Queller[33,34] in assuming that there is no covariance between individual genetic value and (in our case) the three residual error terms: $\mathbb{C}_x\left[G_x, \epsilon_{\mathbb{E}_\pi[w_x]}\right] = \mathbb{C}_x\left[G_x, \epsilon_{\mathbb{E}_\pi[w_x]} \ll^k \overline{w} \gg\right] = \mathbb{C}_x\left[G_x, \epsilon_{\ll w_x, {}^k\overline{w} \gg}\right] = 0$.

The partial regression slopes are constants so can be moved outside of their respective covariances with G_x. We can now express the condition for an expected increase in $\bar{G}$ (i.e. $\mathbb{E}_\pi[\Delta\bar{G}] > 0$) as follows:

$$\frac{1}{\mathbb{E}_\pi[\overline{w}]}\left(\beta_{\mathbb{E}_\pi[w_x],G_x}\mathbb{C}_x[G_x, G_x] + \beta_{\mathbb{E}_\pi[w_x],G_y}\mathbb{C}_x[G_x, G_y] + \sum_{k=1}^{\infty}\frac{(-1)^k}{\mathbb{E}_\pi[\overline{w}]^k}\begin{pmatrix} \ll^k \overline{w} \gg \beta_{\mathbb{E}_\pi[w_x],G_x}\mathbb{C}_x[G_x, G_x] + \ll^k \overline{w} \gg \beta_{\mathbb{E}_\pi[w_x],G_y}\mathbb{C}_x[G_x, G_y] \\ +\beta_{\ll w_x, {}^k\overline{w}\gg, G_x}\mathbb{C}_x[G_x, G_x] + \beta_{\ll w_x, {}^k\overline{w}\gg, G_y}\mathbb{C}_x[G_x, G_y] \end{pmatrix}\right) > 0 \qquad (A11)$$

Dividing both sides of Inequality (A11) by the variance in genetic value across individuals, $\mathbb{V}_x[G_x]$, obtains relatedness[33,35] $\left(r \equiv \frac{\mathbb{C}_x[G_y, G_x]}{\mathbb{V}_x[G_x]}\right)$:

$$\frac{1}{\mathbb{E}_\pi[\overline{w}]}\left(\beta_{\mathbb{E}_\pi[w_x],G_x} + r\beta_{\mathbb{E}_\pi[w_x],G_y} + \sum_{k=1}^{\infty}\frac{(-1)^k}{\mathbb{E}_\pi[\overline{w}]^k}\begin{pmatrix} \ll^k \overline{w} \gg \left(\beta_{\mathbb{E}_\pi[w_x],G_x} + r\beta_{\mathbb{E}_\pi[w_x],G_y}\right) \\ +\beta_{\ll w_x, {}^k\overline{w}\gg, G_x} + r\beta_{\ll w_x, {}^k\overline{w}\gg, G_y} \end{pmatrix}\right) > 0 \qquad (A12)$$

We multiply both sides of this inequality by $\mathbb{E}_\pi[\overline{w}]$. Grouping the coefficients of r gives:

$$r\left(\beta_{\mathbb{E}_\pi[w_x],G_y} + \sum_{k=1}^{\infty}\frac{(-1)^k}{\mathbb{E}_\pi[\overline{w}]^k}\left(\ll^k \overline{w} \gg \beta_{\mathbb{E}_\pi[w_x],G_y} + \beta_{\ll w_x, {}^k\overline{w}\gg, G_y}\right)\right) > -\beta_{\mathbb{E}_\pi[w_x],G_x} + \sum_{k=1}^{\infty}\frac{(-1)^k}{\mathbb{E}_\pi[\overline{w}]^k}\left(-\ll^k \overline{w} \gg \beta_{\mathbb{E}_\pi[w_x],G_x} - \beta_{\ll w_x, {}^k\overline{w}\gg, G_x}\right) \qquad (A13)$$

For clarity, we denote the regression slopes of the individual's genetic value G_x and the genetic value of its social partner G_y on the different parameters of the individual's probability distribution for reproductive success as follows:

$$b_\mu \equiv \beta_{\mathbb{E}_\pi[w_x],G_y} = \beta_{\mathbb{E}_\pi[w_y],G_x} \qquad (A14a)$$

$$b_k \equiv \beta_{\ll w_x, {}^k\overline{w}\gg, G_y} = \beta_{\ll w_y, {}^k\overline{w}\gg, G_x} \qquad (A14b)$$

$$c_\mu \equiv -\beta_{\mathbb{E}_\pi[w_x],G_x} \qquad (A14c)$$

$$c_k \equiv -\beta_{\ll w_x, {}^k\overline{w}\gg, G_x} \qquad (A14d)$$

The general expression for a stochastic Hamilton's rule is then:

$$r\left(b_\mu + \sum_{k=1}^{\infty} \frac{(-1)^k}{\mathbb{E}_\pi[\bar{w}]^k}\left(\ll^k \bar{w} \gg b_\mu + b_k\right)\right) > c_\mu + \sum_{k=1}^{\infty} \frac{(-1)^k}{\mathbb{E}_\pi[\bar{w}]^k}\left(\ll^k \bar{w} \gg c_\mu + c_k\right) \qquad (A15)$$

which is Inequality (1) in the main text.

A2 | Approximation for the first two moments

A bet-hedging genotype reduces the variance in fitness at the expense of reducing arithmetic mean fitness. We obtain the stochastic approximation of Hamilton's rule suitable for bet-hedging effects by ignoring $k > 1$ in Equation (A4) (e.g. skew) to focus only on the arithmetic mean and variance effects. This allows us to approximate the selection covariance of the Price equation as follows:

$$\mathbb{E}_\pi[\Delta\bar{G}] \approx \mathbb{C}_x\left[G_x, \left(\frac{\mathbb{E}_\pi[w_x]}{\mathbb{E}_\pi[\bar{w}]} - \frac{\mathbb{C}_\pi[w_x, \bar{w}]}{\mathbb{E}_\pi[\bar{w}]^2}\right)\right] \qquad (A16)$$

The covariance between individual fitness w_x and population average fitness $\bar{w}$ in Equation (A16) can be alternatively expressed as:

$$\mathbb{C}_\pi[w_x, \bar{w}] = \rho_x \sigma_\pi[w_x]\sigma_\pi[\bar{w}] \qquad (A17)$$

where $\sigma_\pi[w_x]$ is the standard deviation of the individual's reproductive success, $\sigma_\pi[\bar{w}]$ is the standard deviation of the population's average reproductive success over Π, and ρ_x is the product-moment correlation coefficient for w_x and $\bar{w}$ as they fluctuate over Π. Substituting these terms into the approximation of the selection covariance (Equation (A16)) obtains:

$$\mathbb{E}_\pi[\Delta\bar{G}] \approx \frac{1}{\mathbb{E}_\pi[\bar{w}]}\mathbb{C}_x[G_x, (\mathbb{E}_\pi[w_x] - \nu\rho_x\sigma_\pi[w_x])] \qquad (A18)$$

where ν denotes the coefficient of variation of the population's average reproductive success:

$$\nu = \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]} \qquad (A19)$$

ν is independent of the organism's decisions, and quantifies the degree to which the environment is stochastic.

The condition for expected increase in $\bar{G}$ ($\mathbb{E}_\pi[\Delta\bar{G}] > 0$) is then:

$$\mathbb{C}_x[G_x, (\mathbb{E}_\pi[w_x] - \nu\rho_x\sigma_\pi[w_x])] > 0 \qquad (A20)$$

Above (Equation (A5)), we have already defined $\mathbb{E}_\pi[w_x]$ using multiple linear regression. We follow a similar approach with standard deviation, weighted by the degree to which it correlates (ρ_x) with the fluctuating average reproductive success $\bar{w}$ in the population:

$$\rho_x \sigma_\pi[w_x] = \alpha_{\rho_x \sigma_\pi[w_x]} + \beta_{\rho_x \sigma_\pi[w_x],G_x} G_x + \beta_{\rho_x \sigma_\pi[w_x],G_y} G_y + \epsilon_{\rho_x \sigma_\pi[w_x]} \qquad (A21)$$

Substituting Equation (A5) and Equation (A21) into the condition for selection (Inequality (A20)), we obtain:

$$\mathbb{C}_x\Big[G_x, \Big(\alpha_{\mathbb{E}_\pi[w_x]} + \beta_{\mathbb{E}_\pi[w_x],G_x} G_x + \beta_{\mathbb{E}_\pi[w_x],G_y} G_y + \epsilon_{\mathbb{E}_\pi[w_x]} \\ - \nu\left(\alpha_{\rho_x \sigma_\pi[w_x]} + \beta_{\rho_x \sigma_\pi[w_x],G_x} G_x + \beta_{\rho_x \sigma_\pi[w_x],G_y} G_y + \epsilon_{\rho_x \sigma_\pi[w_x]}\right)\Big)\Big] > 0 \qquad (A22)$$

As before, the covariances of G_x with the constants equal 0 (i.e. $\mathbb{C}_x\left[G_x, \alpha_{\mathbb{E}_\pi[w_x]}\right] = \mathbb{C}_x\left[G_x, \nu\alpha_{\rho_x \sigma_\pi[w_x]}\right] = 0$), and the covariances with the error terms are assumed to be zero (i.e. $\mathbb{C}_x\left[G_x, \epsilon_{\mathbb{E}_\pi[w_x]}\right] = \mathbb{C}_x\left[G_x, \nu\epsilon_{\rho_x \sigma_\pi[w_x]}\right] = 0$). For clarity, we denote the effects on the correlated variation of the recipient's reproductive success as follows:

$$b_\sigma \equiv -\beta_{\rho_x \sigma_\pi[w_x],G_y} = -\beta_{\rho_x \sigma_\pi[w_y],G_x} \qquad (A23a)$$

$$c_\sigma \equiv \beta_{\rho_x \sigma_\pi[w_x],G_x} \qquad (A23b)$$

Substituting Equations (A23a) and (A23b) into (A22) gives:

$$b_\mu \mathbb{C}_x[G_x, G_y] - c_\mu \mathbb{C}_x[G_x, G_x] - \nu\left(c_\sigma \mathbb{C}_x[G_x, G_x] - b_\sigma \mathbb{C}_x[G_x, G_y]\right) > 0 \qquad (A24)$$

Dividing both sides of Inequality (A24) by $\mathbb{V}_x[G_x]$ to obtain relatedness $\left(r \equiv \frac{\mathbb{C}_x[G_y, G_x]}{\mathbb{V}_x[G_x]}\right)$, we can rewrite the condition for selection as follows:

$$r\left(b_\mu + \nu b_\sigma\right) > c_\mu + \nu c_\sigma \qquad (A25)$$

which is Inequality (2) in the main text.

Note that a positive benefit b_σ (beneficial for the recipient) will be a *negative* regression slope, since it will be *reducing* the volatility of the recipient's reproduction. Likewise, a positive cost c_σ (deleterious for the actor) will be a *positive* regression slope, since it will be *increasing* the volatility of the actor's reproduction. If the actor can succeed in reducing its own reproductive volatility, c_σ will be negative (i.e. a 'negative cost').

Accordingly, the benefit term B to expected relative fitness is:

$$B \approx b_\mu + \nu b_\sigma \tag{A26}$$

The cost term C to expected relative fitness is:

$$C \approx c_\mu + \nu c_\sigma \tag{A27}$$

These are approximations of the exact benefit and cost terms captured in the general expression for Hamilton's rule (Inequality (1) in the main text), showing that selection can favour paying a cost to expected reproductive success $(c_\mu > 0, c_\sigma = 0)$ to reduce the $\bar{w}$-correlated variation of a relative's reproductive success $(b_\sigma > 0)$ even in the absence of any effect on the expected reproductive success of the recipient $(b_\mu = 0)$. In this situation:

$$\mathbb{E}_\pi[\Delta\bar{G}] > 0 \iff r > \frac{c_\mu}{\nu b_\sigma} \tag{A28}$$

The relative importance of mean effects (b_μ and c_μ) versus volatility effects (b_σ and c_σ) is determined by ν. If we denote the importance of mean effects (i.e. their power to determine the outcome of selection) with the weight a_μ and the importance of volatility effects with the weight a_σ, such that Inequality (A25) can be written as $r(a_\mu b_\mu + a_\sigma b_\sigma) > a_\mu c_\mu + a_\sigma c_\sigma$, these weights are the numerator and denominator of ν (i.e. $a_\mu = \mathbb{E}_\pi[\bar{w}]$ and $a_\sigma = \sigma_\pi[\bar{w}]$). In short, we emphasise that the true benefits and costs in social evolution should be measured using the expectation of relative fitness[1], which we decompose here into mean fecundity and volatility effects, rather than mean fecundity effects alone.

Under the definition of 'bet-hedging', a behaviour must incur a cost to arithmetic mean number of offspring whilst deriving a benefit by reducing the variance associated with the number of offspring[3]. The role of fitness variation reduction in social evolution has long attracted verbal speculation[9,10,13,36–39], but has evaded formalisation. We define 'altruistic bet-hedging' as occurring when the cost (a reduction in arithmetic mean number of offspring) is paid by the actor whilst a recipient derives the benefit (a reduction in the variance associated with the number of offspring). It is, of course, possible that the recipient may also experience either an increase or a decrease in arithmetic mean number of offspring (a b_μ effect). For clarity, we include such cases as 'altruistic bet-hedging' only if the b_μ effect if insufficient to overcome the costs paid by the actor without the additional b_σ effect. A behaviour is altruistic bet-hedging if it (*i*) involves a cost c_μ paid by the actor and (*ii*) would not evolve without a socially mediated reduction in the variation of a recipient's reproductive success (a b_σ effect).

In short, each state has a mean fitness $\bar{w}$, and a distribution of realised fitnesses for every individual. Unlike environmental stochasticity, within-genotype demographic stochasticity (inter-individual variation in fitness within the same environmental state) is shown by Inequality (2) (main text) not to matter to the outcome of selection in large populations, because the regression slopes cut through this variation to obtain the relationship between alleles and fitness visible to natural selection. The one condition in which inter-individual variation in fitness within the same environment state does matter is when population sizes (the scale of the competitive population) are tiny, a well-known result in the bet-hedging literature that we generalise for social interactions in section A5 below (Inequality (3) in the main text).

Note that when the 'natural' distribution for reproductive success is sufficiently skewed (i.e. either good years or bad years are rare), Hamilton's rule will need to be approximated to higher moments (e.g. $k = 2$), using Inequality (1) in the main text, to capture effects on the asymmetry of the probability distributions from which the social partners are sampling their reproductive success (although under such conditions, organisms will tend to be specialised to the common environmental state).

A3 | Non-social bet-hedging and Hamilton's rule

In this section, we show how the stochastic Hamilton's rule (Inequality (A25)) captures familiar forms of bet-hedging as special cases. In the absence of social interaction ($rB = 0$), the rule is simply:

$$c_\mu + vc_\sigma < 0 \tag{A29}$$

Note that $c_\mu = -\beta_{\mathbb{E}_\pi[w_x],G_x}$ (Equation (A14c)), so a reduction in the reproductive success of the actor ($\beta_{\mathbb{E}_\pi[w_x],G_x} < 0$) represents a positive cost c_μ. In terms of regression effects, therefore, the stochastic Hamilton's rule shows the condition for non-social bet-hedging to be as follows, where $\beta_{\mathbb{E}_\pi[w_x],G_x} < 0$ and $\beta_{\rho_x\sigma_\pi[w_x],G_x} < 0$:

$$\beta_{\mathbb{E}_\pi[w_x],G_x} > v\beta_{\rho_x\sigma_\pi[w_x],G_x} \tag{A30}$$

To see how these results recover previous results in the non-social bet-hedging literature, consider a non-social haploid population consisting of two rival alleles, denoted A_1 and A_2. To identify whether selection is expected to favour the A_1 allele ($\mathbb{E}_\pi[\Delta\bar{G}] > 0$), we ask whether there will be a change in genetic value for A_1 (individuals with the allele of interest A_1 have a genetic value $G_x = 1$, whilst those lacking it have a genetic value of $G_x = 0$).

Each individual x has an expected number of offspring μ_x and a standard deviation in number of offspring σ_x. Following Frank & Slatkin[40] and Starrfelt & Kokko[3], set μ_x equal to the value μ_1 for all carriers of allele 1 (and equal to μ_2 for all carriers of allele 2) and σ_x equal to the value σ_1 for all carriers of allele 1 (and equal to σ_2 for all carriers of allele 2). In other words, members of a genotype sample their fitness w_x from a probability distribution shared by all members of the genotype, but they may in principle do so in an uncorrelated fashion with other members of the genotype. The degree to which an individual's fitness w_x correlates with fluctuations in population average reproductive success $\bar{w}$ is given by $\rho_{w_x,\bar{w}}$.

To obtain the exact expected change in gene frequency, Hamilton's rule can be expressed in the following format:

$$\mathbb{E}_\pi[\Delta\bar{G}] = \frac{r(b_\mu + vb_\sigma) - c_\mu - vc_\sigma}{\mathbb{E}_\pi[\bar{w}]} \cdot \mathbb{V}_x[G_x] \qquad (A31)$$

In Equation (A31), we derive Hamilton's rule without dividing by the variance $\mathbb{V}_x[G_x]$ in genetic value; see Equation (2.3) in Okasha & Martens[41].

We now set b_μ and b_σ in Equation (A31) to zero to focus on non-social cases. Fitting the stochastic Hamilton's rule (Inequality (A25)) obtains the following non-social components c_μ and c_σ:

$$c_\mu = \mu_2 - \mu_1 \qquad (A32a)$$

$$c_\sigma = \rho_{1,\bar{w}}\sigma_1 - \rho_{2,\bar{w}}\sigma_2 \qquad (A32b)$$

In other words, there are two horizontal positions (0 and 1) on a graph of μ against genetic value G_x; the two vertical positions are μ_1 and μ_2. The slope $\beta_{\mathbb{E}_\pi[w_x],G_x}$ of μ against G_x is then simply $\mu_1 - \mu_2$. The cost term c_μ is $-\beta_{\mathbb{E}_\pi[w_x],G_x}$, i.e. $\mu_2 - \mu_1$. Likewise, on the graph of $\bar{w}$-correlated volatility against genetic value G_x, there are two vertical positions $\rho_{1,\bar{w}}\sigma_1$ and $\rho_{2,\bar{w}}\sigma_2$, so the slope $\beta_{\rho_x\sigma_\pi[w_x],G_x}$ of $\bar{w}$-correlated volatility against G_x is $\rho_{1,\bar{w}}\sigma_1 - \rho_{2,\bar{w}}\sigma_2$. The coefficient c_σ is equal to $\beta_{\rho_x\sigma_\pi[w_x],G_x}$ (Equation (A23b)).

Recall that v is the coefficient of variation in population average reproductive success ($\frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}$; Equation (A19)). The variance in breeding value, $\mathbb{V}_x[G_x]$, is equal to q_1q_2, as it represents a two-point distribution (i.e. $\mathbb{V}_x[G_x] = q_1q_2(1-0)^2 = q_1q_2$). Equation (A31) can now be written:

$$\mathbb{E}_\pi[\Delta\bar{G}] = \frac{\mu_1 - \mu_2 - \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}(\rho_{1,\bar{w}}\sigma_1 - \rho_{2,\bar{w}}\sigma_2)}{\mathbb{E}_\pi[\bar{w}]} q_1q_2 \qquad (A33)$$

As there are only two genetic values in a haploid world (0 and 1), $\rho_{1,\bar{w}}\sigma_1$ and $\rho_{2,\bar{w}}\sigma_2$ are the expected values of $\rho_{w_x,\bar{w}}\sigma_x$ for members of each genotype, obtained as a predicted value in a least-squares regression (Inequality (A25)). We denote individuals with the index i:

$$\rho_{1,\bar{w}}\sigma_1 = \sigma_1 \cdot \frac{1}{q_1N}\sum_{i=1}^{q_1N}\rho_{w_i,\bar{w}} \qquad (A34a)$$

$$\rho_{2,\bar{w}}\sigma_2 = \sigma_2 \cdot \frac{1}{q_2N}\sum_{i=1}^{q_2N}\rho_{i,\bar{w}} \qquad (A34b)$$

The summation term in Equation (A34a) contains the correlation between individual i's reproductive success w_i and the average reproductive success $\overline{w}$ in the population. Since a correlation can be expressed in the form $\rho_{Z,Y} = \frac{\mathbb{C}[Z,Y]}{\sigma[Z]\sigma[Y]}$, we express this summation term as follows:

$$\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \rho_{w_i,\overline{w}} = \frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \frac{\mathbb{C}_\pi[w_i,\overline{w}]}{\sigma_1 \sigma_\pi[\overline{w}]} \qquad (A35)$$

We then carry the summation inside the covariance:

$$\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \rho_{w_i,\overline{w}} = \frac{\mathbb{C}_\pi\left[\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} w_i\,,\overline{w}\right]}{\sigma_1 \sigma_\pi[\overline{w}]} \qquad (A36)$$

The term $\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} w_i$ is the average reproductive success for carriers of allele A_1. To match notation in Starrfelt & Kokko[3], we call this R_1:

$$R_1 = \frac{1}{q_1 N}\sum_{i=1}^{q_1 N} w_i \qquad (A37)$$

Equation (A37) can now be expressed more simply as:

$$\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \rho_{w_i,\overline{w}} = \frac{\mathbb{C}_\pi[R_1,\overline{w}]}{\sigma_1 \sigma_\pi[\overline{w}]} \qquad (A38)$$

Likewise, since $\overline{w} = q_1 R_1 + q_2 R_2$, we substitute this formula for $\overline{w}$ into Equation (A38) and expand the covariance (since $\mathbb{C}[X+Z,Y] = \mathbb{C}[X,Y] + \mathbb{C}[Z,Y]$):

$$\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \rho_{w_i,\overline{w}} = \frac{q_1 \sigma_\pi[R_1]^2 + q_2 \mathbb{C}_\pi[R_1,R_2]}{\sigma_1 \sigma_\pi[\overline{w}]} \qquad (A39)$$

As Starrfelt & Kokko[3] note in their Equations (7–9), $\mathbb{C}_\pi[R_1,R_2] = \rho_{12}\sigma_1\sigma_2$, and $\sigma_\pi[R_1]^2 = \rho_1 \sigma_1^2$, letting σ_1 denote the standard deviation in reproductive success of an individual carrying allele 1 and ρ_1 denote the correlation in reproductive success between individuals carrying allele 1. Therefore:

$$\frac{1}{q_1 N}\sum_{i=1}^{q_1 N} \rho_{w_i,\overline{w}} = \frac{q_1 \rho_1 \sigma_1^2 + q_2 \rho_{12}\sigma_1\sigma_2}{\sigma_1 \sigma_\pi[\overline{w}]} \qquad (A40)$$

We can perform the same series of rearrangements for carriers of allele A_2:

$$\frac{1}{q_2 N}\sum_{i=1}^{q_2 N} \rho_{w_i,\overline{w}} = \frac{q_2 \rho_2 \sigma_2^2 + q_1 \rho_{12}\sigma_1\sigma_2}{\sigma_2 \sigma_\pi[\overline{w}]} \qquad (A41)$$

Substituted into Hamilton's rule (Equation (A33)), this obtains:

$$\mathbb{E}_\pi[\Delta\bar{G}] = \frac{\mu_1 - \mu_2 + \frac{1}{\mathbb{E}_\pi[\bar{w}]}\left(q_2\rho_2\sigma_2^2 - q_1\rho_1\sigma_1^2 + (q_1 - q_2)\rho_{12}\sigma_1\sigma_2\right)}{\mathbb{E}_\pi[\bar{w}]} q_1 q_2 \qquad (A42)$$

If the population is neither rising nor falling in size, $\mathbb{E}_\pi[\bar{w}] = 1$:

$$\mathbb{E}_\pi[\Delta\bar{G}] = q_1 q_2 \left(\mu_1 - \mu_2 + (q_2\rho_2\sigma_2^2 - q_1\rho_1\sigma_1^2 + (q_1 - q_2)\rho_{12}\sigma_1\sigma_2)\right) \qquad (A43)$$

Equation (A43) recovers Frank & Slatkin's[40] Equation (7) and Starrfelt & Kokko's[3] Equation (10) for the canonical bet-hedging model familiarly used in the literature (i.e. two alleles in a fixed-size population at a haploid locus in a fluctuating environment of two or more states).

From Equation (A43), we can recover the geometric mean heuristic (which provides a prediction of which allele will fixate) by assuming[40] that there is no correlation between genotypes ($\rho_{12} = 0$) and setting the population to equal frequencies[40] of each allele ($q_1 = q_2 = \frac{1}{2}$). These conditions provide the well-known geometric-mean approximation (Equation (12) in Frank & Slatkin[40]; note that, as heuristic approximations, there are at least five different estimates for the geometric mean, all of which perform roughly equally well[42]):

$$\mu_1 - \frac{\rho_1\sigma_1^2}{2} > \mu_2 - \frac{\rho_2\sigma_2^2}{2}$$

A4 | Uncertain relatedness

The potential effects of stochasticity on relatedness have been raised by Goodnight[43] and Lehmann & Balloux[24]. In this section, we explore how uncertainty over relatedness influences the outcome of selection. We show that the mean relatedness of recipients is sufficient to predict the outcome of selection when there is no strong correlation (across environmental states) between the relatedness r of social partners and the average reproductive success in the population ($\bar{w}$). We denote this correlation as $\rho_{r,\bar{w}}$. However, if the relatedness of interactants and average reproductive success are negatively correlated ($\rho_{r,\bar{w}} < 0$), investments in social partners become more valuable as environmental stochasticity increases (i.e. at high values of v). Conversely, investments in social partners become less valuable in a stochastic environment if this correlation is positive ($\rho_{r,\bar{w}} > 0$).

To illustrate this result, we denote the reproductive success of individual x in state π as $w_{x(\pi)}$, and express this quantity as a function of its own genetic value G_x and the genetic value G_y of its social partner:

$$w_{x(\pi)} = \alpha_{w_{x(\pi)}} + \beta_{w_{x(\pi)},G_x} G_x + \beta_{w_{x(\pi)},G_y} G_{y(\pi)} + \epsilon_{w_{x(\pi)}} \tag{A44}$$

We substitute this regression formula into the Price equation (Equation (A2)) to express the condition for selection ($\mathbb{E}_\pi[\Delta\bar{G}] > 0$) as:

$$\mathbb{E}_\pi\left[\mathbb{C}_x\left[G_x, \frac{\alpha_{w_x} + \beta_{w_x,G_x} G_x + \beta_{w_x,G_y} G_y + \epsilon_{w_x}}{\bar{w}}\right]\right] > 0 \tag{A45}$$

For a given environmental state π, $\bar{w}$ is a constant with respect to G_x, so we move it outside the covariance, which is defined only for the environmental state π. By the linearity of covariances ($\mathbb{C}[X+Y,Z] = \mathbb{C}[X,Z] + \mathbb{C}[Y,Z]$), this is equal to:

$$\mathbb{E}_\pi\left[\frac{\left(\mathbb{C}_x[G_x, \alpha_{w_x}] + \mathbb{C}_x[G_x, \beta_{w_x,G_x} G_x] + \mathbb{C}_x\left[G_x, \beta_{w_x,G_y} G_y\right] + \mathbb{C}_x[G_x, \epsilon_{w_x}]\right)}{\bar{w}}\right] > 0 \tag{A46}$$

$\mathbb{C}_x\left[G_x, \alpha_{w_x}\right] = 0$ (since α_{w_x} is a constant) and we assume $\mathbb{C}_x\left[G_x, \epsilon_{w_x}\right] = 0$. Then:

$$\mathbb{E}_\pi\left[\frac{\left(\beta_{w_x,G_x}\mathbb{C}_x[G_x, G_x] + \beta_{w_x,G_y}\mathbb{C}_x\left[G_x, G_y\right]\right)}{\bar{w}}\right] > 0 \tag{A47}$$

We now divide both sides of Inequality (A47) by the variance in genetic value ($\mathbb{V}_x[G_x]$) to obtain state-specific relatedness $r_\pi = \frac{\mathbb{C}_x[G_x,G_y]}{\mathbb{V}_x[G_x]}$:

$$\mathbb{E}_\pi\left[\frac{\beta_{w_x,G_x} + \beta_{w_x,G_y} r}{\bar{w}}\right] > 0 \tag{A48}$$

Retaining the first two terms of the Taylor series expansion of Inequality (A48) gives:

$$\frac{\mathbb{E}_\pi[\beta_{w_x,G_x}] + \mathbb{E}_\pi\left[\beta_{w_x,G_y} r\right]}{\mathbb{E}_\pi[\bar{w}]} - \frac{\mathbb{C}_\pi[\beta_{w_x,G_x}, \bar{w}] + \mathbb{C}_\pi\left[\beta_{w_x,G_y} r, \bar{w}\right]}{\mathbb{E}_\pi[\bar{w}]^2} > 0 \tag{A49}$$

We now consider the case in which the benefit conferred on a recipient and the cost paid by the actor are the same in all environmental states ($\beta_{w_x,G_y} = b$ and $\beta_{w_x,G_x} = -c$ for all π). However, we allow relatedness to the recipient to vary among states π. This captures the possibility that actors associate with either closer or more distant relatives when the conditions change. The covariance of cost and population average fitness is zero, because cost is now a constant across states ($\mathbb{C}_\pi\left[\beta_{w_x,G_x}, \bar{w}\right] = \mathbb{C}_\pi[c, \bar{w}] = 0$). Multiplying both sides by $\mathbb{E}_\pi[\bar{w}]$, Inequality (A49) can be simplified to:

$$b \cdot \mathbb{E}_\pi[r] - b \cdot \frac{\mathbb{C}_\pi[r, \bar{w}]}{\mathbb{E}_\pi[\bar{w}]} - c > 0 \tag{A50}$$

We now rearrange Inequality (A50) by expanding the covariance. The covariance between relatedness and population average fitness ($\mathbb{C}_\pi[r_y, \bar{w}]$) can be written as $\rho_{r,\bar{w}}\sigma_\pi[r]\sigma_\pi[\bar{w}]$, where ρ denotes correlation and σ_π denotes standard deviation across environmental states. We introduce v as the stochasticity coefficient (the coefficient of variation in population average reproductive success, $v = \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}$, which we introduced earlier in Equation (A19)). We also use the following notation for clarity: we denote the expectation of relatedness across environmental states as r_μ, and we denote the standard deviation of relatedness across environmental states as r_σ:

$$r_\mu = \mathbb{E}_\pi[r] \tag{A51a}$$

$$r_\sigma = \sigma_\pi[r] \tag{A51b}$$

Accordingly, Inequality (A50) can be expressed as:

$$(r_\mu - \rho_{r,\bar{w}} v r_\sigma)b > c \tag{A52}$$

Inequality (A52) shows that uncertainty over relatedness (r_σ) only influences selection if relatedness fluctuates strongly in either a positively or negatively correlated fashion with population average reproductive success. A negative correlation, across environmental states, between relatedness and average reproductive success ($\rho_{r,\bar{w}} < 0$) results in an actor's most valuable investments being focused on closer relatives. These investments are the 'most valuable' because an increase in recipient fecundity of a given size b is more valuable when competitors are underperforming (low $\bar{w}$): the recipient will enjoy a proportionally greater market share of reproduction than if the increase had occurred when competitors were overperforming (high $\bar{w}$). Mean relatedness r_μ is sufficient to capture the outcome of selection when population mean reproductive success does not fluctuate ($v \approx 0$), even if actors face high levels of uncertainty r_σ about the kinship of recipients.

A5 | Demographic stochasticity

We now consider the role of risk in a 'static' environment, for which the environment does not fluctuate between states (the influence of the environment is identical across the set Π). In a population of size N in which organisms sample their reproductive success independently ($\mathbb{C}_\pi[w_x, w_{j\neq x}] = 0$), the covariance (defined across possible fitness outcomes) between the focal individual's reproductive success (w_x) and the average reproductive success in the population ($\bar{w}$) is:

$$\mathbb{C}_\pi[w_x,\overline{w}] = \mathbb{C}_\pi\left[w_x,\frac{1}{N}\sum_j^N w_j\right] = \frac{1}{N}\sum_j^N \mathbb{C}_\pi[w_x,w_j] = \frac{1}{N}\left(\mathbb{C}_\pi[w_x,w_x] + \sum_{j\neq x}\mathbb{C}_\pi[w_x,w_j]\right) = \frac{\mathbb{V}_\pi[w_x]}{N}$$
$$= \frac{\sigma_\pi[w_x]^2}{N}$$

$$(A53)$$

We substitute this into the selection covariance of the Price equation (and multiply out $\mathbb{E}_\pi[\overline{w}]$):

$$\mathbb{C}_x\left[G_x,\left(\mathbb{E}_\pi[w_x] - \frac{\sigma_\pi[w_x]^2}{N\mathbb{E}_\pi[\overline{w}]}\right)\right] > 0 \qquad (A54)$$

Applying Queller's[33] regression method (as in Appendix A2) to this equation, we obtain Inequality (3) in the main text (where b_{σ^2} is the effect of the partner's genotype on the organism's within-generation variance in reproductive success $(-\beta_{\mathbb{V}_\pi[w_x],G_y})$, and c_{σ^2} is the effect of the organism's genotype on its own variance in reproductive success $(\beta_{\mathbb{V}_\pi[w_x],G_x})$:

$$r\left(b_\mu + \frac{b_{\sigma^2}}{N\mathbb{E}_\pi[\overline{w}]}\right) > c_\mu + \frac{c_{\sigma^2}}{N\mathbb{E}_\pi[\overline{w}]} \qquad (A55)$$

A6 | Environmental granularity and dispersal in Hamilton's rule

In this section, we show how the spatial scale at which environments fluctuate influences the role of b_σ and c_σ in selection.

The magnitude of the stochasticity coefficient v depends on the correlation among individuals in their exposure to all conditions of the environment. Accordingly, when individuals are distributed across different microenvironments, the degree of correlation in environmental state across microenvironments influences the magnitude of v. Here, we illustrate this principle in a population divided into multiple microenvironments.

Let the population undergoing global competition be distributed across a total of M microenvironment patches, each of which samples its local environmental 'microstate' from an identical distribution. Population-wide environmental state π is, in effect, a specific combination of microstates across a network of spatial patches inhabited by a population. Assuming there are equal numbers of individuals in each patch, the whole-population average reproductive success $\overline{w}$ is equal to the mean of the mean reproductive success $\overline{w}_m$ in each patch m:

$$\overline{w} = \frac{1}{M}\sum_{m=1}^{M}\overline{w}_m \qquad (A56)$$

Since the scale of competition is the whole population, the stochasticity coefficient υ is obtained as the coefficient of variation in whole-population average reproductive success $\bar{w}$ (Equation (A19)):

$$\upsilon = \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]} = \frac{\sigma_\pi\left[\frac{1}{M}\sum_{m=1}^{M}\bar{w}_m\right]}{\mathbb{E}_\pi\left[\frac{1}{M}\sum_{m=1}^{M}\bar{w}_m\right]} \quad (A57)$$

We assume that every patch samples its microstate from an identical distribution with a mean of $\mathbb{E}_\pi[\bar{w}_m] = \mathbb{E}_\pi[\bar{w}]$ and a variance of $\mathbb{V}_\pi[\bar{w}_m]$, but patches can be correlated or uncorrelated in their samples from this distribution. The variance of $\bar{w}$ (i.e. $\sigma_\pi[\bar{w}]^2$) can then be obtained using the general formula for the variance of a mean[3] (since $\bar{w}$ is the mean of a total of M patches, each with its own $\bar{w}_m$ in a particular state π of the population):

$$\mathbb{V}_\pi[\bar{w}] = \mathbb{V}_\pi\left[\frac{1}{M}\sum_{m=1}^{\mathrm{M}}\bar{w}_m\right] = \left(\frac{1}{M} + \frac{M-1}{M}\bar{\rho}\right)\mathbb{V}_\pi[\bar{w}_m] \quad (A58)$$

$\bar{\rho}$ denotes the average between-patch correlation in average reproductive success $\bar{w}_m$. As patch number M approaches infinity, this whole-population variance $\mathbb{V}_\pi[\bar{w}]$ converges to a simple function of $\bar{\rho}$ and the within-patch variance $\mathbb{V}_\pi[\bar{w}_m]$ in average reproductive success:

$$\lim_{M\to\infty} \mathbb{V}_\pi[\bar{w}] = \bar{\rho}\cdot\mathbb{V}_\pi[\bar{w}_m] \quad (A59)$$

$\mathbb{V}_\pi[\bar{w}]$ is the square of the numerator of υ. Therefore, in a population distributed over many patches, υ is as follows, where υ_m is the coefficient of variation in average reproductive success within a single patch (i.e. patch-level stochasticity, $\upsilon_m = \frac{\sigma_\pi[\bar{w}_m]}{\mathbb{E}_\pi[\bar{w}_m]}$):

$$\lim_{M\to\infty} \upsilon = \frac{\sqrt{\bar{\rho}\cdot\mathbb{V}_\pi[\bar{w}_m]}}{\mathbb{E}_\pi[\bar{w}_m]} = \frac{\sigma_\pi[\bar{w}_m]}{\mathbb{E}_\pi[\bar{w}_m]}\cdot\sqrt{\bar{\rho}} = \upsilon_m\sqrt{\bar{\rho}} \quad (A60)$$

Equation (A60) shows that whole-population stochasticity υ approaches within-patch stochasticity υ_m as between-patch correlation approaches 1 (full correlation). This illustrates the fundamental point, emphasised by Starrfelt & Kokko[3] for non-social bet-hedging, that selection on variation effects (b_σ and c_σ in Inequality (A25)) is driven by whole-population environmental fluctuation when the scale of competition is at the level of the whole population (global competition), and that the 'grain size'[3] of environmental fluctuation (the size of completely correlated areas of the population) is key in determining the strength of selection (Main Text Fig. 1b).

doi:10.1038/nature25965

Appendix B | Deriving regression effects

Here, we describe how the benefit and cost terms are obtained in a specific model (implemented as a simulation in *MATLAB*, for which code is given in *Appendix D*).

B1 | Discrete environment states

Let a haploid asexual population consist of two genotypes, with genetic values 0 and 1, at a single locus. Genotype 0 is non-cooperative, whilst genotype 1 pays a cost c to reduce the volatility of its recipients' reproductive success to a proportion η of its natural level. The frequency of genotype 1 in the population is p (and so the frequency of genotype 0 is $1-p$). The environment fluctuates between two states ('good' and 'bad').

Following the assortment rules in the first model in Gardner et al.[1] (p. 1030), we assume that individuals preferentially pair with same type (cooperators or noncooperators) with the probabilities in Table B1.

Without cooperation, individuals have a fecundity of z_1 in a good year and z_2 in a bad year. Good years occur with probability d and bad years with probability $1-d$. The standard deviation of a genotype 0 individual with a genotype 0 social partner is then:

$$\sigma_{00} = \sqrt{(1-d)d(z_1-z_2)^2} \qquad (B1)$$

Supplementary Table B1 | Mean and variation of reproductive success as a function of social partners in a world fluctuating unpredictably between two states

Genotypes		Probability of interaction	Mean reproductive success	Volatility of reproductive success	ρ if $\eta \neq 0$
Focal (x)	Partner (y)		(μ_{xy})	(σ_{xy})	
1	1	$p^2 + \alpha p(1-p)$	$d(z_1-c)+(1-d)(z_2-c)$	$\eta\sqrt{d(1-d)\big((z_1-c)-(z_2-c)\big)^2}$	1
1	0	$(1-\alpha)p(1-p)$	$d(z_1-c)+(1-d)(z_2-c)$	$\sqrt{d(1-d)\big((z_1-c)-(z_2-c)\big)^2}$	1
0	1	$(1-\alpha)p(1-p)$	$dz_1+(1-d)z_2$	$\eta\sqrt{d(1-d)(z_1-z_2)^2}$	1
0	0	$(1-p)^2+\alpha p(1-p)$	$dz_1+(1-d)z_2$	$\sqrt{d(1-d)(z_1-z_2)^2}$	1

Assortment rules follow the first model in Gardner et al.[1], leading to $r = \alpha$.

A focal individual encountering a genotype 1 social partner experiences a reduction in its fecundity variation by the coefficient η.

From Inequality (A25), Hamilton's rule (approximated to the first two central moments) is:

$$r\left(b_\mu + vb_\sigma\right) > c_\mu + vc_\sigma \tag{B2}$$

To find the four partial regression slopes $(b_\mu, b_\sigma, c_\mu, c_\sigma)$, we fit the following equations to Supplementary Table B1:

$$\mathbb{E}_\pi[w_x] = \alpha_{\mathbb{E}_\pi[w_x]} + \beta_{\mathbb{E}_\pi[w_x],G_x} G_x + \beta_{\mathbb{E}_\pi[w_x],G_y} G_y + \epsilon_{\mathbb{E}_\pi[w_x]} \tag{B3}$$

$$\rho_x\sigma_\pi[w_x] = \alpha_{\rho_x\sigma_\pi[w_x]} + \beta_{\rho_x\sigma_\pi[w_x],G_x} G_x + \beta_{\rho_x\sigma_\pi[w_x],G_y} G_y + \epsilon_{\rho_x\sigma_\pi[w_x]} \tag{B4}$$

Thus, we solve two linear regression equations: one for expected reproductive success ($\mathbb{E}_\pi[w_x]$) and one for the correlated variation of reproductive success ($\rho_x\sigma_\pi[w_x]$). The partial regression slopes m_1 and m_2 in a multiple regression with two predictors h_1 and h_2 of l can be found by solving the following simultaneous equations[1]:

$$m_1 = \frac{\mathbb{C}[l, h_1]}{\mathbb{V}[h_1]} - m_2 \frac{\mathbb{C}[h_1, h_2]}{\mathbb{V}[h_1]} \tag{B5}$$

$$m_2 = \frac{\mathbb{C}[l, h_2]}{\mathbb{V}[h_2]} - m_1 \frac{\mathbb{C}[h_1, h_2]}{\mathbb{V}[h_2]} \tag{B6}$$

To find b_μ and c_μ, we simultaneously solve:

$$m_1 = \frac{\mathbb{C}[\mathbb{E}_\pi[w_x], G_x]}{\mathbb{V}[G_x]} - m_2 \frac{\mathbb{C}\left[G_x, G_y\right]}{\mathbb{V}[G_x]} \tag{B7}$$

$$m_2 = \frac{\mathbb{C}\left[\mathbb{E}_\pi[w_x], G_y\right]}{\mathbb{V}\left[G_y\right]} - m_2 \frac{\mathbb{C}\left[G_x, G_y\right]}{\mathbb{V}\left[G_y\right]} \tag{B8}$$

The components of Equations (B7) and (B8) fitted to Table B1 are:

$$\frac{\mathbb{C}[\mathbb{E}_\pi[w_x], G_x]}{\mathbb{V}[G_x]} = -c \tag{C9}$$

$$\frac{\mathbb{C}\left[G_x, G_y\right]}{\mathbb{V}[G_x]} = \alpha \tag{B10}$$

$$\frac{\mathbb{C}\left[\mathbb{E}_\pi[w_x], G_y\right]}{\mathbb{V}\left[G_y\right]} = -c\alpha \tag{B11}$$

$$\frac{\mathbb{C}\left[G_x, G_y\right]}{\mathbb{V}\left[G_y\right]} = \alpha \tag{B12}$$

17

We therefore simultaneously solve:

$$m_1 = -c - m_2\alpha \tag{B13}$$

$$m_2 = -(m_1 + c)\alpha \tag{B14}$$

This obtains $m_1 = -c$ and $m_2 = 0$, which are the partial regression slopes $\beta_{\mathbb{E}_\pi[w_x],G_x}$ and $\beta_{\mathbb{E}_\pi[w_x],G_y}$, respectively. Since the components c_μ and b_μ in Inequality (A25) are $c_\mu = -\beta_{\mathbb{E}_\pi[w_x],G_x}$ and $b_\mu = \beta_{\mathbb{E}_\pi[w_x],G_y}$, these components are therefore:

$$c_\mu = c \tag{B15}$$

$$b_\mu = 0 \tag{B16}$$

We solve an equivalent pair of simultaneous equations to find b_σ and c_σ:

$$m_3 = \frac{\mathbb{C}[\sigma_\pi[w_x], G_x]}{\mathbb{V}[G_x]} - m_4 \frac{\mathbb{C}[G_x, G_y]}{\mathbb{V}[G_x]} \tag{B17}$$

$$m_4 = \frac{\mathbb{C}[\sigma_\pi[w_x], G_y]}{\mathbb{V}[G_y]} - m_3 \frac{\mathbb{C}[G_x, G_y]}{\mathbb{V}[G_y]} \tag{B18}$$

Simultaneously solving Equations (B17) and (B18) obtains:

$$m_3 = 0 \tag{B19}$$

$$m_4 = (\eta - 1)\sigma_{oo} \tag{B20}$$

m_3 is the partial regression slope $\beta_{\rho_x\sigma_\pi[w_x],G_x}$, which provides the component c_σ in Inequality (A25). m_4 is the partial regression slope $\beta_{\rho_x\sigma_\pi[w_x],G_y}$. The component b_σ in Inequality (A25) is equal to $-\beta_{\rho_x\sigma_\pi[w_x],G_y}$. Accordingly, these two components are:

$$c_\sigma = 0 \tag{B21}$$

$$b_\sigma = (1 - \eta)\sigma_{00} \tag{B22}$$

ν is a simple function of allele frequency p:

$$\nu \equiv \frac{\sigma_\pi[\overline{w}]}{\mathbb{E}_\pi[\overline{w}]} = \frac{\big(p\eta + (1-p)\big)\sigma_{00}}{\mu_{00} - pc} \tag{B23}$$

This is an intuitive measure of stochasticity in this environment fluctuating unpredictably between two states: the numerator is the standard deviation of two completely correlated random variables (i.e. the sum of $\eta\sigma_{00}$ and σ_{00}, weighted by the frequency of each allele), whilst the denominator is the average number of offspring across states (again, weighted by allele frequency).

Since p appears in the equation for v, v is frequency-dependent. Differentiating stochasticity (v) with respect to the frequency of altruistic bet-hedgers (p), v decreases with rising p when:

$$1-\eta > \frac{c}{\mu_{00}} \tag{B24}$$

Accordingly, stochasticity v falls as frequency p rises (*Extended Data* Fig. E2) if the effect of variation reduction ($1-\eta$) is greater than the relative size of mean-fecundity reduction ($\frac{c}{\mu_{oo}}$).

When this condition (Inequality (B24)) is met, as the frequency p rises the bet-hedgers begin to render the environment effectively stable. At high frequency, the value of volatility-reducing altruism b_σ therefore falls, because v is low. The result is that the population can be reinvaded by familiar mean-fecundity-maximisers, much as a 'conspiracy of doves' in the well-known hawk–dove game is vulnerable to invasion by hawks[44]. (At low costs (c), intermediate levels of variation reduction η are less constrained from reaching fixation.) Connections between coexistence and bet-hedging have been analysed in non-social settings[45], although not interpreted in terms of frequency-dependent effects on whole-population stochasticity.

B2 | Frequency at which expected change due to selection is zero

We now have all the components of Hamilton's rule ($r=\alpha$, $c_\mu = c$, $b_\mu = 0$, $c_\sigma = 0$, $b_\sigma = -(\eta-1)\sigma_{00}$,$v = \frac{(p\eta+(1-p))\sigma_{00}}{\mu_{00}-pc}$). Putting these components together, we solve for the frequency p^* at which there is no expected change due to selection ($\mathbb{E}_\pi[\Delta\bar{G}] = 0$):

$$p^* = \frac{\alpha(\eta-1)\sigma_{00}^2 + c\mu_{oo}}{c^2-(\eta-1)^2\sigma_{00}^2\alpha} \tag{B25}$$

When $0 < p^* < 1$, the expected frequency of the social bet-hedgers is intermediate. If $rB > C$ for all p, then the population is expected to tend to $p^* = 1$. Likewise, if $rB < C$ for all p, the population is expected to tend to $p^* = 0$.

B3 | Individual-based simulation

To ensure that gene frequency makes incremental changes through generations in numerical simulation for the system in Table B1, we let offspring production across the population be driven by social interactions, and then sample a random 1% of the adult population for replacement in proportion to the balance of genotypes amongst the offspring (i.e. each environmental state, 1% of the breeding spots become available for offspring produced that generation).

p^* is the gene frequency at which $rB - C = 0$ (*Extended Data* Fig. E1). The equilibrium frequency around which the population is expected to fluctuate over the long run, p', is equal to p^* when the changes in gene frequency that occur each generation are small and the sign changes from $rB > C$ below p^* to $rB < C$ above p^* (*Extended Data* Fig. E1c). The first condition reduces displacement from equilibrium: when the population takes extreme leaps in gene frequency each generation, gene frequencies can enter random cycles for which p^* is not the midpoint ($p^* \neq p'$), as gene frequency moves between extremely different values at which the slope of selection differs. Under a regime of weak selection, $p^* = p'$.

B4 | Effects of chance and autocorrelation in the fluctuating environment

Even if both states are equally probable, the environment may by chance have a run of several good or several bad states. At the predicted equilibrium p^*, the change $\Delta\bar{G}$ from a good state is exactly opposite to the change from a bad state, so the expected change in average genetic value is zero ($\mathbb{E}_\pi[\Delta\bar{G}] = 0$). However, the magnitude of the effect is important. If the two types of change $\Delta\bar{G}$ have a very large effect, the frequency of altruists may alter rapidly due to a chance sequence of many of the same states: a chance run of five bad years, for instance, might cause one genotype to crash completely. Sustained runs of the same environmental state are more probable when the environment fluctuates in a temporally autocorrelated fashion.

A stochastic population is predicted to occupy its polymorphic position p^* when $\mathbb{V}_\pi[\Delta\bar{G}] \approx 0$ (i.e. $\mathbb{V}_\pi\left[\mathbb{C}_x\left[G_x, \frac{w_x}{\bar{w}}\right]\right] \approx 0$) and p^* is convergence-stable (i.e. the frequency-dependent stochasticity coefficient v favours altruists at frequencies p_t below p^* but selects against it above). Since selection favours altruists when $v(rb_\sigma - c_\sigma) > c_\mu - rb_\mu$, as long as $(rb_\sigma - c_\sigma) > 0$ we can divide by $(rb_\sigma - c_\sigma)$ without changing the sign of the inequality to find the conditions for a globally convergence-stable population in terms of the magnitude of the stochasticity coefficient v at a given frequency p_t (denoted v_{p_t}):

$$\begin{aligned} p_t < p^* \Rightarrow v_{p_t} > \frac{c_\mu - rb_\mu}{rb_\sigma - c_\sigma} \\ p_t > p^* \Rightarrow v_{p_t} < \frac{c_\mu - rb_\mu}{rb_\sigma - c_\sigma} \end{aligned} \qquad (B26)$$

In the individual-based simulation, we focus on weak selection, where only 1% of the population's genotype frequencies are available to change each generation. Under weak selection, even high levels of temporal autocorrelation (leading to frequent runs of the same environmental states across years) do not necessarily deter the population from its convergence point. In general, we emphasise that the Price equation – and its derivation, Hamilton's rule – focuses on generational changes: accordingly, both the non-stochastic version familiarly used in the literature and the stochastic version presented here can predict the frequency at which there is no change or no expected change (respectively) due to selection. Under appropriate conditions, including low-amplitude fluctuations in allele frequency between generations, this frequency will be realised as an equilibrium state for the population; outside these conditions, the frequency at which there is no expected change due to selection need not represent an equilibrium state.

B5 | Inducible altruism

An actor in a fluctuating environment does not necessarily need to produce a 'constitutive' strategy (e.g. help in all states or defect in all states). If the actor possesses information about the current state π, it may be able to tailor its response to produce an optimal strategy for the given state. In principle, this form of phenotypic plasticity may produce 'inducible' altruism in a stochastic world: help relatives if you know that a drought is imminent, for instance. In this section, we show how the reliability of information in a stochastic world determines whether cooperation should be constitutive or inducible.

We introduce to the population a plastic allele I, such that there are three alleles in competition:

S: 'Selfish': carriers never cooperate

C: 'Constitutive cooperator': carriers cooperate in all states

I: 'Inducible cooperator': carriers cooperate only when they believe they are in the 'bad' state

These alleles have frequencies p_S, p_C, and p_I respectively (i.e. $p_S + p_C + p_I = 1$).

Let an act of cooperation incur a cost c to the actor's fecundity. In 'bad' states (such as drought), receiving cooperation increases an individual's fecundity by δ_+. In 'good' states, we allow the presence of a cooperator to be detrimental to the recipient's fecundity: the cooperator reduces recipient fecundity by δ_- (note that δ_- can equal zero, or even be negative if the co-operator always benefits the recipient).

Let the plastic allele I pay an additional cost ($c_{plastic}$) as the 'cost of plasticity', determined by both the costs of information gathering and utilisation and the costs of maintaining behavioural flexibility. The quality of the information available to carriers of the plastic allele is determined by its accuracy A: the environmental state π is what the actor thinks it is with probability A.

The frequency of each type of pairing is as follows (Supplementary Table B2):

Supplementary Table B2 | Frequencies of interactions

Genotype of focal individual x (allele carried)	Genotype of partner individual y (allele carried)	Frequency (F_{xy})
I	I	$p_I^2 + \alpha p_I(1 - p_I)$
I	S	$(1 - \alpha)p_I p_S$
I	C	$(1 - \alpha)p_I p_C$
S	I	$(1 - \alpha)p_S p_I$
S	S	$p_S^2 + \alpha p_S(1 - p_S)$
S	C	$(1 - \alpha)p_S p_C$
C	I	$(1 - \alpha)p_C p_I$
C	S	$(1 - \alpha)p_C p_S$
C	C	$p_C^2 + \alpha p_C(1 - p_C)$

We can define the fecundity of each of the types of focal individual described in Supplementary Table B2 as follows (Supplementary Table B3):

Supplementary Table B3 | Fecundities of types of focal individual

Focal individual x	Partner y	Fecundity in good years $(Good_{xy})$	Fecundity in bad years (Bad_{xy})
I	I	$Az_1 + (1 - A)(z_1 - c - \delta_-) - c_{plastic}$	$A(z_2 - c + \delta_+) + (1 - A)z_2 - c_{plastic}$
I	S	$Az_1 + (1 - A)(z_1 - c) - c_{plastic}$	$A(z_2 - c) + (1 - A)z_2 - c_{plastic}$
I	C	$A(z_1 - \delta_-) + (1 - A)(z_1 - c - \delta_-) - c_{plastic}$	$A(z_2 + \delta_+ - c) + (1 - A)(z_2 + \delta_+) - c_{plastic}$
S	I	$Az_1 + (1 - A)(z_1 - \delta_-)$	$A(z_2 + \delta_+) + (1 - A)z_2$
S	S	z_1	z_2
S	C	$z_1 - \delta_-$	$z_2 + \delta_+$
C	I	$A(z_1 - c) + (1 - A)(z_1 - c - \delta_-)$	$A(z_2 + \delta_+ - c) + (1 - A)(z_2 - c)$
C	S	$z_1 - c$	$z_2 - c$
C	C	$z_1 - c - \delta_-$	$z_2 - c + \delta_+$

In Fig. 4 of the main text, we plot the expected direction of change in allele frequency under selection for this population. Note that the stochastic Hamilton's rule identifies the points in frequency space $\{p_S, p_C, p_I\}$ at which each allele is expected to increase in frequency under selection.

An instructive empirical example is found in the temperate paper wasp *Polistes annularis*: field data for foundresses suggest that inclusive fitness is positive in a 'bad' state (characterised by drought) but negative in a 'good' state (when drought is absent)[46]. The existence of cooperative foundress groups in the 'good' state, when cooperation is predicted to be deleterious, implies that foundresses do not take up the theoretically-plausible option of being socially-plastic 'bad-year specialists'. In general, constitutive cooperation (cooperation in all states) can outcompete plastic cooperation ('bad-year specialists') when information is insufficiently reliable or the costs of plasticity are too high.

Appendix C | Feasibility of $b_\sigma > 0$

Hamilton's rule is predictive in the sense that it provides a falsifiable criterion to be applied to any specific hypothesis: a proposed combination of measured fitness effects must conform to the rule if they are to explain a given adaptation. In this section, we explore the potential for b_σ effects in social evolution.

In the main text and *Appendix A*, we highlight that the magnitude of v depends on the extent to which environmental fluctuations are correlated across patches in a matrix or metapopulation, and the extent to which temporal fluctuations within the organism's reproductive lifespan are correlated. Our intention here is to highlight the feasibility of b_σ-driven sociality, in principle, in the real world; at present, empirical data on the direct links between stochasticity and sociality are sparse. Direct empirical tests of the principle should aim to quantify the factors influencing v.

C1 | Elimination of parasite pressure

Recently, Rehan et al.[28] have found that observed mean fecundity effects ($rb_\mu - c$) are unable to explain the evolution of cooperation between sisters ($r = 0.75$) in a facultatively social bee (*Ceratina australensis*). This species inhabits a fluctuating environment, and Rehan et al.[38] have previously suggested that bet-hedging could drive the evolution of cooperation: parasite numbers rise and fall between generations, generating 'periods of extreme parasite pressure'[38], but social nests are better able to evade brood loss to parasites. Bees may be effectively blind to environmental state (ambient level of brood loss to parasitism), since parasitoid activity[38] occurs only once larvae and pupae are available for ovipositing. Whether pupae have been parasitized may be essentially unknowable, as they are sealed within the stem nest.

In this section, we model the evolution of sister-to-sister cooperation in a fluctuating world. Although we necessarily remain agnostic about the drivers of cooperation in the particular species *C. australensis*, we show that, in principle, highly stochastic environments (high v) can be more hospitable than static environments for sister–sister cooperation in such species when sociality buffers parasite pressure.

We obtain matching results through an individual-based haplodiploid simulation and an application of Inequality (A25) to the life-history parameters of Supplementary Table C1. To simplify the interpretation, we first consider a single diallelic haploid locus, with assortment following Gardner et al.[1]: individuals are matched with a social partner identical at the focal locus with probability α and a random partner with probability $1 - \alpha$. This obtains $r = \alpha$, which allows us to set $\alpha = 0.75$ to recover assortment levels between haplodiploid sisters. We let the environment fluctuate between high and low parasite states; a solitary individual has z_G offspring in a 'good year' (low parasite pressure) and z_B offspring in a 'bad year' (high parasite pressure, $z_B < z_G$). We let the presence of social partners buffer the breeder from parasite pressure, so that breeders with helpers attain z_G offspring regardless of environmental state.

Supplementary Table C1| Life history

G_x	G_y	Power	Result	Frequency in the population of this focal individual	Mean fecundity (across environmental states) of focal individual	Standard deviation (across environmental states) of focal individual's fecundity
1	1	Dominant	Queen	$\frac{1}{2}(p^2 + \alpha p(1-p))$	z_G	0
1	1	Subordinate	Worker	$\frac{1}{2}(p^2 + \alpha p(1-p))$	0	0
1	0	Dominant	Solitary	$\frac{1}{2}(1-\alpha)p(1-p)$	$dz_G + (1-d)z_B$	$\sqrt{d(1-d)(z_G - z_B)^2}$
1	0	Subordinate	Worker	$\frac{1}{2}(1-\alpha)p(1-p)$	0	0
0	1	Dominant	Queen	$\frac{1}{2}(1-\alpha)p(1-p)$	z_G	0
0	1	Subordinate	Solitary	$\frac{1}{2}(1-\alpha)p(1-p)$	$dz_G + (1-d)z_B$	$\sqrt{d(1-d)(z_G - z_B)^2}$
0	0	Dominant	Solitary	$\frac{1}{2}((1-p)^2 + \alpha p(1-p))$	$dz_G + (1-d)z_B$	$\sqrt{d(1-d)(z_G - z_B)^2}$
0	0	Subordinate	Solitary	$\frac{1}{2}((1-p)^2 + \alpha p(1-p))$	$dz_G + (1-d)z_B$	$\sqrt{d(1-d)(z_G - z_B)^2}$

Solving for the coefficients in Inequality (2) of the main text obtains the following, where $\mu_\bullet$ and $\sigma_\bullet$ are the average and standard deviation respectively (across the two states) of a solitary individual's number of offspring. Detail about obtaining regression coefficients for social effects is provided in *Appendix B*.

$$b_\mu = \frac{(1-d)(z_G - z_B)}{2} \qquad (C1a)$$

$$c_\mu = \frac{\mu_\bullet}{2} \qquad (C1b)$$

$$b_\sigma = \frac{\sigma_\bullet}{2} \qquad (C1c)$$

$$c_\sigma = -\frac{\sigma_\bullet}{2} \qquad (C1d)$$

The means-based Hamilton's rule implies that cooperation will not evolve by mean fecundity effects alone for this system. The condition for the evolution of altruism by mean fecundity effects is:

$$r(1-d)(z_G - z_B) > \mu_\bullet \qquad (C2)$$

When high-parasite and low-parasite years occur with equal frequency ($d = 0.5$), the critical ratio of $z_G{:}\,z_B$ ($= \frac{(1-d)(1+r)}{(1-d)r-d}$) is negative: even with helpers conferring substantial gains in fecundity on breeders in high-parasite years (Table B1), cooperation cannot evolve by mean fecundity effects. When low-parasite states occur in 40% of years ($d = 0.4$), cooperation only evolves due to mean fecundity effects if individuals have at least 21 times more offspring without parasites than with parasites.

However, incorporating volatility effects increases the scope for cooperation when the environment is stochastic (high v):

$$r\big((1-d)(z_G - z_B) + v \cdot \sigma_{\bullet}\big) > \mu_{\bullet} - v \cdot \sigma_{\bullet} \qquad (C3)$$

In Fig. 3 of the main text, we illustrate this increased scope for the evolution of cooperation, both in terms of Inequality (A25) and individual-based simulation. For instance, whilst equal frequencies of high- and low-parasite years are unable to sustain cooperation by mean fecundity effects at any level of z_G and z_B, Fig. 3a reveals a high-stochasticity region in which cooperation invades a solitary population due to volatility effects. The 21-fold difference in fecundity between high-parasite and low-parasite states required for the evolution of cooperation by mean fecundity effects when low-parasite states occur in 40% of years shrinks to a 3-fold difference with the addition of volatility effects. Volatility effects can, accordingly, extend the region of the adaptive landscape in which social traits evolve, and in principle reduce the gap between B and C in paradoxical cases where Hamilton's rule appears to fail. Not all social species evolve from solitary ancestors inhabiting a highly stochastic world, but those that do may in principle obtain hidden b_σ and c_σ effects that increase the payoff from cooperation. Note that when high-parasite states are very frequent, b_μ effects rise in power: when parasites constantly threaten the population, and helpers eliminate parasite pressure, mean fecundity is increased; in this situation, the environment is no longer stochastic (low v). Incorporating volatility effects means that cooperation can still evolve when high-parasite states are not extremely frequent.

C2 | Galapagos mockingbirds

Empirical data are sorely lacking for testing the effects of b_σ. One encouraging dataset, however, comes from the cooperatively breeding Galapagos mockingbird (*Mimus parvulus*). Curry and Grant[47] recorded demographic information over an 11-year period on Isla Genovesa (Ecuador). Helping is polymorphic in *M. parvulus* (occurring at 34% of nests), allowing a comparison of cooperative and non-cooperative nesting attempts.

Using the relevant summary statistics in Curry and Grant[47] (based on 153 helper-present nests and 297 helper-absent nests), we estimate partial regressions of expected recipient fecundity against actor phenotype (helper or non-helper). We play the 'phenotypic gambit', and adopt a phenotypic (as opposed to genotype) variant of the stochastic Hamilton's rule. We therefore regress fitness components against the focal individual's phenotype P_x and the phenotype of a social partner P_y, and we assign the phenotypic values 0 and 1 for non-helping and helping respectively:

$$b_\mu = \beta_{\mathbb{E}_\pi[w_x],P_y} = 0.3 \qquad (C6)$$

Sample size varies considerably between years (from two helper-attended nests in 1984 to 33 in 1987). We cannot calculate b_σ directly from the data, therefore, as we cannot distinguish 'true' population variance from sampling variance. Instead, our approach is to ask whether a b_σ component can significantly change the estimated benefits of cooperation.

Galapagos mockingbirds inhabit a stochastic environment: Curry and Grant[47] report a coefficient of variation in fledgling production of 0.92 across years, a proxy for the coefficient of variation in average reproductive success $\left(\nu = \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}\right)$ across states of nature. We assume that helping has no effect on the volatility of the helper's own reproductive success $(c_\sigma = 0)$, and we consider the payoff for a sibling helper-at-the-nest $(r = 0.5)$:

$$0.5(0.3 + 0.92b_\sigma) > c_\mu \qquad (C7)$$

The cost of cooperation remains to be quantified in *M. parvulus.* If helpers suffer a loss of expected reproductive success exactly equal to the amount they increase the reproductive success of their recipients (i.e. $c_\mu = b_\mu = 0.3$), then (to two decimal places):

$$b_\sigma > 0.33 \qquad (C8)$$

The regression of recipient fecundity volatility against actor phenotype $(\beta_{\rho_x \sigma_\pi[w_x],P_y})$ must have a slope of at least -0.33 to justify altruism if $b_\mu - c_\mu = 0$. The upshot is that, in principle, b_σ can provide missing components of B in a sufficiently stochastic environment. Conclusively demonstrating altruistic bet-hedging in Galapagos mockingbirds will require (as with any empirical test of such models) elucidating how mockingbird-specific demography and population structure determines the relation between phenotype and the separate components of fitness.

Risk plays an important role in behavioural ecology. A stochastic approach is useful even if risk-management strategies affect the mean reproductive successes[40] of actors or their social partners (c_μ and b_μ respectively) without affecting the reproductive variation of either individual. In the social insects, for instance, the so-called 'Wenzel-Pickering effect' proposes that larger groups are able to reduce the variation in the supply of food for the brood, preventing shortfalls in which brood would otherwise die[27,48]. Whether the Wenzel-Pickering effect in real organisms derives its benefit from a consequent reduction in the variation of the production of offspring[13] ($b_\sigma > 0$), an increase in mean offspring production[48] ($b_\mu > 0$), or a combination of both ($b_\mu > 0$ and $b_\sigma > 0$) remains unknown. Similarly, in the mockingbirds, nesting attempts may be more 'risky' in a given state π: this risk may mean that only a proportion of nests will succeed. This more proximate form of 'risk' differing between years influences the payoffs from social behaviour in each type of year, and therefore can affect both expected fecundity and the volatility of fecundity across states. Classifying benefits accruing to different statistical parameters in the stochastic Hamilton's rule offers a framework for diagnosing these diverse forms of risk-management benefits and costs.

Additional References

31. Price, G. R. Selection and covariance. *Nature* 227, 520–521 (1970).
32. Rice, S. H. A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evol. Biol*. 8, 262 (2008).
33. Queller, D. C. A general model for kin selection. *Evolution* 46, 376–380 (1992).
34. Birch, J. & Marshall, J. A. R. Queller's separation condition explained and defended. *Am. Nat*. 184, 531–540 (2014).
35. Rousset, F. Regression, least squares, and the general version of inclusive fitness. *Evolution* 69, 2963–2970 (2015).
36. Reed, J. & Walters, J. Helper effects on variance components of fitness in the cooperatively breeding red-cockaded woodpecker. *Auk* 113, 608–616 (1996).
37. Soucy, S. L., Giray, T. & Roubik, D. Solitary and group nesting in the orchid bee *Euglossa hyacinthina* (Hymenoptera, Apidae). *Insectes Soc*. 50, 248–255 (2003).
38. Rehan, S. M., Schwarz, M. P. & Richards, M. H. Fitness consequences of ecological constraints and implications for the evolution of sociality in an incipiently social bee. *Biol. J. Linn. Soc*. 103, 57–67 (2011).
39. Gonzalez, J.-C. T., Sheldon, B. C. & Tobias, J. A. Environmental stability and the evolution of cooperative breeding in hornbills. *Proc. R. Soc. B* 280, 20131297 (2013).
40. Frank, S. & Slatkin, M. Evolution in a variable environment. *Am. Nat*. 136, 244–260 (1990).
41. Okasha, S. & Martens, J. The causal meaning of Hamilton's rule. *R. Soc. Open Sci*. 3, 160037 (2016).
42. Young, W. E. & Trent, R. H. Geometric mean approximations of individual security and portfolio performance. *J. Financ. Quant. Anal*. 4, 179–199 (1969).
43. Goodnight, K. F. The effect of stochastic variation on kin selection in a budding-viscous population. *Am. Nat*. 140, 1028–1040 (1992).
44. Dawkins, R. The Selfish Gene. (Oxford University Press, 1976).
45. Seger, J. & Brockmann, H. J. What is bet-hedging? Oxford Surveys in Evolutionary Biology 4, 182–211 (1987).
46. Queller, D. C. & Strassmann, J. E. in *Reproductive success: studies of individual variation in contrasting breeding systems* (ed. Clutton-Brock, T.) 79–96 (University of Chicago Press, 1988).
47. Curry, R. L. & Grant, P. R. Demography of the cooperatively breeding Galapagos mockingbird, *Nesomimus parvulus*, in a climatically variable environment. *J. Anim. Ecol*. 58, 441–463 (1989).
48. Poethke, H. J. & Liebig, J. Risk-sensitive foraging and the evolution of cooperative breeding and reproductive skew. *BMC Ecol*. 8, 1–10 (2008).