Jonathan Birch and James A. R. Marshall

Queller's separation condition explained and defended

Article (Accepted version) (Refereed)

Original citation:
Birch, Jonathan and Marshall, James A. R. (2014) Queller's separation condition explained and defended. The American Naturalist, 184 (4). pp. 531-540. ISSN 00030147
DOI: 10.1086/677750

Queller's 'Separation Condition' Explained and Defended*

Jonathan Birch† and James A. R. Marshall‡

June 4, 2014

Abstract: The theories of inclusive fitness and multi-level selection provide alternative perspectives on social evolution. The question of whether these perspectives are of equal generality remains a divisive issue. In a 1992 analysis based on the Price equation, Queller argued (by means of a principle he called the 'separation condition') that the two approaches are subject to the same limitations, arising from their fundamentally quantitative-genetical character. Recently, van Veelen *et al.* have challenged Queller's results, using this as the basis for a broader critique of the Price equation, the 'separation condition' and the very notion of inclusive fitness. Here we show that the van Veelen *et al.* model, when analysed in the way Queller intended, confirms rather than refutes his original conclusions. We thereby confirm (i) that Queller's 'separation condition' remains a legitimate theoretical principle, and (ii) that the standard inclusive fitness and multi-level approaches are indeed subject to the same limitations.

1 Introduction

The theories of inclusive fitness and multi-level selection provide alternative perspectives on social evolution. From an inclusive fitness perspective, the genes for a social behaviour spread because they contribute to the inclusive fitness of their bearers: that is, they make a positive contribution to the actor's genetic representation in future generations (Hamilton 1964, 1970; Gardner *et al.* 2011). From a multi-level selection (or group selection) perspective, the evolution of social behaviour results from the interplay of selection within groups and selection between groups (Price 1972; Wilson 1975; Okasha 2006).

The formal parallels between the two theories are striking. Both approaches typically begin with a partition of the Price equation (Price 1970, 1972). In the inclusive fitness partition, we split the change attributable to selection into components separately attributable to the direct and indirect fitness effects of the social behaviour under consideration (Queller 1992a; Gardner *et al.* 2011). In the multi-level partition, we split the

---

[*] Forthcoming in *American Naturalist*

[†] Christ's College, University of Cambridge, St Andrew's Street, Cambridge, CB2 3BU, UK.

[‡] Department of Computer Science and Kroto Research Institute, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK.

change into components separately attributable to selection within groups and selection between groups (Price 1972; Wade 1985; Okasha 2006). It is straightforward to show that, when formulated in purely genetic terms, these alternative partitions of the Price equation are formally equivalent, in the sense that they can never disagree regarding the direction of the response to selection (Hamilton 1975; Wenseleers *et al.* 2010; Marshall 2011).

Although one might have expected these equivalence results to put to rest any debate about the relative accuracy or generality of the two approaches, recent years have seen numerous arguments to the effect that one approach (usually the multi-level approach) is more accurate and/or applicable to a wider range of cases than the other (Wilson and Hölldobler 2005; van Veelen 2009; Traulsen 2010; Nowak *et al.* 2010; van Veelen *et al.* 2010, 2012). We suspect that one important reason for this continuing disagreement is that results establishing the equivalence of the inclusive fitness and multi-level perspectives tend to rely on purely *genetic* formulations of the two theories. Yet when we apply the theories to particular ecological scenarios, we usually want to 'play the phenotypic gambit' (Grafen 1984; Queller 2011): that is, we want to work with partly *phenotypic* formulations of the theories that focus on the fitness effects and heritabilities of behavioural phenotypes. The problem is that demonstrating the equivalence of the two frameworks when both are formulated genetically does not settle the question of whether they constitute equally valid ways to play the phenotypic gambit.

In an analysis based on the Price equation, D. C. Queller (1992b) directly addressed this question. He argued that both the inclusive fitness and multi-level approaches (as applied to behavioural phenotypes) are, at heart, quantitative-genetical approaches that seek to separate selection gradients (which relate behavioural phenotypes to fitness) from heritabilities (which relate behavioural phenotypes to underlying genotypes)—and that both approaches succeed or fail to achieve this separation under the same conditions and for the same reasons. More specifically, he identified a formal 'separation condition' and showed that, for both the standard inclusive fitness and multi-level approaches (both of which traditionally rely on two phenotypic predictors of fitness), a quantitative-genetic separation of selection gradients from heritabilities is unattainable when genetic relatives interact in ways that yield non-additive payoffs. The upshot is that both approaches require additional predictors in order to accommodate such deviations from additivity. This result has led to the development of extended formulations of inclusive fitness theory in which deviations from additivity are explicitly represented and analysed (Smith *et al.* 2010; Queller 2011; Cornforth *et al.* 2012). These extended formulations have already proved invaluable in lab studies of microbial cooperation, a context in which accounting for deviations from additivity turns out to be particularly important (Smith *et al.* 2010; Cornforth *et al.* 2012).

We see Queller's 'separation condition', and his conclusions concerning when it is satisfied, as results of wide and profound significance for both behavioural ecologists and quantitative geneticists. Moreover, we want to emphasize that these results are valuable not only for theorists, but also for biologists studying social behaviour in the field and in the lab. In brief, this is because many empirical, quantitative-genetic studies of social evolution employ roughly the following methodology: first, use regression analysis to estimate the fitness effects and heritabilities of behavioural phenotypes; second, use the breeder's equation, or some suitably extended version of the breeder's equation, to predict the response to selection from these fitness effects and heritabilities (Falconer and Mackay 1996; McGlothlin *et al.* 2010, 2014). Empirical studies which use an inclusive fitness approach (e.g. Pfennig *et al.* 1999; Krakauer 2005; Chuang *et al.* 2010; Hatchwell *et al.* 2014) and those which use a multi-level selection approach (e.g. Tsuji 1995; Eldakar *et al.* 2010; Formica *et al.* 2010) all fall under this broad description, although (as we show below) the two approaches represent two different ways of extending the traditional breeder's equation. Queller's 'separation condition' has direct import for such studies, because it provides a general statement of the conditions under which the breeder's equation-based methodology they employ will work. If our regression model satisfies the separation condition, then the predictions delivered by this method will be accurate. If it does not, then the predictions are likely to mislead, since there will be a residual 'unseparated' component of the response to selection that our regression model fails to account for. Finding regression models that satisfy the separation condition when simple models violate it is thus a central challenge for empirical studies in behavioural ecology.

Unfortunately, because Queller framed his (1992b) argument in a relatively informal way, he left room for doubt about the meaning of his crucial 'separation condition' and the validity of the results he derived from it. In a recent critique, M. van Veelen *et al.* (2012) argue (by means of a simple game-theoretic model) that the separation condition is *irrelevant* to the question of the comparative generality of inclusive fitness and multi-level selection, and that, moreover, the two approaches are *not* equally general after all. If these new results were valid, they would have far reaching implications for social evolution theory; for in addition to challenging conventional wisdom about the relationship between inclusive fitness and multi-level selection, they would also vitiate the program (pursued by Smith et al. 2010 and Queller 2011) of using the separation condition to develop extended, more general formulations of Hamilton's rule. In this note, however, we come to Queller's defence. We show that, if one interprets the 'separation condition' in the way that Queller intended it to be understood, the formal model advanced by van Veelen *et al.* supports rather than refutes his conclusions regarding the circumstances under which the condition is satisfied. It therefore supports Queller's broader thesis that there is an important sense in which the inclusive fitness and multi-level approaches are subject to the same limitations.

2 What is the 'separation condition'?

Queller's immediate aim in formulating the 'separation condition' was to elucidate the conditions under which a specified regression model of fitness, when used to analyse the evolutionary change in some character, can successfully separate quantities that relate genotype to phenotype from quantities that relate phenotype to fitness. In other words, Queller's 'separations' are *quantitative-genetic* separations of quantities that measure a trait's *heritability* from quantities that measure the *strength of selection* on that trait. The simplest example of such a separation is the breeder's equation (Falconer and Mackay 1996), which expresses the response to selection (R) on some character as the product of a selection differential (S) and the narrow-sense heritability (h^2) of the character:

$$R = Sh^2 \tag{1}$$

Queller's 'separation condition' is intended to capture the conditions under which a given regression model of fitness can achieve a quantitative-genetic separation of this sort.

The reasoning that leads to the 'separation condition' is easiest to see in the simplest case of a one-predictor regression model of fitness. We start with the simple Price equation (Price 1970; Queller 1992b), which identifies the evolutionary change in the breeding value (G) for some character between parental and offspring populations with the covariance between breeding value and fitness (W) in the parental population (we explain the notion of a breeding value below; for now, note simply that it is a quantitative measure of one's genotype):

$$\overline{W}\Delta\overline{G} = \mathrm{Cov}(W, G) \tag{2}$$

The covariance between W and G is affected by both the heritability of the character and the strength of selection on it. Suppose that we hope to separate these effects by means of the following one-predictor regression model:

$$W = \alpha + \beta_{WP}P + \epsilon_W \tag{3}$$

in which α is the intercept of the regression line, P is the phenotypic value of the focal individual with respect to the character of interest, β_{WP} is the slope of fitness on phenotypic value, and ϵ_W is the residual (*i.e.* the extent to which the focal individual's fitness departs from the regression prediction). By substituting (3) into (2) and making the assumption that $\mathrm{Cov}(G, \epsilon_W) = 0$, we can derive the following partition of evolutionary change (Queller 1992b):

$$\overline{W}\Delta\overline{G} = \beta_{WP}\mathrm{Cov}(G, P) \tag{4}$$

We have now expressed the overall W-G covariance as a product of two quantities: one relating fitness to phenotype, and the other relating genotype to phenotype. In fact, we have arrived at a result that is formally equivalent to the breeder's equation. To see this, note that $\beta_{WP}\text{Cov}(G,P) = \beta_{GP}\text{Cov}(W,P)$, and note that the regression coefficient β_{GP} is a measure of narrow-sense heritability h^2, while $\text{Cov}(W,P)$ is a measure of the selection differential S (Queller 1992b; Falconer and Mackay 1996).

Crucially, *our assumption that* $Cov(G,\epsilon_W) = 0$ *is what makes this separation possible.* If $\text{Cov}(G,\epsilon_W) = 0$, then our regression model has fully accounted for the W-G covariance. By contrast, if the residuals in the regression model co-vary with G, then there is a residual component of the W-G covariance that our one-predictor regression model has failed to account for. This residual covariance will remain 'unseparated' (*i.e.* we will not be able to rewrite it as a product of a quantity relating P to W and a quantity relating G to P) unless we add more predictors to our regression model to account for it. For this reason, Queller refers to $\text{Cov}(G,\epsilon_W) = 0$ as the 'separation condition':

$$\textbf{Separation condition: } \text{Cov}(G,\epsilon_W) = 0 \tag{5}$$

We should note from the outset that the mathematics of regression provide no guarantee that the separation condition will be satisfied by any *phenotypic* regression model. It is guaranteed that the residuals in a regression model do not co-vary with any of the predictors, but when we 'play the phenotypic gambit' our predictors are phenotypic, not genetic, and G is not among them (cf. Queller 1992a, b, 2011).

Equation (5) provides an adequate general statement of the separation condition. It is vital, however, to note that the precise meaning of ϵ_W (and hence the precise meaning of the separation condition) is not absolute, but rather depends on the regression model of fitness we are working with. It always denotes the portion of the fitness of the focal individual that our regression model fails to account for, but altering the predictor set will typically change the residuals, and may accordingly affect whether or not the separation condition is satisfied. We cannot overemphasize this point: *the separation condition must always be defined relative to a specified regression model, and the* ϵ_W *variable always refers to the residuals in that model.*

For example, inclusive fitness analyses of social evolution often make use of the following two-predictor regression model (Queller 1992a, b, 2011; Frank 1998, 2013; McGlothlin *et al.* 2014):

$$W = \alpha + \beta_{WP.P'}P + \beta_{WP'.P}P' + \epsilon_W \tag{6}$$

in which $\beta_{WP.P'}$ is the *partial* regression of one's fitness on one's own phenotype, controlling for the effect of one's social partner's phenotype; $\beta_{WP'.P}$ is the *partial* regression of one's fitness on one's social partner's phenotype, controlling for the effect of one's own; α is again the intercept of the regression line; and ϵ_W is again the residual, *i.e.* the extent to which the focal individual's fitness departs from the value predicted by the regression model. On the assumption that $\mathrm{Cov}(G, \epsilon_W) = 0$ (*i.e.* that the separation condition is satisfied), we can use this two-predictor regression to derive the following decomposition of evolutionary change:

$$\overline{W}\Delta\overline{G} = \beta_{WP.P'}\mathrm{Cov}(G, P) + \beta_{WP'.P}\mathrm{Cov}(G, P') \tag{7}$$

from which it is straightforward to derive versions of Hamilton's rule (Queller 1992b; Frank 1998, 2013; McGlothlin *et al.* 2014). As Queller points out, equation (7) may be conceptualized as a natural extension of the breeder's equation to accommodate (additive) social interaction, since the fitness effects of a social trait on oneself and on one's social partners are weighted (respectively) by a measure of the trait's heritability via each pathway. The separation condition (5) again captures what is needed for our regression model to fully separate selection gradients from heritabilities. But note that *its precise meaning is different in this case.* For in this case, it amounts to the condition that the residuals in our *two-predictor* regression model (6) do not co-vary with breeding value.

For a second example, consider the regression models we need in order to 'play the phenotypic gambit' within the framework provided by G. R. Price's (1972) formulation of multi-level selection theory. A multi-level analysis first splits the overall evolutionary change into a between-group component and a within-group component, and then seeks to separate the effects of selection and heritability on each component. One way to achieve this separation is to start with Price's multi-level version of the Price equation, in which the covariance term of the simple equation is partitioned into between-group and within-group components (Price 1972; see Appendix for details). We can then introduce two simple regression models, one expressing a group's fitness as a function of its average phenotypic value, and the other expressing an individual's differential fitness relative to its local group mean as a function of its own phenotypic value. As we show in the Appendix, we can use these two simple regressions to achieve a clean separation of selection gradients and heritabilities at each level, but only if we assume both that the residuals in the group-level regression are independent of group breeding value *and* that the residuals in the individual-level regression are independent of individual breeding value. Hence, once again, a version of Queller's separation condition captures what is needed for our regression model to achieve a clean separation of selection gradients and heritabilities; but once again, its meaning is subtly different in this case. For in this case, it amounts to the condition that (i) the residuals in our regression model of *group* fitness do not co-vary with *group* breeding value, and that (ii) the residuals in our regression model of *individual* fitness (relative to the local group mean) do not co-vary with *individual* breeding value. A multi-level quantitative-genetic separation is attainable only if *both* conditions (*i.e.* one for each level) are satisfied.

3 The van Veelen *et al.* synergy game

We now introduce the probabilistic synergy game presented by van Veelen *et al.* (2012). In the following we have altered the notation slightly from the original, in order to disambiguate allelic values from breeding values. Individuals interact in pairs drawn from an infinite population. Every individual has an allelic value X, such that $X = 1$ if the individual possesses the social allele of interest and $X = 0$ otherwise. Similarly, every individual has a phenotypic value P, such that $P = 1$ if the individual expresses the social phenotype of interest and $P = 0$ otherwise. A fraction r of individuals are assigned a social partner with an allelic value guaranteed to be identical to their own (and thus r can be equated with genetic relatedness; Grafen 1985). A fraction $(1 - r)$ have their social partner drawn uniformly at random from the population. Of individuals with the allele $(X = 1)$, a fraction $\mathbf{P}$ express the cooperative phenotype $(P = 1)$. Individuals who do not possess the allele $(X = 0)$ never express the cooperative phenotype $(P = 0)$. The payoff matrix for interactions is given in Table 1 (where P' denotes the phenotype of the focal individual's social partner), and the frequencies of the various possible character combinations are given in the Appendix (Table A2).

At this juncture, we also need to introduce breeding values. We note in passing that van Veelen *et al.* do not do this; but it is necessary to do so, because the separation condition is defined in terms of breeding values rather than allelic values. Roughly speaking, an individual's breeding value with respect to a character P is a measure of its genetic predisposition to express P. More precisely, it is its value for that character as predicted by a linear combination of its allelic values, weighted by their average effects on the character (Falconer and Mackay 1996; Frank 1998; Gardner *et al.* 2011). If $\beta_{PX} = 1$, we could simply identify breeding values with allelic values, but in this game $\beta_{PX} = \mathbf{P}$. Accordingly, individuals with the allele have a breeding value for P equal to $\mathbf{P}$ $(X = 1 \Rightarrow G = \mathbf{P})$, while individuals without the allele have a breeding value for P equal to 0 $(X = 0 \Rightarrow G = 0)$.

	$P' = 1$	$P' = 0$
$P = 1$	$b - c + d$	$-c$
$P = 0$	b	0

Table 1: Non-additive payoff matrix as used in the probabilistic synergy game analysed by van Veelen *et al.*. The game is symmetric; payoffs to the focal (row) individual are shown in the table. As the table indicates, c denotes a payoff incurred by an agent if and only if it cooperates; b denotes a payoff received by an agent if and only if its social partner cooperates; and d denotes a 'synergistic' payoff that both players obtain if and only if they both cooperate.

4 Does the van Veelen *et al.* model support or refute Queller's argument?

The key claims of Queller's (1992b) analysis are (a) that the two-predictor regression model of fitness that underlies the standard inclusive fitness approach (*i.e.* our equation (6)) satisfies the separation condition when the payoffs of social interactions between relatives are wholly additive, but violates it when these payoffs deviate from additivity (Queller 1992b, p. 551), and (b) that the two-predictor regression model of fitness that underlies the standard multi-level selection approach *also* fails the separation condition under the same circumstances (Queller 1992b, p. 553). From these results, Queller inferred that the standard inclusive fitness and multi-level selection approaches are subject to the same limitations, and that neither is more general than the other.

van Veelen *et al.* argue that the simple synergy game presented above, in which a prosocial behaviour is expressed probabilistically, yields a counterexample to both of these key claims. This is because they take their model to show that the additivity of the payoff matrix is *irrelevant* to whether or not the separation condition is satisfied. More precisely, van Veelen *et al.* (2012, p. 71) assert that:

$$\textit{The separation condition is satisfied} \iff b = 0 \text{ or } r = 0 \text{ or } \mathbf{P} = 1 \text{ or } \mathbf{P} = 0 \tag{8}$$

in which b represents the additive component of the benefit conferred by the behaviour, r represents the coefficient of genetic relatedness, and $\mathbf{P}$ represents the probability that a gene encoding the prosocial phenotype is expressed. Crucially, and contrary to Queller's (1992b) results, the deviation from additivity (d) is absent from the right-hand side of the biconditional.

This alleged refutation of Queller rests on a misunderstanding of the separation condition. The root of the misunderstanding is the point we emphasized in Section 2: because the ϵ_W variable represents the residuals in a specified regression model, the separation condition only has meaning relative to a specified regression model. When Queller asserts that deviations from additivity lead to violations of the separation condition, we take him to be referring specifically to the separation condition *as defined with respect to a two-predictor regression model, whether of the inclusive fitness or multi-level selection variety.* Regardless of Queller's original meaning, it is plain that only the separation condition as defined with respect to a two-predictor regression model will help us to identify the circumstances under which the standard inclusive fitness and multi-level approaches cleanly separate the effects of selection and heritability.

In contrast to Queller, van Veelen *et al.* consider only whether or not the separation condition is satisfied with respect to a *one-predictor* regression model. They argue (correctly) that this has nothing to do with whether or not payoffs are additive, but rather depends only on whether or not social partners are genetically related. This point, although correct, was already noted by Queller (1992b, p. 545), and so it does not constitute a refutation of his argument. Moreover, it implies nothing at all about the conditions under which the separation condition is satisfied by a *two-predictor* regression model, and so does not bear either way on the claim that a two-predictor regression model fails the separation condition if and only if relatives interact non-additively.

Drawing attention to van Veelen and colleagues' misconstrual of the separation condition is enough to show that their alleged refutation of Queller does not succeed. However, it does not settle the question of whether or not Queller was actually correct about the parallel limitations of inclusive fitness and multi-level selection, which is the broader question at stake. To address this question, we here re-analyse van Veelen and colleagues' synergy game, this time evaluating the separation condition as defined with respect to the two-predictor regression models that characterize the standard inclusive fitness and multi-level selection approaches.

To say that the separation condition (SC) is satisfied by the standard inclusive fitness approach in the context of the van Veelen *et al.* synergy game is to say that the residuals in the two-predictor regression (6) do not co-vary with breeding value in this game, and consequently that the quantitative-genetic separation envisioned in equation (7) is indeed a correct statement about the evolutionary change in this game. Hence:

$$\textit{The SC is satisfied by the IF regression model} \iff$$
$$\mathrm{Cov}(W,G) = \beta_{WP.P'}\mathrm{Cov}(P,G) + \beta_{WP'.P}\mathrm{Cov}(P',G) \tag{9}$$

The simplest way to evaluate whether the inclusive fitness approach does indeed satisfy the separation condition in this game is to express $\mathrm{Cov}(W,G)$ and $\beta_{WP.P'}\mathrm{Cov}(P,G) + \beta_{WP'.P}\mathrm{Cov}(P',G)$ as functions of the parameters of the game, and use these expressions to work out the parameter values for which these quantities are equal. The payoff matrix (table A1) and the frequencies of the various possible interactions (given in the Appendix) give us all the information we need to do this. We will not elaborate on the steps in the computation here (see the Appendix for details). The bottom line is that the separation condition is satisfied by the IF regression model only for parameter values that satisfy the following equality:

$$\mathbf{P}(r + (1-r)p)d = \mathbf{P}\left(\frac{1+r}{1+\beta_{P'P}}\right)(r + (1-r)p)d \tag{10}$$

When genetic relatives interact non-additively (so that $d \neq 0$) and genotype does not determine phenotype (implying that $\beta_{P'P} \neq r$), this equality is *not* satisfied, and accordingly the separation condition is *not* satisfied by the inclusive fitness regression model. The equality is satisfied if payoffs are perfectly additive (such that $d = 0$), or if social partners are genetically unrelated (such that $r = 0$), or if genotype unconditionally determines phenotype (such that $\mathbf{P} = 1$ or $\mathbf{P} = 0$), but not otherwise. In short:

$$\textit{The SC is satisfied by the IF regression model} \iff$$

$$d = 0 \text{ or } r = 0 \text{ or } \mathbf{P} = 1 \text{ or } \mathbf{P} = 0 \tag{11}$$

In the Online Appendix, we further show that a two-predictor *multi-level* regression model also satisfies the separation condition only if $d = 0$ or $r = 0$ or $\mathbf{P} = 1$ or $\mathbf{P} = 0$; hence it too will violate the separation condition when genetic relatives interact non-additively and genotype does not determine phenotype. We therefore verify both of Queller's key claims.

5 Conclusion

In sum, our re-analysis of the van Veelen *et al.* synergy game reveals that the game *supports* Queller's central claim that the standard inclusive fitness and multi-level selection approaches (conceptualized in terms of the two-predictor phenotypic regression models they traditionally employ) both fail to separate the effects of selection and heritability on behavioural evolution when relatives interact in ways that yield non-additive payoffs. Our analysis thus supports Queller's broader thesis that the two approaches are subject to the same limitations, arising from their fundamentally quantitative-genetical character. Since deviations from payoff additivity are surely common in nature (Strassmann and Queller 2007), this underlines the value of developing extended formulations of both approaches in which deviations from additivity are explicitly represented (Smith *et al.* 2010; Queller 2011; Cornforth *et al.* 2012).

We should note that we do not take our results to imply that simple, two-predictor phenotypic regression models of fitness should be completely discarded in social evolution theory. This is because there are conditions under which an additive model provides a reasonable approximation of a more complex fitness structure. For example, it may provide a reasonable approximation if selection is weak, depending on how the 'weak selection' assumption is formulated (Uyenoyama and Feldman 1982; Michod 1982; Grafen 1985; Wild and Traulsen 2007). Nevertheless, our results do highlight the limitations of two-predictor phenotypic regression models when payoffs deviate from additivity, and they verify Queller's claim (challenged by van Veelen *et al.*) that such models are subject to the same limitations in both their inclusive fitness and multi-level selection guises. Moreover, our analysis reinforces the value of Queller's separation condition as a formal tool with which to address such questions.

Acknowledgements

This paper grew out of discussions at the Foundations of Social Evolution Workshop, University of Bristol, in April 2013. We thank the organizers of that event, Samir Okasha and Johannes Martens, and we thank Troy Day, Martijn Egas, Scott Nuismer, David Queller, Matthijs van Veelen and two anonymous referees for comments on various versions of the manuscript.

References

Chuang, J. S., O. Rivoire and S. Leibler. 2010. Cooperation and Hamilton's rule in a simple synthetic microbial system. Molecular Systems Biology 6:398.

Cornforth, D. M., D. J. Sumpter, S. P. Brown, and Å. Brännström. 2012. Synergy and group size in microbial cooperation. American Naturalist 180:296-305.

Eldakar, O. T., D. S. Wilson, M. J. Dlugos and J. W. Pepper. 2010. The role of multilevel selection in the evolution of sexual conflict in the water strider *Aquarius remigis*. Evolution 64:3183-3189.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics (4th edition). London: Longman.

Formica, V. A., J. W. McGlothlin, C. W. Wood, M. E. Augat, R. E. Butterfield, M. E. Barnard and E. D. Brodie III. 2010. Phenotypic assortment mediates the effect of social selection in a wild beetle population. Evolution 65:2771-2781.

Frank, S. A. 1998. Foundations of social evolution. Princeton, NJ: Princeton University Press.

Frank, S. A. 2013. Natural selection. VII. History and interpretation of kin selection theory. Journal of Evolutionary Biology 26:1151-1184.

Gardner, A., S. A. West and N. H. Barton. 2007. The relation between multilocus population genetics and social evolution theory. American Naturalist 169:207-26.

Gardner, A., S. A. West and G. Wild. 2011. The genetical theory of kin selection. Journal of Evolutionary Biology 24:1020-1043.

Goodnight, C. 2013. On multilevel selection and kin selection: contextual analysis meets direct fitness. Evolution 67:1539-48.

Grafen, A. 1984. Natural selection, kin selection and group selection. In J.R. Krebs and N.B. Davies (eds), Behavioural ecology (2nd ed). Oxford: Blackwell, pp. 62-84.

Grafen, A. 1985. A geometrical view of relatedness. Oxford Surveys in Evolutionary Biology, 2:28-89.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. Journal of Theoretical Biology 7:1-52.

Hamilton, W. D. 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228:1218-20.

Hamilton, W. D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In R. Fox (ed.), Biosocial anthropology, New York: Wiley, pp. 133-55.

Hatchwell, B. J., P. R. Gullett and M. J. Adams. 2014. Helping in cooperatively breeding long-tailed tits: a test of Hamilton's rule. Philosophical Transactions of the Royal Society of

London B: Biological Sciences 369:20130565.

Heisler, I. L. and J. Damuth. 1987. A method for analyzing selection in hierarchically structured populations. American Naturalist 130:582-602.

Krakauer, A. H. 2005. Kin selection and cooperative courtship in wild turkeys. Nature 434:69-72.

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

McGlothlin, J. W., J. B. Wolf, E. D. Brodie III and A. J. Moore. 2014. Quantitative genetic versions of Hamilton's rule with empirical applications. Philosophical Transactions of the Royal Society of London B: Biological Sciences 369:20130358.

Marshall, J. A. R. 2011. Group selection and kin selection: formally equivalent approaches. Trends in Ecology and Evolution 26:325-332.

Nowak, M. A., C. E. Tarnita and E. O. Wilson. 2010. The evolution of eusociality. Nature 466:1057-62.

Okasha, S. 2006. Evolution and the levels of selection. Oxford: Oxford University Press.

Pfennig, D. W., J. P. Collins and R. E. Ziemba. 1999. A test of alternative hypotheses for kin recognition in cannibalistic tiger salamanders. Behavioral Ecology 10:436-443.

Price, G. R. 1970. Selection and covariance. Nature 227:520-1.

Price, G. R. 1972. Extension of covariance selection mathematics. Annals of Human Genetics 35:485-90.

Queller, D. C. 1992a. A general model for kin selection. Evolution 46:376-80.

Queller, D. C. 1992b. Quantitative genetics, inclusive fitness and group selection. American Naturalist 139:540-58.

Queller, D. C. 2011. Expanded social fitness and Hamilton's rule for kin, kith and kind. Proceedings of the National Academy of Sciences USA 108:10792-10799.

Smith, J., J. D. van Dyken and P. C. Zee. 2010. A generalization of Hamilton's rule for the evolution of microbial cooperation. Science 328:1700-1703.

Strassmann, J. E. and D. C. Queller. 2007. Insect societies as divided organisms: the complexities of purpose and cross-purpose. Proceedings of the National Academy of Sciences USA 104:8619-8626.

Tsuji, K. 1995. Reproductive conflicts and levels of selection in the ant *Pristomyrmex pungens*: contextual analysis and partitioning of covariance. American Naturalist 146:586-607.

Traulsen, A. 2010. Mathematics of kin- and group-selection: formally equivalent? Evolution 64:316-323

van Veelen, M. 2009. Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. Journal of Theoretical Biology 259:589-600.

van Veelen, M., J. Garcia, M. W. Sabelin and M. Egas. 2010. Call for a return to rigour in models. Nature 467:661.

van Veelen, M., J. Garcia, M. W. Sabelin and M. Egas. 2012. Group selection and inclusive fitness are *not* equivalent; the Price equation vs. models and statistics. Journal of Theoretical Biology 299:64-80.

Wade, M. J. 1985. Soft selection, hard selection, kin selection, and group selection. American Naturalist 125:61-73.

Wenseleers, T., A. Gardner and K. R. Foster. 2010. Social evolution theory: a review of methods and approaches. In T. Szkely, A. J. Moore and J. Komdeur (eds), Social behaviour: genes, ecology and evolution. Cambridge: Cambridge University Press, pp. 132-158.

Wild, G. and A. Traulsen. (2007) The different limits of weak selection and the evolutionary dynamics of finite populations. Journal of Theoretical Biology 247:382-390.

Wilson, D. S. (1975) A theory of group selection. Proceedings of the National Academy of Sciences USA 72:143-146.

Wilson, E. O. and B. Hölldobler. 2005. Eusociality: origin and consequences. Proceedings of the National Academy of Sciences USA 102:13367-13371.

Appendix: Regression Analysis of a Synergy Game with Probabilistic Gene Expression

The basic setup for the van Veelen *et al.* synergy game with probabilistic gene expression is described in the main text. For ease of reference, the payoff matrix is given in Table A1 below. The frequencies of different possible interacting pairs in this game is given in Table A2. Here we give technical details on the derivation of our results concerning the parameter values for which two-predictor regression models (whether of the inclusive fitness or multi-level variety) satisfy Queller's separation condition.

	$P' = 1$	$P' = 0$
$P = 1$	$b - c + d$	$-c$
$P = 0$	b	0

Table A1: Non-additive payoff matrix as used in the probabilistic synergy game analysed by van Veelen *et al.*. The game is symmetric; payoffs to the focal (row) individual are shown in the table. As the table indicates, c denotes a payoff incurred by an agent if and only if it cooperates; b denotes a payoff received by an agent if and only if its social partner cooperates; and d denotes a 'synergistic' payoff that both players obtain if and only if they both cooperate.

A. The inclusive fitness case

We start with the two-predictor inclusive fitness (IF) regression model (equation (6) in the main text). To say that the separation condition (SC) is satisfied by this regression model in the context of the van Veelen *et al.* synergy game is to say that the residuals in

	$X'=0$, $P'=0$	$X'=1$, $P'=0$	$X'=1$, $P'=1$
$X=0$, $P=0$	$(r+(1-r)(1-p))(1-p)$	$(1-\mathbf{P})p(1-p)(1-r)$	$\mathbf{P}p(1-p)(1-r)$
$X=1$, $P=0$	$(1-\mathbf{P})p(1-p)(1-r)$	$(1-\mathbf{P})^2(r+(1-r)p)p$	$\mathbf{P}(1-\mathbf{P})(r+(1-r)p)p$
$X=1$, $P=1$	$\mathbf{P}p(1-p)(1-r)$	$\mathbf{P}(1-\mathbf{P})(r+(1-r)p)p$	$\mathbf{P}^2(r+(1-r)p)p$

Table A2: Frequencies of the different possible interactions between individuals having different allelic values and expressed phenotypes, where X' and P' denote (respectively) the allelic value and phenotypic value of the focal individual's social partner.

the regression model do not co-vary with breeding value in this game, and consequently that the quantitative-genetic separation envisioned in equation (equation (7) in the main text) is indeed a correct statement about the evolutionary change in this game. Hence:

The SC is satisfied by the IF regression model $\Longleftrightarrow$

$$\mathrm{Cov}(W,G)=\beta_{WP.P'}\mathrm{Cov}(P,G)+\beta_{WP'.P}\mathrm{Cov}(P',G) \tag{A1}$$

The simplest way to evaluate when the inclusive fitness regression model does indeed satisfy the separation condition in this game is to express $\mathrm{Cov}(W,G)$ and $\beta_{WP.P'}\mathrm{Cov}(P,G)+\beta_{WP'.P}\mathrm{Cov}(P',G)$ as functions of the parameters of the game, and use these expressions to work out the parameter values for which these quantities are equal. The payoff matrix (Table A1) and the frequencies of the various possible interactions (Table A2) give us all the information we need to do this. We start with the three covariances, which we compute to be the following:

$$\mathrm{Cov}(W,G)=(-c+rb+\mathbf{P}(r+(1-r)p)d)\mathbf{P}p(1-p) \tag{A2}$$

$$\mathrm{Cov}(P,G)=\mathbf{P}p(1-p) \tag{A3}$$

$$\mathrm{Cov}(P',G)=r\mathbf{P}p(1-p) \tag{A4}$$

To compute the two partial regression coefficients following the standard least-squares method outlined by Gardner *et al.* (2011), we first compute the following quantities:

$$\mathrm{Cov}(W,P)=[-c+rb+\mathbf{P}(r+(1-r)p)d]\mathbf{P}p(1-\mathbf{P}p) \tag{A5}$$

$$\mathrm{Cov}(W,P')=[b-rc+\mathbf{P}(r+(1-r)p)d]\mathbf{P}p(1-\mathbf{P}p) \tag{A6}$$

$$\mathrm{Var}(P)=\mathbf{P}p(1-\mathbf{P}p) \tag{A7}$$

from which we obtain the simple regressions of fitness on phenotype:

$$\beta_{WP} = \frac{\mathrm{Cov}(W,P)}{\mathrm{Var}(P)} = -c + rb + \mathbf{P}(r + (1-r)p)d \tag{A8}$$

$$\beta_{WP'} = \frac{\mathrm{Cov}(W,P')}{\mathrm{Var}(P)} = b - rc + \mathbf{P}(r + (1-r)p)d \tag{A9}$$

We then apply the formula for the partial regression coefficients in a two-predictor regression model (Gardner *et al.* 2011):

$$\beta_{WP.P'} = \frac{\beta_{WP} - \beta_{WP'}\beta_{P'P}}{1 - \rho^2_{PP'}} \tag{A10}$$

Since phenotypes can only take the values 1 or 0 in this game, the correlation coefficient $\rho_{PP'}$ is equal to the regression coefficient $\beta_{P'P}$. Making this simplification, and substituting in our values for the simple regressions, we obtain:

$$\beta_{WP.P'} = -c + \frac{\mathbf{P}}{1 + \beta_{P'P}}(r + (1-r)p)d \tag{A11}$$

$$\beta_{WP'.P} = b + \frac{\mathbf{P}}{1 + \beta_{P'P}}(r + (1-r)p)d \tag{A12}$$

Combining our expressions from (A3), (A4), (A11) and (A12), we obtain:

$$\begin{gathered}\beta_{WP.P'}\mathrm{Cov}(P,G) + \beta_{WP'.P}\mathrm{Cov}(P',G) = \\ \left[-c + rb + \mathbf{P}\left(\frac{1+r}{1+\beta_{P'P}}\right)(r + (1-r)p)d\right]\mathbf{P}p(1-p)\end{gathered} \tag{A13}$$

Equations (A13) contains a placeholder for $\beta_{P'P}$, a measure of the statistical association between our two phenotypic predictors. Computing $\beta_{P'P}$ from the frequency table, we find that:

$$\beta_{P'P} = \mathbf{P}(r + (1-r)p) - \frac{\mathbf{P}p}{1 - \mathbf{P}p}\left[(1-p)(1-r) + (1-\mathbf{P})(r + (1-r)p)\right] \tag{A14}$$

Note that $\beta_{P'P} = r$ in the special case of $\mathbf{P} = 1$, and consequently in this special case our expressions for $\beta_{WP'.P}$ and $\beta_{WP.P'}$ reduce (as they should) to those derived by Gardner *et al.* (2007, 2011) for a synergy game in which genotype unconditionally determines phenotype.

Taken together, these results vindicate Queller's original claim about the inclusive fitness regression model. For recall that the separation condition is satisfied by this regression model if and only if our expression for $\mathrm{Cov}(W,G)$ (*i.e.* the right-hand side of equation (A2)) is equal to our expression for $\beta_{WP.P'}\mathrm{Cov}(P,G) + \beta_{WP'.P}\mathrm{Cov}(P',G)$ (*i.e.* the right-hand side of equation (A13)). Since the b and c terms in our two expressions cancel, this boils down to the following:

$$\textit{The SC is satisfied by the IF regression model} \iff$$
$$\mathbf{P}(r+(1-r)p)d = \mathbf{P}\left(\frac{1+r}{1+\beta_{P'P}}\right)(r+(1-r)p)d \tag{A15}$$

As stated in the main text, this implies that the separation condition is satisfied by the IF regression if and only if payoffs are perfectly additive (such that $d = 0$), or social partners are genetically unrelated (such that $r = 0$), or genotype unconditionally determines phenotype (such that $\mathbf{P} = 1$ or $\mathbf{P} = 0$). In any other situation, they are not equal, because $\beta_{P'P} \neq r$. In short:

$$\textit{The SC is satisfied by the IF regression model} \iff$$
$$d = 0 \text{ or } r = 0 \text{ or } \mathbf{P} = 1 \text{ or } \mathbf{P} = 0 \tag{A16}$$

B. The multi-level selection case

We turn now to the multi-level selection (MLS) approach. As a preliminary, we need to give a formal characterization of the multi-level separation of selection gradients and heritabilities that we described informally at the end of Section 2 in the main text. A multi-level regression analysis first splits the overall evolutionary change into a between-group component and a within-group component, and then seeks to separate the effects of selection and heritability on each component.

There are two main ways to split the overall change into between- and within-group components: one is to partition the covariance term of the simple Price equation through recursive expansion in the manner of Price (1972); the other (known as 'contextual analysis') is to partition the covariance term through multivariate regression (Heisler and Damuth 1987). We focus on the first method here, in part because Queller (1992b) does, and in part because the alternative 'contextual analysis' method is extremely similar to the inclusive fitness method discussed above (see Okasha 2006 for a comparison of the 'Price approach' with 'contextual analysis'; the parallels between the latter and inclusive fitness are noted by Goodnight 2013).

We therefore start with the following multi-level version of the Price equation, originally derived by Price (1972):

$$\overline{W}\Delta\overline{G} = \mathrm{Cov}(W_i, G_i) + \mathrm{E}\left[\mathrm{Cov}^i(W_{ij}, G_{ij})\right] \tag{B1}$$

in which W_i and G_i represent (respectively) the average fitness and breeding value of the i^{th} group, and W_{ij} and G_{ij} represent (respectively) the individual fitness and breeding value of the j^{th} member of the i^{th} group. Accordingly, $\mathrm{Cov}(W_i, G_i)$ represents the between-group covariance, while $\mathrm{Cov}^i(W_{ij}, G_{ij})$ represents the within-group covariance in the i^{th} group. We then introduce two simple regression models to express a group's fitness as a function of its average phenotypic value (P_i), and an individual's differential fitness relative to its local group mean as a function of its own phenotypic value (P_{ij}):

$$W_i = \alpha_1 + \beta_{W_i P_i} P_i + \epsilon_{W_i} \tag{B2}$$

$$W_{ij} = \alpha_2 + \beta^i_{W_{ij} P_{ij}} P_{ij} + \epsilon_{W_{ij}} \tag{B3}$$

On the assumptions firstly that $\mathrm{Cov}(G_i, \epsilon_{W_i}) = 0$ and secondly that, for all i, $\mathrm{Cov}^i(G_{ij}, \epsilon_{W_{ij}}) = 0$ (*i.e.* that the separation condition is satisfied for both the between-group *and* within-group regressions), we can use these two simple regressions to derive the following decomposition of evolutionary change:

$$\overline{W}\Delta\overline{G} = \beta_{W_i P_i}\mathrm{Cov}(G_i, P_i) + \mathrm{E}\left[\beta^i_{W_{ij} P_{ij}}\mathrm{Cov}^i(G_{ij}, P_{ij})\right] \tag{B4}$$

To say that the separation condition (SC) is satisfied by this approach in context of the van Veelen *et al.* synergy game is to say that the residuals in the two multi-level regressions (equations (B2) and (B3) above) do not co-vary with breeding value in this game, and consequently that the quantitative-genetic separation envisioned in equation (B4) is a correct statement about the evolutionary change in this game. Hence:

$$\textit{The SC is satisfied by the MLS regression model} \iff$$
$$\mathrm{Cov}(W, G) = \beta_{W_i P_i}\mathrm{Cov}(G_i, P_i) + \mathrm{E}\left[\beta^i_{W_{ij} P_{ij}}\mathrm{Cov}^i(G_{ij}, P_{ij})\right] \tag{B5}$$

By focussing only on the between-group component of the overall covariance, and by noting that (by definition) $\beta_{W_i P_i} = \mathrm{Cov}(W_i, P_i)/\mathrm{Var}(P_i)$, we can obtain the following necessary condition for the SC to be satisfied:

$$\textit{The SC is satisfied by the MLS regression model} \implies$$
$$\mathrm{Cov}(W_i, G_i) = \frac{\mathrm{Cov}(G_i, P_i)\mathrm{Cov}(W_i, P_i)}{\mathrm{Var}(P_i)} \tag{B6}$$

This is only a necessary condition for the SC to be satisfied by our multi-level regression model; it is not a *sufficient* condition, since the within-group component of the covariance must also be separable into a (within-group) heritability and a (within-group) selection gradient. But a necessary condition is enough to draw relevant conclusions as to what the satisfaction of the SC requires.

Using van Veelen and colleagues' own computations of the relevant covariances for their synergy game, we can show that our MLS regression model will tend to fail the separation condition when payoffs deviate from additivity. Here are their results:

$$\begin{aligned}\mathrm{Cov}(W_i, P_i) = \mathbf{P}p&\left[\left(\frac{b-c}{2}\right)(1 + r\mathbf{P}) + \mathbf{P}(r + (1-r)p)d\right] \\ -(\mathbf{P}p)^2&\left[\left(\frac{b-c}{2}\right)(1 + r) + \mathbf{P}(r + (1-r)p)d\right]\end{aligned} \tag{B7}$$

$$\mathrm{Cov}(G_i, P_i) = \frac{1}{2}\mathbf{P}p(1-p)(1+r) \tag{B8}$$

$$\mathrm{Cov}(W_i, G_i) = \mathbf{P}p(1-p)\left[\frac{(b-c)}{2}(1+r) + \mathbf{P}(r+(1-r)p)d\right] \tag{B9}$$

These are (respectively) results (B.18), (B.15), and (B.10) in their (2012) paper. To assess whether or not the condition expressed in equation (B5) is satisfied in this model, we need one more quantity—the variance in group phenotype, $\mathrm{Var}(P_i)$—which we compute to be the following:

$$\mathrm{Var}(P_i) = \frac{1}{2}\left[\mathbf{P}p(1+r\mathbf{P}) - (\mathbf{P}p)^2(1+r)\right] \tag{B10}$$

The above results jointly imply the following:

$$\mathrm{Cov}(W_i, G_i) = \frac{\mathrm{Cov}(G_i, P_i)\mathrm{Cov}(W_i, P_i)}{\mathrm{Var}(P_i)} \iff \tag{B11}$$

$$\mathbf{P}p(1-p)\left[\mathbf{P}(r+(1-r)p)d\right] = \frac{\left[\mathbf{P}p(1-p)\right]^2(1+r)}{\mathbf{P}p(1+r\mathbf{P}) - (\mathbf{P}p)^2(1+r)} \cdot \left[\mathbf{P}(r+(1-r)p)d\right]$$

from which we conclude that the SC can be satisfied by the MLS approach only in the special cases of $d = 0$, $r = 0$, $\mathbf{P} = 1$, or (applying l'Hôpital's rule) in the limit as $\mathbf{P} \to 0$. In other words, in the special cases of perfectly additive payoffs or unconditional determination of phenotype by genotype, a two-predictor multi-level selection regression model meets our necessary condition for a clean quantitative-genetic separation. But in the more general case of $d \neq 0$, $r \neq 0$ and $0 < \mathbf{P} < 1$, the two sides of the above equation are *not* equal, and the SC cannot be satisfied. In short:

$$\textit{The SC is satisfied by the MLS regression model} \implies$$

$$d = 0 \text{ or } r = 0 \text{ or } \mathbf{P} = 1 \text{ or } \mathbf{P} = 0 \tag{B12}$$

This result confirms Queller's original claim that both two-predictor regression models fail to satisfy the separation condition when relatives interact socially and payoffs deviate from additivity, except in the special cases of zero relatedness or unconditional determination of phenotype by genotype. In either case, the way to overcome the problem is to add more predictors to the regression model in order to take deviations from additivity explicitly into account (Queller 1992b; Smith *et al.* 2010; Cornforth *et al.* 2012).