THE ROYAL SOCIETY

Received 18 April 2002
Accepted 13 August 2002
Published online 18 October 2002

The evolution of anisogamy: a game-theoretic approach

M. G. Bulmer[1*] and G. A. Parker[2]

[1]*The Old Vicarage, Chittlehampton, Umberleigh, Devon EX37 9RQ, UK*
[2]*Population and Evolutionary Biology Research Group, School of Biological Sciences, University of Liverpool, Liverpool L69 3GS, UK*

A popular theory has proposed that anisogamy originated through disruptive selection acting on an ancestral isogamous population, though recent work has emphasized the importance of other factors in its evolution. We re-examine the disruptive selection theory, starting from an isogamous population with two mating types and taking into account the functional relationship, $g(m)$, between the fitness of a gamete and its size, m, as well as the relationship, $f(S)$, between the fitness of a zygote and its size, S. Evolutionary game theory is used to determine the existence and continuous stability of isogamous and anisogamous strategies for the two mating types under various models for the two functions $g(m)$ and $f(S)$. In the ancestral unicellular state, these two functions are likely to have been similar; this leads to isogamy whether they are sigmoidal or concave, though in the latter case allowance must be made for a minimal gamete size. The development of multicellularity may leave $g(m)$ relatively unchanged while $f(S)$ moves to the right, leading to the evolution of anisogamy. Thus, the disruptive selection theory provides a powerful explanation of the origin of anisogamy, though other selective forces may have been involved in the subsequent specialization of micro- and macrogametes.

Keywords: isogamy; anisogamy; game theory

1. INTRODUCTION

Isogamy is commonplace (but not universal) in unicellular organisms, while anisogamy prevails exclusively in multicellular animals and plants. Several theories exist for the evolution of anisogamy (reviewed by Randerson & Hurst 2001*a*). Parker *et al.* (1972) proposed that males and females originated through disruptive selection acting on an ancestral isogamous (i.e. single sex) population. This arises from three very simple assumptions as follows.

(i) In a primitive marine ancestor, individuals produce a range of gamete sizes, and fusion between pairs of gametes is at random in the sea.
(ii) Each parent has a fixed budget for reproduction, so that there is a size–number trade-off: the number of gametes produced is inversely proportional to their size.
(iii) The success (e.g. viability) of the zygote increases with its size, or provisioning, which equals the sum of the sizes of the two fusing gametes.

In this model, the most frequent fusions would be between the smallest (S) gametes, but the resulting zygotes would experience low viability compared with zygotes arising from the fusion of large gametes (O). Thus, S-producers succeed by gaining most fusions with O gametes, and O-producers succeed because they generate zygotes with high viability, having more energy for development. Individuals producing intermediate-sized gametes (I), though originally commonest, decrease in frequency relative to S- and O-producers, resulting in a population consisting of proto-males and proto-females.

The condition necessary for I gametes to be lost in this model is a very simple one, and relates to how viability and other aspects of a zygote's fitness, f, increases with its size, S. Parker *et al.* (1972) derived their conclusions from computer simulations, but the model has been investigated from a population-genetic and a game-theoretic viewpoint by Charlesworth (1978) and by Maynard Smith (1978, 1982), respectively.

Disruptive selection remains a major theory for the origin of the two sexes in both the plant and animal kingdoms. It has some empirical support, based on the transition from unicellularity to multicellularity in certain algal groups, notably the Volvocales (Knowlton 1974; Bell 1985; Masden & Waller 1983; Randerson & Hurst 2001*b*), though the present mode of reproduction of Volvocales does not immediately fit the assumptions of the model (Randerson & Hurst 2001*a*,*b*).

However, Randerson & Hurst (2001*a*) emphasize the importance of other factors in the evolution of anisogamy because they claim that disruptive selection does not lead to anisogamy unless the form of the relationship between a zygote's fitness and its size, $f(S)$, is accelerating at the origin. They also claim that this is an unusual assumption for which there is no empirical evidence. We have shown elsewhere that both of these claims are incorrect (Bulmer *et al.* 2002). The model of Levitan (2000; see also Vance 1973) for zygote survival in echinoids leads to a zygote fitness function that is accelerating at the origin (see § 2b below). Randerson & Hurst (2001*a*) incorrectly conclude that the zygote fitness function under this model is zero below some finite size, above which it is decelerating, and they claim that anisogamy cannot evolve in this case by disruptive selection. But the latter claim is incorrect even

* Author for correspondence (mg.bulmer@virgin.net).

if the zygote fitness function were of this form since this type of fitness function is very similar to a sigmoidal function, and both can give rise to anisogamy (Bulmer *et al.* 2002). In response to our criticism, Randerson & Hurst (2002) postulate that a biologically plausible model must assume some finite size below which zygote fitness is zero, leading to the paradox that disruptive selection *always* gives rise to anisogamy if our argument is correct. This leads to the opposite conclusion to that of Randerson & Hurst (2001*a*) that disruptive selection *never* gives rise to anisogamy under plausible assumptions about the form of the zygote fitness function. The flaw in the argument of Randerson & Hurst (2002) is that it ignores the fact that if there is a size limit below which zygote fitness is zero, there must also be a size limit below which gamete fitness is zero (see § 2d).

Our present aim is to re-examine the conditions that lead to isogamy and anisogamy under disruptive selection in order to resolve the confusion about the predictions of this theory that now exists. The theory comes in two forms. The first form starts from an isogamous population with random mating between all gametes (Parker *et al.* 1972; Maynard Smith 1978). After the evolution of anisogamy, the evolution of preferential mating between micro- and macrogametes follows as a separate step (Parker 1978). The second form starts from an isogamous population with two mating types, with fusion restricted to gametes from individuals of opposite type (Maynard Smith 1982; Bulmer 1994). Hoekstra (1987) has argued that this is the correct starting point for models of the evolution of anisogamy. It is not necessary in this scenario to consider the evolution of preferential mating between gametes of different size since the mating-type locus is supposed to be linked to the gamete size locus; Charlesworth (1978) has shown that there will be selection to suppress crossing-over between these loci.

We shall use evolutionary game theory to determine the conditions for the evolution of anisogamy under a model that assumes two mating types and takes into account the functional relationship between the fitness of a gamete and its size, as well as the relationship between the fitness of a zygote and its size. We argue that the key to understanding the isogamy–anisogamy dichotomy is the interaction between these functional relationships. During the transition from uni- to multicellularity, the zygotic fitness function moves to the right while the gametic fitness function remains roughly constant; the shift in their relative position together with their shape determines the evolution of anisogamy from isogamy.

2. THEORY AND RESULTS

(a) *The model*

We assume that fusions occur between gametes of '+' and '–' individuals (mating types). All + individuals produce n_1 gametes of size m_1, and – individuals produce n_2 gametes of size m_2, with $n_i = M/m_i$, where M is the fixed budget for reproduction. Zygotes are of size $S = m_1 + m_2$, and the survival to adulthood of a zygote of size S is $f(S)$. The chance that a gamete of size m will survive to mate is $g(m)$.

Under these assumptions, the reproductive fitness of + individuals is

$$w_1(m_1,m_2) = \frac{Mg(m_1)}{m_1} f(m_1 + m_2), \tag{2.1a}$$

and the reproductive fitness of – individuals is

$$w_2(m_1,m_2) = \frac{Mg(m_2)}{m_2} f(m_1 + m_2). \tag{2.1b}$$

We seek an evolutionarily stable strategy (ESS), (a, b), in which a is the best response of + individuals to $m_2 = b$ and b is the best response of – individuals to $m_1 = a$. The conditions for the existence and continuous stability of the isogamous strategy with $a = b$ and of the anisogamous strategy with $a \neq b$ are of particular interest.

In testing for continuous stability we assume that m_1 evolves to its optimal size with m_2 fixed, that m_2 then evolves to its optimal size with m_1 fixed, and so on, until the system converges either to an isogamous or an anisogamous strategy. McNamara *et al.* (2003) develop a more realistic but more complicated method of testing for continuous stability in which m_1 and m_2 coevolve continuously with each other. The two methods are equivalent for the class of models in equations (2.1).

(b) *The Vance survival function*

Vance (1973; see also Levitan 2000) proposed an inverse exponential form for zygote survival, based on the fact that the time taken for a larva to reach a size capable of independent feeding (and hence unreliant on zygotic provisioning) typically relates inversely to egg (= zygote) size. Following Vance (1973), we assume that the survival (and other aspects of success) of gametes and zygotes in relation to their size can be approximated respectively as

$$g(m) = \exp\left(-\frac{\alpha}{m}\right),$$
$$f(S) = \exp\left(-\frac{\beta}{S}\right), \tag{2.2}$$

where α and β are positive parameters. This survival function is sigmoidal, accelerating from the origin until $m = \alpha/2$ ($S = \beta/2$) and then decelerating. It is probably realistic for gamete as well as for zygote survival. (Randerson & Hurst (2001*a*) incorrectly infer a different zygote survival function from the work of Levitan (2000); see Bulmer *et al.* (2002).)

Consider the log fitness of individuals of mating type i ($i = 1, 2$),

$$v_i = \ln w_i = \ln M - \frac{\alpha}{m_i} - \frac{\beta}{m_1 + m_2} - \ln m_i. \tag{2.3}$$

Then

$$\frac{\partial v_i}{\partial m_i} = \frac{\alpha}{m_i^2} + \frac{\beta}{(m_1 + m_2)^2} - \frac{1}{m_i}. \tag{2.4}$$

The isogamous ESS, $m_1 = m_2 = m^*$, satisfies

$$\left.\frac{\partial v_i}{\partial m_i}\right|_{m_1 = m_2 = m^*} = 0, \tag{2.5}$$

whence

$$m^* = \alpha + \frac{\beta}{4}. \tag{2.6}$$

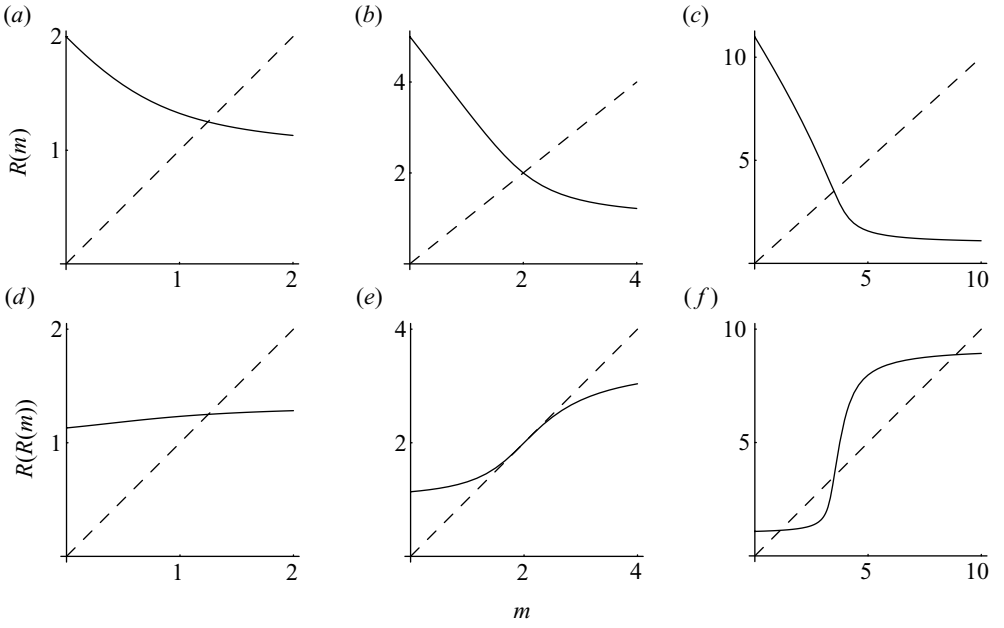

Figure 1. (*a–c*) The best response function $R(m)$ and (*d–f*) its iterate $R(R(m))$ under the model in equation (2.2) when $\alpha = 1$. (*a*,*d*) $\beta = 1$. (*a*) There is an isogamous ESS at (1.25, 1.25), which is continuously stable because the slope of the response function is greater than –1 at this point; (*d*) there is no anisogamous ESS. (*b*,*e*) $\beta = 4$. (*b*) There is an isogamous ESS at (2,2), which is on the brink of continuous instability since the slope of the function at this point is –1; (*e*) an anisogamous is about to be born. (*c*,*f*) $\beta = 10$. (*c*) There is an isogamous ESS at (3.5, 3.5), but it is continuously unstable because the slope of the response function at this point is less than –1; (*f*) there is an anisogamous ESS at (1.13, 8.87).

Although equation (2.6) is an ESS, it may not be continuously stable (Eshel 1983; McNamara *et al.* 2003). An ESS in a game with a continuous strategy set is continuously stable or unstable according as it converges back to the equilibrium after a small perturbation, or diverges from it. To investigate whether the isogamous ESS is continuously stable or unstable, consider the best response of m_1 given m_2, $R(m_2)$, satisfying

$$\left.\frac{\partial v_1}{\partial m_1}\right|_{m_1 = R(m_2)} = \frac{\alpha}{R^2(m_2)} + \frac{\beta}{(R(m_2) + m_2)^2} - \frac{1}{R(m_2)} = 0, \tag{2.7}$$

whence

$$\alpha(R(m_2) + m_2)^2 + \beta R^2(m_2) = R(m_2)[(R(m_2) + m_2)^2]. \tag{2.8}$$

We expect the best response function $R(m)$ to have a negative slope (the larger $R(m)$, the smaller the optimal value of m_2). If this slope is between –1 and 0, there is undercompensation leading to continuous stability; but if it is less than –1, there is overcompensation leading to continuous instability. This can be demonstrated in figure 1*a–c*. If we take a given m on the broken line, the intersect on the curve above or below it gives the best reply, $R(m)$. This value can be extrapolated to its equivalent new m value by finding its intersect, horizontally left or right, on the broken line. To this new m, we can obtain a new $R(m)$, and so on. The resulting figure is called a cobweb diagram (Sandefur 1990). If $R'(m) > -1$ at $m = m^*$ (undercompensation; figure 1*a*), the cobweb converges towards the intersection of the curve and broken line, which is therefore a stable equilibrium. If $R'(m) < -1$ (overcompensation; figure 1*c*), the cobweb diverges away from the intersection, which is an unstable equilibrium.

Differentiating equation (2.8) with respect to m_2 and evaluating it at $m_2 = m^*$ (using the facts that $R(m^*) = m^* = \alpha + \beta/4$), we find that the slope of the best response function at the isogamous ESS is

$$R' = -\frac{\beta}{4\alpha}. \tag{2.9}$$

This slope is always negative, as expected. It is greater than –1, giving continuous stability, when $\beta < 4\alpha$, but it is less than –1, giving continuous instability, when $\beta > 4\alpha$. (When $\beta > 32\alpha$, the strategy (m^*, m^*) is not even a global ESS because the cubic equation (2.8) has three real roots near $m_2 = m^*$. For example, when $\alpha = 1$, $\beta = 50$, so that $m^* = 13.5$, the roots of equation (2.8) at $m_2 = 13.5$ are 1.5, 9 and 13.5; the first and third roots are maxima for the log fitness, with $v_1(1.5,13.5) > v_1(13.5,13.5)$. Thus, $m_1 = 13.5$ is only a locally, not a globally, best response to $m_2 = 13.5$.)

The anisogamous ESS (a, b) with $a \neq b$ is a pair of numbers such that a is the best response to b and b is the best response to a:

$$R(b) = a, \quad R(a) = b, \tag{2.10}$$

so that

$$R(R(b)) = R(a) = b, \quad R(R(a)) = R(b) = a. \tag{2.11}$$

This ESS (if it exists) can be found by plotting the iterated response function, $R_2(m) = R(R(m))$, and finding the points, other than the isogamous ESS, at which it intersects the 45° line through the origin (figure 1*d–f*). It does

not exist when the isogamous ESS is continuously stable (figure 1*d*), but arises when the isogamous ESS becomes continuously unstable (figure 1*f*). This behaviour is analogous to the bifurcation of a stable equilibrium into a stable two-point cycle in discrete time models of a single population (May & Oster 1976; Sandefur 1990; Bulmer 1994).

It follows from the chain rule for differentiation that $(R_2)' = (R')^2$ at the isogamous ESS. Geometrical considerations (see figure 1) show that an anisogamous ESS only exists when $(R_2)' > 1$, so that an anisogamous ESS only exists when the isogamous ESS is continuously unstable. Following the graphical procedure described in the previous paragraph, we see that if $(R_2)' < 1$ (figure 1*d*), the cobweb from any m converges towards the isogamous ESS. But if $(R_2)' > 1$ (figure 1*f*), the cobweb from any m smaller than the isogamous ESS (central intersect) leads downwards towards a (lower intersect) and the cobweb from any m greater than the isogamous ESS leads upwards towards b (upper intersect); in figure 1*c*, the cobweb for $R(m)$ from any m would eventually alternate between a and b. This argument can be extended in general to any pair of survival functions.

Thus, under the Vance survival function in equation (2.2), there is a continuously stable isogamous ESS (m^*, m^*) with $m^* = \alpha + \beta/4$ (equation (2.6)) provided that $\beta < 4\alpha$. When $\beta > 4\alpha$ this ESS becomes continuously unstable, so that it is unlikely to persist, but it is replaced by the anisogamous ESS (a, b) with $a < b$. Numerical calculations show that $a = \alpha + \varepsilon$, $b = \beta - \alpha - \varepsilon$, where ε is a small deviation that tends to zero as β increases. (See the legend to figure 1*f* for an example.) Thus, the zygote size under anisogamy is $S = a + b = \beta$, which is the Smith & Fretwell (1974) optimal offspring size, satisfying $f'(S) = f(S)/S$.

In unicellular organisms, one might expect that $\alpha \approx \beta$, leading to isogamy. In the early stages of the evolution of multicellularity, one might expect that α would stay roughly constant, but that β would increase with the need to provision the embryo; when it has increased more than fourfold the scene for the evolution of anisogamy is set.

(c) *The complementary exponential survival function*

The Vance function in equation (2.2) was used to illustrate a sigmoidal survival function. To illustrate a concave survival function that decelerates continuously from the origin we shall use the function

$$g(m) = 1 - \exp\left(-\frac{m}{\alpha}\right),$$
$$f(S) = 1 - \exp\left(-\frac{S}{\beta}\right). \qquad (2.12)$$

The behaviour of the model was investigated numerically by writing a MATHEMATICA program to evaluate the best response function $R(m)$, to find the isogamous ESS by solving the equation $R(m) = m$ and to determine the system behaviour by iterating $R(m)$ many times until it settles down to a limit point (isogamy) or to a two-point cycle (anisogamy). It was found that there is an isogamous ESS at $(0.58\alpha, 0.58\alpha)$ when $\beta = \alpha$ but that it is continuously unstable because the slope of $R(m)$ at this point is –1.14. Iteration of $R(m)$ showed that the system settled down at the anisogamous ESS $(0, 1.26\alpha)$. Similar behaviour was found for all values of β tested between 0.1α and 100α. The conclusion, that anisogamy is predicted under this model for all parameter values, is however crucially dependent on the assumption that there is no lower limit on gamete size.

(d) *The threshold model for gametic survival*

A popular model for gametic survival (e.g. Maynard Smith 1978; Bulmer 1994) is to suppose that there is a critical minimum size δ for survival, such that all gametes below this size perish while all gametes above it have the same chance of survival to mate. In equation (2.1) we write

$$g(m) = 0 \quad m < \delta,$$
$$g(m) = 1 \quad m \geqslant \delta. \qquad (2.13)$$

Although not entirely realistic, this model permits some general conclusions to be drawn.

First suppose that $\delta = 0$, so that the model becomes

$$w_i(m_1, m_2) = \frac{M}{m_i} f(m_1 + m_2). \qquad (2.14)$$

There will be an isogamous ESS (m^*, m^*), where the zygote size $S^* = 2m^*$ satisfies

$$f'(S) = \frac{2f(S)}{S} \qquad (2.15)$$

(Maynard Smith 1978, 1982). McNamara *et al.* (2003) have studied this model in detail in the context of a model of Parker (1985) for bi-parental care; they have shown that this ESS is always continuously unstable if it is an internal solution, so that in the current context the system moves towards the anisogamous ESS (0, $R(0)$).

The isogamous gamete size m^* nevertheless plays a crucial part in determining how the system behaves when $\delta > 0$. Suppose first that $0 < \delta < m^*$. Then (m^*, m^*) remains unchanged as a continuously unstable ESS, so that the system will move away from it towards an anisogamous ESS, presumably (δ, $R(\delta)$), where $R(\delta) = S^* - \delta$ can be obtained by solving for the zygote size S in

$$f'(S) = \frac{f(S)}{S - \delta}. \qquad (2.16)$$

But when $\delta > m^*$, (m^*, m^*) ceases to be an ESS since gametes of this size are inviable, and the system will move to the isogamous ESS (δ, δ) since $R(\delta) = \delta$.

Thus, we expect to find isogamy or anisogamy depending on whether $\delta > m^*$ or $\delta < m^*$. As an example, suppose that $f(S)$ follows the Vance equation (2.2) with parameter β, so that $m^* = \beta/4$. When $\beta < 4\delta$, we expect the isogamous ESS (δ, δ). When $\beta > 4\delta$, we expect the anisogamous ESS (δ, b), where, from equation (2.16),

$$b = \frac{\beta}{2}\left(1 + \sqrt{\frac{1 - 4\delta}{\beta}}\right) - \delta, \qquad (2.17)$$

which tends to $\beta - \delta$ for large β.

It seems unlikely that gamete survival had a sharp threshold while zygote survival was spread out like the Vance function in the primitive unicellular condition. But if this had been the case, one might expect that $\delta \approx 1.4\beta$, the

50% point of the Vance function. Thus, one would expect that $\beta \ll 4\delta$, producing isogamy. With the evolution of multicellularity, δ would stay roughly constant, but β would increase, pushing the zygote survival function to the right, to meet the need to provision the embryo; eventually β would exceed 4δ, leading to anisogamy.

It seems likely that this argument would hold good for other types of sigmoidal zygotic survival function. In the primitive state in which gametes and zygotes have similar nutritional needs, one would expect that $\delta > m^*$, leading to isogamy; but with multicellularity m^* would increase until $\delta < m^*$, allowing the evolution of anisogamy.

Suppose now that $f(S)$ follows the complementary exponential function in equation (2.12). The isogamous ESS from equation (2.15) is $m^* = 0$, so that only the isogamous ESS (δ, δ) exists for any value of δ. Anisogamy cannot evolve under this model, however large β becomes. However, even if this were the primitive zygotic survival function, it is likely that the survival function would become sigmoidal after multicellularity developed, allowing the evolution of anisogamy, because the zygote needs a minimal food supply to develop into a free-living multicellular form.

Note that a function $f(S)$, which is zero up to some critical point γ after which it is concave, behaves in the same way as a continuous sigmoidal function. For example, the function

$$f(S) = \begin{cases} 1 - \exp\left(\dfrac{(S-\gamma)}{\beta}\right) & S \geqslant \gamma \\ 0 & S < \gamma \end{cases} \qquad (2.18)$$

has a value of m^* from equation (2.15) that is greater than γ, so that it certainly leads to anisogamy when $\gamma > \delta$. This is the type of function considered plausible by Randerson & Hurst (2001*a*, 2002). When it is concave from the origin ($\gamma = 0$), it leads to isogamy whatever the critical size δ for gametic survival. But when $\gamma > 0$, it leads to anisogamy when δ is sufficiently small, contrary to the assertion of Randerson & Hurst (2001*a*) that anisogamy never evolves under this assumption about the zygotic fitness function, but in line with the finding of Bulmer *et al.* (2002); while it leads to isogamy for larger values of δ, contrary to the assertion of Randerson & Hurst (2002) that anisogamy always evolves under this assumption. It is the relationship between the zygotic and gametic fitness functions that determines whether or not anisogamy evolves.

(e) *An alternative model for minimal size restrictions*

A more realistic model might suppose that a gamete of mass m has two components: an obligate part of fixed size δ comprising components such as the chromosomes that are necessary for its existence, and a facultative part of size $m - \delta$ comprising components such as energy reserves that increase survival. A zygote formed by the fusion of gametes of sizes m_1 and m_2 will comprise an obligate part of size 2δ (e.g two sets of chromosomes) that has no effect on survival, and a facultative part of size $m_1 + m_2 - 2\delta$ that has a direct effect on survival. The model for fitness of the two mating types in equation (2.1) can be amended to

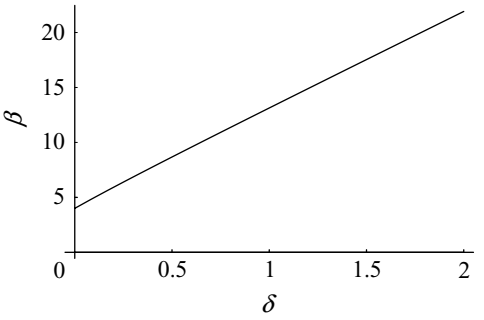

Figure 2. The critical value of β above which anisogamy evolves as a function of δ under the model in equation (2.20) when $\alpha = 1$.

$$w_i(m_1,m_2) = \frac{Mg(m_i - \delta)}{m_i} f(m_1 + m_2 - 2\delta), \qquad (2.19)$$

with the understanding that the functions g and f are zero for negative arguments.

As an example, consider the Vance survival functions in equation (2.2), which become

$$g(m) = \exp\left(-\frac{\alpha}{m-\delta}\right),$$
$$f(S) = \exp\left(-\frac{\beta}{S-2\delta}\right). \qquad (2.20)$$

By the method of § 2b, it can be shown that the isogamous ESS, $m_1 = m_2 = m^*$, is the larger root of the quadratic equation

$$\left(\alpha + \frac{\beta}{4}\right)m^* = (m^* - \delta)^2. \qquad (2.21)$$

The slope of the best response function at the isogamous ESS is

$$R' = \frac{0.5m^*(2m^* - \delta)}{c},$$
$$\text{where } c = (2m^* - \delta)(4m^* - \delta)\alpha + (m^* - \delta)(3m^* - \delta)\beta - 2(m^* - \delta)(2m^* - \delta)(3m^* - 2\delta) \qquad (2.22)$$

Numerical calculations show that this slope is always negative, and that for fixed α and δ it decreases, as β increases, from a value between 0 and –1, giving isogamy, to a value less than –1, giving anisogamy. Critical values of β at which $R' = -1$, corresponding to the transition from isogamy to anisogamy, are shown in figure 2, when size is scaled so that $\alpha = 1$. It will be seen that the critical value of β increases almost linearly from 4 when $\delta = 0$ to about 22 when $\delta = 2$. Thus, the transition to anisogamy requires a stronger shift to the right in the zygote survival function under this model, at least with the Vance curves.

Finally, consider the complementary exponential survival function in equation (2.12) with

$$g(m) = 1 - \exp\left(-\frac{m-\delta}{\alpha}\right),$$
$$f(S) = 1 - \exp\left(-\frac{S-2\delta}{\beta}\right). \qquad (2.23)$$

(Remember that the functions g and f are understood to be zero for negative arguments.) Suppose first that $\beta = \alpha$. With $\delta = 0$, the isogamous ESS (0.58α, 0.58α) is continuously unstable, and there is an anisogamous ESS

$(0, 1.26\alpha)$, as we saw in § 2c. But for $\delta > 0.01\alpha$, numerical calculations show that the isogamous ESS is continuously stable and the anisogamous ESS ceases to exist when $\beta = \alpha$. When $\delta = 0.1\alpha$, the switch from isogamy to anisogamy occurs at $\beta = 13\alpha$, and when $\delta = 0.5\alpha$ there is stable isogamy for all values of β tested up to 1000α.

Under this model, isogamy is expected in the ancestral unicellular state with $\alpha \approx \beta$ provided that $\delta > 0.01\alpha$. With the evolution of multicellularity, one would expect $f(S)$ to become sigmoidal for $S > \delta$, in which case anisogamy will eventually evolve for any value of δ. For example, consider a mixed model in which $g(m)$ has the complementary exponential form in equation (2.23) while $f(S)$ has the Vance form in equation (2.20). With $\delta = 0.5\alpha$, numerical calculations show that there is a switch from isogamy to anisogamy at $\beta = 5\alpha$. Similar behaviour is expected if $f(S)$ is zero up to $2\delta + \gamma$ and is concave from that point onwards.

3. DISCUSSION

In the ancestral unicellular state the gametic and zygotic survival functions, $g(m)$ and $f(S)$, are likely to be similar in shape and location, leading to isogamy. The development of multicellularity may leave $g(m)$ relatively unchanged, but will push $f(S)$ to the right as the need to provision the zygote increases, eventually leading to anisogamy. This is most clearly seen when the survival functions are sigmoidal, exemplified by the inverse exponential Vance function in equation (2.2). The situation is more complicated in the less likely case when the survival functions are concave, exemplified by the complementary exponential function in equation (2.12). The model must then be modified to allow for a minimal gamete size to ensure isogamy when the two survival curves are in the same location. With the development of multicellularity, the zygote survival curve may remain concave but will tend to move to the right by increasing the region for which $f(S) = 0$ in equation (2.23), eventually generating anisogamy. Alternatively, $f(S)$ may become sigmoidal as it moves to the right, which will also generate anisogamy. Thus, our present analysis shows rather generally that a plausible unicellular ancestral state would be isogamy, with anisogamy becoming inevitable once the two functions become sufficiently differentiated, as must always apply with increasing complexity in multicellular organization in plants and animals.

Fungi do not fit so obviously into the disruptive selection theory, but they do not contradict it. There are various modes of sexual reproduction (summarized in Alexopoulos (1962)). Planogametic copulation (involving the fusion of two naked gametes) most closely fits the assumptions of the theory. The morphologically simplest Chytridiomycetes are single-celled, aquatic and holocarpic (i.e. the entire thallus is used to produce gametes), and often have isogamous planogametes. Some *Allomyces* species have a simple thallus, are eucarpic (i.e. have specialized reproductive organs) and there is fusion of motile anisogametes. The most complex Chytridiomycetes have more-developed eucarpic thalli, and non-motile female gametes (e.g. *Monoblepharis*). This trend appears to follow the predictions of the theory. Higher fungi lack motile free-swimming gametes and have various forms of transfer of gametic nuclei. Some forms of transfer may be analogous to anisogamy under the disruptive selection theory. For example, in spermatization, numerous minute, uninucleate, male structures (spermatia) are produced that are carried by insects, wind, water, etc. to female gametangia or to unspecialized somatic hyphae, to which they attach and transfer their contents. Other forms are not analogous. In Basidiomycetes, the basidiospores (haploid spores resulting from meiosis) germinate to form haploid, monokaryotic, hyphae after falling on a moist substrate. These fungi usually possess multiple mating types, and two haploid hyphae of different mating types fuse and pass nuclei into each other so that one, or more typically both mycelia become dikaryotized (the donor nuclei divide and migrate from cell to cell forming a dikaryon in the recipient mycelium). The growing haploid hyphae are morphologically similar, and in that sense could be considered isogametes, but the selective forces are clearly very different from those envisaged in the disruptive selection theory.

Multicellularity in plants and animals may have evolved by the failure of the mitotic products of a single unicell to separate after fission, a pattern that is reflected by colonial forms that consist of a number of zooids, each having the nucleus, shape and organization of an individual of a related solitary species. Division of labour of the zooids can be found in some species as the colonial habit becomes more advanced. That changes in $f(S)$ would accompany the evolutionary transition from uni- to multicellularity, favouring the switch from isogamy to anisogamy, was proposed by Parker *et al.* (1972, p. 551). The correlation between increased complexity and anisogamy has generally been supported by comparative studies (Randerson & Hurst 2001*a*,*b*), beginning with Knowlton (1974) for volvocine algae, though this correlation is less distinct in other chlorophyte algae (Bell 1978). An extensive analysis by Bell (1982) for several algal and protozoan groups showed a clear correlation between the level of vegetative organization and the degree of gamete dimorphism. The most recent analysis (Randerson & Hurst (2001*b*) for the Volvocales), using modern comparative methods to control for phylogenetic effects, also supports the disruptive selection theory for the evolution of anisogamy. Both (i) the anisogamy ratio (macro-/microgamete volume), and (ii) the macrogamete size (see also Bell 1985), increased with adult size, although these results were sensitive to the mode of analysis and the phylogeny used. Randerson & Hurst (2001*b*) proposed a plausible alternative explanation of their results, based on a constraint due to the (present) mode of reproduction, but it remains to be tested whether this is a better explanation of the anisogamy–adult size correlation in the Volvocales than the present theory.

The shift in $f(S)$ to the right of $g(m)$ due to multicellularity cannot yet be quantified explicitly, and discussion of the relationship between the two functions must remain speculative. Consider first the ancestral $f(S)$ under isogamy. We envisage that this would be close to the ancestral $g(m)$, though possibly not identical to it. The sexual processes of unicells are diverse and often remarkably complex. Occasionally, conjugation results from the fusion of two full-sized, ordinary individuals (hologametes), which later divide once or twice, giving rise to products within the size range of normal 'adults' (which

halve in size at mitotic divisions), supporting the notion that $f(S) = g(m)$. Occam's Razor indicates fusion by 'hologametic adults' as the ancestral state, but it has to be noted that this pattern is relatively rare. Much more commonly, isogamous unicells produce merogametes, which are smaller than the adult (though typically not much smaller), by special fissions. The reason for the smaller merogametes may well relate to the fact that they are often formed under special conditions and specialized solely for fusion, rather than for longer-term survival and mitotic divisions, as would be the adult. It is difficult to determine whether some differentiation between $f(S)$ and $g(m)$ was ancestral, or a very common secondary specialization due to a small shift to the left by the $g(m)$ function during the specialization of merogametes.

Smallness and unicellularity are likely to have been ancestral in evolution and would have had advantages in early planktonic forms by increasing the surface area to volume ratio. A high ratio would facilitate mineral absorption in a habitat depleted of minerals, which are lost continuously in dead organisms as they sink to the seabed. This advantage probably still applies for planktonic plants and probably generates selection against large gametes and zygotes. This is a factor militating against the evolution of both multicellularity and anisogamy in planktonic forms. However, the fact that many unicells show anisogamy, albeit with anisogamy ratios typically less than those of higher organisms, demonstrates that it is not always multicellularity *per se* that drives the switch between isogamy and anisogamy. Quite commonly, anisogamous unicells have hologametic macrogametes (i.e. of similar size to the adult), which fuse with merogametes that are considerably smaller than the adult. Two explanations for the evolution of anisogamy in unicells may be suggested. First, the shift of $g(m)$ to the left of $f(S)$ during the specialization of merogametes suggested in the previous paragraph may have been large enough to generate anisogamy. Second, there may have been different $g(m)$ functions for the two types of gametes. Suppose that – gametes produce a pheromone to which the + gametes respond, so that + gametes are selected for motility; then $g(m)$ would be shifted to the left for + gametes compared with – gametes, since small gametes swim faster than large ones, which would give a preadaptation towards the evolution of anisogamy.

In multicellular organisms, it is easy to see how $f(S)$ would shift to the right of the ancestral state as organismal complexity and body size increased. Larger zygotes would require less time to reach a given adult size, and would suffer less juvenile mortality. Higher organisms typically have zygotes (and hence eggs) that are notably bigger than those of unicells; typically, their zygotes could not now survive at all if they were the size of unicellular zygotes. At each shift of $f(S)$ to the right, eggs would become adapted by the current reproductive circumstances to function efficiently around their Smith & Fretwell (1974) optimum (if sperm contribution is negligible), with a corresponding decline in $f(S)$ values away from this region, accentuating the curvature of $f(S)$. The magnitude of the shift will now largely reflect phylogenetic specializations, but will generally increase with body size. For example, although birds produce vastly bigger eggs than mammals, bigger birds produce bigger eggs than small birds. Thus, the exact location of $f(S)$ will be strongly constrained by phylogeny, but the underlying impetus for the shift to the right is related to increased size and complexity.

Once gametes become dimorphic, other selective forces are involved in the subsequent specialization of micro- and macrogametes: sperm become very small (Parker 1982; see the review of Randerson & Hurst (2001*a*)) and ova non-motile (Parker 1979). Thus, two distinct $g(m)$ functions develop, $g_1(m_1)$ for microgametes and $g_2(m_2)$ for macrogametes, even if anisogamy arose by disruptive selection from an isogamous ancestral state with the same $g(m)$ function for + and – gametes.

In higher animals, $g_1(m_1)$ may have shifted somewhat to the left of the ancestral state, generating microgametes (sperm) usually with smaller mass than the hologametes or even the merogametes of the unicellular ancestor. An explanation is that sexual selection and sperm competition have driven males to eject sperm close to, or even into females; sperm are no longer constrained by the need to maintain themselves in the way that the hologametes of a hypothetical isogametic unicellular ancestor would have been. Under internal fertilization, sperm are often provisioned by agents in the female tract, and under external fertilization such as simultaneous spawning, sperm competition places a premium on high sperm motility and low survival where the two components trade-off against each other (Ball & Parker 1996), so that sperm life may be very short. All of these effects allow sperm to become smaller by shifting $g_1(m_1)$ to the left of our hypothetical ancestral state. However, any such changes are likely to have been of much smaller magnitude than comparable shifts to the right by the function $f(S)$.

Once anisogamy has become established, sperm contribute so little to the size of the zygote that we can equate zygote size to ovum size, $m_1 + m_2 \approx m_2$, in equations (2.1). Thus, sperm are selected to maximize $g_1(m_1)/m_1$; the optimal sperm size satisfies the Smith & Fretwell (1974) equation

$$[g_1(m_1)]' = \frac{g_1(m_1)}{m_1}. \tag{3.1}$$

Ova are selected to maximize $g_2(m_2)f(m_2)/m_2$. Remember that $g_2(m_2)$ is the relationship between an ovum's survival probability and its size, up to the moment of fertilization. The zygote survival function $f(m_2)$ ensures that egg size is large for higher organisms. Coupled with the high densities of sperm typically available around eggs, it is inevitable that high survival prospects must usually apply before fusion. For species where good data exist (fish, insect, birds, mammals), quite large changes in the number of sperm do not significantly affect the fertilization probability, which is usually close to 1.0 (Ball & Parker 2000). Thus, $g_2(m_2)$ is likely to be independent of quite large changes in m_2, at least in the general region of typical ovum size, so that ova are selected to maximize $f(m_2)/m_2$; the optimal ovum size satisfies the Smith & Fretwell (1974) equation

$$f'(m_2) = \frac{f(m_2)}{m_2}. \tag{3.2}$$

Effectively, egg size becomes solely determined by $f(m_2)$. This analysis does not apply, however, in certain marine external fertilizers, which are often sperm-limited; here, fusion probability increases with ovum size (Levitan

1993), so that $g_2(m_2)$ is a significantly increasing function of m_2 in the region of the optimal ovum size, which satisfies the Smith–Fretwell equation

$$[g_2(m_2)f(m_2)]' = \frac{g_2(m_2)f(m_2)}{m_2}. \quad (3.3)$$

This gives a larger ovum size than equation (3.2).

We have argued that the origin of anisogamy is inextricably linked with the evolution of the gamete and zygote survival functions, $g(m)$ and $f(S)$. It may be difficult to make more detailed predictions until the theoretical biology of these functions, and their differentiation during the transition to multicellularity, becomes better developed. The disruptive selection theory still remains a candidate as the most powerful explanation of the origin of anisogamy, and so a more detailed understanding of the subsequent changes in $g_1(m_1)$, $g_2(m_2)$ and $f(S)$, might generate further insights into one of the most important transitions in evolution, which ultimately generated the vast diversity of adaptations that we associate with the two sexes.

We thank Dr J. W. Eaton and Dr G. N. Greenhalgh for their advice over algae and fungi, respectively, Professor J. M. McNamara for discussion about stability and two anonymous reviewers for their comments.

REFERENCES

Alexopoulos, C. J. 1962 *Introductory mycology*. New York: Wiley.

Ball, M. A. & Parker, G. A. 1996 Sperm competition games: external fertilization and 'adaptive' infertility. *J. Theor. Biol.* **180**, 141–150.

Ball, M. A. & Parker, G. A. 2000 Sperm competition games: a comparison of loaded raffle models and their biological implications. *J. Theor. Biol.* **206**, 487–506.

Bell, G. 1978 The evolution of anisogamy. *J. Theor. Biol.* **73**, 247–270.

Bell, G. 1982 *The masterpiece of nature*. London: Croom Helm.

Bell, G. 1985 The origin and evolution of germ cells as illustrated by the volvocales. In *Orgin and evolution of sex* (ed. H. O. Halvorson & A. Monroy), pp. 221–256. New York: Allan R. Liss.

Bulmer, M. 1994 *Theoretical evolutionary ecology*. Sunderland, MA: Sinauer.

Bulmer, M. G., Luttikhuizen, P. C. & Parker, G. A. 2002 Survival and anisogamy. *Trends Ecol. Evol.* **17**, 357–358.

Charlesworth, B. 1978 The population genetics of anisogamy. *J. Theor. Biol.* **73**, 347–359.

Eshel, I. 1983 Evolutionary and continuous stability. *J. Theor. Biol.* **103**, 99–111.

Hoekstra, R. F. 1987 The evolution of sexes. In *The evolution of sex and its consequences* (ed. S. C. Stearns), pp. 59–91. Basel: Birkhäuser.

Knowlton, N. 1974 A note on the evolution of gamete dimorphism. *J. Theor. Biol.* **46**, 283–285.

Levitan, D. R. 1993 The importance of sperm limitation to the evolution of egg size in marine invertebrates. *Am. Nat.* **141**, 517–536.

Levitan, D. R. 2000 Optimal egg size in marine invertebrates: theory and phylogenetic analysis of the critical relationship between egg size and development time in Echinoids. *Am. Nat.* **156**, 175–192.

McNamara, J. M., Houston, A. I., Barta, Z. & Osorno, J.-L. 2003 Should young ever be better off with one parent? *Behav. Ecol.* (In the press.)

Masden, J. & Waller, D. M. 1983 A note on the evolution of gamete dimorphism in algae. *Am. Nat.* **121**, 443–447.

May, R. M. & Oster, G. F. 1976 Bifurcations and dynamic complexity in simple ecological models. *Am. Nat.* **110**, 573–599.

Maynard Smith, J. 1978 *The evolution of sex*. Cambridge University Press.

Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge University Press.

Parker, G. A. 1978 Selection on non-random fusion of gametes during the evolution of anisogamy. *J. Theor. Biol.* **73**, 1–28.

Parker, G. A. 1979 Sexual selection and sexual conflict. In *Sexual selection and reproductive competition in insects* (ed. M. S. Blum & N. A. Blum), pp. 123–166. New York: Academic.

Parker, G. A. 1982 Why are there so many tiny sperm? Sperm competition and the maintenance of two sexes. *J. Theor. Biol.* **96**, 281–294.

Parker, G. A. 1985 Models of parent–offspring conflict. V. Effects of the behaviour of the two parents. *Anim. Behav.* **33**, 519–533.

Parker, G. A., Baker, R. R. & Smith, V. G. F. 1972 The origin and evolution of gamete dimorphism and the male–female phenomenon. *J. Theor. Biol.* **36**, 181–198.

Randerson, J. P. & Hurst, L. D. 2001*a* The uncertain evolution of the sexes. *Trends Ecol. Evol.* **16**, 571–579.

Randerson, J. P. & Hurst, L. D. 2001*b* A comparative test of a theory for the evolution of anisogamy. *Proc. R. Soc. Lond.* B **268**, 879–884. (DOI 10.1098/rspb.2000.1581.)

Randerson, J. P. & Hurst, L. D. 2002 Survival and anisogamy. *Trends Ecol. Evol.* **17**, 358.

Sandefur, J. T. 1990 *Discrete dynamical systems: theory and applications*. Oxford: Clarendon.

Smith, C. C. & Fretwell, S. D. 1974 The optimal balance between size and number of offspring. *Am. Nat.* **108**, 499–506.

Vance, R. R. 1973 On reproductive strategies in marine bottom invertebrates. *Am. Nat.* **107**, 339–352.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.