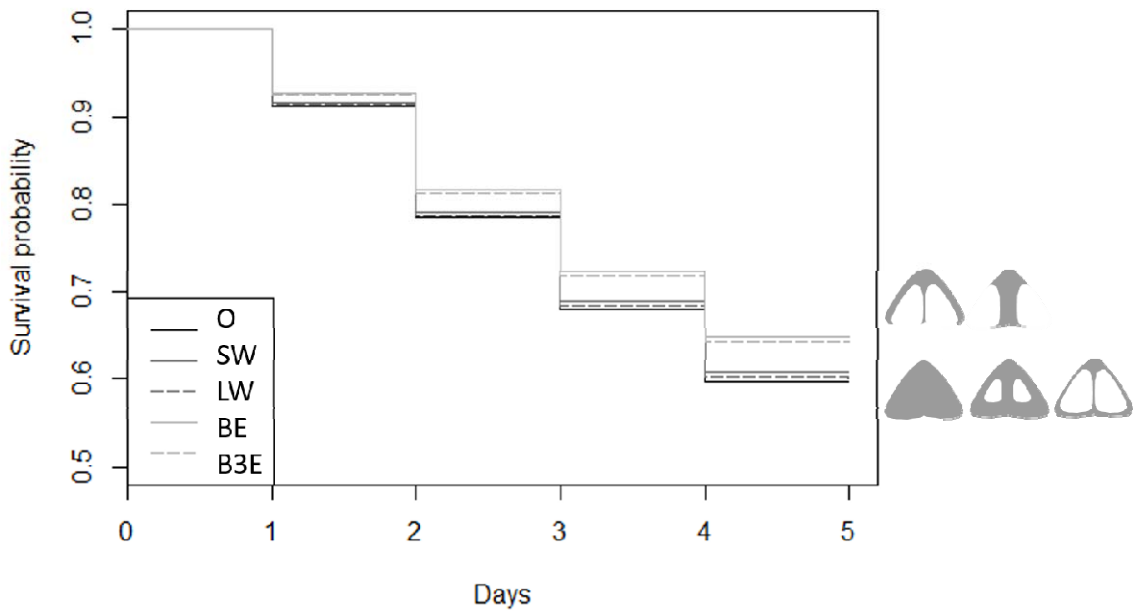


Figure 2. Survival of artificial prey without transparent elements (O-opaque), with small (SW) and large transparent elements touching none (LW), one (B1E) or three prey edges (B3E). Artificial butterflies were placed on tree trunks and monitored for their 'survival' every day for 4 days. Data from the six weeks during which the experiment was conducted are pooled together.

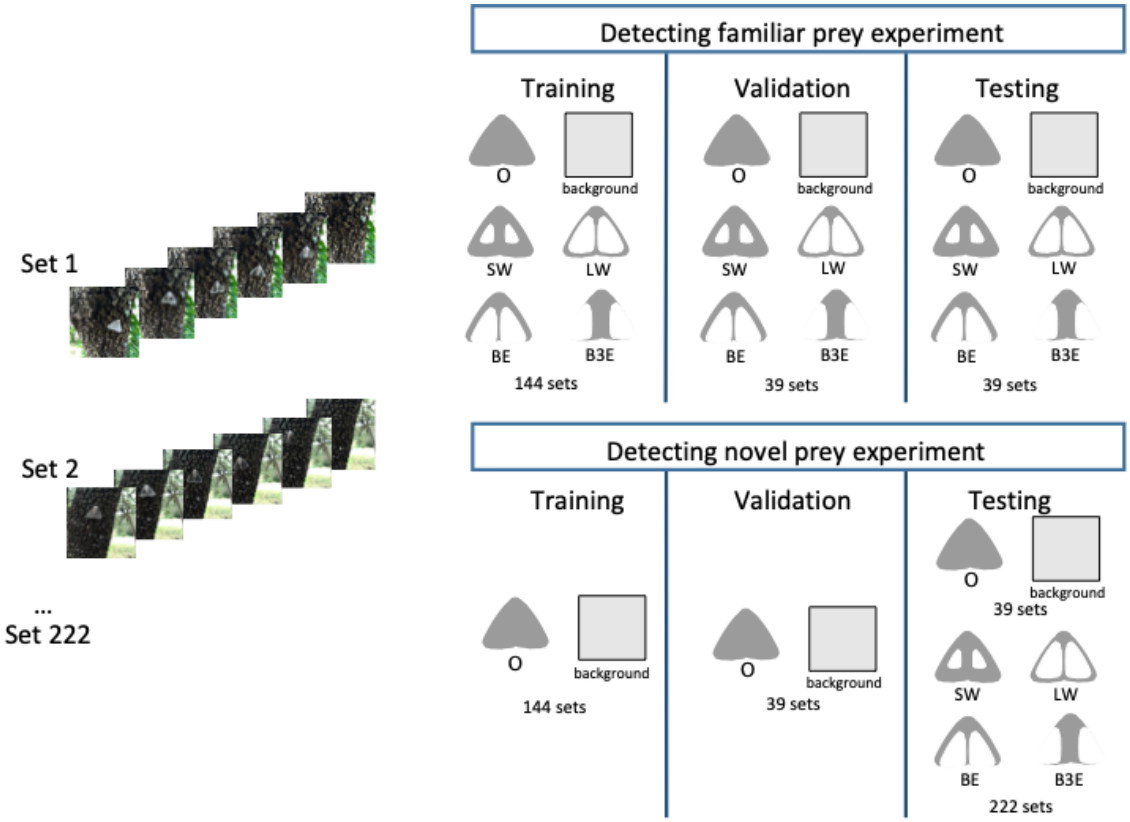


Figure 3. Scheme of training and testing in the detection and the establishment scenarios with the different DNNs. Each cycle was repeated 100 times.

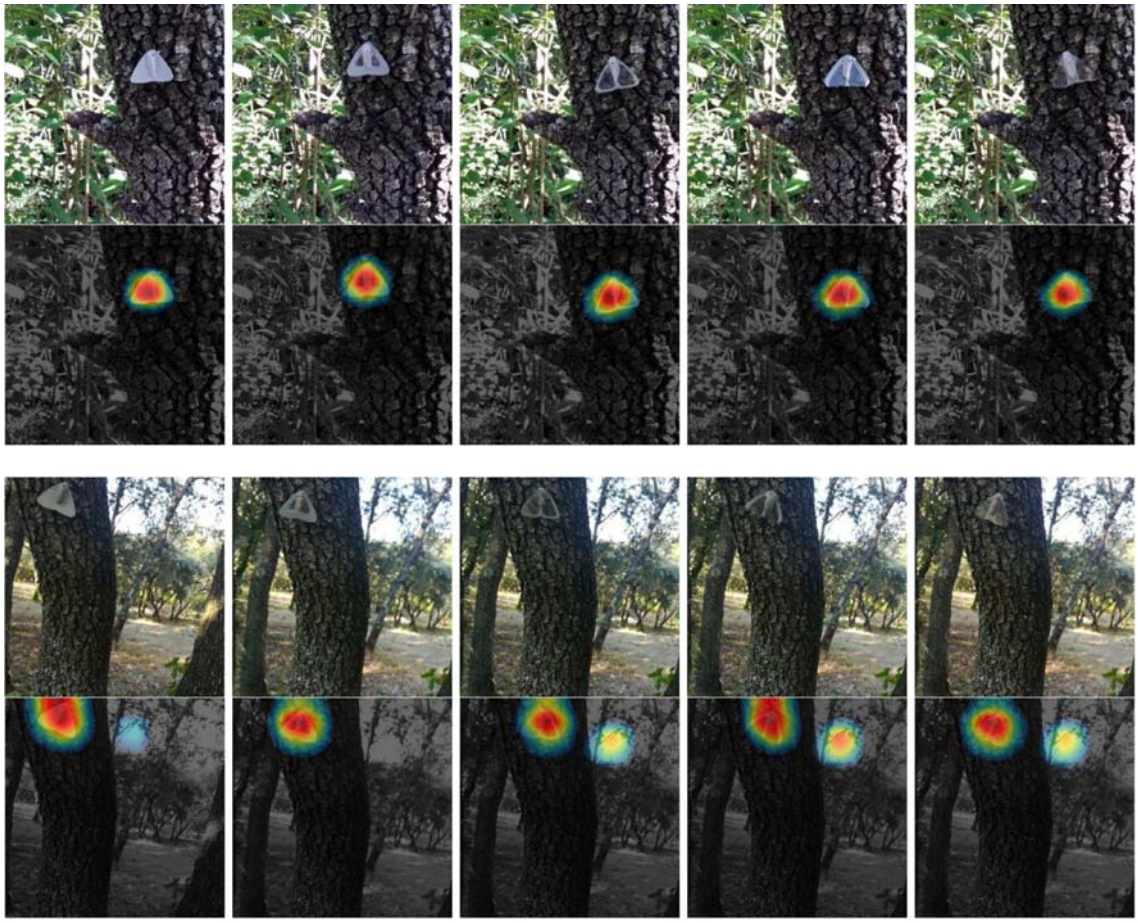

Figure 4. Example images of each of the five morphs against two different backgrounds, together with their corresponding class activation maps (CAMs) produced by SqueezeNet. The CAMs are heatmaps with “hot”, or red, regions corresponding to parts of the image with the highest activations.

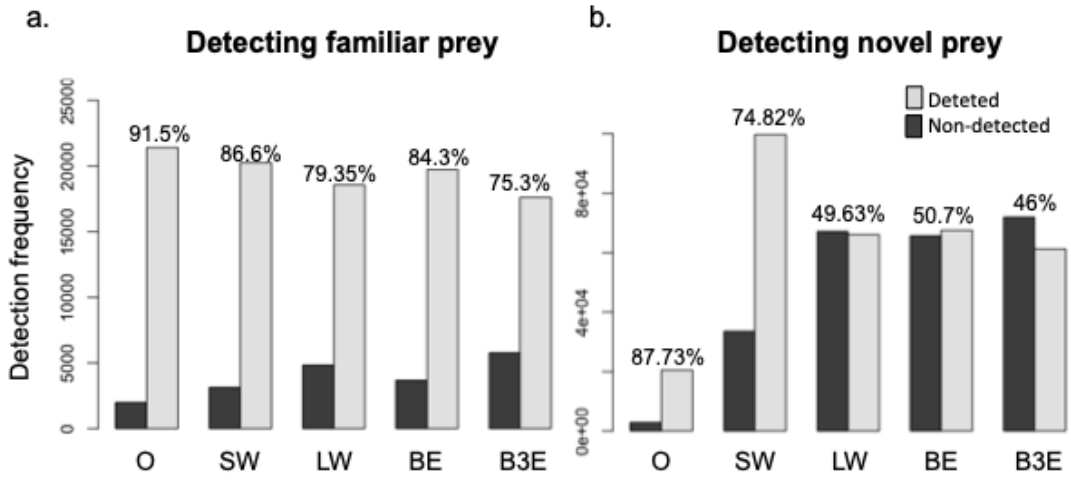


Figure 5. Detection frequency per DNN and morph at detecting familiar prey (a, training DNNs on all prey forms and the background) and novel prey (b, training DNNs on only background and the opaque form). Percentage of detection per morph are represented on top of the bars.

Table 1. Generalised linear mixed model results for average morph detection for all DNNs in both experiments: detection of familiar prey (left, DNN trained on all morphs and background) and detection of novel prey (right, DNN trained on O morph and background).

| | Detection of familiar prey | | | Detection of novel prey | | |
|---|---|---|---|---|---|---|
| | Estimate± SE | z | P | Estimate± SE | z | P |
| (Intercept) | 4.38±0.42 | 10.32 | *** | 1.78±0.15 | 12.22 | *** |
| O >SW, LW, BE, B3E | 0.41±0.04 | 9.8 | *** | 0.79±0.04 | 19.06 | *** |
| SW>LW | 0.57±0.12 | 4.66 | *** | 1.21±0.11 | 11 | *** |
| BE > B3E | 0.68±0.12 | 5.67 | *** | 0.26±0.11 | 2.37 | * |
| full windows > broken windows | 0.25±0.09 | 2.96 | ** | 0.68±0.08 | 8.71 | *** |

Prey morphs are the explanatory variables (O: opaque, SW: small windows, LW: large windows, BE: windows touching one border, B3E: windows touching three borders). Symbols: . * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Electronic supplementary material**

# Deep Neural Network and field experiments reveal how transparent wing windows reduce detectability in butterflies

Mónica Arias*[1,2], Cynthia Tedore*[1,3], Marianne Elias[2], Lucie Leroy[1], Clément Madec[1], Louane Matos[1], Julien P. Renoult[1], Doris Gomez[1, 4]

Affiliations:

[1] CEFE, CNRS, Univ. Montpellier, Univ. Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

[2] ISYEB, CNRS, MNHN, Sorbonne Univ., EPHE, Univ. Antilles, 45 rue Buffon CP50, Paris, France

[3] Univ. Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Institute of Zoology, Hamburg, Germany

[4] INSP, CNRS, Sorbonne Univ., Paris, France

*first co-authors

Corresponding author email: moarias@gmail.com

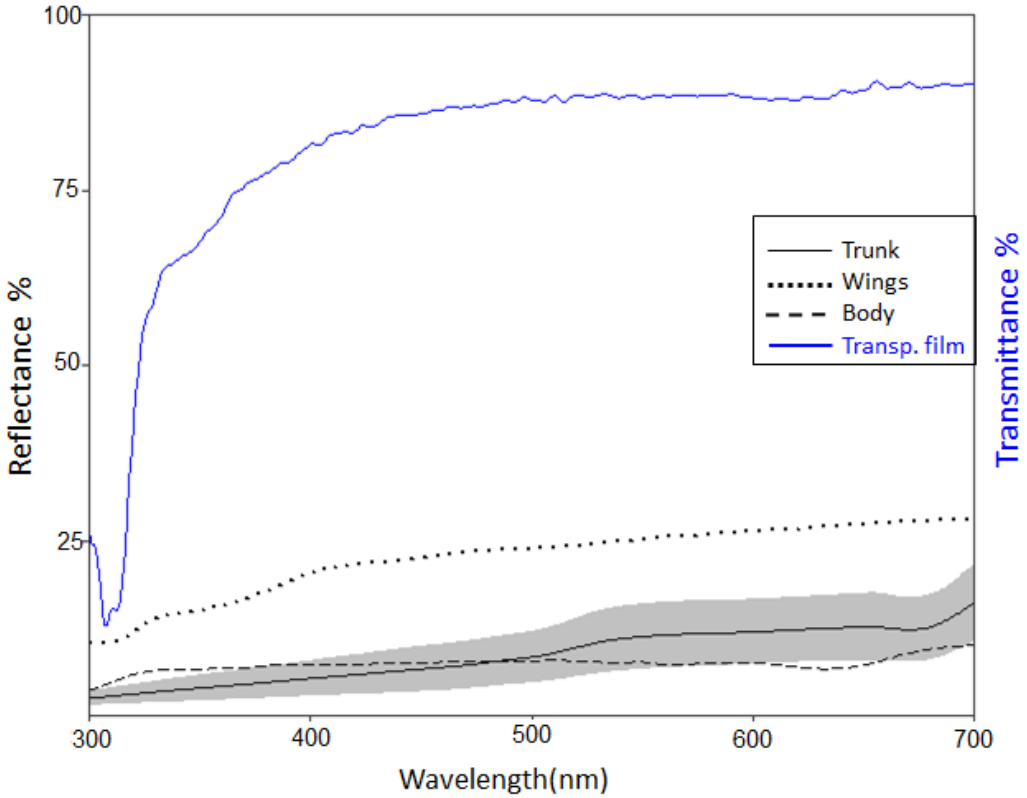


Figure S1. Reflectance spectra of the background (oak tree trunks in solid black line with a confidence interval of ± 1 standard deviation) and opaque areas: body (dashed line) and opaque wings (dotted line), and transmittance of the transparent film (solid blue line).

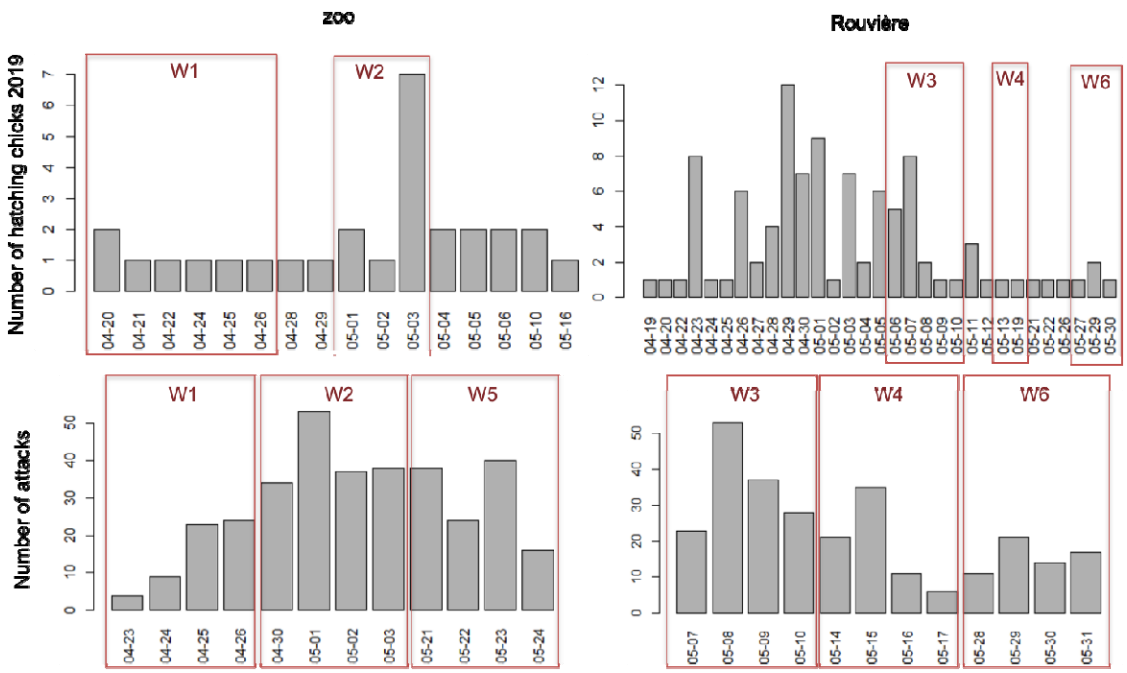


Figure S2. Hatching chicks (top) and attacked prey (bottom) at the two different experimental sites: the zoo (left) and La Rouvière forest (right). Red squares in the top plots indicate the days when the experiment was run in each of the places and in the bottom plot delimit the weeks of the experiment. By week 5 no more hatchlings were registered at the zoo during 2019.

Table S1. False positive rate for different DNNs and training protocols

| Training morph | DNN | False positive rate |
|---|---|---|
| O | ResNet | 0.01 |
| O | SqueezNet | 0.06 |
| O | AlexNet | 0.19 |
| O | GoogLeNet | 0.03 |
| O | vgg16 | 0.03 |
| O | vgg19 | 0.03 |
| all | SqueezNet | 0.18 |
| all | ResNet | 0.12 |
| all | AlexNet | 0.31 |
| all | vgg16 | 0.12 |
| all | vgg19 | 0.14 |
| all | GoogLeNet | 0.09 |

Table S2. Summary statistics on moth detection per DNN architecture and moth morph when DNNs on the detection experiment (trained in all morphs, left) and on the establishment experiment (trained on opaque, right).

| DNN | Morph | Detection of familiar prey (mean±SD) | Detection of novel prey (mean±SD) |
|---|---|---|---|
| AlexNet | O | 0.81±0.39 | 0.75±0.43 |
| AlexNet | SW | 0.72±0.45 | 0.61±0.49 |
| AlexNet | LW | 0.61±0.49 | 0.47±0.5 |
| AlexNet | BE | 0.71±0.45 | 0.55±0.5 |
| AlexNet | B3E | 0.62±0.48 | 0.47±0.5 |
| GoogLeNet | O | 0.93±0.25 | 0.91±0.29 |
| GoogLeNet | SW | 0.88±0.32 | 0.78±0.41 |
| GoogLeNet | LW | 0.81±0.39 | 0.54±0.5 |
| GoogLeNet | BE | 0.86±0.35 | 0.54±0.5 |
| GoogLeNet | B3E | 0.76±0.43 | 0.5±0.5 |
| ResNet | O | 0.93±0.25 | 0.87±0.34 |
| ResNet | SW | 0.9±0.3 | 0.75±0.43 |
| ResNet | LW | 0.85±0.36 | 0.47±0.5 |
| ResNet | BE | 0.88±0.33 | 0.44±0.5 |
| ResNet | B3E | 0.8±0.4 | 0.41±0.49 |
| SqueezNet | O | 0.93±0.26 | 0.9±0.3 |
| SqueezNet | SW | 0.87±0.33 | 0.73±0.44 |
| SqueezNet | LW | 0.8±0.4 | 0.5±0.5 |
| SqueezNet | BE | 0.86±0.34 | 0.53±0.5 |
| SqueezNet | B3E | 0.79±0.41 | 0.49±0.5 |
| vgg16 | O | 0.94±0.24 | 0.92±0.28 |
| vgg16 | SW | 0.9±0.3 | 0.81±0.39 |
| vgg16 | LW | 0.83±0.37 | 0.52±0.5 |
| vgg16 | BE | 0.87±0.34 | 0.52±0.5 |

| | | | |
|---|---|---|---|
| vgg16 | B3E | 0.76±0.43 | 0.45±0.5 |
| vgg19 | O | 0.95±0.23 | 0.92±0.27 |
| vgg19 | SW | 0.91±0.28 | 0.8±0.4 |
| vgg19 | LW | 0.85±0.36 | 0.48±0.5 |
| vgg19 | BE | 0.88±0.32 | 0.46±0.5 |
| vgg19 | B3E | 0.79±0.41 | 0.44±0.5 |

Table S3. Mantel test results between distance matrices of field experiments and all DNNs average mean score per morph on the detection experiment (trained in all morphs) and on the establishment experiment (trained on opaque).

| | Comparisons | Observation (R) | Simulated p |
|---|---|---|---|
| Per morph | Field vs mean DNN (opaque) | 0.106 | 0.23 |
| | Field vs mean DNN (all) | 0.081 | 0.42 |
| | mean DNN (all) vs mean DNN (opaque) | 0.71 | 0.01 |
| Per surface | Field vs mean DNN (opaque) | 0.939 | 0.17 |
| | Field vs mean DNN (all) | 0.828 | 0.17 |
| | mean DNN (opaque) vs mean DNN (all) | 0.97 | 0.16 |
| Per border | Field vs mean DNN (opaque) | 0.99 | 0.16 |
| | Field vs mean DNN (all) | 0.1 | 0.5 |
| | mean DNN (opaque) vs mean DNN (all) | -0.005 | 0.51 |

Tables S4. Distance matrices per morph, border and surface for field, detection of familiar prey and detection of novel prey experiments. For the field experiment they are calculated from the Cox survival coefficients. For the DNNs experiments they are calculated from the average score for all DNNs and replicates. They are always calculated by subtracting values corresponding to column morphs - values corresponding to row morphs.

Tables S4.1. Per morph

Table S4.1.a From the field

| Field | O | SW | LG | B1E | B3E |
|---|---|---|---|---|---|
| O | 0 | -0.02 | -0.02 | -0.16 | -0.22 |
| SW | 0.02 | 0 | 0 | -0.14 | -0.2 |
| LG | 0.02 | 0 | 0 | -0.14 | -0.2 |
| B1E | 0.16 | 0.14 | 0.14 | 0 | -0.06 |
| B3E | 0.22 | 0.2 | 0.2 | 0.06 | 0 |

Table S4.1.b From detection familiar prey experiment

| All | O | SW | LG | BE | B3E |
|---|---|---|---|---|---|
| O | 0 | -0.05 | -0.12 | -0.07 | -0.16 |
| SW | 0.05 | 0 | -0.07 | -0.02 | -0.11 |
| LG | 0.12 | 0.07 | 0 | 0.01 | -0.04 |
| BE | 0.07 | 0.02 | -0.01 | 0 | -0.09 |
| B3E | 0.16 | 0.11 | 0.04 | 0.09 | 0 |

Table S4.1.c. From detection novel prey experiment

| Opaque | O | SW | LG | BE | B3E |
|---|---|---|---|---|---|
| O | 0 | -0.13 | -0.38 | -0.37 | -0.42 |
| SW | 0.13 | 0 | -0.25 | -0.24 | -0.29 |
| LG | 0.38 | 0.25 | 0 | 0.01 | -0.04 |
| BE | 0.37 | 0.24 | -0.01 | 0 | -0.05 |
| B3E | 0.42 | 0.29 | 0.04 | 0.05 | 0 |

Tables S4.2. Per surface

Table S4.2.a From field experiment

| Field | O | SW | LargeW |
|---|---|---|---|
| O | 0 | -0.02 | -0.13 |
| SW | 0.02 | 0 | -0.11 |
| LargeW | 0.13 | 0.11 | 0 |

Table S4.2.b. From detection familiar prey experiment

| All | O | SW | LargeW |
|---|---|---|---|
| O | 0 | -0.05 | -0.12 |
| SW | 0.05 | 0 | -0.07 |
| LargeW | 0.12 | 0.07 | 0 |

Table S4.2.c. From detection novel prey experiment

| Opaque | O | SW | LargeW |
|---|---|---|---|
| O | 0 | -0.13 | -0.39 |
| SW | 0.13 | 0 | -0.26 |
| LargeW | 0.39 | 0.26 | 0 |

Table S4.3. Per border

Table S4.3.a From field experiment

| Field | Complete | BE | B3E |
|---|---|---|---|
| Complete | 0 | -0.15 | -0.21 |
| BE | 0.15 | 0 | -0.06 |
| B3E | 0.21 | 0.06 | 0 |

Table S4.3.b. From detection familiar prey experiment

| All | Complete | BE | B3E |
|---|---|---|---|
| Complete | 0 | -0.02 | -0.11 |

| | | | |
|---|---|---|---|
| BE | 0.02 | 0 | -0.09 |
| B3E | 0.11 | 0.09 | 0 |

Table S4.3.c. From detection novel prey experiment

| Opaque | Complete | BE | B3E |
|---|---|---|---|
| Complete | 0 | -0.14 | -0.18 |
| BE | 0.14 | 0 | -0.05 |
| B3E | 0.18 | 0.05 | 0 |