# Appendix 1 – Mortality due to predation

## 0 Purpose

Here, we describe how the mortality caused by predation was implemented in the model, using a step-by-step explanation.

# 1 Original equation (Joron & Iwasa, 2005)

When facing aposematic prey, predators need to sample a certain number of a prey with a given phenotype per unit time to learn and avoid this particular phenotype. As a result, the predation rate becomes a strongly nonlinear, hyperbolically decreasing function of relative prey density in the population. Under such nonlinear density-dependence, selection against very rare phenotypes should be enhanced at very low frequencies, and should weaken as frequency increases.

Following Joron & Iwasa (Journal of Theoretical Biology 2005, 237:87-103; doi: 10.1016/j.jtbi.2005.04.005), we account for such nonlinear density-dependence and we consider that the baseline mortality factors d is multiplied by 1/(1 + sN)which produces a hyperbolic decrease in mortality with the density N of the focal phenotype when s > 0. Parameter s therefore encompasses any characteristic that determines the efficiency of associative learning (level of toxins, and memorability of the signal).

The predation rate on individuals with the focal phenotype, is:

$$P = \frac{dN}{1+sN} \tag{A1}$$

The numerator of this equation represents the baseline mortality rate due to predation, and the denominator represents the reduction in predation due to associative learning by predators (which is more efficient as the density of individuals with the focal phenotype increases).

## 2 Perfect mimicry between ancestral and derived phenotypes

From equation 1, we can express the mortality of individuals (with ancestral or derived phenotypes) that are perfect mimics as follows:

$$P_{\rm a} = \frac{d N_{\rm a}}{1 + s \left(N_{\rm a} + N_{\rm d}\right)} \tag{A2}$$

$$P_{\rm d} = \frac{d N_{\rm d}}{1 + s \ (N_{\rm a} + N_{\rm d})} \tag{A3}$$

If the predators cannot perceive a difference between the two phenotypes, then all individuals suffer the same predation risk per capita  $\left(\frac{P_{a}}{N_{a}} = \frac{P_{d}}{N_{d}}\right)$  and benefit from a lower predation risk due to mimicry (term  $N_{a} + N_{d}$  in the denominator).

## **3** Imperfect mimicry between ancestral and derived phenotypes

We can now express the mortality of individuals (with ancestral or derived phenotypes) that are imperfect mimics, with resemblance captured by a quantity S, as follow:

$$P_{\rm a} = \frac{d N_{\rm a}}{1 + s \left(N_{\rm a} + S N_{\rm d}\right)} \tag{A4}$$

$$P_{\rm d} = \frac{d N_{\rm d}}{1 + s \left(S N_{\rm a} + N_{\rm d}\right)} \tag{A5}$$

If S = 0, then ancestral and derived phenotypes are perceived as completely different by predators. If S = 1, then ancestral and derived phenotypes are perfect mimics.

We assume that the similarity level depends on the difference in conspicuousness,  $c_{\rm a} - c_{\rm d}$ , and is expressed as a Gaussian generalization function, as defined in the main text:

$$S = \exp\left(-\gamma \sqrt{\left(c_{\rm a} - c_{\rm d}\right)^2 + l^2}\right) \tag{A6}$$

The parameter  $\gamma$  describes the generalization behaviour of predators, i.e. how much they perceive phenotypic differences. The parameter *l* represents the distance between the ancestral and the derived colour patterns; i.e. the phenotypic distance that is not related to differences in conspicuousness. See Ruxton *et al.* (2008) (*Evolution*, **62**(11), 2913-2921) for more details on this function.

## 4 Mimicry rings

To assume that other comimetic species reduce the predation risk by enhancing predator learning, we include additional terms in the denominator. Remember that the denominator represents the reduction in predation due to associative learning by predators.

$$P_{\rm a} = \frac{d}{1 + s \, (N_{\rm a} + S \, N_{\rm d}) + M_{\rm a} + S \, M_{\rm d}} \tag{A7}$$

$$P_{\rm a} = \frac{a}{1 + s \, (S \, N_{\rm a} + N_{\rm d}) + S \, M_{\rm a} + M_{\rm d}} \tag{A8}$$

High parameters values of  $M_{\rm a}$  and  $M_{\rm d}$  lead to decreased predation risks. Those parameters therefore describes the efficiency of the mimicry ring (i.e., the comimetic defended species) in reducing predation risk by enhancing associative learning by predators. The accuracy of mimicry with those other defended species depends on the resemblance between the phenotype carried by individuals and the mean phenotypes displayed in the comimetic species. We thus assume that the contribution of  $M_{\rm a}$  and  $M_{\rm d}$  in reducing predation risks is modulated by a factor that determines this resemblance. For simplicity, we assume that a derived phenotype that is an imperfect mimic of the ancestral phenotype is an equally imperfect mimic of the co-mimetic individuals that resemble the ancestral phenotype. We make the same assumption to describe the resemblance between ancestral phenotype and individuals from the mimicry rings matching the derived phenotype.

## 5 Increased detectability due to high conspicuousness

We assume that high conspicuousness results in a high risk of being detected by predators, thereby increasing the baseline mortality factor. Therefore, individuals with ancestral and derived phenotypes may have different baseline mortality factors,  $d_{\rm a}$  and  $d_{\rm d}$ . In particular we assume increased conspicuousness increases the baseline mortality factor:

$$d_{\rm a} = p \, c_{\rm a} \tag{A9}$$

$$d_{\rm d} = p \, c_{\rm d} \tag{A10}$$

$$P_{\rm a} = \frac{d_{\rm a} N_{\rm a}}{1 + s \ (N_{\rm a} + S N_{\rm d}) + M_{\rm a} + S M_{\rm d}} \tag{A11}$$

$$P_{\rm d} = \frac{d_{\rm d} N_{\rm d}}{1 + s \left(S N_{\rm a} + N_{\rm d}\right) + S M_{\rm a} + M_{\rm d}} \tag{A12}$$

## 6 Increased associative learning due to high conspicuousness

We assume that high conspicuousness may result in more efficient associative learning by predators. Therefore, individuals with ancestral and derived phenotypes are associated with different density-dependence factors,  $s_a$  and  $s_d$ . In particular, we assume that increased conspicuousness increases the density-dependence factor depending on how memorable the colour pattern is (modulated by parameter  $\beta_a$  and  $\beta_d$ ):

$$s_{\rm a} = u \,\beta_{\rm a} \,c_{\rm a} \tag{A13}$$

$$s_{\rm d} = u \,\beta_{\rm d} \,c_{\rm d} \tag{A14}$$

$$P_{\rm a} = \frac{d_{\rm a} N_{\rm a}}{1 + (s_{\rm a} N_{\rm a} + S s_{\rm d} N_{\rm d}) + M_{\rm a} + S M_{\rm d}}$$
(A15)

$$P_{\rm d} = \frac{d_{\rm d} N_{\rm d}}{1 + (S s_{\rm a} N_{\rm a} + s_{\rm d} N_{\rm d}) + S M_{\rm a} + M_{\rm d}}$$
(A16)

# 7 Implications of alternative edible prey

The presence of alternative edible prey can affect the population dynamics by making predators more or less hungry. We assume that a parameter h adjusts the prey baseline mortality rate, so that even highly cryptic prey (with  $c_a$  or  $c_d$  equal to 0) can be attacked. The parameter h can reflect the abundance of alternative edible prey. Indeed, hungry predators have more incentive to search for cryptic prey compared to well-fed predators, thus h = 0 occurs when there is a lot of alternative prey and high h > 0 occurs when there are few alternative edible prey. Parameter h therefore modulate the baseline mortality rate:

$$d_{\rm a} = p \ (c_{\rm a} + h) \tag{A17}$$

$$d_{\rm d} = p \ (c_{\rm d} + h) \tag{A18}$$

## 8 Final system of equation

Overall, we get the system of equations presented in the manuscript:

$$d_{\rm a} = p \ (c_{\rm a} + h) \tag{A19}$$

$$d_{\rm d} = p \ (c_{\rm d} + h) \tag{A20}$$

$$s_{\rm a} = u \,\beta_{\rm a} \,c_{\rm a} \tag{A21}$$

$$s_{\rm d} = u \,\beta_{\rm d} \,c_{\rm d} \tag{A22}$$

$$P_{\rm a} = \frac{d_{\rm a} N_{\rm a}}{1 + (s_{\rm a} N_{\rm a} + S s_{\rm d} N_{\rm d}) + M_{\rm a} + S M_{\rm d}}$$
(A23)

$$P_{\rm d} = \frac{d_{\rm d} N_{\rm d}}{1 + (S s_{\rm a} N_{\rm a} + s_{\rm d} N_{\rm d}) + S M_{\rm a} + M_{\rm d}}$$
(A24)

$$S = \exp\left(-\gamma \sqrt{\left(c_{\rm a} - c_{\rm d}\right)^2 + l^2}\right) \tag{A25}$$

# Appendix 2 – Analytical derivations

# 0 Purpose

We here derive analytically the conditions under which a derived phenotype can invade in the population, assuming either 'perfect mimicry within a single mimicry ring' or a 'complete mimicry shift', as defined in the main text.

Assuming 'perfect mimicry within a single mimicry ring', the derived phenotype differs from the ancestral phenotype by its conspicuousness only. Assuming a 'complete mimicry shift', the derived phenotype differ from the ancestral phenotype by its conspicuousness but also by its colour pattern (as assumed by the existence of an alternative mimicry ring).

We perform here local stability analyses. First, we study the dynamical system without the derived phenotype and we derive the density of individuals carrying the ancestral phenotype at equilibrium (Section 1). Second, we derive the conditions under which individuals with a derived phenotype have a positive growth rate, assuming that they are very rare initially, and that the density of individuals carrying the ancestral phenotype is at its equilibrium value (Section 2 assuming 'perfect mimicry within a single mimicry ring', Section 3 assuming a 'complete mimicry shift')

# 1 Study of the dynamical system without the derived phenotype

Without individuals carrying the derived phenotype  $(n_d = 0)$ , the systems of equations are identical assuming 'perfect mimicry within a single mimicry ring' and assuming a 'complete mimicry shift':

$$\frac{\mathrm{d}n_{\mathrm{a}}}{\mathrm{d}\tau} = n_{\mathrm{a}}(1-n_{\mathrm{a}}) - \frac{\delta c_{\mathrm{a}}n_{\mathrm{a}}}{1+\lambda_{\mathrm{a}}c_{\mathrm{a}}n_{\mathrm{a}}+M_{\mathrm{a}}} \tag{B1}$$

Let  $n_a^*$  be the density of individuals carrying the ancestral phenotype at equilibrium, *i.e.* satisfying the condition  $\frac{dn_a}{d\tau} = 0$ . Here, we are placed in conditions under which the population composed only of individuals with the ancestral phenotype does not get extinct, *i.e.*  $n_a^* > 0$ . We have :

$$n_{\rm a}^*(1 - n_{\rm a}^*) - \frac{\delta c_{\rm a} n_{\rm a}^*}{1 + \lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}} = 0$$
(B2)

then

$$n_{\rm a}^* = 1 - \frac{\delta c_{\rm a}}{1 + \lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}} \tag{B3}$$

We will further use these expressions to simplify analytical calculations.

$$n_{\rm a}^* = 1 - \frac{1 + \lambda_{\rm a} c_{\rm a} + M_{\rm a} - \sqrt{(1 + \lambda_{\rm a} c_{\rm a} + M_{\rm a})^2 - 4\lambda_{\rm a} c_{\rm a}^2 \delta}}{2\lambda_{\rm a} c_{\rm a}} \tag{B4}$$

$$n_{\rm a}^* = \frac{-(1+M_{\rm a}-\lambda_{\rm a}c_{\rm a}) + \sqrt{(1+M_{\rm a}-\lambda_{\rm a}c_{\rm a})^2 + 4\lambda_{\rm a}c_{\rm a}(1+M_{\rm a}-\delta c_{\rm a})}}{2\lambda_{\rm a}c_{\rm a}}$$
(B5)

We are only interested in the case where  $n_{\rm a}^* > 0$ , which holds only when  $1 + M_{\rm a} > \delta c_{\rm a}$ .

#### **Proof:**

 $n_{\rm a}^*$  exists, and  $n_{\rm a}^* > 0$  if:

$$4\lambda_{\rm a}c_{\rm a}(1+M_{\rm a}-\delta c_{\rm a})>0\tag{B6}$$

Which is equivalent to:

$$1 + M_{\rm a} > \delta c_{\rm a} \tag{B7}$$

# 2 Invasion conditions assuming a 'complete mimicry shift'

We assume that the density of individuals carrying the derived phenotype is initially very low, *i.e.* we consider that  $n_{\rm d} = O(\epsilon)$  with epsilon being small. To determine whether having the derived phenotype is advantageous or not, we determine the sign of the derivative of the density of individuals with the derived phenotype when  $n_{\rm a} = n_{\rm a}^*$ . We note  $f(n_{\rm a}, n_{\rm d})$  the derivative of the density of individuals with the derived phenotype:  $f(n_{\rm a}, n_{\rm d}) = \frac{dn_{\rm d}}{d\tau}$ .

Hence, when  $n_{\rm a} = n_{\rm a}^*$ , we have:

$$f(n_{\rm a}^*, n_{\rm d}) = \frac{\mathrm{d}n_{\rm d}}{\mathrm{d}\tau}\Big|_{n_{\rm a}=n_{\rm a}^*} \tag{B8}$$

$$= n_{\rm d}(1 - n_{\rm a}^* - n_{\rm d}) - \frac{\delta c_{\rm d} n_{\rm d}}{1 + \lambda_{\rm a}(c_{\rm a} n_{\rm a}^* + c_{\rm d} n_{\rm d}) + M_{\rm a}}$$
(B9)

Given that  $n_{\rm d} = O(\epsilon)$ , we can approximate:

$$f(n_{\rm a}^*, n_{\rm d}) = n_{\rm d}(1 - n_{\rm a}^*) - \frac{\delta c_{\rm d} n_{\rm d}}{1 + \lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}} + O(\epsilon^2)$$
(B10)

By using equation B3, we get:

$$f(n_{\rm a}^*, n_{\rm d}) = \frac{\delta n_{\rm d}}{1 + \lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}} (c_{\rm a} - c_{\rm d}) + O(\epsilon^2)$$
(B11)

By neglecting the term of the same order as  $\epsilon^2$ , we find that having the derived phenotype is advantageous when:

$$c_{\rm d} < c_{\rm a}$$
 (B12)

# 3 Invasion conditions assuming a 'complete mimicry shift'

Assuming a 'complete mimicry shift', when  $n_{\rm a} = n_{\rm a}^*$ , we have:

$$f(n_{\rm a}^*, n_{\rm d}) = n_{\rm d}(1 - n_{\rm a}^* - n_{\rm d}) - \frac{\delta c_{\rm d} n_{\rm d}}{1 + \lambda_{\rm a} c_{\rm d} n_{\rm d} + M_{\rm d}}$$
(B13)

$$= n_{\rm d}(1 - n_{\rm a}^*) - \frac{\delta c_{\rm d} n_{\rm d}}{1 + M_{\rm d}} + O(\epsilon^2)$$
(B14)

By using equation B3, we have :

$$f(n_{\rm a}^*, n_{\rm d}) = \delta n_{\rm d} \left( \frac{c_{\rm a}}{1 + \lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}} - \frac{c_{\rm d}}{1 + M_{\rm d}} \right) + O(\epsilon^2)$$
(B15)

By neglecting the term of the same order as  $\epsilon^2$ , we find that having the derived phenotype is advantageous when:

$$\frac{c_{\rm d}}{1+M_{\rm d}} < \frac{c_{\rm a}}{1+\lambda_{\rm a}c_{\rm a}n_{\rm a}^*+M_{\rm a}} \tag{B16}$$

and therefore:

$$c_{\rm d} < c_{\rm a} \frac{(1+M_{\rm d})}{1+\lambda_{\rm a} c_{\rm a} n_{\rm a}^* + M_{\rm a}}$$
 (B17)

We call  $\hat{C}$  this threshold value:

$$\hat{C} = c_{a} \frac{(1+M_{d})}{1+\lambda_{a}c_{a}n_{a}^{*}+M_{a}}$$
(B18)

## 3.1 Sensitivity of Ĉ

The threshold value  $\hat{C}$  below which the derived phenotype can invade can be simplified as:

$$\hat{C} = \frac{2(1+M_{\rm d})c_{\rm a}}{1+\lambda_{\rm a}c_{\rm a}+M_{\rm a}+\sqrt{X}}$$
(B19)

With:

$$X = \left(1 + M_{\rm a} + \lambda_{\rm a} c_{\rm a}\right)^2 - 4 \lambda_{\rm a} \delta c_{\rm a}^{\ 2} \tag{B20}$$

We now determine the sensitivity of this threshold value to a change in parameter values, to determine under what conditions a phenotype characterized by a high conspicuousness  $c_{\rm d} > c_{\rm a}$  can invade (which occurs when the threshold value  $\hat{C}$  is high). See Supp. Tab. S1.

#### Effect of $M_d$ on the threshold value $\hat{C}$ :

$$\frac{\partial \hat{C}}{\partial M_{\rm d}} = \frac{2 c_{\rm a}}{1 + \lambda_{\rm a} c_{\rm a} + M_{\rm a} + \sqrt{X}} > 0 \tag{B21}$$

Therefore, increased  $M_d$  increases the invasibility area. Interestingly, the derived phenotype with a higher conspicuousness than the ancestral phenotype can invade the population if:

$$M_{\rm d} > M_{\rm a} + \lambda_{\rm a} c_{\rm a} n_{\rm a}^*. \tag{B22}$$

Effect of  $M_a$  on the threshold value  $\hat{C}$ :

$$\frac{\partial\sqrt{X}}{\partial M_{\rm a}} = \frac{1 + M_{\rm a} + \lambda_{\rm a} \, c_{\rm a}}{\sqrt{X}} > 0 \tag{B23}$$

$$\frac{\partial \hat{C}}{\partial M_{\rm a}} = \frac{-2(1+M_{\rm d})c_{\rm a}}{\left(1+\lambda_{\rm a}\,c_{\rm a}+M_{\rm a}+\sqrt{X}\right)^2} \left(1+\frac{\partial\sqrt{X}}{\partial M_{\rm a}}\right) < 0 \tag{B24}$$

Therefore, increased  $M_{\rm a}$  decreases the invasibility area.

#### Effect of $\lambda_{\mathbf{a}}$ on the threshold value $\hat{\mathbf{C}}$ :

$$\frac{\partial\sqrt{X}}{\partial\lambda_{a}} = \frac{c_{a}\left[1 + M_{a} + (\lambda_{a} - 2\,\delta)\,c_{a}\right]}{\sqrt{X}} \tag{B25}$$

$$\frac{\partial \hat{C}}{\partial \lambda_{\rm a}} = \frac{-2\left(1+M_{\rm d}\right)c_{\rm a}^2}{\left(1+\lambda_{\rm a}\,c_{\rm a}+M_{\rm a}+\sqrt{X}\right)^2\sqrt{X}} \left(\sqrt{X}+1+M_{\rm a}+\lambda_{\rm a}\,c_{\rm a}-2\,\delta\,c_{\rm a}\right) \tag{B26}$$

$$\frac{\partial \hat{C}}{\partial \lambda_{\mathrm{a}}} = \frac{-2\left(1+M_{\mathrm{d}}\right)c_{\mathrm{a}}^{2}}{\left(1+\lambda_{\mathrm{a}}c_{\mathrm{a}}+M_{\mathrm{a}}+\sqrt{X}\right)^{2}\sqrt{X}}\left(\sqrt{X}-\left(1+M_{\mathrm{a}}-\lambda_{\mathrm{a}}c_{\mathrm{a}}\right)+2\left(1+M_{\mathrm{a}}-\delta c_{\mathrm{a}}\right)\right) \tag{B27}$$

$$\frac{\partial \hat{C}}{\partial \lambda_{\mathrm{a}}} = \frac{-2\left(1+M_{\mathrm{d}}\right)c_{\mathrm{a}}^{2}}{\left(1+\lambda_{\mathrm{a}}c_{\mathrm{a}}+M_{\mathrm{a}}+\sqrt{X}\right)^{2}\sqrt{X}}\left(2\lambda_{\mathrm{a}}c_{\mathrm{a}}n_{\mathrm{a}}^{*}+2\left(1+M_{\mathrm{a}}-\delta c_{\mathrm{a}}\right)\right) \tag{B28}$$

Yet,  $1 + M_{\rm a} - \delta c_{\rm a} > 0$  (imposed by the condition of existence of the equilibrium), and  $\sqrt{X} - (1 + M_{\rm a} - \lambda_{\rm a} c_{\rm a}) > 0$ . Therefore:

$$\frac{\partial C}{\partial \lambda_{\rm a}} < 0 \tag{B29}$$

Therefore, increased  $\lambda_{\rm a}$  decreases the invasibility area.

Effect of  $\delta$  on the threshold value  $\hat{\mathbf{C}}$ :

$$\frac{\partial\sqrt{X}}{\partial\delta} = \frac{-2\lambda_{\rm a}c_{\rm a}^{\ 2}}{\sqrt{X}} < 0 \tag{B30}$$

$$\frac{\partial \hat{C}}{\partial \delta} = \frac{-2\left(1+M_{\rm d}\right)c_{\rm a}}{\left(1+\lambda_{\rm a}\,c_{\rm a}+M_{\rm a}+\sqrt{X}\right)^2} \frac{\partial\sqrt{X}}{\partial \delta} > 0 \tag{B31}$$

Therefore, increased  $\delta$  increases the invasibility area.

## Effect of $c_a$ on the threshold value $\hat{C}:$

$$\frac{\partial\sqrt{X}}{\partial c_{\rm a}} = \frac{\lambda_{\rm a}\left(1 + M_{\rm a} + \lambda_{\rm a}\,c_{\rm a} - 4\,\delta\,c_{\rm a}\right)}{\sqrt{X}} \tag{B32}$$

$$\frac{\partial n_{\rm a}^*}{\partial c_{\rm a}} = \frac{(1+M_{\rm a})(1+\lambda_{\rm a}c_{\rm a}+M_{\rm a}-\sqrt{X})}{2\lambda_{\rm a}c_{\rm a}^2} > 0 \tag{B33}$$

$$\frac{\partial \hat{C}}{\partial c_{\rm a}} = \frac{2\left(1+M_{\rm d}\right)}{\left(1+\lambda_{\rm a}\,c_{\rm a}+M_{\rm a}+\sqrt{X}\right)^2\sqrt{X}} \left[\left(1+M_{\rm d}\right)\sqrt{X} + (1+M_{\rm a})(1+M_{\rm a}+\lambda_{\rm a}\,c_{\rm a})\right] > 0 \tag{B34}$$

Therefore, increased  $c_{\rm a}$  increases the invasibility area.

### Calculation:

$$\hat{C} = \frac{2(1+M_{\rm d})c_{\rm a}}{1+\lambda_{\rm a}c_{\rm a}+M_{\rm a}+\sqrt{X}}$$
(B35)

With:

$$X = (1 + M_{\rm a} - \lambda_{\rm a} c_{\rm a})^2 + 4 \lambda_{\rm a} c_{\rm a} (1 + M_{\rm a} - \delta c_{\rm a})$$
(B36)

$$\frac{\partial \hat{C}}{\partial c_{a}} = \frac{1}{\left(1 + \lambda_{a} c_{a} + M_{a} + \sqrt{X}\right)^{2}} \left[2\left(1 + M_{d}\right)\left(1 + \lambda_{a} c_{a} + M_{a} + \sqrt{X}\right) - 2\left(1 + M_{d}\right)c_{a}\left(\lambda_{a} + \frac{\partial\sqrt{X}}{\partial c_{a}}\right)\right]$$
(B37)

$$\frac{\partial \hat{C}}{\partial c_{a}} = \frac{1}{\left(1 + \lambda_{a} c_{a} + M_{a} + \sqrt{X}\right)^{2}} \left[2\left(1 + M_{d}\right)\left(1 + \lambda_{a} c_{a} + M_{a} + \sqrt{X}\right) - 2\left(1 + M_{d}\right)c_{a}\left(\lambda_{a} + \frac{\lambda_{a}\left(1 + M_{a} + \lambda_{a} c_{a} - 4\delta c_{a}\right)}{\sqrt{X}}\right)\right] (B38)$$

$$\frac{\partial \hat{C}}{\partial c_{a}} = \frac{2\left(1+M_{d}\right)}{\left(1+\lambda_{a}c_{a}+M_{a}+\sqrt{X}\right)^{2}} \left[1+\lambda_{a}c_{a}+M_{a}+\sqrt{X}-c_{a}\left(\lambda_{a}+\frac{\lambda_{a}\left(1+M_{a}+\lambda_{a}c_{a}-4\delta c_{a}\right)}{\sqrt{X}}\right)\right]$$
(B39)

$$\frac{\partial \hat{C}}{\partial c_{\mathrm{a}}} = \frac{2\left(1+M_{\mathrm{d}}\right)}{\left(1+\lambda_{\mathrm{a}}\,c_{\mathrm{a}}+M_{\mathrm{a}}+\sqrt{X}\right)^{2}} \left[1+M_{\mathrm{a}}+\sqrt{X}-\frac{\lambda_{\mathrm{a}}\,c_{\mathrm{a}}\left(1+M_{\mathrm{a}}+\lambda_{\mathrm{a}}\,c_{\mathrm{a}}-4\,\delta\,c_{\mathrm{a}}\right)}{\sqrt{X}}\right] \tag{B40}$$

$$\frac{\partial \hat{C}}{\partial c_{a}} = \frac{2\left(1+M_{d}\right)}{\left(1+\lambda_{a}c_{a}+M_{a}+\sqrt{X}\right)^{2}\sqrt{X}} \left[\left(1+M_{a}\right)\sqrt{X}+X-\lambda_{a}c_{a}\left(1+M_{a}+\lambda_{a}c_{a}-4\delta c_{a}\right)\right]$$
(B41)

$$\frac{\partial \hat{C}}{\partial c_{\rm a}} = \frac{2\left(1+M_{\rm d}\right)}{\left(1+\lambda_{\rm a}\,c_{\rm a}+M_{\rm a}+\sqrt{X}\right)^2\sqrt{X}} \left[(1+M_{\rm a})\sqrt{X}+(1+M_{\rm a})^2+\lambda_{\rm a}\,c_{\rm a}\,(1+M_{\rm a})\right] \tag{B42}$$

$$\frac{\partial \hat{C}}{\partial c_{\mathrm{a}}} = \frac{2\left(1+M_{\mathrm{d}}\right)}{\left(1+\lambda_{\mathrm{a}}\,c_{\mathrm{a}}+M_{\mathrm{a}}+\sqrt{X}\right)^{2}\sqrt{X}} \left[\left(1+M_{\mathrm{a}}\right)\sqrt{X}+\left(1+M_{\mathrm{a}}\right)\left(1+M_{\mathrm{a}}+\lambda_{\mathrm{a}}\,c_{\mathrm{a}}\right)\right] \tag{B43}$$

## SUPPORTING INFORMATION FIGURES



Perceived dissimilarity between ancestral and derived phenotypes due to low predator generalization

# Supporting information, Figure 1. Numerical simulations showing that the derived phenotype gets fixed once it has invaded the population.

We consider the same parameter values as in Figure 2. This time, we show the frequency of mutant colour patterns after a runtime equal to  $10^{20}$  using a numerical resolution method (Runge Kutta 4). Note the absence of intermediate shades of grey in all panels. When the derived phenotype invades (as shown in Figure 2 in black), the derived phenotype ultimately replaces the ancestral phenotype, and reaches a frequency equal to 1 (dark purple here). By contrast, when the derived phenotype does not invade (as shown in Figure 2 in grey), the derived

phenotype ultimately reaches a frequency equal to 0 (light purple here). This additional analysis supports our subsection "Numerical analyses" of the article.



Supporting information, Figure 2. Effect of the model parameters on the threshold value below which conspicuousness is favoured (illustrating the results shown in Tab. 2). The derived phenotype can invade when its conspicuousness  $c_d$  is lower than the conspicuousness threshold  $\hat{C} = c_a \frac{1+M_d}{1+\lambda_a c_a n_a^*+M_a}$  (see Eq. [10]). Here, we represent in each subfigure how variations in a given parameter affect this threshold value. If the increase of a parameter value (e.g., increased  $c_a$ ) increases  $\hat{C}$ , this means that the range of conspicuousness values enabling the invasion of the derived phenotype increases. In the case of variations in  $\delta$ , the inset shows

the change in  $\hat{C}$  which is not visible by using the common y-scale ranging from 0 to 1. See default values in Table 1 (remembler that  $\delta = \frac{p}{r}$  is the rescaled baseline mortality rate; and  $\lambda_a = uK\beta_a$  is the rescaled deterrence factor). Note that  $\lambda_d = uK\beta_d$  has no effect on  $\hat{C}$  (not shown here).

This additional analysis supports our results described in subsection "2 – A mimicry shift can promote an increase in conspicuousness (assuming a 'complete mimicry shift'; S = 0)" of the article.



Derived phenotype invades Derived phenotype does not invade

Supporting information, Figure 3. Evolution of conspicuousness depending on differences other than in conspicuousness. We consider different values of parameter l, which controls the phenotypic distance unrelated to conspicuousness. The condition  $l \rightarrow \infty$  is obtained by setting S = 0, just like for  $\gamma \rightarrow \infty$  in Figure 2. When ancestral and derived phenotypes are different  $(l \rightarrow \infty)$ , it is difficult for more conspicuous mutants to invade the population because the ancestral phenotype benefits from a greater number-dependent protection. By contrast, if

the ancestral and derived phenotypes are totally similar (l = 0) or imperfect mimics (intermediate l), more conspicuous mutants are favoured because they benefit from an increased number-dependent protection by resembling both the wild-type and mutant mimetic community. See Figure 2 for more details. Here,  $\gamma = 1$ ,  $\beta_a = 1$ , and h = 0. See other default values in Table 1.

This additional analysis supports our results described in subsection "3 – Imperfect mimicry inhibits the invasion of derived phenotypes, and is more favourable to an increase in conspicuousness" of the article.



Supporting information, Figure 4. Conditions under which higher conspicuousness is favoured. In each of the five subfigures, we vary the values of parameters  $\gamma$  and  $M_d$  in the four panels. Additionally, the x-axis represents the variation of different parameters in each subfigure: r (a),  $\beta_a$  (b),  $\beta_d$  (c), u (d) and l (e). On the y-axis, we show the conspicuousness

values,  $c_a$ , where higher conspicuousness is favoured. In subfigure a, we represent for a combination of parameters what these minimum and maximum values of  $c_a$  correspond to in the graphs presented in the main manuscript. See default values in Table 1.

This additional analysis supports our results described in subsection "4 – Easily-memorable phenotypes that facilitate predator learning inhibit the evolution of increased conspicuousness" of the article.

#### a Baseline predation rate

**b** Increase of the predation rate on cryptic prey



Supporting information, Figure 5. Conditions under which higher conspicuousness is favoured for various predation rates. Same as Supporting information Appendix 3, Fig. 3 but here in each subfigure, the x-axis represents variations of p (a) and h (b). See Supporting information Appendix 3, Fig. 3 for more details. See default values in Table 1.

This additional analysis supports our results described in subsection "4 – Easily-memorable phenotypes that facilitate predator learning inhibit the evolution of increased conspicuousness" of the article.

Baseline predation rate



Supporting information, Figure 6. Evolution of conspicuousness depending on the baseline predation rate. We consider different values of parameter p, which controls the baseline predation rate. The baseline predation rate has very little effect on the conditions of invasion of the derived phenotype. See Figure 2 for more details. Here,  $\gamma \rightarrow \infty$ ,  $\beta_a = 1$ , and h = 0. See other default values in Table 1.

This additional analysis supports our results described in subsection "5 – Predation pressure and availability of alternative prey modulate selection on conspicuousness" of the article.

# SUPPORTING INFORMATION TABLE

Supporting information, Table 1. Sensitivity of the invasion of the derived phenotypes to the different parameters, assuming a 'complete mimicry shift'. The derived phenotype can invade when its conspicuousness  $c_d$  is lower than the conspicuousness threshold  $\hat{C} = c_a \frac{1+M_d}{1+\lambda_a c_a n_a^*+M_a}$  (see Eq. [10] and Appendix B). See also Supp. Fig. 2.

Parameter	Sensitivity of the invasion condition	Meaning
$M_a$ , the protection brought by the <i>ancestral</i> mimicry ring	$\frac{\partial \hat{\mathcal{C}}}{\partial M_a} < 0$	Higher protection provided by the mimicry ring matching the ancestral phenotype $(M_a)$ <b>decreases</b> the range of conspicuousness values enabling the invasion of the derived phenotype.
$M_d$ , the protection brought by the <i>derived</i> mimicry ring	$\frac{\partial \hat{C}}{\partial M_d} > 0$	Higher protection provided by the mimicry ring matching the derived phenotype $(M_d)$ <b>increases</b> the range of conspicuousness values enabling the invasion of the derived phenotype.
$\lambda_a$ , the rescaled deterrence factor associated with the ancestral phenotype $(\lambda_a = uK\beta_a)$	$\frac{\partial \hat{\mathcal{C}}}{\partial \lambda_a} < 0$	Higher deterrence factor $\lambda$ <b>decreases</b> the range of conspicuousness values enabling the invasion of the derived phenotype.
$\delta$ , rescaled baseline mortality rate	$\frac{\partial \hat{\mathcal{L}}}{\partial \delta} > 0$	Increased baseline predation pressure ( $\delta$ ) <b>increases</b> the range of conspicuousness

$(\delta = p/r)$		values enabling the invasion of the derived phenotype.
<i>c<sub>a</sub></i> , conspicuousness of the ancestral phenotype	$\frac{\partial \hat{C}}{\partial c_a} > 0$	Higher conspicuousness of the ancestral phenotype $(c_a)$ increases the range of conspicuousness values enabling the invasion of the derived phenotype.