OXFORD
UNIVERSITY PRESS

Journal	:	BEHECO
Article Doi	:	10.1093/beheco/arw166
Article Title	:	Which traits do observers use to distinguish Batesian mimics from their models?

INSTRUCTIONS

1. **Author groups:** Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online. Please also check all author affiliations.

2. **Figures:** If applicable figures have been placed as close as possible to their first citation. Please check that they are complete and that the correct figure legend is present. Figures in the proof are low resolution versions that will be replaced with high resolution versions when the journal is printed.

3. **Colour reproduction:** These figures are currently intended to appear online in colour and black and white in print. Please check the black and white versions (these may be available at the end of the paper) and contact us if you have any concerns. Please re-word the legend/text to avoid using reference to colour. Alternatively, please let us know if you wish to pay for print colour reproduction or to have both versions in black and white. Please note that there is a £350/$600 charge for each figure reproduced in colour in print.

4. **URLs:** Please check that all web addresses cited in the text, footnotes and reference list are up-to-date, and please provide a 'last accessed' date for each URL.

5. **Funding:** Please provide a Funding statement, detailing any funding received. Remember that any funding used while completing this work should be highlighted in a separate Funding section. Please ensure that you use the full official name of the funding body, and if your paper has received funding from any institution, such as NIH, please inform us of the grant number to go into the funding section. We use the institution names to tag NIH-funded articles so they are deposited at PMC. If we already have this information, we will have tagged it and it will appear as coloured text in the funding paragraph. Please check the information is correct. [red text to be used for suppliers who are tagging the funding]

6. **Conflict of interest:** All authors must make a formal statement indicating any potential conflict of interest that might constitute an embarrassment to any of the authors if it were not to be declared and were to emerge after publication. Such conflicts might include, but are not limited to, shareholding in or receipt of a grant or consultancy fee from a company whose product features in the submitted manuscript or which manufactures a competing product. The following statement has been added to your proof: 'Conflict of Interest: none declared'. If this is incorrect please supply the necessary text to identify the conflict of interest.

7. Please note that the first colour figure of this paper will be published in print colour at no charge.

MAKING CORRECTIONS TO YOUR PROOF

These instructions show you how to mark changes or add notes to your proofs using Adobe Acrobat Professional versions 7 and onwards, or Adobe Reader DC. To check what version you are using go to **Help** then **About**. The latest version of Adobe Reader is available for free from get.adobe.com/reader.

DISPLAYING THE TOOLBARS

Adobe Reader DC

In Adobe Reader DC, the Comment toolbar can be found by clicking 'Comment' in the menu on the right-hand side of the page (shown below).
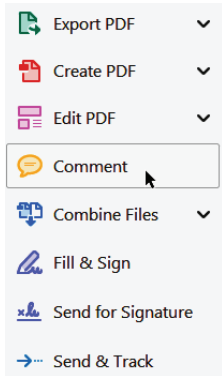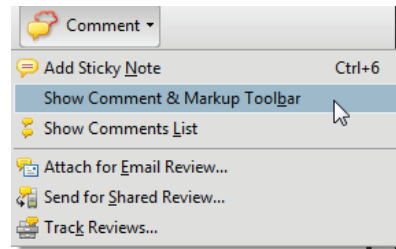

The toolbar shown below will then display along the top.

Acrobat Professional 7, 8, and 9

In Adobe Professional, the Comment toolbar can be found by clicking 'Comment(s)' in the top toolbar, and then clicking 'Show Comment & Markup Toolbar' (shown below).
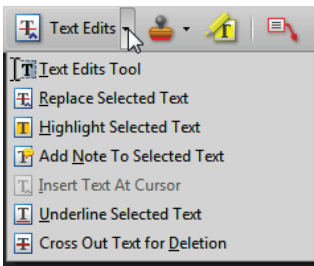

The toolbar shown below will then be displayed along the top.

USING TEXT EDITS AND COMMENTS IN ACROBAT

This is the quickest, simplest and easiest method both to make corrections, and for your corrections to be transferred and checked.

1. Click **Text Edits**
2. Select the text to be annotated or place your cursor at the insertion point and start typing.
3. Click the **Text Edits** drop down arrow and select the required action.

You can also right click on selected text for a range of commenting options, or add sticky notes.

SAVING COMMENTS

In order to save your comments and notes, you need to save the file (**File, Save**) when you close the document.
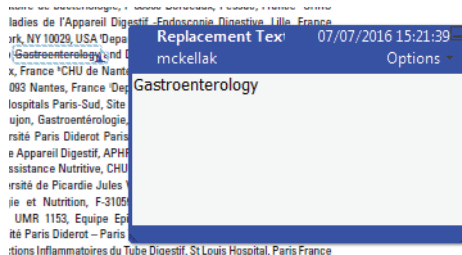
USING COMMENTING TOOLS IN ADOBE READER

All commenting tools are displayed in the toolbar. You cannot use text edits, however you can still use highlighter, sticky notes, and a variety of insert/replace text options.
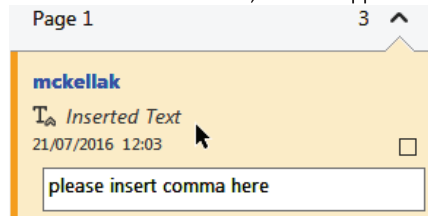

POP-UP NOTES

In both Reader and Acrobat, when you insert or edit text a pop-up box will appear. In **Acrobat** it looks like this:

In **Reader** it looks like this, and will appear in the right-hand pane:

DO NOT MAKE ANY EDITS DIRECTLY INTO THE TEXT, USE COMMENTING TOOLS ONLY.

AUTHOR QUERY FORM

AUTHOR QUERIES - TO BE ANSWERED BY THE CORRESPONDING AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please click on each query number and respond by indicating the change required within the text of the article. If no change is needed please add a note saying "No change."

AQ1	Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and fore-names are tagged correctly, for accurate indexing online.
AQ2	Please provide full road address for the affiliation.
AQ3	Please check and confirm the changes made in the sentence beginning "Humans provide…"
AQ4	Please review the typeset tables carefully against copies of the originals to verify accuracy of editing and typesetting.
AQ5	Figures have been placed as close as possible to their first citation. Please check that they have no missing sections and that the correct figure legend is present.
AQ6	Please provide author name and citation in text for the reference "Author. 2012."
AQ7	Please update the reference "Morris R, Reader T. in review" with year, volume number, and page range.
AQ8	Please check and confirm the change made in the reference "R Core Team. 2014".
AQ9	Figures 1 and 2 are in low resolution, please provide us the high resolution figures for processing.

Original Article

Which traits do observers use to distinguish Batesian mimics from their models?

Christopher H. Taylor,[a] Jonathan Warrin,[a] Francis Gilbert,[a] and Tom Reader[a]

[a]School of Life Sciences, University of Nottingham, Nottingham, NG7 2RD, UK

Batesian mimicry, in which a harmless mimic resembles a more aversive model, can encompass a wide range of morphological traits, but the resemblance is never perfect. Previous studies have used abstract "prey" designs to show that differences in certain traits may not be relevant to mimicry if they are not perceived or recognized by a predator. Here, we extend these results by examining how human "predators" respond to realistic variation in traits of aposematic wasps and their hoverfly mimics. We measured the ability of humans to discriminate between images of wasps and hoverflies in which only certain traits were visible, to determine the contributions of those traits to discrimination decisions. We found that shape is a particularly useful and easily learnt trait for separating the 2 taxa. Subjects did not successfully discriminate on the basis of abdominal patterns, despite those containing useful information. Color similarity between wasps and hoverflies is relatively high in comparison with other traits, suggesting that selection has acted more strongly on color. Our findings demonstrate the importance of consideration of natural variation in the traits of prey and their salience to predators for understanding the evolution of prey defenses.

INTRODUCTION

In Batesian mimicry, a relatively harmless organism evolves to resemble a more dangerous one in order to deceive predators (Bates 1862). A major focus of recent studies on Batesian mimicry has been the somewhat puzzling inaccuracy of many mimics, when we expect natural selection to favor the most accurate ones (Penney et al. 2012; Kikuchi and Pfennig 2013). The resemblance of a mimic to its model might encompass a wide range of traits, including shape (Jones et al. 2013), color (Kraemer and Adams 2014), behavior (Golding et al. 2005a), and sound (Barber and Conner 2007). However, these traits may not all hold the same relevance to mimicry, as some may be overlooked by the predator (Kikuchi and Pfennig 2010).

For a predator to select for increased accuracy in a given mimetic trait, that trait must be perceived by the predator. A predator may not be physically able to detect a trait because of sensory limitations, such as those imposed on color perception by the spectral sensitivity of cone cells (Cuthill and Bennett 1993; Chen and Goldsmith 1986; Kraemer and Adams 2014). Furthermore, a signal may be presented in a way that makes it difficult to resolve in detail, such as in the case of a moving target, or when it is only viewable for a short period of time (Chittka and Osorio 2007).

Even if a trait is perceived, differences between the model and mimic may or may not be learned effectively by the predator. Different traits can have different levels of "salience", which is how much they stand out as part of a complex image or scene, and the relative salience of the presented traits determines how readily they can be learned (Kazemi et al. 2014). Salience is determined by both absolute characteristics, such as the magnitude of a signal (e.g., brightness of a color or volume of a sound), as well as the cognition of the observer (Shettleworth 2010). Thus, a trait could be clearly perceivable by the predator but nonetheless overlooked. Such a trait will be under little or no selection for mimetic accuracy, which may account for inaccurate mimicry of some aspects of the model phenotype (Kazemi et al. 2014; Sherratt et al. 2015).

It is not possible to make universal statements about the salience of particular traits, since their importance will depend upon the signal's context and the sensory and cognitive abilities of the signal receiver (Cuthill and Bennett 1993), as well as the previous experience of a given observer (Blough 1969). However, with birds as predators, experiments using artificial prey generally support the idea that color is an especially salient signal in the context of mimicry, although other traits such as pattern may come into play if color is not informative (Terhune 1977; Kazemi et al. 2014). Humans also give a strong weighting to color in discriminating among rewarding and non-rewarding artificial prey (Kikuchi et al. 2015; Sherratt et al. 2015), to an extent that may be suboptimal (Kikuchi et al. 2015). The use of artificial prey in such experiments allows careful control over the stimuli, allowing different traits to be varied entirely independently. However, the downside is that the

Address correspondence to C.H. Taylor, Life Sciences Building, University Park Campus, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: christopher.taylor@nottingham.ac.uk

magnitude of variation seen among prey items may not be realistic, and there has been a call for experiments that study multiple dimensions of mimicry in a more natural context (Kikuchi et al. 2015).

2.5   An important study of the relative importance of mimetic traits in signals found in nature was conducted by Bain et al. (2007), based on data from Dittrich et al. (1993). In the latter study, groups of pigeons were conditioned to peck at images of flies but not of aposematic wasps (or vice versa). They were then presented with
2.10   images of various wasp-mimicking hoverflies, and the measured peck rates provided an estimate of mimetic accuracy through the pigeons' eyes. Bain et al. (2007) then measured a number of morphometric variables from the presented images and investigated which variables best predicted the pigeons' responses. They found
2.15   that summary measures of color and pattern were good predictors of bird behavior, with influence also from some shape measures such as antennal length. However, this study was essentially observational and, as such, the measured variables may have been confounded with one another. Other more experimental studies
2.20   have focused on small numbers of traits, such as shape and eyespots in caterpillars (Hossie and Sherratt 2014; Skelhorn et al. 2014) or head shape and body pattern in snakes (Valkonen et al. 2011). Kauppinen and Mappes (2003) studied the contribution of wasp odor, pattern, and shape to avoidance by dragonflies but with very
2.25   coarsely delineated treatments.

One of the reasons that experimental tests of the relative importance of the morphological traits of Batesian mimics are rare is that relevant, well-replicated experiments with vertebrate predators are ethically and logistically difficult. Humans provide a useful
2.30   model "predator", with fewer ethical and practical constraints, and experimental games in which humans "forage" for prey have previously generated large datasets to explore hypotheses about mimicry (Golding et al. 2005b; McGuire et al. 2006; Sherratt et al. 2015; Morris and Reader, in review). Humans have been shown to make
2.35   judgments about mimetic accuracy similar to those of birds, which are thought to provide the main selective pressure for mimicry in hoverflies (Dittrich et al. 1993; Penney et al. 2012; Sherratt et al. 2015). While bird vision is sensitive to some ultraviolet wavelengths that are invisible to humans, these colors do not feature in the
2.40   aposematic patterns of wasps or their mimics (Green et al. 1999; Taylor et al. 2016a).

Here, we combine the experimental approach of studies such as Kazemi et al. (2014) with the realistic signals of Bain et al. (2007) to shed further light on the selective pressures on various individual
2.45   ual traits contributing to Batesian mimicry in hoverflies. We present the results of an online game in which human volunteers take the role of "predators" choosing whether or not to "attack" images of hoverflies and wasps, with some images only showing a limited set of traits (e.g., a black and white image showing outline only).
2.50   We use low-resolution images and limited time for observation, to reflect the difficulties that predators face in the wild. Due to our use of images that show realistic levels of variation, not all traits are equally informative and hence we cannot directly test their salience. Instead, we investigate which hoverfly trait(s) are used suc-
2.55   cessfully to discriminate mimics from models. For traits that are not used in making these foraging choices, we examine whether they are overlooked (i.e., contain potentially useful information that is not exploited) or uninformative. Finally, we investigate whether our subjects can be "trained" to make use of different information
2.60   within the images, and hence test whether prior experience shapes
2.61   the relative contributions of different traits in future encounters.

MATERIALS AND METHODS

The following methodology was approved by the School of Life Sciences Research Ethics Committee at the University of Nottingham.   2.65

Specimens

Human volunteers played a game in which they viewed images of hoverflies (Diptera: Syrphidae) and stinging Hymenoptera. We col-
2.70   lected these insects using a hand net from wild communities around Nottingham, UK (particularly the Attenborough Nature Reserve) and Sobreda, Portugal. Our target insects were the common wasp *Vespula vulgaris* and other stinging Hymenoptera (models) or hov-
2.75   erflies (mimics) of a roughly similar appearance (i.e., with a black and yellow or black and orange pattern, and not excessively hairy). In this study, we have not included examples of alternative, nonmimetic prey. As the presence of such prey can influence predator
2.80   behavior (Lindström et al. 2004), here we are specifically interested in the differences between good and poor mimics, rather than between mimics and nonmimics.

We refrigerated specimens at 5 °C for a period of at least an hour after capture in order to slow their movements and temporar-
2.85   ily prevent them from flying. We then placed them in an open petri dish inside a homemade "photo studio"—a white 30 × 18 × 10 cm open-topped box. We placed a 5-mm scale bar near to the insect. We then took 2 types of photograph: for the "live photo", we allowed the specimen to assume a natural resting position and then photographed it with an Olympus E420 DSLR camera. For the
2.90   "abdominal photo", we euthanized the same specimens by freezing and then photographed their abdomens using the same apparatus with, where necessary, legs and wings pinned out to the sides to give a clear view of the abdomen.

Images   2.95

From the insects collected and photographed, we selected 3 model and 9 mimic species for which we had at least 5 individuals with both live and abdomen photos of good quality. The 12 species each had between 5 and 21 individuals with suitable photographs, giving a total of 104 individuals that were used in the game (Table 1).   2.100

We processed photographs using a combination of MATLAB (2012) for automated processing and ImageJ (Abràmoff et al. 2004) for manual editing. All images were saved in GIF format, with the insect image showing for 0.5 s before being replaced by a blank white background. All images were rotated by eye to a standard   2.105
orientation with the axis of the insect vertical and the anterior end at the top, in order to remove this relatively uninteresting source of variation. The images were necessarily displayed in RGB color format, but there are no "hidden" components of the color pattern (such as ultraviolet patches) in these insects that would be omitted   2.110
as a result (Green et al. 1999; Taylor et al. 2016a). We were unable to control for any variation in the specifications of the monitors used by participants and how given colors were represented.

In total, we prepared 7 different images for each individual insect to be used across 2 experiments (see Figure 1), as follows: 2.115

"Complete"

An image of the whole insect, in color and to scale. Using the live photos, body length was measured from the tip of the head between the antennae to the tip of the abdomen, using the scale   2.120
bar for reference. Excess background was then cropped out by eye, including removal of the scale bar from the image. Histogram   2.122

equalization was used to enhance the contrast of the image. Finally, using MATLAB, the image was resized to a standard scale such that 1 mm on the insect was represented by a distance of 5 pixels, with the resulting images varying from 73 to 171 pixels in height.

Note that as we were unable to control the size and resolution of the monitors used by participants, it was impossible to present the images at a completely standardized scale. As a guide, at a resolution of 1280×1024 pixels on a 19-inch monitor, the images appear at roughly twice the true size of the insects. We were also unable to control the distance at which volunteers sat from their monitors. However, very approximately, the level of detail available at a resolution of 5 pixels per mm is about the same as that available to a passerine bird viewing an insect at a distance of half a meter (based on Donner 1951).
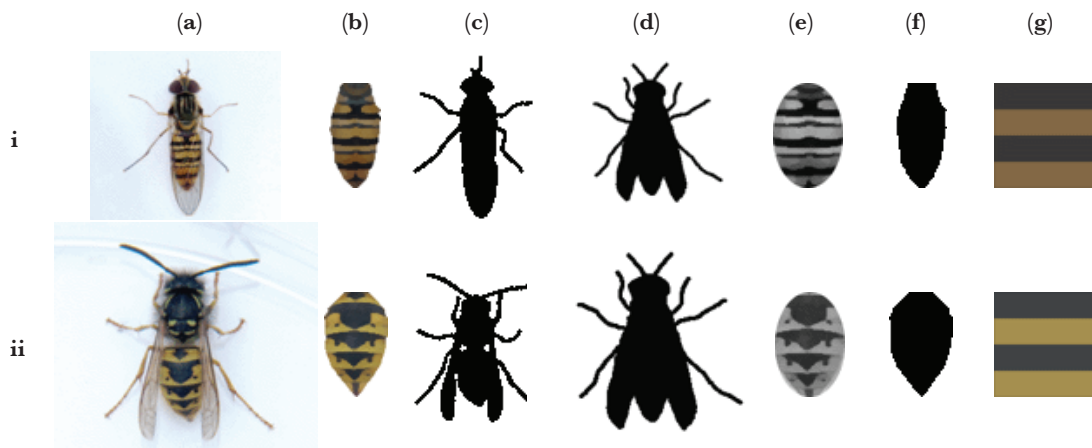
"Abdomen"

An image of just the abdomen of the insect, in color, presented at a standard size. From the abdomen photos, the outline of the abdomen was detected via a user-guided edge detection algorithm and anything outside this boundary was assigned as background and set to pure white. The image was rotated to standard orientation (see above) and resized to a standard height of 60 pixels. Thus, the images varied in terms of the pattern, color, and shape of the abdomen, but showed no variation in size or in appearance of other body parts (which were not displayed).

"Shape"

A black and white image showing the silhouette of the whole insect, presented at a standard size. The outline of the live image was initially generated using "Canny" edge detection (Canny 1986). However, this outline contained some gaps, and some background edges were detected that were not part of the insect. The outline was therefore modified by eye where necessary to be continuous and then filled in with black for the insect and white for the background. The image was rotated to standard orientation (see above) and then resized to a standard height of 100 pixels. Note that the larger size as compared to the "Abdomen" images (60 pixels) was in order to keep the level of detail in the 2 images roughly equal, since the latter show only a portion of the whole insect. The resulting "Shape" images varied in outline, but were identical in all other respects.

"Size"

A black and white image showing the silhouette of an "intermediate" insect, scaled in proportion to the size of the individual in question. We generated a black and white image of an imagined intermediate between a hoverfly and a wasp, guided by superimposing one of the wasp silhouette images (*V. vulgaris*) over one of the

Table 1

Details of the 12 species included in the game

Species	Abbrev.	N	Mean body length (mm)	Example image
Models				
Nomada sp.	Nsp	8	102	
Vespula germanica	Vge	6	149	
Vespula vulgaris	Vvu	8	126	
Mimics				
Chrysotoxum intermedium	Cin	6	125	
Epistrophe eligans	Eel	5	107	
Episyrphus balteatus	Eba	18	104	
Helophilus pendulus	Hpe	21	117	
Helophilus trivittatus	Htr	5	163	
Melanostoma scalare	Msc	5	86	
Myathropa florea	Mfl	11	125	
Sphaerophoria scripta	Ssc	5	111	
Syrphus ribesii	Sri	6	118	

Note that the example images have been cropped down from the full versions used in the game.

Figure 1

Examples of image types used in the online quiz. These images are shown at the same resolution as they appeared to participants. a) Complete. b) Abdomen. c) Shape. d) Size. e) Abdomen pattern. f) Abdomen shape. g) Abdomen color. i) Images of *Episyrphus balteatus*. ii) Images of *Vespula vulgaris*.

hoverflies (*Syrphus ribesii*). This ambiguous shape was then resized in each case to correspond to the length of each individual specimen, again at a scale of 5 pixels per mm. This set of images therefore varied in terms of their size (from 73 to 171 pixels in height), but were identical in all other respects.

"Abdomen Pattern"

A grayscale image of the abdominal pattern, cropped to an oval shape of fixed size. The abdomen photo was converted to grayscale and then cropped to a standard oval shape 60 pixels long and 40 pixels wide, with the removed area set to white. Prior to cropping, the image was scaled so that the oval area was representative of the majority of the abdominal pattern. For some species with very elongated abdomens, such as *Sphaerophoria scripta*, it was necessary to stretch the abdominal image horizontally before cropping in order that the whole oval "window" was filled but the majority of the pattern was still represented. The resulting images varied in their pattern (intensity of grayscale) but were identical in all other respects.

"Abdomen Shape"

A black and white image showing the silhouette of the abdomen, presented at a standard size. The edge of the abdomen was detected using the same method as described in Abdomen above. Small adjustments were made by eye to the outline where the algorithm had not completely separated insect from background. This outline was then filled in with black for the insect and white for the background, and resized to a fixed length of 60 pixels. These images varied in the abdominal outline only.

"Abdomen Color"

An image of a simple, standard pattern with colors based on the abdominal colors from the original abdomen photo. A simple artificial pattern was generated, made up of 4 horizontal stripes within a 60 by 60 pixel area. These stripes were then colored alternately "black" and "yellow" using mean RGB values obtained from the "black" and "yellow" regions of the abdomen photo. The images varied only in the colors of the stripes.

The game

Volunteers were recruited via requests through social media, electronic mailing lists, and word of mouth. Participants were also encouraged to help publicize the game via their own social media accounts. The experiment was described as a "game" or "quiz" in order to instill a sense of fun and challenge, to encourage a greater number of participants.

We invited potential volunteers to view a webpage which gave a brief description of the game: the volunteers would take on the role of a predator discriminating among different types of prey, and they should attempt to maximize their score by "eating" harmless hoverflies and avoiding wasps. The webpage also described the general purpose of the game (i.e., to collect data for research), but did not give the exact aim of the research (i.e., determining which traits are used by observers to distinguish models from mimics) to avoid creating any conscious or unconscious bias in the volunteers. The page also included a disclaimer, stating that data stored from attempts at the game would be anonymous at all stages, and that participants should feel free to withdraw from the game at any time.

If they agreed to participate, volunteers then clicked through to a second page where they entered some general information about themselves: their age, gender, and whether or not they had completed the game before. Any responses marked as not being a first

attempt were discarded from analysis in order to minimize the effect of any previous experience. Next, they completed the game, which comprised 20 questions, separated into 2 equal halves, with the content of each half differing between the 2 experiments (see below). For each question, an image of an insect was displayed on the screen for 0.5 s before disappearing. After presentation of each image, the participant was required to select one of 2 options: "Eat" or "Avoid" (Figure 2). They would then immediately be presented with feedback for that question, based on their response and the identity of the insect. The response was judged "correct" if a mimic was eaten or a model avoided and "incorrect" for the opposite choices. Thus, as a participant progressed through the questions, they had the opportunity to gather information from this feedback and integrate it into their responses, in much the same way as a predator might learn gradually from sampling a range of insects.

For each question, the insect displayed was drawn at random, with replacement, from all 104 possible individuals. Probabilities were adjusted such that each selected individual had an equal chance of being a model or mimic, that within those 2 groups, each species had an equal chance of being selected, and that within each species, each individual had an equal chance of being selected.

At the end of the game, the player could return to the homepage where they were given a score out of 20 according to the number of correct choices they made.

Experiment 1

Here the aim was to investigate how well participants who had been trained using full images could identify mimic and model insects from particular individual traits or combinations of traits. The first 10 questions of the game, the "Training" phase, used the Complete images. In the second half of the game, the "Testing" phase, participants were assigned to one of 7 treatments: Shape, Size, Abdomen, Abdomen Shape, Abdomen Pattern, Abdomen Color, and Complete (control), using the corresponding images as described above. The experiment ran from July 2013 to March 2015.

Figure 2
A screenshot from the online quiz. Note that, in the real quiz, the insect image would only be visible for half a second.

Experiment 2

Here, the aim was to determine whether training participants on a particular trait would influence their performance when the full image was subsequently made available. Participants received one of 3 treatments during the Training phase (questions 1–10): Shape, Size, and Abdomen. Following the Training phase, all participants then experienced a Testing phase in which images were drawn from the Complete set. Data from the Complete treatment from Experiment 1 was used as a control, since participants were given Complete images during both halves of the game. This version ran from February to December 2014.

Statistical analysis

We analyzed the data in R version 3.1.2 (R Core Team 2014). Initial inspection of the data from Experiment 1 showed that on average, participants improved in score over the first 3 questions of the quiz, with the mean scores showing little change after that point (Figure 3). There was a similar effect at the start of the Testing phase for some treatments and during the Training phase of Experiment 2 (Figure 4), while participants adjusted to the new image type. For statistical comparisons, we have therefore discarded data from questions 1–3 and 11–13. We modelled responses to each individual question in the Testing phase using a Generalized Linear Mixed Model (GLMM), with a logit link function to account for the binomial response (correct or incorrect). Predictors were the treatment, the species in the image, and the age and gender of the participant, with participant identity as a random factor. Age was standardized to mean 0, standard deviation 1 as recommended by Zuur et al. (2009). We followed backwards model selection using likelihood ratio tests.

Within data from the Size treatment only, we also tested for an effect of size on attack rate, regardless of whether or not the response was correct, to test whether participants showed consistent bias towards smaller or larger insects. For this, we used a separate GLMM with a fixed effect of size, random effect of participant identity, and attack rate as the response.

For Experiment 2, we fitted a similar binomial GLMM with the predictors treatment, species, phase (Training or Testing), age and gender, and the 2 way interactions phase:treatment and phase:species. Again, we included participant identity as a random factor.

On the basis of results from Experiment 1 (see below), participants were unable to distinguish models and mimics under 3 of the experimental treatments: Size, Abdominal Pattern, and Abdominal Color. There are 2 possible explanations for these results: 1) the treatment images contain no relevant information that can be used to reliably distinguish models and mimics or 2) the images contain relevant information that was not noticed/acted upon by the participants. We therefore decided post hoc to summarize information content in model and mimic images for those 3 treatments in order to help distinguish between those 2 explanations.

The Size treatment was based on a single univariate axis. We therefore compare model and mimic values by means of a histogram. The Abdominal Color treatment was based on a pair of RGB values (for the yellow/orange and black bands), and hence contained 6 different axes of variation. We summarized this information using Principal Components Analysis and then plotted the first 2 components. To summarize the Abdominal Pattern images, we created a pairwise distance map using the distance transform method (Taylor et al. 2013) and then plotted this distance map using nonmetric multidimensional scaling in 2D.

RESULTS

Experiment 1

A total of 602 volunteers completed this version of the game (*N* for each treatment: 158 Shape, 137 Size, 146 Abdomen, 43 Abdomen Pattern, 45 Abdomen Shape, 48 Abdomen Color, and 25 Complete). During model selection for our GLMM, age (Likelihood Ratio [LR] = 1.41, df = 1, $P = 0.235$) and gender of
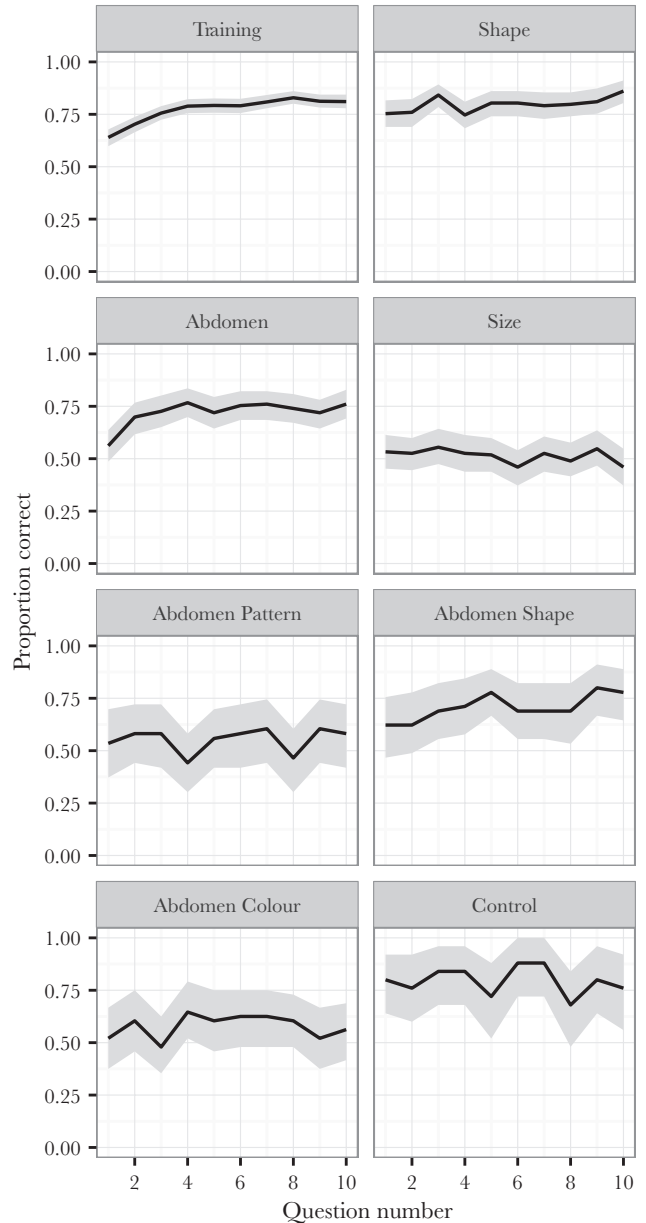

Figure 3
Learning curves according to treatment (Experiment 1). Solid lines show how the proportion of correct answers (*y* axis) changes through the different questions of the quiz (*x* axis) as users learn to distinguish models and mimics. Shaded area shows 95% confidence intervals, calculated via bootstrapping. The first panel shows the learning curve for the Training phase, which was the same for all participants. The remaining panels show the second half of the quiz, separated by treatment. Note that for the purposes of the graph, the question numbers are reset to one at the start of the Testing phase.
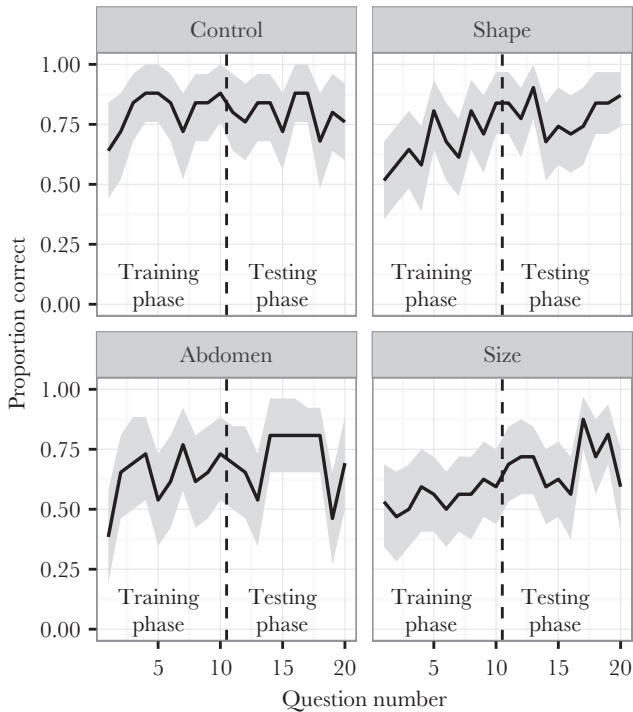
Figure 4

Learning curves according to treatment (Experiment 2). Solid lines show how the proportion of correct answers (*y* axis) changes through the different questions of the quiz (*x* axis) as users learn to distinguish models and mimics. Shaded area shows 95% confidence intervals, calculated via bootstrapping. Panels show the 4 different treatments.

the participant (LR = 1.03, df = 1, *P* = 0.309) were removed from the model, leaving species and treatment as significant fixed effects.

Species identity had a strong effect on the probability of a correct choice by the participant (LR = 129, df = 11, *P* < 0.0001). During the Training phase (i.e., based on the Complete image), volunteers efficiently learned the 2 *Vespula* species as models, attacking them in only 7% of encounters, but *Nomada sp.* was attacked more frequently (39%). Of the mimics, *Myathropa florea* experienced the lowest rate of attack (71%), and was therefore the most effective mimic, whereas *S. scripta* was the worst, experiencing an attack rate of 87% (Figure 5).

The number of correct responses differed significantly among treatments (LR = 206, df = 6, *P* < 0.0001), and according to Tukey post hoc tests, the treatments fell into 2 groups, with Control, Shape, Abdomen, and Abdomen Shape in one and Size, Abdomen Pattern, and Abdomen Color in the other. There were no significant differences within either group (*P* > 0.05), but all pairwise differences between the 2 groups were significant (*P* < 0.01), with the former group consistently higher (Figure 6). For Size and Abdomen Pattern, 95% confidence intervals for success rate included 0.5, indicating that performance by the participants was no better than expected by chance. Note that participants were not necessarily attacking at random even in treatments, which had a 50% success rate: a separate GLMM showed that in the Size treatment, larger insects were significantly less likely to be attacked (LR = 7.09, df = 1, *P* = 0.008).

Experiment 2

A total of 114 volunteers took part in the second version of the game, split roughly evenly among the 4 different treatments (31
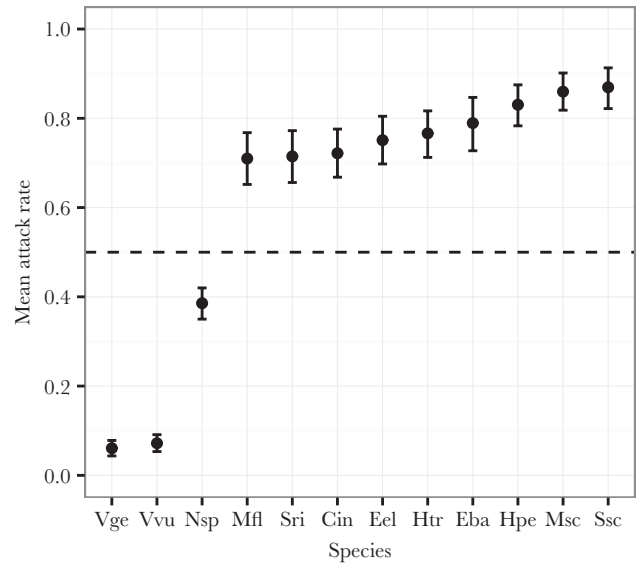

Figure 5

Variation in attack rate according to species (Experiment 1). Points show the mean rate of attack against different species during the Training phase (excluding the first 3 questions of the phase, during which learning was ongoing). Error bars show 95% confidence intervals calculated from bootstrapping. The dashed line shows a score of 0.5, or 50% correct, as predicted if responses were random. For species abbreviations, see Table 1.

Shape, 26 Abdomen, and 32 Size, with 25 Complete taken from Experiment 1 as a control). Again, age (LR = 0.314, df = 1, *P* = 0.575) and gender (LR = 1.75, df = 1, *P* = 0.186) of participants were removed from the model as nonsignificant, as well as the interaction between phase and treatment (LR = 6.09, df = 3, *P* = 0.107). Fixed effects in the final model were species, phase, treatment, and the interaction between phase and species.

Participants' performances varied significantly according to treatment (LR = 22.38, df = 3, *P* < 0.0001), with highest scores in the Control, followed by Shape, then Abdomen, then Size (Figure 7). A Tukey post hoc test showed significant differences between the Control and both Abdomen (*P* = 0.003) and Size (*P* < 0.001), and between Shape and Size (*P* = 0.013). The lack of significant interaction between phase and treatment (see above), indicates that treatment differences persisted in the Testing phase.

There was a significant interaction between phase and species, indicating that the attack rates on different species were not consistent from one phase to the next (LR = 41.74, df = 11, *P* < 0.0001).

Model and mimic phenotypes

When the information from the Abdominal Color is summarized by Principal Components Analysis, the first 2 components explain 85.7% and 6.7% of the variation (Table 2). In these 2D, the model and mimic distributions overlap extensively (Figure 8a). By contrast, information from the abdominal patterns, when plotted in 2D using NMDS (stress = 0.23), shows that model and mimic distributions are clearly separated (Figure 8b). Size distributions for model and mimic individuals are very similar to each other (Figure 8c).

DISCUSSION

Our results show that humans acting as predators in our online game can readily distinguish between mimetic hoverflies and
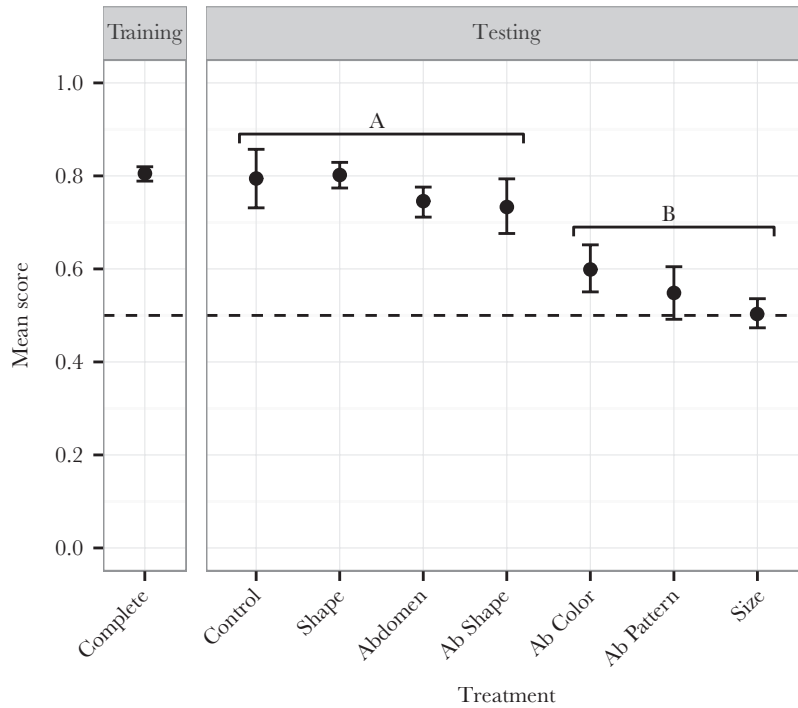
Figure 6

Variation [in] scores across treatments (Experiment 1). Points show the success rate in discriminating hoverflies from wasps during the relevant phase (excluding the first 3 questions of the phase, during which learning was ongoing) averaged across all volunteers for a given treatment. All participants viewed images of the complete insect during the Training phase, and then received one of 7 different treatments in the Testing phase. Error bars show 95% confidence intervals calculated from bootstrapping. The dashed line shows a score of 0.5, or 50% correct, as predicted if responses were random. All treatments from group A are significantly different from all those in group B according to a Tukey post hoc test ($P < 0.01$); there are no significant differences within either group ($P > 0.05$).

their wasp models using some traits, but not using others. When basing their decisions solely on shape and even when restricted to the abdominal shape, they were as successful in distinguishing models from mimics as when they were able to see all of the traits together. By contrast, the insect's size and the color and pattern of the abdomen were of little use to the participants, who failed to accurately identify models and mimics on those criteria. Our experiments build on previous studies that used more abstract stimuli, and support the idea that some traits (shape in particular) are likely to be under stronger selection for accurate mimicry than others. Inaccurate mimicry in other traits such as pattern may come about because predators simply do not perceive or process the relevant information.

In the first experiment, participants were best able to discriminate prey types on the basis of the shape (outline) of the insects. This result tallies with previous evidence that hoverflies and wasp shapes are clearly distinct from one another in multivariate space (Penney et al. 2012). One particular aspect of shape, the length of the antenna, is often noted as being a clear separator between hoverflies and wasps (Waldbauer 1970; Bain et al. 2007; Penney et al. 2012). While this trait could account for high success rates in the Shape treatment, success rates in the Abdomen Shape treatment, in which antennae were not visible, were almost as great. Here, the "pointiness" of the abdomen may have been the diagnostic trait; certainly a number of our participants mentioned in informal feedback that they had looked for the pointiness of the abdomen when making decisions. As a means of discriminating models from mimics, this particular trait has the advantage that it is directly associated with the aversive sting in the case of female Hymenoptera.

If shape is easily exploited to discriminate mimics from models, why then does selection not force mimics to resemble the shapes of their models more closely? Perhaps developmental and physiological constraints are simply too strong to be overcome in this case. For example, the distal segments of the hoverfly abdomen form the genital capsule (Speight 1987) and it is clear that modifying its shape to be more pointed and wasp-like could conflict with its essential primary function and cause a drastic reduction in mating success. In the case of antennae, some hoverflies have developed radical solutions to the problem of mimicking the length of the [ ] segments of female wasp antennae when syrphids have only 3 segments (Waldbauer 1970), but it is not clear why these solutions have not been adopted more widely among the wasp mimics. Alternatively, the response to shape by the humans in our experiments may not be seen in the real predators of mimetic hoverflies. However, birds have been shown to make use of shape information to discriminate among prey types when other traits are not informative, showing that they are able to process information from shape at least under some circumstances (Kazemi et al. 2014).

Three treatments (Abdomen Color, Abdomen Pattern, and Size) had success rates that were little or no different f[rom] picking at random between models and mimics. There are 2 possible explanations in each case: either the images in that treatment contained no information relevant to distinguishing models and mimics, or else there was useful information present that was overlooked. On the basis of summary data from each set of images (Figure 8), it appears that the former case applies to Abdomen Color and Size, as there is a great deal of overlap between model and mimic distributions. This suggests that, at least for the species included in
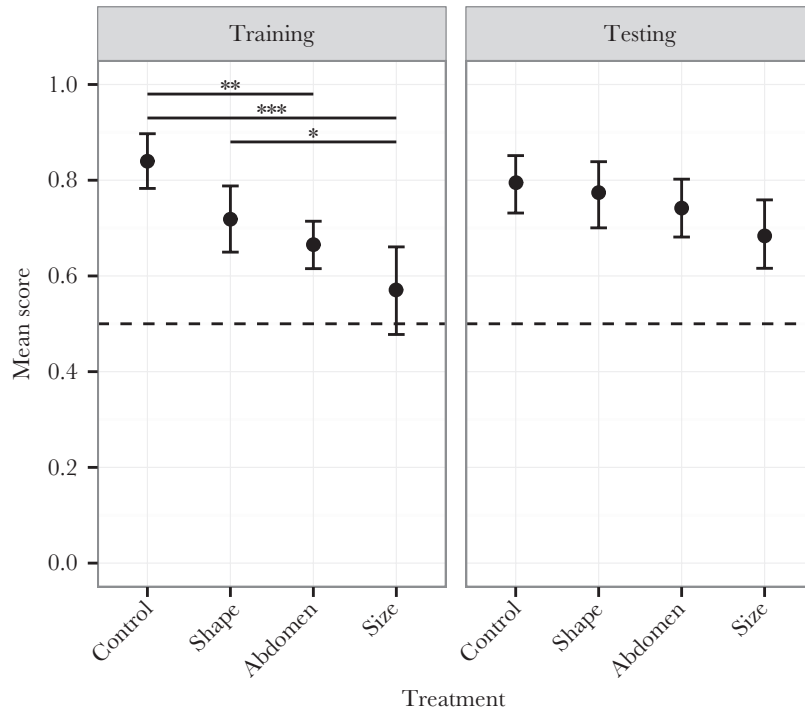
8.5

8.10

8.15

8.20

8.25

Figure 7
Variation in scores across treatments (Experiment 2). Points show the success rate in discriminating hoverflies from wasps during the relevant phase (excluding the first 3 questions of the phase, during which learning was ongoing) averaged across all volunteers for a given treatment. Participants received one of 4 different treatments during the Training phase, and then all viewed images of the complete insect in the Testing phase, but with behavior potentially influenced by the prior treatment. Error bars show 95% confidence intervals calculated from bootstrapping. The dashed line shows a score of 0.5, or 50% correct, as predicted if responses were random. Horizontal lines indicate pairs of treatments that are significantly different from each other (within the phase) according to a Tukey post hoc test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

8.30

8.35

Table 2
A description of the Principal Components created from color data

Variable	PC1	PC2
Black stripe R	0.194	−0.476
Black stripe G	0.229	−0.385
Black stripe B	0.188	−0.442
Yellow stripe R	0.627	0.291
Yellow stripe G	0.633	0.341
Yellow stripe B	0.283	−0.478
Proportion of variance	0.857	0.067

8.40

8.45

Values show the correlation between the original 6 variables (RGB color values for both the dark "black" stripes and the lighter "yellow" stripes) and the first 2 Principal Components. The 2 Principal Components displayed account for 92.4% of the total variance in the data.

8.50

this study, color and size mimicry is already relatively accurate. Color has been found to be highly salient to birds (Kazemi et al. 2014) and to humans (Sherratt et al. 2015), and it is therefore likely that mimetic colors have already undergone strong selection from predators for mimetic accuracy. Note, however, that the variety of species encountered will be a key factor. It is quite possible that a single mimic species would have reliably distinguishable colors from a single model species; but as more species are encountered, it becomes impossible to adopt a simple rule that separates one taxon from the other. Mimicry that appears "inaccurate" based on a single comparison may in fact be sufficient to provide protection from predation in a scenario where a number of model species are

8.55

8.60
8.61

encountered (Ihalainen et al. 2012; Easley and Hassall 2014), as in this study and likely faced by many wild predators.

It is interesting to note that participants did attack larger insects less frequently during the Size treatment, despite gaining no resulting advantage. Clearly our subjects had a pre-existing aversion to larger stimuli, which could indicate an assumption that larger prey are more likely to be dangerous. For example, even though horseflies are not included in our image sets, large hoverflies of similar appearance might elicit an avoidance reaction if a participant associates them with a previous negative experience. The bias in attack rates we observe in relation to size suggests the potential for selection on size as a mimetic trait, a possibility which has rarely been considered (though see Rashed et al. 2005). This will be challenging to study though, as prey size affects a number of potentially confounding factors such as the calorific value of the prey (Penney et al. 2012), as well as search and handling times by the predator.

Surprisingly, our participants were unable to perform better than random on the Abdominal Pattern treatment, despite the availability of information that separates models and mimics almost perfectly (Figure 8). Kazemi et al. (2014) found pattern to be less salient than color, but that birds would make use of pattern in the absence of other useful information. However, our data are not well explained by such an overshadowing effect, since participants did not learn to make use of the pattern information even after it was presented in isolation during the Testing phase. Instead, it appears that participants were genuinely unable to process the pattern information that was presented, perhaps due to the short time given to view the images. If real predators of hoverflies have similar limitations, this might explain inaccurate mimicry of wasp patterns,
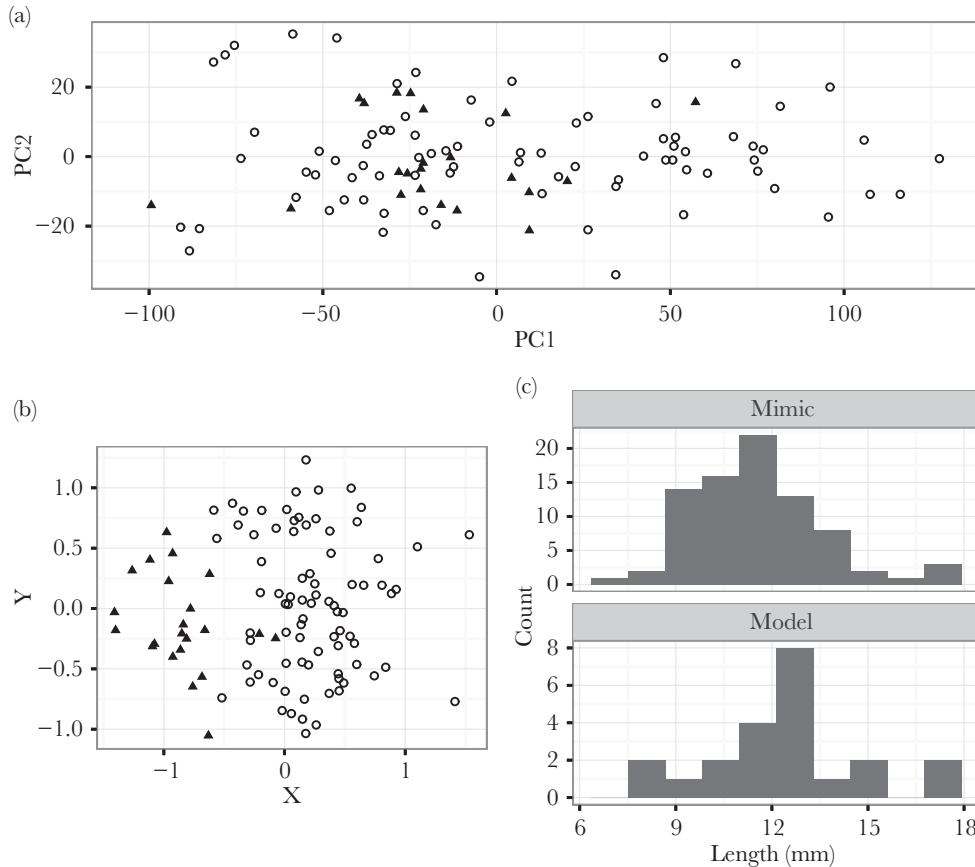
8.65

8.70

8.75

8.80

8.85

8.90

8.95

8.100

8.105

8.110

8.115

8.120

8.122

(a)

(b)

(c)

Figure 8
Variation in model and mimic phenotypes. a) Plot of Principal Components Analysis summarizing the RGB values (for both yellow and black stripes) from the Abdominal Color treatment. Each point represents a single individual model (triangle) or mimic (circle). The first 2 components are shown, accounting for 85.7% and 6.7% of the total variation respectively. See Table 2 for correlations between the Principal Components and the original variables. b) Plot of Non-metric Multi-Dimensional Scaling in 2D summarizing variation from the Abdominal Pattern treatment, based on pairwise distances between individuals calculated using the distance transform method (Taylor et al. 2013). Each point represents a single individual model (triangle) or mimic (circle). The stress value for this ordination is 0.23. c) Histogram summarizing body lengths of models and mimics in the Size treatment.

although at the other end of the spectrum, certain mimics have such accurate patterns that it is hard to imagine the similarity arising by chance.

Our results agree to some extent with those of Bain et al. (2007), who also found that shape played an important part in influencing birds' decisions to attack hoverflies. However, in that study pattern and color also had a role to play in predicting the birds' behavior. The different results with respect to color are probably explained by the fact that the hoverfly images in Bain et al. (2007)'s experiments covered a wider range of colors than the present study, including species such as *Ischyrosyrphus glaucius* whose markings are almost white. The lack of response from our participants to pattern, in contrast to Bain et al. (2007)'s findings, could reflect a difference between the visual processing of birds and humans, or differences in the way that images were presented (e.g., short duration in the present study). Alternatively, it could mean that the effects of color and pattern are not independent from each other, but that they combine synergistically (Marples et al. 1994). There is some scattered evidence that birds are capable of "gestalt" perception of mimetic patterns: that is, they can recognize not just individual traits, but the combinations in which they appear (Ikin and Turner, 1972; Terhune 1977). It has been suggested that internal contrast in a pattern may help to increase the salience of the colors within it (Aronsson and Gamberale-Stille

2013). Certainly, our participants were much more successful in the Full Abdomen treatment, in which pattern and color information were presented together, than the Abdominal Pattern treatment with pattern alone, but the shape information visible in the former treatment could also have been responsible for the improvement.

Results from the first experiment suggest that, in some treatments, participants either overlooked useful information or were unable to process it. We investigated this further in Experiment 2, by initially presenting participants with images showing only some stimulus traits (like the treatment images from the first experiment) in order to "train" them to pay more attention to those particular traits. This was followed by a Testing phase showing the full image. The initial Training phase resulted in a similar outcome to Experiment 1, with the Shape treatment giving the highest scores. These differences persisted into the subsequent Testing phase (although with smaller magnitude), despite all participants being shown the same (Complete) image types. This suggests that participants used differing strategies according to the way in which they had been "trained" in the initial phase. Alternatively, poorer performance following treatments such as Size could simply reflect a lack of opportunity to learn about the more informative traits. However, if that were the case, then participants should be able to perform at least as well in the Testing

phase as did the Control group in the Training phase, in contrast to the observed results. The "carry-over" we observed emphasizes that differences in early learning experiences (which in the wild can be varied and unpredictable) can have persistent effects on later behavior.

Several previous studies have adopted a similar approach of using humans as surrogate predators to judge the accuracy of mimics, and show the utility of using human test subjects in enabling a large sample size (Golding et al. 2005b; Penney et al. 2012; Kikuchi et al. 2015; Sherratt et al. 2015). Of course, results from human trials can only be applied to real mimetic systems with a degree of caution, although their results tally well with other studies that have used avian predators (Dittrich et al. 1993; Kazemi et al. 2014). It is also possible that previous insect encounters, or entomological knowledge, could have shaped participants' responses to the presented images. However, because treatments were assigned at random, the differences that we observed among treatments must be attributable to learning during the experiment rather than any past experience.

Past studies that have used humans as predators (Golding et al. 2005b; Kikuchi et al. 2015; Sherratt et al. 2015) have (to varying degrees) used the somewhat unrealistic scenario of the subjects being given time to inspect the images at their leisure, and in addition, Penney et al. (2012) presented large, high-resolution images. The present work used a short presentation of a low resolution image in order to best resemble the situation in which a predator must make a quick decision based on an insect viewed from a distance, with any hesitation allowing the prey a chance to escape (Chittka and Osorio 2007; Abbott and Sherratt 2013). Furthermore, presenting volunteers with a choice that is based on a limited subset of information allows conclusions to be drawn about which traits are the most relevant to the signal receiver (Gibson et al. 2005). Nonetheless, future studies could follow a similar format to the present work, replacing humans with more relevant predators, and perhaps using 3D models rather than photographs, in order to further increase the realism of the scenario.

Understanding how predators make decisions about which prey to attack will form a crucial part of understanding inaccurate mimicry. Many theoretical predictions regarding predator behavior towards models and mimics are based on Signal Detection Theory (Green and Swets 1966) and assume that the predator has access to global information about the prey populations and can dedicate unlimited cognitive resources to the problem (e.g., Getty 1985; Sherratt 2002). There is experimental evidence to support the validity of models like these, but it tends to come from scenarios that include a limited range of simple stimuli (Lindström et al. 1997; McGuire et al. 2006). There is an increasing appreciation that, to understand mimicry, we must take into account the complexity of the prey community (Easley and Hassall 2014) and the limited knowledge of the predator (Kikuchi and Sherratt 2015). By comparing results of simple experiments with others that confront predators with more complex prey communities and realistic situations, as done here and in other studies (Ihalainen et al. 2012; Easley and Hassall 2014), we can establish the circumstances under which selection will and will not act on inaccurate mimetic signals.

FUNDING

Data accessibility: Analyses reported in this article can be reproduced using the data provided by Taylor et al. (2016b).

Handling editor: Johanna Mappes

REFERENCES

Abbott KR, Sherratt TN. 2013. Optimal sampling and signal detection: unifying models of attention and speed–accuracy trade-offs. Behav Ecol. 24:605–616.

Abràmoff MD, Magalhães PJ, Ram SJ. 2004. Image processing with ImageJ. Biophotonics Intern. 11:36–42.

Aronsson M, Gamberale-Stille G. 2013. Evidence of signaling benefits to contrasting internal color boundaries in warning coloration. Behav Ecol. 24:349–354.

Bain RS, Rashed A, Cowper VJ, Gilbert FS, Sherratt TN. 2007. key mimetic features of hoverflies through avian eyes. Proc Biol Sci. 274:1949–1954.

Barber JR, Conner WE. 2007. Acoustic mimicry in a predator–prey interaction. Proc Natl Acad Sci USA. 104:9331–9334.

Bates HW. 1862. XXXII. Contributions to an insect fauna of the Amazon Valley. Lepidoptera: Heliconidæ. Trans Linn Soc Lond. 23:495–566.

Blough DS. 1969. Attention shifts in a maintained discrimination. Science. 166:125–126.

Canny J. 1986. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell. 8:679–698.

Chen DM, Goldsmith TH. 1986. Four spectral classes of cone in the retinas of birds. J Comp Physiol A. 159:473–479.

Chittka L, Osorio D. 2007. Cognitive dimensions of predator responses to imperfect mimicry. PLoS Biol. 5:e339.

Cuthill IC, Bennett ATD. 1993. Mimicry and the eye of the beholder. Proc R Soc Lond B. 253:203–204.

Dittrich W, Gilbert F, Green P, Mcgregor P, Grewcock D. 1993. Imperfect mimicry: a pigeon's perspective. Proc R Soc Lond B. 251:195–200.

Donner KO. 1951. The visual acuity of some passerine birds. Acta Zool Fenn. 66:1–40.

Easley JL, Hassall C. 2014. Field estimates of survival do not reflect ratings of mimetic similarity in wasp-mimicking hover flies. Evol Ecol. 28:387–396.

Getty T. 1985. Discriminability and the sigmoid functional response: how optimal foragers could stabilize model-mimic complexes. Am Nat. 125:239–256.

Gibson BM, Wasserman EA, Gosselin F, Schyns PG. 2005. Applying bubbles to localize features that control pigeons' visual discrimination behavior. J Exp Psychol Anim Behav Process. 31:376–382.

Golding YC, Edmunds M, Ennos AR. 2005a. Flight behaviour during foraging of the social wasp Vespula vulgaris (Hymenoptera: Vespidae) and four mimetic hoverflies (Diptera: Syrphidae) Sericomyia silentis, Myathropa florea, Helophilus sp. and Syrphus sp. J Exp Biol. 208:4523–4527.

Golding YC, Ennos AR, Sullivan M, Edmunds M. 2005b. Hoverfly mimicry deceives humans. J Zool. 266:395–399.

Green DM, Swets JA. 1966. Signal detection theory and psychophysics. New York: Wiley.

Green PR, Gentle L, Peake TM, Scudamore RE, McGregor PK, Gilbert F, Dittrich WH. 1999. Conditioning pigeons to discriminate naturally lit insect specimens. Behav Processes. 46:97–102.

Hossie TJ, Sherratt TN. 2014. Does defensive posture increase mimetic fidelity of caterpillars with eyespots to their putative snake models? Curr Zool. 60:76–89.

Ihalainen E, Rowland HM, Speed MP, Ruxton GD, Mappes J. 2012. Prey community structure affects how predators select for Mullerian mimicry. Proc Biol Sci. 279:2099–2105.

Ikin M, Turner JRG. 1972. Experiments on mimicry: gestalt perception and the evolution of genetic linkage. Nature. 239:525–527.

Jones RT, Le Poul Y, Whibley AC, Mérot C, ffrench-Constant RH, Joron M. 2013. Wing shape variation associated with mimicry in butterflies. Evolution. 67:2323–2334.

Kauppinen J, Mappes J. 2003. Why are wasps so intimidating: field experiments on hunting dragonflies (Odonata: Aeshna grandis). Anim Behav. 66:505–511.

Kazemi B, Gamberale-Stille G, Tullberg Birgitta S, Leimar O. 2014. Stimulus salience as an explanation for imperfect mimicry. Curr Biol. 24:965–969.

Kikuchi DW, Malick G, Webster RJ, Whissell E, Sherratt TN. 2015. An empirical test of 2-dimensional signal detection theory applied to Batesian mimicry. Behav Ecol. 26:1226–1235.

Kikuchi DW, Pfennig DW. 2010. Predator cognition permits imperfect coral snake mimicry. Am Nat. 176:830–834.

Kikuchi DW, Pfennig DW. 2013. Imperfect mimicry and the limits of natural selection. Q Rev Biol. 88:297–315.

Kikuchi DW, Sherratt TN. 2015. Costs of Learning and the Evolution of Mimetic Signals. Am Nat. 186:321–332.

Kraemer AC, Adams DC. 2014. Predator perception of Batesian mimicry and conspicuousness in a salamander. Evolution. 68:1197–1206.

Lindström L, Alatalo RV, Lyytinen A, Mappes J. 2004. The effect of alternative prey on the dynamics of imperfect Batesian and Müllerian mimicries. Evolution. 58:1294–1302.

Lindström L, Alatalo RV, Mappes J. 1997. Imperfect Batesian mimicry—the effects of the frequency and the distastefulness of the model. Proc R Soc Lond B. 264:149–153.

Marples NM, van Veelen W, Brakefield PM. 1994. The relative importance of colour, taste and smell in the protection of an aposematic insect *Coccinella septempunctata*. Anim Behav. 48:967–974.

Author. 2012. MATLAB. Natick (MA): The Mathworks.

McGuire L, Van Gossum H, Beirinckx K, Sherratt TN. 2006. An empirical test of signal detection theory as it applies to Batesian mimicry. Behav Process. 73:299–307.

Morris, Reader T. in review. The evolution of Batesian mimicry is affected by model profitability and the availability of alternative prey. Am Nat.

Penney HD, Hassall C, Skevington JH, Abbott KR, Sherratt TN. 2012. A comparative analysis of the evolution of imperfect mimicry. Nature. 483:461–464.

R Core Team. 2014. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Rashed A, Beatty CD, Forbes MR, Sherratt TN. 2005. Prey selection by dragonflies in relation to prey size and wasp-like colours and patterns. Anim Behav. 70:1195–1202.

Sherratt TN. 2002. The evolution of imperfect mimicry. Behav Ecol. 13:821–826.

Sherratt TN, Whissell E, Webster R, Kikuchi DW. 2015. Hierarchical overshadowing of stimuli and its role in mimicry evolution. Anim Behav. 108:73–79.

Shettleworth SJ. 2010. Cognition, evolution and behavior. 2nd ed. Oxford: OUP.

Skelhorn J, Dorrington G, Hossie TJ, Sherratt TN. 2014. The position of eyespots and thickened segments influence their protective value to caterpillars. Behav Ecol. 25:1417–1422.

Speight MCD. 1987. External morphology of adult Syrphidae (Diptera). Tijdschr Entomol. 130:141–175.

Taylor CH, Gilbert F, Reader T. 2013. Distance transform: a tool for the study of animal colour patterns. Methods Ecol Evol. 4:771–781.

Taylor CH, Reader T, Gilbert F. 2016. Hoverflies are imperfect mimics of wasp colouration. Evol Ecol 30:567–581.

Taylor CH, Warrin J, Gilbert F, Reader T. 2016. Data from: Which traits do observers use to distinguish Batesian mimics from their models? Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.k7057.

Terhune EC. 1977. Components of a visual stimulus used by scrub jays to discriminate a Batesian model. Am Nat. 111:435–451.

Valkonen JK, Nokelainen O, Mappes J. 2011. Antipredatory function of head shape for vipers and their mimics. PLoS One. 6:e22272.

Waldbauer GP. 1970. Mimicry of hymenopteran antennae by Syrphidae. Psyche. 77:45–49.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. 2009. Mixed Effects Models and Extensions in Ecology with R. New York (NY): Springer-Verlag.