doi:10.1111/j.1558-5646.2007.00212.x

A REAPPRAISAL OF BATEMAN'S CLASSIC STUDY OF INTRASEXUAL SELECTION

Brian F. Snyder,[1,2] and Patricia Adair Gowaty[1,3,4,5]

[1] *Institute of Ecology, 517 Biological Sciences Building, University of Georgia, Athens, Georgia 30602*

[2] *E-mail: snyderb@uga.edu*

[3] *Department of Ecology and Evolutionary Biology, 621 Charles E. Young, Drive South, UCLA, Los Angeles, CA 90095-1606*

[4] *Smithsonian Tropical Research Institute, Unit 0948, APOAA, 34002-0948*

[5] *E-mail: gowaty@eeb.ucla.edu*

Received June 12, 2007
Accepted June 13, 2007

Bateman's (1948) study showing greater variances in number of mates and reproductive success in male than female *Drosophila melanogaster* is a foundational paper in sexual selection. Here we show for the first time that his methods had flaws, including the elimination of genetic variance, sampling biases, miscalculations of fitness variances, statistical pseudo-replication, and selective presentation of data. We conclude that Bateman's results are unreliable, his conclusions are questionable, and his observed variances are similar to those expected under random mating. Despite our analysis, we do not intend this article as a criticism of Bateman; he accomplished his work without modern computational tools, and his approach was groundbreaking emphasizing the significance of fitness variance for sexual selection. However, this reanalysis has implications for what counts as evidence for sexual selection and we believe that our concerns should be of interest to contemporary students of sexual selection. We call for repetitions of Bateman's study using modern statistical and molecular methods.

KEY WORDS: **Bateman gradients, mating success variance, null model, sex differences, variance in number of mates, variance in reproductive success.**

A. J. Bateman's experiment (1948), cited 15 times before and nearly a thousand times after the birth of sociobiology, was rediscovered by Parker (1970) and Trivers (1972), who used Bateman's conclusions to support their theories of sperm competition and parental investment. After Parker and Trivers, investigators frequently cited Bateman as the classic demonstration of sexual selection, whose mechanisms, mate choice, and within-sex competition, create within-sex variance in fitness (Darwin 1871). Bateman's paper also is the key citation, along with Darwin (1871), for generalities about the evolution of sex differences in fitness (Andersson 1994). These generalities, which began as Bateman's hypotheses, became his conclusions, and have more recently become elevated to "Bateman's principles." These generalities were: (1) there is greater variance in numbers of mates among males than among females; (2) fitness is more closely associated with number of mates for males than for females; and (3) differences between the sexes in the variance in number of mates (VNM) is the sine qua non signature of sexual selection. Others (Sutherland 1985; Dewsbury 2005; Tang-Martinez and Ryder 2005) have criticized the interpretations of Bateman's paper. The paper has stimulated similar studies in other species, some of which are consistent (Clutton-Brock et al. 1988; Le Boeuf and Reiter 1988; McLain 1991) with Bateman's hypotheses, whereas others are inconsistent (Hafernik and Garrison 1986; Oring et al. 1991; Ribble 1992; Bartmann and Gerlach 2001) (Table 1). Despite previous criticism and equivocal support from modern studies in some species, Bateman's paper retains its place as the single most important empirical observation in sexual selection. Because of the foundational nature of Bateman's paper, it is important to know that Bateman's data are robust, his analyses are correct and his conclusions are justified.

Table 1. A list of studies with results inconsistent with Bateman's principles. This list is not from a comprehensive search of the literature, thus we expect that it does not include all results inconsistent with Bateman's principles. Nor is it meant to suggest that other studies have not found results consistent with Bateman's principles.

Study	Species	Prediction	Conclusion (significance; if reported)
Bartmann and Gerlach 2001	Wood mice, *Apodemus sylvaticus*	Male VRS > Female VRS	No difference between male and female VRS ($P = 0.75$)
Hafernik and Garrison 1986	Damselfly, *Ischnura gemina*	Male VNM > Female VNM	No difference in male and female mating success variance ($P > 0.50$)
Ribble 1992	California mouse, *Peromyscus califonicus*	Male VNM > female VNM	Female VNM > Male NVM (F-test: $P = 0.01$)
Sheridan and Tamarin 1988	Meadow vole, *Microtus penslyvanicus*	Male VRS > Female VRS	Male VRS < Female VRS
Oring et al. 1991	Spotted Sandpiper, *Actitis macularia*	Female VRS > Male VRS	Male VRS > Female VRS
Becher and Magurran 2004	Trinidadian guppies, *Poecilia reticulata*	Male VNM > Female VNM; Male VRS > Female VRS	Male VNM = Female VNM; Male VRS > Female VRS ($P < 0.01$)
Marti 1997	Barn owls, *Tyto alba*	Male VRS > Female VRS	No difference between male and female VRS
Thomas and Coulson 1988	Kittiwake gulls, *Rissa tridactyla*	Male VRS > Female VRS	No difference between male and female VRS
Jensen et al. 2004	House sparrow, *Passer domesticus*	Male VRS > Female VRS	No difference between male and female VRS
Scott 1988	Bewick's swan, *Cygnus columbianus*	Male VRS > Female VRS	No difference between male and female VRS
Smith 1988	Song sparrow, *Melospiza melodia*	Male VRS > Female VRS	Male variance in number of eggs fertilized > female variance in number of eggs laid ($P = 0.001$); male variance in number offspring fledged < female variance in number offspring fledged ($P = 0.03$)
Merila and Sheldon 2000	Collared flycatcher, *Ficedula albicollis*	Male h^2 of LRS > Female h^2 LRS	Male h^2 of LRS is not different from 0; Female h^2 of LRS is > 0

In studying Bateman's work to attempt to experimentally repeat it, we found that Bateman's paper contains methodological and theoretical flaws, reported here for the first time, that call into question his conclusions.

We do not intend this reanalysis as a criticism of Bateman; his approach was groundbreaking and we are well aware that he accomplished his work without modern computational aids. We are much more concerned with modern researchers, many of whom have accepted Bateman's intuitive results at face value and uncritically cited them as demonstrating sexual selection. We hope our reanalysis will contribute to a more nuanced discussion of sex differences in fitness variances and the relationship of reproductive competition to fitness variances.

Bateman's Study

Bateman put small populations of virgin *Drosophila melanogaster*, consisting of either three males and three females, or five males and five females, in small milk bottles for three or four days. He studied 64 populations organized into six "series." The series varied in the number of replications, in the number of days over which mating could occur, in the number of males and females in the populations, the ages of males and females, and in their dominant marker mutations (Table 2). Bateman treated each series as a separate experiment, although he was not explicit about his reasons. Table 2 lists possible justifications for each series. There were 215 adult males and 215 adult females in Bateman's experiments. Each adult was heterozygous for a different dominant phenotypic mutation (Table 3). One-quarter of an individual's offspring would in theory exhibit the mutations simultaneously from the mother and from the father; a quarter would show a mutation only from the father; a quarter would show a mutation only from the mother. Thus, Bateman was able to assign genetic parentage to, on average, three quarters of the offspring. From these data (Table 3), he inferred how many mates and offspring each male and female had in each of the experimental populations.

Table 2. Bateman's experimental design.

Series	Number of replicates	Number of each sex per replicate	Length of experiment (days)	Age of flies at start (days)[1]	Mutations[2] (females × males)	Pedigree of experimental animals	Justification (inferred from pp. 354–356)
1	5	5	3	1, 3, 6 (mixed)	Hw, Pm, Sb, H & Me × B, Cy, CyL, Bl and Mc	Mass cultures	Difference in age between series 1 and 2 allows for determination of influence of age on variance in fecundity
2	9	5	3	1, 3, 6 (age constant)	Hw, pm, Sb, H & Me × B, Cy, CyL, Bl and Mc	Mass cultures	
3	9	3	4	1, 3, 6 (age constant)	CyL, Cy & Mc × Pm, H & Sb	Mass cultures	Switched to three individuals of each sex because of difficulty genotyping some mutants
4–1	8	3	3	1, 3, 6 (age constant)	Pm, H & Sb × CyL, Cy & Mc	Crossed to Or+ for three generations	Crossed to eliminate genetic variation; reciprocal crosses to control for effects of markers
4–2	9	3	3	1, 3, 6 (age constant)	CyL, Cy & Mc × Pm, H & Sb	Crossed to Or+ for three generations	
5–1	4	3	4	1, 3, 6 (age constant)	Pm, H & Sb × CyL, Cy & Mc	Crossed to Or+ for six generations then crossed to Skd+	Crossed to eliminate genetic variation; backcrossed to increase heterozygosity; reciprocal crosses to control for effects of markers; flies transferred to new vials daily to examine female use of male sperm
5–2	4	3	4	1, 3, 6 (age constant)	CyL, Cy & Mc × Pm, H & Sb	Crossed to Or+ for six generations then crossed to Skd+	
6–1	8	3	3	3	Pm, H & Sb × CyL, Cy & Mc	Crossed to Or+ for 15 generations then crossed to Skd+	Crossed to eliminate genetic variation; backcrossed to increase heterozygosity; reciprocal crosses to control for effects of markers
6–2	8	3	3	3	CyL, Cy & Mc × Pm, H & Sb	Crossed to Or+ for 15 generations then crossed to Skd+	

[1] "mixed" means that the ages of the flies in a population were variable; "age constant" means that the ages of the flies in one population were the same.
[2] See Figure 1.

Although Bateman did not record the behavior of males and females in his experimental populations, he concluded, as Darwin had hypothesized (1871), that male–male competition and female choice were the causes of high variance in male fitness. Based on his claim that male VNM exceeded female VNM, he concluded that males were under stronger sexual selection than females. Bateman reasoned that this difference in the strength of sexual selection could explain the sexual dimorphism observed

Table 3. The phenotypic mutants and their viability effects that Bateman used to identify genetic parents of offspring in his experiments (copied from Bateman 1948).

Chromosome	Symbol	Name	Main effect
I	*Hw*	Hairy-wing	Extra long bristles on wing veins. Homozygote viable and more extreme.
	B	Bar	Reduction of size of eye that becomes a narrow kidney shape. Homozygote fully viable and more extreme.
II	*Pm*	Plum	Eye color brown: slight darkening of body color. Homozygote lethal.
	Cy	Curly	Wings curled upward. Homozygote lethal.
	CyL^4	Curly-Lobe	*Cy* with, in addition, eye reduced in size and with a nick in the ventral edge.
	Bl	Bristle	All bristles shortened and thickened. Homozygote lethal.
III	*Sb*	Stubble	Same as *Bl*. *Sb* + *Bl* more extreme than either.
	Md	Moire	Eyes paler than wild type with shimmering appearance of shot silk. Body color paler. Homozygote lethal.
	H	Hairless	Hairs removed from various parts of the body, particularly the post verticals at back of head. Homozygote lethal.
	Mc	Microcephalous	Eyes reduced or absent. Homozygote viable.

in some species: the evolution of showy, extravagant, and bizarre traits in males, but not in females. It was Bateman who first concluded that differences between the sexes in variance in reproductive success (VRS) are a "sign of intramasculine selection" (Bateman 1948, p. 362).

Bateman's results have since been codified into Bateman's three principles (Arnold 1994b). These are: (1) the sex experiencing the more intense sexual selection has higher variance in the number of mates. (2) The sex experiencing the more intense sexual selection has higher VRS (Wade 1979; Wade and Arnold 1980). (3) The slope of the relationship between number of mates and reproductive success is greater in the sex experiencing stronger sexual selection (Arnold and Duvall 1994; Arnold 1994b).

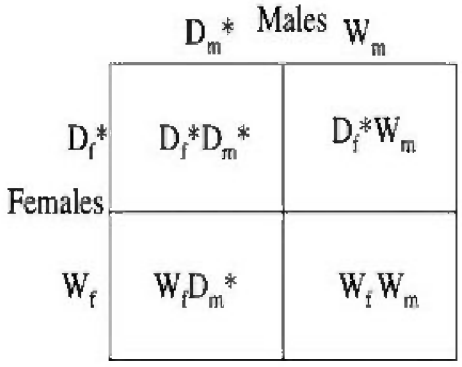

Figure 1. A stylized Punnett square with the progeny types from mating two individuals, each heterozygous for different dominant markers (D) and wild-type (W). Bateman used D_f*D_m* (where D_f and D_m are dominant alleles inherited from females and males, respectively) to estimate VNM of females and males. He used the sum of D_f*D_m* + $W_f D_m$* as an estimate of reproductive success of males and D_f*D_m* + D_f*W_m for reproductive success of females.

Note on Terminology

In contemporary literature, variation in number of mates is often called mating success variance (MSV). However, we, like Bateman (p. 360), prefer the operational terms "number of mates" and VNM to avoid confusing number of mates with number of copulations. Bateman also frequently used the term "mean square" of the number of mates or (mean square of fertility). Bateman's term "mean square" has generally gone out of common usage, but it is synonymous with "variance." Using the data in Bateman (1948, table 7, p. 360), we determined that he used the equation, $\sum (\bar{Y} - Y)^2/n - 1$, to compute mean squares. This is the standard formula for the second moment about the mean, or "variance" in common usage (Sokal and Rolf 1969) and is the same formula for variance in JMP®, which we used for all recalculations.

Throughout the remainder of the article we will use VNM when discussing what Bateman called either VNM or mean square in number of mates, and VRS for what Bateman called the mean square of fertility. In Bateman's study, "fertility" was operationally the number of larvae that eclosed. We use the operational term, "the number of eclosed offspring" or "reproductive success" to refer to what Bateman called "fertility." We use VNM and VRS to remain as precise as possible both with the current literature and with Bateman's terminology.

Methodological Problems with Bateman's Study

Bateman's methods were constrained by the technology and knowledge available to him in the mid 1940s. His use of phenotypic mutants (Table 3) was an ingenious technique and necessary at that time as the only way to determine parentage in populations with more than one individual of each sex. Nevertheless, the use of phenotypic markers may have altered the mating behavior of adult

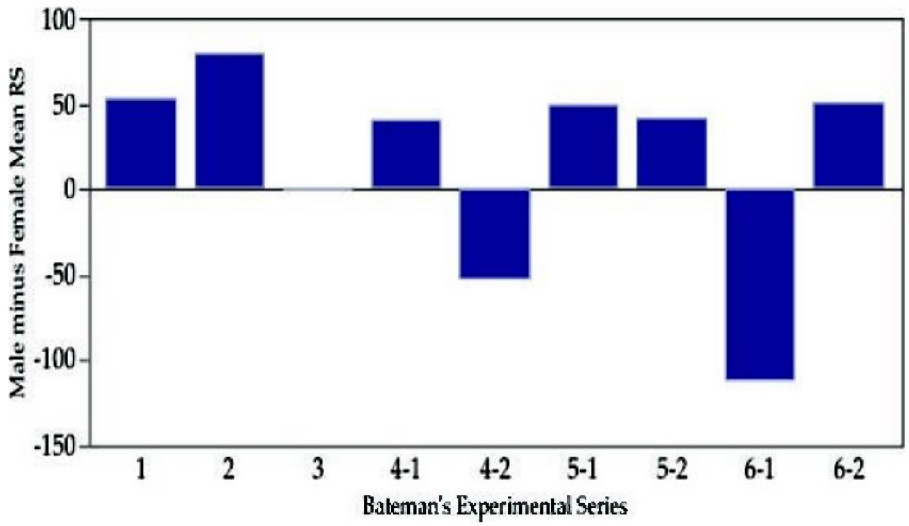

Figure 2. Mean reproductive success of males and females is unequal in Bateman's experiment, indicating rounding error or sampling bias that may have affected the magnitude of VNM and VRS and thus Bateman's conclusions. Data for this figure were in Bateman (1948, table 7, p. 360 and table 8, p. 361).

male and female subjects. In general, inbred, mutant animals mate less often and have fewer viable offspring than out-bred, wild-type animals (Merrel and Underhill 1956; Greer and Green 1962; Sharp 1984). Modern studies in many other species (Andersson 1994) suggest that it is probable that mutations of major effects affect individual mate choice and mating behavior. In fact, Bateman was aware that the Curly mutation, which he used, potentially affected the mating behavior of *D. pseudoobscura* females by making it harder for them to "interpose their wings between their abdominal and approaching males." (Bateman 1948, p. 351). Thus, sex differences in *D. melanogaster* in Bateman's study were probably affected by differences in behavior, mating capacity, or attractiveness of different mutants, and therefore could be artifacts of his methodology, rather than rules of nature.

Bateman was clearly aware of some of these potential problems. He conducted an analysis of variance to determine the effects of his markers, age, and variation between populations (bottles) on variation in reproductive success ("fertility" in Bateman's table 6, p. 358). He reported that the within-group variance was generally greater than the between group variance, suggesting that the size of the effect of his experimental manipulations on VRS exceeded the size of the natural VRS. However, he was also able to use this analysis to demonstrate that, in all but one case, the male within-group VRS was not greater than the female within-group VRS suggesting that the portion of male reproductive success due to the markers did not vary more than female reproductive success due to markers.

Besides the possible effects of mutant phenotypes on behavior, more than half of the mutations Bateman used were lethal when homozygous (Table 3), and some of the heterozygous offspring had reduced viability (Bateman 1948, table 2, p. 355). Because Bateman used the phenotypes of the eclosing (and thus surviving) flies that carried one dominant marker from the male and/or one from the female (Fig. 1) as a way to estimate the number of mates and the reproductive success of each adult fly of the parental generation, his results were likely confounded with viability effects. If viability deficits were greater in offspring having more than one mutation simultaneously (e.g., heterozygous for both curly wings and short bristles), Bateman's counts of surviving offspring to estimate the VNM may have been underestimates for one or both sexes.

In diploid species with an equal sex ratio, an equal mean reproductive success for the sexes is a biological necessity, but only one of Bateman's six series showed equal mean reproductive success of males and females (Fig. 2). This result suggests that differential mortality may have been a problem in Bateman's study. The results in Figure 2 could be partly due to rounding error, but the differences are sufficiently large that this cannot be the full explanation. Bateman used unequal progeny classes to assign paternity and maternity (Fig. 1). That is, Bateman assigned single parentage (Bateman 1948, p. 353 and table 4, p. 357) to offspring that only expressed a mutation from one parent. Had he confined parentage assignments to offspring exhibiting two mutations, he would have been able to unambiguously assign both of an individual's parents, as he noted, and he also would have constrained his estimates of reproductive success for mothers and fathers to be equal. The fact that he did not do this is critical to the reliability of his conclusions about sex differences in VRS, because means and variances are positively correlated. Bateman's summary of mean reproductive success for females and males (Bateman 1948, table 8, p. 361) shows that males in six of the nine comparisons had higher mean reproductive success than females (Fig. 2), a bias

that may have contributed to higher VRS in males (see below). In the wild, when mean reproductive success of males and females of diploid species are unequal, we would conclude that we did not sample all males and all females. In Bateman's laboratory study with equal numbers of males and females in small populations, failure to observe equal mean reproductive success for the two sexes is an indication that sampling was biased.

Bateman's sex differences in VNM may have also been a consequence of sex differences in age of first reproduction. He used one-, three-, and six-day-old virgin males and females in constant and mixed-age populations (Table 2). In general, modern studies (Markow 1996) indicate that male *D. melanogaster* are sexually mature at one day, but females are not sexually mature until they are four days old. Thus, only in the populations with six-day-old females (approximately one-third of experiments is in series 1–4) would modern investigators expect females to have been old enough to mate more than one time. In series one, the only mixed-age population, inclusion of sexually immature females may have artificially increased the VRS of females. Bateman wrote, for males "the mean square due to age is 20.5 whilst in females it is 424.6. There is a probability of 5 per cent that this is random" (Bateman 1948, p. 358). In the remainder of the populations, those with constant aged flies, the use of sexually immature individuals may have decreased female VNM and VRS. In populations that began with one-day-old flies and ran for three days (approximately one-third of experiments in series 1, 2, and 4), females would reach reproductive maturity just as the experiment ended and would be expected to mate only once. The fact that many of the females that were only one-day old at the start of experiments produced offspring is surprising and could be due to forced copulation (Manning 1967; Markow 2000). Thus, an alternative explanation for Bateman's data is that the sex differences in VNM were due to the inclusion of immature females in his breeding populations.

Finally, in half of Bateman's experimental series (i.e., series 4–6, table 2) he purposefully eliminated all genetic variation other than that associated with the mutations (Bateman table 3, p. 356). In series 4–6 Bateman crossed his mutant flies to highly inbred strains for three, six, or 15 generations. In series 5 and 6 he then backcrossed these inbred flies to another highly inbred strain to introduce heterozygosity lost during generations in inbreeding. Due to this inbreeding, the individuals within a sex and series would be nearly genetically identical, except for the parts of the genomes linked to the phenotypic mutations (given that inversions accompanied four of the six mutations he used for these series [see Bateman 1948, p. 354], the section of the genome linked to these mutations may have been relatively large). Bateman did this to control for background genetic variation. However, in doing so he rendered series 4–6 much less useful for the study of sexual selection because he eliminated genetic variation.

Reanalysis of Bateman's Data

Bateman's use of statistics was far ahead of its time; many papers in the field written during Bateman's time did not contain any statistical analyses. Despite this, by modern standards, there are a number of problems with his statistical analysis. Our comments are not directed at Bateman, but rather at modern students of sexual selection. We call for more caution about Bateman's data because, apart from methodological and sampling problems, there are problems of miscalculations, pseudo-replication, and statistically unanalyzed data that render Bateman's conclusions questionable.

Although Bateman did not report statistical tests for sex differences in VNM, we were able to recalculate male and female VNM (Table 4) from a table in Bateman (1948, table 9, p. 360). One of nine female VNM and two of nine male VNM were incorrect in Bateman's table. We easily attributed two of the mistakes to rounding errors, but the other appears to be an error in arithmetic—acceptable in an era before calculators. Comparing Bateman's reported VNM ratios with recalculated VNM ratios (using statistical formulae in JMP[@] that are identical to those Bateman used for his analysis of VRS) reveals that only two of the nine recalculated VNM ratios are statistically significant (Table 4). Bateman did not test their statistical significance; thus, the often-cited result from Bateman of significant sex differences in VNM is, and has always been, weak.

Bateman also reported sex differences in VRS (Bateman 1948, p. 357). It is impossible given the information in his paper to reconstitute a reliable record of individual reproductive success. To retest the VRS ratios that Bateman reported, we used Bateman's reported data (Bateman 1948, table 5, p. 357). Bateman said that eight of the nine tested ratios demonstrated statistically significant sex differences in VRS. There were between four and nine replicates for each of Bateman's treatments (Table 2). However, he used between 11 and 44 degrees of freedom in his statistical analyses (see Bateman 1948, table 5, p. 357). Evidently, he considered each individual fly, and not the populations, to be the experimental unit, because the degrees of freedom Bateman used were exactly what one would expect had he considered each individual rather than each population the experimental units. Because several flies all inhabited the same bottle, their behavior and their reproductive success were not independent. More importantly, VRS is a population statistic that is not a property of individuals. Thus, by treating each fly within a bottle as the experimental unit and not the bottle population, Bateman pseudo-replicated his statistical analysis (Hulbert 1984). Repeating Bateman's analysis using an F-test on untransformed data the same test Bateman used but with the correct degrees of freedom associated with the number of study populations, only two of the nine ratios support his conclusion that male VRS exceeds female VRS (Table 4). (Note that in modern

Table 4. Reanalysis of Bateman's VNM and VRS showing (in bold) calculation errors and pseudo-replication. Starred items include significant *P* values and Bateman's calculation errors. Like Bateman, we used an *F*-test for comparisons of male and female variance in number of mates and reproductive success.

	Variance in numbers of mates (VNM) analysis							Variance in reproductive success (VRS) analysis					
Series	Bateman's Female VNM	Bateman's Male VNM	Calculated Female VNM	Calculated Male VNM	S^2	df	*P*-value	Bateman's Female VRS	Bateman's Male VRS	S^2 ratio	Bateman's df	True df	*P*-value
1	0.36	1.51	0.36	1.52	4.23	4	0.096	474.9	1377.8	2.9	19	4	0.16
2	0.44	***1.48***	0.44	***1.53***	3.5	8	***0.048****	183.9	734.6	3.99	44	8	***0.033****
3	0.55	1.17	0.55	1.17	2.12	8	0.15	858.5	2433	2.83	26	8	0.081
4-1	0.13	0.48	0.13	0.48	3.70	7	***0.053****	215.0	463.7	2.16	23	7	0.166
4-2	0.23	0.85	0.23	0.85	3.70	8	***0.0412****	454.5	1367.7	3.01	26	8	0.07
5-1	0.27	0.45	0.27	0.45	1.66	3	0.34	984.6	1604.4	1.63	11	3	0.35
5-2	0.39	0.93	0.39	0.93	2.41	3	0.24	208.9	1700.4	8.14	11	3	0.059
6-1	***0.27***	***0.96***	***0.34***	***1.04***	3.04	7	0.083	992.7	2798.4	2.82	23	7	0.097
6-2	0.38	0.55	0.38	0.55	1.45	7	0.32	276.7	1098.0	3.97	23	7	***0.045****

studies, investigators would probably log-transform variances before testing for differences between independent variances.)

The most famous parts of Bateman (1948) are his graphs (p. 362), which are closely associated with the second and third of Bateman's principles. Two graphs (Bateman 1948, fig. 1a and b, p. 362) show reproductive success (expressed as relative RS) as a function of the mean number of mates for males and females. The second graph (Bateman 1948, fig. 1b, p. 362) pools results over series 5 and 6 (Table 2), which were inbred (see above). This graph shows the familiar and most cited figure in which female relative reproductive success plateaus after mating with one male. In contrast, the first graph (Bateman 1948, fig. 1a, p. 362) pools results over experimental series 1–4 (Table 2), and it shows that female reproductive success does not in fact plateau, but continues to increase with a female's number of mates. Tang-Martinez and Ryder (2005) point out that series 5 and 6, which gave the intuitively expected result but which were based on fewer number of breeders (Bateman, fig. 1b, p. 362) is the graph most often cited and republished when modern researchers describe sex differences in the relationship between number of mates and reproductive success (Krebs and Davies 1987; Arnold 1994a; Freeman and Herron 2001; Drickamer et al. 2002). Apparently, Bateman's sole rationale for graphing these series separately was that "...series 5 and 6 differed somewhat from the rest" (Bateman 1948, p. 361). This was neither a legitimate nor an a priori justification. Bateman goes on to note that series 5 and 6 were different from series 1–4 because the flies in series 5 and 6 were derived from crosses with inbred strains. However, series 4 was derived from inbred lines in a similar way, and, all six series differed in important ways including the number of flies in each population. A more appropriate way of presenting and analyzing these data would be to analyze each series separately (Fig. 3) and to use a combined probability analysis (see below), or to do an analysis combining individuals over all trials (see below). Unfortunately, because Bateman reported mean reproductive success rather than reproductive success and number of mates for each individual in each population, we cannot complete a single analysis using the most robust of approaches. Consequently, we used the next most reasonable analysis and re-graphed all the data on the same graph (Fig. 4). A regression using these means (Fig. 4) shows that the number of mates explains a significant amount of the variation in mean reproductive success for males *and for females*, contrary to Bateman's conclusion for females, and contrary to Bateman's second principle. The 95% confidence interval for the female slope is very close to the male slope, so the evidence that the male and female curves are different is weak. Moreover, because these regressions are on means, the true 95% confidence intervals for individuals may be wider than the ones we report. The fact that the male line lies above the female line in Bateman's figure and in our reanalysis may be due to the sampling problems we discussed above, which gave rise to a higher apparent mean reproductive success of males (Fig. 2). Thus, we conclude that there is no serious statistical basis in Bateman's data for his conclusion that the reproductive success of females does not increase with the numbers of mates females have. In fact, had Bateman graphed his data on a single figure, he could have claimed the first evidence for a benefit of polyandry for females.

Bateman could have made other semiarbitrary decisions for how to group his series for analysis, which could have affected his conclusions. For example, had Bateman been focused on the effects of population history on female mating decisions with inbred and outbred flies, he might have come to different conclusions. Females faced with mating with inbred males may have chosen less often to mate multiple times because there would

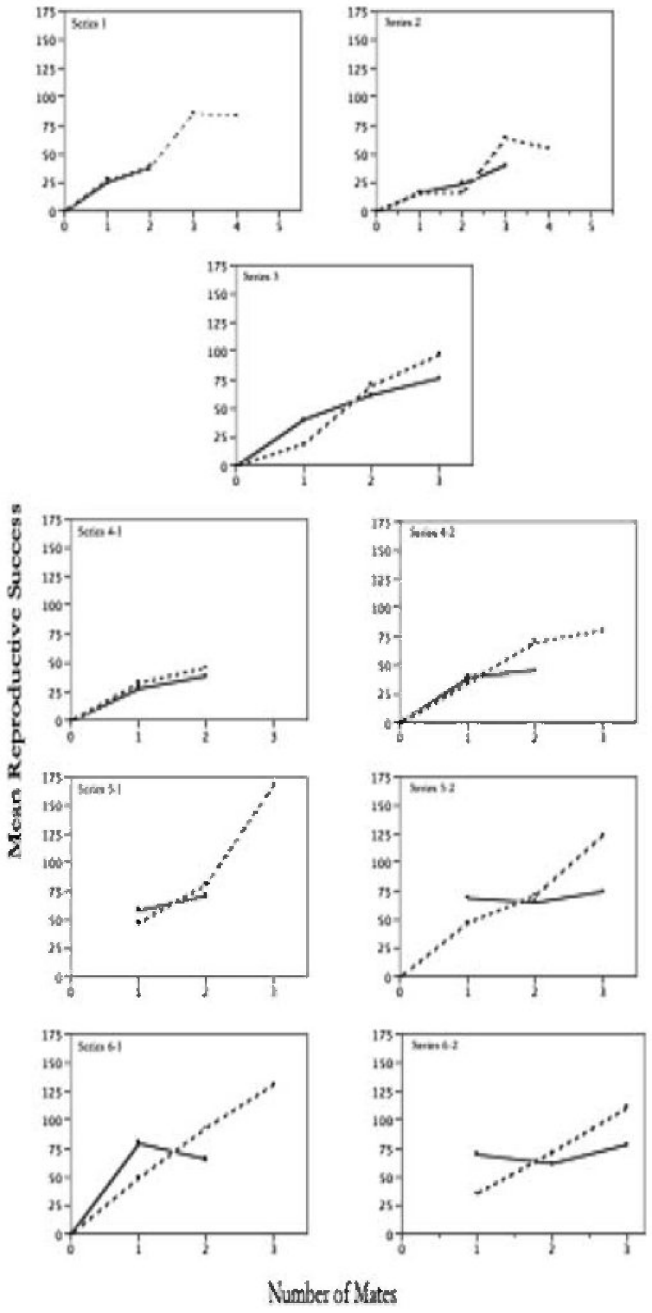

Figure 3. Separate graphs of Bateman's "series," Lines (solid = females and dotted = males) connect each value. Refer to Table 2 for a description of the methodological differences in each series.

be little or no genetic benefit for their offspring, which might explain the apparent plateau in reproductive success in series 4–6 (Fig. 5). However, perhaps the females in series 1–3, mated multiply to secure survival-enhancing genetic variation for their offspring, which could explain why female reproductive success did not plateau with number of mates in these outbred populations.

If one sets aside the methodological difficulties and sampling problems we have described, one can find support for Bateman's results in our reanalyzed VNM and VRS ratios and their probabilities (Table 4). Fisher's test for combined probabilities from independent tests (Sokal and Rohlf 1969) shows that both the recalculated VNM ratios and the VRS ratios in Bateman are significantly different, with greater male than female VNM and VRS (VNM: $\chi^2 = 38.02$, df $= 18$, $P < 0.005$; RSV: χ^2 43.04, df $=$ 18, $P < 0.001$). Yet, the problems with the mutants and unequal reproductive success for males and females (see above) remain, and it is unclear whether these sex differences in fitness variances are due to sexual selection or to some other cause (see below).

Theoretical Problems with Bateman's Study

It takes time to search, mate, and reproduce, and there is chance individual variation in each of these time-dependent processes. A proper theoretical analysis of Bateman's results would ask the question: can his results be explained by chance? Sutherland (1985) showed that a simple stochastic time-allocation model, which included chance effects on search time as well as fixed (nonheritable) sex differences in time between reproductive bouts (latency to receptivity to remating), could explain Bateman's data. Chance differences in life span (Hubbell and Johnson 1987) also affect VNM and VRS.

Because of the importance of Sutherland's theoretical challenge that Bateman's results could be due entirely to chance, we have explored chance effects on time allocation (arising from demographic stochasticity in survival and mate encounter rates and fixed latencies) more thoroughly using an individual-based simulation model, DYNAMATE (Gowaty and Hubbell 2005) for individuals in Bateman-like populations. DYNAMATE simulates mating and reproductive success for unique, named individuals in a population under demographic stochasticity in the context of time available for mating. DYNAMATE is related to Hubbell and Johnson (1987) but in contrast to their model, DYNAMATE models individuals, each described by their survival probability, mate encounter probability, and postmating latencies, which change dynamically over time as the social environment changes. In the null version of DYNAMATE, mating is random: there is no mate choice and no intrasexual competition. It is a null model because in this version of DYNAMATE, no sexual selection is possible. Individuals mate as they encounter receptive opposite sex individuals. Individuals move between states: searching for mates, mating, and in a latency period after a mating when they cannot mate. What determines the availability of an individual to mate is whether it is still alive, searching, or in latency. Whether it encounters another receptive individual depends on the number of opposite sex individuals alive, the encounter rate, and the number of others in latency.

We used DYNAMATE to calculate mating success over a very short mating "season," simulating the short durations of

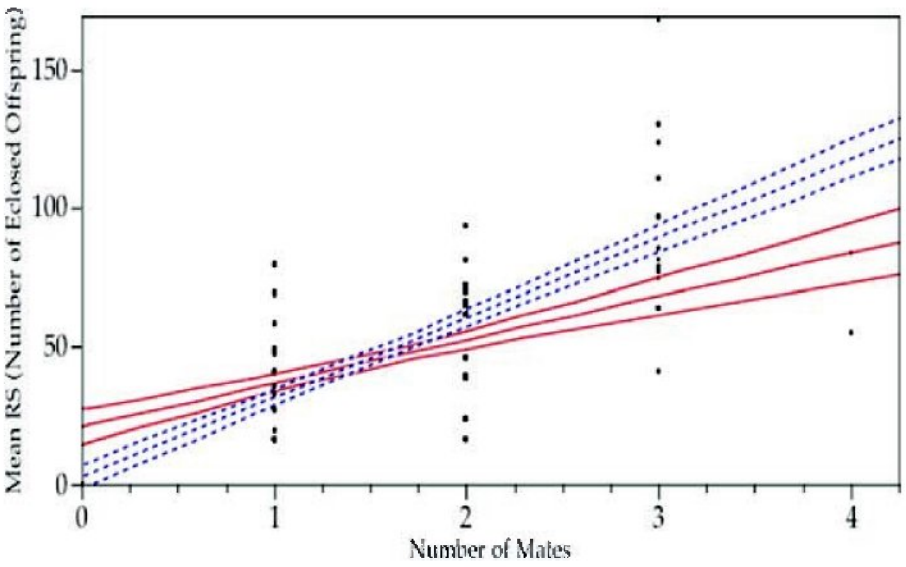

Figure 4. **Regression of the mean number of offspring for females (solid line) and males (dotted line) pooled over Bateman's two figures. The points indicate the means listed in tables from Bateman (1948, table 7 p. 360 and table 8 p. 361). For females Y = 12.95 + 20.83 X; for males Y = 5.4 + 28.5 X. The number of mates explains 53.5% of the variance in mean number of offspring for females ($P < 0.0001$) and 65.5% for males ($P < 0.0001$). The 95% confidence intervals around the mean number of offspring are for females 13 and 28.6; for males 21.2 and 35.9.**

Bateman's trials. We designed three ensemble runs that mimicked the size and sex ratio of Bateman's populations (Table 2). Over his series there were three combinations of population size and duration in days. We set survival probability at 0.995 per day because Bateman reported no deaths in his populations and all of the flies were relatively young. We set male latencies as one h. We set female latencies as 36 h because Bateman's results indicated that most females mated two times and some three times even in the three-day populations. This means that the latencies of Bateman's female flies had to have been shorter than two days. The population sizes in one ensemble of runs were set at three males and three females, with a duration of four days. In the second ensemble of runs, population size was three males and three females for three days; and in the third ensemble, five males and five females for three days. We set initial encounter probabilities at one for each sex, because the vials Bateman used were small. In each run of DYNAMATE, we calculated the number of mates for each individual in 1000 populations. We then calculated the experimentally observed VNM and their SD for each sex separately in each of the three combinations of population size and duration that were in Bateman (Table 2), and compared them to the model results for each population type (Fig. 6).

The expected ratios of male to female VNM due to chance demographic effects were not significantly different from the observed VNM ratios (Table 3) (Bateman mean = 2.87 ± 0.32; DYNAMATE mean = 3.14 ± 0.56 *t*-test; $P < 0.686$). Nor was there any evidence of excess VNM in males. Bateman's VNM exceeded those expected by chance in only one case—but it was for females, not males—from series 1 and 2, populations with five males and five females (Fig. 6). This result could be due to sexual selection among females or to sampling bias (Fig. 3) associated with differential mortality of offspring (discussed above). A way to connect this excess VNM in females to sexual selection would be to demonstrate a correlation of variation in some behavioral or morphological phenotype with variation in the number of mates (Clutton-Brock 1988). Unfortunately, Bateman did not describe the behavioral and morphological phenotypes (aside from the dominant mutations and individuals' ages) of his flies (Dewsbury 2005), and it is impossible to determine posthoc if any phenotypic variation in the females correlated with the observed excess in female VNM.

Thus, on theoretical grounds, even if one is willing to overlook the problems with methodology and data analysis, and accept Bateman's results as reliable, it is possible to explain most of the VNM of both females and males simply by demographic stochasticity, that is, chance variation in mate encounter and by nonheritable (fixed) sex differences in latencies. Another component of demographic stochasticity, survival differences, were minimal in our particular simulations of Bateman's experiments, but chance differences in reproductive life span have large impacts on measurements of VNM in natural populations (Hubbell and Johnson 1987; Gowaty and Hubbell 2005).

Future Directions

We conclude that the methodological and statistical problems with Bateman's experiment throw his conclusions and their generality

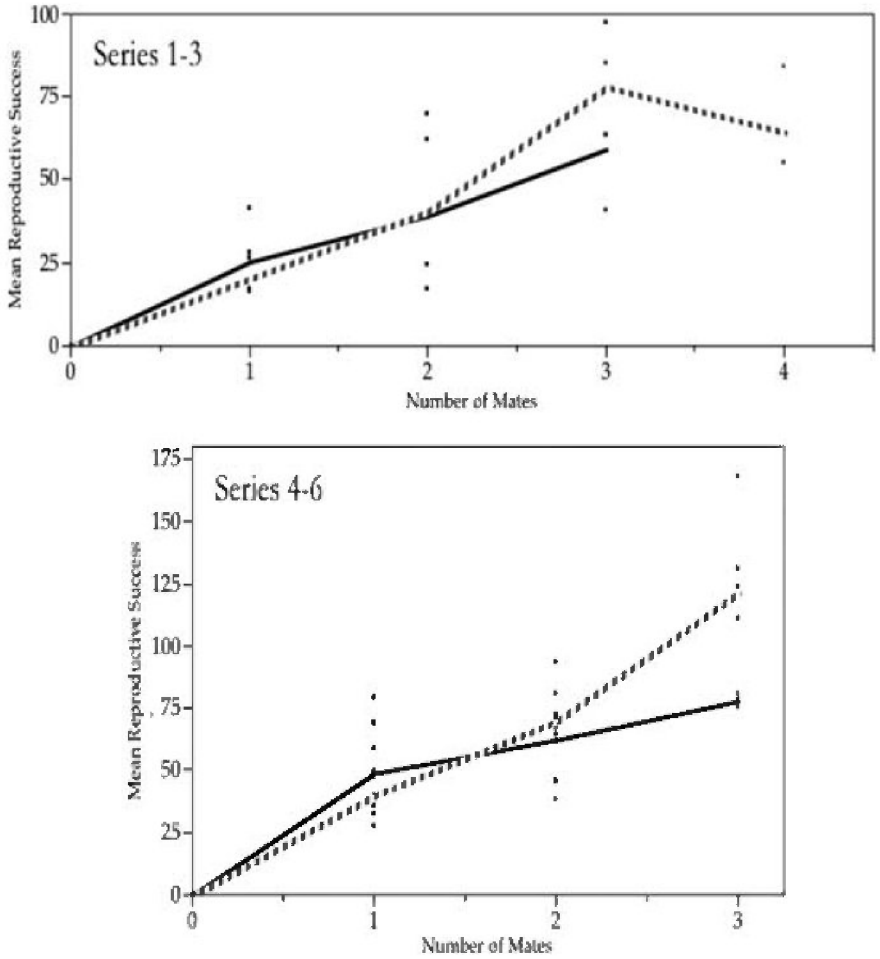

Figure 5. Bateman's data clumped by culture method, either outbred (series 1–3) or inbred (series 4–6). Lines (solid = females and dotted = males) connect each value.

into question. We call for repetitions of Bateman's experiment avoiding sampling biases, using appropriate statistical methods, and comparisons of observed VNM, VRS, and their ratios against expected values under appropriate null models, including demographic stochasticity.

Most significant for future study of sexual selection in *Drosophila* and other organisms will be an understanding of how chance affects VNM and VRS, which null models show will be unequal in the sexes if fixed reproductive latencies are unequal (Sutherland 1985; Hubbell and Johnson 1987; Gowaty and Hubbell 2005). Given that latencies in male and female *D. melanogaster* are different (Markow 1996), we expect that new observations will show greater VNM and VRS in males than females. Yet, we wonder what residual VNM and VRS will remain after chance effects are subtracted off. Will we discover unexplained VNM and VRS in females, for example? In fact, we wonder if sex differences in fitness variances are truly or correctly interpreted as "signs of intramasculine selection" or are merely a sign of fixed sex differences in life histories that contribute to demographic stochasticity.

Based on our results, we believe that Bateman's principles are better expressed as hypotheses or questions: (1) Given chance effects on VNM, when is it true that the sex with the higher VNM experiences more intense sexual selection? (2) Given chance effects on VNM, when is it true that the sex experiencing the more intense sexual selection has the higher VRS? (3) Given chance effects on VNM, when is it true that the slope of the relationship between number of mates and reproductive success is greater in the sex experiencing stronger sexual selection?

ACKNOWLEDGMENTS

We thank S. Hubbell for many discussions, for programming the modified version of DYNAMATE that calculated VNM under random mating, and for comments on the manuscript. Discussions with W. Anderson, Y-K. Kim, and J. P. Drury improved the manuscript and sharpened our analyses. A grant to PAG from the National Science Foundation (IOB-0545597) supported this research. BFS and PAG reanalyzed Bateman's data. BFS and PAG wrote the article.

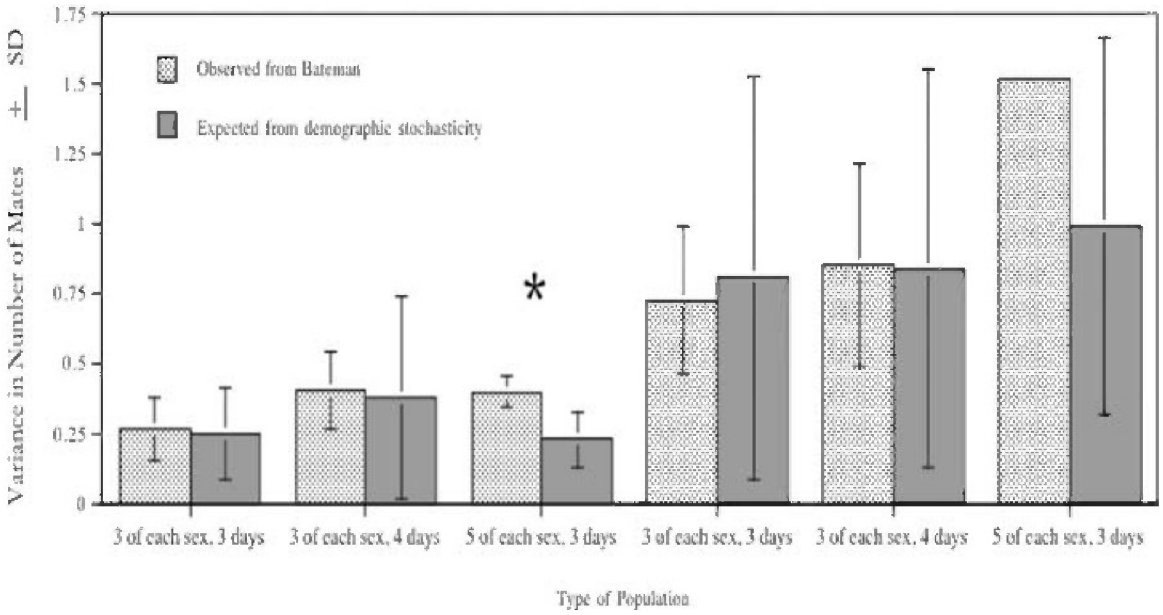

Figure 6. **Plot of Bateman's observed (stippled white) and DYNAMATE's expected (solid dark) variances in number of mates for females (first through third set of bars) and males (fourth through sixth set of bars).**

LITERATURE CITED

Andersson, M. B. 1994. Sexual selection. Princeton Univ. Press, Princeton, NJ.

Arnold, S. J. 1994a. Bateman's principles and the measurement of sexual selection in plants and animals. Am. Nat. 144:S126–S148.

———. 1994b. Is there a unifying concept of sexual selection that applies to both plants and animals? Am. Nat. 144:S1–S12.

Arnold, S. J., and D. Duvall. 1994. Animal mating systems: a synthesis based on selection theory. Am. Nat. 143:317–348.

Bartmann, S., and G. Gerlach. 2001. Multiple paternity and similar variance in reproductive success of male and female wood mice (*Apodemus sylvaticus*) housed in an enclosure. Ethology 107:889–899.

Bateman, A. J. 1948. Intra-sexual selection in *Drosophila*. Heredity 2:349–368.

Becher, S. A., and A. E. Magurran. 2004. Multiple mating and reproductive skew in Trinidadian guppies. Proc. R. Soc. Lond. B. 271:1009–1014.

Clutton-Brock, T. H. 1988. Reproductive success. Univ. of Chicago Press, Chicago.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness. 1988. Reproductive success in male and female red deer. Pp. 325–343 *in* T. H. Clutton-Brock, ed. Reproductive success. Univ. of Chicago Press, Chicago.

Darwin, C. R. 1871. Descent of man and selection in relation to sex. J. Murray, London.

Dewsbury, D. 2005. The Darwin–Bateman paradigm in historical context. J. Integr. Comp. Biol. 45:831–837.

Drickamer, L. C., S. H. Vessey, and E. M. Jakob. 2002. Animal behavior: mechanisms, ecology, evolution. McGraw Hill, New York.

Freeman, S., and J. C. Herron. 2001. Evolutionary analysis. Prentice Hall, Upper Saddle River, NJ.

Gowaty, P. A., and S. P. Hubbell. 2005. Chance, time allocation, and the evolution of adaptively flexible sex role behavior. J. Integr. Comp. Biol. 45:931–944.

Greer, B. W., and M. M. Green. 1962. Genotype, phenotype and mating behavior of *Drosophila melanogaster*. Am. Nat. 96:175–181.

Hafernik, J. E., and R. W. Garrison. 1986. Mating success and survival rate in a population of damselfiies: results at variance with theory. Am. Nat. 128:353–365.

Hubbell, S. P., and L. K. Johnson. 1987. Environmental variance in lifetime mating success, mate choice, and sexual selection. Am. Nat. 130:91–112.

Hulbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54:187–211.

Jensen, H., B. E. Saether, T. H. Ringsby, J. Tufto, S. C. Griffith, and H. Ellegren. 2004. Lifetime reproductive success in relation to morphology in the house sparrow *Passer domesticus*. J. Anim. Ecol. 73:599–611.

Krebs, J. R., and N. B. Davies. 1987. An introduction to behavioural ecology. Blackwell Scientific Publications, Oxford.

Le Boeuf, B. J., and J. Reiter. 1988. Lifetime reproductive success in northern elephant seals. Pp. 344–362 *in* T. H. Clutton-Brock, ed. Reproductive success. Univ. of Chicago, Chicago.

Manning, A. 1967. The control of sexual receptivity in female *Drosophila*. Anim. Behav. 15:239–250.

Markow, T. A. 1996. Evolution of *Drosophila* mating systems. Evol. Biol. 29:73–106.

———. 2000. Forced matings in natural populations of *Drosophila*. Am. Nat. 156:100–103.

Marti, C. D. 1997. Lifetime reproductive success in barn owls near the limit of the species' range. Auk 114:581–592.

McLain, D. K. 1991. Components of variance in male lifetime copulatory and reproductive success in a seed bug. Behav. Ecol. Sociobiol. 29:121–126.

Merila, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. Am. Nat. 155:301–310.

Merrel, D. J., and J. C. Underhill. 1956. Competition between mutants in experimental populations of *Drosophila melanogaster*. Genetics 41:469–485.

Oring, L. W., M. A. Colwell, and J. M. Reed. 1991. Lifetime reproductive success in the spotted sandpiper (*Actitis macularia*): sex differences and variance components. Behav. Ecol. Sociobiol. 28:425–432.

Parker, G. A. 1970. Sperm competition and its evolutionary consequences in the insects. Biol. Rev. 45:525–567.

Ribble, D. O. 1992. Lifetime reproductive success and its correlates in the monogamous rodent, *Peromyscus californicus*. J Anim. Ecol. 61:457–468.

Scott, D. K. 1988. Reproductive success in Bewick's swans. P. 538 *in* T. H. Clutton-Brock, ed. Reproductive success: studies of individual variation in contrasting breeding systems. Univ. of Chicago Press, Chicago.

Sharp, P. M. 1984. The effect of inbreeding on competitive male mating ability in *Drosophila melanogaster*. Genetics 106:601–612.

Sheridan, M., and R. H. Tamarin. 1988. Space use, longevity, and reproductive success in meadow voles. Behav. Ecol. Sociobiol. 22:85–90.

Smith, J. N. M. 1988. Determinants of lifetime reproductive success in the song sparrow. P. 538 *in* T. H. Clutton-Brock, ed. Reproductive success: studies of individual variation in contrasting breeding systems. Univ. of Chicago Press, Chicago.

Sokal, R. R., and F. J. Rohlf. 1969. Biometry. W. H. Freeman and Company, San Francisco.

Sutherland, W. J. 1985. Chance can produce a sex difference in variance in mating success and account for Bateman's data. Anim. Behav. 33:1349–1352.

Tang-Martinez, Z., and B. T. Ryder. 2005. The problem with paradigms: Bateman's worldview as a case study. J. Integr. Comp. Biol. 45:821–830.

Thomas, C. S., and J. C. Coulson. 1988. Reproductive success of Kittiwake gulls, *Rissa tridactyla*. P. 538 *in* T. H. Clutton-Brock, ed. Reproductive success: studies of individual variation in contrasting breeding systems. Univ. of Chicago Press, Chicago.

Trivers, R. L. 1972. Parental investment and sexual selection. P. 136–179 *in* B. Campbell, ed. Sexual selection and the descent of man. Aldine, Chicago.

Wade, M. J. 1979. Sexual selection and variance in reproductive success. Am. Nat. 114:742–747.

Wade, M. J., and S. J. Arnold. 1980. The intensity of sexual selection in relation to male sexual behavior, female choice, and sperm precedence. Anim. Behav. 28:446–461.

Associate Editor: Rausher