The dual benefits of aposematism: predator avoidance and enhanced resource collection

Michael P. Speed,1,2 Michael A. Brockhurst,1 and Graeme D. Ruxton3

1School of Biological Sciences, Biosciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom
2E-mail: speedm@liv.ac.uk
3Faculty of Biomedical and Life Sciences, Graham Kerr Building, University of Glasgow G12 8QQ, United Kingdom

Received February 3, 2009
Accepted October 16, 2009

Theories of aposematism often focus on the idea that warning displays evolve because they work as effective signals to predators. Here, we argue that aposematism may instead evolve because, by enhancing protection, it enables animals to become more exposed and thereby gain resource-gathering benefits, for example, through a wider foraging niche. Frequency-dependent barriers (caused by enhanced conspicuousness relative to other prey and low levels of predator education) are generally assumed to make the evolution of aposematism particularly challenging. Using a deterministic, evolutionary model we show that aposematic display could evolve relatively easily if it enabled prey to move more freely around their environments, or become exposed in some other manner that provides fitness benefits unrelated to predation risk. Furthermore, the model shows that the traits of aposematic conspicuousness and behavior which lead to raised exposure positively affect each other, so that the optimal level of both tends to increase when the traits exist together, compared to when they exist in isolation. We discuss the ecological and evolutionary consequences of aposematism. One conclusion is that aposematism could be a key evolutionary innovation, because by widening habitat use it may promote adaptive radiation as a byproduct of enhanced ecological opportunity.

Key words: Aposematism, key innovation, life history, niche, predator.

Many potential prey show some level of aposematism, a combination of a brightly colored warning display plus one or more repellent secondary defense (toxins, stings, spines etc.). Wasps and bees are obvious examples, combining a black and yellow warning display with a venomous sting. Why do prey adopt this strategy? The commonly given explanation for aposematism is that conspicuous appearance helps to deter predation. Secondary defenses can repel predators during an attack (Whitman et al. 1990) and can educate predators to avoid similar prey, potentially benefiting family members (Fisher 1930). Bright warning displays can enhance protection from predation in a number of ways: reducing recognition errors, reducing attacks by provoking predator wariness (as neophobia or dietary conservatism), and then causing predators to handle bright prey cautiously. Warning displays may also enhance the rate of learning and retard forgetting of learned aversions (review in Ruxton et al. 2004). Hence, for many prey exposed to predation, aposematism is the optimal appearance state compared to crypsis. Indeed avoidance of predation is generally held to be the major reason that aposematism exists.

In this article, we wish to argue that aposematism may evolve in previously cryptic animals for a related but distinct second reason: that it allows prey to expand their niche, enabling them to collect resources in full view of predators without substantially increasing their predation risk. In effect, aposematism generates “predator-free foraging space” within which prey can make greater use of the opportunities in their environments. The idea that aposematism enables prey to forage in full view of predators...
An alternative way of articulating the idea is to consider that there are likely to be opportunity costs to adopting a cryptic antipredatory defense; effective crypsis may limit behaviors (such as movement) and the times of day and microhabitats that an individual can exploit for resource gathering (Stamp and Wilkens 1993; Merilaita and Tullberg 2005; Speed and Ruxton 2005a). Thus, there is a potential trade-off; more resources can benefit growth, reproduction, and avoidance of starvation, but the price may be enhanced conspicuousness to predators through being active at times and places that are not compatible with effective crypsis. Although we focus on the idea of movement, it is also likely that a similar trade-off applies to animals that draw attention to themselves if they forage in a particularly exposed site or become too large to be able to hide effectively in their habitat. Whether by movement, growth, or simply positioning themselves in view of predators, animals may therefore make themselves apparent to predators and for simplicity we term this raised exposure, and contrast it with the raised conspicuousness which comes from coloration such as aposematic display. That is, we consider that the rate at which a prey individual encounters predators is a function both of its visual appearance and of its behavior. Although the potential trade-off between raised exposure and predation costs is well known (Lima and Dill 1990; Sih 1992; Martin et al. 2005; Polo et al. 2005; Whitman and Vincent 2008), it has rarely been linked to the evolution and optimization of aposematic traits. In this article, we want to study how animals might best exploit this trade-off, especially when—as will often be the case—levels of raised exposure can coevolve with pigments on the integument of the animal that might function as warning displays.

At the time of writing, only two published theoretical models have explicitly linked raised exposure with aposematism (although there are pertinent discussions in the literature—Engen et al. 1986; Leimar et al. 1986; Guilford 1988). Merilaita and Tullberg (2005) used stochastic, individual-based simulations to show that when the level of visual heterogeneity of a habitat is high, prey should evolve secondary defense and some sort of aposematic display, rather than crypsis. They used an ingenious phylogenetic analysis to demonstrate that raised exposure (scored as diurnal activity in lepidopterans) is associated with the evolution of aposematism. Speed and Ruxton (2005a) also examined the possibility that optimal levels of costly secondary defenses increase as the gains from foraging in exposed areas increases. What is lacking from these models, in our view, is a detailed consideration of how the pertinent traits (raised exposure and aposematic display) may cooptimize in relation to each other and with respect to variation in key ecological parameters.

Furthermore, recently published phylogenetically based work is providing important insights into the evolutionary history and ecological function of aposematic signaling (Santos et al. 2003; Summers 2003; Vamosi 2005; Przezcek et al. 2008). One suggestion is that aposematism may be a key evolutionary innovation, for example, reducing extinction rates or increasing speciation rates within lineages (Vamosi 2005). If aposematism can confer benefits beyond reducing costs associated with predatory attacks, for example, because it widens diet breadth and thereby provides access to greater ecological opportunity or a new “adaptive zone” (Simpson 1944), then it is possible to see that these benefits could play a significant role in diversification and that aposematism can act as a key evolutionary innovation.

We therefore generated a numerical, deterministic evolutionary model of aposematism (related to that described in Blount et al. 2009) in which chemically defended prey can evolve and simultaneously cooptimize two traits; aposematic signaling and raised exposure. We used the model to evaluate: (1) whether aposematism evolves more easily for prey if there is sufficient incentive for raised exposure within a habitat; (2) how the optimal values of display and behavior are affected by the presence of the other trait; and (3) how variations in ecological conditions affect optimal values. We discuss the results in the context of the ecological function of signal conspicuousness.

**An Optimization Model**

We consider a population of prey individuals that have a fixed level of antipredator toxicity ($D$) but that can vary in their levels of aposematic display ($A$) and the extent to which they make themselves obvious to predators by choosing “actions” such as heightened growth rates or movement around the habitat that are incompatible with crypsis (so-called “raised exposure,” $B$). Levels of display and raised exposure are both assumed to be heritable and we consider how traits coevolve to their optimal level using a deterministic evolutionary model. We use the system to consider whether aposematic prey can evolve from a point of maximum crypsis, and we evaluate the optimal values of its signaling trait and its raised exposure over a range of ecological conditions. We compare the likelihood of the evolution of aposematism with relatively nonexposed prey (for which optimal value of $B$ is zero), and for those for which, for example, some level of raised exposure is common (i.e., optimal $B > 0$). In addition, we consider the role that aposematic display may play in increasing the level...
of exposure further. The model is coded using Matlab (terms are summarized in Table 1 and sample code is presented in the Supporting Information). This model is based on that presented in (Blount et al. 2009), and can replicate the results of a related model (Speed and Ruxton 2007), but has the marked advantage of including frequency-dependent selection.

**Model Outline**

We are not concerned with the evolutionary effects of small population size; hence, we consider change in an infinitely large prey population distributed over the vertices of a matrix in A–B parameter space. Specifically, we have a square matrix with \( N \) nodes per side and \( P_{ij} \) gives the population density of individuals that play strategy \((i, j)\), meaning that they have \( A_i \) and \( B_j \) values given by

\[
A_i = A_{\text{min}} + \frac{i (A_{\text{max}} - A_{\text{min}})}{N}, \quad (1)
\]

\[
B_j = B_{\text{min}} + \frac{j (B_{\text{max}} - B_{\text{min}})}{N}, \quad (2)
\]

where \( A \) is constrained to lie in the range \([A_{\text{min}} = 0, A_{\text{max}} = 100]\) and \( B \) in the range \([B_{\text{min}} = 0, B_{\text{max}} = 100]\), and \( i \) and \( j \) are integers between 0 and \( N \) (which is set to 100 throughout). Essentially, we constrain \( A \) and \( B \) to take \( N \) different evenly distributed values, rather than being free to take continuously distributed values. However, because \( N \) can take a very large value, this is not a biologically important constraint. We utilize different starting distributions according to our aims; for example, we might start the first generation with the population equally distributed across the \( N^2 \) nodes. At the start of every subsequent generation, we normalize the population so that the total population size across all \( N^2 \) nodes sum to one.

Prey fitness is determined by the probability of prey survival over a generation time, \( T \), and the reproductive resource that it has accumulated in that time. We first consider how survival is determined. If encounters with predators occur at a rate \( r_{ij} \), the probability of surviving a given interval is given by \( S_{ij} \) where

\[
S_{ij} = \exp(-r_{ij} s_{ij} T). \quad (3)
\]

We now define \( s_{ij} \) and \( r_{ij} \).

**Probability of Prey Death Given Encounter with the Predator (\( s_{ij} \))**

We make a number of assumptions in determining the probability of death given an encounter, all of which have empirical support. First, we assume that the probability of death during an encounter, \( s_{ij} \), decreases with increases in the individual prey’s level of antipredatory toxicity, \( D \), because we assume that predators can to some extent, detect toxins and reject prey without necessarily killing them (Wiklund and Järvi 1982; Skelhorn and Rowe 2005). In addition, we assume that the level of a prey’s warning signal can also act as a deterrent. Predators may be less likely to attack defended prey that have bright warning displays, because such prey are less easily confused with edible prey (Wallace 1867; Speed 2000; Gamberale-Stille and Guilford 2004; Skelhorn and Rowe 2006). If they do attack, predators may be less likely to injure prey if they are both chemically defended and use bright warning signals, rather than if they are cryptically colored and defended (Wiklund and Järvi 1982; Sillén-Tullberg 1985a,b; Wiklund and Sillén-Tullberg 1985). However, we assume that prey will not gain protection from predation merely by having a bright signal and no secondary defense. Hence, we also require that prey survival depends on the interaction of aposematic display with secondary defense (in effect probability of death also decreases with increases in the product of \( A \) and \( D \)), so that the probability of survival is low for any prey that completely lacks a secondary defense.

Predators are also likely to be ignorant of the defenses of prey of a rare appearance type, and thus more likely to attack and kill them, than they are more abundant prey (Mallet and Singer 1987; Servedio 2000; Puurtinen and Kaitala 2006). We therefore included a sigmoidal function, \( A_j \) (values between zero and one, see Fig. 1), to represent the role of frequency dependence caused by predator learning. \( A_j \) takes relatively low values when a prey signal is rare and higher values when a prey signal is common. We
employed the standard logistic exponential equation to calculate, $A_f$ (details and provided in legend to Fig. 1. The relative shape of the curve is defined by the parameter $H$).

The probability of death given encounter is then given by

$$s_{ij} = 0.05 + 0.94(\exp(-(D + A_{ji}D)A_f),$$

where $s_{ij}$ is bounded between 0.05 and 0.99.

The interaction of prey defense and relative frequency in equation (4) needs some explanation. If a prey form is very common with a frequency near 1, and if the value of $(D + A_{ji}D)$ is greater than 5, then the probability of death given an encounter is close to the minimum (at 0.051, with $H = 1$, see Fig. 1). Because parameter space allows many values of the term $(D + A_{ji}D)$ to be greater than 5, it follows that many prey forms (for $D > 0$) would have very similar survival values if they were sufficiently common. In essence, we assume that knowledgeable predators are very reluctant to sample any prey that they know are defended, because there are abundant alternative prey in the environment. However, when the predator is ignorant because a prey morph is relatively rare (and $A_f$ is low) the probability that a prey is killed during an encounter varies more widely depending on whether it has aposematic signaling and strong secondary defenses (i.e., on the term $(D + A_{ji}D)$ in eq. 4). In these cases, the relatively ignorant predator "goes slow" with aposematic prey, being less likely to kill them if they are brightly colored and turn out to be defended. We apply the same frequency dependence (via $A_f$) curve regardless of prey coloration, as we have already biased survival toward brightly colored and defended prey in the rest of the exponent in equation 4. To simulate alternative forms of frequency dependence we varied the shapes of the curves for $A_f$ (see Fig. 1).

We can of course have alternative relationships between $s_{ij}$ and $D$ and $A_{ji}$; to save space these are discussed in Supporting Information Appendix S1.
**Encounter Rate**

We next need to calculate \( r_{ij} \), the rate at which an individual with properties \( A_i \) and \( B_j \) is encountered by predators. We first calculate the encounter-rate \( (c_{ij}) \) in the absence of frequency-dependent effects (see methods in Blount et al. 2009). For simplicity, we assume that encounter rate is proportional to the sum of raised exposure \( (B_j) \) and aposematic conspicuousness \( (A_i) \). In the graphs shown here, we use a negative exponential term

\[
c_{ij} = 1.5 - \exp(-0.01(B_j + A_i)). \tag{5}
\]

An alternative, sigmoidal function, which provides very similar predictions, is described in Appendix S1.

We now allow that encounter rate can be affected by a different sort of frequency dependence in which prey that are, for example, relatively brighter than the rest of the population are encountered by predators more frequently. To achieve this in the model, \( c_{ij} \) is multiplied by a modifier for frequency-dependent effects \( (f_{ij}) \).

We calculate the average value of \( c_{ij} \) across the population \((C)\)

\[
C = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} P_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} P_{ij}}, \tag{6}
\]

then the frequency dependence modifier is simply given by

\[
f_{ij} = \frac{c_{ij}}{C}, \tag{7}
\]

capturing the idea that an individual’s level of conspicuousness attracts more attention from predators if the brightness of its coloration is high relative to the population as a whole. We might, for example, consider a locally abundant species, such as a ladybird. A brightly colored individual’s risk of detection by a local predator will be relatively low if all other individuals are very brightly colored, whereas its probability of detection will be relatively higher if all other individuals are highly cryptic in their coloration.

We calculate the final rate of encounter between predators and an individual playing strategy \((i, j)\)

\[
r_{ij} = c_{ij} f_{ij}. \tag{8}
\]

Note: qualitatively equivalent results would pertain if this component of frequency dependence did not apply, so that

\[
r_{ij} = c_{ij}. \tag{8a}
\]

We now have two frequency-dependent barriers to the evolution of aposematism in our model: relative encounter rate, making aposematic prey encounter predators more often than more cryptically colored prey, and relative frequency, making rare prey forms unfamiliar to predators and more prone to lethal attack than others.

**Reproduction**

We assume that animals which have maximum cryptis \((A = 0, B = 0)\) accrue resource to the value \( \alpha \) and that animals that reduce cryptis, by, for example, raised mobility, can gain a maximum net additional benefit of \( G \) over the prereproductive season, which is converted into reproductive effort, \( R_{ij} \). The net gain per unit time that accrues from behavioral activity defined by \( B_{ij} \) is simply \( G(1 - \exp(-(B_{ij} \cdot \gamma))) \), where \( G \) sets the upper boundary to the gain in that time unit and \( \gamma \) determines the rate of gain per unit time for values of \( B \) and \( G \). We set the minimum value of \( R_{ij} \) as \( \alpha \), and define the term as

\[
R_{ij} = \alpha + TG(1 - \exp(-(B_{ij} \cdot \gamma))). \tag{9}
\]

If \( P_{ij} \) represent the frequencies of traits at the start of the generation, then after the generation abundance changes to

\[
P'_{ij} = \frac{P_{ij} S_{ij} R_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} P_{ij} S_{ij} R_{ij}}, \tag{10}
\]

thereby simulating asexual reproduction across generations.

To ensure that evolved frequencies are stable against invasion by rare forms, we allow a small level of mutation between generations. Mutation involves taking a small proportion, \( m \), of the population from the total surviving prey with each defense level at the end of a time interval, \( t \), and redistributing these prey evenly across all \( N + 1 \) trait values.

This mechanism ensures that at the start of each generation all phenotypic forms have at least some small abundance level. Results are robust against values between \( m = 1 \times 10^{-9} \) and \( m = 1 \times 10^{-3} \) (and the results shown used mutation rate of \( 1 \times 10^{-7} \)). The process is iterated until stability is approximated when the change in frequencies of aposematic forms is negligible across generations. To determine this we find the summed absolute changes in the frequencies of all trait types, \( i, j \), from one generation to the next. Stability is approximated when the following criteria have been met: (1) the summed change in frequencies is less than \( 2 \times 10^{-7} \) per generation, (2) more than 99.99% of the prey locate at one pair of \( A-B \) values, and (3) there have been at least \( 5 \times 10^6 \) generations (so that condition ii can not be met simply by the starting conditions). We now develop the model, considering how prey evolution is affected by variation in assumptions about the responses of predators to aposematism.


**Results**

**QUESTION 1: DOES RAISED EXPOSURE FACILITATE THE INITIAL EVOLUTION OF APOSEMATISM?**

We use the model outlined above to simulate four alternative conditions, outlined in Table 2. In Condition 1, aposematic prey are very rare and there is no incentive for prey to evolve raised exposure ($G = 0$, so optimal $B = 0$). Condition 2 is the same, except that all permitted values of aposematic display are initially equally common. With equality in initial distribution, aposematism easily evolves in these conditions, and our aim here is to determine what the optimal value of display would be in the absence of raised exposure. Condition 3 again has very rare aposematic prey but now $G > 0$, so that there may be sufficient incentive for prey to take decisions which are incompatible with crypsis and the optimal value of $B$ can be greater than zero. Finally, Condition 4 is as Condition 3, except that $A$ is set to zero in all cases. Comparing results in Conditions 3 and 4 enables us to evaluate the difference that aposematic signaling makes to the values of optimized raised exposure. We evaluated the optimal levels for $A$ and $B$ across all four conditions using four different frequency dependence curves (Fig. 1) and set the other parameters to $T = 10$, $D = 10$, and $\gamma = 0.1$, $\alpha = 1$.

Figure 2 illustrates how prey converge (after only 100 generations) toward their optimal phenotype levels (for the case in which $H = 1$, with quite strong frequency dependence). As evolution progresses, prey congregate around their optimal phenotypic values (shown by the red contours), and at stability (virtually) all prey are located at the same, optimum value. Table 2 records optima for $A$ and $B$ with different frequency-dependent curves (described by $H$).

Looking first at the case of aposematism evolving without the benefits of raised exposure (Conditions 1 and 2), aposematism did not evolve when prey were initially located at $A = 0$, $B = 0$, but it did evolve when the frequency-dependent barrier is overcome by allowing all prey forms to initially be equally common. As we descend the table, the value of $H$ increases, indicating that relative rarity is increasingly punished (with a low value of $A_f$ and hence a raised probability of death given detection, eq. 4, and see Fig. 1). The result is that the optimal value of aposematic display increases with the severity of frequency dependence. As the conditions become increasingly harsh on the rare aposematic prey, the minimum value that ensures superiority for an aposematic signal over the common cryptic form increases (c.f. Engen et al. 1986).

Looking now at Condition 3 (all prey initially located at $A = 0$, $B = 0$, and $G = 1$), optimal aposematism ($A$) again tends to rise with $H$ (up to $H = 1$) and it takes higher levels than the equivalent values in which $G = 0$ (Condition 2). $B$ now evolves nonzero values and increases with the optimal level of $A$ (again up to $H = 1$) and it also takes higher values in the presence of aposematic display (Condition 3) than in its absence (Condition 4). Hence, aposematic display and raised exposure appear to reinforce each other, causing both to take higher values than they would if they evolved alone.

It is important to appreciate that in the model, encounter rate is assumed to increase in a decelerated, curvilinear manner in relation to display and behavior. If predator–prey encounter rate is raised because prey are increasingly detected, then the marginal increase in encounter rate caused by additional aposematic signaling is diminished. In essence aposematic displays are less costly to prey that are already easily observed by predators by virtue of their mobility, hence they evolved more easily. In addition, when aposematism is optimal (i.e., $A > 0$), this reduction in marginal costs means that the optimised signalling trait value tends to be higher if $B > 0$. For example, in Condition 2, if we set $B_{\text{min}}$ to 100 (increased from zero), then the optimal value of $A$ increased from 64 to 75.

**Table 2.** Optimal values for display ($A$) and raised exposure ($B$) for different values of $H$.

<table>
<thead>
<tr>
<th>$H$</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
<th>Condition 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>$A$=0, B=0</td>
<td>$A$=0, B=0</td>
<td>$A$ and B</td>
<td>$B$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>10</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>$A$=0, B=0</td>
<td>$A$=0, B=0</td>
<td>$A$ and B</td>
<td>$B$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>24</td>
<td>27</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>$A$=0, B=0</td>
<td>$A$=0, B=0</td>
<td>$A$ and B</td>
<td>$B$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>64</td>
<td>71</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>$A$=0, B=0</td>
<td>$A$=0, B=0</td>
<td>$A$ and B</td>
<td>$B$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>94</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. Contour plot to illustrate how the frequencies converge toward optimal values. Contours map frequency of prey: blue, low frequency; red, high frequency. $T = 10$, $D = 10$, $\gamma = 0.1$. Results shows after 100 generations; blue areas have lowest frequency, red highest.

By reducing predation, effective signaling can also enable prey to risk a higher rate of encounter with predators (higher $B$), and hence enable them to accrue more reproductive resources. This can be demonstrated in Condition 4, where we prevented aposematic displays evolving, here the optimal value for raised exposure reduced to 33 (from $B = 41$ in Condition 3).

The exception to these general predictions is seen when frequency dependence causes highly punitive costs of signaling in rare prey, that is, $H = 1.2$. Here the optimal state is raised exposure ($B > 0$) and no aposematic signaling at all. With very strong frequency dependence, only very bright signaling tends to evolve in the absence of raised exposure (Condition 2, $A = 93$, see Table 2). Combined with some nonzero level of raised exposure (Condition 3) very bright signaling incurs too high a mortality cost to make it viable.

Thus, the model generally predicts that increased exposure (e.g., through high activity levels) should aid in the evolution and maintenance of aposematic appearance traits.

**QUESTION 2: WHEN TO BE CRYPTOIC AND WHEN TO BE APOSEMATIC?**

In this next scenario we consider the sensitivity of the system to the minimum level at which resources are gained from the environment ($\alpha$, eq. 9). We varied $\alpha$ between 1 and 10, and set $G$ at either 1 or 10 (Table 3). When $\alpha$ is comparatively low ($\alpha = 1$ and $G = 1$ or 10) then both raised exposure and aposematic display can evolve in the conditions we set here ($D = 10$, $T = 10$, $\gamma = 0.1$). However, if the minimum resource value from the environment is sufficiently high compared to that gained in addition from raised exposure (e.g., $\alpha = 10$ and $G = 1$) then we see that it can again be optimal for the prey to have raised values of exposure, but here there is no aposematic signaling. As the minimum reward for cryptic prey increases, so the relative importance of resource gathering through raised exposure declines in importance. Consequently, the costs of aposematic display as a means of facilitating resource acquisition from the environment diminish to the point where bright coloration is no longer profitable. Raising $G$ now to 10 (so that $\alpha = G$), we again return to the state in which nonzero values of raised exposure and signaling are again optimal. These simulations indicate that the relative scaling of the parameters $\alpha$ and $G$ can have a profound effect on the ecological outcome for the prey. If the environment provides a relatively high return for nonexposed prey then the additional costs of signaling are excessive and cryptism is consequently selectively favored.

In the General Discussion, we consider the implications of this finding.
**Table 3. Optimal values for display (A) and raised exposure (B) for different values of G and α.**

<table>
<thead>
<tr>
<th></th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
<th>Condition 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of environmental gain (G)</td>
<td>G=0, no raised exposure</td>
<td>G=0, no raised exposure</td>
<td>G=1, raised exposure may evolve</td>
<td>G=1, raised exposure may evolve</td>
</tr>
<tr>
<td>Traits that can evolve</td>
<td>A</td>
<td>A</td>
<td>A and B</td>
<td>B</td>
</tr>
<tr>
<td>Initial prey location</td>
<td>Initial prey location evenly distributed across A×B matrix</td>
<td>Initial prey location evenly distributed across A×B matrix</td>
<td>Initial prey location evenly distributed across A×B matrix</td>
<td>Initial prey location evenly distributed across A×B matrix</td>
</tr>
<tr>
<td>G=1, α=1</td>
<td>A</td>
<td>0</td>
<td>64</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>G=10, α=1</td>
<td>A</td>
<td>0</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td>42</td>
<td>34</td>
</tr>
<tr>
<td>G=1, α=10</td>
<td>A</td>
<td>0</td>
<td>64</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>G=10, α=10</td>
<td>A</td>
<td>0</td>
<td>64</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td>41</td>
<td>33</td>
</tr>
</tbody>
</table>

**QUESTION 3: WHAT IS THE RESPONSE OF OPTIMA TO CHANGES IN ECOLOGICAL CONDITIONS?**

We next considered how the behavioral and aposematic phenotypes may change with changes in levels of secondary defenses (D) and rates of resource acquisition (γ). In the first instance, D is increased from zero to 15; in the second γ is increased from 0.05 to 1.

When all independent variables except D were kept constant, then, as before, in Condition 1 aposematism never evolved (i.e., optimal A = 0). Figure 3 shows the results for Condition 2 in Panel (1), (only A is shown as optimal B here is always zero), Condition 3 (A is co-optimized with respect to B) in Panels (2), (3), and Condition 4 in Panel (4), (B is optimized in the absence of any signaling, A is set to 0).

In the absence of aposematic display, we found a decelerated rise in the optimal levels of raised exposure (B) as D increases (Condition 4, Fig. 3A), so that, for example, prey with better defenses (but no aposematic signals) are more mobile and encounter predators more often. However, when aposematic display evolves with or without raised exposure, we observe that its optimized values tend to decline in value as D increases from nonzero values. Here “nastier” prey are predicted to be less colorful and less mobile than those with weaker secondary defenses.

A decline in the level of aposematic display with increases in secondary defenses has been predicted and discussed several times before (Leimar et al. 1986; Speed and Ruxton 2005b, 2007). This results because the two defensive traits (A and D) contribute to a single component of fitness (survival) but incur unequal costs. In our model, aposematic display is more costly than toxicity because it incurs costs of raised encounter rate. Thus, if there is an increase in the cheapest component (here toxicity), the animal often maximizes its fitness by reducing the value of the more expensive one. Thus, the prey encounters predators less often, and because toxicity (D) is higher it is still well defended when it does so. Empirical evidence to support this prediction is, for example, described in Darst et al. (2006). The apparent a counter example (Summers and Clough 2001) has recently been explained by Blount et al. (2009) who proposed a mechanism which would encompass both kinds of result and we refer interested readers to this source. In our model, prey that signal less intensely tend also to reduce their exposure to predation (B), and consequently optimal values of A and B can both fall as D increases. We note although, that the optimal level of raised exposure is still higher with aposematic signaling than without it.

Finally, we set D = 10 and changed γ, the rate at which resources are collected per unit time. As can be seen in Figure 3B, if the rate of collection of resource increases (here by changing γ between 0.05 and 1), then the optimal levels of raised exposure and aposematism both decline. As the environment becomes increasingly rewarding, prey have less need to risk encounters with predators, hence both traits decline in value. An analogous result is obtained by keeping γ constant, but increasing the season length, T. Again, aposematic display and raised exposure declined in value over the parameters used (result shown in Appendix S1).

**General Discussion**

In this article, we suggest that one important aspect of antipredator defenses, such as aposematism, is to enable animals to make better use of the opportunities in their environments. We applied an evolutionary optimization method to examine the evolutionary relationships between raised exposure and aposematic display. The results make a number of general points about the evolution and ecological role of aposematism and we now deal with these in turn.

**THE EVOLUTION OF APOSEMATISM MAY BE EASY TO EXPLAIN**

The large majority of papers that consider the initial evolution of aposematism assume that the trait evolved initially in cryptically
Figure 3. Optimization of traits in relation to changed ecological conditions. (A) $\gamma = 0.1$, $T = 10$, $\alpha = 1$. Aposematism did not evolve in Condition 1. Panel (1)—Condition 2, optimized display ($A$) as $D$ changes in the absence of raised exposure ($G = 0$, $B = 0$; optimized $B$ is zero and not shown for Condition 2); Panel (2)—Condition 3, optimized display ($A$) as $D$ changes in the presence of raised exposure ($G > 0$, $B > 0$); Panel (3)—Condition 3, optimized behavior ($B$) as $D$ changes in the presence of aposematic display; Panel (4)—Condition 4, optimized behavior ($B$) as $D$ changes in the absence of aposematic display ($A = 0$; there is no panel for optimized $A$ in condition 4).

(B) Results changes in rate of collection of resources, $\gamma (T = 10)$. Aposematism did not evolve in Condition 1. Panel (1)—Condition 2, optimized display ($A$) as $D$ changes in the absence of raised exposure ($G = 0$, $B = 0$; optimized $B$ is zero and not shown for Condition 2); Panel (2)—Condition 3, optimized display ($A$) as $D$ changes in the presence of raised exposure ($G > 0$, $B > 0$); Panel (3)—Condition 3, optimized behavior ($B$) as $D$ changes in the presence of aposematic display; Panel (4)—Condition 4, optimized behavior ($B$) as $D$ changes in the absence of aposematic display ($A = 0$; there is no panel for optimized $A$ in condition 4).
colored species (see review in Ruxton et al. 2004). Frequency-dependent barriers (based on relative crypsis and predator ignorance) therefore make the explanation of aposematism difficult. However, as has been argued elsewhere (Ruxton et al. 2004; Merilaita and Tullberg 2005; Speed and Ruxton 2005a) crypsis may impose considerable opportunity costs on prey, because existing in varied microhabitats is incompatible with hiding from predators; one way to avoid these costs is through aposematism. Indeed there is a general consensus that obligate crypsis can be costly, in the sense that it constrains the space over which a prey can move and still be hidden from predators. This idea is found relatively frequently within the aposematism and the plant–herbivore literature (Guilford 1988; Stamp and Wilkens 1993; Chambers et al. 1996; Bernays et al. 1997; Sword et al. 2000; Bernays and Singer 2002; Despland and Simpson 2005a,b).

The simulations described in this article are the first that demonstrate that aposematism can easily evolve from rarity in the face of strong frequency-dependent selection, if prey have sufficient incentive (in terms of resource or reproductive gain) to encounter predators by moving through their habitats, or perhaps by growing to a size at which crypsis is ineffective. If a prey’s behavior or life-history characteristics such as size, tend to make it exposed to predators then, as discussed earlier, apomorphic displays add a smaller marginal conspicuousness cost compared to the situation in which the prey have little intrinsic exposure to predation. Given that many prey animals are mobile and therefore fairly conspicuous to predators, it therefore seems reasonable to argue that the initial evolution of conspicuous appearance traits (aposematic signaling) is therefore not particularly difficult to explain.

THE ECOLOGICAL ROLE OF APOSEMATIC DISPLAY

If prey mobility can explain the initial evolution of aposematism, it follows that aposematic prey should be typically ones whose behavior brings them into contact with predators. Furthermore, a novel prediction of the model is that coevolution of aposematic display and behavioral conspicuousness leads to exaggeration of both traits. Displays allow prey to expose themselves to higher rates of encounter with predators, so that resources can be exploited, and in turn, this facilitates exaggeration of aposematic displays. Hence, a correspondence between movement in front of predators and aposematism is clearly made. It has, in fact, been recognized in the literature for some time that the switch from nocturnal to diurnal activity is one of the major mechanisms by which this may happen (Leimar et al. 1986). There is growing empirical evidence that this is the case; for example, in ditrysian Lepidoptera (Merilaita and Tullberg 2005) and in dendrobatid (poison arrow) frogs (Santos et al. 2000). A similar idea exists within studies of marine invertebrates, in which the acquisition of chemical defenses this has been linked to profound changes in life-history pattern made possible by behavioral conspicuousness (Lindquist and Hay 1996). It is interesting to note that recent comparative studies of an amphibian and an insect group have found positive correlations between measures of body size and aposematic coloration (e.g., Forsman and Hagman 2003; Whitman 2008; Whitman and Vincent 2008), suggesting that the evolution of aposematism may indeed be involved in significant change in niche (see related dynamic models in Higginson and Ruxton 2010).

Our simulations predict some quantitative variation in the conspicuousness of aposematic signaling, which relates positively to the mobility of animals (or more generally their ecological level of exposure to predatory threat). One qualitative test of this idea is to examine the degree of aposematic coloration within different life-history stages, testing the prediction that the most mobile stages should have the brightest colouration. Pupal forms of lepidopteran species that are aposematic in larval or adult stages are often cryptic (see references in Wiklund and Sillén-Tullberg 1985), consistent with the fact that they are completely stationary and hidden away from predators. In addition, it may be possible to test our general hypothesis by examining whether aposematically colored animals make use of a greater range of food plants than closely related, but cryptically colored species. Thus, it may be that our general qualitative prediction of a positive correlation between strength of aposematic signaling and movement by prey could be evaluated using phylogenetically controlled datasets.

THE EVOLUTIONARY CONSEQUENCES OF APOSEMATIC DISPLAYS

Aposematism may be a key evolutionary innovation, in the sense of being a “novel trait . . . that influences ability to exploit resources hitherto little used” (Schluter 2000, p. 181), and thereby could facilitate adaptive radiation by providing access to greater ecological opportunity or a new “adaptive zone.” Evidence supporting the idea that aposematism is in fact a key innovation has recently been published. Przeczek et al. (2008) examined data from a range of animals including amphibians, hymenopterans, lepidopterans, and coleopterans. They found that in 11 of 14 cases, groups that contained aposematism had a higher rate of diversification than their sister groups that lack aposematism. The authors conclude that reduction in predation may lead to increased divergence within aposematic groups. However, we argue that it is the broader ecological consequences of aposematism that are more likely to make it a key evolutionary innovation. First, moderate increases in range size can increase the likelihood of speciation whereas concomitantly reducing the risk of extinction (Gaston 1988). Second, evolutionary transitions from crypsis to aposematism may often be selected for because this provides access to a wider range of ecological opportunities. Ecological opportunity is a prerequisite for adaptive radiation, where competition for novel
resources drives niche specialization, which may in turn lead to specialization as a byproduct (i.e., ecological speciation Rundle and Nosil 2008). Animals that achieve relatively low levels of predation can make use of many different facets of the environment: for example, specializing their diet on valuable foods that cannot be collected by cryptic animals. One example of this is seen in the Dendrobatid frogs, in which aposematism is associated with changes in diet (from generalist to dietary specialist, with aposematic species foraging on ants and similar sized prey) and from nocturnal to diurnal feeding patterns (Santos et al. 2003; Summers 2003). With relatively free movement around a habitat, prey can also make use of non-dietary opportunities such as escape from competition, parental care, enhanced thermoregulatory efficiency, and sexual signaling. Thus, crypsis and aposematism as modes of antipredator defense are unlikely to be equivalent in their evolutionary consequences for animal lineages: aposematism is likely to act as a key innovation whereas crypsis is not. Similar points have recently been made by Whitman and Vincent (2008) in relation to the highly toxic lubber grasshoppers, and in relation to the acquisition of dual secondary defenses in Platyphora leaf beetles by Termonia et al. (2002). Therefore, we might expect to observe “escape and radiate” coevolutionary dynamics (Ehrlich and Raven 1964) between predators and aposematic but not cryptic prey lineages.

**Conclusions**

We have argued in this article that an emphasis on aposematism purely as a signaling system is oversimplistic in ecological terms. Given that aposematic display imposes a major cost of raised encounter rates with predators, it may often exist as a response to ecological opportunity. Furthermore, there may be coevolutionary exaggeration of behavioral and aposematic traits, leading to a general prediction that more active prey may have brighter aposematic signals. Finally, the evolutionary function of aposematism as a key innovation is only just beginning to be explored. We argue that if aposematism is a key innovation, it is because of its ecological role in widening ecological opportunity.

**ACKNOWLEDGMENTS**

M. Speed was supported by a Leverhulme Research Fellowship and by NERC grants NE/D01667/1 and NE/E018521/1. We thank P. Stephens for advice with the modeling method and two anonymous reviewers for their very helpful suggestions. Modeling was performed by MS in conjunction with GDR. Article was written by MS, GDR and MB.

**LITERATURE CITED**


Poulton, E. B. 1890. The colours of animals: their meaning and use especially in relation to animal economy and habit. London.


Associate Editor: C. Jiggins

Supporting Information

The following supporting information is available for this article:

Appendix S1. Alternative formulations for equations used.

Figure S1. Optimization of traits in relation to changed season lengths.

Figure S2. Re-simulating Figure 1 with equation (A6) substituting for equation (6), the results are qualitatively very similar, as are the other results in the main paper.

Figure S3. Re-simulation of Figure 2 (main paper) but with equation 4a and ε = 0 and ε = 1,

Supporting Information may be found in the online version of this article.

(This link will take you to the article abstract).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.