

Stabilising selection in Weldon's snails: a reappraisal

John M. C. Hutchinson

Department of Biology, University of York,
York, YO1 5DD, U.K.

This paper reconsiders Weldon's (1901, 1904) and Di Cesnola's (1907) oft-quoted research on stabilising selection on shell shape of terrestrial snails. Although I recognise problems in the experimental design, the commonest criticism is probably inapplicable. The statistical analyses were flawed, but a crude reanalysis supports Di Cesnola's conclusions, whilst Weldon's results are also suggestive. It is argued that the technique is feasible but a positive result difficult to interpret.

INTRODUCTION

This is a critique of research published 80 years ago in the first few volumes of *Biometrika* by W. F. R. Weldon (1901, 1904), and his student A. P. di Cesnola (1907). Most evolutionary research of that period would not be worth detailed criticism because modern biologists ignore or distrust it anyway; but Weldon (1901) has continued to be quoted as an example of stabilising selection. Amongst the more influential books that use it are Huxley (1942), Maynard-Smith (1958), Sheppard (1958) and Kimura (1983). (As befalls the famous, it has also been misquoted: the snails involved were terrestrial, not collected from Plymouth Sound (Mather, 1973, p. 90); and not 6 inches long (Berry, 1977, p. 136), only 15 mm!). My purpose is to question whether the work should again be mentioned so uncritically.

Classic papers such as this enhance their reputation by the very fact that they are repeatedly quoted, but I think that the sheer beauty of Weldon's technique is also responsible. His exploitation of the snail shell as a record of its own immature shape has deservedly been described as "ingenious" (Haldane, 1959) and "brilliant" (Pearson, 1906). And this is another reason to examine his methods in detail; to consider whether the technique is worth a more careful repetition.

THE ORIGINAL CLAIM

Weldon and di Cesnola worked on terrestrial snails, although the method is applicable to other organisms exhibiting accretionary growth. In all three papers, each of which concerned a different species, the experimental design was very similar. Shells of different ages, 100 adult and 100 immature, from the same locality, were sectioned down the axis about which the shell coils (fig. 1). The earliest few whorls could then be measured in both adult and immature, these measurements indicating the shape of the entire shell when the animals were younger. Weldon and di Cesnola claimed that the *mean* shape of these early whorls was the same whether they came from adult or immature shells, but that early whorls of adult shells *varied* less than the corresponding whorls of immatures. This implies that the more unusually shaped immatures had been weeded out by stabilising selection. Weldon repeated his experiment in 1904 with another snail species but this time was unable to claim any consistent difference in variation.

CRITICISMS OF THE EXPERIMENTAL DESIGN

The literature already contains three arguments critical of Weldon's and di Cesnola's work. First,

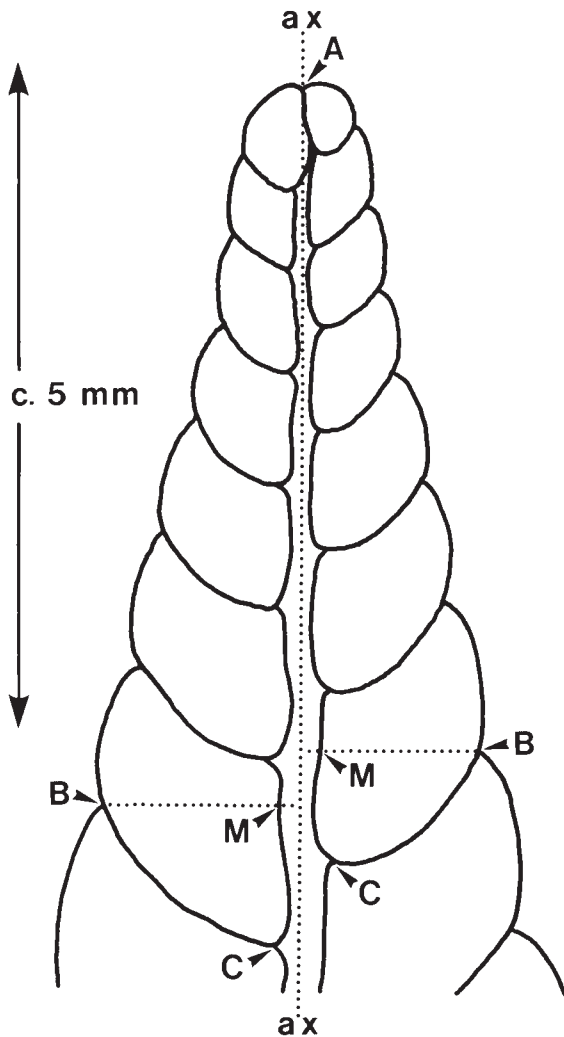


Figure 1 Apical portion of *Cochlodina laminata* sectioned down the coiling axis (redrawn from Weldon, 1901). ax = coiling axis; AB = peripheral radius; AC = columellar radius; BC = base line; BM = perpendicular.

Robson and Richards (1936, p. 211) objected that variation might decline simply because the shape of immatures is less canalised than that of adults. But as Huxley (1942, p. 449) realised, this completely misses the point; it is whorls of the same maturity that are compared and their shapes cannot be altered after deposition.

Spatial heterogeneity

A much more frequently quoted argument was suggested by Clarke *et al.* (1978). They observed

that it is often harder to find immature land snails than adults; so, when trying to collect a specific number of each age group, one tends to collect immatures from a wider area. They then quoted evidence on the size of the panmictic unit of *Cepaea* to claim that the immatures may consequently be drawn from a genetically more heterogeneous population. (The larger area is in any case likely to be environmentally more heterogeneous, which might also increase phenotypic variance.)

Di Cesnola did admit the difficulty of finding specimens and it is certainly unfortunate that his collection along 2 km of the Thames must have spanned many populations. But Clarke *et al.*'s criticism is negated if every individual seen is collected, regardless of size. Di Cesnola collected 600 specimens, but measured only 100 of each age class, so it seems unlikely that he ignored the adults after enough had been collected. In fact, when scoring the size of the umbilicus he considered only 130 adults, so perhaps it was they that were in the minority. Weldon gives us fewer clues about his collecting technique. But we know that he was originally interested in comparing morphology between sites and that some of his samples were huge (Pearson, 1906), so I consider it likely that he too would not have become satiated with adults.

Temporal heterogeneity

This problem is briefly mentioned by Parkin (1979). Although the immatures and adults might not have differed in their sites of collection, they necessarily differ in age. *Cochlodina laminata* (Montagu), the first species that Weldon examined, is believed to lay eggs in late summer and to take two years to grow to maturity, (Steenberg, 1911; Cameron, 1982). Judging by their sizes, the immatures were a year old and the adults one or more years older. The problem is that, for instance, a longer breeding season one year could quite conceivably lead to one generation being more variable than those preceding. It is not impossible that this would nevertheless leave the mean unaffected.

This is a fundamental weakness of the method. Because adults may survive several years it would not even solve the problem completely to sample the immature population one year and the adults the next. But, on average, just as often as weather makes one generation less variable than the next, a generation will be made more variable. Thus repeating the experiment over several years would reduce the problem. (The samples for Weldon's 1901 paper did happen to have been collected over

two years; not so with the other two collections.) However, repeated sampling really negates the advantages of the technique. If one is to go to the effort of sampling repeatedly, and put up with the concomitant disturbance of the study site, one might as well avoid the cross-sectional approach altogether and use a mark-recapture technique to monitor selection on a single cohort.

Other clausiliids, such as *Clausilia itala* (von Martens) of Weldon's second paper, are thought to have similar life cycles to *C. laminata* (Cameron, 1982). But with Cesnola's *Arianta arbustorum* (L.) the problem may be worse because it has a long breeding season (Taylor, 1914, p. 421), and within some populations the time taken to reach maturity may vary from 1 to 3 years (Baur, 1984). So the same whorls in adults and immatures will not only have formed in different years, but might also tend to have formed in different seasons (fewer of the fast-growing individuals will be represented in the immature age class).

Inexact sectioning

The next criticism reflects my own experience of grinding gastropod shells so that the coiling axis lies in the plane of section. Despite the narrow umbilici of the species that they used, it is still a matter of judgement when to stop grinding; but the longer the umbilicus, the easier the alignment (Weldon, 1904). So the early whorls of adult snails should be more precisely aligned than those of the immatures. It is hard, without having seen the specimens, to judge the importance of poor alignment in making the measurements of immatures more variable. Counterbalancing it to some extent is the fact that the coiling axis may deviate from a straight line. Consequently, especially in adults, the section may require a compromise and does not correspond as precisely as it might to the coiling axis of the earliest whorls.

Selective breakage

Another potential worry is di Cesnola's admission that twice as many shells were broken as could be sectioned successfully. Might not selection be operating at that stage rather than during their life in the wild? The immatures would be easier to break; so if the aberrantly-shaped shells were stronger, their proportion in the immature sample might be enriched more than with the adults. Parasites can distort and thicken shells (Sturrock and Sturrock, 1971; Cheng, 1971), but this criticism remains somewhat hypothetical.

CRITICISM OF STATISTICAL METHODS

Much more damning criticisms concern their statistical analyses. In mitigation, we should remember that the *F* test for comparing variances was not developed until the 1920s. The *t* test also just post-dates this work, but the sample sizes are large enough for their use of the normal deviate to be valid (although modern readers may be disconcerted by the use of "probable error", a multiple of the standard error, to compare means, and by the failure to convert the corresponding multiple of the normal deviate into a probability). The large and equal sample sizes also mean that it makes little practical difference if I use a *t* test when its assumption of the homogeneity of variances is later shown to be violated.

Weldon (1901) measured the distance between the apex and the outer suture (the "peripheral radius", fig. 1) every half revolution. Because snail shells approximate to a logarithmic spiral, the plot of peripheral radius against the number of revolutions (whorl number) is very non-linear. Rather than transforming the measures to make it more linear, Weldon split the graph into sections spanning half a revolution each and considered the relationship to be linear over each such interval. He calculated the mean whorl number and mean peripheral radius over each interval; and, so that the latter could be directly compared between young and old, the values for both were converted to a standard whorl number using linear interpolation. Assuming the peripheral radius and whorl number are linearly related, when they are not, does affect these values; but I have checked, using a simulation technique, that the possible bias to adult or immature is small enough to ignore.

These standardised mean peripheral radii inevitably differ between immatures and adults, but Weldon judged the differences to be insignificant compared with the standard deviations of the peripheral radii over each half whorl. But this standard deviation includes the size variation inevitable in shells differing by up to half a revolution in whorl number; as the two means have the same whorl number, we should use the variation at a constant whorl number. This can be obtained from the error mean square of the regression of peripheral radius on whorl number over each interval. Fortunately this is a quantity that Weldon uses later. In fact when I calculate *t* values using this smaller standard deviation, the differences are still non-significant, except for the largest whorl (table 1).

Table 1 Data from *Cochlodina laminata* (Weldon, 1901). Mean peripheral radii (m) and their standard deviations (s), from the error mean squares of regressions on whorl number, for successive half whorls. For this and the succeeding tables: the values of P are for tests on individual whorls *prior* to application of the Bonferroni inequality; "ns" implies that P is above 10 per cent; an "*" indicates 5 per cent significance *after* Bonferroni application. Dimensions are in mm. "Whorl number" is the number of right angles from a whorl of standard size. " s_{trunc} " is the standard deviation of a distribution, originally normal with the immatures' mean and variance, then truncated to match the mean with the adults'

Whorl number	m_{adult}	m_{imm}	t	P	s_{adult}	s_{imm}	$F = \frac{s_{\text{imm}}^2}{s_{\text{adult}}^2}$	P	s_{trunc}	$F = \frac{s_{\text{trunc}}^2}{s_{\text{adult}}^2}$	P
-9	2.3463	2.3630	-1.43	ns	0.07956	0.08597	1.168	ns	0.07258	0.832	ns
-7	2.7336	2.7521	-1.61	ns	0.07901	0.08388	1.127	ns	0.06955	0.774	ns
-5	3.1529	3.1666	-1.30	ns	0.06991	0.07833	1.255	ns	0.06704	0.919	ns
-3	3.6265	3.6297	-0.30	ns	0.07143	0.08062	1.274	ns	0.07703	1.162	ns
-1	4.1397	4.1254	+1.39	ns	0.05784	0.08096	1.959	0.06%*	0.06921	1.431	5%
+1	4.7181	4.6952	+2.38	2%	0.05732	0.06087	1.127	ns	0.04575	0.637	ns

Possibly directional selection operated on only the largest whorl. But we could not have made any prior predictions about which whorls would exhibit the effects of directional—or stabilizing—selection; so I had to test all six whorls. This has increased the opportunity for chance alone to make one individual t test produce a "significant" result. The measurements of each whorl will be far from independent, but, as their intercorrelations are not available, we must assume independence as a worst case. If we perform N independent tests, only if an individual t test reaches a $(5/N)$ per cent significance level can we reject at the 5 per cent level the hypothesis that no whorls show a difference (the Bonferroni inequality; see Cooper, 1968). In this case the t test for the largest whorl does not produce this level of significance (0.83 per cent), so there is no evidence for directional selection.

Satisfied that the data suggested no directional selection, Weldon proceeded to detect stabilising selection by comparing adult and immature variation. As a measure of variation, Weldon calculated the error sum of squares, which describes the scatter about the regression line relating peripheral radius to whorl number. For all whorls this scatter is greater in immature than adult snails. However the only argument Weldon advanced that this difference was statistically significant was that in six out of six whorls it was the same age class (the immatures) that had the greater variance. This argument is invalid because the peripheral radii of successive whorls of a shell are obviously not independent. Instead I have calculated the F ratios of the two error mean squares for each whorl (table 1). Only one out of six is, when considered separately, significant at the 5 per cent level; but this F ratio is so very significant that, using the Bonferroni inequality, we must reject at the 5 per cent level the overall null hypothesis that no whorls exhibit a decrease in variation from immature to adult.

There is a worry about my use of the F test: it is extremely sensitive to deviations from normality, and alternative tests for the equality of variances are always preferable (Van Valen, 1978). But to deduce anything from the information available, I have had to assume normality. Weldon's scattergrams do not indicate marked non-normality.

The sceptic might raise one further statistical objection to this as evidence of stabilising selection. We failed to demonstrate directional selection, but it is still conceivable that some directional selection really is responsible for the non-significant difference in means, and therefore might also have affected the variance. As Schluter (1988) points out, to calculate the reduction in variance caused by directional selection requires detailed information, including the mean curvature of the fitness function. We lack this data, but can calculate an upper limit on the reduction. I again assume that the characters in the immatures are normally distributed and consider the worst case, of truncation selection displacing the mean to the extent observed. Truncation selection removes the extreme values and thus reduces the variance to the largest possible extent. I calculate that this form of directional selection could often have caused the observed decrease in variance from immature to adult; and, where the adjusted immature variance remains greater than the observed adult variance, it is not sufficiently so for their F ratio (minimum variance following directional selection of immatures/observed adult variance) to be significant (table 1).

The situation is thus rather unsatisfactory. We have no evidence for directional selection; random sampling could be responsible for the non-significant difference in means, in which case we would have to invoke stabilising selection to explain the significant decrease in variance. But we cannot dismiss the possibility of some direc-

Table 2 Data from *Arianta arbustorum* (di Cesnola, 1907). Means and standard deviations (from the error mean squares of their regressions on whorl number) of three characters measured on successive half whorls. Units and abbreviations as in Table 1

Whorl number	m_{adult}	m_{imm}	t	P	s_{adult}	s_{imm}	s_{trunc}	$F = \frac{s_{trunc}^2}{s_{adult}^2}$	P
Columellar radius									
-5	1.5481	1.5648	-1.89	10%	0.06400	0.06070	0.04855	0.711	ns
-3	1.9224	1.9179	+0.43	ns	0.07031	0.07686	0.07214	1.052	ns
-1	2.5130	2.5389	-1.33	ns	0.12210	0.15158	0.13011	1.135	ns
+1	3.5801	3.5126	+4.26	0.01%*	0.06696	0.14380	0.10247	2.341	0.01%*
Peripheral radius									
-5	1.0546	1.0642	-0.62	ns	0.11222	0.10775	0.09850	0.906	ns
-3	1.5535	1.5456	+0.39	ns	0.13887	0.14814	0.13972	1.012	ns
-1	2.3035	2.2948	+0.31	ns	0.17830	0.21646	0.20671	1.344	10%
+1	3.2436	3.2143	+1.05	ns	0.15935	0.22960	0.20353	1.631	0.76%
Base line									
-5	1.4533	1.4744	-1.99	5%	0.08048	0.06900	0.05413	0.837	ns
-3	1.7900	1.7844	+0.35	ns	0.09020	0.13413	0.12789	2.010	0.03%*
-1	2.3219	2.3344	-0.73	ns	0.08834	0.14540	0.13326	2.275	0.01%*
+1	3.2537	3.1895	+2.14	5%	0.16926	0.24834	0.20078	1.407	5%

tional selection, which, if of a sufficiently extreme form, could also account for the variance decreasing.

Di Cesnola scored two further characters but analysed the data similarly to Weldon. He also did not properly compare the means and so did not realise that the largest whorl considered showed very significant differences between adult and immature (table 2), significant enough for us to reject the overall hypothesis that no means differ. So it is even more important with these data to calculate the extent to which directional selection could decrease the variance. Again I have calculated the F ratios, comparing the minimum variance to which directional selection could reduce the immature variance, with the observed adult variance about the regression line. In three instances these are highly significant (table 2); sufficiently so as to reject (even at the 0.5 per cent level) the null hypothesis that no whorls exhibit an adult decrease in variation over and above that which directional selection could have caused. Di Cesnola was right for the wrong reasons; his data do imply stabilising selection.

The caveat is that a normal distribution prior to selection was an assumption. But I expect any non-normality to be towards a longer tail for the larger sizes (because the shell dimensions were not log-transformed to remove the likely heteroscedasticity—see below). Thus when I assume a normal distribution and it is the smallest values that are eliminated by truncation selection, as in two of the three significant instances, reduction of variance is overestimated. So this procedure happened to be conservative. Also the high level of significance

makes less critical the extent of disruption of the F test by non-normality.

For the sake of completeness I have also calculated the corresponding t values and F ratios from the data in Weldon's 1904 paper (table 3). As Weldon pointed out, successive whorls differ irregularly in whether it is the adult or immature shells whose mean or variance is greater, suggesting that no selection is operating (although it is disconcerting how many of the individual differences are significant).

Weldon's 1901 paper just fails conclusively to demonstrate stabilising selection. But, assuming that it had succeeded, Haldane (1954) calculated that the average decrease in variation indicated a selection intensity as high as 0.12. The whorl where the variance was most reduced indicates a 0.34 selection intensity. It may thus seem surprising that this did not lead to statistical significance in a sample as large as 200 individuals with many measurements being made on each.

The problem is that no data appears to survive other than the published summary statistics. If I knew the distribution of the characters in adults and immatures, I would not have to assume that any directional selection had such an extreme effect on the variance as truncation selection. In fact I should then perhaps forget about comparing variances and instead attempt a full description of how fitness changes with the size of the character (Schluter, 1988). Detailed knowledge of the shape of this fitness function supersedes crude descriptions such as "stabilising", which covers fitness functions of all sorts of shapes having a maximum within the range of phenotypes. With cross-

Table 3 Data from *Clausilia itala* (Weldon, 1904). Means and their standard deviations for two characters measured on successive whorls. The data are grouped into intervals by the size of their columellar radii, and the standard deviations calculated from the error mean square of a multiple regression over each interval, using columellar radius and the third character as two predictor variables. Units and abbreviations as in Table 1

Columellar radius	m_{adult}	m_{imm}	t	P	s_{adult}	s_{imm}	$F = \frac{s_{imm}^2}{s_{adult}^2}$	P
Peripheral radius								
2·2	1·9428	1·9783	-5·36	0·01%*	0·0452	0·0496	1·204	ns
2·8	2·4770	2·4383	+1·28	ns	0·0519	0·0449	0·748	ns
3·5	3·0427	3·0020	+4·55	0·01%*	0·0611	0·0669	1·198	ns
4·1	3·6747	3·6527	+3·05	1%	0·0546	0·0484	0·785	ns
5·0	4·4703	4·4711	-0·07	ns	0·0701	0·0863	1·515	2%
6·1	5·4216	5·4268	-0·40	ns	0·0912	0·0949	1·082	ns
7·1	6·4632	6·4951	-2·29	5%	0·1106	0·0863	0·608	ns
Perpendicular								
2·2	0·5832	0·5919	-1·88	10%	0·0334	0·0327	0·958	ns
2·8	0·6886	0·6927	-0·81	ns	0·0367	0·0359	0·956	ns
3·5	0·8191	0·8134	+0·91	ns	0·0467	0·0424	0·824	ns
4·1	0·9707	0·9569	+1·79	10%	0·0556	0·0550	0·978	ns
5·0	1·1512	1·1358	+1·74	10%	0·0647	0·0621	0·921	ns
6·1	1·3272	1·3555	-2·69	1%	0·0756	0·0750	0·984	ns
7·1	1·5326	1·5432	-0·94	ns	0·0747	0·0858	1·319	5%

sectional data, Schluter suggests putting $W = 0$ for, in this case, the immature snails, and $W = 1$ for the adults. Using cubic splines and iterative regression techniques, he fits a curve to the scattergram of W plotted against the character. This curve should be monotonically related to the fitness function. By making no assumptions about the distribution of the character or shape of the fitness function, Schluter's procedure becomes more hypothesis generating than hypothesis testing, although he suggests bootstrap techniques to judge whether a maximum is a reliable feature of the fitted fitness function.

Schluter reanalyses various standard sets of data, the most interesting to us being the classic example of stabilising selection on human birth weight (Karn and Penrose, 1951). The fitness function does decrease at both extremes, so Karn and Penrose were justified in claiming stabilising selection. But by instead fitting a purely directional fitness function, Schluter shows that stabilising selection is responsible for very little of the variability in fitness. If this data had instead been analysed by comparison of population variances before and after selection, and we had had to assume the worst case of truncation selection as a null hypothesis, the stabilising component might well not have shown as significant.

If Weldon's original data sets were available to transform, I also could probably avoid separate comparisons of each half whorl by comparing the position of, and scatter about, regression lines describing several whorls at a time. Snail shells

roughly follow a logarithmic spiral which remains similar throughout growth. Therefore regression lines should be fitted to log-transformed shell dimensions in order for whorl number to be related linearly, in order to space the observations evenly and to avoid heteroscedasticity. This transformation is also likely to linearise the relationships if growth is in fact allometric.

Alternatively the same measurements from separate whorls could be viewed as different but correlated characters; the different measurements made on each whorl certainly should be viewed in this way. A multivariate analysis of these characters is desirable, if only because of the greater reliability of a statistic obtained by amalgamating many more observations, as well as the need to perform fewer tests concurrently (thus reducing the loss of power associated with the Bonferroni inequality). One technique, a multivariate version of the t and F tests, would be to compare the centres of the adult and immature distributions using Hotelling's T^2 , and compare their dispersions using a likelihood-ratio test of determinants (Kendall and Stuart, 1976, chapter 42). Endler (1986) reviews alternative multivariate techniques.

A prominent technique is Lande and Arnold's (1983), which uses multiple regression to disentangle directional and stabilising components of selection operating on correlated characters, also identifying which are the characters, or combinations of characters, directly selected. Unfortunately, as Schluter (1988) points out, their stabilising selection coefficient in fact distinguishes

whether the fitness function is curved, not whether it has a maximum; also their quadratic fitness function may contain a maximum within the range of phenotypes not because fitness declines either side, but in order to better fit the curvature elsewhere. Another problem is the technique's inaccuracy when traits are highly correlated, as raw shell dimensions certainly will be. Therefore welcome as a potential alternative to Lande and Arnold, is Schluter's promise (1988) to develop a multivariate generalisation of his non-parametric approach.

DISCUSSION

Even when Weldon's and di Cesnola's work is not formally convincing, it does strongly suggest stabilising selection. Now that statistical techniques have advanced, should we try to confirm the reduction in variation by gathering a new set of data, this time being sure to avoid sampling and other sources of bias?

From the practical aspect, the labour of sectioning shells would remain as onerous as for Weldon (Pearson, 1906). X-raying shells might prove a short cut (*e.g.*, Foote and Cowie, 1988), so long as the pictures so obtained have sharp enough boundaries for accurate measurement. But, with high-spired shells, several dimensions, for example the maximum width and the distance from the widest point to the preceding suture, can be measured on all whorls from the silhouette, without any need to section (*fig. 1*). This means that data could be gathered rapidly and in quantity, for instance permitting comparisons between sites. Numerous replications would also partially overcome a problem considered earlier, that the corresponding whorls in immatures and adults have formed in different years.

But would proving Weldon right be of any interest to modern biologists? It surprises no Darwinian to know that stabilising selection exists and there are already enough examples to satisfy the needs of genetics textbooks (*e.g.*, Bumpus, 1899; Karn and Penrose, 1951; Perrins, 1964). More surprising to me are the indications of measureable selection coefficients acting on small differences in shell shape. Ought we not to be curious to understand what is happening in this particular system? For one thing, many other workers have been attracted by intraspecific differences in shell shape, but have mostly proved unable to advance from description to explanation (Goodfriend, 1986). Not only would it be exciting to know what selec-

tive agent is maintaining shell shape, but also why the snails are not more precisely adapted. Perhaps they cannot get their development fully under control; or is each individual making the best of slightly different environmental conditions?

The temptation is to imagine a predator finding the aberrantly shaped shells easier to crack. But, to his great credit, Weldon (1901) realised that selection might instead be acting directly only on characters correlated with shell shape. For instance, both rate of growth (Kemp and Bertness, 1984) and the presence of parasites (Sturrock and Sturrock, 1971) are known to affect shell shape and will themselves be correlated with other traits directly affecting survival. Lande and Arnold's (1983) technique should be able to disentangle which are the characters directly selected, but the problem is that, without a mechanism of selection in mind, we cannot be sure of including these characters in the analysis. Indeed, when we rely on sectioned shells for our data, the characters directly selected may no longer be measurable.

It would matter if selection on shell shape were merely indirect because the influence of the correlated character need not be reciprocal; we could readily conceive of shell shape changing without affecting rate of growth or parasite load. The significance of this is that, if the stabilising selection on shape is only indirect, shell shape can evolve without that selection being any constraint. Genes must exist that affect shell shape but not the character directly stabilised. They could respond to any directional direct selection on shell shape, which lies equally unprotected against drift.

Furthermore, apparent stabilising selection need not even reflect any direct stabilising selection acting on another character. Boycott and Diver (1930) bred genetic strains that produced a high proportion of abnormally shaped shells, but each strain produced both unusually flat and unusually tall snails. A disease might similarly merely reduce the canalisation of shape and then direct directional selection against diseased individuals would appear as stabilising selection on shape.

Demonstrating apparent stabilising selection on shell shape by Weldon's method would therefore mean very little. We would learn the phenotypic effect on this character of the current selective regime, but not what really interests us—on what the selection acts and whether it might be a cause of stasis. The same is true of studies of stabilising selection on size (*e.g.* Bumpus, 1899; Karn and Penrose, 1951), which are particularly susceptible to these arguments, since size is a product of many traits themselves of high adaptive

significance. In fact, the possibility of stabilising selection being indirect is always hard to dismiss, if only because of linkage. Sometimes a feasible solution is to manipulate characters artificially, for instance by transferring fledglings to increase and decrease clutch size (Perrins and Moss, 1975). This has the further advantage of providing more data on the extreme phenotypes, which are those most affected by stabilising selection, but usually rare in natural populations (Schluter, 1988).

Acknowledgements Many thanks to Professor J. D. Currey and Dr G. S. Oxford for commenting on drafts of this paper, and to a referee for his relevant suggestions. I was in receipt of a S.E.R.C. studentship.

REFERENCES

- BAUR, B. 1984. Early maturity and breeding in *Arianta arbustorum* (L.) (Pulmonata: Helicidae). *J. Moll. Stud.*, 50, 241-242.
- BERRY, R. J. 1977. *Inheritance and Natural History*. Collins, London.
- BOYCOTT, A. E. AND DIVER, C. 1930. Abnormal forms of *Limnaea peregra* obtained in artificial breeding and their inheritance. *Proc. Malac. Soc. Lond.*, 19, 141-146.
- BUMPUS, H. C. 1899. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. *Biol. Lectures, Woods Hole Marine Biol. Station*, 6, 209-226.
- CAMERON, R. A. D. 1982. Life histories, density and biomass in a woodland snail community. *J. Moll. Stud.*, 48, 159-166.
- CHENG, T. C. 1971. Enhanced growth as a manifestation of parasitism and shell deposition in parasitized mollusks. In Cheng, T. C. (ed.) *Aspects of the Biology of Symbiosis*. Univ. Park Press, Baltimore, pp. 103-137.
- CLARKE, B., ARTHUR, W., HORSLEY, D. T. AND PARKIN, D. T. 1978. Genetic variation and natural selection in pulmonate molluscs. In Fretter, V. (ed.) *Pulmonates, Vol. 2a. Systematics and Ecology*. Academic Press, London, pp. 219-270.
- COOPER, D. W. 1968. The significance level in multiple tests made simultaneously. *Heredity*, 23, 614-617.
- DI CESNOLA, A. P. 1907. A first study of natural selection in "*Helix arbustorum*" (Helicogena). *Biometrika*, 5, 387-399.
- ENDLER, J. A. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton.
- FOOTE, M. AND COWIE, R. H. 1988. Developmental buffering as a mechanism for stasis: evidence from the pulmonate *Theba pisana*. *Evolution*, 42, 396-399.
- GOODFRIEND, G. A. 1986. Variation in land-snail shell form and size and its causes: a review. *Syst. Zool.*, 35, 204-223.
- HALDANE, J. B. S. 1954. The measurement of natural selection. *Proc. 9th Int. Congr. Genet.*, 1, 480-487.
- HALDANE, J. B. S. 1959. Natural selection. In Bell, P. R. (ed.) *Darwin's Biological Work*. Cambridge University Press, Cambridge, pp. 101-149.
- HUXLEY, J. S. 1942. *Evolution: the Modern Synthesis*. Allen and Unwin, London.
- KARN, M. N. AND PENROSE, L. S. 1951. Birth weight and gestation time in relation to maternal age, parity and infant survival. *Ann. Eugen.*, 16, 147-164.
- KEMP, P. AND BERTNESS, M. D. 1984. Snail shape and growth rates: evidence for plastic shell allometry in *Littorina littorea*. *Proc. Natl Acad. Sci. USA* 81, 811-813.
- KENDALL, M. G. AND STUART, A. 1976. *The Advanced Theory of Statistics. Vol. 3.* 3rd edn. Charles Griffin, London.
- KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- LANDE, R. AND ARNOLD, S. J. 1983. The measurement of selection on correlated characters. *Evolution*, 37, 1210-1226.
- MATHER, K. 1973. *Genetical Structure of Populations*. Chapman and Hall, London.
- MAYNARD SMITH, J. 1958. *The Theory of Evolution*. Penguin Books, London.
- PARKIN, D. T. 1979. *An Introduction to Evolutionary Genetics*. Edward Arnold, London.
- PEARSON, K. 1906. Walter Frank Raphael Weldon, 1860-1906. *Biometrika*, 5, 1-52.
- PERRINS, C. M. 1964. Survival of young swifts in relation to brood size. *Nature*, 201, 1147-1148.
- PERRINS, C. M. AND MOSS, D. 1975. Reproductive rates in the Great Tit. *J. Anim. Ecol.*, 44, 695-706.
- ROBSON, G. C. AND RICHARDS, O. W. 1936. *The Variation of Animals in Nature*. Longmans, London.
- SCHLUTER, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution*, 42, 849-861.
- SHEPPARD, P. M. 1958. *Natural Selection and Heredity*. Hutchinson, London.
- STEENBERG, C. M. 1911. *Bløddyril Landsnegle*. G.E.C. Gads, Copenhagen.
- STURROCK, R. F. AND STURROCK, B. M. 1971. Shell abnormalities in *Biomphalaria glabrata* infected with *Schistosoma mansoni* and their significance in field transmission studies. *Jour. Helminthol.*, 45, 201-210.
- TAYLOR, J. W. 1914. *Monograph of the Land and Freshwater Mollusca of the British Isles*. Part 20. Taylor Brothers, Leeds.
- VAN VALEN, L. 1978. The statistics of variation. *Evol. Theory*, 4, 33-43.
- WELDON, W. F. R. 1901. A first study of natural selection in *Clausilia laminata* (Montagu). *Biometrika*, 1, 109-124.
- WELDON, W. F. R. 1904. Note on a race of *Clausilia itala* (von Martens). *Biometrika*, 3, 299-307.