Disentangling taste and toxicity in aposematic prey

Øistein Haugsten Holen

Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066 Blindern, 0316 Oslo, Norway

Many predators quickly learn to avoid attacking aposematic prey. If the prey vary in toxicity, the predators may alternatively learn to capture and taste-sample prey carefully before ingesting or rejecting them (go-slow behaviour). An increase in prey toxicity is generally thought to decrease predation on prey populations. However, while prey with a higher toxin load are more harmful to ingest, they may also be easier to recognize and reject owing to greater distastefulness, which can facilitate a taste-sampling foraging strategy. Here, the classic diet model is used to study the separate effects of taste and toxicity on predator preferences. The taste-sampling process is modelled using signal detection theory. The model is applicable to automimicry and Batesian mimicry. It shows that when the defensive toxin is sufficiently distasteful, a mimicry complex may be less profitable to the predator and better protected against predation if the models are moderately toxic than if they are highly toxic. Moreover, taste mimicry can reduce the profitability of the mimicry complex and increase protection against predation. The results are discussed in relation to the selection pressures acting on prey defences and the evolution of mimicry.

1. Introduction

Predators face the challenge of securing a high intake of nutrients while at the same time limiting their intake of toxins. Many chemically defended prey use bright, aposematic signals to warn that they are toxic, and predators can quickly learn to avoid attacking aposematic prey [1]. However, if prey that share an aposematic signal vary substantially in toxicity, the predators may instead ‘go slow’ and capture and taste each prey carefully before deciding whether or not to reject it [2]; such taste-sampling allows the predators to preferentially consume the least toxic ones [2–6]. Variation in toxicity is found within populations of visually identical prey (i.e. automimicry, [1,3,7–12]), and between different prey species that share the same warning signal (i.e. Batesian and Müllerian mimicry, [1,13–15]). Since many prey are able to survive taste-rejection [16,17], selection should favour prey defences that increase the chance of being taste-rejected upon capture. Such defences include toxins [16], but also compounds that exploit innate taste aversions in predators or mimic the taste of toxins [18,19].

Aversions to the taste of harmful compounds can be learned or innate [3,19]. If a predator lacks innate aversion to the taste of a specific toxin, avoidance of prey items containing it may be acquired through social learning or post-ingestive feedback [3,5,19,20]. Post-ingestive feedback allows a predator to associate any ill effects that follow consumption of a novel food with a wide range of food cues, including gustatory, olfactory and visual ones [21–23]. In rats, such aversion learning can occur after a single trial, even if the delay between ingestion and the onset of ill effects is several hours [21].

Although protective mimicry and aposematism have been studied for more than a century [13,14,24], the effects of prey taste and prey toxicity on predator preferences have rarely been separated (but see [5]). In fact, terms referring to distastefulness are often used as synonyms for toxicity in the literature [18]. This is surprising given their different effects: increased model toxicity is likely to increase the cost of ingesting model prey, while increased model distastefulness...
may facilitate taste recognition and thus enable the predator to reduce its consumption of model prey, all else being equal.

Here, I present an optimality model of foraging that separates the effects of taste and toxicity on predator preferences. A signal detection theory model [25] of taste discrimination is incorporated into the optimal diet model [26,27]. Signal detection theory has been much used in experimental studies of stimulus discrimination [28], and when applied to the taste-sampling process, it allows the probability of taste-rejection to depend on the distribution of toxic and non-toxic prey in the population at large, as was found in a recent study [29]. The optimal diet model predicts which prey types should be included in a predator’s diet, and is well suited to account for the opportunity costs of attack and effects of alternative prey. This is important, since the presence of alternative prey influences a predator’s willingness to forage on mimicry complexes [1,30]. The model predicts whether the predator should avoid the mimicry complex completely or rely on a taste-sampling strategy.

2. The model

Consider a predator that encounters three kinds of prey while foraging. Prey type 1 (the mimic) is non-toxic and visually identical to prey type 2 (the model), which contains toxins. Prey type 3 (the alternative prey) is visually distinct from prey types 1 and 2. According to the classic diet model, the predator’s rate of energy intake is [26,27]:

\[
c_{1}N_{1}E_{1}h_{1} + c_{2}N_{2}E_{2}h_{2} + c_{3}N_{3}E_{3}h_{3}
\]

\[
\frac{1}{1 + c_{1}N_{1}h_{1} + c_{2}N_{2}h_{2} + c_{3}N_{3}h_{3}},
\]

(2.1)

where \(E_{i}\) denotes the expected energy intake and \(h_{i}\) the expected handling time when attacking one prey item of type \(i \in \{1,2,3\}\). The product \(c_{i}N_{i}\) is the number of prey of type \(i\) encountered by the predator in one unit of search time, with \(N_{i}\) denoting prey density and \(c_{i}\) search efficiency, while \(a_{i}\) is the probability that the predator will attack the prey item. The energetic expenditure of the predator is assumed to be the same over all parts of the foraging process [26]; thus the long-term rate of energy gain is maximized by maximizing expression (2.1). For mathematical tractability, I will simply assume that the cost of ingesting a toxic prey item can be measured in terms of the time needed for recovery before the predator can resume search (cf. [31]). The cost thus enters into the handling time of the model prey.

The attack sequence is as follows: a predator that visually detects a mimic or model will either attack it or ignore it. Since the mimics and models are perfectly mimetic (i.e. visually identical), they must be equally prone to being attacked. Let \(a_{12}\) denote the probability of attack for an item in the mimicry complex (i.e. \(a_{12} = a_{2} = a_{3}\)). If the predator attacks, some time will be spent on pursuit. If the prey is captured, the predator will taste the prey and decide whether to ingest it or reject it. If the predator decides to ingest the prey, some time will be spent from the start of ingestion until the predator is ready to resume search. I adopt the standard assumption that the time spent by the predator recognizing and deciding whether to attack a prey is negligible [26,27]. Analogously, I assume that the time spent by the predator to taste and decide whether to consume or reject a prey is negligible.

Cognitive constraints and perceptual noise cause the distributions of perceived taste to overlap for the mimics and models, making perfect taste discrimination impossible.

![Figure 1](http://rspb.royalsocietypublishing.org/)

Figure 1. The receiver operating characteristic (ROC) curve, denoted \(R(x)\). It gives the probability of ingesting a mimetic prey after tasting as a function of the probability of ingesting a model prey after tasting. Three different normal–normal equal–variance ROC curves are shown, each with a different index of discriminability, \(d\). Dashed line, \(d = 3\); solid line, \(d = 1.2\); dotted line, \(d = 0\).

According to signal detection theory, the different compromises that the predator can make when discriminating can be summarized in a ‘receiver operating characteristic’ (ROC) curve [25]. Let \(x\) denote the probability of ingesting models upon tasting, and \(R(x)\) the probability of ingesting mimics upon tasting. The function \(R(x)\) is the ROC curve (figure 1). A ROC curve starts at the point (0,0) and ends at the point (1,1), and is non-decreasing [25]. The predators gain information from the taste stimuli, and if they use this information in an optimal way—a routine assumption in foraging theory [27]—the ROC curve is also necessarily concave ([25], p. 35). I return to the shape of the ROC curve in a section below.

The taste-sampling process is incorporated into the optimal diet model by expanding the terms describing the expected energy gain and handling time from attacking mimic and model prey items. Formally, let \(E_{1} = q_{1} R(x) B_{1}\) and \(E_{2} = q_{2} x B_{2}\), where \(B_{i}\) is the expected energy from ingesting a prey item of type \(i\) and \(q_{i}\) is the probability of capture given attack. Let \(h_{1} = h_{1p} + q_{1} R(x) h_{1c}\) and \(h_{2} = h_{2p} + q_{2} x h_{2c}\), where \(h_{ij}\) is the expected time spent pursuing prey type \(i\) and \(h_{ic}\) is the expected time spent from the start of ingestion until the predator is ready to resume search. The cost of ingesting a toxic model is incorporated via \(h_{2c}\).

The rate of energy intake can now be stated as:

\[
E_{t} = \frac{c_{1}N_{1}B_{1}q_{1}R(x)h_{12} + c_{2}N_{2}B_{2}q_{2}xh_{12} + c_{3}N_{3}E_{3}h_{3}}{1 + c_{1}N_{1}h_{1} + c_{2}N_{2}h_{2} + c_{3}N_{3}h_{3}}.
\]

(2.2)

The probabilities \(a_{12}\) and \(a_{3}\) comprise the predator’s attack strategy, while the probability \(x\) comprises the predator’s taste-rejection strategy. The probabilities \(a_{12}\) and \(a_{3}\) and \(x\) can take any value on the unit interval \([0,1]\), while all other parameters in equation (2.2) are strictly positive. The optimal foraging strategy \((a_{12}, a_{3}, x)\) is found by maximizing \(F(a_{12}, a_{3}, x)\). According to the zero–one rule [27], the optimal forager will neither exhibit partial preferences for attacking the alternative prey nor for attacking the mimicry complex (i.e. \(a_{12}\) and \(a_{3}\) will not take values other than 0 or 1).

The model analysis is restricted to parameter values that are biologically interesting and relevant to the subject.
By definition, it is more profitable for the predator to ingest a mimic than a model. Throughout the paper it is therefore assumed that the energy gain per unit time is higher when ingesting mimics than models, or
\[
\frac{B_1}{h_1} > \frac{B_2}{h_2}. \tag{2.3a}
\]

In addition, the mimics are assumed sufficiently profitable that ingestion is always beneficial (given capture). In other words, a predator that attacks all prey types but only ingests alternative prey would increase its energy intake rate if it consumed a mimic. Formally, the inequality
\[
\frac{B_1}{h_1} > \frac{c_BN_1E_3}{1 + c_1N_1h_1 + c_2N_2h_2 + c_3N_3h_3} \tag{2.3b}
\]
is assumed to hold.

In the numerical explorations, mimic and model prey items are assumed equal in all respects except toxin content. Formally, let \( h_{2c} = h_1 + h_{toxin} \) where the non-negative parameter \( h_{toxin} \) reflects the time needed to recover from the effects of toxins before search can be resumed (the ‘toxin recovery time’). Moreover, let \( B_1 = B_2 \) and \( h_{1c} = h_2 \). The probability that prey type \( i \) survives taste-rejection is denoted \( q_i \).

The perceived distastefulness of mimics and models is assumed to be normally distributed with equal variance, as is consistent with studies of taste discrimination [32]. An optimal taste-discriminating predator will then use a simple acceptance threshold and consume all prey perceived as being less distasteful than the threshold and taste-reject all prey perceived as more distasteful. Each possible threshold will correspond to an interior point on the ROC curve [25]. This resulting ROC curve is called a normal–normal equal-variance ROC, and the distance between the means of the two distributions of perceived distastefulness measured in units of the standard deviation is termed the index of discriminability, conventionally denoted \( d' \) [25]. The normal–normal equal–variance ROC for a given taste discriminability \( d' \) (figure 1) is the parametric curve (with parameter \( t \)) obtained when plotting \( Z(t) \) against \( Z(t - d') \), where \( Z \) is the cumulative standard normal distribution (figure 1).

Methods for calculating the optimal foraging strategy are given in the electronic supplementary material, appendix S1, including a graphical method applicable to any ROC and a numerical procedure that can be used for the normal–normal equal–variance ROC.

3. Results

(a) The optimal diet

Illustrative examples of the optimal diet of a predator are given in figure 2, for a range of values of taste discriminability \( d' \) and recovery time \( h_{toxin} \). The predator should typically include the mimicry complex in the diet when mimics and models are discriminable on the basis of taste (i.e. \( d' \) is high) and when the toxin recovery time is low (i.e. \( h_{toxin} \) is low). As expected, the parameter range for which the mimicry complex is included in the diet decreases when the alternative prey is more profitable (i.e. \( h_3 \) is low) and when it is more abundant (i.e. \( N_3 \) is high). The higher value of \( h_3 \) chosen in figure 2a,b may be interpreted as more time being needed to pursue, subdue or consume the alternative prey.

It is possible to get some analytical insight into how the optimal diet changes with different parameters (see the electronic supplementary material, appendix S1). Such a change will widen the range of (other) parameters for which the mimicry complex is included from the diet, and reduce the range for which the alternative prey is included from the diet. In figure 2, the dark grey region will contract or disappear entirely while the light grey region will expand. When either \( N_3 \) or \( c_3 \) increases, foraging on the alternative prey will become a more profitable option, causing the mimicry complex to be excluded from the diet for a wider range of parameters. The effect of changing \( c_1 \), \( N_i \), \( c_2 \), \( N_2 \) or \( q_2 \) depends on other parameters in a more complex way (see the electronic supplementary material, appendix S1). In cases where a predator that forages only on the mimicry complex benefits from taste-rejecting models, the profitability of the mimicry complex decreases with \( c_2 \), \( N_2 \) and \( q_2 \) and increases with \( c_1 \) and \( N_1 \). If a predator that forages only on the mimicry complex benefits from consuming models, then further investigation will be necessary to determine the effect of changes in the parameters \( c_1 \), \( N_1 \), \( c_2 \), \( N_2 \) and \( q_2 \). This latter situation may arise if mimics and models have very low abundance and/or search efficiency is low.

(b) The functional relationship between toxicity and taste

A reasonable assumption is that prey containing a higher amount of a distasteful toxin will appear more distasteful to taste-consuming predators than prey containing a lower
Figure 3. The optimal diet consists of the mimicry complex and the alternative prey (white regions) or only the alternative prey (grey regions). (i) The solid lines show the hypothetical functional relationships between taste discriminability ($d'$) and toxin recovery time ($h_{\text{toxin}}$) that are used in columns (ii–iv). (ii) The energy intake rate of a predator that attacks only alternative prey (solid line), only the mimicry complex (dotted line) and both alternative prey and the mimicry complex (dashed line). (iii) The probabilities by which a taste-sampling predator that attacks all prey will ingest mimics (dashed line) and models (solid line) upon capture. (iv) The number of mimics and models that a taste-sampling predator that attacks all prey kills per unit time (dashed lines, mimics; solid lines, models; thick lines, intake rate of a predator that attacks only alternative prey (solid line), only the mimicry complex (dotted line) and both alternative prey and the mimicry complex (dashed line). (ii) The energy intake rate of a predator that attacks only alternative prey (solid line), only the mimicry complex (dotted line) and both alternative prey and the mimicry complex (dashed line). (iii) The probabilities by which a taste-sampling predator that attacks all prey will ingest mimics (dashed line) and models (solid line) upon capture. (iv) The number of mimics and models that a taste-sampling predator that attacks all prey kills per unit time (dashed lines, mimics; solid lines, models; thick lines, toxin recovery time ($h_{\text{toxin}}$)).

Figure 3 provides some illustrative examples. When the toxin level is high. Because of this, the predators spend less time recovering from toxic effects and more time searching for prey when model toxicity is high, and thus taste-sample more prey per unit time. Mortality is calculated as kill rate (kills per predator per unit time, see the electronic supplementary material, appendix S1). Mimic mortality is lowest when model toxicity is intermediate. If the probability of surviving taste-sampling and rejection is low (i.e. $s_1 = s_2 = 1/4$, intermediate model toxicity gives the lowest kill rate also for models (figure 3b). If the probability of surviving taste-sampling and rejection is high (i.e. $s_1 = s_2 = 3/4$), the models (but not the mimics) are best protected by high model toxicity (figure 3a).

A qualitatively different situation arises in figure 3b, where the curve relating taste discriminability $d'$ to recovery time $h_{\text{toxin}}$ is moderately steep. Here, intermediate model toxicity lends the strongest protection to prey because it forces the predator to exclude the mimicry complex from the diet (the region of exclusion is indicated by the grey area, figure 3b). The mimicry complex is profitable when the models are weakly toxic and thus not too harmful (i.e. recovery time is short), and when the models are highly toxic and thus easy to reject by taste. When the models are moderately toxic, however, they will neither be sufficiently safe to consume

amount of the same toxin, everything else being equal. An evolutionary change in toxin level will therefore tend to change model distastefulness (and thus taste discriminability, $d'$) and harmlessness/recovery time (i.e. $h_{\text{toxin}}$) in the same direction. It may be useful to think of (monomorphic) populations of model prey that differ only in toxin load as points lying on a monotonically increasing curve in the ($h_{\text{toxin}}, d'$)-plane. The effect of a population increase in toxin load on predator preferences (and ultimately on prey survival) will depend on the exact shape of this curve, which in turn depends on numerous factors, including the predator’s ability to detect the toxin at different concentrations and the prey tissue(s) in which the toxin is stored.

Figure 3 provides some illustrative examples. When the curve is steep, a small increase in toxin recovery time will be accompanied by a great increase in toxin detectability. In figure 3a, the predator keeps the mimicry complex in the diet at all toxin levels, since taste-rejection suffices to keep toxin consumption below acceptable levels. The mimics are least likely to be ingested upon capture when the toxin level in the model population is intermediate. By contrast, the models are least likely to be ingested upon capture when the toxin level is high. Because of this, the predators

\begin{align*}
\alpha &= \frac{1}{2} (1 + \frac{1}{2} s_1 + s_2) \quad \beta = \frac{1}{2} (1 + \frac{1}{2} s_1 + s_2) \\
\gamma &= \frac{1}{2} \quad \delta = 1
\end{align*}
Figure 4. (a) The optimal diet can include the mimicry complex only (dark grey region), the alternative prey and the mimicry complex (white region), or the alternative prey only (light grey region). The dashed line shows the toxin recovery time and the range of taste discriminability explored in b and c. (b) The energy intake rate of a predator that only attacks alternative prey (solid line), only attacks the mimicry complex (dotted line), and attacks all prey (dashed line). (c) The number of mimics and models killed per unit time by a predator (dashed lines, mimics; solid lines, models; thick lines, $s_1 = s_2 = 1/2$; thin lines, $s_1 = s_2 = 95/100$). The predator attacks only alternative prey ($c$, left line sections) attacks both alternative prey and the mimicry complex ($c$, middle line sections) and attacks only the mimicry complex ($c$, right line sections). Other parameters: as in figure 3, except $N_1 = 3$, $h_3 = 5$, and in (b,c) $h_{	ext{mimic}} = 20$.

(c) Taste mimicry
In figure 4, toxin recovery time is kept constant, while the taste discriminability $d'$ is allowed to vary. In this example, the optimal diet includes alternative prey only (when $d'$ is low), both alternative prey and the mimicry complex (when $d'$ is intermediate), and the mimicry complex only (when $d'$ is high; figure 4a,b). The shifts in diet are caused by the mimicry complex becoming more profitable as taste-discriminability increases and are characterized by discontinuous jumps in prey mortality (figure 4c). Both mimics and models obtain best protection against predation when taste discriminability is low and the mimicry complex is excluded from the diet. When the mimicry complex is included in the diet, the mimic model mortality increases with taste discriminability. As before, a higher taste discriminability can reduce model ingestion and thus the time spent recovering from toxins, freeing the predator to attack more prey per unit time. For a given predator diet, model mortality could therefore both increase or decrease with taste discriminability, depending on the probability of surviving taste-sampling and rejection (figure 4c).

(d) Alternative assumptions
The optimal diet of the predator (shown in figure 2) was briefly explored under two alternative assumptions. First, I relaxed the assumption that $B_1 = B_2$ to see what happens if the predator is unable to use the energy in the model prey, i.e. when $B_2 = 0$. Qualitatively similar results to those in figure 2 were obtained (not shown). Second, I explored an alternative cost of toxin consumption: I assumed that the metabolism and excretion of toxins were energetically costly to the predators (i.e. model consumption could have a negative effect on energy reserves, $B_2 < 0$), but that the predators did not spend any time on recovery. Again, qualitatively similar results were obtained (see the electronic supplementary material, figure S2 and appendix S1). These results show that a protective advantage of intermediate toxicity can arise also for metabolic costs (see the electronic supplementary material, appendix S1).

4. Discussion
The model presented here explores optimal strategies for predators that forage on deceptive mimicry complexes (i.e. Batesian mimicry and automimicry). The predators can either avoid the mimicry complex entirely, or ‘go slow’ (sensu [2]) and capture and taste-sample each prey before deciding whether or not to reject it. As expected, the model predicts that avoidance of the mimicry complex is the best strategy if the mimicry complex is unprofitable and the alternative prey highly profitable and/or common. These are classic predictions of the diet model that are in accordance with available empirical data [1,27,30,33].

A more surprising result is that the mimicry complex may be more unprofitable to the predator and thus better protected against predation if the models have an intermediate level of a distasteful toxin than if they have a high level. This contrasts with the long-standing consensus perspective that better-defended model populations should provide better protection from predation, everything else being equal [1,7,15,34–40]. This novel result has an intuitive explanation. Distasteful toxins affect the profitability of a mimicry complex in two different ways: first, they increase the harm associated with model consumption, which reduces profitability; second, they facilitate taste recognition. This allows the predator to reduce its probability of consuming models for a given probability of consuming mimics, and thus
increases profitability. When the latter effect outweighs the former, the profitability of the mimicry complex increases with model toxicity.

The model shows that the shape of the function relating the taste discriminability of the mimicry complex to model harmfulness plays a key role in determining whether a protective advantage of intermediate toxicity might exist. The shape depends on many factors, including the predator's ability to detect the toxin at different concentrations and its susceptibility to the toxin; in addition, it will depend on the tissue(s) in which the prey store the toxin and on whether the toxin is excreted in response to predators [9,41–45]. Since little is currently known about the shape of this function, some hypothetical examples have been explored (figure 3). The first two examples illustrate two different ways in which a protective advantage of intermediate toxicity may arise.

In the first example, taste discriminability increases quickly with model harmfulness (figure 3a). Here, the mimicry complex remains sufficiently profitable to be kept in the diet at all toxin levels. The predator commits fewer recognition errors when model toxicity (and thus distastefulness) is high than when it is intermediate. This leads to a higher probability of consuming mimics upon capture at high than at intermediate model toxicity, but a lower probability of consuming models. As a result of the reduced model consumption, the predator also spends less time recovering from toxic effects and more time searching for prey at high than at intermediate model toxin levels, which ultimately causes it to attack and taste-sample more prey per unit time. Mimics are better protected by intermediate model toxicity than by high model toxicity. The models are attacked more often but also taste-rejected more often when the model population is highly toxic. The models are therefore better protected by intermediate model toxicity when the chance of surviving taste-rejection is low (25%) and by high model toxicity when the chance of surviving taste-rejection is high (75%) (figure 3a). Empirical studies have shown a high probability of survival: one experimental study reported 100 per cent survival for aposematic bugs that were seized and dropped (n = 14) and 52 per cent for an equally defended cryptic form (n = 27) [17]; another study reported an average survival of 84 per cent for five different aposematic insect species seized and held in the beak by naive individuals of four different bird species [16].

In the second example, the protective advantage of intermediate toxicity is more dramatic (figure 3b). Here, the taste discriminability increases moderately fast with model harmfulness, and the predators benefit from avoiding mimicry complexes with intermediate model toxicity and attacking mimicry complexes with high model toxicity (figure 3b). Model populations that are intermediate toxic can be sufficiently difficult to recognize on the basis of taste and at the same time sufficiently harmful to make avoidance the best option. Although highly toxic models will be even more harmful to ingest for predators, they might also be much easier to recognize by taste, thus favouring a taste-sampling strategy in the predator. It seems likely that taste-sampling will always be of potential harm to a prey, even when the prey is ultimately rejected. Soft-bodied prey with internal toxin storage are particularly likely to be killed or seriously damaged. Even tough prey that tolerate handling well are likely to suffer smaller costs, such as minor injuries, toxin depletion or opportunity costs [46]. Both mimics and models will therefore enjoy a protective advantage of intermediate model toxicity.

In the third example, the taste discriminability of the mimicry complex increases slowly with model harmfulness (figure 3c). When taste discriminability increases sufficiently slowly with model harmfulness (or not at all), a protective advantage of intermediate toxicity is not expected. Taste-sampling will be ineffectual and the mimicry complex should be excluded from the diet whenever the toxicity of the model population exceeds a certain level. Innate aversions or learned aversions based on post-ingestive feedback are then expected to form.

A key assumption in the model is that the consumption of toxic prey items incurs a cost in terms of handling time (i.e. recovery time). In the first example (figure 3a), the protective advantage of intermediate toxicity arises through changes in the taste-sampling strategies of the predators. For the model prey, a protective advantage of intermediate model toxicity is possible only because the time costs of model consumption leads to a reduction in predator attack rates. Other types of costs associated with toxin consumption (e.g. metabolic costs) will not have the same effect. In the second example (figure 3b), intermediate toxicity provides a protective advantage to mimics and models because it causes the predator to cease attacking the mimicry complex. This mechanism works equally well with metabolic costs of toxin consumption (see the electronic supplementary material, appendix S1).

Toxic defences that are distasteful may yield benefits to the individual carriers, since they facilitate taste rejection. Toxic defences also yield benefits to the group via education of predators. Individual prey obtain these group-level benefits regardless of their own defence level, and the benefits therefore comprise a common good [10]. The model identifies the level of toxicity that gives the strongest protection to the prey when both individual benefits and group-level benefits are taken into account. In general, this level of toxicity will not be evolutionarily stable [47]: if toxins are costly [48], mutant model prey may invade that 'cheat' and reap some of the group-level advantages of defence (e.g. being avoided by predators) without incurring as heavy in defences as the others [7,49]. By contrast, if defences are cheap, mutant model prey with toxin loads (and levels of distastefulness) above the level that is optimal for the population may invade. Such mutants are also 'cheaters', since their own increase in the probability of being taste-rejected comes at a cost to the population level of protection (by increasing the taste-discriminability and thus the profitability of the mimicry complex). An investigation of evolutionarily stable levels of defence is outside the scope of this paper; it would require a different model framework in which costs of defence production were taken into account.

Natural selection acting on toxic prey may favour various traits that increase the probability of taste recognition, such as secretion of toxins upon attack [9,41,42], storage of the toxins in outer tissues (e.g. wings, [43,44]) and production of non-toxic compounds with distinct taste. Undeferred prey may be under selection to mimic the taste of model prey to reduce their chances of being recognized. The compounds responsible for the resemblance in taste need not be toxic themselves: in fact, it may be cheaper to store distasteful non-toxins within the body than distasteful toxins, since the latter may require costly alterations in prey physiology in order to avoid autotoxicity [48]. Mimics that contain compounds with a distinct taste that are not shared by the models may be under selection to suppress the taste, either...
by ceasing production of the compound or by storing it internally. The result is a coevolutionary race in taste characteristics that in many ways is analogous to a (visual) mimicry chase. Although taste mimicry will tend to make the mimicry complex less profitable to a predator and may lend protection to both mimics and models (figure 4), it will typically be favoured only in mimics.

The taste-sampling process might ultimately affect the evolution of warning signals. It is commonly held that the better defended the model prey are, the less visually accurate the Müllerian mimics need to be to obtain significant protection [15,36,38]. Likewise, better defended models are thought to be able to protect a higher number of Batesian mimics [38,39,50]. This presupposes, however, that the best defended models create the greatest aversions in the predators. If we take ‘best defended’ to mean the most toxic models, this is not necessarily the case; intermediate toxic models may potentially lend more protection and therefore support a higher number of Batesian mimics.

The model also has ramifications for Müllerian mimicry. According to classical theory, Müllerian mimicry is a mutualistic interaction with two or more defended prey species benefiting from sharing the same warning signal and thus the costs of predator education [1,15,37,39]. If a pair of Müllerian mimics are equally abundant, the better-defended one will provide most (but not all) of the protection [1,37]. By contrast, the theory of quasi-Batesian mimicry predicts that a better-defended model sometimes may suffer from the presence of a less defended co-mimic [1,51]; this has some empirical support [52]. Neither of these theories takes the taste-sampling process into account. The results obtained here suggest that the presence or the absence of palatable mimics could determine which species acts as the model in a Müllerian pair. In their presence, a highly toxic model can receive more protection from a moderately toxic Müllerian co-mimic than from a highly toxic co-mimic, for exactly the same reasons discussed above: a moderately toxic co-mimic may render taste-sampling risky and unprofitable, and force the predators to exclude the mimicry complex from their diet. The less defended co-mimic may thus contribute to all of the protection enjoyed by a Müllerian pair. Perhaps this could partly explain observed cases of Müllerian mimicry where a less toxic species has acted as the model and a more toxic species has adverged towards it [37].

The diet model identifies the foraging strategy that maximizes the long-term average energy intake of the predator [27]. The results therefore concern the asymptotic foraging behaviour of the predators. In early stages of learning (before post-ingestive feedback has shaped aversions), predators may instead be more likely to show stronger avoidance to more distasteful prey. Unless the proportion of inexperienced predators in the population is high, the behaviour of the average predator may nevertheless be close to that predicted by the model. In addition, the potential protective advantage of intermediate toxicity applies only when predators have an incentive for taste-sampling; aposematic prey that are not subject to automimicry or Müllerian mimicry will typically be better protected against predation the more toxic they are. The diet model also predicts that appearance types should either be fully excluded or fully included in the diet (the ‘zero—one rule’, [27]); the effect of this is evident in the sharp diet boundaries in figures 2 and 4. Such sharp boundaries will not be observed in nature, where individual variation will smooth out the transitions (discussed in [27, pp. 20–21]).

There is a long tradition for modelling the perception of warning signals in a psychologically realistic way, using either gradient-interaction theory [53–55] or signal-detection theory [38,56,57], both of which are frameworks firmly based in experimental psychology [28,58]. By contrast, prey defences are typically modelled in an implicit fashion, with higher toxin levels assumed to give better protection to prey and create stronger aversions for predators (but see [59]). It is high time to also bring psychological realism into models of the taste-sampling process. The model presented here is a step in that direction.

In conclusion, I have shown that an increase in toxicity in a model population may increase mortality for both mimics and models. Taste mimicry may reduce prey mortality, since it makes taste-sampling difficult. The taste-sampling process clearly deserves more attention in the study of protective coloration.

I thank Ø. Langangen, J. Mappes, C. Rowe, G. D. Ruxton and T. O. Svennungsen for discussion and comments on the manuscript. I also thank the two reviewers for their helpful comments. Funding was provided by CEES (Centre for Ecological and Evolutionary Synthesis), University of Oslo.

References
