Social runaway: Fisherian elaboration (or reduction) of socially selected traits via indirect genetic effects

Nathan W. Bailey[1,2] **and Mathias Kölliker**[3]

[1] Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, United Kingdom

[2] E-mail: nwb3@st-andrews.ac.uk

[3] Natural History Museum Fribourg, 1700 Fribourg, Switzerland

RUNNING TITLE

Social Runaway via Indirect Genetic Effects

AUTHOR CONTRIBUTIONS

NWB and MK conceived the idea for the study, constructed the models, and wrote the manuscript.

ACKNOWLEDGEMENTS

We are grateful to Graeme D. Ruxton and Michael B. Morrissey for valuable discussion and advice. Two associate editors, Louis Miguel-Chevin and Jarle Tufto, three anonymous reviewers, and the Editor in Chief Mohamed Noor provided helpful remarks that improved the study. NWB was funded by fellowships from the UK Natural Environment Research Council [NE/G014906/1 and NE/L011255/1].

DATA ACCESSIBILITY

No data are associated with the manuscript.

Social runaway: Fisherian elaboration (or reduction) of socially selected traits via indirect genetic effects

Our understanding of the evolutionary stability of socially-selected traits is dominated by sexual selection models originating with R. A. Fisher, in which genetic covariance arising through assortative mating can trigger exponential, runaway trait evolution. To examine whether non-reproductive, socially-selected traits experience similar dynamics—social runaway—when assortative mating does not automatically generate a covariance, we modelled the evolution of socially-selected badge and donation phenotypes incorporating indirect genetic effects (IGEs) arising from the social environment. We establish a *social runaway criterion* based on the interaction coefficient, ψ, which describes social effects on badge and donation traits. Our models make several predictions. (1) IGEs can drive the original evolution of altruistic interactions that depend on receiver badges. (2) Donation traits are more likely to be susceptible to IGEs than badge traits. (3) Runaway dynamics in non-sexual, social contexts can occur in the absence of a genetic covariance. (4) Traits elaborated by social runaway are more likely to involve reciprocal, but non-symmetrical, social plasticity. Models incorporating plasticity to the social environment via IGEs illustrate conditions favouring social runaway, describe a mechanism underlying the origins of costly traits such as altruism, and support a fundamental role for phenotypic plasticity in rapid social evolution.

KEY WORDS: **Altruism, Fisher process, parental investment, runaway, phenotypic plasticity, social selection**

Traits involved in social interactions are expected to have different evolutionary dynamics than non-social traits (West-Eberhard 1983). Such dynamics include how fast they evolve, whether different traits coevolve in a mutually reinforcing or antagonistic manner, how elaborated they can ultimately become, and their evolutionary persistence. While much theoretical effort has clarified the evolutionary dynamics of a wide variety of socially-selected traits over the last century, our understanding of the evolutionary *stability* vs. *instability* of socially-selected traits has arguably been dominated by sexual selection models. That is because for sexually-selected traits, conditions favouring instability – i.e. non-linear, exponential, or otherwise chaotic exaggeration or decay of traits – have been found to depend on genetic covariance arising through assortative mating, triggering exponential, "runaway" trait evolution. However, the potential for similar runaway dynamics when assortative mating does not automatically generate a covariance is less clearly understood. Here, we explore this scenario by modelling 'social runaway' in non-sexual, social traits using a quantitative genetic framework that accounts for several of the distinctive characteristics of such traits. Without loss of generality, we focus our analyses on a specific, tractable example of such traits, resource donation and elicitation, which are relevant for understanding many social processes in behavioural and evolutionary ecology such as parent-offspring conflict, sibling rivalry, altruism, cooperative breeding, and others.

One unique characteristic of traits with social functions is that they can generate social selection, that is, "differential reproductive success…due to differential success in social competition, whatever the resource at stake" (West-Eberhard 1983, p. 158). Another is that the social environment an individual experiences can influence the expression of traits under selection (as opposed to influencing the selection itself), through socially-mediated plastic responses (Cardoso et al. 2015). For example, variation in the social environment can affect the expression of: aggression during agonistic encounters (Rodenburg et al. 2008), mate preferences (Collins 1995, Hughes et al. 1999), parental care (Royle et al. 2012), social learning (Battesti et al. 2012), and offspring solicitation in species where parents or helpers provide care (Velando et al. 2013; Mas and Kölliker 2008). Social environments may vary because of genes that are expressed by interacting social partners, and indirect genetic effects (IGEs) arise when individuals experiencing such social environments express different trait values as a result (Moore 1997). The evolutionary consequences of IGEs can be distinct and unusual, because the social environment may contain a genetic component that itself evolves. For this reason, an IGE framework is useful for exploring evolutionary dynamics of social traits such as aggression, reproduction, conflict and cooperation (Wolf et al. 1998).

The central predictions of numerous verbal models of social evolution are that social behaviours should experience evolutionary dynamics that are frequently more rapid, more volatile, and more susceptible to influencing macroevolutionary patterns via divergence and diversification (West-Eberhard 1983, 1989, 2006). However, debate about the distinctiveness of evolutionary processes involving social interactions has now persisted for well over a century. In the late 1800's, Baldwin (1896) suggested a major role for "psycho-genetic" modifications—social flexibility in contemporary terminology—in producing adaptive responses at the level of an individual responding to its social environment, in addition to influencing macroevolutionary patterns. Seminal publications thereafter (reviewed by: Wcislo 1989; West-Eberhard 2006; Duckworth 2009; Bailey 2012; Bailey et al. 2018) have reinforced the idea that phenotypic plasticity (e.g. developmental plasticity, learning, behavioural flexibility) arising from variation in the social environment can cause evolutionary feedback that exaggerates evolutionary rates upward or downward. The influence of this debate has been strongest in the field of sexual selection, which can be considered a special form of social selection (Lyon and Montgomerie 2012), and a central focus has been on unstable conditions generated during trait-preference coevolution, the 'runaway' process first described by Fisher (1915; 1958) and later mathematically formalised by Lande (1981).

As envisioned by Fisher, the runaway process comprises two steps: the evolution of a stable equilibrium between two traits, followed by rapid coevolution resulting in exponential elaboration or diminution of each trait until checked by countervailing natural selection. Fisher verbally described this process as starting when, stochastically, a male trait variant conferred an initial survival advantage, was then perceived by females and elaborated through the action of female choice—becoming an ornament—then subsequent genetic covariance built up as a result of assortative mating and gametic phase disequilibrium even though the ornament conferred a fitness cost to its bearer and no direct benefit to females (Fisher 1915, 1958). The key characteristic of runaway is that it describes an unstable process and results in exponential trait elaboration or reduction depending primarily on the genetic variances and covariance for male and female traits (Lande 1981, Kirkpatrick 1982).

Fisher-like runaway instabilities or unstable cycling in the evolution of other social traits have been proposed to explain evolutionary trait elaboration in social contexts (West-Eberhard 1989; Kölliker et al. 2000; Kölliker and Richner 2001). Nesse (2007) provides an overview of social selection models, and argues that runaway dynamics arising from evolutionary feedbacks may underpin much of social evolution for traits ranging from cooperation to human self-domestication. An earlier quantitative genetic model of social selection (Tanaka 1997) examined runaway dynamics for social responsiveness and signalling traits. It re-captured social trait dynamics analogous to the more specific case of sexual selection models in which the relative importance of genetic covariances in determining runaway conditions is related to models of female preference (Lande 1981). In the present study, we consider how genes expressed in the social environment impact evolutionary stability of socially-selected badge and donation traits. We consider the scenarios when i) the resource donation made by an individual depends on its experience of badge traits in social partners and ii) resource donation and badge traits reciprocally depend on each other. An example of the former occurs when a more intense begging behaviour elicits more provisioning by non-parental escorts, as has been found in a communally breeding mongoose, *Mungos mungo* (Bell 2008), and for the latter the more generally-found pattern of parental responsiveness to offspring begging and begging adjustments to obtained care (Kilner and Johnstone 1997). Partitioning quantitative genetic causes of trait variation into direct and indirect effects—DGEs and IGEs—provides a useful modelling framework to consider how dynamics arising from the social environment impact the stability of social selection, and allow us to derive testable, empirical predictions for behavioural and evolutionary genetics research.

Runaway Social Selection Model

Defining the conditions under which unstable evolution is expected is a separate goal from establishing the expected values of coevolving traits in a population at equilibrium. In some cases, the only equilibrium might occur when neither trait is expressed, i.e. both average trait values are zero. Alternatively, there might be several equilibria, which could include a lack of trait expression, separated by fitness valleys, in which case focussing on equilibria does not provide answers to how traits shift from one equilibrium to another, an issue pointed out by Rodríguez-Gironés et al. (1996) in the context of offspring begging and parental care. If the conditions for unstable runaway are satisfied, it is theoretically possible to observe rapid exponential evolution characteristic of the unstable phase of the Fisher process and, in cases of multiple equilibria, social runaway may drive trait evolution through the fitness valleys.

In this study, we examined whether evolutionary instabilities that characterise the Fisher process also contribute to the emergence and elaboration of an altruistic social trait, and whether such dynamics support the proposed rapid evolution of social traits in general (West-Eberhard 1983). We adapted a quantitative genetic modelling approach to include feedback arising from the social environment through IGEs, focusing on the stability of coevolutionary outcomes between a 'donation' trait that enhances the fitness of a conspecific and potentially decreases the fitness of its bearer and a 'badge' trait that elicits such donation in other individuals (González-Forero and Gavrilets 2013). In a sexual selection context, the genetic covariance between male ornament and female preference is of central importance for a runaway process. In a non-sexual, social context, satisfying this condition is more difficult. Linkage disequilibrium might be facilitated through population viscosity favouring assortative mating among badge loci-carrying and donation loci-carrying individuals (Biernaskie et al. 2011), or correlational selection gradually tightening physical linkage of relevant loci into coadapted gene complexes (Sinervo et al. 2006; Kölliker et al. 2012).

Our model makes three important distinctions relevant to this problem. First, we considered the genetic architecture of badge and donor traits to be quantitative and not sex-limited, and therefore influenced by the combined effects of many loci. Second, donor traits are, at least initially, assumed to have an independent genetic architecture from badge traits and are therefore genetically unlinked and freely recombining. We assume no population genetic structure or genetic relatedness, and consider a population of freely-mating diploid individuals. Third, we model IGEs on donor and badge traits, to assess the consequences of genetic variance arising from the social environment. We then focus on the questions of whether coevolution of donor and badge traits can occur via runaway dynamics, and whether IGEs arising from the social environment influence the scope for such social runaway.

DEFINING TRAITS

An individual's propensity to donate resources is described by the trait z_d. Such donation would incur a direct fitness cost to the donating individual, for example by removing food or habitat resources that could otherwise be allocated to survival or reproduction. A recipient individual's ability to gain resources from others is denoted z_b. Specifically, z_b refers to a quantitative badge trait that elicits resource provisioning by the donor. It is important to note that badges can take many forms in animals. An elicitation badge could be morphological, such as a colouration (e.g. Hunt et al. 2003); an acoustic signal, such as a begging vocalisation (e.g. Muller and Smith 1978); a behaviour or pheromone, such as the stereotyped movements and physical contact of juvenile burying beetles, *Nicrophorus vespilloides* or the cuticular hydrocarbons of juvenile earwigs *Forficula auricularia*, used to solicit food from parents (Mas and Kölliker 2008). Similar to other animal signals, physical badges could vary in chromatic spectrum, contrast, intensity, size, or match to an internal template or pattern (e.g. Lynn et al. 2019). For conceptual simplicity, we refer to badges as a physical property of an animal which can take different 'sizes', i.e. values, but behaviour such as begging could be similarly conceived as an elicitation signal, substituting for the badge trait z_b in our model. Traits are influenced by additive genetic effects, a, and all other non-additive and environmental effects, e, following standard quantitative genetic formulations:

$$z_d = a_d + e_d \tag{1a}$$

$$z_b = a_b + e_b \tag{1b}$$

We consider the case where the expression of donation is affected not only by additive genetic effects and abiotic features of the environment, but where it is also sensitive to the social environment in which it is expressed. Indirect genetic effects generated by such social flexibility can be illustrated by decomposing the environmental contribution to z_d as follows:

$$e_d = e_{d_p} + e_{d_s} \tag{2}$$

The social environment, e_{d_s}, comprises interactions that one focal donating individual has with a conspecific partner, and e_{d_p} comprises all other non-social effects, including stochastic environmental and developmental noise. For simplicity, we assume a single dyadic interaction occurs between these individuals. In general, therefore, the social environment can contain a genetic component arising from genes expressed by interacting partners (Moore et al. 1997). We consider the case where the propensity to donate resources depends on the interaction with the elicitation badge traits that are present in the social environment:

$$e_{d_s} = \psi_{bd}\, z'_b \tag{3}$$

The prime indicates that the effect arises from an interacting partner, and ψ_{bd} is a linear coefficient where $-1 < \psi_{bd} < 1$ and the subscript indicates that trait z_b influences the expression of z_d via the social environment. When ψ_{bd} is positive, a larger badge value in a social partner causes increased donation over and above the direct genetic effects on donation expressed by the focal individual. Alternatively, ψ_{bd} would be negative if the efficacy of an elicitation badge decreases with badge size, which might be expected in the case of retaliation by potential donors (Royle et al. 2002). When ψ_{bd} is zero, the social environment has no effect on donor trait expression and correspondingly there is no IGE. Consistent with prior models (reviewed in Bailey et al. 2018), we treat the interaction coefficient as a fixed parameter. IGEs themselves are likely to evolve (Kazancioğlu et al. 2013), but our aim here is to evaluate conditions under which IGEs are likely to drive runaway dynamics, as opposed to modeling ψ as an evolutionary outcome. In the following we present two models, the first assuming that the donation trait responds to the size of the elicitation badge, but that the elicitation badge is not responsive to the amount of donations received. The second model allows for reciprocal interactions where the donation traits responds to badge size, and badge size in turn responds to the amount obtained from donors.

UNILATERAL EFFECT OF BADGE SIZE ON DONATION

By substitution, the donation phenotype affected by IGEs, and the badge trait unaffected by IGEs, can be expressed as follows:

$$z_d = a_d + e_{d_p} + \psi_{bd}(a'_b + e'_{b_p}) \tag{4a}$$

$$z_b = a_b + e_{b_p} \tag{4b}$$

The primes indicate effects arising from the interacting partner, scaled by the interaction coefficient ψ_{bd}. We assume physical environment effects are independent of additive effects and normally distributed with a mean of zero (Falconer and Mackay, 1996), and the distinction between focal and interacting individuals identified in eqns. (4) disappears when averaging across all individuals in a population (Moore et al. 1997). The expectations for mean trait values are then:

$$\bar{z}_d = \bar{a}_d + \psi_{bd}\bar{a}_b \quad (5a)$$

$$\bar{z}_b = \bar{a}_b \quad (5b)$$

SELECTION ON TRAITS

We considered sources of selection on z_d and z_b (Figure 1). In general, natural selection should disfavour trait values that incur fitness costs. The marginal fitness costs associated with altruistic behaviours tend to increase at ever-increasing rates, such that the costs of a unit increase in the range of small donations may be relatively negligible but a unit increase in the range of large donations detrimental (Brown and Vincent 2008). We correspondingly define the donor fitness function using a generalized Gaussian function borrowed from the sexual selection literature (Iwasa and Pomiankowski 1995):

$$W_d = e^{-cd^4} \quad (6a)$$

Positive versus negative sign of the donation trait value indicates whether donors contribute to partners with larger or smaller than average badge trait values, respectively. For analogous treatments see Lande (1981) and Pomiankowski and Iwasa (1993). The cost of donating resources increases as an individual donates more resources. Initial contributions have a relatively small impact on fitness, but this increases as the donation size increases, and the steepness of that fitness decrease is captured by the coefficient c. This fitness function produces a nonlinear selective force (Lande 1981) that can ultimately stabilise exponential evolution of donation sizes if elicitation badges become overly-effective. We later assess the sensitivity of the model to donation costs by substituting a steeper fitness function.

The total fitness consequence of elicitation badges depends on their viability costs and socially-selected benefits and is defined by the fitness function:

$$W_b = e^{k\bar{d}(b-\bar{b})}\, e^{-mb^2} \quad (6b)$$

Here, an interacting partner's fitness gain is represented in the first exponential term of equation (6b), and depends not only on the value of his or her badge, but also on the relative deviation from the population average badge value $(b - \bar{b})$. This is further scaled by the average donation across the population, $\bar{d}$, and a scaling constant, k, which determines the strength of the association between the donation and recipient fitness. Badges are assumed to function as 'social' signals, and therefore experience natural selection in a manner analogous to other types of animal communication signals. Correspondingly, we adopt a symmetrical viability cost function around an optimum value $b = 0$ (Lande 1981, Pomiankowski & Iwasa 1993). Badge costs could include, for example, energetic expenditure in the case of begging displays, or enhanced predation costs in the case where a badge increases the conspicuousness of its bearer. These costs are reflected in the second exponent of equation (6b), where m describes how quickly these costs mount as the recipient's badge increases in intensity. Defining total fitness as the product of W_d and W_b and assuming weak selection (Iwasa et al. 1991), partial derivatives of the natural log of fitness functions with respect to each trait at the population means yield selection gradients:

$$\beta_d = -4c\bar{d}^3 \tag{7a}$$

$$\beta_b = k\bar{d} - 2m\bar{b} \tag{7b}$$

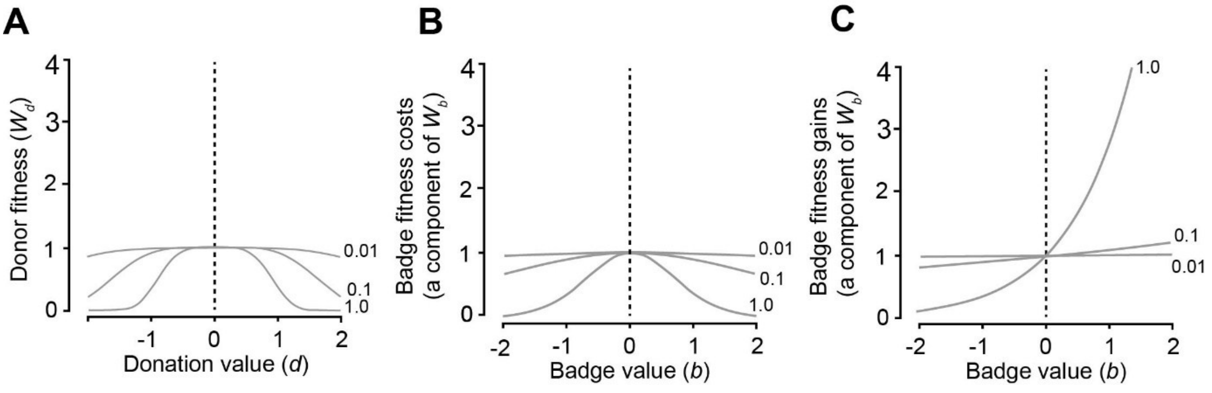

Figure 1. Badge and donation fitness components, with variation in scaling parameters illustrated. (A) Donation is costly for donors. The sign of the donation trait value indicates whether donors contribute to partners with larger or smaller badges than the population mean. Small donations carry relatively smaller costs than large donations, and the steepness of this relationship is determined by the scaling factor $\boldsymbol{c}$, for which three values are illustrated. (B) Badge-displayers pay symmetric fitness cost as badge values deviate from a viability optimum, here given by $\boldsymbol{b = 0}$. These costs are scaled by the constant $\boldsymbol{m}$. When $\boldsymbol{m}$ is small, selection against badges is weak and viability decreases slowly. (C) Badges elicit beneficial resource donation from others. Here, the fitness benefit of displaying a badge depends on how big the badge is relative to the population average badge size, in addition to a scaling factor $\boldsymbol{k}$,

which determines the marginal benefit of increased badge size. We illustrate the situation where the population mean badge value is equivalent to the viability optimum, i.e. $\bar{\boldsymbol{b}} = \mathbf{0}$. Fitness is highest for benefit functions with large $\boldsymbol{k}$, and when badges are large compared to the population mean. Note that the fitness benefits of displaying a badge are also scaled by the average donation trait value in the population, $\bar{\boldsymbol{d}}$, as in equation (6b), which we assume here for illustrative purposes to be positive and equal to 1. The significance of this is that if most donors in the population preferentially donate resources to holders of relatively small badges instead, then $\bar{\boldsymbol{d}} < \mathbf{0}$ and social runaway could drive the rapid diminution or evolutionary loss of elicitation badges as the direction of selection illustrated in panel C would be reversed.

TRAIT EVOLUTION

The per-generation change in average donor and recipient traits can be modelled using Price's theorem (Taylor 1996; Falconer and Mackay 1997). Following Moore and Pizzari (2005; eqn. 4) and McGlothlin et al. (2010; eqn. 4), the evolutionary expectation is obtained by examining the action of selection β_i on the covariance of the breeding value A_i and phenotypic value z_i for a given trait i:

$$\Delta\, \bar{z}_i = cov(A_i, z_i)\beta_i \tag{8}$$

When IGEs are present, the total breeding value for the focal donation trait z_d is comprised of both the direct additive genetic value for the trait plus the indirect additive genetic value arising from the interacting partner (Moore et al. 1997), thus:

$$\Delta\, \bar{z}_d = cov(a_d + \psi_{bd} a_b, z_d)\beta_d + cov(a_d + \psi_{bd} a_b, z_b)\beta_b \tag{9}$$

In this first model, evolution of the badge trait z_b is not affected by indirect genetic effects. Substituting the trait definitions from eqn. (4a) into eqn. (9), taking covariances and grouping terms, and similarly treating z_b, gives expressions for total evolutionary change for each trait:

$$\Delta\, \bar{z}_d = (G_{dd}\beta_d + G_{db}\beta_b) + \psi_{bd}(G_{bb}\beta_b + G_{db}\beta_d) \tag{10a}$$

$$\Delta\, \bar{z}_b = G_{bb}\, \beta_b + G_{db}\beta_d \tag{10b}$$

Here, G_{dd}, G_{bb}, and G_{db} are the additive genetic variances for donation and badge and their genetic covariance, respectively, and these are assumed to be constant. The first term in parentheses on the right side of equation (10a) illustrates the direct genetic component of evolution, and the second term illustrates the indirect genetic component modulated by ψ_{bd}, the influence of badges on donation. As can be seen from equation (10b), badge evolution in this model is unaffected by IGEs. Equations (10a) and (10b) provide a univariate model of donor and elicitation trait coevolution that incorporates a genetic covariance between the donor and elicitation traits, as well as IGEs caused by the influence of elicitation badges on donation propensity.

EQUILIBRIUM AND INSTABILITY

We assume that $\bar{z}_d$ and $\bar{z}_b$ are in equilibrium when there is no cross-generational evolutionary change in either trait, i.e. $\Delta\bar{z}_d = \Delta\bar{z}_b = 0$. If the equilibrium is locally stable, then a small perturbation of the population away from equilibrium will result in a return to that equilibrium. However, evolutionary runaway away from an equilibrium occurs when a population perturbed off that equilibrium continues to evolve away from it. We adapted methodology described in Hall et al. (2000) and elsewhere (Otto and Day 2007; Matthiopoulos 2011), to perform linear stability analyses. For illustrative purposes we redefine equations (10a) and (10b) in terms of a system of two ordinary differential equations describing the generational change in donation and badge traits:

$$\Delta\,\bar{z}_d = f(d,b) \qquad (11a)$$

$$\Delta\,\bar{z}_b = g(d,b) \qquad (11b)$$

Stability at the trivial equilibrium (0,0) in this system is more straightforward to analyse and can be generalised by the Hartman-Grobman Theorem (Hartman 1960), therefore we define an equilibrium state when $f\,(d^*,b^*) = 0$ and $g\,(d^*,b^*) = 0$. We model a perturbation in both traits by including a small deviation away from the trait values at equilibrium and then assessing whether the population evolves back towards equilibrium (stable) or away from it (unstable). Using Taylor expansion of eqns. (11a) and (11b) and disregarding higher-order terms expected to make negligible contributions to stability dynamics, we construct the Jacobian matrix $\boldsymbol{J}$ of first-order partial derivatives describing the incremental change in donor and badge traits around the equilibrium. Eigenvalues of the Jacobian at the equilibrium (d^*,b^*) give an indication about the stability of the point of equilibrium. The Jacobian at equilibrium, $\boldsymbol{J}^*$, is:

$$\boldsymbol{J}^* = \begin{bmatrix} G_{db}(k) + \psi_{bd}G_{bb}(k) & G_{db}(-2m) + \psi_{bd}G_{bb}(-2m) \\ G_{bb}(k) & G_{bb}(-2m) \end{bmatrix} \qquad (12)$$

The determinant equation $|\boldsymbol{J}^{*} - \lambda \boldsymbol{I}| = 0$ is solved to find eigenvalues of $\boldsymbol{J}^{*}$:

$$0 = \lambda^{2} - \lambda[G_{bb}(2m) - G_{db}(k) - \psi_{bd} G_{bb}(k)] \tag{13}$$

If any part of the eigenvalues λ_1 or λ_2 are positive, then instability around the point of equilibrium is indicated. Likewise, if any part contains a complex component, then the system can experience cyclical dynamics. The solution $\lambda_1 = 0$ corresponds to a lack of deterministic movement of the system at equilibrium (Lande 1981), so we focus on situations for which the root is positive, indicating instability. In the latter case $\lambda_2 > 0$ when:

$$\psi_{bd} + \frac{G_{db}}{G_{bb}} > \frac{2m}{k} \tag{14}$$

Inequality (14) defines the *social runaway criterion*: unstable dynamics during the joint evolution of donation and badge traits depends on genetic variance of badge traits, which will have a dampening effect on runaway, relative to the genetic covariance between donation and badge, which when strong will enhance the likelihood of runaway. It also depends on IGEs, indicated by the term ψ_{bd}. Crucially, runaway can occur when there is no genetic covariance, provided indirect genetic effects of badges on donation in interacting partners are strong and positive. However, the potential also exists for IGEs to counteract genetic covariance, for example if badges elicit an aversive reaction in interacting donors. The criterion also depends on how effective donation is at increasing badge-holder fitness (k) and the intensity with which badge costs mount with increasing badge size (m) (see Figure 1). Runaway is more likely when fitness is highly responsive to donation and when badge costs are relatively minor.

RECIPROCAL INTERACTIONS

There are situations in which both traits may be socially flexible, especially when both donor and badge trait are behaviours, such that donation by a target individual is influenced by badges they experience in the social environment, and elicitation itself depends on donor phenotypes that have been experienced. A clear example is offspring begging, where the eagerness with which offspring solicit food from parents depends on parental generosity (Kilner and Johnstone 1997; Royle et al. 2012). Modelling IGEs when both z_d and z_b are reciprocally affected by one another requires incorporating genetic influences on each trait's expression contributed by the other trait's presence in the social environment. We assume as before that environmental effects are independent of additive effects and normally distributed with a mean of zero, but that both focal and interacting individuals are capable of expressing badge and donation phenotypes. Trait values are thus partitioned into direct and indirect influences for focal individuals:

$$z_d = a_d + e_{d_p} + \psi_{bd}(a'_b + e'_{b_p}) \qquad (15a)$$

$$z_b = a_b + e_{b_p} + \psi_{db}(a'_d + e'_{d_p}) \qquad (15b)$$

As before, donation depends in part on IGEs arising from variation in badges within the social environment of donors, but the expression of badge traits is now also modified by IGEs arising from interaction with donors. ψ_{bd} describes how strongly interaction with badge traits affects the expression of donation, whereas ψ_{db} describes how strongly interaction with donor traits affects the expression of the badge, i.e. elicitation. It can then be shown (Moore et al. 1997, eqn. 12) that the evolutionary change in both traits is:

$$\Delta\, \bar{z}_d = \left(\frac{1}{1-\psi_{bd}\psi_{db}}\right)^2 \ [(G_{dd}\beta_d + G_{db}\beta_b) + \psi_{bd}(G_{bb}\beta_b + G_{db}\beta_d)] \qquad (16a)$$

$$\Delta\, \bar{z}_b = \left(\frac{1}{1-\psi_{db}\psi_{bd}}\right)^2 \ [(G_{bb}\beta_b + G_{db}\beta_d) + \psi_{db}(G_{dd}\beta_d + G_{db}\beta_b)] \qquad (16b)$$

Two features of this model are immediately apparent. The first is that IGEs affect the evolution of both traits, scaled by their respective interaction coefficients, and the second is that trait evolution now has an additional dependency on the interaction coefficients, which act multiplicatively in the denominator of the squared term of each expression. When ψ_{bd} and ψ_{db} are both large and positive (or both large and negative), the reciprocal IGEs on badge and trait are mutually reinforcing and they provide an additional enhancement to trait evolution. Likewise, if their directions oppose one another, IGEs on one trait may cancel the effects of IGEs on the other, in which case evolutionary responses remain the same as they would be without IGEs, even though IGEs are exerted on both traits. Such a situation might arise, for example, if donors positively respond to badge size, and badge size decreases when donations increase. To evaluate whether these different dynamics introduced by reciprocal donation and badge IGEs influence evolutionary stability, a linear stability analysis can again be performed.

Establishing stability criteria uses the Jacobian $\boldsymbol{J_R}$ (where the subscript indicates reciprocity of IGEs in this case), evaluated at the equilibrium $(d^*, b^*) = (0,0)$:

$$\boldsymbol{J}_R^* = \left(\frac{1}{1-\psi_{bd}\psi_{db}}\right)^2 \begin{bmatrix} G_{db}(k) + \psi_{bd}G_{bb}(k) & G_{db}(-2m) + \psi_{bd}G_{bb}(-2m) \\ G_{bb}(k) + \psi_{db}G_{db}(k) & G_{bb}(-2m) + \psi_{db}G_{db}(-2m) \end{bmatrix}$$

(17)

Solving the determinant equation $|\boldsymbol{J}_R^* - \lambda_R \boldsymbol{I}| = 0$ for conditions under which the leading eigenvalue $\lambda_{R(1)} > 0$ identifies conditions of evolutionary instability:

$$\frac{G_{db}+\psi_{bd}G_{bb}}{G_{bb}+\psi_{db}G_{db}} > \frac{2m}{k} \tag{18}$$

Requirements for unstable coevolutionary dynamics are more complicated when IGEs are reciprocal (eqn. 18; Figure 2). Provided $G_{bb} + \psi_{db}G_{db} \neq 0$, instability and runaway depend, as before, on the genetic variances and covariances between badge and donor traits, but they also depend on the characteristics of IGEs exerted on both traits, i.e. ψ_{bd} and ψ_{db}. Runaway is still possible in the absence of a genetic covariance, in which case evolutionary instability is expected when $\psi_{bd} > \frac{2m}{k}$. This inequality is the same as for the non-reciprocal IGE condition (Figure 2A). However, when a genetic covariance is present between badge and donation traits, it allows an influence of ψ_{bd}, and effectively scales this influence.

STEEPER DONATION COSTS

In the above two scenarios, we modelled social runaway using a donation fitness function which assumes initially negligible donation costs (Figure 1A). This was selected to reflect a situation in which, during the initial stages of coevolution between donation and badge traits, the negative fitness impacts of donation increase only shallowly until reaching a more critical threshold, reflecting donation traits that are relatively, and consistently, inexpensive to express for donors. However, cost functions themselves can evolve, and the fitness function may not remain the same over the course of the evolutionary increase in donation (Kölliker et al. 2012). We therefore examined stability dynamics of social traits experiencing non-reciprocal IGEs (the first scenario above) using a donor cost function in which fitness falls off more steeply as donation increases:

$$W_d = e^{-cd^2} \tag{19}$$

When the steeper donation cost function in equation (19) is adopted, it can be shown (*Appendix 1*) that the social runaway criterion becomes:

$$\frac{-2c[G_{dd}+\psi_{bd}G_{db}]}{G_{bb}} + \frac{k[G_{db}+\psi_{bd}G_{bb}]}{G_{bb}} > 2m \tag{20}$$

Conditions for evolutionary instability for donation and elicitation badge traits are now more complicated. When donor fitness is invariant regardless of donation size ($c = 0$), then the social runaway criterion is the same as that in expression (14). Otherwise, conditions for social runaway now depend on donation costs, reflected in the first part of the left-hand side of expression (20). As might be expected, the more steeply costs increase (i.e. the larger c is), then the lower the likelihood of unstable runaway trait evolution. Nevertheless, IGEs can play an important role in overcoming those costs. By evaluating the scenario presented in Figure 2, it can be shown that IGEs permit conditions favouring runaway even when donation costs are present under this model (*Appendix 2*). Thus, under more restrictive conditions in which donation costs mount more steeply as individuals donate more to badge-holders, the presence of strong IGEs can permit social runaway when it would otherwise not occur.

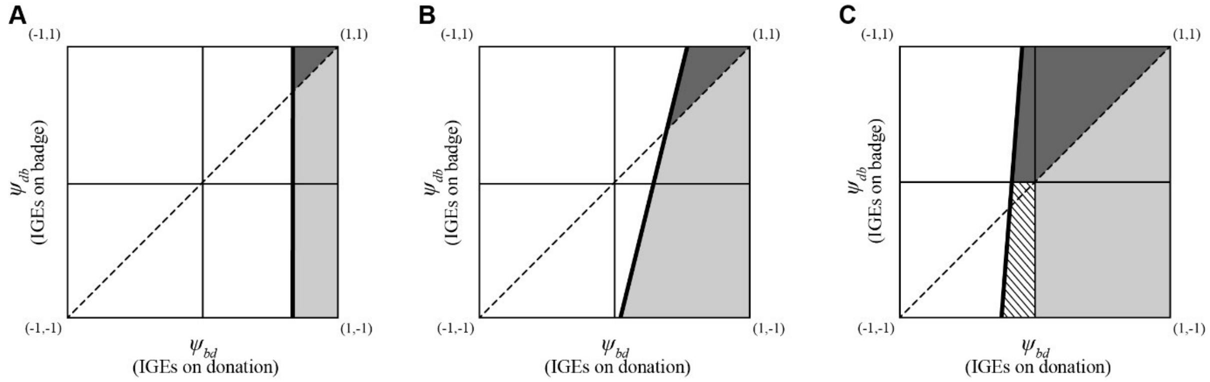

Figure 2. The influence of IGEs on potential for social runaway. IGEs describing the influence of donation by social partners ($\boldsymbol{\psi_{db}}$) are plotted against IGEs describing how badges in the social environment change the expression of donation behaviour ($\boldsymbol{\psi_{bd}}$). Shaded regions indicate where the social runaway criterion in expression (18) is satisfied, making unstable runaway coevolution between badge and donation traits possible. The dashed line indicates parity between $\boldsymbol{\psi_{db}}$ and $\boldsymbol{\psi_{bd}}$, and darker shading indicates instability when $\boldsymbol{\psi_{db}} > \boldsymbol{\psi_{bd}}$. (A) Runaway conditions when there is no genetic covariance between badge and trait. Here, $\boldsymbol{G_{bb}}$ = 0.4, $\boldsymbol{m}$ = 0.1, and $\boldsymbol{k}$ = 0.3, reflecting the expectation that the fitness costs of badges scale less rapidly than the fitness gains of expressing badges, i.e. $\boldsymbol{m} < \boldsymbol{k}$. In this scenario, the social runaway criterion is equivalent to that when IGEs are not reciprocal and only affect donation (expression 14), i.e. for any value of $\boldsymbol{\psi_{db}}$, instability occurs when $\boldsymbol{\psi_{bd}} > \frac{\boldsymbol{2m}}{\boldsymbol{k}}$. (B) Runaway conditions with a genetic covariance. Holding the other parameters in (A) constant, this scenario assumes a genetic covariance between badge and donation traits of $\boldsymbol{G_{db}}$ = 0.15. The existence of a genetic covariance further relaxes conditions for runaway, although instability tends to be predicted when $\boldsymbol{\psi_{bd}} > \boldsymbol{\psi_{db}}$, that is, when IGEs affecting donation are stronger than IGEs affecting badges. (C) Example of a plausible scenario in which unstable social runaway occurs when IGEs on both badge and donation traits are negative ($\boldsymbol{\psi_{db}} < \mathbf{0}$ and $\boldsymbol{\psi_{bd}} < \mathbf{0}$; striped shading). Such conditions are more likely to occur when the fitness costs of badges increase very slowly with increasing badge values,

compared to the rate of fitness gain associated with increasing badge values. In the illustration, we have set additive genetic variance to a moderate value of $\boldsymbol{G_{bb}}$ = 0.4, genetic covariance to be moderate to weak with $\boldsymbol{G_{db}}$ = 0.15, the scaling coefficient for badge costs $\boldsymbol{m}$ = 0.05 consistent with an assumption of weak selection, and the strength of association between donation and recipient fitness scaled by $\boldsymbol{k}$ = 0.5 (see Figure 1).

Discussion

THE SOCIAL RUNAWAY CRITERION

The foregoing models provide several general insights. First, they analytically demonstrate how Fisherian runaway processes can drive rapid evolution of socially selected traits outside the context of sexual selection. Second, IGEs can play an important role in overcoming barriers imposed by direct selection against resource donation phenotypes (z_d). The inequality represented by expressions (14) and (18)—the social runaway criterion—defines conditions under which unstable coevolutionary dynamics are predicted (Figure 2). The criterion is a positivity condition for the sum of two independent slopes, one describing the social interaction coefficient ψ_{bd} and the other the additive genetic regression slope of donation trait on badge trait, and this implies that social runaway can occur in the absence of genetic covariance between badge and donor traits. Instability in the neighbourhood of the equilibrium that we modelled, in which badges are at a viability optimum and donations are made to average-sized badge-holders (cf. Figure 1A), informs the evolutionary origins of socially-selected, potentially altruistic, traits. A steeper donation cost function impedes runaway (expression 20), though again this can be offset by IGEs acting on donating individuals. Our findings thus imply an important role for social runaway when donation has relatively mild negative impacts on donor fitness, which might be expected during the early evolutionary origins of altruistic traits (Le Galliard et al. 2005).

The social runaway criterion bears an obvious relationship to sexual selection models, in that the propensity for unstable evolutionary dynamics is influenced by the magnitude of genetic variance for badge traits and by the genetic covariance between badge and donation traits. However, even if we make a restrictive assumption that no such genetic covariance can be formed by assortative mating in the manner proposed by Fisher (1915, 1958) and modelled by others (Lande 1981, Kirkpatrick 1982), IGEs can nevertheless induce social runaway of badge and donor traits (Figure 2A). The influence of IGEs on this process is indicated by the interaction coefficient ψ_{bd}, and unstable conditions are more likely when ψ_{bd} is strong and positive. This is also generally the case when IGEs on donation and badge traits are reciprocal (Figure 2B). Assuming realistic conditions such as a moderate cost:benefit ratio of expressing a badge, moderate genetic variation in badge traits, and a weak genetic covariance consistent with estimates from the sexual selection literature (Greenfield et al. 2014), there is greater potential for runaway when IGEs on donation exceed those on badges, i.e., $\psi_{bd} > \psi_{db}$.

How likely is it that IGEs arise in cases of resource donation and elicitation, and that they are of a sufficient strength to affect the social runaway criterion? Our results highlighting the importance of IGEs arising from the social environment are consistent with theoretical findings that plasticity in partner choice can lead to runaway cooperation – but only until it reaches an optimum level where it is then counterbalanced by reduced payoffs (Geoffroy et al. 2019). Recent empirical studies suggest that IGEs on resource donation and elicitation can be of a considerable magnitude and affect the tempo and direction of evolutionary change. For example, in the burying beetle *N. vespilloides*, experimental evolution under different regimes of parental provisioning to offspring (donation) affected the heritability and response to selection of larval body size, indicating that genes in the social environment exert strong effects on both expression and evolution of a key fitness trait (Jarrett et al. 2017). These effects are mediated through resource donation and begging phenotypes (Jarret et al. 2017), and researchers are beginning to identify genes underlying such phenotypes, for example *neuropeptide F* (Cunningham et al. 2016). In another insect, the dung beetle *Onthophagus taurus*, females provision eggs by donating a brood mass of dung they have gathered, and IGEs have been shown to affect the size of this donation (Hunt and Simmons 2002). Finally, in earwigs *Forficula auricularia*, nymphs influence the likelihood that females produce a second clutch through a paternally inherited effect (Meunier and Kölliker 2012).

More recent empirical work suggests that IGEs may be widespread across different behavioural, morphological and physiological traits, and in some cases affect trait expression more strongly than DGEs. In outbred lab mice, Baud et al. (2017) detected IGEs in over a third of behavioural and fitness-related phenotypes assayed, and in nearly a fifth of those, IGEs had a greater influence on phenotypic expression than DGEs. Estimates of the interaction coefficient ψ have varied widely to date, suggesting significant scope for a variety of impacts on stability dynamics during social evolution. For example, values of ψ ranged from -0.96 to 0.93 in the guppy *Poecilia reticulata* (Bleakley and Brodie 2009), from -0.486 to 0.419 in a study of *Drosophila melanogaster* (Bailey and Hoskins 2014), and from 0.18 to 0.54 for different inbred genotypes of *D. melanogaster* (Signor et al. 2017). The considerable variation in ψ reported in the literature suggests that evolutionary dynamics are likely to occupy a large portion of the parameter space represented in Figure 2, including counterintuitive situations in which IGEs cause reduction of both badge and donation expression yet provoke unstable runaway (striped shading in Figure 2C), or when elicitation badges are more socially flexible than donation (dark shading in all panels of Figure 2). Acquiring additional estimates of the strength and direction of IGEs on socially selected traits is therefore a key priority which will advance our understanding of the way IGEs might – or might not – affect evolutionary stability of socially-selected traits.

Superficially, it would seem intuitive that reciprocal IGEs should increase the potential for unstable conditions, given their enhanced impact on evolutionary potential above and beyond that of unidirectional IGEs (Moore et al. 1997; Bijma 2014). However, closer examination suggests a robust biological interpretation for this finding. Strong and reinforcing IGEs would occur when *either*: greater badge values are stimulated by increased donation from interacting partners and interacting partners donate more when interacting with bigger badges, *or*: badges decrease with increasing donation and donation decreases with larger badges. In the former case, both interaction coefficients ψ are positive, and in the latter they are both negative, and standard IGE theory (eqns. 16a and 16b) predicts that evolutionary responses should be accelerated. Understanding the impact of IGEs on evolutionary potential under social selection does not in and of itself inform us about the stability of the associated evolutionary dynamics, however, as the latter is influenced by asymmetry in the strength of IGEs. Observations from other coevolutionary contexts support a role for asymmetrical partner interactions in causing dynamical instabilities. For example, in interspecific coevolution, asymmetrical species interactions can not only drive unstable coevolutionary dynamics, but also contribute to instability across broader ecological networks (Bascompte et al. 2006). It is notable that asymmetrical interactions in other coevolutionary systems modulate the destabilising effects of trait covariances, for example when interspecific interactions involve an exploiter species and a victim species (Débarre et al. 2014).

While IGEs are reciprocal in the second set of badge/donation models above—because both badge and donation traits are affected by IGEs—the effect of asymmetry in the strength of those IGEs on evolutionary stability is scaled by the genetic covariance in the social runaway criterion (eqn. 18). More generally, asymmetry of the Jacobian $\boldsymbol{J}$ predisposes unstable evolutionary dynamics (Leimar 2009); in the models above, this arises owing to both IGEs and the nature of selection operating on badges and donation. IGEs arising from social interactions between individuals expressing badge and donation traits are thus implicated in the evolutionary stability of those traits, but additional non-linearities, such as can be introduced by reciprocity of IGEs which do not scale linearly (Bijma 2014), or the evolution of ψ itself (Chenoweth et al. 2010; Bailey and Zuk 2012; Kazancioğlu et al. 2013; Marie-Orleach et al. 2017) can add further unpredictability to long-term evolutionary processes (Doebeli and Ispolatov 2014). Future work would benefit from relaxing the assumption that the interaction coefficient ψ is a fixed parameter and modelling the evolution of G_{bb} and G_{db}. Evaluating how the joint coevolution of DGEs and IGEs influences evolutionary stability would also be of key importance, given empirical evidence that phenotypic effects of genes in the social environment may coevolve with direct additive genetic effects (Pascoal et al. 2018).

The social runaway criterion is also determined by the overall strength of selection acting on badges, where m describes the gradient of natural selection against badges and k indicates how efficiently fitness benefits accrue to badge-holders as a result of resource donation by donors. Inequality (14) therefore represents a cost-to-benefit ratio in which unstable evolution is expected when direct genetic effects and IGEs contributed by the social environment override the fitness costs of expressing an elicitation badge. If badge traits are costly to produce, for example because of the need to sequester resources from the environment or divert allocation to other life-history traits, the threshold for unstable runaway dynamics will be increased and instability will become less likely. However, instability may be more likely at an evolutionary origin for donation and badge size when trait values are small (i.e. near zero), and badge costs are likely to be low.

Nevertheless, there is evidence for such costs, although mixed for some elicitation traits. Carotenoid pigments, for example, are costly to sequester and process into signals (Olson and Owens 1998). Although IGEs could enhance the likelihood of such costly donor traits evolving through unstable, runaway coevolutionary feedback with badge traits, they could equally inhibit such evolution if ψ_{bd} is negative, that is, when donors retaliate against large badge sizes. When might such a circumstance arise? If begging individuals evolve to manipulate and exploit donors, antagonistic selection may favour donor resistance. If resistance traits invade a population and depend on the detection of manipulative elicitation behaviours, it is plausible that z_d will tend to decrease with increasing values of z_b, that is, $\psi_{bd} < 1$. In other words, more vigorous begging will be met with increasing resistance, causing a negative feedback from the social environment that lessens the probability of unstable runaway.

GENETIC COVARIANCES AMONG SOCIALLY SELECTED TRAITS

Despite the similarity the social runaway criterion of expression (14) suggests to sexual selection models, there is a fundamental difference between Fisherian dynamics in sexually-selected versus socially-selected trait evolution. Sexual advertisement traits and preferences expressed by the opposite sex will automatically generate genetic covariances provided there is heritable variation underlying both trait and preference in a population. If trait and preference are pleiotropic effects of the same loci, the covariance is implicit. If they are not, then the covariance is generated by gametic phase disequilibrium arising as a result of assortative mating. The origins and build-up of this covariance have been discussed by Fisher (1915, 1958), Lande (1981), and numerous others (see Mead and Arnold 2004).

In the case of social runaway, the establishment of the covariance G_{db} is not as straightforward. Firstly, donor and badge traits may not be sex-limited and can thus be expressed within the same individual simultaneously or at different times or life history stages. Secondly, there exists no obvious automatic process of generating linkage disequilibrium between donation and badge loci. Tanaka (1996) suggested that the requirement for a genetic covariance may be minimal during runaway evolution under some conditions, but it is important to note that Tanaka used a different definition of runaway, as: “rapid evolution to a stable equilibrium” (p. 518), as opposed to a process destabilizing an equilibrium and driving the co-evolution of traits beyond their equilibrium values, *exponentially* elaborating or diminishing traits away from the equilibrium, cf. Fisher (1958), Lande (1981), and others. In addition, Tanaka’s (1996) models did not include indirect genetic effects, which in our model are key for unstable runaway co-evolution in the absence of a genetic covariance. Thus, while stochastic population processes may generate a genetic covariance between badge and donation trait sufficient to satisfy expression (14), this is not guaranteed to happen as an emergent property of the traits themselves, as it is in a system of coevolving sexual traits and preferences. The main hurdle for runaway social selection is thus for the terms on the left-hand side of expression (14) to overwhelm the cost-to-benefit ratio indicated on the right. Apart from the influence of IGEs, several potential factors may mitigate this problem.

The first is the link between social selection and kin selection. One possible source of direct genetic covariance between socially-selected traits is relatedness that arises in kin structured populations (e.g. among siblings inheriting genes underlying both donation and badge traits). These have been modelled elsewhere (e.g. McGlothlin et al. 2010; Queller 2011), but for our purposes it is illustrative to consider conditions under which relatedness contributes to runaway evolution. It has been previously shown that relatedness r and IGEs interact, such that the quantity $\frac{r+\psi}{1+r\psi}$ modulates the likelihood of altruistic traits evolving (expression 25 in McGlothlin et al. (2010)). Relatedness also has important implications for stability dynamics in our model, because it influences the expected genetic covariance between individuals (Lynch and Walsh 1998). For example, in the case of an elicitation badge and resource donation trait, the additive × additive epistatic covariance summed across all loci for a pair comprising a badge-bearing (x_b) and donating (x_d) individual, $\sigma_{AA(x_b,x_d)}$, contributes significantly to their total genetic covariance, $\sigma_{G(x_b,x_d)}$. This can be expressed in terms of the probability that genes from the two individuals are identical by descent, $\theta_{x_b x_d}$, and the total additive genetic variance associated with both traits (Lynch and Walsh 1998):

$$\sigma_{AA(x_b,x_d)} = (2\theta_{x_b x_d})^2 \, \sigma_{AA}^2 \qquad (21)$$

Thus, additive × additive epistatic covariance between loci influencing badge and donation traits will play a determining role in the magnitude of covariance between relatives (Lynch and Walsh 1998, pp. 144--145). Put another way, relatedness augments the genetic covariance between b and d, increasing the likelihood of exceeding the cost:benefit ratio defined by m and k in expression (14).

The likelihood of social runaway is therefore predicted to be enhanced under conditions of high relatedness, for example in strongly kin-structured populations, or under local inbreeding. Our models suggest these relatedness conditions are not so strict when IGEs are strong, suggesting an important role of IGEs during the evolutionary origin of altruism. A careful distinction from the generality of Hamilton's rule must be made: it is well-established that high relatedness favours the evolution of altruism. Our interest here is in the conditions favouring unstable evolutionary dynamics, and our argument implies that social runaway may also be favoured under high relatedness between donors and recipients. The relative impact of relatedness and IGEs in pushing badge and trait coevolution into instability is likely to be equivalent, given the symmetrical contributions of r and ψ (McGlothlin et al. 2010).

Additional mechanisms favouring a build-up of the direct genetic covariance G_{bd} include situations where badged recipients with strong elicitation traits preferentially associate socially with responsive individuals, or if responsive donors associate with eliciting recipients, both resulting in assortment of donor and badge trait values. The former is probably the more intuitive of the two. Even if mating is *per se* not assortative with respect to these two traits, social assortment each generation through behavioural preferences would create a genetic covariance if individuals tend to mate within groups formed through such social mechanisms. This behaviourally induced genetic covariance could then enhance the likelihood of social runaway, assuming that the behavioural preference remains stable over generations.An extreme case of this occurs with greenbeards, in which a genetic variant recognises copies of itself in other individuals and acts altruistically towards them. Under typical greenbeard scenarios, it is also possible for donor and badge traits to be controlled by different loci, but there is some debate about the degree to which strong linkage disequilibrium between them is required for greenbeard traits to successfully invade a population (see for example Jansen and Van Baalen 2006; Gardner and West 2010). Models of conditional helping behaviour also examined coevolution between resource donation traits and elicitation badges, i.e. phenotypic markers that indicate relatedness (Axelrod et al. 2004). Evolution of altruistic resource donation is also enhanced when loci underlying the traits are linked through gametic phase disequilibrium. Our model does not rely on badges conferring information about relatedness, but the influence of genetic covariance between the two traits has a common effect. In addition to representing a particularly stable form of kin structure, parent-offspring associations may also lead to a covariance through coadaptation due to selection favouring a range of equivalent combinations of offspring strategies to elicit, and parental strategies to provide, care, and there is now evidence for genetic covariances between provisioning and begging potentially driven by coadaptation across a range of species (Kölliker et al. 2012).

Predictions and Conclusions

Our findings can guide empirical work with four key predictions (Table 1). (1) IGEs and runaway dynamics may drive the evolution of donor and badge traits – i.e. altruistic interactions that depend on receiver elicitation signals more generally. (2) The social runaway criterion predicts that when examples of donation traits are detected in empirical study systems, they should be more susceptible to IGEs arising from social interactions than non-donation traits. The reason for this implication is that our model predicts that the unstable runaway dynamics associated with accelerated social trait elaboration are more likely to occur when strong, positive IGEs affect such traits (i.e. ψ_{bd}>0), or in cases where IGEs are reciprocal, when $\psi_{bd} > \psi_{db}$. Therefore, when we observe such traits, there is an expectation that their past evolution is more likely to have involved IGEs than for other types of traits. If it is the case that socially elaborated traits have arisen owing, in part, to the action of strong IGEs, such social lability may predispose the traits to manipulation and exploitation by social partners. (3) Genetic covariances are not required for social runaway, but they can enhance its likelihood. In a trivial sense, an enhanced likelihood of altruism in systems with significant kin structure and local relatedness is clearly predicted by Hamilton's rule and has been extensively studied (Hamilton 1964; Queller 2011). However, our model of social runaway provides the additional prediction that IGEs should be strong in such situations: kin-selected altruism, or altruism evolving through other mechanisms such as greenbeard effects should involve particularly socially labile traits. Intriguingly, traits such as cast determination in the social hymenoptera are among the most dramatic examples of socially-cued phenotypic plasticity (Huang and Robinson 1996), and neurogenesis in the eusocial mammal, the naked mole rat, has also been found to be highly susceptible to the social environment (Peragine et al. 2014). (4) Reciprocity of IGEs does not have the same effect as symmetry in the strength of IGEs, but asymmetrical strength of IGEs increases the likelihood of evolutionary instability and social runaway. As a result, we predict that elaborate, socially selected traits should tend to show reciprocal, but asymmetrical, IGEs.

The contribution of IGEs to runaway dynamics in non-sexual, social evolution—social runaway—informs empirical evolutionary study of traits coevolving in a variety of social contexts, not just reproduction. These have been the focus of much study and debate, and have stimulated entire fields of theoretical biology: examples include parent-offspring conflict (parental investment theory: Trivers 1972), extreme division of labour and caste determination (inclusive fitness and kin selection: Hamilton 1964), altruism (kin or group selection: Hamilton 1964; Maynard Smith 1964; Nowak 2006). As a result of these efforts, we have an appreciation of some of the evolutionary processes that can maintain such traits in their present state. IGEs arising from the social environment may play a key role during the initial stages of such coevolution by facilitating unstable runaway dynamics. Our models suggest the possibility that IGEs causing runaway dynamics in non-sexual, socially selected traits may make an outsized contribution to such evolutionary volatility, supporting arguments for a role of plasticity in capacitating evolutionary change that predate the Modern Synthesis but continue unresolved in the contemporary field of evolutionary biology (Baldwin 1896; Wcislo 1989; West-Eberhard 2006; Ghalambor et al. 2015; Bailey et al. 2018).

Table 1. Summary of predictions, with expectations for empirical studies

Prediction		Empirical expectation
1	IGEs facilitate the initial evolution of badge and donation traits	*De novo* mutations influencing provisioning are less likely to be lost from populations due to drift if their expression is socially flexible, and affected by begging behaviours.
2	Donation phenotypes should be particularly susceptible to IGEs	IGEs should have a stronger effect on costly donation traits than on badge traits. Also, the former are expected to be more frequently exploited by intraspecific or interspecific individuals.
3	Strong genetic covariances should accompany unstable social evolution	In experimental evolution studies, strong badge-donation genetic covariance should predict extreme evolutionary outcomes such as extinction and rapid trait fixation.
4	Asymmetrical IGEs enhance social runaway	Asymmetrical IGEs should be particularly common for elaborate, socially-selected traits, compared to less extreme socially-selected, or non-social traits.

DATA ARCHIVING

No data are associated with the manuscript.

LITERATURE CITED

Axelrod, R., Hammond, R. A., and A. Grafen. 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. Evolution 58:1833-1838.

Bailey, N. W. 2012. Evolutionary models of extended phenotypes. Trends Ecol. Evol. 27:561-569.

Bailey, N. W., and J. L. Hoskins. 2014. Detecting cryptic indirect genetic effects. Evolution 68:1871-1882.

Bailey, N. W., Marie-Orleach, L., and A. J. Moore. 2018. Indirect genetic effects in behavioral ecology: does behavior play a special role in evolution? Behav. Ecol. 29:1-11.

Bailey, N. W., and A. J. Moore. 2012. Runaway sexual selection without genetic correlations: social environments and flexible mate choice initiate and enhance the Fisher process. Evolution. 66:2674-2684.

Bailey, N. W., and M. Zuk. 2012. Socially flexible female choice differs among populations of the field cricket *Teleogryllus oceanicus*: geographic variation in the interaction coefficient (Ψ). Proc. R. Soc. Lond. B. 279:3589-3596.

Baud, A., Mulligan, M. K., Cesale, F. P., Ingels, J. F., Bohl, C. J., Callebert, J., Launay, J.-M., Krohn, J., Legarra, A., Williams, R. W., Stegle, O. 2017. Genetic variation in the social environment contributes to health and disease. PLoS Genet. 13:e1006498.

Baldwin, J. M. 1896. A new factor in evolution. Am. Nat. 30:441-451, 536--553.

Bascompte, J., Jordano, P., and J. M. Olesen. 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science. 312:431-433.

Battesti, M., Moreno, C., Joly, D., and F. Mery. 2012. Spread of social information and dynamics of social transmission within *Drosophila* groups. Curr. Biol. 22:309-313.

Bell, M. B. V. 2008. Receiver identity modifies begging intensity independent of need in banded mongoose (*Mungos mungo*) pups. Behav. Ecol. 19: 1087--1094.

Biernaskie JM, West SA, and A. Gardner. 2011. Are greenbeards intragenomic outlaws? Evolution. 65:2729-2742.

Bijma, P. 2014. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. Heredity. 112:61-69.

Bleakley, B. H. and E. D. Brodie III. 2009. Indirect genetic effects influence antipredator behavior in guppies: estimates of the coefficient of interaction *psi* and the inheritance of reciprocity. Evolution 63:1796-1806.

Brown, J. S., and T. L. Vincent. 2008. Evolution of cooperation with shared costs and benefits. Proc. Roy. Soc. Lond. B. 275:1985-1994.

Cardoso, S. D., Teles, M. C., and R. F. Oliviera. 2015. Neurogenomic mechanisms of social plasticity. J. Exp. Biol. 218:140-149.

Chenoweth, S. F., H. D. Rundle, and M. W. Blows. 2010. Experimental evidence for the evolution of indirect genetic effects: changes in the interaction coefficient, Psi (Ψ), due to sexual selection. Evolution 64:1849-1856.

Collins, S. A. 1995. The effect of recent experience on female choice in zebra finches. Anim. Behav. 49:479-486.

Cunningham, C. B., VanDenHeuvel, K., Khana, D. B., McKinney, E. C., and A. J. Moore. 2016. The role of neuropeptide F in a transition to parental care. Biol. Lett. 12:20160158.

Dawkins, R. 1982. The extended phenotype. Oxford University Press: Oxford.

Débarre, F., Nuismer, S. L., and M. Doebeli. 2014. Multidimensional (co)evolutionary stability. Am. Nat. 184:158-171.

Doebeli, M., and I. Ispolatov. 2014. Chaos and unpredictability in evolution. Evolution. 68:1365-1373.

Duckworth, R. A. 2009. The role of behavior in evolution: a search for mechanism. Evol. Ecol. 23:513-531.

Falconer, D. S, and T. F. C. Mackay. 1997. Introduction to Quantitative Genetics. Pearson Education Ltd., Harlow.

Fisher, R. A. 1915. The evolution of sexual preference. Eugenics Rev. 7:184-192.

Fisher, R. A. 1958. The genetical theory of natural selection, 2nd ed. (Dover Press).

Gardner, A. and S. West. 2010. Greenbeards. Evolution. 64:25-38.

Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., and K. A. Hughes. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature. 525:372-375.

Geoffroy, F., Baumard, N., and J.-B. André. 2019. Why cooperation is not running away. bioRxiv https://doi.org/10.1101/316117

González-Forero, M., and S. Gavrilets. 2013. Evolution of manipulated behavior. Am. Nat. 182:439-451.

Greenfield, M. D., Alem, S., Limousin, D., and N. W. Bailey. 2014. The dilemma of Fisherian sexual selection: Mate choice for indirect benefits despite rarity and overall weakness of trait-preference genetic correlation. Evolution. 69:3524-3536.

Hall, D. W., M. Kirkpatrick, and B. West. 2000. Runaway sexual selection when female preferences are directly selected. Evolution 54:1862-1869.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. I. J. Theoret. Biol. 7:1-16.

Hartman, P. 1960. A lemma in the theory of structural stability of differential equations. Proc. Amer. Math. Soc. 11:610-620.

Huang, Z. Y., and G. E. Robinson. 1996. Regulation of honey bee division of labor by colony age demography. Behav. Ecol. Sociobiol. 39:147-158.

Hughes, K. A., Du, L., Rodd, F. H., and D. N. Reznick. 1999. Familiarity leads to female mate preference for novel males in the guppy, *Poecilia reticulata.* Anim. Behav. 58:907-916.

Hunt, J., Simmons, L. W. 2002. The genetics of maternal care: Direct and indirect genetic effects on phenotype in the dung beetle *Onthophagus taurus*. Proc. Natl. Acad. Sci. USA 99:6828-6832.

Hunt, S., Kilner, R. M., Langmore, N. E., and A. T. D. Bennett. 2003. Conspicuous, ultraviolet-rich mouth colours in begging chicks. Proc. R. Soc. Lond. B. 270:S25-S28.

Iwasa, Y., and A. Pomiankowski. 1995. Continual change in mate preferences. Nature 377:420-422.

Iwasa, Y., Pomiankowski, A., and S. Nee. 1991. The evolution of costly mate preferences II. The 'handicap' principle. Evolution 45:1431-1442.

Jansen, V. A. A., and van Baalen, M. 2006. Altruism through beard chromodynamics. Nature 440:663-666.

Jarrett, B. J. M., Schrader, M., Rebar, D., Houslay, T. M., and R. M. Kilner. 2017. Cooperative interactions within the family enhance the capacity for evolutionary change in body size. Nat. Ecol. Evol. 1:0178.

Kazancioğlu, E., Klug, H., and S. H. Alonzo. 2012. The evolution of social interactions changes predictions about interacting phenotypes. Evolution. 66:2056-2064.

Kilner, R., and R. A. Johnstone. 1997. Begging the question: are offspring solicitation behaviours signals of needs. Trends Ecol. Evol. 12:11-15.

Kirkpatrick, M. 1982. Sexual selection and the evolution of female choice. Evolution 36:1-12.

Kirkpatrick, M., and N. H. Barton. 1997. The strength of indirect selection on female mating preferences. Proc. Natl. Acad. Sci. USA 94:1282-1286.

Kölliker, M., Brinkhof, M. W. G., Heeb, P., Fitze, P. S., and H. Richner. 2000. The quantitative genetic basis of offspring solicitation and parental response in a passerine bird with biparental care. Proc. R. Soc. Lond. B. 267:2127-2132.

Kölliker, M. and H. Richner. 2001. Parent-offspring conflict and the genetics of offspring solicitation and parental response. Anim. Behav. 62:395-407.

Kölliker, M., Royle, N. J. and P. T. Smiseth. 2012. Parent-offspring coadaptation. In: The evolution of parental care (Royl, N. J., P. T. Smiseth and M. Kölliker eds.). Oxford: Oxford University Press.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. USA 78:3721-3725.

Le Galliard, J.-F., Ferrière, R., and U. Dieckmann. 2005. Adaptive evolution of social traits: Origin, trajectories, and correlations of altruism and mobility. Am. Nat. 165:206-224.

Leimar, O. 2009. Multidimensional convergence stability. Evol. Ecol. Res. 11:191-208.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, MA.

Lynn, J. C. B., and G. L. Cole. 2019. The effect of against-background contrast on female preferences for a polymorphic colour sexual signal. Anim. Behav. 150:1-13.

Lyon, B. E., and R. Montgomerie. 2012. Sexual selection is a form of social selection. Phil. Trans. R. Soc. Lond. B. 367:2266-2273.

Marie-Orleach, L., Voght-Burri, N., Mouginot, P., Schlatter, A., Vizoso, D. B., Bailey, N. W., and L. Schärer. 2017. Indirect genetic effects and sexual conflicts: Partner genotype influences multiple morphological and behavioral reproductive traits in a flatworm. Evolution. 71:1232-1245.

Mas, F., and M. Kölliker. 2008. Maternal care and offspring begging in social insects: chemical signalling, hormonal regulation and evolution. Anim. Behav. 76:1121-1131.

Matthiopoulos, J. 2011. How to be a quantitative ecologist. John Wiley & Sons, Ltd. Chichester, UK.

Maynard Smith, J. 1964. Group selection and kin selection. Nature. 201:1145-1147.

Mayr, E. 1988. Toward a new philosophy of biology: observations of an evolutionist. Harvard University Press: Cambridge, Massachusetts.

McGlothlin, J. W., Moore, A. J., Wolf, J. B., and E. D. Brodie III. 2010. Interacting phenotypes and the evolutionary process. III. Social evolution. Evolution. 64:2558-2574.

Mead, L. S., and S. J. Arnold. 2004. Quantitative genetic models of sexual selection. Trends Ecol. Evol. 19:264-271.

Meunier, J. and M. Kölliker. 2012. Parental antagonism and parent-offspring co-adaptation interact to shape family life. Proc. R. Soc. Lond. B. 279:3981-3988.

Moore, A. J., E. D. Brodie III, and J. B. Wolf. 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution 51:1352-1362.

Muller, R. E., and D. G. Smith. 1978. Parent-offspring interactions in zebra finches. Auk. 95:485-495.

Nesse, R. M. 2009. Runaway social selection for displays of partner value and altruism. In: Verplaetse, J., Schrijver, J., and J. Braeckman (Eds.) The Moral Brain. Springer, Dordrecht.

Nowak, M. A. 2006. Five rules for the evolution of cooperation. Science. 314:1560-1563.

Olson, V. A., and I. P. F. Owens. 1998. Costly sexual signals: are carotenoids rare, risky or required? Trends Ecol. Evol. 13:510-514.

Otto, S. P., and T. Day. 2007. A Biologists Guide to Mathematical Modeling in Ecology and Evolution. Princeton University Press. Princeton, U.S.A.

Pascoal, S., Liu, X., Fang, Y., Paterson, S., Ritchie, M. G., Rockliffe, N., Zuk, M., and N. W. Bailey. 2018. Increased socially mediated plasticity in gene expression accompanies rapid adaptive evolution. Ecol. Lett. 21:546-556.

Peragine, D. E., Simpson, J. A., Mooney, S. J., Lovern, M. B., and M. M. Holmes. 2014. Social regulation of adult neurogenesis in a eusocial mammal. Neurosci. 268:10-20.

Pomiankowski, A., Y. Iwasa, and S. Nee. 1991. The evolution of costly mate preferences I. Fisher and biased mutation. Evolution 45:1422-1430.

Pomiankowski, A., and Y. Iwasa. 1993. Evolution of multiple sexual preferences by Fisher's runaway process of sexual selection. Proc. R. Soc. Lond. B 253:173-181.

Queller, D. C. 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. Proc. Natl. Acad. Sci. USA 108:10792-10799.

Rodenburg, T. B., H. Komen, E. D. Ellen, K. A. Uitdehagg, and J. A. M, van Arendonk. 2008. Selection method and early life-history affect behavioural development, feather pecking and cannibalism in laying hens: a review. Appl. Anim. Behav. Sci. 110:217-228.

Rodríguez-Gironés, M. A., Cotton, P. A. and A. Kacelnik. 1996. The evolution of begging: signalling and sibling competition. Proc. Natl. Acad. Sci. USA 93:14673-14641.

Rousset, F., and D. Roze. 2007. Constraints on the origin and maintenance of genetic kin recognition. Evolution 61:2320-2330.

Royle, N. J., Hartley, I. R., and G. A. Parker. 2002. Begging for control: when are offspring solicitation behaviours honest? Trends Ecol. Evol. 17:434-440.

Royle, N. J., Smiseth, P. T., and M. Kölliker. 2012. The evolution of parental care. Oxford University Press: Oxford.

Signor, S. A., Abbasi, M., Marjoram, P., and S. V. Nuzhdin. 2017. Social effects for locomotion vary between environments in *Drosophila melanogaster*. Evolution 71:1765-1775.

Sinervo, B., Chaine, A., Clobert, J., Calsbeek, R., Hazard, L., Lancaster, L., McAdam, A. G., Alonzo, S., Corrigan, G., and M. E. Hochberg. 2006. Self-recognition, color signals, and cycles of greenbeard mutualism and altruism. Proc. Natl. Acad. Sci. USA 103:7372-7377

Tanaka, Y. 1996. Social selection and the evolution of animal signals. Evolution. 50:512-523.

Taylor, P. D. 1996. The selection differential in quantitative genetics and ESS models. Evolution 50:2106-2110.

Trivers, R. L. 1972. Parental investment and sexual selection. In: “Sexual selection and the descent of man.” Pp. 136--179. (B. Campbell, ed.) Aldine: Chicago.

Velando, A., Kim, S.- Y., and J. C. Noguera. 2013. Begging response of gull chicks to the red spot on the parental bill. Anim. Behav. 85:1359-1366.

Wcislo, W. T. 1989. Behavioral environments and evolutionary change. Annu. Rev. Ecol. Syst. 20:137-169.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58:155-183.

West-Eberhard, M. J. 1989. Phenotypic plasticity and the origins of diversity. Annu. Rev. Ecol. Syst. 20:249-278.

West-Eberhard, M. J. 2006. Developmental plasticity and evolution. Oxford University Press: Oxford.

Wolf, J. B., Brodie III, E. D., Cheverus, J. M., Moore, A. J., and M. J. Wade. 1998. Evolutionary consequences of indirect genetic effects. Trends Ecol. Evol. 13:64-69.

Appendix

1. SOCIAL RUNAWAY WITH A QUADRATIC DONATION COST FUNCTION

To evaluate the social runaway condition when selection acts more strongly on initially small donations, we performed stability analysis following Otto and Day (2007). We substitute the quadratic fitness function in equation (19) in the Main Text for the donor cost function and assume non-reciprocal IGEs arising from the effects of elicitation badges on donors. Eigenvalues of the Jacobian at the equilibrium, $\boldsymbol{J}_{\boldsymbol{\omega}}^{*}$ (where the subscript indicates the Jacobian under a steeper donation fitness function), give an indication of stability. Thus by substitution, at $(d^*, b^*) = (0,0)$ we obtain:

$$\boldsymbol{J}_{\boldsymbol{\omega}}^{*} = \begin{bmatrix} G_{dd}(-2c) + G_{db}(k) + \psi_{bd}(G_{bb}(k) + G_{db}(-2c)) & G_{db}(-2m) + \psi_{bd}G_{bb}(-2m) \\ G_{bb}(k) + G_{db}(-2c) & G_{bb}(-2m) \end{bmatrix}$$

(A1)

The eigenvalues $\lambda_{\omega(1,2)}$ of $\boldsymbol{J}_{\boldsymbol{\omega}}^{*}$ are:

$$\lambda_{\omega(1,2)} = \frac{\gamma \pm \sqrt{\gamma^2 - 16mc(G_{bb}G_{dd} - (G_{db})^2)}}{2} \tag{A2}$$

where $\gamma \ = \ tr(\boldsymbol{J}_{\boldsymbol{\omega}}^{*})$. We make the assumption that $G_{bb}G_{dd} > (G_{db})^2$. This need not always be the case, but is not unreasonable when $G_{bb} \approx G_{dd}$ and implies $|\boldsymbol{J}_{\boldsymbol{\omega}}^{*}| > 0$. In this situation, instability is determined by the sign of the trace (Otto and Day 2007), such that unstable conditions occur when $\gamma > 0$. Thus the social runaway criterion becomes:

$$G_{dd}(-2c) + G_{db}(k) + \psi_{bd}G_{bb}(k) + \psi_{bd}G_{db}(-2c) + G_{bb}(-2m) > 0 \tag{A3}$$

With rearrangement, this takes the form of expression (20) in the Main Text:

$$\frac{-2c[G_{dd} + \psi_{bd}G_{db}]}{G_{bb}} + \frac{k[G_{db} + \psi_{bd}G_{bb}]}{G_{bb}} > 2m \tag{A4}$$

2. IGEs PERMIT SOCIAL RUNAWAY WITH COSTLY DONATION

By evaluating the social runaway criterion in expression (A4) under the parameter values used in Figure 2 of the Main Text, it can be demonstrated that IGEs permit unstable evolutionary dynamics that would otherwise not happen. First, we consider the social runaway criterion in expression (A4) without IGEs by setting $\psi_{bd} = 0$:

$$\frac{-2c(G_{dd}) + k(G_{db})}{G_{bb}} > 2m \tag{A5}$$

We assume equivalent genetic variances in badge and donor traits, such that $G_{bb} = G_{dd} \ = \ 0.4$, respectively, and a genetic covariance of $G_{db} = \ 0.15$. As in the Main Text, fitness costs of expressing a badge are assumed to mount less steeply than the fitness benefits gained through its resource elicitation effects. Thus, $m < k$, and we let $m \ = \ 0.1$ and $k \ = \ 0.3$. In this case, there is no positive value of c that satisfies the social runaway criterion. However, assuming a moderate positive IGE such that the experience of elicitation badges during interactions causes greater donation ($\psi_{bd} = 0.5$), then the social runaway criterion in (A4) is met when $c > 0.026$, and the system is unstable. More generally, stronger IGEs allow greater tolerance of donation costs for runaway dynamics (Figure A1A). This effect is amplified when the covariance between badges and donation is greater. For example, maintaining the condition in Appendix 1 that $G_{bb}G_{dd} > (G_{db})^2$, letting $G_{db} = 0.39$ allows for runaway under a greater range of IGE strengths and directions (Figure A1B).

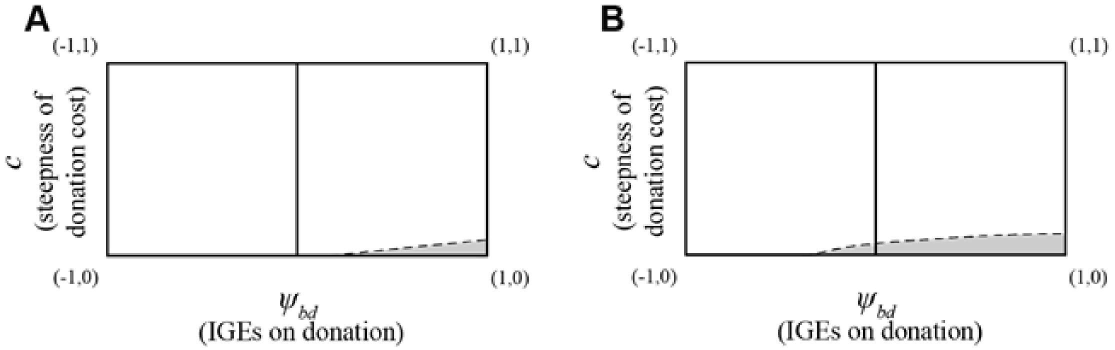

Figure A1. The interaction between unidirectional IGEs on donation and the donation cost function affects the social runaway criterion. Here, selection on donation traits is modelled using a quadratic fitness function. The shaded region indicates values satisfying the social runaway criterion in expression (A4). Only positive values of c are shown. As in Figure 2 of the Main Text, we depict a scenario in which $G_{bb} = G_{dd} = 0.4$, $m = 0.1$, and $k = 0.3$. (A) When there is a weak genetic covariance between badge and donation traits ($G_{db} = 0.15$), IGEs permit runaway provided the donation fitness function is not steep (shading). (B) However, a stronger genetic covariance between badge and donation ($G_{db} = 0.39$) widens the range of IGEs that can provoke unstable evolutionary dynamics.