Why aren’t warning signals everywhere? On the prevalence of aposematism and mimicry in communities

David W. Kikuchi1,2*, Marie E. Herberstein1,3, Michael Barfield4, Robert D. Holt4 and Johanna Mappes1,5,6

1Wissenschaftskolleg zu Berlin, Wallotstraße 19, Berlin, Germany
2Evolutionary Biology, Universität Bielefeld, Konsequez 45, Bielefeld, 33615, Germany
3Department of Biological Sciences, Macquarie University, North Ryde, New South Wales, 2109, Australia
4Department of Biology, University of Florida, Gainesville, FL, 32611-8525, U.S.A.
5Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Helsinki University, Helsinki, Finland
6Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, FI-40014, Finland

ABSTRACT

Warning signals are a striking example of natural selection present in almost every ecological community – from Nordic meadows to tropical rainforests, defended prey species and their mimics ward off potential predators before they attack. Yet despite the wide distribution of warning signals, they are relatively scarce as a proportion of the total prey available, and more so in some biomes than others. Classically, warning signals are thought to be governed by positive density-dependent selection, i.e. they succeed better when they are more common. Therefore, after surmounting this initial barrier to their evolution, it is puzzling that they remain uncommon on the scale of the community. Here, we explore factors likely to determine the prevalence of warning signals in prey assemblages. These factors include the nature of prey defences and any constraints upon them, the behavioural interactions of predators with different prey defences, the numerical responses of predators governed by movement and reproduction, the diversity and abundance of undefended alternative prey and Batesian mimics in the community, and variability in other ecological circumstances. We also discuss the macroevolution of warning signals. Our review finds that we have a basic understanding of how many species in some taxonomic groups have warning signals, but very little information on the interrelationships among population abundances across prey communities, the diversity of signal phenotypes, and prey defences. We also have detailed knowledge of how a few generalist predator species forage in artificial laboratory environments, but we know much less about how predators forage in complex natural communities with variable prey defences. We describe how empirical work to address each of these knowledge gaps can test specific hypotheses for why warning signals exhibit their particular patterns of distribution. This will help us to understand how behavioural interactions shape ecological communities.

Key words: community ecology, predator–prey interactions, ecological niche, Batesian mimicry, Müllerian mimicry, aposematism

CONTENTS

I. Introduction ................................................................. 2
II. Prevalence of warning signals .............................................. 4
III. Mechanistic underpinnings of prey defence .......................... 4
IV. Predator foraging behaviour in the face of prey defences ............ 5
I. INTRODUCTION

Despite over 150 years of investigation into the evolutionary biology of warning signals (Bates, 1862; Wallace, 1867; Müller, 1879), we know surprisingly little about their role in ecological communities, or how, in turn, the characteristics of an ecological community determine the prevalence and diversity of warning signals. Laboratory studies find that once a warning signal is established, it is often more successful than other methods of avoiding attack, such as cryptis (e.g. Alatalo & Mappes, 1996; Lindström, 1999). Yet natural communities are not overrun with conspicuous prey. Among amphibians, approximately 5% of species display obvious warning signals (Arbuckle & Speed, 2015). Larvae of Lepidoptera species exhibit warning colouration at similarly low levels in Finland (~5%; Mappes et al., 2014). There is also variability in the prevalence of warning signals among different biomes. For example, only 8% of western Palearctic snake species are hypothesized to mimic venomous snakes (de Solan et al., 2020), yet estimates among New World snakes suggest that ~26% of species may participate in coral snake mimicry (Savage & Slowinski, 1992). Neither experiments nor theory can yet explain why the success of warning signals might differ among communities. Most models of warning signals consider only a subset of species within a community (reviewed in Ruxton, Sherratt & Speed, 2004; Ruxton et al., 2018). More general models of foraging predict no upper limit to the proportion of individuals (across all prey species) that can participate in warning signals (e.g. Holling, 1965). Generally, the discordance between laboratory experiments, theory, and natural history observations raises the question: what limits the prevalence of warning signals in a community?

Here, we discuss several factors that could promote or limit the success of warning signals within diverse natural communities (summarized in Fig. 1). These include the availability and costs of chemical or other defences, characteristics of individual predator and prey species, the structure of predator and prey assemblages, ecosystem productivity, deceptive mimics, and interspecific competition. Our synthesis of these concepts identifies research gaps and opportunities to integrate warning signals into ecological processes.

Warning signals, in the broadest sense, refer to prey signals that deter predators by indicating qualities that suggest prey might be difficult or dangerous to capture, subdue, or consume (Jamie, 2017; Ruxton et al., 2018). Our review focuses primarily on terrestrial animals, but we would be remiss not to mention the importance of warning signals and mimicry in aquatic and marine systems. Many larval fishes may depend upon mimicry for survival, some resembling unprofitable zooplankton (Greer et al., 2016). Nudibranchs are another exciting emerging system for the study of warning signals and mimicry in the marine realm (Edmunds, 1991; Paduła et al., 2016; Winters et al., 2017, 2018). Additionally, in the context of this review, we consider mostly visual warning signals and hence use the terms warning signals and warning colours interchangeably. However, many warning signals are non-visual (e.g. Ratcliffe & Nydam, 2008; Raška, Štys & Exnerová, 2018), and the questions we pose also apply to sensory modalities besides vision.

Warning signals that honestly indicate defences are called aposematic signals (Fig. 2; Wallace, 1867; Poulton, 1890; for detailed review, see Summers et al., 2015). The success of aposematism depends on its efficacy in deterring attack after prey have been detected. For this to happen, predators must either learn about the association between the aposematic signal and prey defence (Gittleman & Harvey, 1980; Lindström et al., 2001a; Exnerová et al., 2007), or evolve to avoid the signal (Smith, 1975, 1977; Caldwell & Rubinoff, 1983). The higher the abundance of an aposematic prey, the lower its per-capita risk of being attacked by a predator, in effect because the cost of predator learning (in prey mortality or injury) is shared and thus diluted among more individual prey. Therefore, aposematism is positively density dependent (Endler & Greenwood, 1988; Mallet & Joron, 1999; Lindström et al., 2001b; Mallet, 2010; Chouteau, Arias & Joron, 2016). Because aposematism is disfavoured when rare, its initial evolution is a famously difficult problem (Fisher, 1930; Mallet & Singer, 1987; Guilford, 1988; Alatalo & Mappes, 1996; Lindström et al., 2001b; Riipi et al., 2001). Yet once a warning signal has achieved high abundance, it should be well protected (Alatalo & Mappes, 1996; Rowland et al., 2010a). A striking example of positive density dependence in a warning-coloured species drives home this point. When desert locusts achieve a certain threshold of density, they transition from an undefended, cryptic morph to a chemically defended, aposematic morph (Sword et al., 2000; Bazzari et al., 2008). Aposematism aids the locusts in becoming a toxic swarm that ravages whole countrysides. In 2020, swarms reached huge numbers, threatening crops that feed millions of people (United Nations, 2020). These biblical hordes only collapse once they...
have exhausted their available food. Fortunately, such eruptions of aposematic organisms are rare.

Similarly, there is also selection for multiple defended species to share an aposematic signal – this is called Müllerian mimicry (Müller, 1879). Müllerian mimics benefit from sharing a signal because fewer individuals per species die while educating predators to avoid just a single aposematic signal instead of many different signals (Müller, 1879; Mallet & Joron, 1999; Rowland et al., 2010c). This underscores the paradox of the rarity of aposematism in communities: positive density dependence does not predict an upper limit to the number of Müllerian mimics. Müllerian mimicry has been described across the globe, and involves species from most major groups of insects (Borer et al., 2010; Wilson et al., 2012; Jiggins, 2018), amphibians (Darst & Cummings, 2006), reptiles (Greene & McDiarmid, 1981; Valkonen, Nokelainen & Mappes, 2011), fish (Alexandrou et al., 2011), molluscs (Winters et al., 2018), and even birds (Dumbacher & Fleischer, 2001). Müllerian mimics, together with other aposematic species that do not participate in Müllerian mimicry, constitute the honest senders of warning signals (Fig. 2).

Warning signals across communities

Fig 1. A heuristic summary of factors considered in this review as determinants of warning signal prevalence (numerical abundance) in ecological communities. They take the form of ecological and evolutionary feedbacks.

Fig 2. Functional groups of prey within a community with respect to warning signals. We are interested in factors that control the relative densities of warning-coloured prey (red and orange boxes) relative to alternative, non-warning prey (white box) in this simplified depiction of a community. From left to right, photographs are of *Zygaena filipendulae* (Quartl; CCA-3), *Cercropia vulnerata* (DWK), and *Pterophoridae* spp. and *Cydalima perspectalis* (Rudolf & Elisabeth Hofer).
between taxa as different as Lepidoptera, Serpentes, and Aves (e.g. Brown, 2006; Londoño, García & Sánchez Martínez, 2015). We return to the role of Batesian mimics in determining the prevalence of warning signals below (Section VI). For now, we explain what we mean by the prevalence of warning signals in a community.

II. PREVALENCE OF WARNING SIGNALS

We are primarily interested in the proportion of individuals in a prey assemblage that exhibit warning signals. Most theory on warning signals deals with individuals, rather than species, as do most empirical experiments. However, surveys of the proportion of prey that exhibit warning signals in a natural community usually use numbers of species as a proxy for individual abundances, making the tacit assumption that all prey species have equal abundances. This is clearly false, but for most taxonomic groups, we lack abundance data. As a matter of practicality then, we consider individuals and species interchangeably in most of this review, with the understanding that individuals are our underlying focus.

How the diversity of species that share particular warning signals changes among communities is an interesting question in its own right. Many of the processes that control the abundance of individuals will interact with those that affect the success of warning-signalling species. Although the majority of this review is concerned with individual abundance, we explicitly delve deeper into macroevolutionary processes in Section VIII. Ultimately, we hope that more abundance data from the field and higher taxonomic coverage in phylogenetic studies will help effect an explicit separation between individuals and species.

A final way of describing the characteristics of warning signals in communities is through their phenotypic diversity (disparity, sensu Futuyma, 2015). There has been much work on mechanisms that maintain phenotypic diversity in warning signals, as this diversity contradicts basic theories that predict convergence (reviewed in Joron & Mallet, 1998; Mallet & Joron, 1999; Briolat et al., 2019). Only some of these mechanisms are directly related to predicting the number of individuals and species that use warning signals within a community. Therefore, we do not discuss mechanisms promoting or maintaining diversity in warning signals directly, although we touch on many of them in the course of explaining the general prevalence of warning signals.

III. MECHANISTIC UNDERPINNINGS OF PREY DEFENCE

Secondary defences in at least some species are a prerequisite for warning signals in a community. Many, but not all, defences are chemical. Among those that are chemical, one can imagine that prey defences arise loosely from two sources: exogenous acquisition and endogenous manufacture. Any explanation for the rarity of warning signals has to consider the potential for constraints on both exogenous and endogenous prey defences. The former could, for instance, describe herbivorous insects that ingest and sequester chemicals from their host plants, which then become toxic or otherwise noxious to potential predators [e.g. monarch butterflies Danaus plexippus (Brower & Moffit, 1974); Glanville fritillary Melitaea cinxia butterflies (Laurentz et al., 2012)]. Endogenous defences occur when a prey species itself produces a defence [e.g. the venoms of coral snakes or ants (Roze, 1996; Nelson et al., 2006)].

The ability of a prey species to acquire defences from the environment (exogenous defences) does not depend solely on prey traits, but also on the availability of such defences in the first place. Plants employ a great diversity of defences against herbivory (Mithöfer & Boland, 2012), but some are unlikely to be exploitable by herbivores (e.g. tannins, structural defences; Feeny, 1976). Moreover, plant communities vary in the quantity and quality of chemical defences that herbivorous prey can exploit to defend themselves (the ‘phytochemical landscape’; Hunter, 2016). This phytochemical landscape is governed by evolutionary and ecological factors that operate directly and indirectly on plants and their herbivores (Strauss et al., 2002; Neilson et al., 2013; Anstett et al., 2016; Oksanen et al., 2020). These factors could constrain the scope for the evolution of warning signals. Insects in plant communities with low levels of chemical defence might not encounter any chemicals that could successfully deter predators (Brower et al., 1982; Malcolm & Brower, 1989; Bowers, 1992). Therefore, one reason aposematism could be rare in some communities is that prey may simply not have defences worth advertising. For example, the leaves of tropical plants may contain more secondary metabolites than temperate plants (Coley & Barone, 1996), as well as the tissues of plants that grow in resource-poor environments (Coley, Bryant & Chapin, 1985). One might then anticipate higher levels of aposematism among herbivorous insects in the tropics, and in resource-poor settings, provided insects had adapted to tolerate and sequester those plants’ secondary compounds. Insects that specialize on particular plant species might be more likely to evolve to utilize secondary plant compounds than would generalist herbivores that consume a wide range of plant species. So the incidence of warning signals associated with exogenous defences might be related to the patterns of trophic specialization versus generalization in herbivore guilds.

By contrast, herbivorous prey with endogenous defences might be freed from dependence on specific host plant secondary metabolites, and so might not show such associations. The distribution of defended herbivorous prey among plant groups might then be determined primarily by phylogenetic history and the costs of producing specific defences (Termonia et al., 2001), resulting in no correlation with plant defences. Furthermore, non-herbivorous prey must synthesize their own toxins (e.g. detritivorous fungi; Sherratt, Wilkinson & Bain, 2005) or find them from a non-plant source.
Prey defences; if they are provided with abundant alternative forms of prey defences.

IV. PREDATOR FORAGING BEHAVIOUR IN THE FACE OF PREY DEFENCES

Predation is the agent of selection on warning signals, and therefore the ultimate determinant of their evolutionary success. In this section, we describe variability in predator responses to warning-signalling prey. It may be driven by sensitivity to the signals themselves, or by sensitivity to different forms of prey defences.

Not all predators are the same. Some modes of predation may not involve detecting prey signals at all (e.g. the traps of ant-lions, the webs of orb-weaving spiders, the sifting of water by baleen whales). Other predators that do receive signals while searching for prey will not be affected by a particular defence if they have specialized tactics for dealing with it. These specialists may even preferentially seek out prey with conspicuous warning signals, possibly constraining the evolution of such signals (Endler & Mappes, 2004; Pekár et al., 2011). This contrasts with other predators (generalists) that feed on a variety of prey and typically choose to avoid aposematic prey because they can target alternative, undefended prey. A higher frequency of generalist predators in a community is predicted to increase the probability that conspicuous aposematic signals evolve (Endler & Mappes, 2004).

In the field, Valkonen et al. (2012) found that a high abundance of raptorial birds that can safely prey upon venomous snakes (specialists) resulted in higher attack rates on aposematic vipers, disfavouring aposematism. Therefore, under some conditions, the specialist:generalist predator ratio may indeed influence the evolutionary viability of aposematism.

Generalist predators can exhibit behavioural variability that is likely to be critical to the success of warning signals. Generalist predators have no special way of circumventing prey defences; if they are provided with abundant alternative, undefended prey, they are expected to focus their attention on those easier meals. Such behaviour might be deemed discriminating (predators that do this are called ‘specialists’ in the optimal foraging literature, but we here use the term ‘specialist’ to connote a predator that deals easily with specialized prey defences). When a generalist predator accepts a wide range of prey, including warning-signalling prey, we could say that its behaviour is undiscriminating. A predator might be discriminating when food is abundant, but then become undiscriminating when food grows scarce (Holt & Kotler, 1987). In other words, when the density of alternative prey species drops to the point that it is no longer advantageous for predators to target them exclusively, predators could shift their attention towards a broader swathe of the prey community, including prey with warning signals (Getty, 1985).

Qualities of prey defences may have other effects on predator behaviour. For example, some toxins can be metabolized more easily than others (Speed & Ruxton, 2014). Consider prey that are nutritious, but mildly toxic. Even if consuming an individual prey means that a predator ingests a toxin, it might be better to ingest some toxin than starve to death, as long as the predator can metabolize the toxin load (Skelhorn & Rowe, 2007). Theory predicts that a predator’s metabolic state (its current toxin load, rate of toxin metabolism, and nutritional reserves) will determine its willingness to attack or reject warning-signalling prey (Kokko, Mappes & Lindström, 2003; Sherratt, 2003; Sherratt, Speed & Ruxton, 2004). If this is true, the ability to metabolize toxins could erode the advantages of warning signals when predators are hungry. This prediction is supported by experimental evidence. Hungry starlings (Sturnus vulgaris) are more willing to attack prey that they know to be defended with a mild toxin (quinine) than are satiated starlings (Barnett, Bateson & Rowe, 2007), and hungry great tits (Parus major) have been shown ignore some warning signals (Sandre, Stevens & Mappes, 2010). Thus, although a mild toxin may sometimes function as a defence, it may not always work. We call such defences weak, because they are only conditionally effective.

Some defences are not so weak. Prey with chemical defences that the predator cannot metabolize may always be a net fitness loss to the predator. For instance, emetic chemicals that cause vomiting might result in no nutritional gain or even a net loss (Brower & Mofitt, 1974). Additionally, some avian predators have evolved strong, specific aversions to the colour patterns of deadly snakes (Smith, 1975, 1977; Caldwell & Rubinoff, 1983). These innate behaviours may imply that attacking venomous snakes is an assured fitness loss for generalist avian predators (although innate aversions could change with nutritional state or experience; e.g. Schuler & Hesse, 1985). Defences such as emetic chemicals or lethal venoms would be strong defences if predators are unlikely to change their behaviours towards them with changing ecological conditions. In other words, weak defences can break down when predators are hungry, but strong ones do not. Thus, depending on their defence types, some warning-signalling prey will be more vulnerable than others when predators exhaust their alternative food.

Another form of prey defence that is critical to consider is handling time (Holen & Sherratt, 2020). Some prey simply require a long time to be captured and consumed. For example, toxic prey do not always have to be eaten whole – some predators learn to avoid toxin glands (Parrott et al., 2019) or...
irritating hairs (Sherry & McDade, 1982) before consuming more palatable tissue, but this probably lengthens the handling time. Still other kinds of prey may never be profitable to handle—trying to catch fast-moving flies could reduce fitness if it results in a net energy loss. This is probably why slow-moving weevils have evolved to mimic flesh flies (Guerra, 2019) in what is now called evasive mimicry (van Someren & Jackson, 1959; Ruxton et al., 2018). All else being equal, there are two types of costliness that long handling times could produce. First, difficult-to-handle prey could still provide a net energy gain, but represent an opportunity cost if alternative prey that can be handled more quickly are sufficiently available (Holling, 1965; MacArthur & Pianka, 1966). Second, some prey may never be profitable to handle regardless of alternative prey availability—i.e. consuming that prey is always a net loss, or absolute cost, for the predator. Conceptually, weak defences and opportunity costs are similar, and may have similar consequences for the prevalence of warning signals. Predators with either kind of defences may be worth incorporating into predators’ diets occasionally, depending on handling times or metabolic state, and the availability of alternative prey (Kokko et al., 2003; Sherratt et al., 2004). However, weak defences versus opportunity costs may produce different predictions for predator behaviour in particular experimental designs—e.g. acceptance of weakly defended prey may depend on a predator’s internal state, while acceptance of prey defended by opportunity costs may depend on the availability and quality of alternative prey. On the other hand, strongly defended prey or prey defended by absolute costs are unlikely ever to become meals of last resort.

Prey populations often exhibit variability in defences (Speed et al., 2012), and predators can vary in their tolerance to toxins (Fig. 3; Speed & Ruxton, 2014). Predators that die from even a small load of toxins will suffer absolute costs and should behave accordingly, while predators that can metabolize more toxins might respond to them in a manner more consistent with weak defences. Predator learning could also be slowed by variation in toxin load if only some of the aposematic prey are unpalatable to that predator (Brower, Pough & Meck, 1970; Speed, 1993; Sherratt, 2011). Variability in toxin levels among species, and predator willingness to accept those toxins, can be critical in deciding whether the relationship between mimetic species is Müllerian or Batesian (Kokko et al., 2003; Rowland et al., 2010b). Indeed, prey with weak defences or opportunity costs might usually be aversive to predators (and hence aposematic), but still suffer some attacks. This could increase attack rates upon better-defended prey that they mimic (in other words, they could exhibit aspects of ‘quasi-Batesian’ rather than Müllerian mimicry; Speed, 1993; Speed & Turner, 1999). Results with artificial prey show that quasi-Batesian dynamics can occur under some circumstances, especially when alternative prey are scarce (Speed et al., 2000; Rowland et al., 2007, 2010b). This might happen in the wild when it becomes profitable for starving predators to attack weakly defended mimics, increasing the risk of attack for all members of a mimicry complex. In turn, this process could play a role in limiting the prevalence of warning-signalling prey.

The final aspect of prey phenotype we consider is conspicuousness (which should account for a predator’s sensory capabilities in its sensory environment; Endler, 1990). Warning signals are usually conspicuous (but see Wüster et al., 2004; Barnett & Cuthill, 2014; reviewed in Mappes, Marples & Endler, 2005). Conspicuousness may be costly because it makes prey more likely to be detected (Alatalo & Mappes, 1996; Hanlon, Forsythe & Joneschild, 1999; Lindström et al., 1999; Endler & Mappes, 2004; Speed & Ruxton, 2010). An intriguing possibility was raised by Holling (1965) who pointed out that when warning signals indicate opportunity costs, predators might not attack warning-signalling prey until alternative prey are depleted, meaning that less-palatable options are held in reserve until times are lean. The conspicuousness of warning signals could exaggerate this process, because they make prey easier to locate. This might be especially likely if there is scramble competition among predators for depletable prey in a patch (Holt & Kotler, 1987). One example comes from conspicuous white Pieris butterflies in the northwestern USA, which are mildly chemically defended and usually only incorporated into birds’ diets when nestlings are about to fledge (Srygley & Kingsolver, 1998). Peak nestling food demand may trigger the birds’ indiscriminating behaviour (Srygley & Kingsolver, 1998). Challenging ecological conditions could play a similar role (Lindstecht et al., 2011). Another example comes from Mappes et al. (2014), who found that artificial warning-coloured caterpillars were most at risk of predation during the period in the breeding season when most young birds were fledged.

---

**Fig 3.** (A) Relationships between predator fitness and toxin load could vary depending on the metabolism of the predator: the predator fitness curve could be convex (dashed) or concave (dotted). (B) If the predator fitness curve is convex, the expected toxin loads of most prey with a particular warning signal might be lethal to a predator (all shaded areas under the curve). If it is concave, only a few such prey might have a lethal toxin loads (crosshatched area). The difference in risk between convex and concave fitness curves could make a predator avoid or attack all warning-signalling prey, respectively. Individual predators might also exhibit variation in their metabolism.
Knowing how prey appearance and defence interacts with predator foraging decisions might help us unravel how prevalent warning signals should be. Currently, we have splendid laboratory experiments that have demonstrated the behavioural phenomena described in this section, and tantalizing ecological data. Gathering more quantitative data from ecological communities – and behavioural observations from predators in such settings, or more realistic laboratory proxies – is needed to test the relevance of these behaviours to warning signal prevalence.

V. NUMERICAL RESPONSES BY PREDATORS

The overall ratio of predators to prey modulates the predation pressure that prey experience, so here we consider the effect of predator numbers. Predator numbers may directly depress the abundance of warning-signalling prey, or disfavour warning signals on evolutionary timescales if predation pressure is consistent (or periodic) in the long term.

Mobile predators can change in abundance locally by choosing the areas in which they forage. For example, optimally foraging predators should aggregate in particularly valuable foraging patches, and leave patches with scant prey (Holt & Kotler, 1987). The value of the patch to the predator depends on the availability of suitable prey not just in that particular patch, but across a broader landscape of multiple patches. A patch will remain profitable for longer if predators switch from discriminating against warning-signalling prey to undiscriminating attacks. If predators do make this switch, they may subsequently depress populations of warning-signalling prey.

The temporal scale of predator aggregation may also change attack rates on prey with warning signals. Immediately after arrival, predators in a patch where there is an overabundance of prey may quickly become sated (a dilution effect; Rowland et al., 2007, 2010c). Also, having many prey present should result in few attacks on prey with warning signals because predators are free to focus their attention on high-quality alternative prey (Lindström et al., 2001b). However, the ability of predators to do so may depend upon whether they arrive from an area where prey have the same warning signals or different ones. Predators that are naïve to local warning signals can inflict a cost of education as they discover which prey are good and which are bad to eat, mitigating the benefits of prey overabundance. Still, information about which prey are palatable that spreads through predator social networks could help to educate naïve predators quickly at minimal cost to warning-signalling prey (Thorogood, Kokko & Mappes, 2018).

In practice, predators are unlikely to leave a patch instantaneously once its resources are depleted enough for departure to be optimal. Predators may not be able to rapidly assess, or respond to, changes in the quality of their current patch and alternative patches. Thus, in the medium term, before departing for better hunting grounds, aggregated predators may forage on whatever prey remains. This could be unfortunate for warning-coloured prey because they are often conspicuous and slow-moving (Pinheiro, 1996) (Section IV); hungry predators may raid the larder that is full of prey they previously ignored (i.e. switch from discriminating to undiscriminating behaviour) before moving to patches that might have more profitable prey.

In the long term, predator numerical responses will also include reproductive output. New predators may be born at a very high rate at particular times of the year due to timed reproduction (e.g. temperate regions). Timed reproduction may produce predictable increases in attack rates on warning-signalling prey, leading to cyclical losses to predators (see the example of Pieris butterflies above). It could additionally select for life-history adaptations in aposematic prey that allow them to avoid naïve predators. Mappes et al. (2014) found that in Finland, artificial caterpillars with warning signals were attacked less than cryptic caterpillars early and late in the avian breeding season, but were attacked more in the middle. In the middle of the season, more newly fledged birds were present; naïve fledglings would not recognize warning signals and also need lots of food, reducing discriminating behaviour. The timing of bird reproduction in temperate regions is driven by an influx of food sources as primary productivity increases (Hurlbert & Haskell, 2003), particularly emerging caterpillars (Perrins, 2008). This means that elevated attacks on warning-coloured caterpillars during the Finnish breeding season are probably caused by apparent competition with other, alternative caterpillar prey (Holt, 1977). As a consequence, warning-coloured caterpillars may have adjusted their life-history strategies; the middle of the season coincided with a drop in the emergence of warning-coloured caterpillars (Mappes et al., 2014). Clearly, it is better not to be conspicuous when predators do not recognize you as defended. This underscores the positive density dependence of aposematism.

In other types of communities, naïve predators may be added at a fairly constant rate (e.g. tropical rainforests). This may prevent a pulse of apparent competition from depressing the number of warning-signalling individuals following an influx of alternative prey. The relatively stable environment of the wet tropics may also select for slower predator life histories: long-lived tropical birds may have much more opportunity to learn about warning signals than do short-lived migrants that live at high latitudes (McNamara et al., 2008). This may favour a higher prevalence of warning signals in tropical communities, because predators spend a smaller fraction of their lives being naïve.

VI. PREY COMMUNITY STRUCTURE: ALTERNATIVE PREY AND BATESIAN MIMICS

Most profitable prey species in natural communities do not display obvious warning signals – i.e. they are potential alternative prey (Fig. 2). Laboratory studies show that the
diversity and relative abundance of alternative prey influence how likely predators are to generalize among different kinds of warning-signalling prey (Beatty, Beirinckx & Sherratt, 2004; Lindström et al., 2004; Ihalainen et al., 2012; Kikuchi et al., 2019). For example, in species-rich communities where abundance is relatively evenly distributed among prey types, predators are likely to generalize avoidance among broadly similar types (Beatty et al., 2004; Kikuchi et al., 2019), but when species abundances are less even, predators are more likely to specialize on alternative prey (Ihalainen et al., 2012; Kikuchi et al., 2019). Like other laboratory studies, however, these experiments deal only with individual predators foraging on a fixed population of prey. A missing link in our understanding is to translate these behaviours into theoretical expectations about how populations of predators will influence many dynamic prey populations. Better understanding predator generalization – in a community context – would allow us to make predictions about the prevalence of Batesian mimics and honest warning signallers as functions of prey community diversity.

Batesian mimics comprise an important fraction of warning-signalling species (Fig. 2). Often, authors refer to Batesian mimics as parasites (e.g. Speed & Turner, 1999; Rowland et al., 2010b; Kikuchi & Pfennig, 2013). In disease ecology, rich communities can provide abundant hosts for parasites, increasing parasite success in some circumstances (Hudson, Dobson & Lafferty, 2006). By analogy, Batesian mimics may rise to higher frequencies (as a fraction of the community) when warning signals are phenotypically diverse because Batesian mimics have more potential models. An example (albeit at a broad spatial scale) comes from the Batesian mimic butterfly Papilio dardanus, which has evolved a polymorphism mimicking various model species in different regions of Africa (Clarke & Sheppard, 1963). Without the availability of diverse models, P. dardanus would likely be geographically restricted, as is the mimetic morph of Limenitis arthemis, which has a single model (Ries & Mullen, 2008; Pfennig & Mullen, 2010). Therefore, Batesian mimics at least have the potential to augment the community-wide prevalence of warning signals.

Extending the analogy between parasites and Batesian mimics raises two questions. First, how many Batesian mimics can be sustained by a given number of honest signallers? Limits on the prevalence of Batesian mimics are thought to be governed by negative frequency dependence. The higher the frequencies of Batesian mimics in a mimicry complex, the more likely predators will be to incorporate members of the complex into their diet (Charlesworth & Charlesworth, 1975; Oaten, Pearce & Smyth, 1975; Getty, 1983; Sherratt, 2002). This means that as Batesian mimics become more prevalent, they risk becoming less successful, which may impose an upper threshold on their abundance. There has been much discussion in the literature about what this means for the abundance of Batesian mimics relative to that of their models. Many authors have stated that Batesian mimics must always be rarer than their models (e.g. Dunn, 1954; Brattstrom, 1955), although theory (Nicholson, 1927; Oaten et al., 1975; Sherratt, 2002) and empirical data (Brower & Brower, 1962; Harper & Pfennig, 2007; Davis Rabosky et al., 2016) show that if models are well defended, Batesian mimics can often become more abundant. One way or another, however, there is consensus that Batesian mimics place some upper limit on their own abundance. This is reflected by seasonal patterns of abundance in Batesian mimics. Hoverfly mimics of stinging hymenopterans in northern regions tend to emerge early in the season when naïve young birds are rare, and then again late in the season after birds have presumably been educated through encounters with models (Waldbauer, 1988). Similarly, among salamanders, mimics exhibit greater activity later in the season than their models (Brodie, 1981). These phenological patterns support the general argument that the selective advantage of mimicry depends upon predators having gained experience with the model. The unanswered question is how tightly correlated model and mimic populations are, and which factors in nature determine the ratio between them.

The second – and related – question is whether Batesian mimics actually function as parasites that negatively affect their models. If Batesian mimics act as parasites, they should increase attacks on the entire mimicry complex by making it difficult for predators to tell whether or not its members are defended (Lindström, Alatalo & Mappes, 1997). This question has been addressed mostly from two evolutionary perspectives: one assumes changes only to relative fitness, and the other assumes changes to mean absolute fitness. When only relative fitness is assumed to change, alleles in a population have different fitnesses, leading to evolutionary change without any change in total population size. Theories that make this assumption predict that models will experience selection to evolve away from their mimics in phenotype space in a so-called ‘chase-away’ process (Fisher, 1930; Nur, 1970; Gavrilets & Hastings, 1998; Holmgren & Enquist, 1999; Franks & Noble, 2004). This may sometimes result in models reverting to crypsis (Franks, Ruxton & Sherratt, 2009). The evolutionary transience of warning signals in amphibians and snakes is consistent with this prediction (Arbuckle & Speed, 2015; Davis Rabosky et al., 2016) (Section VIII). However, a direct test of predation on coral snake warning signals found no evidence for chase-away selection (Akcali, Kikuchi & Pfennig, 2018). Therefore, it is unclear whether Batesian mimics decrease the prevalence of warning signals by lowering the relative fitness of their models.

An alternative perspective is that increasing the number of mimics in a community could lower the mean absolute fitness of models, which would decrease model populations. A shift away from an ecological optimum can depress population size, rather than merely altering allele frequencies. To illustrate how this could affect warning signals, imagine a population of aposematic prey that is almost entirely avoided by predators. The aposematic prey has a slow life history, producing few offspring that take a long time to develop. Now imagine that a similar species with no defence and a faster life
Warning signals across communities

VII. THE INFLUENCE OF ECOLOGICAL CONDITIONS AND LIFE-HISTORY TRAITS

The overall strategy of aposematism may only be favoured in certain ecological conditions or with specific prey traits. Ruxton et al. (2018) list exposure to predators on the prey’s host plant (Prudic, Oliver & Sperling, 2007), diurnality (Merilaita & Tullberg, 2005) and large prey body size relative to predator sizes (Prudic et al., 2007; Hossie et al., 2015) as predictors of aposematism and mimicry in lepidopteran larvae. One might venture to guess that aposematism is generally favoured in high-risk microhabitats or niches. In support of this hypothesis, aposematic carnivores (that signal their noxious anal scent glands, and perhaps behavioural ferocity) often utilize more open habitats than other carnivores, and have chunkier bodies less suited to fleeing predators (Stankovich, Caro & Cox, 2011). Their visibility may be mitigated by nocturnality and spotted patterns that engender a degree of camouflage, however (Stankovich et al., 2011). Furthermore, in butterflies, Batesian mimicry is likely to be more advantageous to females than to males because females exhibit slow-flying behaviour while egg laying (Ohsaki, 1995). Yet a positive correlation between warning colors and visible, risky behaviours are not evident in all warning-signalling taxa or at all life stages. In dendrobatid frogs the association between warning signals and habitat is unclear (Rojas, 2017). Coral snakes and their mimics are primarily fossorial, generally being hidden from sight in the leaf litter or in decaying logs (Campbell, Lamar & Brodie, 2004). Some lepidopteran larvae are cryptic from a distance, and only appear aposematic when predators are close (Rothschild, 1975; Tullberg, Merilaita & Wiklund, 2005; Barnett & Cuthill, 2014). All of this variety suggests that if warning signals are only worthwhile in some ecological contexts and not others, those contexts may be heterogeneous and conceptually discontinuous, where the predictive power of any ecological variable is conditional on several others.

Even if no single ecological axis presently appears to explain all warning signals, some correlations are at least occasionally true. The requisite chemical defences can be associated with slow life histories, as in amphibians (Arbuckle & Speed, 2015) or toxic Heliconius butterflies, which are exceptionally long-lived for Lepidoptera (Turner, 1971). We do not know if longevity is permitted by toxicity, or whether instead low-quality resources leading to slow life-styles have made toxicity and aposematism selectively advantageous. Aposematic species may also be limited to environments with available toxins if they are incapable of producing their own chemical defences (e.g. Dohler et al., 1996; Jiggins, 2018; see Section III). Generally speaking, because predators are likely to respond over both ecological and evolutionary timescales to the total abundance of prey with a particular set of warning signals, there might be a kind of competition for enemy-free space (Jeffries & Lawton, 1984). This competition could drive warning-coloured prey species to colonize or evolve within the disparate ecological niches where their signals are favoured, ultimately limiting their numbers.

It seems logical to postulate that the number of distinct aposematic signals would have an upper limit determined by environmental context, organismal constraints, and the sensory ecology of the relevant predators, and that this would then impose constraints on the diversity of Batesian mimics. Batesian mimics may also at times be limited by exploitative competition with their models (Kumazawa et al., 2006), particularly if environmentally derived metabolites that are limited in supply (such as brightly coloured carotenoids) are required for the production of warning signals used by both models and mimics (Pfenning & Kikuchi, 2012). There is a dearth of information on competitive relationships between model species and their Batesian mimics. It would be very useful to explore how much fine-scale overlap in space and time there is between models and Batesian mimics, since this could govern the likelihood of competitive interactions between them. Mullerian mimicry may also be affected by competition (Kumazawa et al., 2009). Aubier & Elias (2020) predict that either microhabitat or resource-use divergence
is required for Müllerian mimics to coexist. This prediction is supported by some empirical evidence (Elias et al., 2008; Alexandrou et al., 2011; Chazot et al., 2014). However, more studies would be useful, as non-mutually exclusive hypotheses for coexistence have not been ruled out, such as consumption by a shared predator (Holt, Grover & Tilman, 1994) or specialist predators for each Müllerian co-mimic (Grover, 1994).

VIII. BIRTH AND DEATH OF WARNING SIGNALS IN EVOLUTIONARY TIME

Community ecology and evolution intersect at the broad scale of macroevolutionary processes (Cavender-Bares et al., 2009; Hendry, 2017; Weber et al., 2017). Species diversification and trait gains and losses (macroevolutionary processes) do not completely fit the definition of ‘warning signal prevalence’ that we give above (based on individuals). However, the two questions share many conceptual similarities and may be affected by the same processes, so we discuss them here. We begin with emphasizing the microevolutionary problem of positive density-dependent selection on warning coloration as a potential barrier to its evolution. Then we expand the discussion to macroevolutionary phenomena.

The positive density dependence that makes successful warning signals even more likely to succeed may also be a substantial impediment to their origin, since rare mutants will not be recognized by predators (Fisher, 1930; Guilford, 1988; Alatalo & Mappes, 1996; Lindström et al., 2001b). Mallet & Singer (1987) propose several mechanisms by which this problem might be circumvented. These mechanisms include predator neophobia towards brightly coloured prey (Sillén-Tullberg, 1985; Marples & Kelly, 1999), ‘pre-adaptation’ of bright warning colours used for sexual signalling or other purposes, enhancement of cryptic appearances, joining an extant mimicry ring, kin aggregation, and the interplay of drift and selection in Sewall Wright’s shifting balance process (Mallet, 2010). We note that the origins of new warning signals in isolated populations may be aided by the geography of speciation. The positive density-dependent efficacy of warning signals might disfavour populations at low densities. In population ecology terms, this is a form of Allee effect. Such effects can lead to patchiness and gaps in species ranges, which could under some circumstances facilitate divergence of isolated populations into new species (Keitt, Lewis & Holt, 2001). In general terms, however, inability to access any of the above mechanisms—in addition to other factors we described in preceding sections—could all decrease the rate at which new warning signals accumulate among existing taxa. Even if selective hurdles could be surmounted, developmental or phylogenetic constraints may also play an important role in warning signal origins (Charlesworth, 1994). In some cases, warning signals may rely upon special genetic architecture, as is potentially the case among some mimetic butterflies (Kunte, 2009; Jiggins, 2018). Others signals can be evolutionarily labile, arising repeatedly within a clade when a few basic pigment elements are present (Kikuchi, Seymoure & Pfennig, 2014; Davis Rabosky et al., 2016), or via novel metabolic pathways (Ford, 1944). Thus, the rate of origination of warning signals can have different probabilities across biomes, clades, and biogeographical regions.

Once warning signals evolve in a lineage, they may continue to have an impact on macroevolution. They can do this by increasing the rate of new species formation in lineages that bear them. This appears to occur in warning-coloured amphibians, which have been found to speciate two to three times faster than cryptic taxa (Arbuckle & Speed, 2015). Extinction rates were unaffected by warning signals, but warning signals were also quickly lost, with lineages reverting to cryptic without going extinct (Arbuckle & Speed, 2015). Thus, warning signals appear to be rare among amphibians because they are evolutionarily transient. Similarly, Batesian mimicry of coral snakes (Elapidae) by colubrid snakes is rapidly gained and lost (Davis Rabosky et al., 2016), although Batesian mimicry has not been lost in Papilio butterflies (Kunte, 2009).

Why should warning signals increase speciation rates? One explanation for high speciation rates is that warning signals can cause reproductive isolation when they are also targets of sexual selection. Heliconius cyanus, which diverged from H. melpomene to join a different mimicry ring, became isolated because its coloration was also the target of sexual selection (Jiggins et al., 2001). Across the genus, the colour patterns of Heliconius appear to have helped—in some instances at least—to facilitate adaptive radiation due to a combination of sexual selection and ecological selection on aposematism (Jiggins, 2018). In other words, Heliconius warning signals may be a ‘magic trait’ (Servedio et al., 2011) that promotes their diversification, although other traits also play important roles (Jiggins, 2018). The general importance of warning signals to reproductive isolation in other groups is unclear.

Some of the factors we list in preceding sections probably also influence diversification rates and the rate of gains and losses of warning signals. Likely candidates include specialist predators that target particular warning signals, and parasitism by Batesian mimics. If warning signals are favoured in particular niches, warning signals may be dragged along by other ecological forces that drive speciation and extinction.

A final macroevolutionary consideration would be evolutionary responses by the predator community to the array of defences and signalling strategies in the prey community. One could imagine a kind of Red Queen dynamic, where aposematism is kept in check, because if predators evolve to ‘crack’ prey defences (Hanfin, Brodie & Brodie, 2000), signals of those defences lose their selective advantage and might even be selected against (i.e. generalist predators could evolve to become specialists).

IX. FUTURE DIRECTIONS

First, our ability to quantify warning signals across communities is hamstrung by a deficit in phenotypic data, and in
Warning signals across communities

particular, abundance data for warning signal phenotypes at a community scale. Describing the distribution of traits within a community is a heroic task, and often only attempted for traits such as body size (Damuth, 1981; Thi-bault et al., 2011), or structural features of plants (Wright et al., 2004), although this trend is changing (Adams, Kang & June-Wells, 2014; Kantsa et al., 2017; Dalrymple et al., 2018; Pigot et al., 2020). In an ideal world, we would like to be able to obtain data not only on the proportion of individuals in a community with warning signals, and how similar or diverse those signals are (i.e. how many mimicry complexes do they form, and how conspicuous are they), but also how those aspects change across time and space.

The second item on our wish list is to understand how populations of warning-coloured species are limited. If they are constrained by density-dependent competition, as would be likely if they are successful in deterring most predation, it may be through a combination of intraspecific competition, interspecific competition between models and Batesian or Müllerian mimics, and interspecific competition of models and mimics with alternative prey species. We would like to know especially under what conditions, if any, predation limits the populations of warning-signalling prey. We suspect that predators that learn to avoid prey with warning signals will be unlikely to limit the populations of those prey, although in learning to avoid warning signals they may exert a degree of soft selection on them. Desperately hungry generalist predators, or specialist predators of aposematic species (i.e. those against which the warning signal is a liability; Endler & Mappes, 2004) are more likely to limit populations. Alternatively, warning-signalling prey may be limited by predators that are entirely indifferent to warning signals and prey defences, such as parasitoid wasps that prey upon larval Lepidoptera (Nicholson, 1927) – natural enemies to whom the warning colours of adult Lepidoptera are likely irrelevant.

Third, we would love to know more about the distribution of defences within and among prey populations, and how this variation is related to predators’ abilities to deal with these defences. Both will influence predation on warning-signalling prey. Although there are data available on variation in prey toxicity (reviewed in Speed et al., 2012), connecting them to fitness outcomes is less common, especially when predators themselves may be variable in their capacities to overcome defences. Moreover, it would be very useful to see data on the distribution of handling times and nutritional rewards offered by different kinds of prey. Our need for information on prey defences applies not just to warning-coloured species, but to cryptic prey as well. Although we suspect that a far higher proportion of warning-coloured prey are defended compared with those that are camouflaged, no study to our knowledge has yet made this comparison.

Fourth – and crucially – we need more data on how predators interact with prey in the wild. Most of our understanding of predator behaviour towards aposematic prey comes from laboratory studies where birds are trained intensively, given rather few prey types among which to choose, and not provided the option of easily abandoning the task (as they can in the wild). Such laboratory studies have been designed in this way for obvious reasons – when birds are held captive, we have the opportunity to isolate certain aspects of their behaviour and measure them repeatedly. Yet it is not clear that results from such experiments translate directly into behavioural interactions in nature. We would like to know how often predators meet aposematic prey in the wild, when they meet, how they respond to those prey, and in particular what determines the rate of successful predation. This is critically important because realistic descriptions of how predators learn and then forget can influence our theoretical predictions of how mimicry evolves (Hueheey, 1988; Leimar, Tullberg & Mallet, 2013). Observations using camera traps may be one way to explore this question (Akcali et al., 2019), although the problem is more difficult than simply measuring snapshots of behaviour. As laboratory studies and models of associative learning make clear, the history of a predator’s encounters with prey determines its foraging decisions. In addition, empirical work to measure the fitness of warning-signalling prey in environments that vary in the abundance and frequency of alternative prey would be helpful in understanding how the phenotypic complexity of a prey assemblage interacts with conspicuous warning signals.

X. CONCLUSIONS

(1) Our aim in this review was to articulate a question of general interest regarding the phenotypic composition of prey communities. That the proportion of warning-coloured prey is so low is surprising, considering its apparent advantages over other strategies in laboratory studies, and the positive density-dependent nature of aposematism and Müllerian mimicry. After outlining many of the possible reasons that the prevalence of warning signals might be limited in nature, their lack of universal dominance appears less paradoxical, but we do not know which hypotheses are actually important.

(2) The availability of chemical plant defences in the environment is one potential determinant of the viability of warning signals for herbivores that can evolve the technology needed to sequester those plant chemicals. We need more quantitative data on the relative representation of exogenous versus endogenous defences across environmental gradients to help us understand the differences in the capacity of aposematism to evolve between different communities.

(3) The relative proportion of specialist to generalist predators may be a critical factor in limiting the success of warning signals as a defensive strategy. Furthermore, generalist predators can exhibit discriminating or undiscriminating behaviour towards warning-signalling
species depending on whether prey defences are weak or strong, relative or absolute, the availability of alternative prey, and the predator’s own metabolic state.

4. Numerical responses of predator populations – i.e. their population dynamics regulated by both movement and reproduction – could have a strong influence on whether or not warning signals are advantageous, yet we have little investigation into this topic either theoretically or experimentally in the context of the evolutionary ecology of warning coloration.

5. The role of prey community structure in the success of warning signals has been addressed by a few empirical experiments showing that the phenotypes of prey in a community affect predator decision-making, and subsequent attack rates on warning-signalling prey. Additionally, Batesian mimics may impinge upon the prevalence of aposomatic species, although empirical evidence for this is hard to find. A major gap in our knowledge on this topic is of the population dynamics of models and mimics.

6. The ecological space occupied by warning-signalling species has been characterized within some taxonomic groups, e.g. mustelids, but the relationships between ecology and warning signals do not appear to hold among taxonomic groups. Information on the relationships between warning-signalling species that is not related to their signals for satiety is lacking. Competitive interactions and mutualistic interactions unrelated to mimicry are poorly understood, particularly outside the Heliconius butterflies. Competition between Müllerian mimics is theoretically predicted to require some ecological divergence.

7. Macroevolutionary lability in warning signals may be common. Higher speciation rates may occur following the evolution of warning signals, particularly if they are also targets of sexual selection. Extinctions of warning-signalling species – and gains and losses of warning signals – occur, but are less well understood.

8. R. A. Fisher called aposematism and mimicry thereof ‘the greatest post-Darwinian application of natural selection’ (Fisher, 1930). Although warning signals are intimately tied to ecology, an understanding of such signals – key traits in defining the links of predators and prey in food webs – is incompletely merged into our understanding of the dynamic interactions within ecological communities. We hope that this review provokes readers to leverage our rich evolutionary and organismal knowledge of warning signals to interpret the structure of ecological communities, and in turn to elucidate how community interactions may modulate the evolution and prevalence of such interspecies signals. We also hope that it inspires evolutionarily oriented researchers to consider the diversity of prey defences, predator behaviours, and ecological contexts that jointly produce these remarkable adaptations.

XI. ACKNOWLEDGEMENTS

This manuscript was greatly improved by the comments of Mike Speed and an anonymous colleague. We thank the Wissenschaftskolleg zu Berlin for funding our focus group ‘Integrating mimicry into community ecology’ and providing a wonderful place to work. We would also like to thank Susanne Dobler for mentioning the Zygama–Cercropis mimicry complex used in Fig. 2. M.B. and R.D.H. thank the University of Florida Foundation for support. M.E.H. was supported by the Australian Research Council through DP190101028. D.W.K. was partially funded by the German Research Foundation (DFG) as part of the SFB TMR 212 (NC3) – Project number 316099922.

Open access funding enabled and organized by Project DEAL.

XII. REFERENCES

Warning signals across communities


