Electronic supplement. Calculation of some mathematical results for "Reciprocal mimicry: kin selection can drive defended prey to resemble their Batesian mimics" .

A1. Calculation of $\left[\partial t^*_{mix} / \partial m'_d\right]_{m'_d=\hat{m}_d}$.

The optimal attack threshold $t^*_{mix}(m'_d, \hat{m}_d, \hat{m}_m)$ is the unique root of

$f(t, m'_d) = rZ'(t - m'_d) + (1-r)Z'(t - \hat{m}_d) - KZ'(t - \hat{m}_m)$. Let t^*_{mix} be shorthand for

$t^*_{mix}(m'_d, \hat{m}_d, \hat{m}_m)$. According to the implicit function theorem, we have

$$\frac{\partial t^*_{mix}}{\partial m'_d} = \left[-\frac{\partial f / \partial m'_d}{\partial f / \partial t}\right]_{t=t^*_{mix}} = -\frac{-rZ''(t^*_{mix} - m'_d)}{rZ''(t^*_{mix} - m'_d) + (1-r)Z''(t^*_{mix} - \hat{m}_d) - KZ''(t^*_{mix} - \hat{m}_m)},$$

which using the fact that $Z''(x) = -Z'(x)x$ can be written as

$$\frac{\partial t^*_{mix}}{\partial m'_d} = -\frac{rZ'(t^*_{mix} - m'_d)(t^*_{mix} - m'_d)}{rZ'(t^*_{mix} - m'_d)(-1)(t^*_{mix} - m'_d) + (1-r)Z'(t^*_{mix} - \hat{m}_d)(-1)(t^*_{mix} - \hat{m}_d) - KZ'(t^*_{mix} - \hat{m}_m)(-1)(t^*_{mix} - \hat{m}_m)}.$$

As by definition $f(t^*_{mix}, m'_d) = 0$, we may substitute in $rZ'(t^*_{mix} - m'_d) + (1-r)Z'(t^*_{mix} - \hat{m}_d)$ for

$KZ'(t^*_{mix} - \hat{m}_m)$ and rearrange, yielding

$$\frac{\partial t^*_{mix}}{\partial m'_d} = \frac{rZ'(t^*_{mix} - m'_d)(t^*_{mix} - m'_d)}{rZ'(t^*_{mix} - m'_d)(t^*_{mix} - m'_d - t^*_{mix} + \hat{m}_m) + (1-r)Z'(t^*_{mix} - \hat{m}_d)(t^*_{mix} - \hat{m}_d - t^*_{mix} + \hat{m}_m)}$$
$$= \frac{rZ'(t^*_{mix} - m'_d)(t^*_{mix} - m'_d)}{rZ'(t^*_{mix} - m'_d)(\hat{m}_m - m'_d) + (1-r)Z'(t^*_{mix} - \hat{m}_d)(\hat{m}_m - \hat{m}_d)}.$$

We are interested in the case in which $m'_d = \hat{m}_d$, in which case t^*_{mix} is given explicitly by t^* (2) in the main text. We obtain

$$\left[\frac{\partial t^*_{mix}}{\partial m'_d}\right]_{m'_d=\hat{m}_d} = \frac{rZ'(t^* - \hat{m}_d)(t^* - \hat{m}_d)}{rZ'(t^* - \hat{m}_d)(\hat{m}_m - \hat{m}_d) + (1-r)Z'(t^* - \hat{m}_d)(\hat{m}_m - \hat{m}_d)}$$
$$= \frac{r(t^* - \hat{m}_d)}{r(\hat{m}_m - \hat{m}_d) + (1-r)(\hat{m}_m - \hat{m}_d)} = \frac{r\left(\dfrac{\hat{m}_m + \hat{m}_d}{2} - \dfrac{\ln K}{\hat{m}_m - \hat{m}_d} - \hat{m}_d\right)}{d} = r\left(\frac{1}{2} - \frac{\ln K}{d^2}\right).$$

A2. Determining the sign of the fitness gradient using attack function A_{alt}.

The rate of attack is given by:

$$A_{alt}(m, m'_d, \hat{m}_d, \hat{m}_m) = \left(1 - Z(t^*_{mix}(m'_d, \hat{m}_d, \hat{m}_m) - m)\right)\left(w_{opt}(m'_d, \hat{m}_d, \hat{m}_m)\right).$$

The two scenarios are treated one after the other.

Scenario 1.

The fitness gradient has the same sign as $-\mathrm{E}_r\left[\frac{\partial}{\partial m'_d} A_{alt}\left(m'_d, m'_d, \hat{m}_d, \hat{m}_m\right)\right]_{m_d = \hat{m}_d}$.

Multiplying this with the positive expression $1/\left((1-p)c\right)$, we obtain the sign-equivalent expression

$$-\mathrm{E}_r\left[\frac{\partial}{\partial m'_d}\left(1 - Z(t^*_{mix} - m'_d)\right)\left((1 - Z(t^*_{mix} - \hat{m}_m))K + rZ(t^*_{mix} - m'_d) + (1-r)Z(t^*_{mix} - \hat{m}_d) - 1\right)\right]_{m'_d = \hat{m}_d}.$$

Performing the differentiation, we obtain:

$$\begin{aligned}-\mathrm{E}_r\Bigg[& -Z'(t^*_{mix} - m'_d)\left(\frac{\partial t^*_{mix}}{\partial m'_d} - 1\right)\left((1 - Z(t^*_{mix} - \hat{m}_m))K + rZ(t^*_{mix} - m'_d) + (1-r)Z(t^*_{mix} - \hat{m}_d) - 1\right) \\ & + (1 - Z(t^*_{mix} - m'_d))\left(-Z'(t^*_{mix} - \hat{m}_m)K\frac{\partial t^*_{mix}}{\partial m'_d} + rZ'(t^*_{mix} - m'_d)\left(\frac{\partial t^*_{mix}}{\partial m'_d} - 1\right) + (1-r)Z'(t^*_{mix} - \hat{m}_d)\frac{\partial t^*_{mix}}{\partial m'_d}\right)\Bigg]_{m'_d = \hat{m}_d}.\end{aligned}$$

When $m'_d = \hat{m}_d$, we can (in turn) make the substitutions $t^*_{mix}(\hat{m}_d, \hat{m}_d, \hat{m}_m) = t^*(\hat{m}_d, \hat{m}_m)$ (see main text), $Z'(t^* - \hat{m}_m) = Z'(t^* - \hat{m}_d)/K$ (the latter can be checked by substituting in $t^* = \frac{\hat{m}_m + \hat{m}_d}{2} - \frac{\ln K}{\hat{m}_m - \hat{m}_d}$ and rearranging), and $\left.\frac{\partial t^*_{mix}}{\partial m'_d}\right|_{m'_d = \hat{m}_d} = r\left(\frac{1}{2} - \frac{\ln K}{d^2}\right)$ (see section A1).

Making these substitutions and simplifying we obtain

$$\mathrm{E}_r\left[Z'(t^* - m'_d)\left(\left(\frac{\partial t^*}{\partial m'_d} - 1\right)\left((1 - Z(t^* - \hat{m}_m))K + Z(t^* - \hat{m}_d) - 1\right) + r(1 - Z(t^* - \hat{m}_d))\right)\right]_{m'_d = \hat{m}_d}$$

$$= Z'(t^* - \hat{m}_d)\ \mathrm{E}_r\left[\left(r\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)\left((1 - Z(t^* - \hat{m}_m))K + Z(t^* - \hat{m}_d) - 1\right) + r(1 - Z(t^* - \hat{m}_d))\right].$$

The fitness gradient has the same sign as the expectation, and by making substitutions $t^* - \hat{m}_d = d/2 - \ln(K)/d$ and $t^* - \hat{m}_m = -d/2 - \ln(K)/d$ (see main text) and rearranging we obtain (9) in the main text:

$$\left(\bar{r}\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)\left(1 - Z\left(-\frac{d}{2} - \frac{\ln K}{d}\right)\right)K - \left(\bar{r}\left(-\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)\left(1 - Z\left(\frac{d}{2} - \frac{\ln K}{d}\right)\right).$$

Scenario 2.

The fitness gradient has the same sign as

$$-\mathrm{E}_{\rho,f}\left[\rho \frac{\partial}{\partial m'_d} A_{alt}\left(m'_d, \hat{m}_d, \hat{m}_d, \hat{m}_m\right) + \frac{\partial}{\partial m'_d} A_{alt}\left(\hat{m}_d, m'_d, \hat{m}_d, \hat{m}_m\right)\right]_{m'_d = \hat{m}_d}.$$

Multiplying this with the positive expression $1/\left((1-p)c\right)$ and using $t^*_{mix}(\hat{m}_d, \hat{m}_d, \hat{m}_m) = t^*(\hat{m}_d, \hat{m}_m)$, we obtain the sign-equivalent expression

$$-\mathrm{E}_{\rho,f}\Bigg[\rho \frac{\partial}{\partial m'_d}\left(1 - Z(t^* - m'_d)\right)\left((1 - Z(t^* - \hat{m}_m))K + rZ(t^* - \hat{m}_d) + (1-r)Z(t^* - \hat{m}_d) - 1\right)$$
$$+ \frac{\partial}{\partial m'_d}\left(1 - Z(t^*_{mix} - \hat{m}_d)\right)\left((1 - Z(t^*_{mix} - \hat{m}_m))K + rZ(t^*_{mix} - m'_d) + (1-r)Z(t^*_{mix} - \hat{m}_d) - 1\right)\Bigg]_{m'_d = \hat{m}_d}.$$

Performing the differentiation, we obtain:

$$-\mathrm{E}_{\rho,f}\Bigg[\rho Z'(t^* - m'_d)\left((1 - Z(t^* - \hat{m}_m))K + rZ(t^* - \hat{m}_d) + (1-r)Z(t^* - \hat{m}_d) - 1\right)$$
$$- Z'(t^*_{mix} - \hat{m}_d)\frac{\partial t^*_{mix}}{\partial m'_d}\left((1 - Z(t^*_{mix} - \hat{m}_m))K + rZ(t^*_{mix} - m'_d) + (1-r)Z(t^*_{mix} - \hat{m}_d) - 1\right)$$
$$+ \left(1 - Z(t^*_{mix} - \hat{m}_d)\right)\left(-Z'(t^*_{mix} - \hat{m}_m)K\frac{\partial t^*_{mix}}{\partial m'_d} + rZ'(t^*_{mix} - m'_d)\left(\frac{\partial t^*_{mix}}{\partial m'_d} - 1\right) + (1-r)Z'(t^*_{mix} - \hat{m}_d)\frac{\partial t^*_{mix}}{\partial m'_d}\right)\Bigg]_{m'_d = \hat{m}_d}.$$

By first making the substitutions $t^*_{mix}(\hat{m}_d, \hat{m}_d, \hat{m}_m) = t^*(\hat{m}_d, \hat{m}_m)$ and

$Z'(t^* - \hat{m}_m) = Z'(t^* - \hat{m}_d)/K$, and then $\left.\dfrac{\partial t^*_{mix}}{\partial m'_d}\right|_{m'_d=\hat{m}_d} = r\left(\dfrac{1}{2} - \dfrac{\ln K}{d^2}\right)$ and $r = f\rho$, we can write

and simplify this as:

$$
\begin{aligned}
&-\mathrm{E}_{\rho,f}\Bigg[\rho Z'(t^* - \hat{m}_d)\big((1 - Z(t^* - \hat{m}_m))K - (1 - Z(t^* - \hat{m}_d))\big) \\
&-Z'(t^* - \hat{m}_d)\left(\left.\frac{\partial t^*_{mix}}{\partial m'_d}\right|_{m'_d=\hat{m}_d}(1 - Z(t^* - \hat{m}_m))K + \left(r - \left.\frac{\partial t^*_{mix}}{\partial m'_d}\right|_{m'_d=\hat{m}_d}\right)\big(1 - Z(t^* - \hat{m}_d)\big)\right)\Bigg] \\
&= \mathrm{E}_{\rho,f}\left[Z'(t^* - \hat{m}_d)\left(\left(\rho f\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) - \rho\right)(1 - Z(t^* - \hat{m}_m))K + \left(\rho f - \rho f\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) + \rho\right)\big(1 - Z(t^* - \hat{m}_d)\big)\right)\right] \\
&= Z'(t^* - \hat{m}_d)\mathrm{E}_{\rho,f}\left[\rho\left(\left(f\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)(1 - Z(t^* - \hat{m}_m))K + \left(f\left(\frac{1}{2} + \frac{\ln K}{d^2}\right) + 1\right)\big(1 - Z(t^* - \hat{m}_d)\big)\right)\right].
\end{aligned}
$$

The fitness gradient has the same sign as the expectation, which can be written:

$$
\begin{aligned}
&\mathrm{E}_{\rho,f}\left[\rho f\right]\left(\left(\frac{1}{2} - \frac{\ln K}{d^2}\right)(1 - Z(t^* - \hat{m}_m))K + \left(\frac{1}{2} + \frac{\ln K}{d^2}\right)\big(1 - Z(t^* - \hat{m}_d)\big)\right) + \\
&\mathrm{E}_{\rho}\left[\rho\right]\big(1 - Z(t^* - \hat{m}_d) - (1 - Z(t^* - \hat{m}_m))K\big) \\
&= \bar{r}\left(\left(\frac{1}{2} - \frac{\ln K}{d^2}\right)(1 - Z(t^* - \hat{m}_m))K + \left(\frac{1}{2} + \frac{\ln K}{d^2}\right)\big(1 - Z(t^* - \hat{m}_d)\big)\right) \\
&+\bar{\rho}\big((1 - Z(t^* - \hat{m}_d)) - (1 - Z(t^* - \hat{m}_m))K\big).
\end{aligned}
$$

Assuming ρ and f are uncorrelated, this equals

$$
\bar{\rho}\left(\left(\bar{f}\left(\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)\left(1 - Z\left(-\frac{d}{2} - \frac{\ln K}{d}\right)\right)K - \left(\bar{f}\left(-\frac{1}{2} - \frac{\ln K}{d^2}\right) - 1\right)\left(1 - Z\left(\frac{d}{2} - \frac{\ln K}{d}\right)\right)\right).
$$