Aposematism: what should our starting point be?

Michael P. Speed^{1*} and Graeme D. Ruxton²

¹*School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, UK*

²*Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, UK*

The evolution of aposematism is considered to be a major evolutionary problem because if new aposematic forms emerged in defended cryptic populations, they would face the dual problems of rarity and conspicuousness. We argue that this commonly assumed starting point might not have wide validity. We describe a novel evolutionary computer model in which prey evolve secondary defences and become conspicuous by moving widely over a visually heterogeneous habitat. Unless crypsis imposes high opportunity costs (for instance, preventing prey from efficient foraging, thermoregulation and communication), costly secondary defences are not predicted to evolve at all. However, when crypsis imposes opportunity costs, prey evolve secondary defences that facilitate raised behavioural conspicuousness as prey exploit opportunities within their environment. Optimal levels of secondary defence and of behavioural conspicuousness increase with population sizes and the costs imposed by crypsis. When prey are already conspicuous by virtue of their behaviours, the evolution of aposematic appearances (bright coloration, etc.) is much easier to explain because aposematic traits add little further costs of conspicuousness, but can bring large benefits.

Keywords: aposematism; crypsis; predator; frequency-dependent selection; secondary defence

1. INTRODUCTION

A. R. Wallace originally invented the concept now known as aposematism to describe prey that combine warning displays with secondary defences (Wallace 1867, 1889). More than a century later, the evolution of aposematism remains a remarkably fertile and controversial area of research (Ruxton *et al.* 2004). Warning displays are still of interest to researchers, in part, because the proximate mechanisms by which they operate tell us much about predator behaviour and predator–prey coevolution. As originally envisaged by Wallace (1867) and Poulton (1890), warning displays function to enhance discrimination, to accelerate learning and perhaps slow down forgetting.

However, aposematic displays remain the focus of considerable attention because, for many researchers, their initial origins contain at least two important evolutionary paradoxes. First, it is generally assumed that before the first aposematic traits evolved, prey were both highly cryptic and had effective secondary defences. If secondary defences are costly (and they often are—see Ruxton *et al.* (2004)), then their presence in prey already highly protected by crypsis is paradoxical: why pay for repellent secondary defences if your enemy rarely finds you (Speed & Ruxton 2002)? Second, there is a better-known paradox of warning signals, which also emerges from commonly held assumptions about initial conditions. Ever since the seminal theoretical model of Harvey *et al.* (1982), it is widely taken that aposematic mutants must emerge from defended cryptic species (Sillén-Tullberg & Bryant 1983; Sundberg 1987; Alatalo & Mappes 1996; Yachi & Higashi 1998; Lindström *et al.* 1999; Servedio 2000; Brodie & Agrawal 2001; Speed 2001; Sherratt 2002; Sherratt & Beatty 2003; Thomas *et al.* 2003; Franks & Noble 2004). When this is

the case, new aposematic forms suffer combined and highly effective barriers to survival because of their rarity and their conspicuousness.

This second problem can be easily illustrated with a simple example. In a scenario in which a prey passes a predator during a specified period, a new rare aposematic mutant has a high chance of being detected (e.g. 0.5). Because the aposematic form is rare and unfamiliar, it may also have a high probability of attack given detection (say $p = 1$), giving a high overall chance that the prey does not survive the period ($0.5 \times 1 = 0.5$). By contrast, the much more numerous cryptic conspecific has a lower probability of detection (e.g. 0.15), and probability of an attack given detection is also lower (e.g. 0.6), because the prey is relatively familiar. Hence, for the common and familiar cryptic conspecific the probability of death during the period is much smaller ($0.6 \times 0.15 = 0.09$) than for a rare, conspicuous conspecific.

Given such strong frequency-dependent selection against a new aposematic form, it seems remarkable that such prey could evolve at all. However, aposematism patently exists and is common in the natural world. In addition, the benefits of aposematic display to very common aposemes may be very large. Imagine, for example, that the probability of detection remains at 0.5, but that because of accurate discrimination, fast learning, etc. attack given detection is only 0.01, now the aposematic prey has a greater likelihood of survival (i.e. probability of death is now 0.005, less than for its cryptic conspecifics).

Hence, the initial evolution of aposematism provides an important evolutionary challenge that has been taken up by several researchers who have used laboratory and theoretical simulations to explain the evolution of conspicuousness via the ‘rarity and conspicuousness’ scenario (reviewed in Speed 2003). Most explanations use a combination of predator psychology (wariness of novel

* Author for correspondence (speedm@liverpool.ac.uk).

foods, learning, memorability and biased generalization) and prey population structures (especially spatial clumping and structures conducive to random drift). Nonetheless, several authors from Poulton (1890) onwards have suggested alternative routes, which may be more feasible. These include sexual selection (and other forms of pre-adaptation, in which existing traits are turned into aposematic traits when a secondary defence arises; Mallet & Singer 1987) and facultative aposematism (in which aposematic displays are only manifest at high densities; Sword 2002).

In this paper, however, we focus on the validity of the commonly used 'cryptic and defended' starting point. We argue that if defences are costly and crypsis is already highly effective, secondary defences are unlikely to exist, except by pre-adaptation. Alternatively, defences may evolve because they facilitate environmental exploitation, allowing prey to use important opportunities in their environments, often in direct view of predators. In the main part of this paper, we evaluate how defences and a form of 'behavioural conspicuousness' may evolve when crypsis imposes opportunity costs. To do this we present a novel stochastic evolutionary model. Subsequently, we argue that if prey without evolved aposematic displays are quite conspicuous anyway, the subsequent evolution of aposematic traits are then quite easy to explain.

(a) *Costs of crypsis and the evolution of secondary defence*

We assume that the conspicuousness of a prey (and also the effectiveness of its crypsis) has two components that we treat as independent: external appearance and behaviour. Both components may incur costs. Thus, to match a particular background (or otherwise generate a cryptic appearance) a prey may have to forgo external characteristics that could, for example, enhance thermoregulatory efficiency (e.g. dark, black patches), and facilitate social/sexual signalling and aposematic display. Similarly, crypsis may incur opportunity costs by constraining the full range of an animal's behaviour (see the excellent discussion in Stamp & Wilkens (1993)). Crypsis may be incompatible with movement across microhabitats from which significant benefits can be obtained (here termed 'environmental exploitation benefits'). Hiding under a rock may provide excellent protection from predation, but can prevent essential behaviours such as effective foraging, thermoregulation, mating, communication and parasite avoidance (Dunham & Tierney 1983; Stamp & Wilkens 1993; Lindquist & Hay 1996; Moses & Sih 1998; Carrascal *et al.* 2001). Hence, to greater or lesser extents, crypsis is costly because of the constraints it places both on appearance and on behaviour (and see related discussion in Wiklund & Silén-Tullberg (1985); Merilaita & Tullberg (2005)).

The trade-off between behaviour and environmental exploitation is well known in parts of the literature on foraging and it is not our intention here to repeat detailed modelling of spatial distribution of animals (Abrams 1991; Sih 1992; Grand 2002). Instead, we first describe a model that evaluates the importance of determinants of variation in 'behavioural conspicuousness' and then, in our discussion, apply the predictions of the model to the problem of aposematic evolution.

Figure 1. Illustration of the prey's conspicuousness against a background. We assume that the habitat's background varies continuously along a single dimension, so that as the prey moves over more backgrounds it becomes more conspicuous. Benefits to a prey increase to a maximum gained if it divides its time evenly across the whole range of backgrounds. In this illustration we arbitrarily choose an initially dark prey (squares) that matches dark areas of the habitat. However, the assumption is that the costs of conspicuousness are similar regardless of the initial coloration of the prey along this stimulus dimension.

(b) *A model*

We consider a prey population with a very simple life-history strategy: asexual reproduction occurring in discrete generations. The generation time T is fixed, taking a value of 100 unless otherwise stated. After this time, all survivors reproduce (clonally) then die, their progeny beginning a new generation.

Individual prey items all have the same appearance but can potentially differ in two heritable characteristics: their 'behavioural conspicuousness' to predators (C) and their level of defence (D), such that the genotype of individual i is fully specified by the pair $\{C_i, D_i\}$.

C_i can be interpreted as the rate at which individual i encounters the predator. For simplicity, we assume that the locality within which predators and prey exist varies continuously along a single visual dimension (see illustration in figure 1) and hence, other things being equal, the value of C_i is set by the average degree of contrast between the prey and its chosen backgrounds within the habitat. The background is assumed to be sufficiently heterogeneous that for prey of fixed appearance crypsis is not compatible with comprehensive use of the habitat (cf. Merilaita *et al.* 1999). We run the model with a single predator, but our results are not sensitive to this assumption.

We assume a non-zero lower bound to C_i such that even maximally cryptic prey (those that stay in one locality of the habitat in which they match the background very closely) can be detected by the predator, albeit at a low rate. As a default, we set this minimum ($C_{\min}$) to 0.0005, so that the

expected time for the predator to find an individual, maximally cryptic prey item is 20 times the generation time T . We begin each generation at time $t = 0$, with a fixed initial number of individual prey items N_1 . We first calculate the time (t_1) until the predator first encounters a prey item. This is simply a random number drawn from a negative exponential distribution with mean value equal to the inverse of the total encounter rate with all prey alive at that time:

$$t_1 \rightarrow \frac{1}{\exp(-\sum_{i=1}^{N_1} C_i)}.$$

We then move forward to time $t = t_1$, and stochastically select the prey individual that the predator encounters. The probability that individual i is the one encountered is based on the relative conspicuousness of that individual: specifically this probability (P_i) is given by

$$P_i = \frac{C_i}{\sum_{j=1}^{N_1} C_j}.$$

When the predator encounters individual i , the outcome is dependent on the defence level of that individual, D_i , in two ways. First, the time taken in the encounter increases with defence level. For simplicity we define defence in this manner: if the predator encounters an individual with defence level D_i , then it takes a time D_i for the interaction with that prey item to be concluded. During this time, no encounters with other prey items can occur. This can be interpreted in several ways. For example, D_i can be the time needed to overcome a physical defence such as spines or a tough integument. Alternatively, D_i can be interpreted as the time needed to recover physiologically from ingested toxins during which searching for prey is curtailed. If the value of D_i increases then the probability that a focal individual will be attacked over the season declines. This provides the kind of selective environment we need to simulate prey evolution, so that poor secondary defences make a species vulnerable to predation after detection. However it is interpreted, if the first prey encountered (after time t_1) is individual i , then that encounter takes a further time D_i , so the predator begins looking for further prey at time $t = t_1 + D_i$.

A second outcome of variation in prey defence is variation in the probability of prey survival. In some simulations we assume that any attack leads to death of the prey. However, in others D_i determines whether individual i survives the encounter with the predator (cf. Wiklund & Järvi 1982). We expect that probability of surviving an encounter S_i should increase with D_i , and assume the functional form

$$S_i = 1 - \exp(-sD_i),$$

for some positive constant s . The survivorship of a given encounter is determined stochastically. However, increasing values of the survivorship constant s are associated with improved survival. Unless otherwise specified, we assume that $s = 0.5$. This means that if there is no investment in defence, then the probability of surviving an attack is zero, rising to 5% if $D_i = 0.1$, 22% if $D_i = 0.5$ and 40% if $D_i = 1$.

If the first prey item encountered in a given generation survives, then the number of prey items from which the

second encountered individual will be randomly drawn remains unchanged ($N_2 = N_1$), otherwise the first encountered individual will be removed from the population and $N_2 = N_1 - 1$ remain. We then proceed to calculate the further time until the predator encounters a second prey item, drawn stochastically from the appropriate negative exponential distribution:

$$t_2 \rightarrow \frac{1}{\exp(-\sum_{i=1}^{N_2} C_i)}.$$

The identity of that individual is also obtained stochastically as before, using the probabilities

$$P_i = \frac{C_i}{\sum_{j=1}^{N_2} C_j}.$$

The duration and outcome of this second interaction is determined stochastically, exactly as for the first interaction. We continue in this fashion until the end of the generation (i.e. when time $t > T$). At this point there will be some number of individuals surviving, N_i . We will select N_1 progeny from these, to begin the next generation (resetting t to zero again). We assume that the reproductive ability of individual i (R_i) decreases with its level of defence (representing the idea that defences are costly), and may increase with conspicuousness (representing opportunity costs). We will define the exact functional form of R_i later. The probability that individual i is the parent of the j th offspring is simply

$$\frac{R_i}{\sum_{j=1}^{N_i} R_j}.$$

An individual is selected randomly (with replacement) independently to be a parent for each of the N_1 offspring. The genotype of offspring j $\{C_j, D_j\}$ will be determined in part stochastically and in part by the genotype of its parent $\{C_i, D_i\}$. If we assume that the values C and D can take are continuous, then we assume that individuals inherit their parent's value of a given characteristic plus a small perturbation drawn from the uniform distributions $[-\epsilon_c, \epsilon_c]$ and $[-\epsilon_d, \epsilon_d]$, subject to resampling to ensure that C and D values remain above their minimum values. For C we assume a non-zero $C_{\min}$ (see above) and $\epsilon_c = 0.0001$; For D we assume that all non-negative values are possible and $\epsilon_d = 0.01$.

To run a simulation, the number of individuals at the start of a generation N_1 must be specified, and $\{C_i, D_i\}$ values given to all these individuals. We assume $N_1 = 300$ as a default, and that the population starts with C values distributed evenly between $C_{\min}$ and $1.1C_{\min}$. D values are initially distributed uniformly between 0 and 0.01. That is, the population begins with individuals that are close to maximally cryptic and do not invest substantially in defence. Simulations generally run for 10^4 generations over which we expect the distribution of C and D values to settle down to equilibrium values (regardless of the starting values); periodic checks have shown that for the parameter values used this is the case.

We assume that the costs of defence and release from the opportunity costs of crypsis act independently on reproductive ability so that $R_i(D_i, C_i)$ can be decomposed into a product of two functions $f(D_i)$ and $g(C_i)$. An appropriate

Figure 2. The mean and standard deviation (across the population) of the (a) defensive investment (D_i) and (b) conspicuousness (C_i) of individual in the population at the 10^4 generation, as a function of the parameter controlling the benefits of conspicuousness (α).

form for f might be a simple negative exponential

$$f(D_i) = \exp(-C_D D_i)$$

for some positive constant C_D . Increased costs of defence are represented by increasing C_D . This clearly has the property that increasing investment in defence (increasing D_i) reduces reproductive ability R_i . As a default, we give C_D a value of 0.1, such that f is 1 when $D_i = 0$, dropping to 0.9 when $D_i = 1$ and 0.37 when $D_i = 10$.

The function g could also have an exponential form:

$$g(C_i) = 1 - \alpha \exp(-\beta(C_i - C_{\min})),$$

where α and β are positive constants. The constant α lies between zero and one and can be seen as describing maximum contribution that saving opportunity costs of crypsis can make to reproductive ability. For example, if $\alpha = 0.2$ then the benefits of giving up crypsis contribute 20% of the reproductive ability of a (highly conspicuous) maximally reproductive individual. The other parameter β controls how quickly this benefit from saving the opportunity costs of crypsis is felt with increasing conspicuousness levels. We set $\beta = 400$; thus when $C = 10C_{\min} = 0.005$, then 83% of the benefits of increased conspicuousness can be collected.

2. RESULTS

To explore the effects of opportunity costs on the evolution of conspicuousness and defence, we hold all other parameters at their default values and vary the benefit of giving up crypsis (α). The model predictions can be seen in figure 2 for situations in which survival after an attack (S_i) is related to the level of prey defence D_i . When $\alpha = 0$, and there are no opportunity costs of crypsis, then the evolutionarily stable strategy is to be maximally cryptic and invest little in defence. This remains generally true for low values of opportunity costs (α values less than 0.4 in figure 2). However, for substantial opportunity costs, the population evolves to levels of conspicuousness significantly higher than those associated with maximal crypsis ($C_{\min} = 0.0005$). This occurs because although increasing

conspicuousness increases an individual's risk of being attacked by the predator, this becomes a price worth paying when there are substantial opportunity costs to crypsis. If the more conspicuous individual survives, then it has a reproductive advantage over less conspicuous survivors. However, because the risk of being attacked increases linearly with increasing conspicuousness and the benefits of crypsis saturate, the population does not evolve ever higher levels of conspicuousness, but rather settles to an equilibrium where the costs and benefits of conspicuousness are most advantageously traded off. This equilibrium value increases as α (and so the opportunity costs of crypsis) increase.

When α is 0.5, and so a maximally cryptic individual that survives has only 50% of the reproductive capability of a maximally conspicuous individual, then the population average value of conspicuousness is over five times that of a maximally cryptic individual. However, interestingly, the increase in conspicuousness is mirrored by an increase in investment in defence as α increases. This occurs because investment in defence is more attractive for individuals that are more conspicuous and so more likely to encounter the predator. The more conspicuous an individual is, the more attractive investment in defence becomes. The reproductive cost of investment in defence means that there is not continual evolution towards ever-higher levels of defence. Investment is relatively modest in figure 2, with average values of D rising to approximately 3 when opportunity costs of crypsis are very high. A value of 3 for D_i gives a 78% chance of surviving an attack at a cost of a 26% reduction in reproductive ability (in the order of costs shown by Hetz & Slobodchikoff (1990)). These results are qualitatively unchanged by changes to the exact values given to various model parameters.

We can see a general trend of increasing conspicuousness and defence with increasing opportunity costs of crypsis, and this leads to a strong positive correlation between conspicuousness and defence. Note that although the overall trend of increasing defence with increasing opportunity costs of crypsis is clear from figure 2, the rise appears not to be monotonic. For example, there is a slight decrease in defence as α rises from 0.5 to 0.55. This is an artefact of figure 2, showing a snapshot of the state of the population at generation 10^4 . The small population size and stochastic nature of the model means that the population levels of both defence and conspicuousness do fluctuate a little over time, although figure 3 illustrates that this fluctuation is relatively minor, and after an initial phase of rapid evolution in the first 1000 generations, the system remains relatively settled.

Reduction in the population size (from $N_1 = 300$) has the effect of lowering optimal levels of defence and conspicuousness (figure 4). This result is perhaps predictable, because the level of inhibition of predator attacks is reduced as the size of the prey population decreases; hence conspicuousness becomes more costly in smaller populations.

If defences confer no survival advantage to an individual when attacked by a predator ($S_i = 0$), then we anticipate that they will not evolve, because, as Fisher originally pointed out, kin selection would be necessary for such defences to evolve as adaptations (Fisher 1930). Our model confirms Fisher's argument. Without kin-structuring secondary defences that provide no direct individual protection

Figure 3. The mean levels of (a) defence (D_i) and (b) relative conspicuousness ($C_i/C_{\min}$) for each generation for a realization of the model with $\alpha = 0.55$.

only exist in the population as a result of random drift in a small population and high levels of mutation. Such secondary defences do not therefore evolve adaptively in response to increases or decreases in the opportunity costs of crypsis (i.e. variation in α ; figure 5). Nonetheless, even without selection for secondary defence prey do vary their optimal rates of conspicuousness, especially for values of α greater than 0.4, albeit to lowered maximum levels (figures 2 and 5). Comparison of figures 2 and 5 shows that although evolved defences are not necessary for prey to raise their optimal level of behavioural conspicuousness, they are necessary for behavioural conspicuousness to reach high optimal levels.

3. DISCUSSION

We consider first the evolution of secondary defences and then the evolution of conspicuousness and aposematic signalling.

(a) Evolution of secondary defences

Because both behavioural conspicuousness and levels of secondary defence are allowed to evolve in this system, we can see that both traits are optimized in relation to, and vary with, the benefits that can be gained by forgoing high levels of crypsis (defined by α and β in the model). In our simulations, costly secondary defences evolved primarily because they facilitated exploitation of opportunities within a visually heterogeneous habitat (Bakus 1981). Within the parameters of our model, optimal investment in defences generally increased with the value of environmental benefits (or, alternatively, increased with the opportunity costs of crypsis) and with the size of the prey population, provided they conferred a survival benefit to individuals during an attack. Hence the presence of a costly secondary defence in highly cryptic prey, as so often assumed in theoretical models of aposematism (reviewed in Ruxton *et al.* (2004)), is unstable within the parameters of our system.

In the only other published model that, to our knowledge, considers the combined evolution of secondary defences and conspicuousness (the seminal Evolutionary Stable Strategy model described by Leimar *et al.* (1986)), a similar conclusion about the instability of costly secondary

Figure 4. As figure 2 except that α is fixed at 0.5 and the maximum population size (N_1) is varied between simulations.

defences in highly cryptic prey was drawn. In the model described by Leimar *et al.* (1986), it becomes optimal for prey to invest in secondary defences when crypsis is degraded because the number of predators and/or the duration of exposure to predators increases. Though they did not include opportunity costs of crypsis in their model, Leimar *et al.* (1986) suggested that 'when a species shifts from profitable to unprofitable, some major change in conditions should have preceded the shift' (p. 481). In our evolutionary simulations this would be consistent with increases in the opportunity costs of crypsis, giving prey ecological reason to invest in secondary defences (cf. Guilford 1988). Leimar *et al.* (1986) further suggest that a switch from nocturnal to diurnal activity might provide sufficient cause for unprofitability to evolve. Even within daylight conditions in aquatic and terrestrial environments there may still be many opportunities lost to prey without suitable secondary defences (Schmidt 1990; Lindquist & Hay 1996).

Another important possibility is that highly cryptic prey are pre-adapted to the generation of secondary defences, for instance because they happen to produce chemicals repellent to predators. Potential examples include chemicals used in sexual communication mechanisms, in the subjugation of prey, anti-parasite/parasitoid defences, antimicrobial properties, wetting or waterproofing agents and intraspecific conflict (Whitman *et al.* 1990; Weller *et al.* 1999). Similarly, certain morphological traits such as spines or tough skins may have primary functions other than predator deterrence (such as interspecific aggression, protection from abrasion, etc.), but may also aid in survival from predation, in which case a prototype secondary defence exists without additional costs. The possibility of pre-adaptation may explain several instances of the initial evolution of chemical secondary defences. However, we note that pre-adaptation to generation of secondary defence is compatible with the expansion of habitat to use environmental opportunities (cf. discussion in Guilford (1988)). Furthermore, if prototype secondary defences that exist because of pre-adaptation require investment to

Figure 5. Identical simulations to those of figure 2 except that we assume that all attacked prey (regardless of their level of defence) are killed by the predator ($S_i = 0$).

maximize their effectiveness as defences, then optimization of the effectiveness of the defence will occur in a manner similar to that simulated here.

(b) *Behavioural conspicuousness and the evolution of aposematic displays*

The central aim of our paper is to demonstrate that the common assumption that aposematic displays evolved in initially cryptic populations is unlikely; hence explaining the initial evolution of aposematism is not a major evolutionary problem. Species that are behaviourally conspicuous are good candidates for the evolution of novel aposematic displays. Consider an extreme case in which individuals with secondary defences in an ancestral population are already very highly conspicuous by virtue of their behaviours; changes in their appearance can now have little further effect on their risk of detection and in these circumstances aposematic displays can add few extra costs of conspicuousness.

Suppose that we measure both behavioural conspicuousness (c_b) and conspicuousness by individual appearance (c_a) on a scale between zero and one, such that overall conspicuousness c_t is, for simplicity, defined in a linear fashion, e.g. $c_t = c_b + c_a(1 - c_b)$; then if c_b is high, 0.9, the effect of wide variation in c_a on the value of overall conspicuousness is necessarily small. By contrast, if the prey were initially highly cryptic (e.g. $c_b = 0.01$), the costs from increased conspicuousness of even modestly conspicuous aposematic displays (e.g. $c_a = 0.4$) would be relatively great. This suggests that there may be some relationship between the level of behavioural conspicuousness that a prey manifests and the conspicuousness of the aposematic signal that it can sustain. The remarkable finding by Summers & Clough (2001) that conspicuousness of dendrobatid coloration is correlated with the level of a prey's defence may be thus explained if there is also a correlation between

behavioural conspicuousness and defence levels in these animals. Clearly, there is scope for phylogenetically controlled investigations of the relationship between the level of a secondary defence and the behavioural conspicuousness of prey.

If we accept that full-blown aposematic displays are more likely to evolve in behaviourally conspicuous prey than in behaviourally cryptic ones, we still need to be able to explain why such traits evolve at all. Over the years, many explanations have been put forward including combinations of receiver psychology effects (of accelerated learning, memory and wariness), stochastic population effects and characteristics of prey distribution (reviewed in Ruxton *et al.* 2004). Each of these explanations remains valid but should in our view be re-evaluated, relaxing the narrow assumption that prey were initially cryptic before the evolution of aposematic displays.

Furthermore, as Wallace originally envisaged, warning displays might be conspicuous so as to be 'very distinct from the protective tints of the defenceless animals allied to them' (Wallace 1889, p. 232). Hence, a good reason that aposematism may evolve initially is to prevent confusion with undefended prey (Guilford 1990; Sherratt & Beatty 2003). On its own, behavioural conspicuousness itself may not be a sufficiently reliable signal of non-profitability to function as an aposematic display. As we found in figure 5, prey can evolve some heightened levels of behavioural conspicuousness even when they do not evolve adaptive secondary defences. Hence, some additional discriminative cue may be necessary for defended prey to minimize erroneous attacks by educated predators.

Such cues may initially be traits beneficial to individuals in their own right, that in addition function as discriminative cues. Probable examples include: conspicuous stimuli that direct predator attacks toward dispensable organs, perhaps those in which secondary defences are localized; and traits that enhance, for example, thermoregulatory efficiency, parasite resistance and sexual signalling (see Poulton 1890; Mallet & Singer 1987; Sword 1999). Again, if prey are already behaviourally conspicuous, such traits can add relatively small increases in overall conspicuousness while in each case bringing direct benefits. When such traits are beneficial in themselves, they can evolve from a point of rarity and take on qualities of aposematic display (*sensu* Wallace) as they become common. Furthermore, as Sherratt (2002) recently argued, predators may often select conspicuous traits to use as discriminative aposematic cues (rather than less conspicuous but similarly discriminative traits), because they are likely to function as reliable indicators of defence: the costs of conspicuous aposematic displays are borne much more easily by individuals with effective secondary defences than those with poor defences (see also Sherratt & Beatty 2003). In summary, we suggest that one plausible route to aposematism (but by no means the only plausible route) is roughly: behavioural conspicuousness plus secondary defence; acquisition of traits that can be borne by behaviourally conspicuous prey and are beneficial in themselves; use by predators of a conspicuous subset of these traits as reliable cues aiding in the discrimination of defended from edible prey; coevolution of conspicuous traits and predator preferences towards highly evolved warning displays (cf. Turner 1984; Mallet & Singer 1987).

4. CONCLUSIONS

Most theoretical models of aposematism focus on the evolution of a conspicuous component of an animal's appearance, and ignore the evolution of secondary defences (though see the important exception in Leimar *et al.* (1986)) and of behavioural conspicuousness. These models almost deliberately use the most difficult scenario to explain the starting point of aposematism: that of rarity and conspicuousness. However, the model outlined here indicates that when crypsis imposes strong opportunity costs and when habitats are visually heterogeneous, we may expect that secondary defences evolve to facilitate the behavioural conspicuousness necessary for optimal exploitation of a habitat. When these conditions pertain, aposematism is relatively easy to explain and is not a major evolutionary problem.

The authors are grateful to two anonymous referees for their comments, and to Tom Sherratt for numerous discussions.

REFERENCES

- Abrams, P. A. 1991 Life-history and the relationship between food availability and foraging effort. *Ecology* **72**, 1242–1252.
- Alatalo, R. V. & Mappes, J. 1996 Tracking the evolution of warning signals. *Nature* **382**, 708–710.
- Bakus, G. J. 1981 Chemical defence mechanisms on the Great Barrier Reef, Australia. *Science* **211**, 497–498.
- Brodie, E. D. & Agrawal, A. F. 2001 Maternal effects and the evolution of aposematic signals. *Proc. Natl Acad. Sci. USA* **98**, 7884–7887.
- Carrascal, L. M., Diaz, J. A., Huertas, D. L. & Mozetich, I. 2001 Behavioral thermoregulation by treecreepers: trade-off between saving energy and reducing crypsis. *Ecology* **82**, 1642–1654.
- Dunham, D. W. & Tierney, A. J. 1983 The communicative cost of crypsis in a hermit crab *Pagurus marshi*. *Anim. Behav.* **31**, 783–785.
- Fisher, R. A. 1930 *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Franks, D. W. & Noble, J. 2004 Warning signals and predator-prey coevolution. *Proc. R. Soc. B* **271**, 1859–1865. (doi:10.1098/rspb.2004.2795)
- Grand, T. C. 2002 Foraging-predation risk trade-offs, habitat selection, and the coexistence of competitors. *Am. Nat.* **159**, 106–112.
- Guilford, T. 1988 The evolution of conspicuous coloration. *Am. Nat.* **131**, S7–S21.
- Guilford, T. 1990 The evolution of aposematism. In *Insect defenses: adaptive mechanisms and strategies of prey and predators* (ed. J. O. Schmidt & D. L. Evans), pp. 23–61. New York: State University of New York Press.
- Harvey, P. H., Bull, J. J., Pemberton, M. & Paxton, R. J. 1982 The evolution of aposematic coloration in distasteful prey—a family model. *Am. Nat.* **119**, 710–719.
- Hetz, M. & Slobodchikoff, C. N. 1990 Reproduction and the energy-cost of defense in a Batesian mimicry complex. *Oecologia* **84**, 69–73.
- Leimar, O., Enquist, M. & Sillén-Tullberg, B. 1986 Evolutionary stability of aposematic coloration and prey unprofitability—a theoretical analysis. *Am. Nat.* **128**, 469–490.
- Lindquist, N. & Hay, M. E. 1996 Palatability and chemical defense of marine invertebrate larvae. *Ecol. Monogr.* **66**, 431–450.
- Lindström, L., Alatalo, R. V., Mappes, J., Riipi, M. & Vertanen, L. 1999 Can aposematic signals evolve by gradual change? *Nature* **397**, 249–251.
- Mallet, J. & Singer, M. C. 1987 Individual selection, kin selection, and the shifting balance in the evolution of warning colors—the evidence from butterflies. *Biol. J. Linn. Soc.* **32**, 337–350.
- Merilaita, S. & Tullberg, B. S. 2005 Warning colouration and constrained crypsis. *Evolution*. (In the press.)
- Merilaita, S., Tuomi, J. & Jormalainen, V. 1999 Optimization of cryptic coloration in heterogeneous habitats. *Biol. J. Linn. Soc.* **67**, 151–161.
- Moses, J. L. & Sih, A. 1998 Effects of predation risk and food availability on the activity, habitat use, feeding behavior and mating behavior of a pond water strider, *Gerris marginatus* (Hemiptera). *Ethology* **104**, 661–669.
- Poulton, E. B. 1890 *The colours of animals: their meaning and use especially considered in the case of insects*. London: Kegan Paul, Trench, Trubner & Co.
- Ruxton, G. D., Sherratt, T. N. & Speed, M. P. 2004 *Avoiding attack: the evolutionary ecology of crypsis, aposematism and mimicry*. Oxford University Press.
- Schmidt, J. O. 1990 Hymenopteran venoms: striving toward the ultimate defense against vertebrates. In *Insect defenses: adaptive mechanisms and strategies of prey and predators* (ed. D. L. Evans & J. O. Schmidt), pp. 387–420. New York: State of New York University Press.
- Servedio, M. R. 2000 The effects of predator learning, forgetting, and recognition errors on the evolution of warning coloration. *Evolution* **54**, 751–763.
- Sherratt, T. N. 2002 The coevolution of warning signals. *Proc. R. Soc. B* **269**, 741–746. (doi:10.1098/rspb.2001.1944)
- Sherratt, T. N. & Beatty, C. D. 2003 The evolution of warning signals as reliable indicators of prey defense. *Am. Nat.* **162**, 377–389.
- Sih, A. 1992 Prey uncertainty and the balancing of anti-predator and feeding needs. *Am. Nat.* **139**, 1052–1069.
- Sillén-Tullberg, B. & Bryant, E. H. 1983 The evolution of aposematic coloration in distasteful prey—an individual selection model. *Evolution* **37**, 993–1000.
- Speed, M. & Ruxton, G. D. 2002 Animal behaviour: evolution of suicidal signals. *Nature* **416**, 375–375.
- Speed, M. P. 2001 Can receiver psychology explain the evolution of aposematism? *Anim. Behav.* **61**, 205–216.
- Speed, M. P. 2003 Theoretical developments in the understanding of warning signals. *Comments Theor. Biol.* **8**, 207–224.
- Stamp, N. E. & Wilkens, R. T. 1993 On the cryptic side of life: being unapparent to enemies and the consequences for foraging and growth in caterpillars. In *Caterpillars: ecological and evolutionary constraints on foraging* (ed. N. E. Stamp & T. M. Casey), pp. 283–330. New York: Chapman & Hall.
- Summers, K. & Clough, M. E. 2001 The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proc. Natl Acad. Sci. USA* **98**, 6227–6232.
- Sundberg, P. 1987 A possible mechanism for the evolution of aposematic coloration in solitary nemerteans (phylum—Nemertea). *Oikos* **48**, 289–296.
- Sword, G. A. 1999 Density-dependent warning coloration. *Nature* **397**, 217–217.
- Sword, G. A. 2002 A role for phenotypic plasticity in the evolution of aposematism. *Proc. R. Soc. B* **269**, 1639–1644. (doi:10.1098/rspb.2002.2060)
- Thomas, R. J., Marples, N. M., Cuthill, I. C., Takahashi, M. & Gibson, E. A. 2003 Dietary conservatism may facilitate the initial evolution of aposematism. *Oikos* **101**, 458–466.
- Turner, J. R. G. 1984 The palatability spectrum and its consequences. In *The biology of butterflies* (ed. by R. I. Vane-Wright & P. Ackery), pp. 141–161. Princeton University Press.

- Wallace, A. R. 1867 *Proceedings of the Entomological Society of London*, 4 March, pp. IXXX–IXXXi.
- Wallace, A. R. 1889 *Darwinism—an exposition of the theory of natural selection with some of its applications*. London: Macmillan.
- Weller, S. J., Jacobson, N. L. & Conner, W. E. 1999 The evolution of chemical defences and mating systems in tiger moths (Lepidoptera: Arctiidae). *Biol. J. Linn. Soc.* **68**, 557–578.
- Whitman, D. W., Blum, M. S. & Alsop, D. W. 1990 Allomones: chemicals for defense. In *Insect defenses: adaptive mechanisms and strategies of prey and predators* (ed. D. L. Evans & J. O. Schmidt), pp. 229–251. New York: State University of New York Press.
- Wiklund, C. & Järvi, T. 1982 Survival of distasteful insects after being attacked by naive birds—a reappraisal of the theory of aposematic coloration evolving through individual selection. *Evolution* **36**, 998–1002.
- Wiklund, C. & Sillén-Tullberg, B. 1985 Why distasteful butterflies have aposematic larvae and adults, but cryptic pupae—evidence from predation experiments on the monarch and the European swallowtail. *Evolution* **39**, 1155–1158.
- Yachi, S. & Higashi, M. 1998 The evolution of warning signals. *Nature* **394**, 882–884.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.