# Selection, structure and the heritability of behaviour

D. G. STIRLING,* D. RÉALE* & D. A. ROFF†

*Department of Biology, McGill University, Montreal, Quebec, Canada
†Department of Biology, University of California at Riverside, Riverside, CA, USA

*Keywords:*

behaviour;
coefficients of variation;
heritability;
selection;
variance compounding.

## Abstract

Characters which are closely linked to fitness often have low heritabilities ($V_A/V_P$). Low heritabilities could be because of low additive genetic variation ($V_A$), that had been depleted by directional selection. Alternatively, low heritabilities may be caused by large residual variation ($V_R = V_P - V_A$) compounded at a disproportionately higher rate than $V_A$ across integrated characters. Both hypotheses assume that each component of quantitative variation has an independent effect on heritability. However, $V_A$ and $V_R$ may also covary, in which case differences in heritability cannot be fully explained by the independent effects of elimination-selection or compounded residual variation.

We compared the central tendency of published behavioural heritabilities (mean = 0.31, median = 0.23) with morphological and life history data collected by Mousseau & Roff (1987). Average behavioural heritability was not significantly different from average life history heritability, but both were smaller than average morphological heritability. We cross-classified behavioural traits to test whether variation in heritability was related to selection (dominance, domestic/wild) or variance compounding (integration level). There was a significant three-way interaction between indices of selection and variance compounding, related to the absence of either effect at the highest integration level. At lower integration levels, high dominance variance indicated effects of selection. It was also indicated by the low $CV_A$ of domestic species. At the same time $CV_R$ increased disproportionately faster than $CV_A$ across integration levels, demonstrating variance compounding. However, neither $CV_R$ nor $CV_A$ had a predominant effect on heritability. The partial regression coefficients of $CV_R$ and $CV_A$ on heritability were similar and a path analysis indicated that their (positive) correlation was also necessary to explain variation in heritability. These results suggest that relationships between additive genetic and residual components of quantitative genetic variation can constrain their independent direct effects on behavioural heritability.

## Introduction

Falconer (1981) argued there is a connection between the nature of a character and the magnitude of its heritability: characters with the lowest heritabilities are those most closely connected with fitness, whereas characters with the highest heritabilities are the least connected to fitness. This has been repeatedly confirmed (Mousseau & Roff, 1987; Roff & Mousseau, 1987; Hartl & Clark, 1989, p. 470) and has been widely interpreted as evidence for Fisher's fundamental theorem. The standard interpretation of Fisher's fundamental theorem is that additive genetic variation ($V_A$) of fitness-related traits should be low as alleles directly regulating fitness will be rapidly fixed by elimination selection (Merilä & Sheldon, 1999). It follows that heritability ($h^2 = V_A/V_P$) of a character

*Correspondence:* D. Gray Stirling, Department of Biology, McGill University, 1205 ave. Dr Penfield, Montreal, Quebec H3A1B1, Canada.
Tel.: 514 398 6450; Fax: 514 398 5069;
e-mail: stirling@bio1.lan.mcgill.ca

should decrease with the strength of its relationship to fitness, and empirical studies have repeatedly supported this prediction (Gustafsson, 1986; Mousseau & Roff, 1987; Houle, 1992; Kruuk *et al*., 2000; Merilä & Sheldon, 2000). However, Price (1972) and Ewens (1989) argue that Fisher's fundamental theorem did not apply to mean reproductive fitness of a population, but describes the partial change in fitness due to changes in gene frequency at a single locus. Regardless of Fisher's intent, directional selection on a phenotypic character can reduce its additive genetic variation (Falconer & Mackay, 1996; Roff, 1997). Thus, negative relationships between a character's heritability and its relationship with fitness (Gustafsson, 1986; Kruuk *et al*., 2000; Merilä & Sheldon, 2000) suggest the magnitude of heritability reflects the nature of selection on a character (elimination-selection hypothesis – Houle, 1992, 1998) and indicates the level of a character's additive genetic variation ($V_A$).

At the same time, quantitative trait structure (number of component characters and their relationships) has both a positive effect on additive genetic variance (thereby increasing $V_A/V_P$), and a negative effect on heritability estimates through increased residual variance ($V_R = V_P - V_A$) in its denominator. Both additive genetic and residual life history variation may be integrated across related component (i.e. morphological) traits (Houle, 1991; Price & Schluter, 1991) and residual variation may be compounded through ontogeny (Cabana & Kramer, 1991). Houle (1992) reported that mean standardized estimates of additive genetic trait variation ($CV_A$ – additive genetic coefficient of variation) are higher on average in life history than in morphological characters and this result has been confirmed in two wild populations (Kruuk *et al*., 2000; Merilä & Sheldon, 2000). The idea that life history traits compound effects of larger numbers of loci is also supported by the higher mutational variation of life history characters compared with morphological characters (Houle *et al*., 1996; Houle, 1998). Thus, the effect of compounded residual variation in life history characters must be disproportionately large and practically independent of (compounded) additive genetic variation to explain why their heritabilities are smaller on average than those of morphological characters (Merilä & Sheldon, 2000). Repeated observation of larger differences in residual variation (Houle, 1992; Pomiankowski & Moller, 1995; Kruuk *et al*., 2000; Merilä & Sheldon, 2000) imply that the difference between heritabilities of these character categories occurs in spite of their higher average levels of additive genetic variation, and because of their differences in $V_R$, measured by the coefficient of residual variation ($CV_R$ – Houle, 1992; Houle *et al*., 1996; Merilä & Sheldon, 1999, 2000). However, in each of the studies cited above, there are strong positive correlations between $CV_A$ and $CV_R$ which may constrain their independent effects on heritability.

Tests of elimination-selection and variance compounding hypotheses contrast expectations of the difference between morphological and life history characters. Behavioural characters are excluded from these analyses, possibly because their relationship to fitness is not certain (Houle, 1992), and in practice because the number of available estimates is typically low (Mousseau & Roff, 1987; Hoffmann, 2000). Nonetheless, Henderson (1990) summarized the presumed relationships of behavioural categories, classified by the number of other component behaviours that they integrate (integration level) and their relationship with fitness. For instance, Henderson argues that mating behaviour, parental care and foraging behaviour integrate a large number of component characters such as reflex and display behaviours. Henderson (1990) also argues that highly integrated behaviours are closely related to fitness. All else being equal, he suggests that elimination-selection should reduce additive variation most strongly in highly integrated behaviours. In other words, they should have lower $CV_A$ and heritability. Applying variance compounding hypotheses to behavioural traits suggests that residual variation should be the highest among highly integrated behaviours, or they should have higher $CV_R$ and lower heritability. At the same time, the positive covariance between residual and additive genetic variation may constrain the expected independent effects of additive genetic and residual variation on heritability of behavioural characters.

We compiled estimates of heritability published after 1988 to collect a large enough sample size to test Falconer's hypothesis that the nature of a character determines its heritability. We first tested whether (1) behavioural heritabilities are different on average than either life history or morphological heritabilities (Mousseau & Roff, 1987; Roff, 1997) and (2) that $CV_A$ and $CV_R$ estimates have significant, independent ($CV_A$ – positive, $CV_R$ – negative) effects on behavioural heritability estimates. We then (3) used behavioural ecology categories to classify heritabilities, $CV_A$ and $CV_R$ by their presumed relationship with reproductive fitness. We also (4) indirectly test whether selection regulates heritability given that stronger selection should increase dominance variation (Crnokrak & Roff, 1995; Merilä & Sheldon, 1999). Finally, we cross-classified these traits by their integration level (or differences in structure) to test whether (5) selection and structure have independent effects on average heritability. In summary, our interest is not only in determining whether quantitative estimates of genetic variation vary between different categories of behaviour, but also in testing in one study the presumed effects of both selection and structure on the genetic architecture of behaviour.

## Methods

### Data sources

We used heritability of morphology and life history estimates reported by Mousseau & Roff (1987) to test whether behavioural heritability is larger on average

than life history heritability and smaller than average morphological heritability (Mousseau & Roff, 1987; Hoffmann, 2000). We collected 428 behavioural heritability estimates from 79 reports not cited in Mousseau & Roff (1987) or Roff & Mousseau (1987), published up to the year 2000 (Table 1). Unlike Mousseau & Roff (1987) we included both broad and narrow sense estimates of heritability and we collected estimates made on both domestic and wild species. Behavioural heritability estimates were reported for vertebrates (mammals, fish, reptiles, birds) and invertebrates (insects and crustaceans). Mousseau & Roff (1987) and Roff & Mousseau (1987) previously collected 154 estimates. We first tested whether both data sets (recent and older) differed in central tendency.

We used an average heritability estimate if there were multiple estimates of heritability for one population in the same environment. Similarly, if trait heritability was estimated separately for males and females in a single study, then we averaged the two estimates. But, if the heritability of a trait was estimated for two different populations of the same species, or for the same population in different environments, then we included both estimates. Mousseau & Roff (1987) reported that differences between invertebrates and vertebrates were nonsignificant. We found similar results for behavioural characters listed in Table 1 (invertebrate: $h^2 \pm \mathrm{SE} = 0.30 \pm 0.02$ vertebrate: $0.27 \pm 0.02$; $t_{580} = 0.85$, $P =$ n.s.). In general, we do not have a priori expectations that species or higher level taxa should differ in their average heritability. But, heritability estimates may be affected by domestication or long-term maintenance under laboratory conditions.

## Analysis

Mousseau & Roff (1987) found that central tendency of behavioural and life history heritabilities did not differ when they used means and medians as the test statistic, and they suggested the low sample size of behavioural traits may have led to this result. Based on Mousseau & Roff's (1987) data we estimated that the minimum sample size required to reject the null hypothesis at $\alpha = 0.05$ should be 946 (Zar, 1996). Our combined life history and behavioural trait sample size was 931. We added the life history estimates reported by Kruuk *et al*. (2000), Merilä & Sheldon (2000), Hoffmann (2000) and Mousseau (2000) so that our final sample size was 948. Thus, our parametric hypothesis test should be powerful enough to reject the null hypothesis that life history and behavioural heritabilities are not significantly different. We used both parametric (ANOVA, Welch ANOVA, *t*-test) and a median test for testing the significance of differences in central tendency. Welch's ANOVA uses an $F$ statistic with degrees of freedom that are adjusted for unequal variances. We used the Welch ANOVA for our parametric tests of $CV_A$ and $CV_R$ categories, as category variance estimates differed widely and Welch ANOVA is not sensitive to violations of variance homogeneity (Day & Quinn, 1989). The median test compares the significance of differences between medians of each category relative to a combined or grand median. Although it is not as powerful as a parametric test, it tests the significance of differences in population medians (Zar, 1996), which in a skewed population can be a better measure of central tendency.

## Classification methods

Our data set comprised a wide variety of behavioural traits, ranging from behaviour closely associated with life history (e.g. predator avoidance, mating preferences, oviposition behaviour) to traits with domestic or economic importance, but little direct ecological relevance (e.g. trot stride).

### *Wild, domestic, semidomestic*

There were few (3) field estimates of behavioural heritability. Almost all of the measurements were collected in a laboratory or farm environment. If a species or a population had not been selected for an economic purpose, we classified their heritability estimates as 'wild'. In this respect our terminology is consistent with Mousseau & Roff's (1987) classification, although 'nondomestic' may be a better description. Included in this category were a number of heritabilities estimated for populations which may have been maintained under laboratory conditions for a large number of generations (e.g. *Drosophila melanogaster*). These may not have been subjected to artificial selection intentionally, but heritability estimates could reflect ongoing adaptation to their maintenance conditions (Harshman & Hoffmann, 2000). For example, heritability estimates for *Drosophila* (Roff & Mosseau, 1987; Hoffman, 2000) are lower in each trait category than the averages for wild (nondomestic) populations (Mousseau & Roff, 1987). We classified these populations as semidomestic and the remaining populations as domestic.

### *Estimation method*

Additive genetic variation could be estimated using broad sense and full sib breeding designs which include all or part of the dominance variance component and the between – family environmental component. These variance components are not included in half sib, maximum likelihood or regression estimates. Consequently, we expect the former heritability estimates should be higher on average than the latter and interpret the difference between them as an index of the strength of recent selection (Curtsinger *et al*., 1994; Lynch & Deng, 1994; Crnokrak & Roff, 1995). Similar reasoning suggests domestic behaviours should also have higher levels of dominance than wild behaviours (Mousseau & Roff, 1987; Crnokrak & Roff, 1995). There were eight estimation methods in our combined data set: ANOVA ($n = 69$); broad sense ($n = 49$); full sib ($n = 63$); half sib

**Table 1** Species, it's classification and the literature source for the recent behavioural heritabilities and CV's compiled in this study. Each species was classified as wild (w), domestic (d), or semidomestic (s) based on the source of the population, the description of its maintenance and the purpose of the study.

| Species | Name | Domestic/wild | Reference |
|---|---|---|---|
| Invertebrates | | | |
| *Allonemobius fasciatus + A. Socius* | Cricket | w | Roff, D.A. *et al.* 1998. *Evolution* **53**: 216–224 |
| *Anagrus delicatus* | Fairyfly parasitoid | w | Cronin, J.T. & Strong, D.R. 1996. *Heredity* **76**: 43–54 |
| *Apis melifica* | Honeybee | d | Bienefeld, K.P.F. 1991. *Ann. Entomol. Soc. Am.* **84**: 324–331 |
| *Apis melifica* | Honeybee | d | Brandes, C. 1988. *Behav. Genet.* **18**: 119–132 |
| *Apis mellifera* | Honeybee | d | Harbo, J.R. & Harris, J.W. 1999. *J. Econom. Entomol.* **92**: 261–265 |
| *Aquarius remigis* | Waterstrider | w | Blanckenhorn, W.U. & Perner, D. 1994. *Anim. Behav.* **48**: 169–176 |
| *Asobara tabida* | Drosophila parasitoid | w | Mollema, C. 1991. *Neth. J. Zool.* **41**: 174–183 |
| *Callosobruchus maculatus* | Bruchid beetle | w | Fox, C.W. 1993. *Evolution* **47**: 166–175 |
| *Callosobruchus maculatus* | Bruchid beetle | w | Messina, F.J. 1993. *Heredity* **71**: 623–629 |
| *Callosobruchus maculatus* | Bruchid beetle | w | Savalli, U.M. & Fox, C.W. 1998. *Anim. Behav.* **56**: 953–961 |
| *Ceratitis capitata* | Fruit fly | w | Whittier, T.S. & Kaneshiro, K.Y. 1995. *Evolution* **49**: 990–996 |
| *Choristoneura rosaceana* | Oblique banded leadfoller | w | Carrière, Y. & Roitberg, B.D. 1995. *Heredity* **74**: 357–368 |
| *Daphnia galeata* | Daphnia | w | Stirling, D.G. & Roff, D.A. 2000. *Anim. Behav.* **59**: 929–941 |
| *Daphnia magna* | Daphnia | w | Van Uytvanck, J. & De Meester, L. 1990. *J. Plankt. Res.* **12**: 1089–1098 |
| *Depressaria pastinacell*a | Parsnip webworm | w | Berenbaum, M.R. & Zangerl, A.R. 1992. *Evolution* **46**: 1373–1384 |
| *Drosophila busckii* | Fruit fly | s | Courtney, S.P. & Hard, J.J. 1990. *Heredity* **64**: 371–376 |
| *Drosophila melanogaste*r | Fruit fly | s | Hoffmann, A.A. 1994. In: (C. R. B. Boake, ed.), pp. 188–205. |
| *Drosophila melanogaste*r | Fruit fly | s | Hoffmann, A.A. 1999. *Heredity* **82**: 158–162 |
| *Drosophila melanogaste*r | Fruit fly | s | Kamping, A. & van Delden, W. 1990. *Behav. Genet.* **20**: 645–659 |
| *Drosophila melanogaste*r | Fruit fly | s | Mackay, T.F.C. *et al.* 1996. *Genetics* **144**: 727–735 |
| *Drosophila melanogaste*r | Fruit fly | s | Ritchie, M.G. & Kyriacou, C.P. 1996. *Anim. Behav.* **52**: 603–611 |
| *Drosophila silvestris* | Fruit fly | s | Boake, C.R.B. & Konigsberg, L. 1998. *Evolution* **52**: 1487–1492 |
| *Drosophila melanogaste*r | Fruit fly | s | Godoy-Herrera, R. *et al.* 1994. *Anim. Behav.* **48**: 251–262 |
| *Gromphadorhina portent*osa | Cockroach | w | Clark, D.C. & Moore, A.J. 1995. *Anim. Behav.* **50**: 719–729 |
| *Gryllus firmus* | Sand cricket | w | Crnokrak, P. & Roff, D.A. 1998. *Evolution* **52**: 1111–1118 |
| *Gryllus firmus* | Sand cricket | w | Webb, K.L. & Roff, D.A. 1992. *Anim. Behav.* **44**: 823–832 |
| *Gryllus integer* | Field cricket | w | Hedrick, A.V. 1994. In: (C. R. B. Boake, ed.), pp. 228–250 |
| *Harmonia axyridis* | Ladybird beetle | w | Wagner, J.D. *et al.* 1999. *Evol. Ecol. Res.* **1**: 375–388 |
| *Ips pini* | Pine engraver beetle | w | Hager, B.J. & Teale, S.A. 1996. *Heredity* **77**: 100–107 |
| *Musca domestica* | House fly | w | Aragaki, D.L.R. & Meffert, L.M. 1998. *Anim. Behav.* **55**: 1141–1150 |
| *Musca domestica* | House fly | w | Collins, R.D. *et al.* 1994. *Physiol. Entomol.* **19**: 165–172 |
| *Mythimna separata* | Moth | w | Han, E.N. & Gatehouse, A.G. 1993. *Physiol. Entomol.* **18**, 183–188 |
| *Nauphoeta cinerea* | Cockroach | w | Moore, A.J. 1994. *Behav. Ecol. Sociobiol.* **35**: 235–241 |
| *Ophraella communa* | Chrysomelidae | w | Futuyma, D.J. *et al.* 1993. *Evolution* **47**: 888–905 |
| *Phytoseiulus persimilis* | | w | Margolies, D.C. *et al.* 1997. *J. Insect Behav.* **10**: 695–709 |
| *Plutella xylostella* | Diamonback moth | w | Head, G. *et al.* 1995. *J. Econom. Entomol.* **88**: 447–453 |
| *Plutella xylostella* | Diamonback moth | w | Head, G. *et al.* 1998. *Ann. Entomol. Soc. Am.* **91**: 217–221 |
| *Rhizoglyphus robini* | Bulb mite | w | Radwan, J. 1999. *J. Evol. Biol.* **11**: 321–327 |
| *Ribautodelphax imitans* | Planthopper | w | De Winter, A.J. 1995. *Res. Pop. Ecol.* **37**: 99–104 |
| *Scatophaga stercoraria* | Yellow dung fly | w | Muhlhauser, C. *et al.* 1996. *Anim. Behav.* **51**: 1401–1407 |
| Fish | | | |
| *Gasterosteus aculeatus* | Stickleback | w | Bakker, T.C.M. 1986. *Behaviour* **98**: 1–144 |
| *Macropodus opercularis* | Paradise fish | s | Gerlai, R. & Csanyi, V. 1989. *Acta Biologica Hungarica* **40**: 67–106 |
| *Macropodus opercularis* | Paradise fish | s | Miklosi, A. *et al.* 1997. *Behav. Genet.* **27**: 191–200 |
| Reptiles | | | |
| *Thamnophis ordinoides* | Garter snake | w | Brodie, E.D. 1993. *Evolution* **47**: 844–854 |
| *Thamnophis sirtalis* | Garter snake | w | Garland T. 1994. In: (C. R. B. Boake, ed.), pp. 251–277 |
| Birds | | | |
| *Coturnix japonica* | Japanese quail | d | Nol, E., Cheng, K. & Nichols, C. 1996. *Anim. Behav.* **52**: 813–820 |
| *Coturnix japonica* | Japanese quail | d | Yang, N. *et al.* 1998. *Poultry Sci.* **77**: 1469–1477 |
| *Gallus domesticus* | Chicken | d | Campo, J.L. & Carnicer, C. 1993. *Poultry Sci.* **72**: 2193–2199 |
| *Gallus domesticus* | Chicken | d | Craig, J.V. & Mui, W.M. 1993. *Poultry Sci.* **72**: 411–420 |
| *Gallus domesticus* | Chicken | d | Craig, J.V. *et al.* 1965. In: Bakker 1986 |
| *Gallus domesticus* | Chicken | d | Gerken, M.P.J. 1992. *Poultry Sci.* **71**: 779–788 |

**Table 1** (Contd.)

| Species | Name | Domestic/wild | Reference |
|---|---|---|---|
| *Gallus domesticus* | Chicken | d | Guhl. A.M. *et al.* 1960. In: Bakker 1986 |
| *Gallus domesticus* | Chicken | d | Hagger, C. 1994. *Poultry Sci.* **73**: 381–387 |
| *Gallus domesticus* | Chicken | d | Heil, G.S. *et al.* 1990. *Poultry Sci.* **69**: 1231–1235 |
| *Sylvia atricapilla* | Blackcap | w | Berthold, P. & Pulido, F. 1994. *Proc. R. Soc. Lond. B* **257**: 311–315 |
| Mammals | | | |
| *Peromyscus maniculatus* | Vole | w | Thompson, D.B. 1990. *Evolution* **44**: 952–965 |
| *Bos taurus* | Cow | d | Fan, L.Q. *et al.* 1996. *Can. J. Anim. Sci.* **76**: 73–79 |
| *Bos taurus* | Cattle | d | Fordyce, G. *et al.* 1996. *Austr. J. Exp. Agri.* **36**: 9–17 |
| *Bos taurus* | Cattle | d | Le Neindre, P. *et al.* 1996. *Appl. Anim. Behav. Sci.* **29**: 73–81 |
| *Bos taurus* | Cattle | d | Morris, C.A. *et al.* 1994. *New Zealand J. Agri. Res.* **37**: 167–175 |
| *Canis familiaris* | Dog | d | Liinamo, A.E. *et al.* 1997. *J. Anim. Sci.* **75**: 622–629 |
| *Canis familiaris* | Dog | d | Wilsson, E. & Sundgren, P.E. 1997. *Appl. Anim. Behav. Sci.* **54**: 235–241 |
| *Canis familiaris* | Dog | d | Wilsson, E. & Sundgren, P.E. 1998. *Appl. Anim. Behav. Sci.* **58**: 151–162 |
| *Clethrionomys glareolus* | Bank vole | w | Horne, T.J. & Ylonen, H. 1998. *Evolution* **52**: 894–899 |
| *Clethrionomys glareolus* | Bank vole | w | Oksanen, T.A. *et al.* 1999. *Proc. R. Soc. Lond. B* **266**: 1495–1499 |
| *Dicrostonyx groenlandicus* | Collared lemming | w | Boonstra, R. & Hochachka, W.M. 1997. *Evol. Ecol.* **11**: 169–182 |
| *Equus domesticus* | Shetland ponies | d | Van Bergen, M.J.M. & Van Arendonk, J.A.M. 1993. *Livest. Prod. Sci.* **36**: 273–284 |
| *Mus domesticus* | House mouse | d | Dohm, M.R. *et al.* 1996. *Evolution* **50**: 1688–1701 |
| *Mus domesticus* | House mouse | d | Lagerspetz, K.J.M. 1964. In: Bakker 1986, pp. 74 |
| *Mus domesticus* | House mouse | d | Lynch, C.B. 1994. In: (C. R. B. Boake, ed.), pp. 278–301 |
| *Mus domesticus* | House mouse | d | Mogil, J.S. *et al.* 1999. *Pain* **80**: 67–82 |
| *Mus domesticus* | House mouse | d | Swallow, J.G. *et al.* 1998. *Behav. Genet.* **28**: 227–237 |
| *Mus domesticus* | House mouse | d | Walker, C. & Byers, J.A. 1991. *Anim. Behav.* **42**: 891–898 |
| *Ovis canadensis* | Bighorn sheep | w | Réale, D. *et al.* 2000. *Anim. Behav.* **60**: 589–597 |
| *Rattus norvegicus* | Norway rat | w | Hewit, J.K. & Fulker, D.W. 1984. *Behav. Genet.* **14**: 125–135 |
| *Sus scrofa* | Pig | d | Hall, A.D. *et al.* 1999. *Anim. Sci.* **68**: 43–48 |
| *Sus scrofa* | Pig | d | McGlone, J. *et al.* 1991. *Appl. Anim. Behav. Sci.* **30**: 319–322 |
| *Sus scrofa* | Pig | d | Von Felde, A. *et al.* 1996. *Livest. Prod. Sci.* **47**: 11–22 |
| *Sus scrofa* | Pig | d | Hemsworth, P.H. *et al.* 1990. *Appl. Anim. Behav. Sci.* **25**: 85–96 |

($n = 172$); maximum likelihood animal models ($n = 25$); regression techniques ($n = 69$); diallel cross ($n = 58$); and selection studies ($n = 76$). We dropped the diallel cross data from this analysis because it consisted almost entirely of data from one study.

### *Behavioural ecology trait types*

Different phenotype categories (morphology, life history, behaviour, etc.) may have different heritabilities because of their relationship with fitness (Falconer, 1981; Mousseau & Roff, 1987; Hoffman, 2000). Following this logic, different categories of behaviour thought to be more or less closely related to fitness should also have different heritabilities. We classified traits into seven behavioural ecology categories (foraging, movement, predator avoidance, reproduction, social, temperament, plus one category for other traits). In testing variation among levels of this factor, we are interested in category averages rather than the heritability of a behaviour in a particular population that may or may not have been under directional, correllated or stabilizing selection.

### *Integration level*

Henderson (1990) argued that there is a hierarchy of behavioural phenotypes that differ in their level of analysis (see his Fig. 14-1). Basic, narrowly defined phenotypes such as sensory responses, conditioned responses or reflexive behaviour having relatively few morphological or other behavioural components, were put into the low integration level category. Alternatively, sexual, maternal and foraging behaviour integrate morphological and other behavioural traits. We assigned these to the high integration level category. Traits which comprised behavioural clusters (e.g. activity level) but that we could not unambiguously assign to high integration level, we designated as 'medium' integration level. Henderson (1990) argues that selection should counteract the compounded (genetic) effects of pleiotropy and epistasis at higher integration levels. But, variance compounding hypotheses imply that residual variation should increase disproportionately faster than additive genetic variation with integration level. Thus, both mechanisms predict heritability should decrease with an increase in integration level.

### Comparisons using mean-standardized CV

Variation in heritability could be due to additive genetic or environmental variation (Houle, 1992). Houle (1992) proposed reporting coefficients of genetic [$CV_A = 100(\sqrt{V_A}/\bar{x})$] variation as an index of a trait's evolutionary potential (evolvability) and level of genetic variation, in addition to the residual variation [$CV_R = 100(\sqrt{(V_P - V_A}/\bar{x})$] which measures the non-additive genetic and environmental variation. From our literature-compiled data we collected reported CVs, or calculated $CV_A$ and $CV_R$ when heritability, phenotypic variance, or standard deviation, and the trait means traits were reported. Based on the literature, we expect that: $CV_A$ will have a positive relationship with heritability; $CV_R$ will have a negative relationship with heritability; and that $CV_A$ and $CV_R$ will covary positively. We used mean-standardized CV ($CV_A$, $CV_R$) as the dependent variable with the classification factors described above, to test the expected effects of selection and structure.

### Scale effects

Price & Schluter (1991) point out that comparing the additive variance of life history and morphologial traits may be misleading because they are measured in different phenotypic units. Within a phenotype category such as behaviour, this problem is less important. Furthermore the data ($h^2$, $CV_A$, $CV_R$) are all on a ratio scale and are independent of the unit of measurement. However, error measurements (Rohlf *et al.*, 1983), mathematical constraints (Lande, 1977) and proportional relationships between mean and variance on an arithmetic scale may each generate apparent differences in coefficients of variation (Falconer, 1981), rather than the biological processes such as selection or variance compounding (Bryant, 1986; Houle, 1992). If error measurements do not scale to character size, there will be a negative relationship between means and variances (Rohlf *et al.*, 1983). A mathematical constraint may also cause a negative correlation between component and composite coefficients of variation (Lande, 1977). Falconer (1981) points out that if coefficients of variation are similar on an arithmetic scale, then a log transformation or a scale change would eliminate scale-related differences in variance. Thus a log transformation may stabilize the residual variation and strengthen the statistical relationship (Falconer, 1981; Bryant, 1986), but it also provides a convenient test for scaling effects.

## Results

### Phenotype heritability distributions and differences in central tendency

Central tendency of behavioural heritability in our data set (0.31 ± SE = 0.013, median = 0.23) was not significantly different ($t_{581} = 0.37$, $P$ = n.s.: median test ${\chi_1}^2 = 0.35$, $P$ = n.s.) from Roff & Mousseau's (1987) and Mousseau & Roff's (1987) data ($\bar{X} = 0.30 \pm 0.023$, median = 0.25), so we combined the data sets. The combined data set was positively skewed [mean = 0.31 ± 95% confidence limits (CL) = 0.02, median = 0.23]. Eleven heritability estimates (1.9%) were larger than 1.0, six were smaller than 0 (1%) and the interquartile range was between 0.09 and 0.45 (Fig. 1).

Central tendency of the behavioural heritability distribution was similar to central tendency of the life history distribution (Fig. 1). A one-way ANOVA indicated that there were significant differences ($F_{2,1527} = 72.0$, $P < 0.0001$) between average behavioural, life history and morphological heritabilities. A *post hoc* Tukey–Kramer HSD indicated average life history (0.26 ± 95% CL = 0.024) and behavioural heritabilities differed

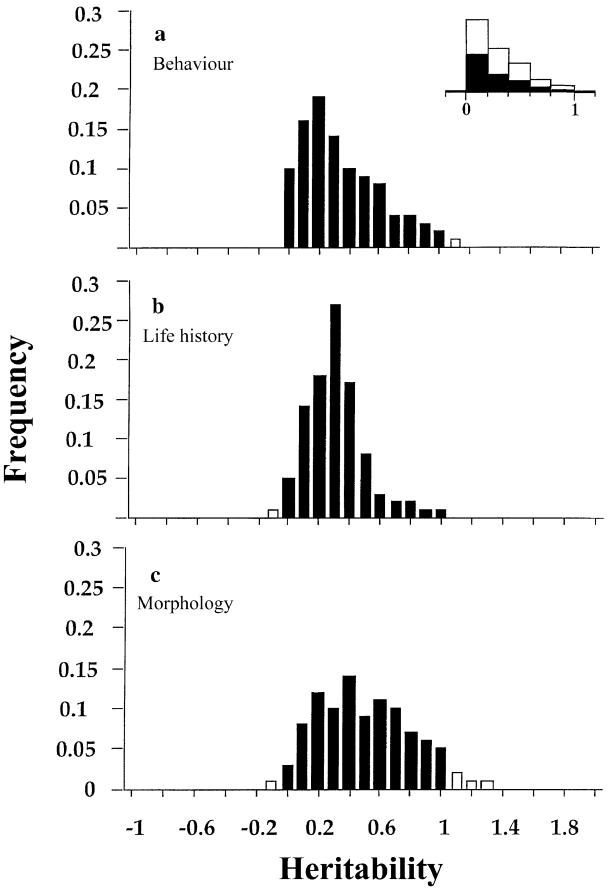


**Fig. 1** Frequency distributions for behavioural traits in the upper panel (a), life history traits in the middle panel (b), and morphological traits in the panel at the bottom (c). The inset in panel (a) indicates the proportions of domestic (dark bar) and wild (light bar) in each heritability interval. Behavioural heritability estimates that are either less than zero or greater than 1 are indicated by the light bars at the distribution extremes.

significantly from average morphological heritability. But neither the Tukey test nor a median test ($\chi_1^2 = 0.26$, $P$ = n.s.) indicated that the difference in central tendency between behavioural and life history heritabilities was significant.

### Variation reported

Mousseau & Roff (1987) noted that standard errors (SE) are frequently not reported. Houle (1992) recommended routine reporting of $CV_A$ and $CV_R$, or at least the data required to estimate them. Between 1988 and 2000, 75% of the studies (59/79) have reported SEs or CLs associated with heritability estimates. But only 227 of the 428 (53%) estimates which we collected in addition to Mousseau & Roff (1987) and Roff and Mousseau's 154 estimates had associated SEs or CLs. Between 1993 and 2000, 29.4% studies (20/68) reported CVs or both the mean and variance of the traits analysed. However, we could obtain $CV_A$ for 22% of the estimates in these studies (67/304) and $CV_R$ for only 18% (54/304).

### Relationships between $CV_R$, $CV_A$ and heritability

Average $CV_A$ (22.83 ± SE = 4.14) was smaller than the average $CV_R$ estimate (43.7 ± 7.37). The $\log_{10}$-transformed variables were positively related ($\log CV_R = 0.73 + 0.63 \log CV_A$, $r^2 = 0.49$, $P < 0.0001$). Thus, although $CV_R$ is larger on average than $CV_A$, the slope indicates $CV_R$ does not generally increase disproportionally faster than $CV_A$.

Both $CV_A$ and $CV_R$ affect heritability in the expected directions, although their coefficients were not very different in magnitude. The multiple regression with heritability as the dependent variable on log $CV_A$ and log $CV_R$ was highly significant ($F_{3,52} = 94.8$, $P < 0.0001$, $R^2 = 0.85$), although the interaction term was not significant ($t = -0.58$, d.f. = 1, 52, $P = 0.56$). We removed the interaction term and recalculated the regression, but neither the significance of the relationship nor the coefficient of determination changed. Both regression coefficients were highly significant, although $CV_A$ was positive ($t = 19.0$, d.f. = 1, 53, $P < 0.0001$, $CV_A = 0.61$) whereas $CV_R$ was negative ($t = -17.9$, d.f. = 1, 53, $P = 0.0001$, $CV_R = -0.64$).

A simple path analysis indicates that log $CV_A$ and log $CV_R$ do not have independent effects on heritability (Fig. 2). The magnitude of the direct effects of log $CV_A$ and log $CV_R$, indicated by the standardized partial regression coefficients (Fig. 2b), can be compared because they are on the same scale. First, their magnitudes are similar enough that they do not provide compelling evidence of a disproportionate effect of either log $CV_A$ or log $CV_R$. Furthermore the correlation between log $CV_A$ and log $CV_R$ is highly significant and it has a critical statistical effect of reducing the variance explained by both log $CV_A$ and log $CV_R$, as both partials are larger than 1 (Fig. 2b). Only a causal model that includes the addition of an intercorrelated path between log $CV_A$ and log $CV_R$ (correlated effects model) is consistent with these observations (Fig. 2a). This result indicates that a disproportionate, independent effect of log $CV_R$ (or log $CV_A$) on behavioural heritability is not a sufficient explanation of their causal relationships (Li, 1975).

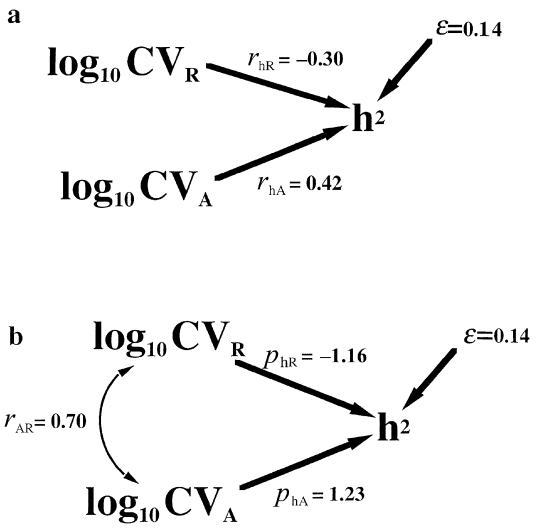


**Fig. 2** Two path analytic, causal models relating $CV_A$, $CV_R$ and $h^2$ (uncorrelated model above and intercorrelated effects model below). In the uncorrelated cause model above (a), the independent effects of $CV_R$ and $CV_A$ are assumed to govern variability in $h^2$ (Li, 1975). In this case, the partial regression coefficients ($p_{hR}$, $p_{hA}$) should equal their correlation coefficients ($r_{hA}$, $r_{hR}$) and one or both of their squared values (coefficients of determination) should largely explain the nonresidual variance ($r_{hA}^2 + r_{hR}^2 = 1.0 - r^2_{residual}$). In the intercorrelated causes model (b), the common causes of $h^2$ are correlated ($r_{AR}$), so the effects of $CV_R$ and $CV_A$ on heritability are through their direct effects measured by the standardized partial regression coefficients ($p_{hR}$, $p_{hA}$) and their intercorrelated effect [i.e. $r_{hR} = p_{hR} + p_{hA} \times r_{AR}$; variance explained = $r^2_{residual} + p_{hR}^2 + p_{hA}^2 + 2(p_{hR} \times r_{AR} \times p_{hA}) = 1.0$].

### Cross-classification of heritabilities

All of the behavioural heritabilities were cross-classified into categories related to selection or structure. Therefore we used a fully factorial ANOVA with heritability as the dependent variable to test for significant effects of differences in: (1) two estimation method categories, (2) wild/domestic categories, (3) three levels of integration and (4) seven behavioural ecology categories. With the full seven behavioural ecology categories, we could not estimate the three and four way interactions. However, heritabilities of the seven behaviour ecology categories did not differ significantly (one-way ANOVA: $F_{6,574} = 1.6$, $P$ = n.s.). Consequently, we collapsed the seven behavioural ecology categories into two metacategories:

first, categories presumed to have consistent, directional fitness relationships (foraging, predator avoidance, reproduction); and second, categories more loosely related to fitness (social, temperament, movement and other). We repeated the analysis using the fully factorial model with the two metacategories. The four-way interaction was not significant. The three-way interaction between level of analysis, estimation method and wild/domestic categories was significant, in addition to the estimation method main effect. We removed the nonsignificant four-way interaction and repeated the analysis, which produced the same result (three-way interaction (level of analysis × estimation method × wild/domestic): $F_{2,554} = 5.4$, $P < 0.01$; estimation method: $F_{1,554} = 12.4$, $P < 0.001$). All the other effects had probability values greater than 0.16. To illustrate determinants of this three-way interaction, we plotted average values governing the three, two-way combinations (Figs 3 and 4).

Average heritability estimates of wild/domestic categories grouped by estimation method indicate dominance effects are larger in domestic (and semidomestic species) than in nondomestic (wild) species (Fig. 3). This observation is consistent with the idea that directional selection related to domestication reduces additive genetic variation. The distribution of wild and domestic species did not affect central tendency of $h^2$, for both categories had a similar distribution across the range of heritability values (Fig. 1 inset). The difference between estimation method categories in wild (nondomestic) species was not significant ($t_{295} = 1.24$, $P =$ n.s.). Roff & Mousseau (1987) and Mousseau & Roff (1987) reported similar

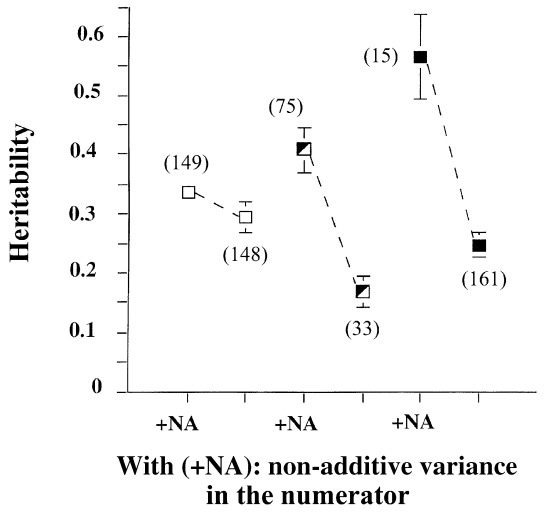


**Fig. 3** Average behaviour heritabilities of two estimation method categories (with or without dominance) in wild (□), semidomestic (◩) and domestic species (■). Estimation methods with dominance effects included in the numerator (+NA) are to the left of every pair joined by the dashed line. The bars indicate 1 SE. Sample sizes for each estimation method/wild-domestic category are in the brackets. Detailed category descriptions are in the text.

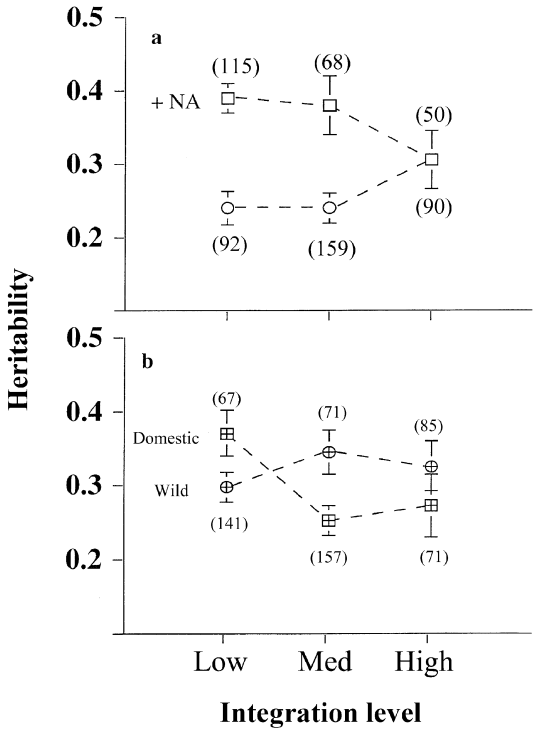


**Fig. 4** Average heritability of different estimation methods (with or without dominance in the denominator) in the upper diagram (a) and domestic-wild categories in the lower diagram (b). In diagram (a), +NA indicates heritability estimates that include dominance (square symbols), whereas the round symbols indicate estimates that do not include dominance. In diagram b, the square symbol indicates domestic categories and the round indicates wild categories. In both diagrams, average heritability of the different classes is estimated at three levels of integration (low, medium and high). The bars indicate 1 SE and sample sizes for each estimation method/wild-domestic category are in brackets.

observations. Finally, our estimate of semidomestic average heritability that does not include dominance variance components (0.17) was significantly smaller than additive estimates for domestic and wild species (Fig. 3). But, the semidomestic estimate was similar to average heritabilities (0.18) reported by Roff & Mousseau (1987) and Hoffmann (2000) for *Drosophila* behaviour.

Average heritabilities of estimation method categories (Fig. 4a) or wild/domestic classes (Fig. 4b) did not differ significantly at the high integration level, but did at lower integration levels. In particular, dominance effects in the numerator significantly biased estimation methods upwards at low and medium levels of integration, but not at the high integration level (Fig. 4a). This did not appear to be because of different proportions of wild and domestic species (Fig. 4b). If selection is stronger on highly integrated traits, or residual variation is disproportionately high compared with additive variation in highly integrated characters, then there should

be divergence in these category averages across integration levels. But, they converged (Fig. 4a). If these effects of selection and structure are stronger at higher integration levels, then we would also expect a decline in average heritabilities across integration level. Neither plot indicated there was a decrease in heritability at a higher integration level (Fig. 4a,b). In summary, neither the expected differences related to selection nor variance compounding was observed at the highest integration level, although the dominance-related differences at lower integration levels indicates the effects of directional selection (Lynch & Deng, 1994; Crnokrak & Roff, 1995).

## Selection, structure and CV

There was less data available for testing hypotheses involving $CV_A$ and $CV_R$, so we restricted out statistical analysis to a median test and one-way Welch ANOVA analyses because of heterogeneous variances (Day & Quinn, 1989).

We first compared $CV_A$ between wild and domestic species (no CVs were available for semidomestic species) and found differences between categories in both $CV_A$ and $CV_R$. Assuming selection reduces additive genetic variation and selection is stronger in domestic species, $CV_A$ should be lower on average in the domestic category compared with the wild category. $CV_A$ of domestic species was lower ($CV_A$: domestic = 14.72 ± 4.9, $n = 30$; wild = 23.23 ± 2.5, $n = 40$), and a median test indicated this difference was highly significant ($\chi_1^2 = 14.720$; $P < 0.0001$). A *t*-test provided a marginal result ($t_{68} = 2.777$, $P = 0.10$), unless one extreme outlier was removed ($CV_A$ domestic = 10.34 ± 2.2: $t_{67} = 3.69$, $P < 0.001$). At the same time, average $CV_R$ of domestic species (30.8 ± 6.4, $n = 23$) was also significantly lower ($t_{54} = -2.9$, $P < 0.01$; median test, $\chi_1^2 = 8.8$, $P < 0.01$) than wild $CV_R$ (58.1 ± 6.4, $n = 33$). In other words, the difference between $CV_R$ of domestic and wild species changes in the same direction as the $CV_A$, but in this case, lower residual variation characterizes traits which appear to be under stronger selection.

### *Behaviour ecology categories*

Differences in $CV_A$ and $CV_R$ among behavioural ecology categories were marginally significant ($CV_A$; Welch ANOVA $F_{6,9.2} = 6.5$, $P < 0.01$; median test $\chi_6^2 = 17.9$; $P < 0.01$; $CV_R$; Welch ANOVA $F_{6,7} = 4.14$, $P < 0.05$; median test $\chi_6^2 = 15.31$; $P < 0.05$). $CV_A$ and $CV_R$ of life history-related traits (reproductive, foraging, antipredator) were higher than the grand average whereas $CV_A$ and $CV_R$ of the other traits (social, movement, temperament, other) were lower than the grand mean (Table 2). The slope of the regression of $\log_{10} CV_R$ on $\log_{10} CV_A$ across these seven categories was less than 1 ($\log CV_R = 0.75 + 0.70 \times \log CV_A$, $r^2 = 0.71$, $P < 0.05$, 95% CL = 0.18–1.22), indicating $CV_R$ did not increase disproportionately.

### *Integration level*

Both $CV_A$ and $CV_R$ increased from low to high integration level (Fig. 5). These results support Henderson's (1990) argument that structural differences classified by integration level are associated with differences in additive genetic variation and Roff's (1997) suggestion that these sources of variation covary ($CV_A$; Welch ANOVA $F_{2,34.9} = 11.49$, $P < 0.0001$; median test $\chi_1^2 = 14.768$; $P < 0.001$; $CV_R$; Welch ANOVA $F_{2,21} = 22.3$, $P < 0.0001$; median test $\chi_1^2 = 9.2$; $P < 0.01$). In particular, the slope of their relationship suggests the increase in $CV_R$ is greater on average than $CV_A$ ($\log CV_R = -0.33 \times 1.6 \log CVA$, $r^2 = 0.99$, $P < 0.05$, Fig. 5), consistent with variance compounding hypotheses.

**Table 2** Means ($\overline{X}$), standard errors (SE) and sample sizes (*n*) for heritability, $CV_A$, and $CV_R$ estimates classified as behaviour ecology or estimation method.

| | | Heritability | | | $CV_A$ | | | $CV_R$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Classification | Category | $\overline{X}$ | SE | *n* | $\overline{X}$ | SE | *n* | $\overline{X}$ | SE | *n* |
| Behavioural ecology | Anti-predator | 0.33 | 0.038 | 41 | 23.0 | 9.8 | 2 | 73.0 | 24.0 | 2 |
| | Reproductive | 0.32 | 0.025 | 137 | 25.4 | 5.0 | 10 | 50.5 | 10.2 | 11 |
| | Foraging | 0.29 | 0.026 | 93 | 23.6 | 3.8 | 25 | 67.7 | 8.5 | 16 |
| | Temperament | 0.31 | 0.039 | 48 | 14.2 | 3.4 | 2 | 22.8 | 20.0 | 3 |
| | Movement | 0.30 | 0.026 | 124 | 14.1 | 8.7 | 15 | 41.5 | 9.1 | 14 |
| | Social | 0.30 | 0.041 | 69 | 8.1 | 9.7 | 6 | 20.1 | 24.0 | 2 |
| | Other | 0.30 | 0.042 | 69 | 4.7 | 1.2 | 8 | 18.6 | 12.0 | 8 |
| Estimation method | Full sib | 0.40 | 0.036 | 63 | | | | | | |
| | Diallel | 0.37 | 0.037 | 58 | | | | | | |
| | ANOVA | 0.37 | 0.034 | 69 | | | | | | |
| | Broad | 0.35 | 0.040 | 49 | | | | | | |
| | Regression | 0.30 | 0.034 | 69 | | | | | | |
| | Selection | 0.26 | 0.032 | 76 | | | | | | |
| | Half sib | 0.26 | 0.022 | 172 | | | | | | |
| | Maximum likelihood | 0.17 | 0.034 | 25 | | | | | | |

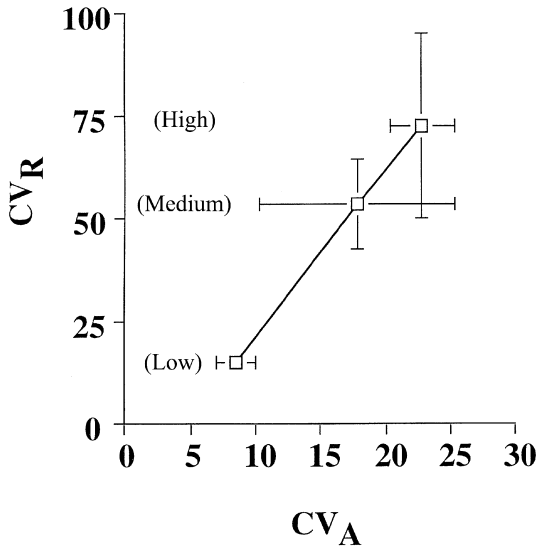


**Fig. 5** Regression of $CV_R$ on $CV_A$ across low, medium and high integration levels. The regression coefficient is significant, positive and greater than one ($CV_R = -18.9 \times 4.01\ CV_A$, $F_{1,1} = 1405.6$, $P = 0.01$, $r^2 = 0.99$; $\log CV_R = -0.33 + 1.6 \times CV_A$, $r^2 = 0.99$, $P = 0.04$) It is also the same as a geometric mean regression (Ricker, 1984), which takes variation in both axes into account. The horizontal and vertical bars indicate 1 SE, and error bars may be smaller than the symbol.

### Scale, means and variation

Because of the small number of studies reporting $CV_A$, $CV_R$ and particularly means ($\bar{x}$), we used a data set that included behavioural, life history and morphological character parameters ($n = 191$) to test whether in general, a log transformation equilibrates residual variation of relationships between $V_A$ and character means. In addition, log-transformed values of $V_A$ and $\bar{x}$ can be organized into a linear model, wherein the slope coefficient ($\log V_A$ = intercept + coefficient $\times \log \bar{x}$) tests whether additive genetic variance is proportional to the mean or its square. Character means $V_R$ and $V_A$ are positively and significantly related on both the arithmetic and log scale, however, in the $\log_{10}$ transformed relationships, both residual and additive genetic variation have closer relationships with character averages ($V_R = -109.7 + 6.02 \times \bar{x}$, $r^2 = 0.34$, $F_{1,186} = 97.5$, $P < 0.0001$; $\log V_R = 1.82 + 1.87 \times \log \bar{x}$, $r^2 = 0.71$, $F_{1,184} = 451.3$, $P < 0.0001$; $V_A = -634.8 + 16.99 \times \bar{x}$, $r^2 = 0.31$, $F_{1,187} = 84.19$, $P$ 0.0001; $\log V_A = 2.25 + 1.83 \times \log \bar{x}$, $r^2 = 0.80$, $F_{1,187} = 693.9$, $P < 0.0001$). Furthermore, the slope coefficient on the log scale indicates both variance components are approximately proportional to the square of their mean.

On an arithmetic scale, average $CV_A$ and $CV_R$ differed significantly between integration levels ($CV_A$: Welch ANOVA $F_{2,35} = 11.27$, $P < 0.001$; $CV_R$: $F_{2,\ 19.4} = 22.6$, $P < 0.0001$), indicating that differences between these categories are not entirely because of proportionality of their means and variances (Falconer, 1981). $CV_A$ and $CV_R$ were positively correlated across behavioural ecology categories and integration levels. This suggests that neither measurement error (Rohlf *et al.*, 1983) nor component/composite correlations (Lande, 1977) can account for differences within these categories.

## Discussion

There may not be a general expectation about the average relationship between behaviour and fitness (Houle, 1992), although behavioural ecologists often assume many foraging and reproductive behaviours are closely related to fitness. In general, this assumption appears justified, for average (or median) behavioural heritability is similar to average life history heritability. At the same time, this indicates mechanisms causing (low) heritabilities of behaviour pose much the same problem as low life history heritabilities (Price & Schluter, 1991; Houle, 1992, 1998; Roff, 1997; Merilä & Sheldon, 1999). Low heritabilities could be caused by an erosion of additive genetic variation, by selection (elimination-selection hypothesis) or, by a disproportionate increase in residual variation (disproportionate variance compounding hypothesis: Price & Schluter, 1991; Houle, 1992, 1998; Merilä & Sheldon, 2000). Both hypotheses predict that each mechanism should lower heritability and each hypothesis assumes its effect on heritability is independent. Directional differences in both $CV_A$ and heritability indicated elimination-selection had a detectable effect on heritability, but there was also a disproportionate increase in $CV_R$ with integration level. However, heritability did not differ between behavioural categories thought to be more or less closely related to fitness, nor was it low at the highest integration level. These results suggest that the nature of a behavioural character predisposes it to differences in elimination selection and its structure governs how it integrates variation, but selection, structure and their interaction appear to determine variation in heritability.

### Evidence of selection

Relatively high dominance variation suggests directional selection on behaviour may be common in domestic and semidomestic environments, but not in wild populations. We can estimate the relative contributions of dominance and additive variance using dominance variance standardized by the sum of dominance and additive variance $D_\alpha$ [$D_\alpha = V_d/(V_d + V_a)$, Crnokrak & Roff, 1995]. Our estimates for semidomestic (0.58) and domestic (0.55) species were both not significantly different from the life history estimate reported by Crnokrak & Roff (1995). This suggests behavioural traits in these categories are under directional selection that is as strong as selection on life history traits. At the same time, our

estimate of $D_\alpha$ for behaviour of wild populations (0.23) is much smaller, although it is not significantly different from Crnokrak & Roff's (1995) estimate for the same category. The relatively low proportion of dominance variance in wild or nondomestic populations means that heritability estimates are not significantly biased by the estimation method. In other words, full sib breeding designs could be used for estimating additive genetic variation of wild population behaviour. Dominance variance of domestic and semidomestic populations was high and $CV_A$ of both categories was smaller than wild species. Finally, there was a concomitant difference in $CV_R$, which in this case is not related to variance compounding effects of structure (e.g. Houle, 1992, 1998; Pomiankowski & Moller, 1995; Houle *et al.*, 1996; Kruuk *et al.*, 2000; Merilä & Sheldon, 2000). These results indicate that additive genetic variation of behavioural characters of animals under domestication is generally being eroded by directional selection and that there may be related changes in residual variation.

### Variance compounding effects of structure

Both low $CV_A$ and low heritabilities of behaviour in semidomestic species suggest that elimination-selection may regulate (in part) behavioural heritabilities by reducing the size of additive genetic variation in the numerator. At the same time, low heritabilities have generally been explained by their large residual variance measured by $CV_R$ ($= CV_P - CV_A$) in the denominator (Houle, 1992; Pomiankowski & Moller, 1995; Merilä & Sheldon, 2000; but see Kruuk *et al.*, 2000). Merilä & Sheldon (2000) point out that if large residual variation governs low heritability estimates, then the difference in residual variation between life history and morphological traits should be disproportionately larger than the corresponding difference in additive genetic variation. They report a disproportionate increase in $CV_R$ in an extensive data set on the collared flycatcher (*Ficedula albicollis*). They also note that $CV_A$ and $CV_R$ are positively and strongly correlated in both males and females, but argue this covariation does not influence the disproportionate increase in $CV_R$. In other words, they hypothesize that $CV_R$ and $CV_A$ are independent causes of variation in heritability. We used their published results to construct a causal analysis of relationships between these three variables ($CV_A$, $CV_R$, $h^2$). First we note that their partial regression coefficient of $CV_R$ on $h^2$ is larger than one. On one hand, this result is consistent with their argument of a disproportionately strong effect of residual phenotypic variation on heritability. But, on the other hand, this result is inconsistent with an independent effects model (Fig. 2), because there must be an intercorrelated (complementary variable) which diminishes the variance of $h^2$, to at least the complement of the unexplained error variation (Li, 1975). Second, we found that intercorrelated or indirect effects of both $CV_A$ and $CV_R$ had to be included in their causal analysis, similar to our analysis of relationships between $CV_A$ and $CV_R$ and behavioural heritability (Fig. 2). But, what is the nature of these intercorrelated effects?

Our results indicate a distinction should be made between the effects of selection or structure on phenotypic variance components and mechanisms governing relationships between them. The evidence for effects of selection was discussed above. $CV_R$ of behavioural trait categories was larger than corresponding $CV_A$ regardless of how they were classified, similar to studies cited above on morphology, life history and sexually selected characters. At the same time, both $CV_A$ and $CV_R$ increased across these categories. We proposed that if the slope of the relationship of log $CV_R$ on log $CV_A$ is greater than one, then there is evidence for a disproportionate increase in $CV_R$. $CV_R$ does increase disproportionately faster than $CV_A$ across different integration levels, indicating the importance of structure. This observation also suggests Price & Schluter's (1991) model should apply to variation between behavioural categories, in addition to differences between life history and morpholological phenotypes (Merilä & Sheldon, 2000). But, we did not see different heritabilities in behavioural ecology categories that had significantly different $CV_A$ and $CV_R$ (Table 2), nor was lower heritability observed at the highest integration level (Fig. 5). In addition, both the path analysis and the interactions between factors related to selection and structure indicate the intuitive relationship between $h^2$, $CV_R$ and $CV_A$ was not solely governed by direct effects of selection or structure as assumed in independent effects models. We suggest the intercorrelated effect, which appears to have its strongest influence at the highest integration level, is related to the influence of mechanisms causing phenotypic plasticity. However, different measurement scales, transformations, mathematical constraints, scaling of errors, arbitrary categories, or sampling effects are thought to cause or confound the interpretation of similar relationships.

### Scale effects

Our estimates of behavioural heritabilities in domestic, semidomestic and wild (nondomestic) species are very similar to those reported in other, independently organized studies. Furthermore, the magnitude of variation in $CV_A$ and $CV_R$ across behavioural ecology classes (Table 1) or integration levels (Fig. 5) is similar to variation reported across life history and morphology categories (e.g. Houle, 1992; Kruuk *et al.*, 2000; Merilä & Sheldon, 2000). At the same time, scaling effects strictly due to various phenotypic measurement scales (Price & Schluter, 1991), must be less of a problem within a phenotype category (i.e. behaviour). Heritabilities are scale invariant (Bryant, 1986). But, Roff (1997) suggests comparisons of $CV_A$ and $CV_R$ may involve differences related to a scale transformation between these variance components.

Thus, one problem is determining a scale where means and variances of $CV_A$ and $CV_R$ have similar, proportional relationships. We found log-transformed means and variances had similar relationships in both coefficients ($CV_A$ and $CV_R$). In addition, the log-transformed relationships between $CV_A$ and $CV_R$ were positive, indicating on one hand that their relationship across behavioural ecology categories was not caused by measurement errors (Rohlf *et al*., 1983) and on the other that part – whole correlations (Lande, 1977) cannot explain the relationship across integration levels. It could also be argued that our categories are arbitrary with respect to fitness or structure. This cannot be discounted in principal, and within each category we acknowledge the problems with assigning behavioural characters to ordinal categories. However, we observed that $CV_R$ was higher than $CV_A$ in each category class and we found differences in $CV_A$ between behavioural ecology classes which are commonly thought to differ in both their structure and their relationship with fitness (Henderson, 1990). Similar results have been reported in studies on differences between morphology and life history (Houle, 1992; Kruuk *et al*., 2000; Merilä & Sheldon, 2000). At the same time, the sample population of studies reporting coefficients of variation is small relative to the heritability sample population. To this we can only reinforce Houle's (1992) appeal that these statistics be routinely reported.

### Selection, structure and genotype by environment interactions

How does the nature of a behavioural trait determine it's heritability? Behavioural character heritabilities differ with respect to effects of selection, for which we found more evidence in characters considered to be less closely related to fitness and less affected by variance compounding. Similar to Pomiankowski & Moller's (1995) study which showed that sexually selected morphological traits had a different structure than other morphological traits, behavioural character categories that differ in structure have different $CV_A$ and $CV_R$. Life history traits such as fecundity genetically integrate (Cheverud, 1996) both morphological and physiological characters within a single life stage (e.g. Stirling *et al*., 2001). This is also the case for complex behavioural characters (Henderson, 1990). On one hand, both $CV_A$ and heritability are standardized measures of additive genetic variation (Houle, 1992), but on the other hand, nonsignificant correlations between $CV_A$ and heritability suggest they are not surrogate measures of genetic variation (Houle, 1992; Pomiankowski & Moller, 1995). Furthermore, $CV_R$ is typically larger than $CV_A$ (Merilä & Sheldon, 1999), implying that compounded environmental variation governs differences in heritability (Cabana & Kramer, 1991; Price & Schluter, 1991). However, our study shows that $CV_A$ and heritability are positively related in behavioural characters, and disproportionate variance compounding does not generally explain variation in behavioural heritabilities. Our analysis points to the importance of relationships between additive genetic and nonadditive genetic residual variation. Although additive genetic variation is a distinct variance component, residual variation comprises dominance, genotype by environment interactions, epistatic and other nonadditive genetic variation, in addition to environmental variation. Merilä & Sheldon (1999) point out that dominance variance is an important variance component in selection studies and that the magnitude of this component would negatively covary with additive genetic variation. But, this does not explain our observation that average heritability of category classes with different dominance variation converged at the highest integration level and that $CV_A$ and $CV_R$ are positively related. Alternatively, highly integrated behaviours such as foraging, mate choice and predator avoidance are characterized by conditional strategies or phenotypically plastic responses to environmental cues. Genetic variation for phenotypic plasticity may be because of both additive genetic variance and genotype by environment interaction variance. In this light, heritability may be viewed not as a misleading standardized index of additive genetic variation, but as a statistic which integrates effects of selection, structure and the interrelationships between variance components, all of which may contribute to the genetic architecture of different quantitative characters.

## Acknowledgments

Many thanks to both reviewers for their constructive comments on the manuscript as well as to the authors of the publications we used to generate the data for this study.

## References

Bryant, E.H. 1986. On the use of logarithms to accommodate scale. *Syst. Zool.* **35**: 552–559.

Cabana, G. & Kramer, D.L. 1991. Random offspring mortality and variation in parental fitness. *Evolution* **45**: 228–234.

Cheverud, J.M. 1996. Developmental integration and the evolution of pleiotropy. *Am. Zool.* **36**: 44–50.

Crnokrak, P. & Roff, D.A. 1995. Dominance variance: associations with selection and fitness. *Heredity* **75**: 530–540.

Curtsinger, J., Service, W.P.M. & Prout, T. 1994. Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *Am. Nat.* **144**: 210–228.

Day, R.W. & Quinn, G.F. 1989. Comparisons of treatments after an analysis of variance in ecology. *Ecol. Monogr.* **59**: 433–463.

Ewens, W.J. 1989. An interpretation and proof of the fundamental theorem of natural selection. *Theor. Popul. Biol.* **36**: 167–180.

Falconer, D.S. 1981. *Introduction to Quantitative Genetics*, 2nd edn. Longman, New York.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Longman, New York.

Gustafsson, L. 1986. Lifetime reproductive success and heritability: empirical support for Fisher's fundamental theorem. *Am. Nat.* **128**: 761–764.

Harshman, L.G. & Hoffmann, A.A. 2000. Laboratory selection experiments using *Drosophila*: what do they really tell us ? *Trends Ecol. Evol.* **15**: 32–36.

Hartl, D.L. & Clark, A.G. 1989. *Principals of Population Genetics*, 2nd edn. Sinauer, Sunderland.

Henderson, N.O. 1990. Quantitative genetic analysis of neurobehavioural phenotypes. In: *Developmental Behaviour Genetics: Neural, Biomedical and Evolutionary Approaches* (M. E. Hahn, J. K. Hewit & N. D. Henderson, eds). Oxford University Press, New York, pp. 283–287.

Hoffmann, A.A. 2000. Laboratory and field heritabilities: some lessons from *Drosophhila*. In: *Adaptive Genetic Variation in the Wild* (T. A. Mousseau, B. Sinervo & J. A. Endler, eds). Oxford University Press, New York, pp. 200–218.

Houle, D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution* **45**: 630–648.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Houle, D. 1998. How should we explain variation in the genetic variance of traits. *Genetica* **102/103**: 241–253.

Houle, D., Morikawa, B. & Lynch, M. 1996. Comparing mutational variabilities. *Genetics* **143**: 1467–1483.

Kruuk, L.E.B., Clutton-Brock, T.H., Slate, J., Pemberton, J.M., Brotherstone, S. & Guinness, F.E. 2000. Heritability of fitness in a wild mammal population. *Proc. Natl. Acad. Sci. USA* **97**: 698–703.

Lande, R. 1977. On comparing coefficients of variation. *Syst. Zool.* **26**: 214–217.

Li, C.C. 1975. *Path Analysis – a Primer*. Pacific Grove, California.

Lynch, M. & Deng, H.W. 1994. Genetic slippage in response to sex. *Am. Nat.* **144**: 242–261.

Merilä, J. & Sheldon, B.C. 1999. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity* **83**: 103–109.

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

Mousseau, T. 2000. Intra- and interpopulation genetic variation: explaining the past and predicting the future. In: *Adaptive Genetic Variation in the Wild* (T. A. Mousseau, B. Sinervo & J. A. Endler, eds), pp. 219–250. Oxford, New York.

Mousseau, T.A. & Roff, D.A. 1987. Natural selection and the heritability of fitness components. *Heredity* **59**: 181–197.

Pomiankowski, A. & Moller, A.P. 1995. A resolution of the lek paradox. *Proc. R. Soc. Lond. B* **260**: 21–29.

Price, G.R. 1972. Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet.* **36**: 129–140.

Price, T. & Schluter, D. 1991. On the low heritability of life history traits. *Evolution* **45**: 853–861.

Ricker, W.E. 1984. Computation and uses of central trend lines. *Can. J. Zool.* **62**: 1897–1905.

Roff, D.A. 1997. *Evolutionary Quantitative Genetics*. Longman, New York.

Roff, D.A. & Mousseau, T.A. 1987. Quantitative genetics and fitness: lessons from *Drosophila*. *Heredity* **58**: 103–118.

Rohlf, F.A.J., Gilmatrin, A.J. & Hart, G. 1983. The Klug–Kerfoot phenomenon – a statistical artifact. *Evolution* **37**: 180–202.

Stirling, D.G., Fairbairn, D.J. & Jensen, S. & Roff, D.A. 2001. Does a negative genetic correlation between wing morph and early fecundity imply a functional constraint in *Gryllus firmus* ? *Evol. Ecol. Res.* **3**: 157–177.

Zar, J.H. 1996. *Biostatistical Analysis*. Simon and Schuster, Upper Saddle River, NJ.

*Received 30 August 2001; accepted 20 November 2001*