The Comparative Biology of Second Sex Ratio Evolution Within a Natural Population of a Parasitic Wasp, *Nasonia vitripennis*

Steven Hecht Orzack

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

Manuscript received April 5, 1989
Accepted for publication November 4, 1989

ABSTRACT

Correlation and regression analyses indicate that isofemale strains extracted from a population of the parasitic wasp, *Nasonia vitripennis*, differ in the fit of their second sex ratios (those produced in previously parasitized hosts) to the predictions of the theory of optimal facultative sex ratio adjustment. Under the theory's simple assumptions about population structure, there is significant heterogeneity of fitnesses among the isofemale strains. The reasons underlying these types of heterogeneity must be understood before we can make statements about the nature of sex ratio evolution in this species. These results suggest that comparative analyses are essential for testing the qualitative predictions of optimality models.

FURTHER advance in our understanding of sex ratio evolution depends upon developments in at least two areas. First, we need better information on the nature of genetic variation for sex ratio traits in wild populations (ORZACK and PARKER 1990). Second, we must develop our understanding of how to relate data and theory. Such an effort is essential for judging the success of theory at explaining natural patterns.

An important issue in relating data and theory is how predictions are to be interpreted. This issue takes on real meaning when one examines the claims made about the model of optimal facultative sex ratio adjustment developed by SUZUKI and IWASA (1980) and WERREN (1980). Consider the behavior of, say, a female wasp who encounters a host that has been previously parasitized. How should this second female adjust her sex ratio given her knowledge of the previous visit? These authors showed that the optimal proportion of males in the second brood decreases as the contribution of the second female increases relative to the first. Following the convention of ORZACK and PARKER (1986), this sex ratio is called the second sex ratio. To test the theory, one can plot the optimal second sex ratio versus the ratio of brood sizes, and compare this curve with data. Three such plots are shown in Figure 1. There are two problems. First, appearance of fit depends very much upon the scale of presentation. The abscissa scale used here was chosen simply to accommodate all three data sets. The second problem is obvious from inspection of the figure: the authors with the best "qualitative" fit of the theory to the data are agnostic about the theory. Moreover, the authors with data that "fit" less well regard the theory as confirmed [see also THORNHILL and ALCOCK (1983), page 71]. Clearly this ambiguity about the nature of appropriate conclusions is undesirable.

One purpose of this paper is to suggest solutions to this problem of interpretation. My general belief is that such ambiguity is almost impossible to avoid with a purely qualitative interpretation of theoretical predictions, especially since investigators inevitably differ in their notions about how accurate a theory in biology should be. Consequently, a quantitative framework must be used to test a theory. The framework I propose has two components. The first is a quantitative comparison of predictions and data since it is important to know whether a theory is accurate. Yet, one can object that it is wrong to take seriously the exact predictions of a simple theory. Accordingly, the second component is the testing of a reasonable set of biological entities for heterogeneity of fit to a qualitative prediction of a theory. The comparative aspect of this test is essential for statistical and biological reasons (see below). Such a comparative assessment of fit provides important insights. Knowing, for example, whether genotypes within the same population differ in their fit to a qualitative prediction of a theory helps determine both the theory's domain of explanation and its success at explanation.

I present various methods that can be used to make these strong and weak tests. The question I ask is whether there is variation among isofemale strains of *Nasonia vitripennis* in the fit of their second sex ratios to the predictions of SUZUKI and IWASA (1980) and WERREN (1980). This species is of particular interest. These authors as well as CHARNOV (1982) and THORN-

---

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

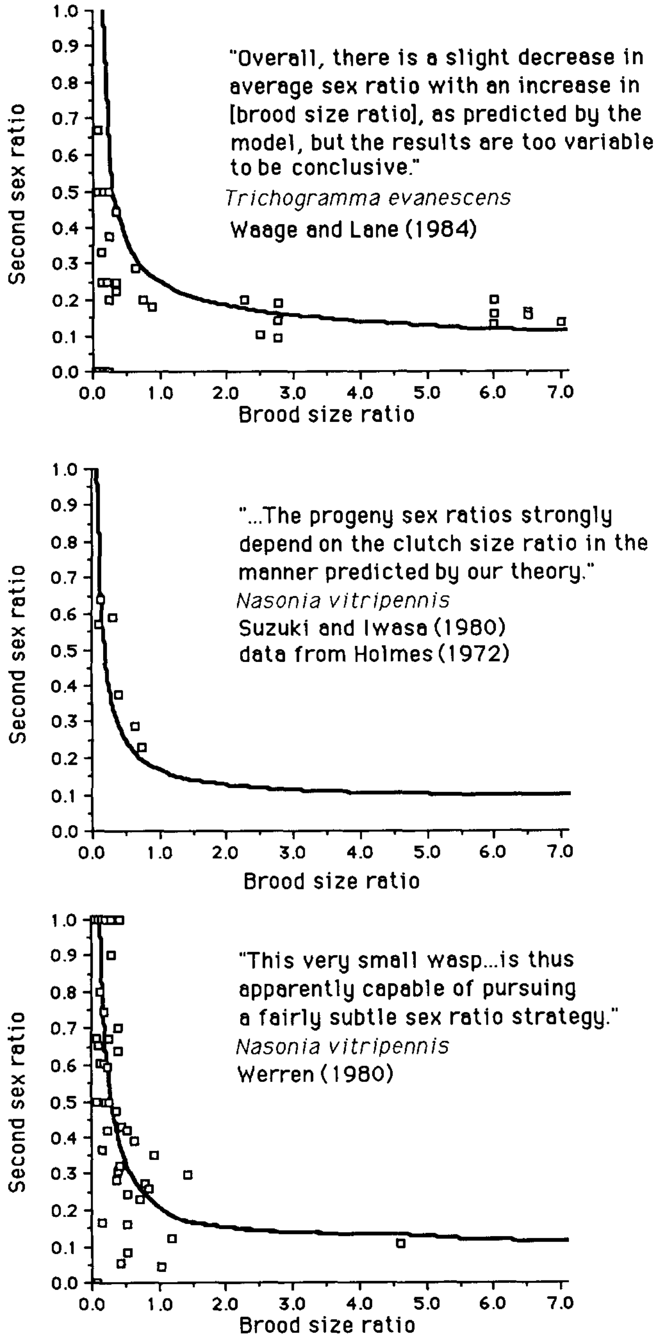


FIGURE 1.—Three claims about the relationship between observations and predictions of the theory of optimal facultative sex ratio adjustment. The optimal proportion of males is denoted by the curve.

HILL and ALCOCK (1983) claim that females of this species have an optimal ability to adjust their second sex ratios (see also WERREN 1987). As will be shown, there is circumstantial evidence of natural selection on this trait but I find no indication that it has evolved to an optimal state. I note that the suitability of this model for this species is not well demonstrated although I believe it is reasonably applied. Verification of the model assumptions that offspring of the two females mate randomly amongst themselves and that no other matings occur is clearly needed. Presumably, a model based upon different assumptions might lead to different evolutionary conclusions about the second sex ratios of this species.

MATERIALS AND METHODS

Experimental procedures: ORZACK and PARKER (1990) carried out sequential oviposition experiments with isofemale strains extracted from a population near Södertälje, a city approximately 20 km WSW of Stockholm, Sweden. We used strains *MS1*, *MS23*, *MS33*, *MS37*, *MS43*, *MS51*, *MS58*, *MS67*, and *MS92* in these experiments. (In Figure 7 I refer to these strains without using the *MS* prefix.) As mentioned in ORZACK and PARKER (1990), the intrapopulational analysis is important for the purposes of making relatively unambiguous comparisons among strains because there are no between-population differences in inbreeding to affect selection. Also, gross differences in ecological details presumably do not confound comparisons.

The basic experimental unit was a single host first parasitized for 24 hr by a female homozygous for the *stDR* eye color allele. She produces a "first brood" having a "first sex ratio." After removal from this female, the same host is than parasitized for 24 hr by a wildtype female who produces a "second brood" having a "second sex ratio." Offspring of each genotype are distinguishable by eye color.

All first and second sex ratios analyzed in this paper are calculated from the totals of direct-developing and diapause offspring (if any) in the brood.

In addition to the *stDR* first sex ratios and the wild-type second sex ratios, additional sex ratios of importance are those produced by *stDR* females in hosts subsequently offered to but *not* parasitized by a particular strain of wild-type female. They are first sex ratios but I call them "first-only sex ratios" to distinguish them from the *stDR* first sex ratios in doubly parasitized hosts. *stDR* first-only sex ratios are *associated* with particular wild-type strains by definition. ORZACK and PARKER (1990) showed that the *stDR* first-only sex ratios and brood sizes are statistically homogeneous whereas *stDR* first sex ratios and brood sizes are heterogenous. The implication is that there are differences among strains in the way they affect the first brood. These differences are of evolutionary interest (ORZACK and PARKER 1990) and are relevant to tests of the optimality model.

ORZACK and PARKER (1990) provides additional details about experimental design, strain history, and protocols.

Statistical procedures: Despite the simple appearance of the plots in Figure 1, complex statistical issues arise in the analysis of the relationship between second sex ratio and brood size ratio.

To develop a framework for testing the theory one must first ask: given a point in Figure 1, what is the expected point? Faced with such graphical presentations, the eye tends to "map" an observed point to the *nearest* point on the theoretical curve. However, one could reasonably use any of several mappings between observed and expected sex ratios. All are compatible with the notion that the sole object of natural selection is the second sex ratio. Each mapping reflects a different assumption about the precision of female behavior and of the measurements. For example, as shown in Figure 2, given a point one could assume that the expected point is at the intersection of the vertical line connecting the point and the curve (as the theory posits), or is the nearest point on the curve, or is at the intersection of the horizontal line connecting the point and the curve. The vertical mapping implies that the brood size ratio is meas-

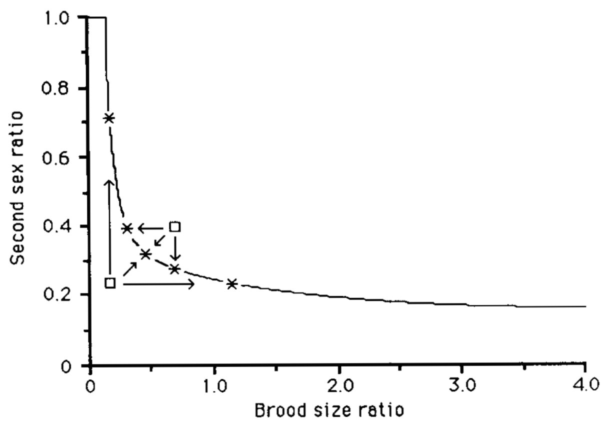


FIGURE 2.—Three mappings between observed points (□) and expected points (*) along the optimal curve.

ured correctly but that measurements of sex ratio are imprecise and/or that a female can only crudely adjust her sex ratio. The nearest mapping implies that both measurements are imprecise and/or that a female can only crudely adjust sex ratios and brood sizes. The horizontal mapping implies that the second female is optimally adjusting her sex ratio but that measurements of brood size are imprecise and/or that a female can only crudely adjust her brood size. I regard the last alternative as the least realistic since the expected brood size ratio is large given a low second sex ratio. Indeed, the expected brood size ratio is infinite given an all-female sex ratio. Such sex ratios were omitted from analyses based upon the horizontal mapping.

One feature of the mappings deserves emphasis. As shown in Figure 2, two points that are approximately equidistant from the curve can have different or similar expected points depending upon the mapping. Most analyses involve the first two mappings.

Given a mapping, one can use correlation analysis to make the two comparisons alluded to above. The first comparison is whether the 95% confidence interval of the correlation between the observed and expected points overlaps, say, 0.95. The second is whether there is between-strain heterogeneity of such correlations. In such a comparison, it is probably appropriate to make an assumption about the sign of the expected correlation, *i.e.*, that it be positive. Hence, this comparative test is powerful because it makes a weak presumption about the exact prediction of the theory.

In constructing tests of the theory it is essential to decide on a domain of explanation. In particular, previous tests have assumed that females have evolved to respond to the average environment. Accordingly, the *average* first sex ratio was used to generate the optimal curve. This may be reasonable but it is worth knowing whether females have an ability to adjust their second sex ratios in response to the *individual* first sex ratio. There is as much rationale for this as there is for using individual brood size ratios. Consequently, some correlation analyses used the expected second sex ratio given the first sex ratio in that host. This can result in negative expected second sex ratios or brood size ratios when the *stDR* first sex ratio is male-biased. Such second sex ratios were set equal to zero and the observed points retained. In the case of such brood size ratios, the observed points were eliminated from the analysis.

Given the nonnormality of the variables, Spearman rank correlations were calculated between observed and expected points. I used the procedures described by FIELLER, HARTLEY, and PEARSON (1957) and FIELLER and PEARSON (1961) for analyses of this correlation coefficient.

Despite its usefulness, there are problems with correlation analysis in this context. First, there is probably a violation of the assumption of constant error variance that underlies tests of correlations. Second, correlations have large sampling variances and type II errors may occur. Third, correlation measures only linear dependence.

To complement this analysis I used linear regression on the logit scale or logistic regression (a type of generalized linear model, NELDER and WEDDERBURN 1972; MCCULLAGH and NELDER 1983) to determine whether the expected second sex ratios derived from the theory (SUZUKI and IWASA 1980; WERREN 1980) predict the observed second sex ratios. Expected second sex ratios were generated using the vertical or the nearest mapping. The logit is $\ln(P/(1 - P))$ where P is the probability of being male in a particular brood. P is regarded as the parameter of a binomial random variable that describes the process by which sex ratio is determined. The estimates of regression coefficients for a particular regression model are those which maximize the log-likelihood function given that regression model and the observed second sex ratios (MCCULLAGH and NELDER 1983, p. 79).

I used this technique to test for a significant overall relationship between predictions and observations as well as for significant between-strain heterogeneity in the relationship between predictions and observations. In contrast to correlation analysis, this technique cannot be used to determine the absolute degree of concordance between predictions and observations. There are two reasons. First, this technique shares with all regression techniques the assumption that only the dependent variable (*i.e.*, observed second sex ratio) is measured with error. However, the expected second sex ratio is a function of the first sex ratio and brood size ratio which are most likely measured with error. Such errors bias the estimates of the regression coefficients (Cochran 1968). My important assumption is that there are no differences between strains in the degree of this bias. Second, it is difficult to determine the absolute degree of fit of such models (MCCULLAGH and NELDER 1983, p. 28).

There are two advantages to the logistic regression approach. The first is that, all other things being equal, it is biologically reasonable to regard a larger brood as providing a more precise estimate of sex ratio and an intermediate sex ratio as having more variance. Logistic regression allows one to incorporate this information since each point is weighted by its expected binomial sampling variance in the estimation procedure (NELDER and WEDDERBURN 1972, p. 372). Second, models are fit hierarchically. This allows for comparative testing of the theory and also provides a quantitative framework in which one can crudely judge the relative fit of competing hypotheses. In the present context, the last objective is not possible because the exact predictions of the competing hypotheses cannot be generated (see below).

An important question to ask is whether second sex ratios are accurately modelled as binomial random variables. Second sex ratios of this species tend to be overdispersed relative to the binomial expectation (ORZACK 1986). Consequently, the estimation of regression coefficients formally becomes a problem of finding the maximum solutions of a quasi-likelihood function (MCCULLAGH and NELDER 1983; WEDDERBURN 1974). In practice, these estimates are identical to the solutions of the log-likelihood function and one need only adjust their variances to account for overdispersion.

Determining the significance of the regression between the observed and expected second sex ratios proceeds by the analysis of deviance (MCCULLAGH and NELDER 1983, pp. 24–28). Changes in deviance are measured *across* models and are approximately distributed as $\sigma^2\chi^2$ where σ^2 is the dispersion parameter and χ^2 denotes the chi-square distribution (McCullagh and Nelder 1983, p. 174). σ^2 can be estimated for a particular regression model by the average deviance (= absolute deviance/degrees of freedom). The significance of this regression model relative to another is determined by comparing the scaled change in deviance (= change in deviance/average deviance) with the χ^2 table. The number of degrees of freedom for the test is equal to the number of additional parameter estimates in the current model relative to the previous model.

All second sex ratios were expressed as proportion males plus 0.01 so that all logits were finite. Calculations were carried out with GLIM, the Generalized Linear Interactive Modelling System Release 3.77 (available from the Numerical Algorithms Group, Inc.).

I also analyzed fitnesses of second sex ratios and fitness ratios (the fitness of an observed sex ratio divided by the fitness of the expected sex ratio). Nonparametric one-way analysis of variance (the Kruskal-Wallis test) was used to detect between-strain heterogeneity. Given a significant test statistic, a multiple comparison procedure (Conover 1980) was used to assess pairwise differences. This test is analogous to the least significant difference procedure developed for Gaussian data by Fisher and it controls the family-wise error rate of the comparisons at the given significance level.

RESULTS

Correlation analyses: Figure 3 shows scattergrams for the Waage and Lane data and the Werren data. One is hard-pressed to decide visually whether either species fits the predictions or even whether one fits better than the other. Both correlations are significant (Waage and Lane data $r_s = 0.389$, 95% confidence interval: 0.022–0.664, $n = 30$; Werren data $r_s = 0.597$, 95% confidence interval: 0.412–0.735, $n = 68$) but the theory fails the strong test since neither correlation has a 95% confidence interval which includes 0.95. The correlations are not significantly different ($\alpha = 0.05$, $\chi^2 = 1.39$, 1 d.f.) with a two-sided test or with a one-sided test predicated upon the different conclusions the authors reached about their respective data sets. Although the distinct biologies of the two species makes this an artificial comparison, one can conclude that the theory passes the weak test since both species meet a uniform qualitative criterion of fit. Interpretation of specific results is clearly another matter. My opinion is that both species fit the theory *weakly* although the data may also fit the predictions of other hypotheses (see below).

The results of correlation analyses of the Orzack and Parker data are presented in Tables 1 and 2. It is clear that the theory again fails the strong test since few of the correlations have a 95% confidence interval which includes 0.95. As expected, the correlations based upon the nearest mapping are higher than those based upon the vertical or horizontal mappings.

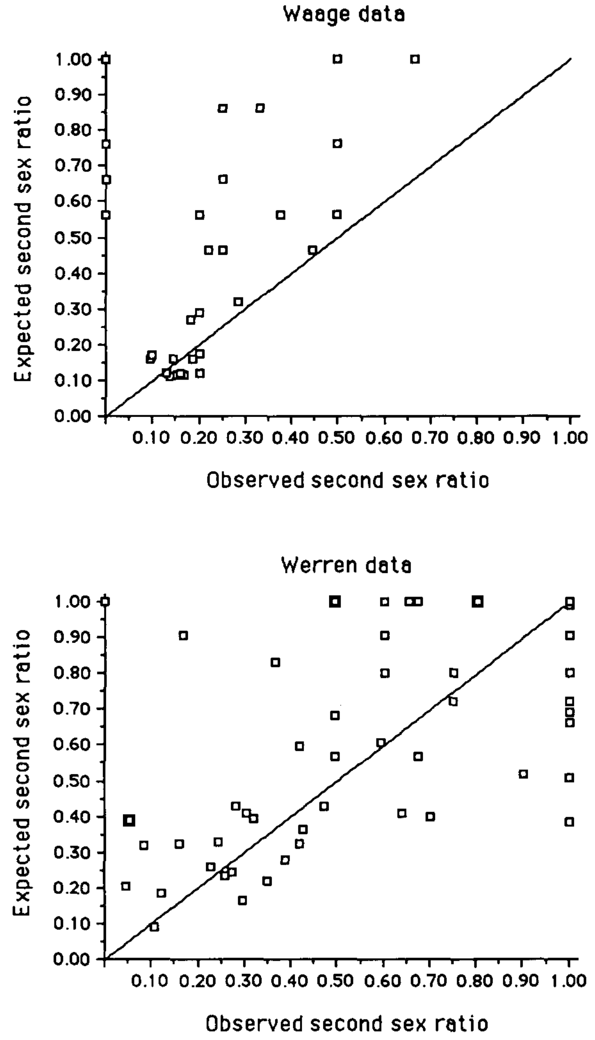


Figure 3.—Observed versus expected second sex ratios. The latter are derived from the vertical mapping (cf. Figure 2) implicit in the theory of optimal facultative sex ratio adjustment. Data from Waage and Lane (1984) and Werren (1980). The line denotes equal observed and expected sex ratios.

There is significant between-strain heterogeneity in five of eight sets of correlations. This heterogeneity reflects the different degrees to which sex ratios of the various strains conform to the theory.

There is an additional important question: would censoring of genetic data obscure the significant relationship between observed and expected sex ratios observed given the strain distinctions? This question is motivated by the claim underlying optimality theory that genetic information can be ignored in many types of analyses. It can be answered by comparing the overall correlations (no strain information) and the common correlations (strain information intact). In one instance (expected second sex ratios generated by individual *stDR* first sex ratios and the vertical mapping) the censoring of strain information would obscure the relationship between the observed and expected. However, the general effect of censoring strain information is to obscure significant between-strain heterogeneity and *not* to obscure the overall relationship between the observed and expected. The

TABLE 1

Spearman rank correlations between observed and expected second sex ratios

		Expected from the vertical mapping			Expected from the nearest mapping		
Expected based on	Strain	r_s	95% C.I.	n	r_s	95% C.I.	n
Individual *stDR* first sex ratio	*MS1*	0.113	−0.117–0.332	79	0.310	0.089–0.502	79
	MS23	−0.029	−0.286–0.232	61	0.051	−0.211–0.306	61
	MS33	0.168	−0.089–0.404	64	0.195	−0.061–0.427	64
	MS37	−0.035	−0.315–0.251	51	0.161	−0.128–0.425	51
	MS43	0.283	0.024–0.507	60	0.476	0.245–0.656	60
	MS51	−0.183	−0.415–0.071	65	−0.074	−0.319–0.180	65
	MS58	0.368	0.024–0.634	34	0.418	0.083–0.668	34
	MS67	0.227	−0.023–0.450	66	0.408	0.177–0.596	66
	MS92	0.061	−0.232–0.344	49	0.245	−0.047–0.499	49
	Overall	0.078	−0.010–0.165	529	0.216	0.131–0.298	529
	Common	0.099	0.010–0.188		0.243	0.157–0.326	
		$\chi^2 = 12.22$, 8 d.f., $P > 0.05$			$\chi^2 = 16.14$, 8 d.f., $P < 0.05$		
Average *stDR* first only sex ratio	*MS1*	0.469	0.270–0.629	79	0.838	0.754–0.895	79
	MS23	0.402	0.160–0.599	61	0.651	0.472–0.779	61
	MS33	0.535	0.326–0.694	64	0.843	0.750–0.903	64
	MS37	0.246	−0.040–0.495	51	0.517	0.274–0.698	51
	MS43	0.239	−0.024–0.471	60	0.594	0.394–0.740	60
	MS51	0.416	0.184–0.604	65	0.573	0.376–0.720	65
	MS58	0.568	0.275–0.765	34	0.764	0.567–0.878	34
	MS67	0.372	0.136–0.568	66	0.590	0.400–0.732	66
	MS92	0.010	−0.280–0.298	49	0.352	0.070–0.582	49
	Overall	0.251	0.167–0.332	529	0.660	0.607–0.707	529
	Common	0.376	0.296–0.450		0.673	0.621–0.720	
		$\chi^2 = 13.89$, 8 d.f., $P > 0.05$			$\chi^2 = 37.51$, 8 d.f., $P < 0.001$		

C.I. = confidence interval.

same conclusion is reached with regression analysis (see below). It will be interesting to determine whether this conclusion holds for other applications of optimality theory.

Logistic regression analyses: One aspect of the biology of second sex ratios has been neglected in previous tests of the theory of optimal facultative sex ratio adjustment (*e.g.*, SUZUKI and IWASA 1980; WERREN 1980; WAAGE and LANE 1984; ORZACK 1986). When testing for independent assortment in a backcross, for example, one has no a priori interpretation about the selective consequences of the deviations themselves. Consequently, one can simply test the predictions in an unweighted manner. In contrast, the theory of facultative sex ratio adjustment is inherently "about" fitness. As a result, testing of fit and interpretation of results must be conditioned upon the fitness consequences of deviations. In essence this amounts to answering the following questions: is the selective surface associated with the optimal prediction steep enough that I should expect close adherence to the theory? Alternatively, is it flat enough that *any* sex ratio achieves a high proportion of the fitness associated with the optimum sex ratio?

The fitness surface associated with Suzuki and Iwasa's and Werren's model is shown at the top of Figure 4. One way to judge its steepness is by plotting isopleths of the fitness ratios. The positions of these isopleths are dependent upon the mapping used. Those shown in the bottom of Figure 4 are generated by the vertical mapping implicit in the theory.

An important observation to be drawn from the isopleths is that a nonoptimal sex ratio has different selective consequences depending upon the brood size ratio. So, for example, given an observed second sex ratio greater than the optimum by 0.20, the fitness ratio is greater than 0.99 when brood size ratio is 0.3 whereas it is approximately 0.90 if the brood size ratio is 4.0. These differences give some indication of the expected nature of fit of the theory. Thus, all other things being equal, the observed second sex ratios should better match the expected when the brood size ratio is high. This is one reason why it is essential to measure second sex ratios over a reasonably wide range of brood size ratios.

I incorporated the fitness scale into the analyses by doing weighted regressions in addition to an unweighted regression. The weights were the fitness ratio, the fitness of the expected point, and the second brood size. Each has a different rationale. Given the vertical mapping, the fitness ratio can be regarded as a measure of efficiency since all such ratios are less than or equal to 1.0. Given the nearest mapping, this weight is less interpretable since the ratio can be greater than one and, consequently, points deviating from the optimum can be given greater relative

TABLE 2

Spearman rank correlations between observed and expected brood size ratios

		Expected from the horizontal mapping			Expected from the nearest mapping		
Expected based on	Strain	r_s	95% C.I.	n	r_s	95% C.I.	n
Individual *stDR* first sex ratio	*MS1*	0.709	0.564–0.812	70	0.821	0.730–0.884	79
	MS23	0.528	0.303–0.697	57	0.946	0.910–0.968	61
	MS33	0.430	0.190–0.621	60	0.865	0.784–0.917	64
	MS37	0.528	0.276–0.712	47	0.893	0.816–0.939	51
	MS43	0.473	0.216–0.669	50	0.927	0.879–0.957	60
	MS51	0.300	0.040–0.522	59	0.959	0.933–0.975	65
	MS58	0.748	0.523–0.876	30	0.974	0.947–0.987	34
	MS67	0.281	0.019–0.507	59	0.920	0.870–0.951	66
	MS92	0.362	0.043–0.614	39	0.908	0.839–0.948	49
	Overall	0.490	0.416–0.558	471	0.915	0.900–0.928	529
	Common	0.494	0.419–0.563		0.918	0.903–0.931	
		$\chi^2 = 19.81$, 8 d.f., $P < 0.025$			$\chi^2 = 37.81$, 8 d.f., $P < 0.001$		
Average *stDR* first only sex ratio	*MS1*	0.592	0.417–0.724	76	0.793	0.690–0.865	79
	MS23	0.492	0.263–0.669	59	0.876	0.798–0.925	61
	MS33	0.533	0.324–0.693	64	0.822	0.719–0.890	64
	MS37	0.405	0.128–0.623	48	0.887	0.806–0.935	51
	MS43	0.527	0.297–0.699	55	0.878	0.800–0.927	60
	MS51	0.429	0.196–0.616	63	0.944	0.908–0.966	65
	MS58	0.691	0.447–0.839	33	0.909	0.821–0.955	34
	MS67	0.480	0.259–0.654	64	0.927	0.882–0.955	66
	MS92	0.321	0.005–0.578	41	0.890	0.809–0.938	49
	Overall	0.514	0.445–0.577	503	0.882	0.861–0.900	529
	Common	0.503	0.431–0.569		0.885	0.864–0.903	
		$\chi^2 = 6.72$, 8 d.f., $P > 0.05$			$\chi^2 = 22.70$, 8 d.f., $P < 0.01$		

C.I. = confidence interval.

weight. The rationale for the fitness weight is that a female should be more precise when the potential fitness payoff is greater. The rationale behind the brood size weight is similar: a female should be more precise when making a sex ratio decision involving a greater number of offspring.

Figure 5 contains all analyses involving the vertical mapping while Figure 6 contains those involving the nearest mapping. Only the scaled changes in deviance are shown.

For each weight a hierarchical set of regression models was fitted. The basal or null model (C) is a constant relationship (no significant regression) between observed and expected second sex ratios. From this one can either test for strain-specific constants ($C + \alpha_i$) or for an overall regression between observed and expected second sex ratios ($C + \beta x$). Given $C + \alpha_i$, one can test $C + \alpha_i + \beta x$. Finally, one can test for strain-specific regressions between observed and expected second sex ratios ($C + \beta x$ versus $C + (\beta + \gamma_i)x$; $C + \alpha_i + \beta x$ versus $C + \alpha_i + (\beta + \gamma_i)x$).

All analyses reveal a significant regression between the overall expected and observed second sex ratios (C *vs.* $C + \beta x$; $C + \alpha_i$ *vs.* $C + \alpha_i + \beta x$). For example, given the vertical mapping (see Figure 5), the χ^2 value associated with the former transition in the unweighted analysis is 35.26. This value is highly significant for one degree of freedom. Hence, as in the correlation analyses, the theory predicts the overall trend of observed second sex ratios. Yet, in seven of eight analyses ($C + \beta x$ *vs.* $C + (\beta + \gamma_i)x$) or in five of eight analyses ($C + \alpha_i + \beta x$ *vs.* $C + \alpha_i + (\beta + \gamma_i)x$) there is also a significant change in deviance when fitting strain-specific regression coefficients. Hence, there is heterogeneity among strains in the manner in which they conform to theoretical predictions.

Unfortunately, in the brood size weighted analysis in Figure 5 and the unweighted analysis in Figure 6 the significance of the complete regression model relative to a simpler model is dependent upon the sequence of models fitted. This is a result of nonorthogonality of the estimates (McCullagh and Nelder 1983, p. 27). Although not unexpected on biological grounds, this dependency does make complete interpretation of these analyses difficult.

The relative magnitudes of changes in deviance associated with model transitions can be used to judge the importance of components of the regression. These changes are reflected in the significance level of transitions. Thus, in the fitness ratio weighted analysis in Figure 6, the significance of fitting an overall regression (C *vs.* $C + \beta x$) is much greater than the significance of fitting strain-specific regression coefficients ($C + \beta x$ *vs.* $C + (\beta + \gamma_i)x$). One can *crudely* state that the overall relationship is more important than the strain-specificity. This conclusion is not gen-

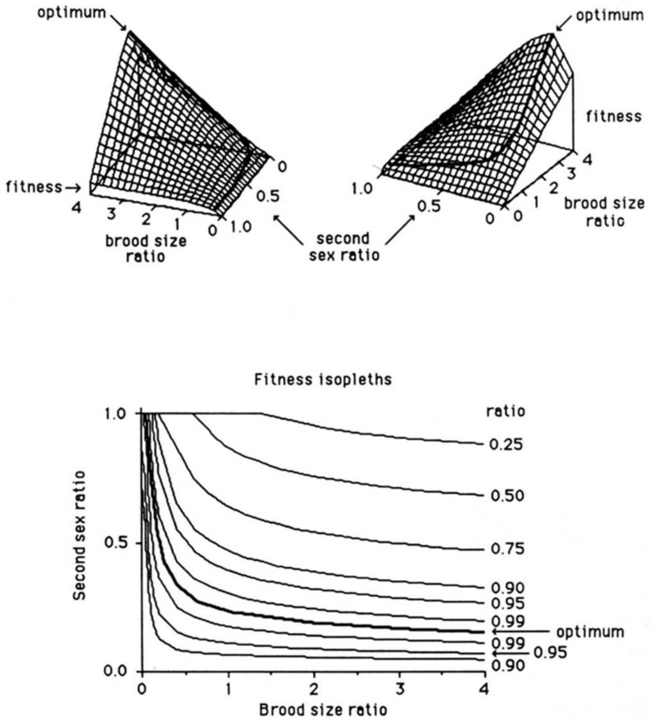


FIGURE 4.—A, Two views of the fitness surface associated with the theory of optimal facultative sex ratio adjustment. The optimal proportion of males is denoted by the dark line. This surface was generated using the overall average *stDR* first only sex ratio (proportion males = 0.382). B, Lines of constant fitness ratios based on the fitness surface in A. The expected second sex ratios are derived from the vertical mapping.

eral as is clear from inspection of this particular transition in all of the analyses. To some extent then the conclusion one draws about the evolutionary significance of the overall relationship and of strain-specificity depends upon which weight one regards as most realistic. This issue is addressed in the DISCUSSION.

Analyses of fitnesses: Testing for heterogeneity of absolute fitness distributions amounts to testing for differences in location of sex ratios on the fitness surface. Since changing the first sex ratio changes the height of the implied fitness surface, I used the average first sex ratio to make the test more conservative. One could use the average *stDR* first-only sex ratio or the average *stDR* first sex ratio (see above and Orzack and Parker 1990). I used the former average (proportion males = 0.382, $n = 248$) on the presumption that it reflected the average environment to which wild-type females responded. Again, this is a conservative assumption since it eliminates differences in the local mating environment *created* by the second female for her offspring.

There is significant heterogeneity of observed fitnesses among strains. The results of the multiple comparison procedure are shown in Figure 7. Most of the significant pairwise comparisons involve *MS1* and *MS33*. The fitnesses of these two strains do not differ but their second sex ratios do (Orzack and Parker 1990, Figure 4). Conversely, another strain (*MS23*) differs in second sex ratios from most other strains but is generally not distinct at the fitness level. What then is the proximate cause of the fitness heterogeneity? Given the theory, the fitness of a second sex ratio is a function of the first sex ratio, the brood size ratio (= second brood size/first brood size), and the second sex ratio. Consider these components in turn. The fitnesses were calculated with the average *stDR* first-only sex ratio. Hence, the implied fitness surface is the same for all strains and this does not contribute to the fitness heterogeneity. What about the brood size ratio? There is no significant between-strain heterogeneity among the second brood sizes (Kruskal-Wallis test $H = 12.154$, 529 d.f., $P > 0.1$) although there is among the *stDR* first broods (Orzack and Parker 1990). Finally, there is significant heterogeneity among second sex ratios and among *stDR* first sex ratios (Orzack and Parker 1990). Hence, the fitness heterogeneity appears to be due to a size-independent interaction between the second brood and the first brood size and sex ratio.

Analyses of fitness ratios: Testing for heterogeneity of fitness ratio distributions amounts to testing whether strains differ in the degree to which they can achieve the expected second sex ratios.

The fitness ratios generated by the vertical and horizontal mappings are not significantly heterogenous ($\alpha = 0.05$: vertical mapping $H = 6.197$, 529 d.f.; horizontal mapping $H = 11.969$, 503 d.f.). There is between-strain heterogeneity of the fitness ratios generated by the nearest mapping ($H = 19.018$, 529 d.f., $P < 0.025$). Multiple comparison tests (not shown) indicate that seven of nine significant ($\alpha = 0.05$) pairwise differences in ratios involve either *MS1* or *MS92*. Hence, there is no necessary overlap between those strains with distinct fitness ratios and those with distinct observed fitnesses (*cf.* Figure 7).

The interpretation of these results is not clear. As noted above, the ratio associated with the vertical mapping can be interpreted as a measure of efficiency. It is intriguing that strains do not differ in this measure of efficiency but do differ in their observed fitnesses.

DISCUSSION

Alternative hypotheses: The theory of optimal facultative sex ratio adjustment predicts an inverse relationship between second sex ratio and brood size ratio as shown in Figure 1. However, such an inverse relationship could result from a second female's adjustment of her sex ratio in response to resource quality (Charnov *et al.* 1981, A. Taylor, personal communication). Also, second sex ratio and brood size ratio can have a spurious negative correlation (Orzack 1986).

In general, one can use both correlation analysis

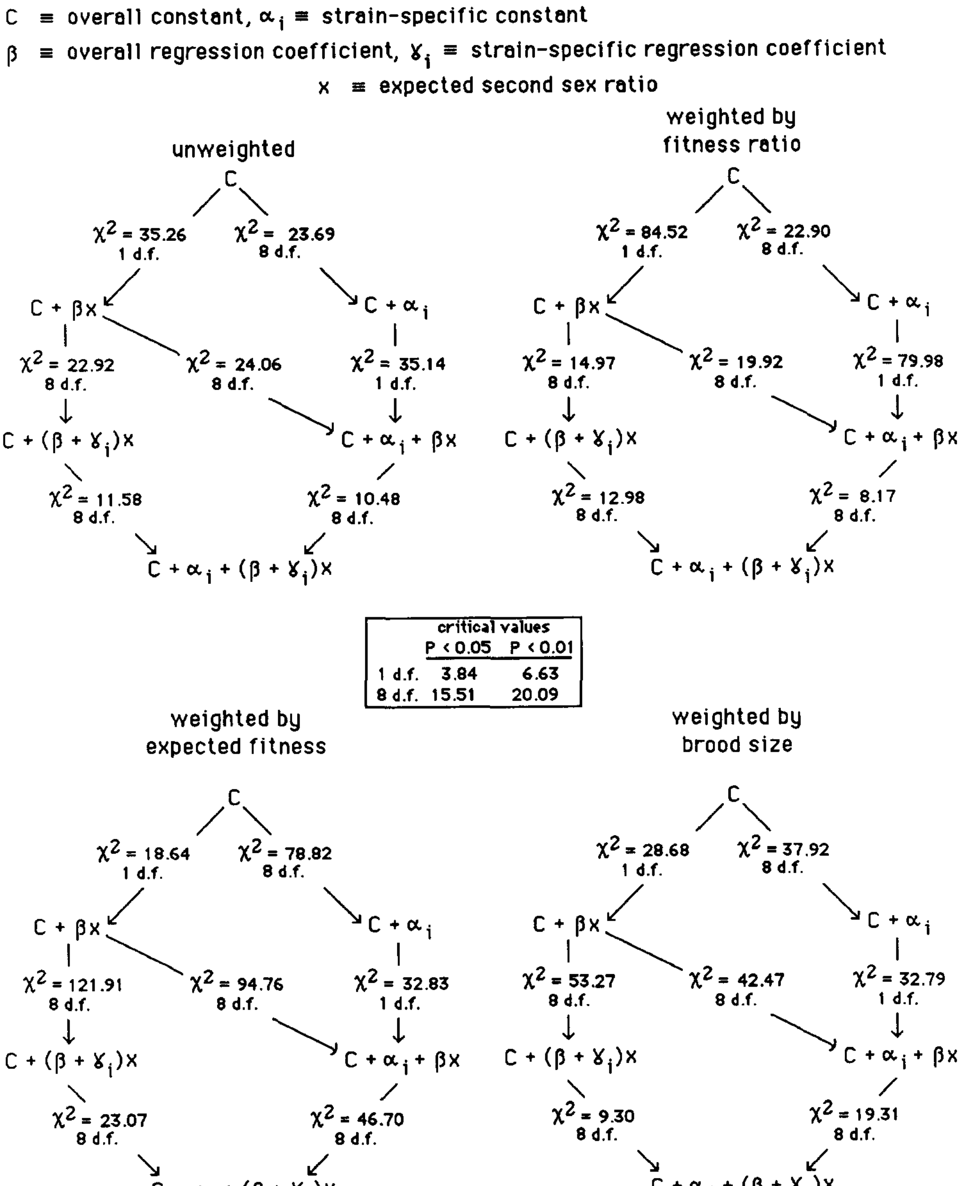
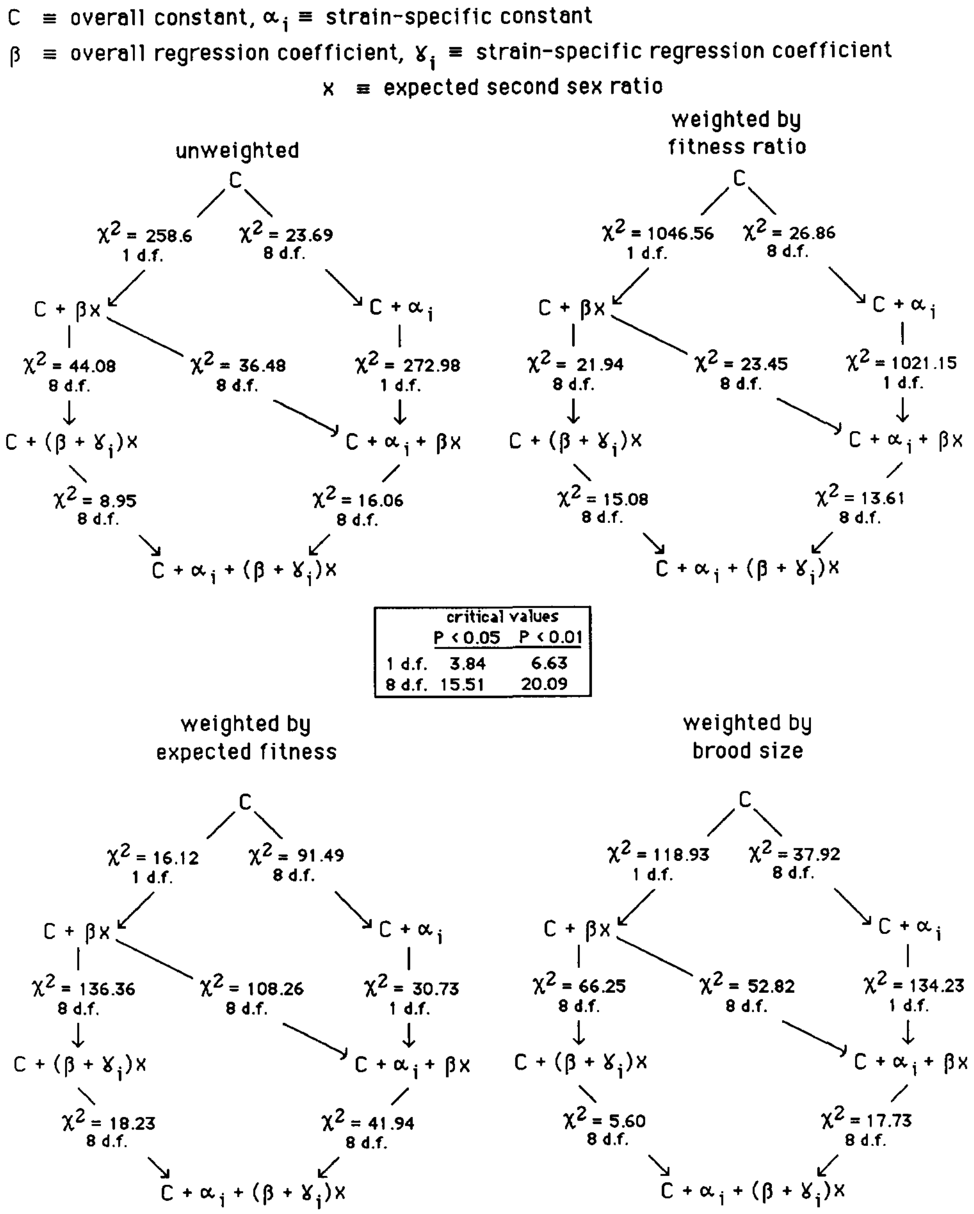


FIGURE 5.—Logistic regression analyses of the relationship between observed and expected second sex ratios. The latter are derived from the vertical mapping.

and regression analysis to test rival hypotheses. Thus, for example, if the correlation between observed and expected second sex ratios is of the magnitude expected of a spurious correlation then it would be parsimonious to accept this nonselective hypothesis in lieu of the selectively based facultative sex ratio adjustment hypothesis. Similarly, suppose one fits a regression model between the observed second sex ratios and those expected given a spurious correlation between second sex ratio and brood size ratio. One can then calculate the change in deviance relative to a model assuming no regression between observed and expected second sex ratios. Further suppose one fits a regression model between the observed second sex ratios and the expected sex ratios given the facultative sex ratio adjustment hypothesis. The change in deviance for *this* model is calculated relative to the model based on the spurious correlation. If the former change in deviance is significant and the latter is not then it is reasonable to choose the spurious correlation explanation for the observed relationship between second sex ratio and brood size ratio. Even if both changes in deviance are significant and the former change is much larger than the latter, one could crudely judge that the nonselective explanation accounts for a larger component of the relationship.

Unfortunately, in the present context, the expected second sex ratios given either the spurious correlation or resource limitation hypotheses are not calculable for empirical and theoretical reasons. There is, however, experimental evidence mitigating against these hypotheses (although additional evidence is definitely needed). First, not all strains of this species produce second sex ratios which are positively correlated with the expected sex ratios (Table 1 and ORZACK 1986). Hence, unless one wants to invoke strain-specificity in



	critical values	
	$P < 0.05$	$P < 0.01$
1 d.f.	3.84	6.63
8 d.f.	15.51	20.09

FIGURE 6.—Logistic regression analyses of the relationship between observed and expected second sex ratios. The latter are derived from the nearest mapping.

sensitivity to density or in the potential for spurious correlations, it is parsimonious to believe that positive correlations are at least partially due to the action of natural selection for facultative sex ratio adjustment ability under conditions of local mate competition. In addition, exposure times of hosts to females in sequential oviposition experiments (WERREN 1980; ORZACK and PARKER 1986, 1990) have ranged from five hours to 48 hr. These time differences presumably result in different degrees of resource competition since brood size is generally correlated with exposure time. Nonetheless, the results of all of these experiments are generally concordant as would not be expected given the resource competition hypothesis.

Relevance of these results to nature: It is not clear whether laboratory and natural conditions during superparasitism are similar. One often presumes that laboratory conditions are less variable than natural conditions. To that extent, laboratory conditions should make it easier to detect facultative sex ratio adjustment and they presumably allow wasps to make more precise sex ratio decisions. Consequently, the observations of imprecision of second sex ratios (ORZACK 1986) and of between-strain heterogeneity of second sex ratios are more meaningful. However, all of these arguments beg the question as to whether laboratory conditions inherently confuse wasps to such a degree that all sex ratio experiments are suspect. This may be true but one must then devise a convincing reason why first sex ratios are generally so much more consistent than second sex ratios. Given this observation, I regard laboratory conditions as relatively congenial to wasps.

The use of *stDR* females in these experiments also deserves comment. This is unavoidable if one wants to unequivocally distinguish between the offspring of

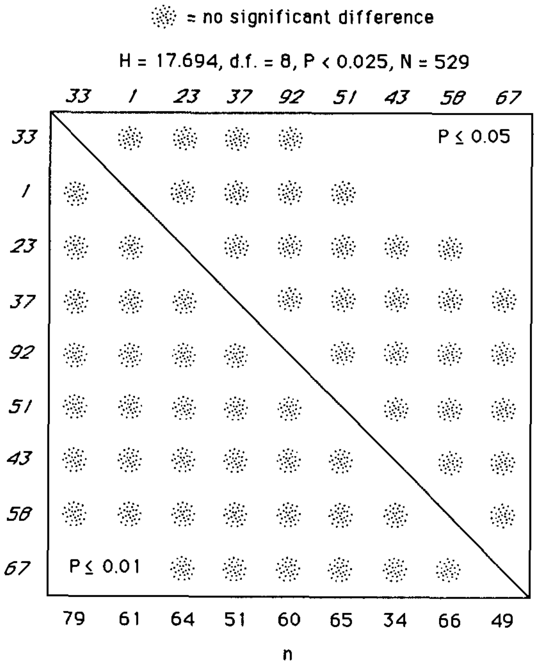


FIGURE 7.—Results of a multiple-comparison procedure applied to the fitnesses of the second sex ratios produced by each strain. The procedure is based upon a significant Kruskal-Wallis test statistic (*H*). I show the value of *H*, the associated probability, the total sample size (*N*), and the sample size for each strain (*n*). Strains are arranged from lowest (*MS33*) to highest (*MS67*) average rank.

different females without disrupting the host. However, this is not a genotype from the Södertälje population and one can reasonably wonder: would second sex ratios be more precise if the first female were from that population? There are two ways to answer this question. First, not all previously measured strains exhibit highly variable second sex ratios when *stDR* females are first (ORZACK 1986). Hence, females do not appear to be inherently "confused" by this situation. Second, these experiments can be viewed as relevant only to previous attempts with this species to assess the validity of the theory. All such attempts have used this genotype as first or second female.

Individual *vs.* average behavior: This is an important distinction. Perhaps a female can only crudely adjust a particular second sex ratio but her average behavior is optimal. Accordingly, one might test the theory with average second sex ratios. I did not do so for two reasons. The first is that visual inspection of the fit of such averages to the predictions of the theory reveals no obvious improvement in fit (ORZACK 1986). Second, at best there are only four measurements of second sex ratios for each female in the present experiments.

Alternative regression models: Figures 5 and 6 each contain four analyses which differ in their weights. Which regression analysis is most realistic? Significance levels cannot be assigned to absolute deviances (MCCULLAGH and NELDER 1983, p. 28) so one cannot decide among the four analyses on this basis. An additional consequence is that it is impossible to know, given the same regression terms, how a particular weight improves model fit relative to another weight.

Are there biological reasons for choosing among the various weights? I believe that the expected fitness weighting is the most realistic given its simple logic: a female should do better when the stakes are higher. In this regard, it is satisfying that this weight and the brood size weight have similar motivations and result in almost identical patterns of significant scaled deviances (Figures 5 and 6) although the latter weight is not "theory-bound" like the former.

The unweighted and weighted models have similar patterns of significant scaled deviances. That weights make no qualitative difference in this instance does not have general implications about the importance of incorporating such information. The fitness dimension should always be incorporated into tests of evolutionary models.

The evolutionary significance of these results: An obvious question raised by these results is: what accounts for between-strain heterogeneity in fit to the theory and for between-strain heterogeneity of fitnesses? There are at least five answers that are plausible. First, there may be little selection in nature for precise adjustment of second sex ratios. Most hosts may be singly parasitized for behavioral or ecological reasons. Second, such selection may be common but females cannot produce optimal second sex ratios. For example, a female may not be able to precisely control release of sperm. Third, second sex ratio may be genetically correlated with another trait in such a way that genetic variation is maintained at equilibrium. Fourth, selection on second sex ratios may proceed in the manner posited by the theory but there may be stable selective equilibria not revealed by the optimality analysis. Finally, selection on second sex ratios may proceed in a different manner than implied by the theory.

At present I regard the first alternative to be the most plausible for the following reason. Females given the choice between unparasitized and previously parasitized hosts generally prefer to oviposit in the former (S. H. ORZACK and E. D. PARKER, unpublished results). Such a behavior could have evolved for many different reasons. Females may not be able to accurately judge the number of previous visitors or how many eggs were oviposited. Indeed, females may not be able to determine whether a previous visitor was conspecific. Females will produce more males in hosts previously parasitized by females of other species (WYLIE 1973). That females do so although no local mate competition can occur suggests that they cannot distinguish these hosts from those parasitized by conspecifics. (An alternative adaptive explanation is that females are not "expecting" local mate competition and produce more males because this sex will do better given the smaller amount of resources per offspring caused by the presence of the other brood). An assumption of the hypothesis that avoidance is preferred is that the cost to fitness of an incorrect second sex ratio outweighs the cost of searching for a virgin host. It is not obvious that this should be generally true.

The absence of natural selection on second sex ratios would be in accord with the traditional explanation (*e.g.*, FISHER 1930) of apparent fitness variation, *i.e.*, that the trait contributes little or nothing to overall fitness. This may be true but it strikes me as special pleading especially given the fitness variation associated with first sex ratios (ORZACK and PARKER 1990). Resolution of this issue clearly depends upon the acquisition of better information on the frequency of multiple parasitization in this and other natural populations. If this frequency is reasonably high then it will be harder to argue that selection on second sex ratios does not occur.

Although I think the alternative explanations listed above are less plausible they do relate to important experimental and theoretical issues. Alternative two highlights the need for better information on the biology of sex ratio traits at the individual level. For example, there is no published information on the nature of within-female correlations between first and second sex ratios. There is no information on genetic correlations. Indeed, nothing is known about the genetic structure of these traits in the Södertälje strains. Alternative three highlights the need for better information on the biology of sex ratio traits at the population level. ORZACK and PARKER (1990) showed that the average values of first and second sex ratios are uncorrelated between strains. Alternative four highlights the need for expansion of our understanding of the present model of facultative sex ratio adjustment while alternative five highlights the need for consideration of more realistic models of sex ratio evolution. For example, observed second sex ratios may be better accounted for by a model of behavior in which the female makes a decision about sex ratio and brood size.

The results in this paper raise an important question: is the theory of optimal facultative sex ratio adjustment successful? My motivation for asking is that theory predicts the overall trend of observed second sex ratios. Yet, there is between-strain heterogeneity in the "success" of this explanation (*cf.* Tables 1 and 2, Figures 5 and 6, ORZACK 1986). Moreover, given the theory, the strains have different observed fitnesses which, of course, imply that the population is not at selective equilibrium. Since the theory is reasonably construed as intended to describe the equilibrium *statistical* behavior of individuals, it is a failure relative to this conception. (Note that one could be more stringent in assessing the theory by requiring that it account for *individual* behavioral acts). Nevertheless, the theory is a success in a more general sense (leaving aside questions about alternative explanations). To paraphrase WAAGE (1986, p. 85), the theory has ordered the data. There is a significant *overall* relationship between predictions and observations. This implies that the theory captures something about the nature of selection on this trait. One can then assess what the theory does not explain (*i.e.*, imprecise second sex ratios, the heterogeneity of fit, and the heterogeneity of fitnesses). In this way, the model may lead to further analyses. Unfortunately, this view of the role in theory is very problematic. Recall that optimality theory is not a theory in which there could be first and second order analyses of individual behavior but a theory with the assumption that *a* stable behavioral strategy will evolve. Heterogeneity of fit to such a prediction among the genotypes within a population contradicts this theory in several ways. (Clearly it is troublesome to speak of an equilibrium population of phenotypes having perfect heritability and differing fit to an optimum). In my opinion, a decision to ignore such heterogeneity and to use optimality theory as a tool for ordering data amounts to a rejection of causation. Such a decision is not necessarily bad and may be quite successful (as in quantum mechanics, for example) but it is important to note that one's perception of the meaning of optimality theory must change accordingly. The theory ceases to be a description of the outcome of optimizing natural selection and becomes simply a measuring device.

I hope it is clear from the analyses in this paper how comparative testing is essential in this assessment of the success of sex ratio theory (see also HERRE 1987). SOBER (1988) makes this point more generally (p. 106):

> When we imagine ourselves explaining a characteristic found in a single population, it can be quite easy to dream up a new adaptationist story, if the old one is disconfirmed. However, this is less easy to do when one works with comparative data. Rather than explaining a single trait in a single population, let the unit of inquiry be a pattern of character diversity in a group whose common ancestry (monophyly) is antecedently known. This does not guarantee that the one true adaptationist explanation must shine through. However, this kind of data clearly counts as *better* data; it provides a check . . . against idle speculation.

I believe this approach is essential given the subjectivity which can enter into simple judgement of fit and lead to a lack of rigor in the analysis of a single data set (*cf.* Figure 1).

An additional problem in reconciling theory and data arises because dynamical analysis of sex ratio evolution is as yet incomplete. For example, there is at present no genetic model of the evolution of facultative sex ratio adjustment. Hence, we do not even know whether there are alternative stable equilibria under selection. Consequently, there is ambiguity about the interpretation of discrepancies between observations and predictions. Do the discrepancies in Figure 3, for example, indicate that the optimal behavior cannot evolve or that the assumptions of the model (*e.g.*, a constant environment) are incorrect? Some ambiguity about the nature of such discrepancies is always going to be present in evolutionary analyses. Nonetheless, there is a very real sense in which much ambiguity in the analysis of sex ratio evolution will be resolved with a much fuller theoretical understanding. In an important way this relieves simple theories from the burden of data. I argue that it is incorrect a priori to make *strong* claims about how data fit the predictions of the theory of optimal facultative sex ratio adjustment, for example. Indeed, I hope the reader is convinced by the present analysis that discrepancies between theory and data are of substantial biological interest. This is not an appeal to particularism. Instead, my motivation is the belief that a clear assessment of what we do and do not know is essential for greater understanding of the natural world.

I thank J. WAAGE and J. WERREN for supplying data, P. PAMILO for a valuable suggestion, and M. ANTOLIN, D. PARKER, S. SKINNER, M. TURELLI, and M. WADE for comments. I am especially grateful to R. THISTED for encouragement and help in statistical analysis and for teaching me how to GLIM. This project was partially supported by U.S. Public Health Service National Research Service Award GM11424.

LITERATURE CITED

CHARNOV, E. L., 1982 *The theory of sex allocation.* Princeton University Press, Princeton, N.J.

CHARNOV, E. L., R. L. LOS-DEN HARTOGH, W. T. JONES and J. VAN DEN ASSEM, 1981 Sex ratio evolution in a variable environment. Nature **289:** 27–33.

COCHRAN, W. G., 1968 Errors of measurement in statistics. Technometrics **10:** 637–666.

CONOVER, W. J., 1980 *Practical Nonparametric Statistics.* John Wiley & Sons, New York.

FIELLER, E. C., H. O. HARTLEY and E. S. PEARSON, 1957 Tests for rank correlations. I. Biometrika **44:** 470–481.

FIELLER, E. C., and E. S. PEARSON, 1961 Tests for rank correlations. II. Biometrika **48:** 29–40.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford, U.K.

HERRE, E. A., 1987 Optimality, plasticity and selective regime in fig wasp sex ratios. Nature **32:** 627–629.

HOLMES, H. B., 1972 Genetic evidence for fewer progeny and a higher percent males when *Nasonia vitripennis* oviposits in previously parasitized hosts. Entomophaga **17:** 79–88.

MCCULLAGH, P., and J. A. NELDER, 1983 *Generalized Linear Models.* Chapman & Hall, London.

NELDER, J. A., and R. W. M. WEDDERBURN, 1972 Generalized linear models. J. R. Statist. Soc. A **135:** 370–384.

ORZACK, S. H., 1986 Sex-ratio control in a parasitic wasp, *Nasonia vitripennis.* II. Experimental analysis of an optimal sex-ratio model. Evolution **40:** 341–356.

ORZACK, S. H., and E. D. PARKER, JR., 1986 Sex-ratio control in a parasitic wasp, *Nasonia vitripennis.* I. Genetic variation in facultative sex-ratio adjustment. Evolution **40:** 331–340.

ORZACK, S. H., and E. D. PARKER, JR., 1990 Genetic variation for sex ratio traits within a natural population of a parasitic wasp, *Nasonia vitripennis.* Genetics **124:** 373–384.

SOBER, E., 1988 What is adaptationism? pp. 105–118 in *The Latest on the Best*, edited by J. DUPRE. MIT Press, Cambridge, Mass.

SUZUKI, Y., and Y. IWASA, 1980 A sex ratio theory of gregarious parasitoids. Res. Popul. Ecol. **22:** 366–382.

THORNHILL, R., and J. ALCOCK, 1983 *The Evolution of Insect Mating Systems.* Harvard University Press, Cambridge, Mass.

WAAGE, J. K., 1986 Family planning in parasitoids: adaptive patterns of progeny and sex allocation, pp. 63–95 in *Insect Parasitoids*, edited by J. K. WAAGE and D. GREATHEAD. Academic Press, New York.

WAAGE, J. K., and J. A. LANE, 1984 The reproductive strategy of a parasitic wasp. II. Sex allocation and local mate competition in *Trichogramma evanescens.* J. Anim. Ecol. **53:** 417–426.

WEDDERBURN, R. W. M., 1974 Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika **61:** 439–447.

WERREN, J. H., 1980 Sex ratio adaptations to local mate competition in a parasitic wasp. Science **208:** 1157–1159.

WERREN, J. H., 1987 Labile sex ratios in wasps and bees. BioScience **37:** 498–506.

WYLIE, H. G., 1973 Control of egg fertilization by *Nasonia vitripennis* (Hymenoptera: Pteromalidae) when laying on parasitized fly pupae. Can. Entomol. **105:** 709–718.

Communicating editor: M. TURELLI