Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The evolutionary stability of automimicry

Thomas Owens Svennungsen and Øistein Haugsten Holen

Proc. R. Soc. B 2007 **274**, 2055-2063
doi: 10.1098/rspb.2007.0456

Supplementary data "Data Supplement"
http://rspb.royalsocietypublishing.org/content/suppl/2009/03/13/274.1621.2055.DC1.html

References **This article cites 23 articles, 4 of which can be accessed free**
http://rspb.royalsocietypublishing.org/content/274/1621/2055.full.html#ref-list-1

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Proc. R. Soc. B* go to: **http://rspb.royalsocietypublishing.org/subscriptions**

This journal is © 2007 The Royal Society

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2007) **274**, 2055–2062
doi:10.1098/rspb.2007.0456
Published online 13 June 2007

The evolutionary stability of automimicry

Thomas Owens Svennungsen* **and Øistein Haugsten Holen**

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

Internal defences such as toxins cannot be detected from a distance by a predator, and are likely to be costly to produce and maintain. Populations of well-defended prey may therefore be vulnerable to invasion from rare 'cheater' mutants that do not produce the toxin themselves but obtain some protection from their resemblance to their better defended conspecifics (automimicry). Although it is well established that well-defended and weakly defended morphs may coexist stably in protected dimorphisms, recent theoretical work suggests that such dimorphisms would not be resistant to invasion by novel mutants with defence levels intermediate to those present. Given that most defences (including toxins) are likely to be continuous traits, this implies that automimicry may tend to be a transitory phenomenon, and thus less likely to explain variation in defence levels in nature. In contrast to this, we show that automimicry can also be evolutionarily stable for continuous traits, and that it may evolve under a wide range of conditions. A recently developed geometric method allows us to determine directly from a trade-off curve whether an evolutionarily stable defence dimorphism is at all possible, and to make some qualitative inferences about the ecological conditions that may favour it.

Keywords: automimicry; predation; alternative strategies; secondary defences; evolutionary game theory; polymorphism

1. INTRODUCTION

Many prey species carry defences such as toxins and stings, which are harmful to predators, but which cannot be detected by predators before the prey is captured. Since such defences are effective only after a predator has decided to attack the prey, they are referred to as secondary defences (Edmunds 1974). Predators that attack defended prey may however learn to avoid similar-looking prey by associating the prey's visual characteristics (the warning signal) with its defence. Because the warning signal is typically structurally unrelated to the secondary defence, it is not necessarily an honest indicator of a prey's true defence level. Therefore, secondary defences are readily exploitable by undefended 'cheats' that carry only the deceptive warning signal. Such exploitation is illustrated by all the reported cases of Batesian mimicry, where one or several undefended species benefit from mimicking the appearance (to some degree) of a defended model species. Another intriguing possibility, which is the focus of this article, is that of automimicry (after Brower *et al.* 1967), in which some fraction of the individuals in a defended species produces little or nothing of the potentially costly defence, but nevertheless obtain some protection from predation because they are identical in appearance to their defended conspecifics.

The production and storage of a secondary defence would in many cases be likely to incur some cost to the individual, and there is indeed empirical support for such costs (Ruxton *et al.* 2004). When a predator's decision of whether or not to attack an individual prey item is based on an assessment of its appearance, and not on direct inspection of its defence, it is not obvious how a costly defence could be maintained by selection. One possibility is a direct individual benefit of the defence: predators that catch a prey often inspect it more closely (e.g. by careful tasting) before devouring it, and if the prey is found undesirable (e.g. if toxins are detected) they may reject it (Gamberale-Stille & Guilford 2004; Skelhorn & Rowe 2006), often unharmed (Wiklund & Järvi 1982 and references therein; Sillén-Tullberg 1985; Lindquist & Hay 1996). Thus, carriers of the defence may have a higher probability of surviving capture than non-carriers. This direct benefit probably plays a key role in the evolution of secondary defences (Leimar *et al.* 1986; Guilford 1994).

Over the years, many studies have documented the presence of intraspecific variation in defence level (reviewed in Ruxton *et al.* 2004), and automimicry has been suggested as a stabilizing factor of such variation (Brower *et al.* 1967; Bowers 1992; Moranz & Brower 1998). Answers to how and under which conditions automimicry may evolve and persist should be a prerequisite for such considerations. Several authors have discussed how an undefended and a defended prey morph may be maintained by natural selection in a stable protected dimorphism (Brower *et al.* 1970; Pough *et al.* 1973; Gibson 1984; Guilford 1988; Augner *et al.* 1991; Till-Bottraud & Gouyon 1992; Ruxton & Speed 2006). More recently, however, the evolutionary stability of automimicry has been questioned in two theoretical papers (Broom *et al.* 2005; Speed *et al.* 2006). These works conclude that such defence strategy dimorphisms are unlikely to be resistant to invasion by novel mutants when the defence level is a continuous trait; instead, a

* Author for correspondence (t.o.svennungsen@bio.uio.no).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2007.0456 or via http://www.journals.royalsoc.ac.uk.

Received 30 March 2007
Accepted 23 May 2007

This journal is © 2007 The Royal Society

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

monomorphic prey population fixed for an intermediate level of defence will be the outcome (Broom *et al.* 2005; Speed *et al.* 2006). Since many defences, such as toxin content, are unarguably continuous in nature, these results have seemingly removed much of the grounds for invoking automimicry as an explanation for observed cases of large intraspecific variation in defence level in natural populations.

In this paper, we use tools from game theory and adaptive dynamics to investigate the evolution of secondary defences, and show that automimicry may be evolutionarily stable also when defence traits are continuous. We conclude that automimicry has a solid theoretical foundation, and recommend that automimicry should not be discarded as a potential explanation for observed variation in secondary defences.

2. THE MODEL

We build on the model of Broom *et al.* (2005) and consider a prey species where individuals have to survive a fixed period of time before reaching the reproductive stage. Individuals are defined by their defence level D, a continuous, non-negative variable, and are otherwise identical. Individual fitness is defined as the product of fecundity and the probability of surviving until reproduction, both functions of D. Any difference in fitness between individuals in the species is thus attributed to their different levels of defence.

We start with fecundity. Let $F(D)$ denote fecundity as a function of defence level. Our assumption is simply that investment in defence has a cost that reduces fecundity, i.e. $F'(D)<0$ until F possibly reaches zero. The highest possible reproductive output is obtained by prey that do not invest resources in defence (provided they survive predation), and we express the fecundity of defended individuals as a proportion of this maximum. Thus, we standardize and let $F(0)=1$.

The probability of surviving until reproduction is first of all dependent on the number of attacks the prey are subject to. We assume that the prey are foraged upon by a community of generalist predators, whose density is unaffected by our focal prey species. The predators search randomly and independently for prey, and the number of predator attacks that an individual must survive in order to reproduce is drawn from a Poisson distribution with mean A. We will refer to A as the 'attack rate' and will return to it shortly. Let $S(D)$ be the probability of surviving an attack. We assume that the probability that an individual survives an attack increases with its defence level (a better defended individual is more likely to be rejected after capture), i.e. $S'(D)>0$ until S possibly reaches 1. If the probability of surviving one attack is $S(D)$, the probability of surviving, say, n attacks is $S(D)^n$. Consider an individual with defence level D in a population experiencing attack rate A. From our assumptions of a Poisson distribution, it follows that the probability of surviving all attacks and reaching reproduction is $S_R(D,A)=\sum_{n=0}^{\infty}[S(D)^n e^{-A}A^n/n!]$. Using the relationship $\sum_{n=0}^{\infty}[(S(D)A)^n/n!]=e^{S(D)A}$, we may simplify and write $S_R(D,A)=e^{-A(1-S(D))}$.

A reasonable assumption is that the attack rate will be negatively related to the typical level of defence in the population, i.e. poorly defended populations receive more attacks than well-defended populations (cf. Broom *et al.* 2005). This may for instance happen because predators more quickly learn to avoid well-defended populations (e.g. Ruxton *et al.* 2004), or because a well-defended prey type may be less profitable and thus incorporated in the optimal diet of a smaller subset of predators than a weakly defended one (e.g. Stephens & Krebs 1986). Let $A(D)$ denote the attack rate experienced by a population that is monomorphic for the defence level D; we assume that $A'(D)<0$.

We define fitness as the product of the probability of surviving until reproduction (S_R) and fecundity (F), and assume that the prey population is large, so that a rare mutant strategy will have a negligible effect on the attack rate experienced by the resident population. Using the above results, the fitness of a rare mutant strategy D in a population monomorphic for strategy $\hat{D}$ is

$$W(D,A(\hat{D}))=F(D)S_R(D,A(\hat{D}))=F(D)e^{-A(\hat{D})(1-S(D))}. \tag{2.1}$$

Note that the only way in which different strategies can affect each other's fitness is through their influence on the attack rate. We assume that density-dependent mortality affects all prey types equally (and independently of their individual defence strategies), so that density dependence may be represented by a function $g(N,\hat{D})$, which decreases with population size, N. A complete description of the fitness of a rare mutant will thus be $U(D,A(\hat{D}),N)=W(D,A(\hat{D}))g(N,\hat{D})$, while that of the resident strategy is $U(\hat{D},A(\hat{D}),N)$. We assume equilibrium population dynamics, i.e. $U(\hat{D},A(\hat{D}),N)=1$ (provided $W(\hat{D},A(\hat{D}))>0$). However, since the density-dependent effects are the same for resident and mutants, the success of an invading mutant can be evaluated from W only. Density-dependent factors are therefore omitted in the remaining of this paper.

3. COEXISTENCE OF TWO STRATEGIES

A prerequisite for automimicry to be evolutionarily stable is that pairs of defence strategies can coexist at some stable equilibrium frequency. The entire set of such pairs will therefore define the region of the strategy space relevant for automimicry. At equilibrium frequency, both defence strategies obtain equal fitness. The less defended strategy will suffer higher mortality in the case of an attack than the better defended, but this effect will be counteracted owing to the larger number of offspring it will produce in the case where it survives until reproduction.

Consider two levels of defence, D_1 and D_2, which by convention satisfy $D_1<D_2$. A population of D_2 individuals will elicit fewer attacks than a population of D_1 individuals, i.e. we have $A(D_2)<A(D_1)$. Let $A(D_1, D_2, \rho)$ denote the attack rate in a dimorphic population in which a fraction ρ of the individuals has defence level D_1, and a fraction $(1-\rho)$ has defence level D_2. A mixed population will experience an intermediary attack rate; we assume that $A(D_1, D_2, \rho)$ changes monotonically from $A(D_2)$ to $A(D_1)$ as ρ tends from 0 to 1. It follows that $\partial/\partial\rho(A(D_1, D_2, \rho))>0$. Moreover, we assume that $A(D_1, D_2, \rho)$ is decreasing in its two first arguments.

Suppose that the two morphs D_1 and D_2 have equal fitness at some equilibrium ρ^*, i.e. $W(D_1,A(D_1,D_2,\rho^*))=W(D_2,A(D_1,D_2,\rho^*))$. Although solving for ρ^* is not possible without making specific assumptions about the

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

functional form of A, we may still make some general observations concerning the equilibrium. By equating fitnesses, we can solve for the equilibrium attack rate $A^* \equiv A(D_1, D_2, \rho^*)$, and obtain

$$A^* = \frac{\ln[F(D_1)] - \ln[F(D_2)]}{S(D_2) - S(D_1)}. \tag{3.1a}$$

For any given pair of D_1 and D_2, the right-hand side of equation (3.1*a*) is constant. Since $A(D_1, D_2, \rho)$ changes monotonically from $A(D_2)$ to $A(D_1)$ as ρ tends from 0 to 1, ρ^* must be the unique value of ρ at which the attack rate is given by (3.1*a*), and thus the unique equilibrium at which D_1 and D_2 coexist. Since ρ^* must satisfy $0<\rho^*<1$, a necessary and sufficient condition for the existence of a dimorphic equilibrium is that the right-hand side of (3.1*a*) is between $A(D_2)$ and $A(D_1)$, or

$$A(D_2) < \frac{\ln[F(D_1)] - \ln[F(D_2)]}{S(D_2) - S(D_1)} < A(D_1). \tag{3.1b}$$

Will a unique equilibrium ρ^* satisfying $0<\rho^*<1$ be stable? Because we have $\partial/\partial\rho(A(D_1, D_2, \rho))>0$, it is clear that the fitness of both D_1 and D_2 decreases with ρ. In the case of a small perturbation of ρ away from ρ^*, stability requires the fitnesses of the two morphs to be differentially altered, and in such a way that the morph that after the perturbation is present in a marginally higher fraction than at the equilibrium ρ^* has the lower fitness. This will always be the situation if the fitness function of D_1 has a steeper negative slope at ρ^* than that of D_2. Formally, the stability condition is $\partial/\partial\rho(W(D_1, A(D_1, D_2, \rho)))|_{\rho=\rho^*} < \partial/\partial\rho(W(D_2, A(D_1, D_2, \rho)))|_{\rho=\rho^*}$. This simplifies to $S(D_1) < S(D_2)$, which always holds under our assumption that per attack survival increases with defence level. This suffices to ensure stability under continuous allele frequency dynamics (as when generations are overlapping). Under discrete allele frequency dynamics, it can be shown that this suffices to ensure stability when the additional condition $\partial/\partial\rho(A(D_1, D_2, \rho)) < 8/(S(D_2) - S(D_1))$ is satisfied for all ρ (electronic supplementary material); in other words, as long as the attack rate does not change extremely quickly with ρ. In the rest of the paper, we restrict our attention to the cases in which this condition is satisfied. We may then safely assert that, conditional upon existence, the equilibrium is always stable.

If we combine this last result with the result that the equilibrium must be unique, it follows that the fitness of D_1 must be higher than that of D_2 for all $\rho<\rho*$, and lower for all $\rho>\rho*$. Consequently, a rare D_1 strategy could invade a population fixated for D_2, and vice versa; the strategies are mutually invasible (*sensu* Geritz *et al.* 1998) and comprise a protected dimorphism. The formal conditions for D_1 and D_2 to be mutually invasible are simply $W(D_1, A(D_2)) > W(D_2, A(D_2))$ and $W(D_2, A(D_1)) > W(D_1, A(D_1))$, which, when written out and combined, gives $A(D_2) < (\ln[F(D_1)] - \ln[F(D_2)])/(S(D_2) - S(D_1)) < A(D_1)$, which is identical to the condition for the existence of an equilibrium given in (3.1*b*).

Hence, without having to explicitly specify any of the functions A, F and S, we have shown that an equilibrium ρ^* at which two strategies have the same fitness will exist if and only if the two strategies are mutually invasible. Any such dimorphic equilibrium must also be stable and unique. It is straightforward to check whether two strategies belong to the set of pairs that are mutually invasible using condition (3.1*b*). The attack rate experienced by a population residing at a dimorphic equilibrium may be calculated without knowing the exact proportion of the two morphs present at the equilibrium, using equation (3.1*a*).

4. EVOLUTIONARY STABILITY OF AUTOMIMICRY

In our search for evolutionarily stable dimorphisms, we consider the fitness of rare mutant prey strategies that invade a predatory environment characterized by the equilibrium attack rate determined by the established resident prey population. We thus assume a separation of time scales, i.e. the resident prey population always reaches its equilibrium state before new variants arrive. A dimorphic population that at its equilibrium state cannot be invaded by any other strategy will constitute an evolutionarily stable dimorphism. Graphically speaking, the component strategies D_1^* and D_2^* of such a dimorphism will reside on top of two separate but equally high peaks in a fitness landscape; a landscape that changes according to the strategies considered.

A necessary (but not sufficient) condition for our model to exhibit an evolutionarily stable dimorphism is thus that the fitness function is capable of producing a landscape with at least two peaks separated by a valley. In our case, this is equivalent to requiring a fitness landscape with a minimum of two extrema in the interior (since one peak may reside on the lower boundary of the strategy space, $D=0$). Thus, a fitness function may be discarded if the equation $\partial/\partial D(W(D, A^*))=0$ does not have more than one solution for any attack rate A^*, where A^* is treated as a free (positive) constant. Simplifying and rearranging $\partial/\partial D(W(D, A^*))=0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}D}\ln(F(D)) = -A^* \frac{\mathrm{d}}{\mathrm{d}D}S(D). \tag{4.1}$$

We proceed by qualitatively evaluating (4.1) for general functional forms for fecundity, $F(D)$, and per attack survival probability, $S(D)$, that satisfy our earlier stated criteria. We restrict ourselves to three general functional forms for $F(D)$: log-concave, log-convex and log-linear, which respectively makes the left-hand side of (4.1) a decreasing, increasing and constant function of D. Similarly, we consider three general functional forms for $S(D)$: concave, convex and sigmoid, which respectively makes $\mathrm{d}/\mathrm{d}D(S(D))$ a decreasing, increasing and unimodal function of D. (These stated properties apply where $S<1$ and $F>0$, respectively; we restrict our search for evolutionarily stable dimorphisms to this region.) A combination of $F(D)$ and $S(D)$ can be discarded if the left-hand side and the right-hand side functions of (4.1) under no circumstances can intercept twice (or more). It should be easy to confirm the results summarized in figure 1, where four of the nine combinations fail to meet this test.

For a pair of mutually invasible strategies D_1 and D_2 in the interior of the strategy space to constitute an evolutionarily stable dimorphism, it is necessary that they form an evolutionarily singular coalition (Geritz *et al.* 1998), i.e.

$$\left[\frac{\partial}{\partial D} W(D, A^*(D_1, D_2))\right]_{D=D_i} = 0 \quad \text{for } i = 1, 2, \tag{4.2a}$$

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

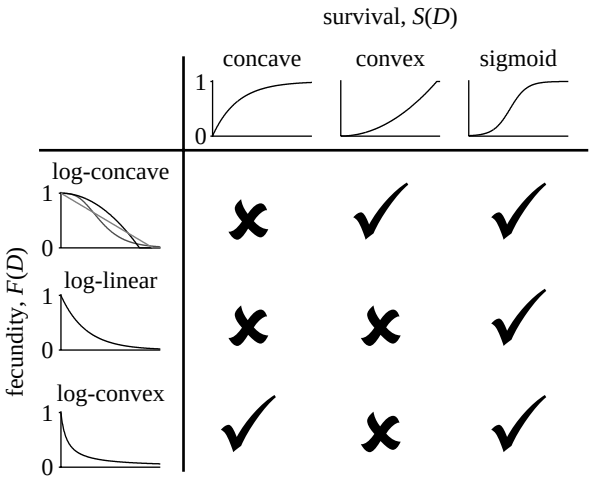

Figure 1. Four out of nine combinations of the functional classes considered cannot support evolutionarily stable dimorphic levels of defence (crosses). For the remaining five combinations (ticks), it is easy to produce examples that do.

and that the singular coalition is evolutionarily stable, i.e.

$$\left[\frac{\partial^2}{\partial D^2} W(D, A^*(D_1, D_2))\right]_{D=D_i} < 0 \quad \text{for } i = 1, 2. \tag{4.2b}$$

If D_1 is on the boundary (i.e. $D_1 = 0$), it is necessary that (4.2*a*) and (4.2*b*) is satisfied for $i = 2$, and that the following inequality holds:

$$\left[\frac{\partial}{\partial D} W(D, A^*(D_1, D_2))\right]_{D=D_1} < 0. \tag{4.2c}$$

We now proceed by constructing examples of explicit functions $F(D)$ and $S(D)$ that allow for evolutionarily stable dimorphisms.

(a) *Example 1*

We start with a combination of a sigmoid $S(D)$ and a log-linear $F(D)$, for which a particular evolutionarily stable dimorphism can be found analytically,

$$S(D) = \begin{cases} D^2(3-2D) & 0 \le D \le 1 \\ 1 & D > 1 \end{cases}$$
$$F(D) = \mathrm{e}^{-2D}.$$

Both functions are shown in figure 2*a*. It is easy to verify that for any function $A(D)$ satisfying $A(3/4) < (16/9) < A(0)$, the strategies $D_1^* = 0$ and $D_2^* = 3/4$ will be mutually invasible and coexist stably at the attack rate $A^* = 16/9$. Using criteria (4.2*a*)–(4.2*c*), it is straightforward to show that they also constitute a dimorphism that is locally evolutionarily stable. The exact evolutionary dynamics will depend on the choice of $A(D)$. The region of mutual invasibility for the choice $A(D) = \max(5 - 10D,\ 0)$ is shown in figure 2*c* (white region). The superimposed arrows indicate the permissible directions for evolutionary change reflected by the local fitness gradients of each morph. The evolutionarily stable dimorphism can be attained through small mutational steps from a monomorphic population if the population has a sufficiently high initial defence level ($D = 0.09$, marked with a cross). Evolution will then proceed towards the convergence stable branching point ($D = 0.35$, indicated by the open circle) at which any nearby mutant can invade and coexist stably with the resident strategy, and subsequently

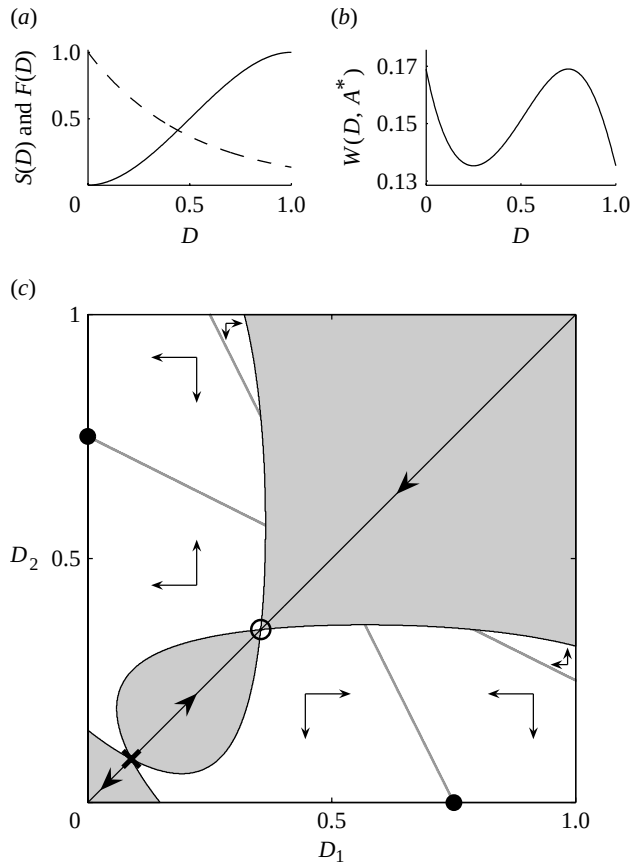

Figure 2. Example 1. (*a*) The log-linear fecundity function (broken line, $F(D)$) and the sigmoid survival function (solid line, $S(D)$). (*b*) Mutant fitness when invading the evolutionarily stable dimorphic population. The resident strategies reside on two equally high maxima; one on the boundary and the other in the interior of the strategy space. (*c*) The mutual invasibility plot with the dimorphic evolutionary endpoint marked with filled circles. Isoclines (grey lines) divide areas of mutual invasibility into separate regions with arrows indicating the permissible directions for evolutionary change within each of them. Arrows on the diagonal indicate the direction of evolution in a monomorphic population; the cross indicates a repellor and the open circle a branching point. At the endpoint, the probability that the defended and undefended morphs will survive until reproduction ($S_R(D)$) is 0.76 and 0.17, respectively. See main text for expressions and further discussion.

the population turns dimorphic. The two strategies then continue to diverge until the evolutionarily stable dimorphism is reached.

(b) *Example 2*

A combination of a sigmoid $S(D)$ and a log-concave $F(D)$ (figure 3),

$$S(D) = \begin{cases} \dfrac{3}{2}\left(D + \dfrac{3}{100}\right)^2 \left(1 - \dfrac{1}{1 + \mathrm{e}^{-25(2D-1)}}\right) & 0 \le D \le 1 \\ \quad + (1 - (D-1)^2)\left(\dfrac{1}{1 + \mathrm{e}^{-25(2D-1)}}\right) & \\ 1 & D > 1 \end{cases}$$

$$F(D) = 1 - \mathrm{e}^{-15\mathrm{e}^{-5D}}.$$

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

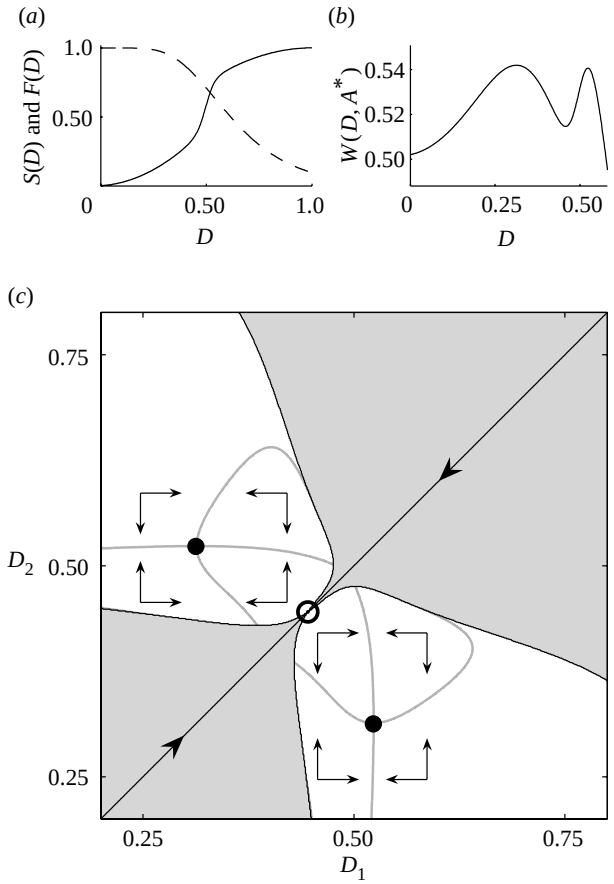

Figure 3. Example 2. (*a*) The log-concave fecundity function (broken line, $F(D)$) and sigmoid survival function (solid line, $S(D)$). (*b*) Mutant fitness when invading the evolutionarily stable dimorphic population. The resident strategies reside on two equally high maxima; both in the interior of the strategy space. (*c*) The mutual invasibility plot showing the dimorphic evolutionary endpoint (filled circles). A convergence stable branching point at $D=0.48$ (open circle) allows the dimorphic endpoint to be attainable from an initially undefended state. At the endpoint, the probability that the well-defended and less defended morphs will survive until reproduction ($S_R(D)$) is 0.81 and 0.57, respectively. See figure 2 and main text for further explanation.

In this example, both component strategies in the evolutionarily stable dimorphism are from the interior of the strategy space and invest significantly in defence. For the functional choice $A(D)=1.1-\tanh(15D-6.4)$ the dimorphism is attainable through small mutational steps from an initially undefended population. Note that we found it more difficult to find functions giving rise to evolutionarily stable dimorphism with both morphs in the interior than with one morph at the boundary, and that the functions used in this example merely reflects our desire to provide an illustrative example of the former.

For the remaining three combinations of $S(D)$ and $F(D)$ that allows for evolutionarily stable dimorphisms (figure 1), we can without difficulty find functions that give rise to evolutionarily stable dimorphisms (not shown).

5. A GRAPHICAL INTERPRETATION

Finding an evolutionarily stable dimorphism analytically is possible only for a small subset of fecundity and survival functions. Given a specific choice of $S(D)$ and $F(D)$,

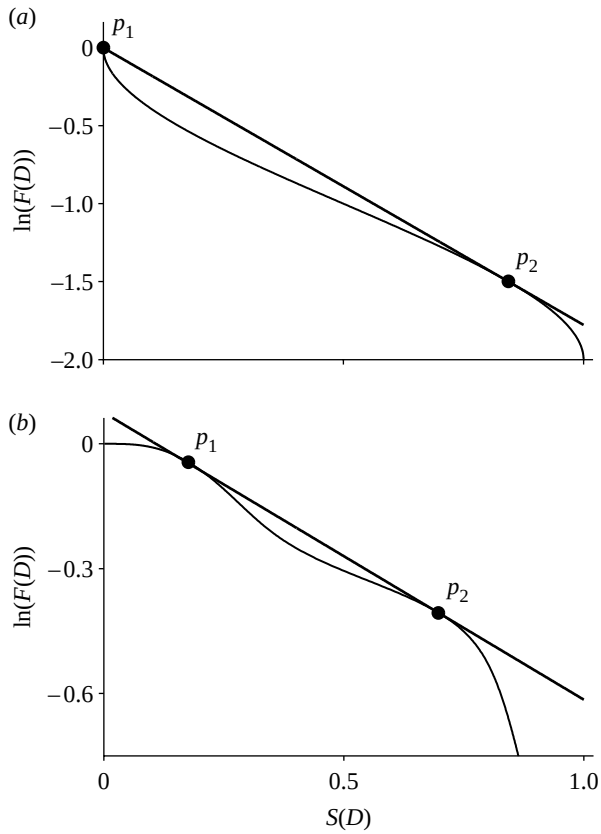

Figure 4. The trade-off curves obtained by plotting $\ln(F)$ against S from (*a*) example 1 and (*b*) example 2. The filled circles mark the position of the component strategies of the evolutionarily stable dimorphism (p_1, p_2), and the slope of the straight line drawn between them corresponds to the equilibrium attack rate (A^*) operating on them: (*a*) $A^*=1.78$, (*b*) $A^*=0.69$.

however, recently developed geometric methods (de Mazancourt & Dieckmann 2004; Kisdi 2006) make it a straightforward task to determine whether a (locally) evolutionarily stable dimorphism is possible, and if so, the component strategies that might comprise it and the functions $A(D)$ that may maintain it. We illustrate this method by looking at the specific functions used in the two previous examples. Let f denote the curve obtained by plotting $\ln(F)$ against S for all possible defence strategies D (figure 4*a*,*b*). The trade-off curve f contains all the possible combinations of our two life-history traits and, clearly, any defence level D corresponds to a unique point on f.

Suppose that we have identified a mutually invasible pair D_1 and D_2, both in the interior of the strategy space, corresponding to the points p_1 and p_2 on f. A sufficient condition for D_1 and D_2 to constitute an evolutionarily singular coalition is that a straight line connecting p_1 with p_2 is tangential to f at both points (figure 4*b*). This can be seen as follows. From (3.1*a*), it is clear that the slope of the straight line connecting p_1 and p_2 must equal $-A^*$. Using the chain rule, it is easy to show that the slope of f at p_1 and p_2 must equal $(\mathrm{d}/\mathrm{d}D(\ln(F)))/(\mathrm{d}/\mathrm{d}D(S)$ evaluated at D_1 and D_2, respectively. Equation (4.2*a*), which may be rearranged into $(\mathrm{d}/\mathrm{d}D(\ln(F)))/(\mathrm{d}/\mathrm{d}D(S))=-A^*$, must hold for both D_1 and D_2. Thus, at both p_1 and p_2, the slope of f must equal $-A^*$, which is the slope of the straight line connecting p_1 and p_2. This can only happen if the straight line connecting p_1 and p_2 is tangential to f at p_1 and p_2 (figure 4*b*).

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

The evolutionarily singular coalition will be evolutionarily stable if the trade-off curve is sufficiently concave around the component strategies that comprise it (Rueffler *et al.* 2004; Kisdi 2006). In our model, it turns out that any degree of concavity suffices, and that evolutionary stability also implies absolute convergence stability (*sensu* Leimar 2001). Using condition (4.2*a*) to simplify condition (4.2*b*), we find that a singular coalition will be evolutionarily stable if, for both component strategies, $\mathrm{d}^2/\mathrm{d}D^2(\ln(F))\mathrm{d}/\mathrm{d}D(S) - \mathrm{d}^2/\mathrm{d}D^2(S)\mathrm{d}/\mathrm{d}D(\ln(F)) < 0$. This always hold if f is concave, i.e. if $\mathrm{d}/\mathrm{d}D[(\mathrm{d}/\mathrm{d}D(\ln(F)))/(\mathrm{d}/\mathrm{d}D(S))] < 0$. Moreover, for this type of trade-off model, the following two requirements ensure absolute convergence stability (Kisdi 2006). The evolutionarily singular coalition has 'isoclinic stability' (generalization of convergence stability from the monomorphic case), i.e.

$$\left[\frac{\partial^2}{\partial D^2} W(D, A^*(D_1, D_2)) + \frac{\partial^2}{\partial D\, \partial D_i} W(D, A^*(D_1, D_2))\right]_{D=D_i} < 0 \quad \text{for } i = 1, 2, \tag{5.1a}$$

and at least one of two following cross-derivatives equals zero, i.e.

$$\left[\frac{\partial^2}{\partial D\, \partial D_j} W(D, A^*(D_1, D_2))\right]_{D=D_i} = 0 \quad \text{for } i = 1,\ j = 2 \text{ or } i = 2,\ j = 1. \tag{5.1b}$$

By using the first-order condition (4.2*a*) and equation (3.1*a*) to simplify (5.1*a*) and (5.1*b*), it can be shown that (5.1*a*) reduces to the condition for evolutionary stability, and that the two cross-derivatives in (5.1*b*) both equal zero. Thus, all evolutionarily stable dimorphisms in the interior are also absolutely convergence stable.

Suppose now instead that the mutually invasible pair has D_1 on the boundary and D_2 in the interior of the strategy space. As before, evolutionary stability requires that conditions (4.2*a*) and (4.2*b*) hold for D_2, so the straight line connecting p_1 and p_2 must be tangential to f at p_2, and f must be concave at p_2 (figure 4*a*). Since D_1 is on the boundary of the strategy space, however, equations (4.2*a*) and (4.2*b*) need not hold for D_1, and concavity at p_1 is therefore not required. Instead, it is required that the slope of f at p_1 is steeper than the slope of the straight line connecting p_1 and p_2 (as in figure 4*a*). This follows from condition (4.2*c*), which will be satisfied if $(\mathrm{d}/\mathrm{d}D(\ln(F)))/(\mathrm{d}/\mathrm{d}D(S)) < -A^*$ holds for D_1. The conditions for absolute convergence stability are simplified when D_1 is on the boundary: it then suffices that (5.1*a*) holds for D_2, and that D_1 is evolutionarily stable.

It follows from the above discussion that we can determine directly from the shape of the trade-off curve f whether it may support a (locally) evolutionarily stable dimorphism. If a straight line may be drawn that is tangential to f at two interior points, p_1 and p_2, that both lie in concave regions of f, the trade-off curve may support an evolutionarily stable dimorphism in the interior. If a straight line may be drawn that is tangential to f at some interior point p_2 at which f is concave, and that is less steep than f at the point, p_1, where they intercept at the boundary, the trade-off curve may support an evolutionarily stable dimorphism with one component strategy at the boundary and one in the interior. However, the realization of an evolutionarily stable dimorphism depends on whether the defence strategies associated with the two relevant points are mutually invasible. Mutual invasibility of two defence strategies D_1 and D_2 is ensured if the slope of the straight line connecting p_1 and p_2 is between $-A(D_2)$ and $-A(D_1)$ (this can be seen from (3.1*b*)). Thus, for any trade-off curve that may support an evolutionarily stable dimorphism, mutual invasibility for the relevant strategies may be ensured by choosing an appropriate attack function.

6. DISCUSSION

There has been, to our knowledge, only one previous attempt (Broom *et al.* 2005) at investigating whether automimicry can persist as a defence level dimorphism that is evolutionarily stable, i.e. resistant to invasion from mutants. That study concluded that such a defence level dimorphism cannot be stable when the defence trait is continuous, which seriously challenges the role of automimicry as an explanation for intraspecific variation in defence level. Like Broom *et al.* (2005), we have assumed a trade-off between an individual's ability to survive an attack by a predator and its fecundity, but by considering a range of functional forms, we have shown that their conclusion does not hold in general. A qualitative analysis has allowed us to identify combinations of functional forms that do not allow for evolutionarily stable dimorphisms (including the one used by Broom *et al.* 2005). However, we have also been able to verify that many choices of functions do indeed lead to evolutionarily stable dimorphisms, and that these functions can take very different shapes.

The shapes of the survival and fecundity functions must ultimately be determined empirically. Nevertheless, the most reasonable representative of a survival function seems to be a sigmoid one. The sigmoid form would easily arise whenever the predators' ability to detect or overcome a defence varies unimodally around some non-zero mean. Moreover, even if all predators adopted the same acceptance threshold for a toxin, perceptual noise would give rise to a sigmoid relationship between actual toxin concentration and the probability that perceived toxicity falls above this threshold (Joron 2003). A sigmoid shape is in fact a standard prediction from signal detection theory, which has been confirmed in numerous stimulus control experiments (e.g. Blough 2001). Since a sigmoid survival function can give rise to evolutionarily stable dimorphisms in combination with representatives of all the classes of fecundity functions that we have considered, the possibility of having evolutionarily stable dimorphisms seems less constrained by the shape of the fecundity function.

The relationship between defence level and fecundity, or between defence level and the probability of surviving a predator encounter, strongly depends on the nature of the secondary defence we are considering. If we want to make educated guesses about these relationships or measure them empirically, it does matter whether we talk about a toxin, a tough integument, or escape abilities. Mathematically speaking, however, what matters for the outcome of the model is only the direct relationship between fecundity and the probability of surviving an attack, which can be summarized in a trade-off curve. Using recent results from the field of adaptive dynamics, we have illustrated how such a trade-off curve allows us to directly evaluate the

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

potential existence and location of evolutionarily stable dimorphisms. The trade-off curve is easy to generate if the two functional forms relating defence level to fecundity and survival are known, and is thus a useful tool to quickly test whether the functional forms may potentially support evolutionarily stable dimorphisms under our model assumptions.

The mathematical transparency obtained when analysing the model in terms of a single trade-off curve also allows us to make some rough generalizations about the ecological circumstances under which an evolutionarily stable dimorphism supported by particular trade-off curves can be realized. As we have shown, the attack rate that is needed to maintain the two component strategies of an evolutionarily stable dimorphism is closely linked to the slope of the trade-off between survival and fecundity (figure 4). Thus, it is clear that evolutionarily stable dimorphisms can be maintained under both high- and low-predation pressure. Suppose that the predator community detects the focal prey at a low rate (e.g. due to crypsis). If the trade-off between survival and fecundity is weak (the trade-off curve is not steep), a low attack rate may suffice to maintain an evolutionarily stable dimorphism. If the trade-off is strong, however, the high attack rate necessary to maintain a protected dimorphism may not be achievable for the predator community. To take the opposite case, suppose that the focal prey is highly conspicuous and thus discovered at a high rate by predators. Several factors may then lead to a high attack rate that changes little in response to different levels of defence: a constant high influx of naive predators; predators that learn slowly; or a section of the predator community that can tolerate the defence. In such cases, a protected dimorphism may not be possible if the trade-off between fecundity and survival is weak. However, a dimorphism may arise if the trade-off is strong. We emphasize, though, that such generalizations may be compromised by large deviations from the model assumptions of random predator encounters and that predator behaviour is unaffected by the abundance of the focal prey species.

Although it has not been shown before that automimicry may persist as an evolutionarily stable dimorphism when defence traits are continuous, there is nothing unique about our model framework that makes it possible. Leimar *et al.* (1986) developed a model to investigate the evolutionary stability of aposematism that incorporated detection rate, predator learning and generalization. Although the focus in that paper was on monomorphic evolutionarily stable strategies, it can be shown that the model will readily demonstrate convergence stable branching points in the defence trait for realistic functional representations of the probability of dying when attacked. To fully model the evolution of a dimorphism within that framework would however require some model extensions.

Broom *et al.* (2005) showed that a dimorphism could be evolutionarily stable when the defence trait was restricted to discrete levels, but only when the defence trait could not take any value intermediate to those constituting the dimorphism. We have shown for continuous defence traits that dimorphisms may be resistant to invasion by any mutant strategy. It follows that dimorphisms can be evolutionarily stable in the discrete case also when the defence trait can take values intermediate to those present in the dimorphism.

The concept of automimicry was originally motivated by observations of variation in toxin content between prey due to the use of different food plants (Brower *et al.* 1967), and toxin content is perhaps the most commonly used example in the literature of a defence that may be vulnerable to invasion by automimics. Our model requires only that the defence level is under genetic control, and applies in principle to both de novo synthesis of toxins by the prey and sequestering of toxins from host plants. However, our assumption of frequency-dependent maintenance of defended and non-defended morphs at an equilibrium in which fitnesses are equal requires that the nutrients necessary for synthesizing toxins or the host plants needed for sequestering them are not limiting resources. Otherwise, different defence levels may simply reflect that some prey do not have access to the most favourable resource (i.e. a 'best of a bad job' strategy). The phenomenon of automimicry is not limited to specific modes of secondary defence, however; the essential requirement is only that the defence level is not obvious to predators upon prey discovery. Relevant defences may therefore include morphological deterrents like a thick cuticle or spines and behavioural defences like rapid escape or retaliation. Note that if the defence has many components, it is in principle possible for a prey to be an automimic for the defensive components that are most costly to produce, but not for others, as long as the different components can be produced independently. Given the prevalence and diversity of systems involving a secondary defence, we will argue that our model will have wide applicability.

We especially thank Evá Kisdi for rewarding discussions and valuable comments which have greatly improved the final paper. M. Broom, T. F. Hansen, G. D. Ruxton, M. P. Speed G.-P. Sætre and the referees provided helpful comments. We also thank R. A. Johnstone for comments and discussion on an early version of the manuscript. Financial support was provided by the University of Oslo (to T.O.S.) and the Research Council of Norway (to Ø.H.H.).

REFERENCES

Augner, M., Fagerström, T. & Tuomi, J. 1991 Competition, defense and games between plants. *Behav. Ecol. Sociobiol.* **29**, 231–234. (doi:10.1007/BF00163979)

Blough, D. S. 2001 Some contributions of signal detection theory to the analysis of stimulus control in animals. *Behav. Process.* **54**, 127–136. (doi:10.1016/S0376-6357(01)00154-1)

Bowers, M. D. 1992 The evolution of unpalatability and the cost of chemical defense in insects. In *Insect chemical ecology: an evolutionary approach* (eds B. D. Roitberg & M. B. Isman), pp. 216–244. New York, NY: Chapman and Hall.

Broom, M., Speed, M. P. & Ruxton, G. D. 2005 Evolutionarily stable investment in secondary defences. *Funct. Ecol.* **19**, 836–843. (doi:10.1111/j.1365-2435.2005.01030.x)

Brower, L. P., Brower, J. V. & Corvino, J. M. 1967 Plant poisons in a terrestrial food chain. *Proc. Natl Acad. Sci. USA* **57**, 893–898. (doi:10.1073/pnas.57.4.893)

Brower, L. P., Pough, F. H. & Meck, H. R. 1970 Theoretical investigations of automimicry. I. Single trial learning. *Proc. Natl Acad. Sci. USA* **66**, 1059–1066. (doi:10.1073/pnas.66.4.1059)

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

de Mazancourt, C. & Dieckmann, U. 2004 Trade-off geometries and frequency-dependent selection. *Am. Nat.* **164**, 765–778. (doi:10.1086/424762)

Edmunds, M. A. 1974 *Defence in animals: a survey of anti-predator defences*. Harlow, UK: Longman.

Gamberale-Stille, G. & Guilford, T. 2004 Automimicry destabilizes aposematism: predator sample-and-reject behaviour may provide a solution. *Proc. R. Soc. B* **271**, 2621–2625. (doi:10.1098/rspb.2004.2893)

Geritz, S. A. H., Kisdi, É., Meszéna, G. & Metz, J. A. J. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57. (doi:10.1023/A:1006554906681)

Gibson, D. O. 1984 How is automimicry maintained? In *The biology of butterflies* (eds R. I. Vane-Wright & P. R. Ackery), pp. 163–165. London, UK: Academic Press.

Guilford, T. 1988 The evolution of conspicuous coloration. *Am. Nat.* **131**, S7–S21. (doi:10.1086/284764)

Guilford, T. 1994 "Go-slow" signalling and the problem of automimicry. *J. Theor. Biol.* **170**, 311–316. (doi:10.1006/jtbi.1994.1192)

Joron, M. 2003 Aposematic coloration. In *Encyclopedia of insects* (eds Rs T. Cardé & V. H. Resh), pp. 39–45. New York, NY: Academic Press.

Kisdi, É. 2006 Trade-off geometries and the adaptive dynamics of two co-evolving species. *Evol. Ecol. Res.* **8**, 959–973.

Leimar, O. 2001 Evolutionary change and darwinian demons. *Selection* **2**, 65–72. (doi:10.1556/Select.2.2001.1-2.5)

Leimar, O., Enquist, M. & Sillén-Tullberg, B. 1986 Evolutionary stability of aposematic coloration and prey unprofitability: a theoretical-analysis. *Am. Nat.* **128**, 469–490. (doi:10.1086/284581)

Lindquist, N. & Hay, M. E. 1996 Palatability and chemical defense of marine invertebrate larvae. *Ecol. Monogr.* **66**, 431–450. (doi:10.2307/2963489)

Moranz, R. & Brower, L. P. 1998 Geographic and temporal variation of cardenolide-based chemical defenses of queen butterfly (*Danaus gilippus*) in northern Florida. *J. Chem. Ecol.* **24**, 905–932. (doi:10.1023/A:1022329702632)

Pough, F. H., Brower, L. P., Meck, H. R. & Kessell, S. R. 1973 Theoretical investigations of automimicry: multiple trial learning and the palatability spectrum. *Proc. Natl Acad. Sci. USA* **70**, 2261–2265. (doi:10.1073/pnas.70.8.2261)

Rueffler, C., van Dooren, T. J. M. & Metz, J. A. J. 2004 Adaptive walks on changing landscapes: Levins' approach extended. *Theor. Popul. Biol.* **65**, 165–178. (doi:10.1016/j.tpb.2003.10.001)

Ruxton, G. D. & Speed, M. 2006 How can automimicry persist when predators can preferentially consume undefended mimics? *Proc. R. Soc. B* **273**, 373–378. (doi:10.1098/rspb.2005.3238)

Ruxton, G. D., Sherratt, T. N. & Speed, M. P. 2004 *Avoiding attack: the evolutionary ecology of crypsis, warning signals, and mimicry*. Oxford, UK: Oxford University Press.

Sillén-Tullberg, B. 1985 Higher survival of an aposematic than of a cryptic form of a distasteful bug. *Oecologia* **67**, 411–415. (doi:10.1007/BF00384948)

Skelhorn, J. & Rowe, C. 2006 Avian predators taste–reject aposematic prey on the basis of their chemical defence. *Biol. Lett.* **2**, 348–350. (doi:10.1098/rsbl.2006.0483)

Speed, M. P., Ruxton, G. D. & Broom, M. 2006 Automimicry and the evolution of discrete prey defences. *Biol. J. Linn. Soc.* **87**, 393–402. (doi:10.1111/j.1095-8312.2006.00577.x)

Stephens, D. W. & Krebs, J. R. 1986 *Foraging theory*. Monographs in behavior and ecology. Princeton, NJ: Princeton University Press.

Till-Bottraud, I. & Gouyon, P.-H. 1992 Intra- versus interplant Batesian mimicry? A model on cyanogenesis and herbivory in clonal plants. *Am. Nat.* **139**, 509–520. (doi:10.1086/285341)

Wiklund, C. & Järvi, T. 1982 Survival of distasteful insects after being attacked by naive birds: a reappraisal of the theory of aposematic coloration evolving through individual selection. *Evolution* **36**, 998–1002. (doi:10.2307/2408077)

Downloaded from rspb.royalsocietypublishing.org on 7 July 2009

NOTICE OF CORRECTION

Figure 3 is now presented in the correct form.

18 June 2007