Article:

Hassall, C orcid.org/0000-0002-3510-0728, Billington, J orcid.org/0000-0003-0995-8875 and Sherratt, TN (2019) Climate-induced phenological shifts in a Batesian mimicry complex. Proceedings of the National Academy of Sciences of the United States of America, 116 (3). pp. 929-933. ISSN 0027-8424

https://doi.org/10.1073/pnas.1813367115

Climate-induced phenological shifts in a Batesian mimicry complex

Christopher Hassall[1*], Jac Billington[2], Tom N. Sherratt[3]

[1]School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, UK
[2]School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, LS2 9JT, UK
[3]Department of Biology, Faculty of Science, Carleton University, Ottawa, Ontario, K1S 5B6, Canada

Christopher Hassall [c.hassall@leeds.ac.uk; +44 01133435578]
Jac Billington [J.Billington@leeds.ac.uk; +44 0113 343 6686]
Tom Sherratt [Tom.Sherratt@carleton.ca; +1 613 520 2600 x1748]

[*]Author for correspondence
Telephone: +44 1133435578
Fax: +44 113 343 2835

Running title: Phenological shifts in a mimicry complex

Summary

Climate-induced changes in spatial and temporal occurrence of species, as well as species traits such as body size, each have the potential to decouple symbiotic relationships. Past work has focused primarily on direct interactions, particularly those between predators and prey and between plants and pollinators, but studies have rarely demonstrated significant fitness costs to the interacting, co-evolving organisms. Here, we demonstrate that changing phenological synchrony in the latter part of the 20th century has different fitness outcomes for the actors within a Batesian mimicry complex, where predators learn to differentiate harmful "model" organisms (stinging Hymenoptera) from harmless "mimics" (hoverflies, Diptera: Syrphidae). We define the mimetic relationships between 2,352 pairs of stinging Hymenoptera and their Syrphidae mimics based on a large-scale citizen science project and demonstrate that there is no relationship between the phenological shifts of models and their mimics. Using computer game-based experiments, we confirm that the fitness of models, mimics, and predators differs among phenological scenarios, creating a phenologically-antagonistic system. Finally, we show that climate change is increasing the proportion of mimetic interactions in which models occur first and reducing mimic-first and random patterns of occurrence, potentially leading to complex fitness costs and benefits across all three actors. Our results provide the first strong evidence for an overlooked example of fitness consequences from changing phenological synchrony.

Keywords: hover flies, Batesian mimicry, phenology, mismatch, climate change

Significance statement

Climate change can degrade ecological interactions by separating interacting species in space and time, but this is not the case in one of the best-studied examples of mimicry in which hoverflies (mimics) imitate stinging wasps and bees (models). While there is no evidence of the emergence of mimics and models tracking climate change in the same way, historical records suggest that the mimicry complex is undergoing complex shifts in evolutionary pressures under climate change through changes in the relative emergence patterns of model-mimic pairs. This finding is based on the first community-level description of mimetic relationships (comparing 2,352 pairs of species) and the most comprehensive demonstration of the importance of phenology for the fitness of mimics, models, and predators.

/body

Introduction

The biological consequences of climate change for individual species have been documented in detail: distributions are moving poleward (1), phenological events are changing (most often documented as an advance in spring phenological events (2)), and organism size is reducing (3). These changes also have consequences for ecological interactions, as interacting species may become separated in space (4) or time (5, 6). The challenge for the study of these eco-evolutionary processes is the linking of (i) ecological data on the nature of relationships among taxa, (ii) the shifting spatiotemporal associations among those taxa, and (iii) data on the fitness costs and benefits that result from changes in the strength of interactions. Here, we provide a comprehensive evaluation of a putative symbiosis that has the potential to provide significant insights into community-level, climate-driven ecological change: model-mimic complexes. Mimetic relationships can be cooperative (e.g. Müllerian mimicry, where multiple defended species evolve a common phenotype as a result of selection to share the burden of educating predators (7)), or parasitic (e.g. Batesian mimicry, where an undefended species exhibits the phenotype of a defended species in order to benefit from the learned aversion of predators to that phenotype (8)). Batesian mimics (8) exploit a range of sensory modalities to enhance their similarity to defended models (9). One taxon that exploits multiple sensory cues is the hoverflies (Diptera: Syrphidae), which have evolved to produce visual (10), behavioural (11) and acoustic (12) cues that resemble those of stinging Hymenoptera.

Previous work has suggested that the order of appearance of mimics and models is important to the success of mimicry. In a classic study, Mostler demonstrated that avian predators learn to avoid stinging Hymenoptera quickly, but that being presented with palatable Diptera can reverse that learning (13). High fidelity syrphid mimics of stinging Hymenoptera in North America emerge predominantly in spring before their models (14, 15). It has been suggested that this phenological pattern has evolved so that the mimics emerge before naïve fledglings begin to feed independently and, to begin with, indiscriminately (16). Once the predators have been educated by the Hymenoptera models, the mimics show a second peak in emergence later in the summer. While consistent with theory on Batesian mimicry, other work has failed to support this idea, showing that mimics and models are largely synchronous and independent of fledging dates (17). However, no studies have attempted to test this idea comprehensively over large datasets and temporal ranges. A further source of uncertainty stems from the role of contemporary climate change in altering the spatiotemporal patterns of occurrence of mimics and models (18), with consequences for predator learning and, ultimately, the fitness of mimics, models, and the predators themselves.

Our understanding of the evolutionary consequences of global environmental change remains poorly developed, partly because of the difficulty in measuring the consequences of natural selection and partly due to interactions between phenotypic plasticity and adaptive change (19). In this paper we make use of alternative approaches to the study of evolutionary change under warming climates to quantify shifting phenological antagonism in a Batesian mimicry complex. This is achieved through (i) defining the mimetic relationships within a large pool of Batesian models and mimics, (ii) quantifying the phenological trends between those models and mimics, (iii) estimating empirically the fitness consequences of phenological asynchrony for mimics, models and predators, and finally (iv) demonstrating changes in optimal phenological patterns under recent climate change.

Results and Discussion

Study 1: Community-level mimetic networks

We selected 42 species of Syrphidae and 56 species of bees and wasps from UK species lists based on their abundance in UK biological records and their taxonomic and morphological distinctness (see Table S1 for the species list). All of the 2,352 species pairs are known to have co-occurred spatially in at least one 10 x 10 km grid square based on biological records, with Jaccard overlap indices (the ratio of shared squares to the number of squares in which one or both species was found) of 0.9-33.3% (mean 12.6%). A web interface was created that randomly paired single representative images from the Syrphidae and Hymenoptera and requested users to rate the pairing on a scale from 1 (not at all similar) to 10 (extremely similar). The experiment ran from 18 March 2016 to 28 March 2017 during which the 2,352 potential pairwise combinations had been rated a total of 30,300 times, with a minimum of 3 ratings and a maximum of 29 ratings for individual pairs. The modal rating was 1 (indicating negligible similarity) and the mean rating 3.0 ± 2.2 SD (see SI Appendix, fig. S1). We consider any mean ratings ≥5 to be indicative of a mimetic relationship, which was the case for 237 pairwise combinations (hereafter "high-fidelity pairs"). These 237 species pairs had Jaccard overlap indices of 2.6-28.4% (mean 13.5%). fig. 1 shows a matrix of those mimetic similarities, highlighting these "islands of mimicry" in the wider Batesian complex that form the basis for our subsequent analyses, and example pairs for different mean similarity scores (see SI Appendix, fig. S2 for a larger version of fig. 1 with species labelled). Our results correlate significantly with data from experiments using pigeons (20) ($r=0.757$, $p=0.030$, see SI Appendix, fig. S3), indicating that our human ratings are meaningful..

Study 2: Comparative phenology of models and mimics

We use the rank biserial correlation (RBC, a correlation between rank data [e.g. emergence date] and a categorical variable [e.g. species] (21)) to quantify the degree of overlap in phenology, where RBC=0 is random occurrence, RBC=-1 is all Hymenoptera models emerging before Syrphidae mimics and RBC=1 is all Syrphidae mimics emerging before Hymenoptera models. Average RBC was significantly lower than zero (median=-0.015, $V = 1709500000$, $p<0.001$) for the entire community of 56 Hymenoptera and 42 Syrphidae, showing that, on average, the Syrphidae emerge later than the Hymenoptera. When only high-fidelity model-mimic pairs were included, median RBC was slightly but significantly greater from zero (median=0.093, $V = 12092000$, $p<0.001$; see SI Appendix, fig. S7 for illustration of the relationship between RBC and difference in median flight date). The rate of phenological advance per unit increase in mean annual central England temperature in the leading edge (5th percentile) of the flight period is smaller in Syrphidae (-14.3 days·$°C^{-1}$ ± 0.9 SE) than in Hymenoptera (-18.9 days·$°C^{-1}$ ± 1.1 SE; t=22.806, p<0.001), but there is no difference in the rate of shift in the median flight date (Syrphidae: -8.1 days·$°C^{-1}$ ± 0.8 SE; Hymenoptera: -8.1 days·$°C^{-1}$ ± 0.8 SE; t=0.365, p=0.715) and the Syrphidae are advancing the trailing edge (95th percentile) of their flight period faster than Hymenoptera (Syrphidae: 1.3 days·$°C^{-1}$ ± 0.8 SE; Hymenoptera: 2.6 days·$°C^{-1}$ ± 1.1 SE; t=-6.491, p<0.001; see Supplementary Information for details). When considering only high-fidelity pairs, the results were quantitatively similar with Syrphidae (-14.4 ± 0.4 days·$°C^{-1}$) advancing the leading edge of the flight period more slowly than Hymenoptera (-19.1 ± 0.7 days·$°C^{-1}$; t=5.793, p<0.001), no difference in median flight date (Syrphidae: -7.0 ± 0.4 days·$°C^{-1}$; Hymenoptera: -7.1 ± 0.4 days·$°C^{-1}$; t=0.054, p=0.957) or the trailing edge of the flight period (Syrphidae: 2.1 ± 0.4 days·$°C^{-1}$; Hymenoptera: 2.8 ± 0.6 days·$°C^{-1}$; t=1.052, p=0.294). There is no significant relationship between the rates of phenological shift in high-fidelity models and mimics (leading edge: Spearman's ρ=-0.039, p=0.604; median: ρ=0.026, p=0.728; trailing edge: ρ=-0.004, p=0.959; dark points in SI Appendix, fig. S6). In other words, we find no evidence that models and mimics are advancing their phenology at the same rate.

Study 3: Fitness consequences of phenological mismatch

We quantified the fitness consequences of phenological change using a computer game-based behavioural experiment within which human participants could act as "predators" and make decisions concerning the profitability of three pairs of prey stimuli: (i) Apis mellifera and Eristalis tenax; (ii) Vespula vulgaris and Chrysotoxum cautum; and (iii) Bombus terrestris and Criorhina ranunculi (see SI Appendix, fig. S8 for stimuli and fig. S9 for an example screenshot). Participants were presented with all three model-mimic pairs in one of three phenological scenarios involving 25 models and 25 mimics: (i) mimics on average first,

(ii) models on average first, or (iii) random presentation with equal mean order of presentation. Mimic- and model-first scenarios were created by increasing the relative probability of the later species from 0% to 100% in increments of 2% over the 50 screens (example sequences are shown in Table S2), resulting in a mean RBC of 0.677 or -0.677 (SD=0.103) depending on whether the model or mimic occurs earlier. Participants gained 5 points for clicking a mimic, lost 10 points for clicking a model, and leaving the insect did not change the score.

The phenological scenario had a significant effect on predation rates on models (χ^2=49.218, df=2, p<0.001) and mimics (χ^2=34.544, df=2, p<0.001), and on the score achieved by human predators (χ^2=51.282, df=2, p<0.001; see Table S3 for full model results). Random presentation produced the highest fitness (survival rate) in mimics, and these were significantly higher than in model-first (z=3.073, p=0.006) or mimic-first scenarios (z=5.773, p<0.001; fig. 2A). The model-first sequence of prey items produced the highest fitness (survival rate) in models, and those outcomes were significantly higher than random (z=3.050, p=0.006) or mimic-first (z=6.983, p<0.001; fig. 2B), in agreement with the theory underpinning Batesian mimicry. However, random presentation produced significantly lower predator scores than mimic-first (z=5.390, p<0.001) or model-first (z=6.849, p<0.001; fig. 2C). These results highlight the phenological antagonism among the three actors within the mimicry system: models benefit from educating predators (model-first), mimics benefit from unpredictability (random), and predators benefit from consistent education on either prey item (model-first or mimic-first).

Temporal trends in fitness

Finally, we can infer temporal trends in fitness outcomes for model-mimic pairs under contemporary climate change based on how their RBCs change over time. From Study 3 we know that there are differences in fitness outcomes where the order of appearance of models and mimics corresponds to mean RBCs of <-0.677 (“model-first”), -0.677<RBC<0.677 (“random”) or >0.677 (“mimic-first”), and so we can apply those thresholds to the biological recording data to infer fitness consequences of real world sequences. Therefore, for each year, we classify each of the 237 high-fidelity pairs based on their RBC into a “model-first”, “random” or “mimic-first” pattern. There was a significant increase between 1960 and 2005 in the proportion of interactions in which the RBC corresponded to a model-first pattern (ρ=0.454, p=0.001, fig. 3A), a significant decline in the proportion of mimic-first sequences (ρ=-0.427, p=0.003, fig. 3B), and a weakly significant decline in the proportion of pairs in which the species occurred randomly (ρ=-0.295, p=0.044, fig. 3C; see SI Appendix, fig. S10 for sensitivity analysis around thresholds). Hence, we can infer a positive fitness trend for models from increasing model-first pairings (where models perform best) and decreasing mimic-first pairings (where models perform worst), a mixed fitness trend for mimics due to the decreasing proportion of random pairs (where mimics perform best) and a decrease in mimic-first pairs (where mimics perform worst), and positive fitness benefits for predators from the increasing proportions of model-first pairs (where predators perform best) and decreasing proportion of random pairs (where predators perform worst).

The evidence is building for a significant impact of phenological decoupling in a wide variety of systems. Snowshoe hares that moult after snowmelt show significant increases in mortality (22), many studies have shown that avian fitness is compromised if peak abundance of food does not coincide with chick rearing (e.g. (23)), and tritrophic studies suggest that oak-caterpillar-bird systems may have little room for buffering from warming springs (24). These exemplar studies are being carried out against a back drop of dynamic shifts in the degree of phenological synchrony over the past few decades (5, 25). However, while previous studies have tended to find the negative aspects of phenological shifts, our data suggest that climate change will result in an increase in phenologically-optimal emergence patterns that benefit (at least in part) all three actors within the mimicry system.

Building on past work (e.g. 13), we have now developed comprehensive evidence to support the theory (26) that the evolutionary costs and benefits of mimicry to models, mimics, and predators depend upon the relative phenological patterns of models and mimics. We have used these findings to help understand the implications of changes in the temporal overlap of models and mimics in a classical Batesian mimicry system. The different actors within the mimicry system each experience costs and benefits from different phenological patterns: hymenopteran models benefit in all cases from increased model-first, decreased mimic-first and decreased random patterns because these simplify – and, hence, accelerate – predator learning of aposematic signals. Mimics benefit from the decrease in mimic-first patterns and increase in model-first patterns because these accelerate predator learning of aposematic cues. However, mimics may suffer from a decline in randomness (their optimal phenological scenario according to Study 2), if predators are able to shift prey preferences as mimics increase in relative abundance later in the season. Finally, predators benefit from the reduction of randomness if they are able to respond by exploiting mimics when they become numerically dominant, but also show more subtle responses to declines in mimic-first and increased in model-first patterns. The results illustrate the benefits of integrating mechanistic and observational data to study large-scale eco-evolutionary processes within a phenologically antagonistic Batesian mimicry complex.

Methods

Study 1: Community-level mimetic networks

Rather than relying upon subjective, isolated descriptions of mimetic relationships between particular species from published work, we derived a full matrix of mimetic associations using an extensive, online citizen science project. Human ratings of visual similarity correlate with data from experiments with birds and with ratings based on morphometric analysis (10) and provide a Gestalt perspective on similarity that avoids issues with the definition of particular traits and with variations in both size and shape that can complicate computational image analysis (e.g. (27)). The online mimicry experiment, which can still be found at www.mimicryexperiment.net, was based around a simple PHP script that randomly selected two images – one from a pool of 42 Hymenoptera images and one from a pool of 56 Syrphidae images (one image for each species). The landing page of the online study contained brief details about the project and contact details for the lead author if participants required any further information. Participants were instructed to click on a button if they consented to take part. Study 1 was approved by the University of Leeds Faculty of Biological Sciences Research Ethics Committee (ref BIOSCI 16-006). Species were selected for inclusion based on a hierarchical process: first, the most common species were selected from the highest ranked abundance in the Hoverfly Recording Scheme (HRS) and the Bees, Wasps, and Ants Recording Scheme (BWARS) datasets (see below for details of those schemes). The rationale behind this criterion was that more common species are more likely to interact and, therefore, to have a true mimetic relationship if species with a similar morphology were present. We calculated the proportion of these putative model-mimic pairs that are known to have co-occurred in the same 10km grid square in the HRS and BWARS dataset. Secondly, congeners with close morphological similarity were excluded. There are many genera of UK Syrphidae and Hymenoptera with similar morphology and we refrained from using physically similar congeners to reduce phylogenetic autocorrelation and redundant morphological variation in the dataset (e.g. Melanostoma mellinum was included but M. scalare was excluded, while Lasioglossum leucozonium was included but L. villosulum was excluded). Thirdly, additional species were added to incorporate further morphological diversity where particularly distinct morphologies were known to the authors (particularly the rarer Chrysotoxum cautum, Criorhina ranunculi and Arctophila superbiens from among the Syrphidae and Vespula rufa from among the Hymenoptera, which are all thought to be involved in mimetic relationships (28)). Since the pairing of images was done at random, the number of comparisons between pairs of images was not equal across the dataset. The ratings gathered during the experiment exhibited a highly positively skewed distribution, with 53.2% of raw ratings and 56.1% of mean ratings being <3, while only 5.1% of raw ratings and 0.3% of mean pair ratings were >7 (see SI Appendix, fig. S1). Representative pairs of images are shown in fig. 1 for mean pair similarities of 7.5, 4.9, 3.1 and 1.0, along with a matrix of similarity values for all 2,352 comparisons. To validate the online experiment, we compared the ratings given to comparisons between eight of the hoverflies in our study that had previously been compared to Vespula vulgaris in pigeon experiments (20). In that previous experiment, pigeons were trained to peck at an image of Vespula vulgaris in return for a food reward, and then shown different Syrphidae images. The peck rate in response to the Syrphidae images was assumed to be proportional to the pigeon's perception of the similarity of the Syrphidae to the original training stimulus. Despite the low number of species in the pigeon data, as in previous studies that have used human ratings there was a significant correlation between our human ratings experiment and the pigeon peck rate data (r=0.757, p=0.030, see SI Appendix, fig. S3), indicating that our human ratings reflect morphological similarity in much the same way that birds might assess it.

Study 2: Comparative phenology of models and mimics

We extracted biological records from two extensive, long-term biological recording schemes, the HRS and BWARS datasets, to recreate past trends in phenology in model and mimic communities. Both HRS and BWARS are citizen science projects that rely on ad hoc recording by a community of recorders, and require compulsory fields relating to the species identity, location of sighting, and name of recorder, with desirable fields that describe ecological variables (e.g. flowers visited, pollen collected, prey/host). While abundance data is very occasionally recorded for each submitted record on a broad categorical scale, these data are rarely available and so have not been incorporated into this analysis. Validation of records for both schemes involves checking the validity of species names and geographical coordinates, while subsequent verification is based on identification difficulty, known spatial distributions, known seasonal phenology, and scarcity (29, 30). Prior to analysis, we cleaned the biological records to remove records from before 1960, records without ordinal dates, and records without valid species names. After this processing, the HRS contained 620,460 records of 288 hoverfly species from between 1960 and 2014 at time of analysis (27 January 2015). The BWARS dataset contained 451,624 records of 547 species from between 1960 and 2013 at time of analysis (27 January 2015). Biological recording data from the HRS and BWARS datasets show spatiotemporal patterns that are consistent with other UK biological recording data: a strong increase in the number of records through the latter part of the 20th Century (see SI Appendix, fig. S4) and a strong concentration of records in the south of the country and around centres of population density (see SI Appendix, fig. S5). Further descriptive statistics for the Syrphidae can be found in (31).

For each year in which a species was recorded, we calculated the 5th, 50th and 95th percentile flight dates (representing the leading edge, median, and trailing edge of the flight period, respectively) and then conducted linear regressions of these dates against a general measure of mean annual UK temperature (central England temperature [CET] (32)). CET uses an average of values from a set of three long-term meteorological stations in central England to create an averaged trend over the country. We calculated Pearson correlation coefficients to provide a measure of the strength of the trends, and regression coefficients to provide a measure of the rate of change in phenology. A comparison of model and mimic shifts across the three flight dates can be seen in SI Appendix, fig. S6.

To quantify the relative phenology of the model and mimic communities, we assumed that the sequence of occurrence is the relevant phenological metric for learning within a Batesian mimicry system. We calculated rank biserial correlations (RBC) of the flight dates of all potential model-mimic pairs and for the subset of pairs for which mean similarity ratings were >=5. The RBC approach gives a correlation coefficient ranging from -1 (no overlap, first sample entirely before second sample) to +1 (no overlap, first sample entirely after second sample). An RBC of 0 indicates random occurrence. To illustrate the relationship between RBCs and differences in flight dates within the BWARS and HRS datasets, the RBCs calculated for the real-world species pairs from the BWARS and HRS datasets are plotted against the difference in median flight dates in SI Appendix, fig. S7.

Study 3: Fitness costs of phenological mismatch

We designed a computer game within which human participants could act as “predators” and make decisions concerning the profitability of different “prey”. A psychological approach was selected above computational algorithms for a number of reasons. First, as we note elsewhere in the manuscript, we have demonstrated that human ratings of similarity are correlated with those of avian model systems. Second, the available algorithms for quantifying model-mimic similarity such as the distance transform method (27), and the neural net approach of Bain et al. (33) are highly data hungry, requiring extensive data to parameterize them compared to our “Gestalt” system based on human assessments. There is no reason to expect that these computational methods would produce a more ecologically relevant result than the use of human scores. Indeed, the neural network classifier of Bain et al. (33) ranked the similarity of hoverfly species to wasps in a very similar manner to humans (R2 = 0.74, P <0.001; 34). Third, we were interested in generating a fitness measure that was more than just similarity but also included speed-accuracy trade-offs that are important to prey survival (35). The game was built in the Vizard (WorldViz) virtual reality environment, programmed in Python, and displayed using an Oculus Rift DK2 immersive virtual reality headset (full field of view horizontal visual angle = 100°) while participants were seated. Participants were given an opportunity to read an information sheet describing the study and ask any questions before giving written informed consent to take part. Study 3 was approved by the University of Leeds Faculty of Biological Sciences Research Ethics Committee (ref BIOSCI 15-021). The experiment began with a short training phase during which participants were asked to “eat” (click) red triangles and leave blue circles, and a scoreboard kept track of their points. All participants performed well during this phase.

Following the training phase, participants were presented with the experimental stimuli. These stimuli comprised three pairs of insect images chosen to represent the three broad groups of model-mimic relationships within the UK Syrphidae-Hymenoptera community: (i) Apis mellifera and Eristalis tenax; (ii) Vespula vulgaris and Chrysotoxum cautum; and (iii) Bombus terrestris and Criorhina ranunculi (see SI Appendix, fig. S8 for stimuli). Participants were presented with a series of sequences of insect images against a grass background to enhance the noise in the image and reduce the use of image artefacts (e.g. cropping of images) as cues to the identity of the insects (see SI Appendix, fig. S9). Each screen contained only one insect at any time. The insects were displayed in random positions and orientations on each trial, all within a 30° central visual angle (horizontally and vertically). The insect images covered approximately 5° visual angle. The three pairs of insects were presented in one of three phenological scenarios involving 25 models and 25 mimics: (i) mimics more likely to be first, (ii) models more likely to be first, or (iii) random presentation (example sequences are shown in Table S2). Sequences were generated by increasing the probability of occurrence of one image from 0% to 100% in increments of 2% over the 50 presentations, producing a series of unique model-first sequences with a mean RBC of 0.678 (SD=0.103). Sequences were reversed to give the mimic-first sequences in order to retain the same RBC. Each participant was asked to decide whether or not to "eat" an insect by pressing a key on a keyboard, after which a score counter would change to reflect whether the decision was "correct": consuming harmless Syrphidae increased the score by 5 points, consuming stinging Hymenoptera reduced the score by 10 points, and leaving the insect did not change the score. Participants were told that "eating" some types of insects would reduce their score while eating others would increase it, and that the aim was to score as many points as possible.

We analysed the data from 45 participants to evaluate the consequences of phenological scenario on (i) mimic survival, (ii) model survival, and (iii) final participant scores (as a proxy for predator fitness). Each participant experienced all three of the phenological scenarios, where each scenario was presented using a different model-mimic pair so that the participants also saw all three model pairs (e.g. a participant may have seen B. pratorum and C. ranunculi in the model-first scenario, A. mellifera and E. tenax in the mimic-first scenario, and V. vulgaris and C. cautum in the random scenario). For tests (i), and (ii) we used generalised linear mixed effects models with binomial errors in the lme4 package in R to analyse the survival or predation of each target as a binary variable. The trial number (from 1 to 150) was a covariate and phenological scenario was a fixed effect, while the model-mimic pair and the participant ID were entered as random effects. For analysis (iii) the same models were run but with general linear mixed effects models with the participant score as the response and normal error distribution. The residuals of the model for analysis (iii) were checked to ensure that the data met the assumptions of normality and homogeneous variance. Full model results can be found in Table S3.

Temporal Trends in Fitness

To explore the sensitivity of the temporal trends (fig. 3, main text) to the choice of thresholds, we conducted a sensitivity analysis by recalculating the RBC trends for thresholds that were 1 standard deviation above (-0.781 and 0.781, for model-first and mimic-first, respectively) and below (-0.574 and 0.574) the mean of the distribution of RBCs (-0.677 and 0.677). The results were qualitatively robust to the variation in threshold, with the number of model-first scenarios always increasing (-1SD: rho=0.428, p=0.002; mean: rho=0.455, p=0.001; +1SD: rho=0.489, p=0.001), mimic-first scenarios always decreasing but not always significantly so (-1SD: rho=-0.607, p<0.001; mean: rho=-0.427, p=0.003; +1SD: rho=-0.120, p=0.431), and random scenarios always decreasing but not always significantly so (-1SD: rho=-0.158, p=0.285; mean: rho=-0.295, p=0.044; +1SD: rho=-0.414, p=0.005). SI Appendix fig. S10 shows the comparison of time series.

Acknowledgements

C.H. was supported by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Programme (EcoEvoMimic). T.N. S. is supported by a Natural Sciences and Engineering Research Discovery Grant. We would like to thank Richard Cairns and Max Cairns for assistance with the online citizen science experiment and Steve Falk for kindly allowing us to use his insect photographs.

Author contributions

C.H. and T.N.S. conceived the project; C.H. conducted the citizen science experiment, conducted the behavioural experiment, analysed the biological records, and analysed the data; J.B. coded the behavioural experiment and advised on behavioural methods. All authors discussed the results and provided comments on the manuscript.

References

1. Hickling R, Roy DB, Hill JK, Fox R, & Thomas CD (2006) The distributions of a wide range of taxonomic groups are expanding polewards. Global Change Biology 12:1-6.
2. Thackeray SJ, et al. (2010) Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. Global Change Biology 16(12):3304-3313.
3. Daufresne M, Lengfellner K, & Sommer U (2009) Global warming benefits the small in aquatic ecosystems. Proceedings of the National Academy of Sciences 106(31):12788-12793.
4. Visser ME, Holleman LJM, & Gienapp P (2006) Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. Oecologia 147:164-172.
5. Iler AM, et al. (2013) Maintenance of temporal synchrony between syrphid flies and floral resources despite differential phenological responses to climate. Global Change Biology 19(8):2348-2359.

6. Rafferty NE, CaraDonna PJ, & Bronstein JL (2015) Phenological shifts and the fate of mutualisms. Oikos (Copenhagen, Denmark) 124(1):14-21.
7. Müller F (1879) Ituna and Thyridia; a remarkable case of mimicry in butterflies. Transactions of the Entomological Society of London 1879:20–29.
8. Bates H (1862) Contribution to an insect fauna of the Amazon Valley, Lepidoptera: Heliconidae. Transactions of the Linnean Society of London 23:495-566.
9. Quicke DLJ (2017) Mimicry, crypsis, masquerade and other adaptive resemblances (Wiley Blackwell, Hoboken, NJ).
10. Penney HD, Hassall C, Skevington JH, Abbott KR, & Sherratt TN (2012) A comparative analysis of the evolution of imperfect mimicry. Nature 483(7390):461-464.
11. Penney HD, Hassall C, Skevington JH, Lamborn B, & Sherratt TN (2014) The relationship between morphological and behavioral mimicry in hover flies (Diptera: Syrphidae). The American Naturalist 183(2):281-289.
12. Moore CD & Hassall C (2016) A bee or not a bee: an experimental test of acoustic mimicry by hoverflies. Behavioral Ecology 27:1767-1774.
13. Mostler G (1935) Observations on the question of wasp mimicry. Z Morph Okol Tiere 29:381-454.
14. Waldbauer GP & LaBerge WE (1985) Phenological relationships of wasps, bumblebees, their mimics and insectivorous birds in northern Michigan. Ecological Entomology 10(1):99-110.
15. Waldbauer GP, Sternburg JG, & Maier CT (1977) Phenological relationships of wasps, bumblebees, their mimics, and insectivorous birds in an Illinois sand area. Ecology 58(3):583-591.
16. Waldbauer GP (1988) Asynchrony between Batesian mimics and their models. The American Naturalist 131:S103-S121.

17. Howarth B & Edmunds M (2000) The phenology of Syrphidae (Diptera): are they Batesian mimics of Hymenoptera? Biological Journal of the Linnean Society 71(3):437-457.

18. Hassall C (2015) Odonata as candidate macroecological barometers for global climate change. Freshwater Science 34(3):1040-1049.

19. Merilä J & Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evolutionary Applications 7(1):1-14.

20. Dittrich W, Gilbert F, Green P, McGregor P, & Grewcock D (1993) Imperfect mimicry - a pigeon's perspective. Proceedings of the Royal Society: Series B (Biological Sciences) 251:195-200.

21. Cureton EE (1956) Rank-biserial correlation. Psychometrika 21(3):287-290.

22. Zimova M, Mills LS, & Nowak JJ (2016) High fitness costs of climate change-induced camouflage mismatch. Ecology Letters 19(3):299-307.

23. Reed TE, Jenouvrier S, & Visser ME (2013) Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. J. Anim. Ecol. 82(1):131-144.

24. Burgess MD, et al. (2018) Tritrophic phenological match–mismatch in space and time. Nature Ecology & Evolution 2(6):970-975.

25. Kharouba HM, et al. (2018) Global shifts in the phenological synchrony of species interactions over recent decades. Proceedings of the National Academy of Sciences 115(20):5211-5216.

26. Bobisud LE (1978) Optimal Time of Appearance of Mimics. The American Naturalist 112(987):962-965.

27. Taylor C, Gilbert F, & Reader T (2013) Distance transform: a tool for the study of animal colour patterns. Methods in Ecology and Evolution 4(8):771-781.

28. Howarth B, Clee C, & Edmunds M (2000) The mimicry between British Syrphidae (Diptera) and aculeate Hymenoptera. British Journal of Entomology and Natural History 13:1-40.

29. Dean H (2013) Verification Rules for NBN Record Cleaner (National Biodiversity Network, Nottingham, UK).

30. Ball SG & Morris RKA (2012) The Hoverfly Recording Scheme - putting Diptera on the map. Antenna 36(3):177-185.

31. Hassall C, Owen J, & Gilbert F (2017) Phenological shifts in hoverflies (Diptera: Syrphidae): linking measurement and mechanism. Ecography 40(7):853-863.

32. Parker DE, Legg TP, & Folland CK (1992) A new daily Central England Temperature Series, 1772-1991. International Journal of Climatology 12:317-342.

33. Bain RS, Rashed A, Cowper VJ, Gilbert FS, & Sherratt TN (2007) The key mimetic features of hoverflies through avian eyes. Proceedings of the Royal Society: Series B (Biological Sciences) 274:1949-1954.

34. Rashed A (2006) Evolution and maintenance of Batesian Mimicry, with particular references to hover flies (Diptera: Syrphidae). PhD (Carleton University, Ottawa).

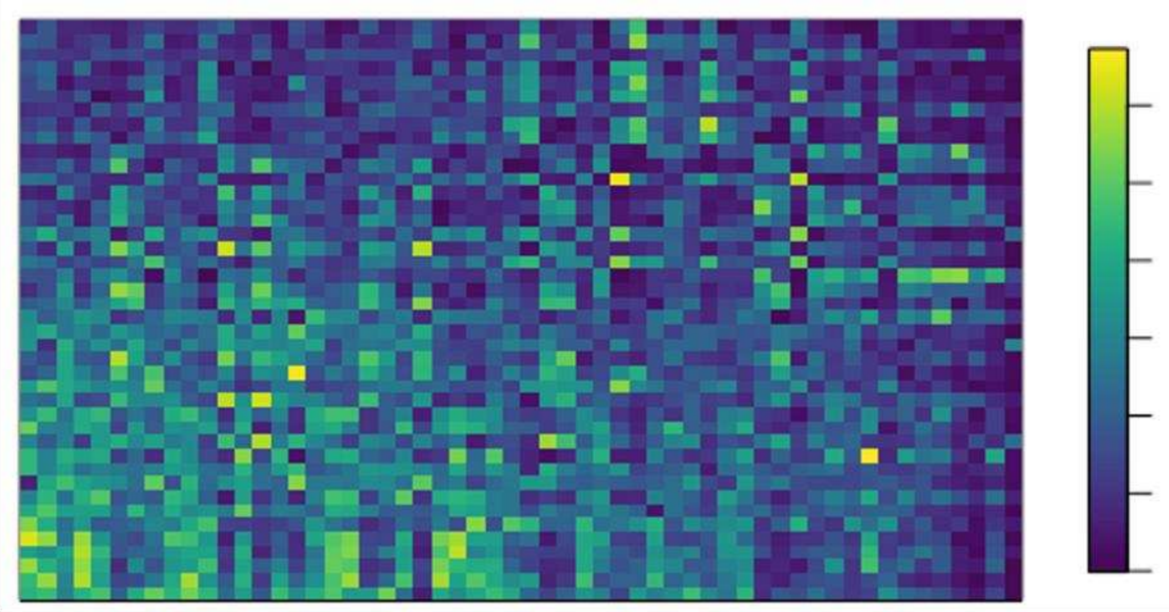
35. Chittka L & Osorio D (2007) Cognitive dimensions of predator responses to imperfect mimicry. PLoS Biology 5:e339.

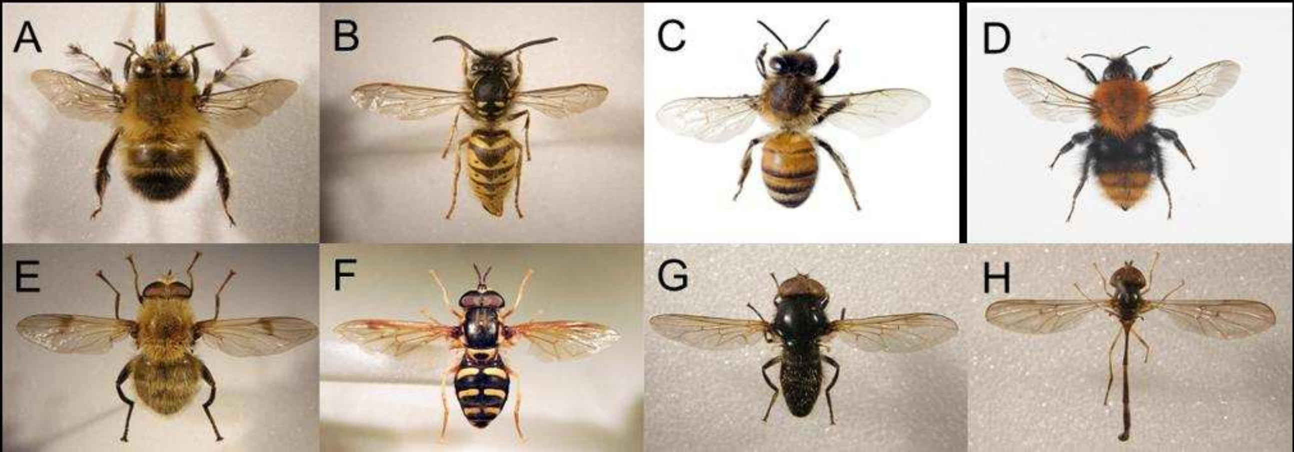
Figure legends

Fig. 1. Above: Heat map of mimetic ratings between 56 Syrphidae and 42 Hymenoptera. Colours indicate the mean similarity rating for each pair. The heat map has been ordered to place the Syrphidae with the highest mean similarity on the left of the plot and the Hymenoptera with the highest mean similarity at the bottom of the plot. See SI Appendix, fig. S2 for an enlarged version with species names. Below: Representative pairs of Hymenoptera (upper row) and Syrphidae mimics (lower row) that were rated as high similarity (A=Anthophora plumipes, E=Arctophila superbiens, rating=7.5), medium-high similarity (B=Vespula vulgaris, F=Chrysotoxum festivum, rating=4.9), low-medium similarity (C=Apis mellifera, G=Cheilosia impressa, rating=3.1), and low similarity (D=Bombus pascuorum, H=Baccha elongata, rating=1.0). Photographs copyright Steven Falk and used with permission.

Fig. 2. Fitness consequences of phenological asynchrony in Batesian mimics, models, and predators. Dotted horizontal line at y=1 in (A) and (B) shows model-first odds ratio, against which the other two phenological patterns are compared: (A) Batesian mimics are predated more under mimic-first scenarios, benefitting most from model-first scenarios. (B) Models are predated more often under mimic-first scenarios, benefitting more from random scenarios. (C) Predator fitness is greatest under model-first scenarios, then mimic-first, and lowest under random scenarios. Error bars are 95% confidence intervals.

Fig. 3. Trends in the number of high-fidelity model-mimic pairs in which (A) the model emerges first, (B) the mimic emerges first, and (C) emergence is random. Mimetic fidelity is derived from a large citizen science study, phenological trends are derived from >1 million biological records, and the three categories of emergence are defined using known fitness consequences from novel behavioural experiments. Shaded areas are 95% confidence intervals.




A
B
C
D
E
F
G
H


A
Mimic odds ratio
5.0
2.0
1.0
0.5
0.2
Model first (reference)
Mimic first
Random
Scenario
B
Model odds ratio
5.0
2.0
1.0
0.5
0.2
Model first (reference)
Mimic first
Random
Scenario
C
Predator fitness (mean score)
450
425
400
375
350
Model first
Mimic first
Random
Scenario


A
Proportion mimic first pairs
0.0
0.1
0.2
0.3
0.4
0.5
1960
1980
2000
2020
Year
B
Proportion random pairs
0.00
0.05
0.10
0.15
0.20
0.25
0.30
1960
1980
2000
2020
Year
C
Proportion model first pairs
0.4
0.5
0.6
0.7
0.8
1960
1980
2000
2020
Year