



## Short- and long-term repeatability of docility in the roe deer: sex and age matter



L. Debeffe <sup>a,b,\*</sup>, J. F. Lemaître <sup>a</sup>, U. A. Bergvall <sup>c,d</sup>, A. J. M. Hewison <sup>b</sup>, J. M. Gaillard <sup>a</sup>,  
N. Morellet <sup>b</sup>, M. Goulard <sup>e</sup>, C. Monestier <sup>a,b</sup>, M. David <sup>f,g</sup>, H. Verheyden-Tixier <sup>b</sup>,  
L. Jäderberg <sup>d</sup>, C. Vanpé <sup>a,\*</sup>, P. Kjellander <sup>d</sup>

<sup>a</sup> CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, Villeurbanne, France

<sup>b</sup> INRA, UR035 Comportement et Ecologie de la Faune Sauvage, Institut National de la Recherche Agronomique, Castanet-Tolosan, France

<sup>c</sup> Department of Zoology, Stockholm University, Stockholm, Sweden

<sup>d</sup> Grimsö Wildlife Research Station, Department of Ecology, Swedish University of Agricultural Science, Grimsö, Sweden

<sup>e</sup> INRA, UMR1201 DYNAFOR, Institut National de la Recherche Agronomique, Castanet-Tolosan, France

<sup>f</sup> Department of Biology-Ethology, University of Antwerp, Wilrijk, Belgium

<sup>g</sup> Centre for Ecology & Conservation, School of Biosciences, College of Life & Environmental Sciences, University of Exeter, Cornwall Campus, Penryn, U.K.

### ARTICLE INFO

#### Article history:

Received 21 December 2014

Initial acceptance 22 January 2015

Final acceptance 21 July 2015

Available online

MS. number: 14-01045R

#### Keywords:

*Capreolus capreolus*

capture

individual differences

mammal

personality

repeatability

Behavioural consistency is a key assumption when evaluating how between-individual differences in behaviour influence life history tactics. Hence, understanding how and why variation in behavioural repeatability occurs is crucial. While analyses of behavioural repeatability are common, few studies of wild populations have investigated variation in repeatability in relation to individual status (e.g. sex, age, condition) and over different timescales. Here, we aimed to fill this gap by assessing within-population variation in the repeatability of docility, as assessed by the individual's response to human handling, in a free-ranging population of European roe deer, *Capreolus capreolus*. Docility was an equally repeatable behaviour at both short- and long-term timescales, suggesting that this behavioural trait is stable across time. Repeatability did not differ markedly between age and sex categories but tended to be higher in juvenile males than in juvenile females. Finally, contrary to expectation, individual variation in the repeatability of docility was not correlated with individual body mass. Further studies are required to assess the life history consequences of the individual variation in docility we report here.

© 2015 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Interest in the study of behavioural differences between individuals has increased markedly over recent years (Bell, 2007; Sih, Cote, Evans, Fogarty, & Pruitt, 2012) and consistent behavioural differences between individuals over time and/or across different ecological contexts have been documented within populations in a large variety of taxa (Bell, 2007; Bell, Hankison, & Laskowski, 2009). These consistent behavioural differences can indeed have profound effects on life histories (Bell, 2007; Biro & Stamps, 2008; Réale, Reader, Sol, McDougall, & Dingemanse, 2007). For instance,

annual adult survival is related to exploration behaviour in great tits, *Parus major* (Dingemanse, Both, Drent, & Tinbergen, 2004). In the common lizard, *Lacerta vivipara*, socially tolerant individuals grow faster and reproduce better than intolerant individuals (Cote, Dreiss, & Clobert, 2008). Similarly, boldness and docility are linked to survival and reproductive success in bighorn sheep rams, *Ovis canadensis* (Réale, Martin, Coltman, Poissant, & Festa-Bianchet, 2009), and to weaning success in bighorn sheep ewes (Réale, Gallant, Leblanc, & Festa-Bianchet, 2000).

Behavioural consistency can be assessed by evaluating repeatability (Hayes & Jenkins, 1997), where repeatability is defined as the proportion of observed behavioural variation within a population that is accounted for by individual differences (Bell et al., 2009). This then requires that behavioural traits be assessed repeatedly in the same individuals. The vast majority of studies that have assessed behavioural repeatability have been conducted under

\* Correspondence: C. Vanpé, Laboratoire Biométrie et Biologie Evolutive, UMR CNRS 5558, UCB Lyon 1, Bât. Grégor Mendel, 43 bd du 11 novembre 1918, 69622 Villeurbanne Cedex, France; L. Debeffe, Department of Biology, University of Saskatchewan, 112 Science Place, Saskatoon, SK S7N 5E2, Canada.

E-mail addresses: [lucie.debeffe@gmail.com](mailto:lucie.debeffe@gmail.com) (L. Debeffe), [cecile.vanpe@univ-lyon1.fr](mailto:cecile.vanpe@univ-lyon1.fr) (C. Vanpé).

somewhat unrealistic laboratory conditions (e.g. see Bell et al., 2009 for a review), whereas repeatability measures of behavioural traits in the wild are less common (e.g. Martin & Réale, 2008a in eastern chipmunks, *Tamias striatus*; Réale et al., 2000 in bighorn sheep; Montiglio, Garant, Thomas, & Réale, 2010 in house mice, *Mus musculus*, and eastern chipmunks; Garamszegi, Eens, & Janos, 2009 in collared flycatchers, *Ficedula albicollis*, Nakagawa, Gillespie, Hatchwell, & Burke, 2007 in house sparrows, *Passer domesticus*).

Despite the importance of individual behavioural differences for understanding the evolution of life history tactics (Sih et al., 2012), the individual characteristics that influence repeatability are still to be fully understood and require further investigation (Bell et al., 2009). Using a meta-analysis based on 114 studies (466 estimates assessed in the laboratory versus 293 in the field) investigating the repeatability of various behavioural traits, Bell et al. (2009) found that within-population variation in repeatability depends on the focal behavioural trait, as well as the species under study, and is affected by (1) time elapsed between measurements, (2) age, (3) sex and (4) ecological conditions.

Behavioural repeatability generally decreases as the time elapsed between successive measurements increases (Bell et al., 2009; David, Auclair, & Cezilly, 2012; Gifford, Clay, & Careau, 2014). This is probably because, with time, there is more opportunity for individuals to experience developmental modifications and fluctuating environments (Stamps & Groothuis, 2010), and also because their phenotype may be influenced by different genes at different ages (Charmantier, Perrins, McCleery, & Sheldon, 2006). Thus, experimental trials performed over short time intervals (i.e. short-term repeatability) generally yield higher estimates of repeatability than those carried out over longer time intervals (i.e. long-term repeatability) for a large range of traits, including docility (Bell et al., 2009; David et al., 2012; Gifford et al., 2014).

Repeatability also varies in relation to age, although predictions for age-dependent variations are not straightforward (Bell et al., 2009). Assuming that behavioural repeatability has an ontogenetic component (Stamps & Groothuis, 2010; i.e. it develops progressively over a lifetime), the behaviour of older individuals should be more repeatable than that of younger ones (De Kort, Eldermire, Valderrama, Botero, & Vehrencamp, 2009). For instance, behavioural repeatability in humans increases with age up to 50 years old before reaching a plateau (Roberts & DelVecchio, 2000). Hence, higher repeatability of docility is expected among adults.

Behavioural repeatability may differ between the sexes under the action of sexual selection (Schuett, Tregenza, & Dall, 2010). Indeed, behaviours shaped by sexual selection should provide reliable, honest, and thus consistent, cues, so that they can reliably be interpreted by conspecifics (Bell et al., 2009; Schuett, Godin, & Dall, 2011). For instance, in two kangaroo rat species (*Dipodomys* spp.), food-hoarding behaviour, a behavioural trait that may partially underpin alternative mating tactics, has been reported to be more repeatable in males than in females (Jenkins, 2011). Surprisingly, however, based on the published literature, Bell et al. (2009) concluded that females generally tend to display more repeatable behaviours (except for mate choice-related behaviours) than males. Hence, between-sex differences in repeatability may depend on the behavioural trait under scrutiny, with males being more repeatable for behaviours linked to mate choice and subject to intrasexual competition (Schuett et al., 2010).

Repeatability may also vary among individuals within a given sex or age class. At the population level, a behavioural consistency estimate that is statistically different from 0 for a given trait in a given age and sex category does not mean that all individuals are equally consistent for that trait (Bell et al., 2009). Indeed, individual variation in predictability, or within-individual variability,

once variation due to sex and/or age differences has been accounted for, is widespread (Briffa, 2013; David et al., 2014; Stamps, Briffa, & Biro, 2012) and may have consequences for fitness-related traits. For instance, in hermit crabs *Pagurus bernhardus*, within-individual variation was higher in a risky situation, suggesting that animals benefit by behaving unpredictably under predation threat (Briffa, 2013). Such flexibility is assumed to be costly (DeWitt, Sih, & Wilson, 1998; Niemela, Vainikka, Forsman, Loukola, & Kortet, 2013) and should therefore vary as a function of individual body condition (David & Giraldeau, 2012; David et al., 2014). For example, zebra finches, *Taeniopygia guttata*, in higher body condition are less consistent in their producer tactic use in a producer–scrounger foraging game (David et al., 2014). A link between body condition and variation in docility may thus be expected.

The objective of this study was thus to test the effect of the above factors on the repeatability of docility in European roe deer, *Capreolus capreolus*. Docility is a commonly measured behavioural trait (Benhajali et al., 2010; Martin & Réale, 2008a; Réale et al., 2000) often associated with the shy–bold behavioural gradient (Réale et al., 2007). Indeed, docility has often been linked with other behavioural traits such as activity and exploration (Ferrari et al., 2013; Martin & Réale, 2008a) and with physiological parameters such as hypothalamo–pituitary–adrenal reactivity (Montiglio et al., 2010) or the level of cortisol (Martin & Réale, 2008b). Docility has been reported to be a repeatable behavioural trait in several studies, both in captivity (e.g. David et al., 2012; Mazurek et al., 2011) and in the field (Ferrari et al., 2013; Petelle, McCoy, Alejandro, Martin, & Blumstein, 2013; Réale et al., 2000, 2009). Moreover, docility is linked to fitness components in bighorn sheep (Réale et al., 2000, 2009), while it also influences the productivity of livestock farming enterprises (Norris, Ngambi, Mabelebele, Alabi, & Benyi, 2014). For instance, fearful cows produce lighter calves at birth that grow less well than those produced by nonfearful (i.e. docile) cows (Turner, Jack, & Lawrence, 2013). Because docility may be part of a behavioural syndrome (i.e. a suite of correlated behaviours reflecting between-individual consistency in behaviour across multiple situations; Sih, Bell, & Johnson, 2004), variation in docility may thus have important consequences for life history traits. Information about which factors influence the repeatability of docility remains scarce, however. In yellow-bellied marmots, *Marmota flaviventris*, Petelle et al. (2013) reported that boldness was repeatable only in yearlings, whereas docility was repeatable in all age classes. Repeatability of docility was also shown to be higher when measured over the short term than over the long term (Bell et al., 2009; David et al., 2012).

In this study, we investigated the repeatability of docility taken as the behavioural responses to human manipulation at capture in a free-ranging, wild population of European roe deer in central Sweden. The response of roe deer to human handling during capture is assumed to reliably reflect docility (Le Neindre et al., 1995; Réale et al., 2000). The roe deer is a medium-sized (20–30 kg), slightly dimorphic and weakly polygynous ungulate in which adults of both sexes are highly sedentary (Andersen, Duncan, & Linnell, 1998). Roe deer life expectancy in the wild at 1 year of age is around 10–12 years in the absence of hunting, but varies among populations (Veiberg et al., 2007). Despite being primarily a forest-dwelling species, roe deer have colonized almost all types of habitats since the early 1980s, including human-modified landscapes (Andersen et al., 1998; Hewison et al., 2001), and express a high degree of behavioural plasticity (Jepsen & Topping, 2004). Here, we investigated whether the repeatability of docility varies in relation to the time elapsed between successive measurements, age, sex and body mass. We expected (1) short-term repeatability estimates to be higher than long-term ones, (2) repeatability to be

higher in adults than juveniles, and (3) repeatability to be similar in both sexes as docility is a priori not related to mate choice (Andersen et al., 1998). Finally, as individual body mass is stable over most of a roe deer's lifetime (Hewison et al., 2011) and constitutes a reliable indicator of body condition and, more generally, of phenotypic quality (Toïgo, Gaillard, Van Laere, Hewison, & Morellet, 2006), we expected (4) within-individual variation in behaviour (i.e. predictability) to decrease with decreasing body mass due to the presumed costs of behavioural flexibility.

## METHODS

### Ethical Note

All captures and handling were done in accordance with Swedish and European laws for animal welfare. All procedures were approved by the Ethical Committee for Animal Experiments in Uppsala, Sweden (permit numbers C302/12 and C289/09) and adhered to legal and ethical requirements for research on wild animals in Sweden and the European Union. Roe deer captures in our study site have been performed since 1973, and all animals used for this study were captured in the context of a long-term ongoing project on roe deer ecology. Some animals were fitted with a VHF or GPS collar to meet the objectives of other projects, including predator–prey interactions, herbivore–plant interactions, migration, disease ecology and veterinary medicine, and roe deer–human interactions using a variety of scientific methods and technical and analytical approaches.

### Study Area

The study was carried out from 2008 to 2013 at Grimsö Wildlife Research Area (GWRA; 130 km<sup>2</sup>), located in south-central Sweden (59°40' N, 15°25' E) within the southern part of the Nordic taiga. The area is covered mainly by mixed conifer forest (74%), while bogs, mires and fens cover 18% and farmland 3%. For a more detailed description of GWRA see Kjellander, Gaillard, and Hewison (2006). Daily mean temperatures range from 20 °C in summer to –10 °C in winter. The ground is generally snow covered from December to March.

### Behaviour at Capture

Each winter from 2008 to 2013, roe deer were caught in wooden box traps (L6 Rådjursfälla M/Öster Malma, dimensions: 130 × 62.5 cm and 100 cm high) baited with specific forage intended for deer (Renfor, manufacturer Lantmännen, Nyköping, Sweden). Roe deer captures spanned 185 days, from 12 November to 25 March, with 11 February as the median date. Hence, this capture period did not overlap with the phases of reproduction (i.e. the birth period in May–June and the rutting period in mid-July to mid-August; Sempéré, Mauget, & Mauget, 1998) or the period of male

territoriality (late March to early September; Liberg, Johansson, Andersen, & Linnell, 1998). Each individual was sexed, marked with plastic ear tags and weighed to the nearest 0.1 kg. Juveniles (<1 year old) were identified from the presence of a tricuspid third premolar milk tooth that is only present in individuals less than 1 year old (Cederlund, Kjellander, & Stålfelt, 1991) and distinguished from adults (≥1.5 years old). During manipulation, we measured docility for each roe deer using a handling score ranging from 0 to 4, a commonly used method for quantifying this behaviour (Benhajali et al., 2010; Le Neindre et al., 1995; Norris et al., 2014; Réale et al., 2000). The score was based on several behavioural component measures that indexed how much resistance roe deer showed during handling (see Table 1 for more details), and was assessed by the same experienced handler (L.J.) at each capture. Note that the absolute value of this score has no direct interpretation; rather it represents the judgement of an experienced handler as to how stressed an animal was during handling relative to other individuals in the population.

Between winter 2007–2008 and winter 2012–2013, 262 roe deer were caught for a total of 1569 capture events, but only individuals captured at least twice with recorded behavioural data were included in the analyses ( $N = 130$  individuals for a total of 532 capture events, mean  $\pm$  SE number of captures per individual =  $4.09 \pm 2.81$ , range 2–22; Fig. 1). A given individual could potentially be caught several times during a given winter ( $N = 117$  individuals) and/or during successive winters ( $N = 65$  individuals).

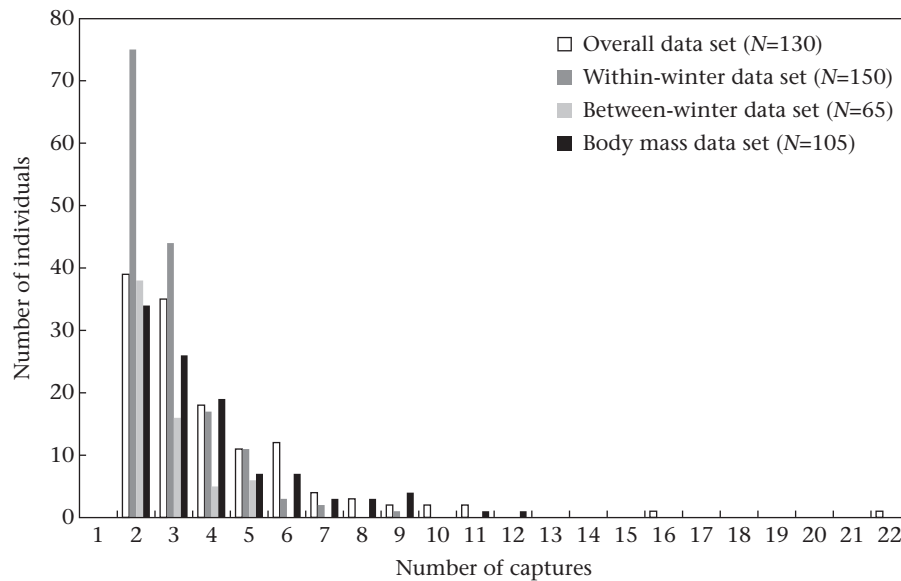
### Docility and Habituation

As pointed out by Martin and Réale (2008a), individuals might habituate to a behavioural assay when they are measured several times. This habituation may lead to a decrease in the strength of the behavioural response over trials and, as a result, bias repeatability estimates. For instance, previous experience with handling increases docility during subsequent capture in domestic cattle, *Bos taurus* (Le Neindre, Boivin, & Boissy, 1996), or decreases the intensity of the behavioural response in hole-board and handling bag tests in eastern chipmunks (Martin & Réale, 2008a) and captive zebra finches (David et al., 2012). To test whether docility was related to an individual's previous capture experience (number of capture events and time elapsed since the previous capture event), we used linear mixed models with handling score as the dependent variable, and with individual identity included as a random factor on the intercept to control for repeated measures. When comparing the model including the number of capture events for a given individual (on a log scale to achieve normality) with the null model, docility was linked to previous capture experience such that the handling score of a given individual decreased with the number of previous capture events ( $N_{\text{capture}} = 532$ ,  $N_{\text{individuals}} = 130$ ,  $t = -10.2$ , slope =  $-0.49 \pm 0.05$ ). To account for this habituation, we included the number of previous capture events for a given individual, as well as the time elapsed since the previous capture event, as fixed

**Table 1**  
Behavioural traits recorded at capture with the corresponding derived handling score

Handling score at capture	Resistance during handling	Kicking with legs	Screaming	Handling	Calm
0	No	No	No	Easy	Yes
1	No	Very few	0–2 times	Easy	Yes
2	Few	Yes	Yes	Fairly easy	Yes between screaming and kicking events
3	Yes	Yes	Yes	Difficult	No
4	Yes	Yes	Yes	Impossible	No

The score represents the judgement of an experienced handler (the same individual for all measures) as to how stressed an animal is during handling relative to other individuals in the population. The table does not include all possible combinations of behaviour but rather displays the more representative combinations of the subjective score (see Methods).



**Figure 1.** Sample size (number of individuals) and number of repeated measures (number of capture events for a given individual) of docility for each data subsample.

effects in the subsequent models. When this approach was not possible, we computed a ‘standardized handling score’ defined as the residual values from the linear regression between handling score and the log-transformed number of capture events using all available individuals and with the time elapsed since the previous capture event included as a fixed effect to control for the rate of habituation in relation to time between capture events (see [Appendix Fig. A1](#) for the linear regression and [Fig. A2](#) for the distributions of the standardized handling scores and handling scores, respectively). Note that because the rate of habituation did not differ between individuals (i.e. the model including individual identity with random slopes did not provide a better fit than the model without random slopes: difference in the Akaike information criterion corrected for small sample size,  $\Delta\text{AICc} = 1.3$ ), a common slope from a linear model was used to compute the standardized handling scores for all individuals. In addition, we estimated the capture rate of each individual as the total number of captures experienced by the individual during the study period divided by the number of days between the first and the last capture event and tested whether more docile individuals were captured more often than less docile ones that are potentially trap-shy. However, we did not find any relationship between an individual’s handling score at first capture and its capture rate (estimated coefficient from the multinomial regression for capture rate: slope  $\pm$  SE =  $0.28 \pm 2.67$ ;  $N = 130$ ,  $Z = 0.10$ ,  $P = 0.92$ ).

#### Variation in Docility

To investigate the pattern of docility among the sexes and age classes, we fitted six candidate mixed multinomial models with handling score as a response variable using the ‘ordinal’ R package ([Christensen, 2015](#); [R Development Core Team, 2010](#);  $N_{\text{observations}} = 521$ ,  $N_{\text{individuals}} = 129$ , because we lacked age information for one individual). The most complex model included the two-way interaction between sex and age class to evaluate the support for differences in docility between the sexes and across age classes, as well as the number of previous capture events for that individual and the time elapsed since the previous capture event to control for habituation. We then computed all possible simpler models. Individual identity was included in all models as a random

intercept to control for repeated measures of docility obtained for each individual. We used a model selection procedure based on the AICc and we retained the model with the lowest AICc value, reflecting the best compromise between precision and accuracy ([Burnham & Anderson, 2002](#)). According to the rule of parsimony, when the AICc of two competing models differed by less than 2, we retained the simplest one. We also calculated AICc weights (AICcWt) as a measure of the likelihood that a given model was the best among the set of fitted models.

#### Repeatability of Docility

Repeatability and its 95% credibility intervals (95% CI) were then assessed using multivariate generalized linear mixed models with the MCMCglmm package ([Hadfield, 2010](#)) of R ([R Development Core Team, 2010](#)), and implemented in this package under the item Categorical Random interactions. To control for differences between the sexes and age classes and/or in relation to habituation, adjusted repeatability, sensu [Nakagawa and Schielzeth \(2010\)](#), was assessed by adding the fixed effect of age class and sex and/or the number of previous capture events experienced and the time elapsed since previous capture to the linear mixed models. The resultant values estimate repeatability as if all measurements had been taken from a single class of animals and for a given number of capture events ([Nakagawa & Schielzeth, 2010](#)). Repeatability and adjusted repeatability ( $R$ ) were calculated as the proportion of the total variance accounted for by between-individual differences using multivariate linear mixed-effect models for raw handling scores and linear mixed-effect models for standardized handling scores. First, global adjusted repeatability of raw handling scores was estimated using the entire data set ( $N_{\text{observations}} = 521$ ,  $N_{\text{individuals}} = 129$ ) and by fitting a multinomial regression for ordinal data. For comparative purposes, global adjusted repeatability for standardized handling scores was estimated using a linear mixed-effect model for Gaussian data.

#### Short- and Long-term Repeatability

Within-winter adjusted repeatability of standardized handling scores was then assessed as a measure of short-term repeatability

while controlling for sex and age differences in repeatability and using a subsample of 116 individuals that were captured several times during a given winter (range 2–9). Because some individuals were also captured several times in successive winters, the total number of data points available was 150; hence a random effect of winter nested within individual was fitted to take into account pseudoreplication. Between-winter adjusted repeatability of standardized handling scores was assessed as a measure of long-term repeatability while controlling for sex differences in repeatability and using a subsample of 65 individuals recaptured during successive winters (mean  $\pm$  SE =  $2.68 \pm 0.97$  winters, range 2–5). For each individual, the standardized handling scores were averaged for a given winter so as to retain only one measure per individual and per winter. Within-winter and between-winter adjusted repeatabilities were also estimated separately for males and females while controlling for age differences at the short-term scale only. The short-term scale corresponds to around 10% of an individual roe deer's life span, whereas the long-term scale corresponds to between 10% and 45% of an individual's life span (with an average close to 20%).

#### Age and Sex Effects on Repeatability

Finally, to evaluate differences in the repeatability of raw handling scores between the sexes and age classes, repeatability was measured for the four age–sex categories (i.e. juvenile females, juvenile males, adult females and adult males) from a single multivariate model also including the number of previous capture events experienced and the time elapsed since the previous capture event to account for habituation, and the corresponding 95% CI were compared. Individual identity was included as a random effect with an unstructured  $4 \times 4$  (co)variance matrix estimating four variance components, one for each sex and age–class category. A total of 500 000 iterations were run, including a burn-in of 50 000.

#### Within-individual Variation in Docility

To test the hypothesis that within-individual variation in docility was linked to between-individual variation in body mass, we assessed the relationship between the standard deviation of the standardized handling score of a given individual and its average body mass, while controlling for the time elapsed between measurements and the number of observations. In this analysis, a subsample of 105 individuals (for a total of 403 capture events) was used for which body mass and handling scores were recorded at least twice during the study period. We standardized body mass to account for sex and age differences by taking the residuals of the regression between body mass and these two factors (referred to as 'corrected body mass'). The standard deviation of the standardized handling score was square-root transformed to achieve normality. Because the expected relationship between within-individual variation in docility and body mass could be sex dependent due to sex-specific behavioural flexibility, the two-way interaction between sex and corrected body mass was also included in the full model. We considered the model including only the number of observations and the number of days between the first and last observations as fixed effects as our baseline model. We then fitted four more complex candidate linear models to explain variation in the standard deviation of the standardized handling score in relation to (1) mean individual corrected body mass, (2) sex, (3) both these factors and (4) their interaction, as well as the constant model (i.e. including no effect) and all the models without controlling for the number of observations and the number of days between the first and last observations. We then used a model selection procedure as described above to select the best of these 10 models.

## RESULTS

### Variation in Docility

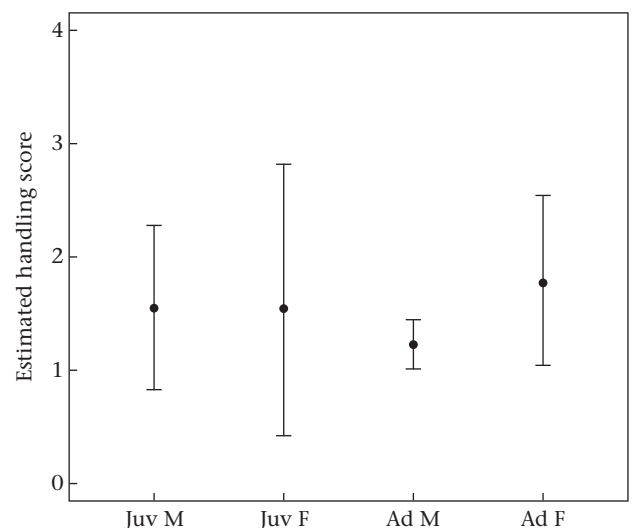
The handling scores varied between 0 and 4 with a mean  $\pm$  SD of  $1.62 \pm 0.95$  for the whole data set (Appendix Fig. A2). Estimated mean docility was between score 1 and 2 with a high standard deviation for all the four age–sex categories (mean docility  $\pm$  SD =  $1.53 \pm 0.90$  for juvenile females,  $1.71 \pm 0.95$  for juvenile males,  $1.68 \pm 0.94$  for adult females and  $1.26 \pm 0.95$  for adult males; see Fig. 2 for the numbers of scores and individuals for each category). The model with the lowest AICc included a two-way interaction between sex and age classes, plus the number of previous capture events experienced, as well as the time elapsed since the previous capture event (Table 2). Overall, females had lower docility (i.e. higher handling score) than males, but this was mainly apparent in adults (Fig. 2).

### Repeatability of Docility

Overall, docility was repeatable ( $N = 129$  individuals for 521 measures, adjusted repeatability when accounting for sex and age class variation:  $R = 0.41$ , 95% CI [0.30, 0.52] and  $R = 0.39$ , 95% CI [0.30, 0.50] using the raw and standardized handling scores, respectively).

### Short- and Long-term Repeatability

Estimates for within-winter adjusted repeatability ( $N = 116$  individuals for 150 individuals/winter and 463 measures, adjusted repeatability when accounting for sex and age classes:  $R = 0.41$ , 95% CI [0.30, 0.51]; using a random effect of winter nested within individual to account for pseudoreplication, because 20 individuals had repeated data for several winters, gave similar results:  $R = 0.43$ ) and between-winter adjusted repeatability ( $N = 65$  individuals for 170 measures; adjusted repeatability when accounting for sex:  $R = 0.31$ , 95% CI [0.13, 0.49]) of docility did not differ



**Figure 2.** Estimated mean handling score and 95% credibility intervals for each sex and age class category from the best multinomial regression. The best model included the two-way interaction between sex and age class and the number of previous capture events for that individual, as well as the time elapsed since the previous capture event, to control for habituation (number of scores and individuals:  $N_{\text{observations}} = 85$ ,  $N_{\text{individuals}} = 31$  juvenile males (Juv M);  $N_{\text{observations}} = 98$ ,  $N_{\text{individuals}} = 35$  juvenile females (Juv F);  $N_{\text{observations}} = 151$ ,  $N_{\text{individuals}} = 42$  adult males (Ad M);  $N_{\text{observations}} = 187$ ,  $N_{\text{individuals}} = 45$  adult females (Ad F)).

**Table 2**  
Performance of the six candidate mixed multinomial models for explaining variation in docility, over the 521 capture events

	<i>k</i>	AICc	ΔAICc	AICcWt
<b>No_catch+Δt_capture+Sex*Age_class+(1 ID)</b>	<b>11</b>	<b>1285.77</b>	<b>0.00</b>	<b>0.76</b>
No_catch+Δt_capture+Sex+(1 id)	9	1288.94	3.17	0.15
No_catch+Δt_capture+Sex+Age_class+(1 ID)	10	1290.69	4.92	0.06
No_catch+Δt_capture+(1 ID)	8	1293.32	7.55	0.02
No_catch+Δt_capture+Age_class+(1 ID)	9	1295.14	9.37	0.01
1+(1 ID)	6	1337.96	52.19	0.00

Individual identity (1|ID) was included in all models as a random factor. The number of captures previously experienced by the individual ('No\_catch') and the time elapsed since the previous capture event ('Δt\_capture') are included as a fixed effect in all models to control for an habituation effect. Models are ranked in increasing order of AICc. ΔAICc corresponds to the difference in AICc between the best model and a given model and AICcWt corresponds to the AIC weight of a given model. *k* refers to the number of model parameters. The selected model is in bold.

markedly (95% CI overlapped). However, when age class variation was accounted for, males tended to have a lower long-term than short-term repeatability (Fig. 3a).

#### Age and Sex Effects on Repeatability

Docility for juvenile females ( $R = 0.38$ , 95% CI [0.18, 0.57]), juvenile males ( $R = 0.60$ , 95% CI [0.36, 0.78]), adult females ( $R = 0.47$ , 95% CI [0.24, 0.66]) and adult males ( $R = 0.55$ , 95% CI [0.35, 0.72]) were all repeatable. Based on overlap of 95% CI, repeatability did not differ markedly between sex and age classes. However, the repeatability of docility among juvenile males tended to be higher than that of juvenile females (see Fig. 3b for the numbers of scores and individuals for each category).

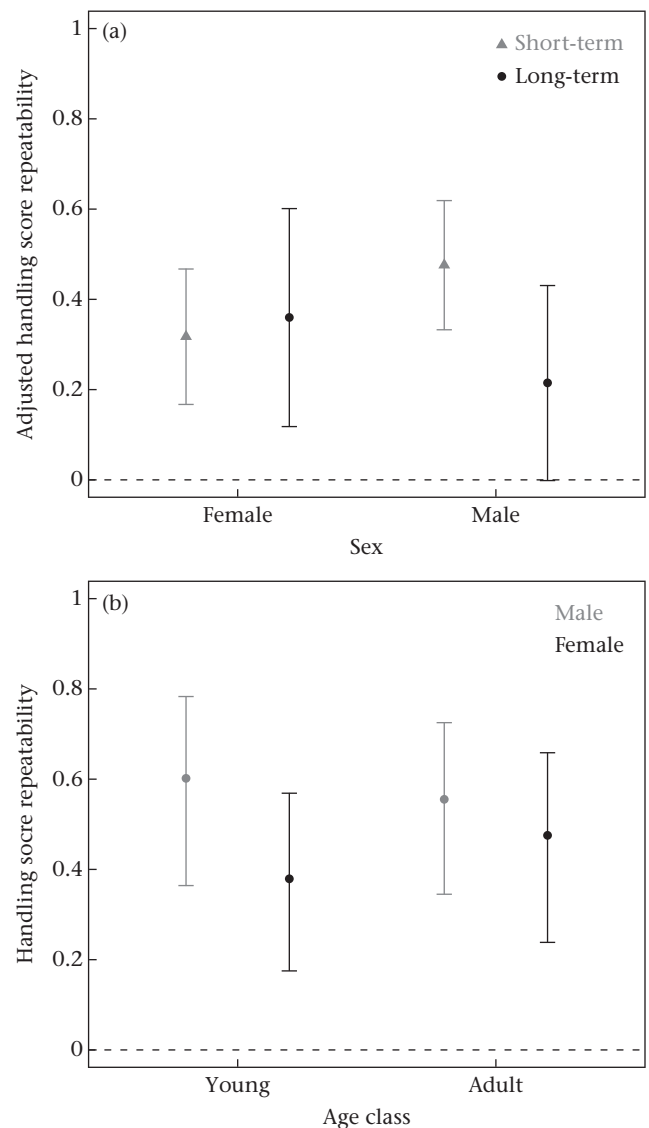
#### Within-individual Variation in Docility

The magnitude of the consistency of docility seemed to vary among individuals. From a subset of six individuals that were recaptured at least 10 times, individual trajectories of docility were indeed highly variable (see Appendix Fig. A3).

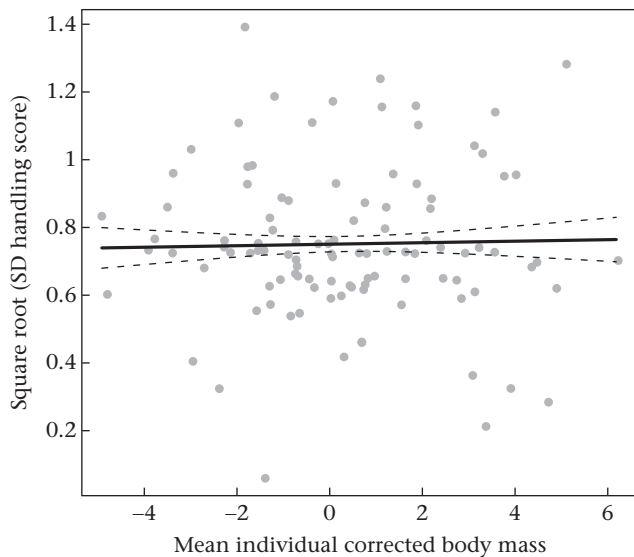
Overall, individual docility was not related to corrected individual body mass (there was no difference between the model including the effect of corrected individual body mass and the basic model correcting for habituation by including the number of previously experienced captures and the time elapsed between the previous capture event; ΔAICc between the two models = 0.01). Moreover, temporal variation in docility (estimated as the standard deviation of standardized handling scores) for a given individual was not related to its corrected body mass (slope ± SE = 0.002 ± 0.011; Fig. 4). Indeed, the best model of variation in the standard deviation of docility was the constant model (Table 3), whereas the model including the fixed effect of individual mean corrected body mass only or coupled with number of observations and time elapsed between consecutive measurements (Table 3), had no support.

## DISCUSSION

We investigated behavioural repeatability in an unusually comprehensive data set composed of a large number of individual roe deer for which behaviour at capture was repeatedly assessed in the wild over a reasonably long time period. Docility, a behavioural trait that we indexed through individual handling scores at capture, did not differ substantially between age classes, but female adults tended to be somewhat less docile than males. Importantly, this behavioural trait was found to be repeatable in this free-ranging population. In comparison, a very high repeatability of docility during capture was reported in a wild population of bighorn sheep ( $R = 0.65$ ; Réale et al., 2000), while a somewhat lower repeatability



**Figure 3.** Repeatability estimates and 95% credibility intervals for (a) the index of docility (standardized handling scores) over the long-term ( $N = 150$  individuals/winter for 463 measures) and short-term ( $N = 65$  individuals for 178 measures) in both sexes (with repeatability adjusted for age differences) and (b) the index of docility (handling scores) for each sex and age class derived from a single model taking into account habituation by including the number of captures previously experienced and the time elapsed since the previous capture event as fixed effects. Juvenile females:  $N = 35$  individuals for 98 measures; juvenile males:  $N = 31$  individuals for 85 measures; adult females:  $N = 45$  individuals for 187 measures; adult males:  $N = 42$  individuals for 151 measures. Juveniles are less than 1 year old; adults are 1.5 years old or more.



**Figure 4.** Relationship between temporal variation in docility for a given individual measured as the standard deviation of its standardized handling scores (square root transformed) in relation to mean individual corrected body mass ( $N = 105$ ). The factors 'number of observations' and 'time elapsed between measurements' were fixed to their mean values. The line represents the estimated relationship and the dashed lines its associated standard error.

was found in yellow-bellied marmots ( $R = 0.25$ ; Petelle et al., 2013). Although scarce, current empirical evidence thus indicates that docility is a repeatable trait in several mammal species (Montiglio, Garant, Pelletier, & Réale, 2012; Petelle et al., 2013; Réale et al., 2000, 2009), and based on some other studies is likely to be part of an overall behavioural syndrome (e.g. Montiglio et al., 2012).

Recently, Bell et al. (2009) highlighted the importance of studying patterns of variation in behavioural repeatability and of exploring the influence of factors that could influence repeatability. In accordance with previous studies (Martin & Réale, 2008a; Van Oers, Klunder, & Drent, 2005), we found that docility increased with the number of behavioural assays (i.e. an habituation effect). However, unexpectedly, we found that short-term and long-term repeatabilities did not differ markedly when accounting for age and sex. In addition, repeatability did not differ markedly between age and sex categories but tended to be higher in juvenile males than in juvenile females. Finally, also contrary to expectation, individual variability in docility was not related to body mass, a reliable measure of phenotypic quality in this species (Toïgo et al., 2006).

Surprisingly, and in contrast to other studies (Bell et al., 2009; David et al., 2012; Gifford et al., 2014), the repeatability estimates of docility over both the short and long term were very similar, although males did tend to have a lower long-term repeatability. In their meta-analysis, Bell et al. (2009) reported higher repeatability for behaviours that were measured closer in time. Interestingly, variation in repeatability in relation to the size of the time window has been shown to depend on the focal behavioural trait (David et al., 2012; Gifford et al., 2014). In a study on captive zebra finches, exploratory tendencies were repeatable over the long term, whereas struggling rate was not, even though it was repeatable over the short term (David et al., 2012). The high long-term repeatability of docility we observed has also been reported in other vertebrate species (Petelle et al., 2013; Réale et al., 2000) and suggests that this behavioural trait may be stable, indexing an ecologically relevant dimension of behaviour that may be affected by natural selection. Indeed, docility may be part of a behavioural syndrome linking exploratory, social and aggressive behaviours, as well as physiological features such as long-term stress reactivity or metabolism (Careau, Thomas, Humphries, & Réale, 2008; Debeffe et al., 2014; Montiglio et al., 2012; Réale et al., 2009; Taylor et al., 2012). Moreover, like other behavioural traits, docility may have a genetic basis (Dingemanse et al., 2004; Poissant, Réale, Martin, Festa-Bianchet, & Coltman, 2013; Réale et al., 2009). Given these features, docility may be affected by natural selection. For example, in bighorn sheep, age-specific selection on docility and a moderate level of selection on boldness were reported under predation pressure (Réale & Festa-Bianchet, 2003). Furthermore, because consistent behavioural traits can affect how individuals react to challenging situations (Réale et al., 2007), they may also influence individual fitness (Dingemanse & Réale, 2005; Smith & Blumstein, 2008). Indeed, docility has been linked to differences in individual fitness components in bighorn sheep, with bold ewes reproducing earlier and having higher weaning success than shy ones (Réale et al., 2000, 2009). In domestic cattle, a similar relationship between docility and fitness components has also been reported (Norris et al., 2014; Turner et al., 2013). Overall, these studies indicate that docility is likely to be an ecologically relevant behaviour shaped by natural selection. This behavioural trait should therefore be recorded in field studies designed to explore the link between repeatable behavioural differences and fitness or its components.

Recent work has highlighted the importance of studying differences in consistency in relation to phenotypic attributes (Bell et al., 2009; Schuett et al., 2010), but how behavioural consistency varies in relation to age has not been intensively explored so

**Table 3**

Model selection for the linear models describing the link between temporal variation in docility for a given individual in relation to mean corrected body mass and sex for 105 roe deer

	$k$	AICc	$\Delta$ AICc	AICcWt
<b>Constant</b>	<b>2</b>	<b>-10.63</b>	<b>0.00</b>	<b>0.34</b>
Sex	3	-10.15	0.48	0.26
Mean corrected body mass	3	-8.55	2.08	0.12
Mean corrected body mass+Sex	4	-8.07	2.56	0.09
No. days+No. observations	4	-7.32	3.31	0.06
Sex+No. days+No. observations	5	-6.57	4.06	0.04
Mean corrected body mass+No. days+No. observations	5	-6.12	4.51	0.04
Mean corrected body mass*Sex	5	-5.16	5.47	0.02
Mean corrected body mass+Sex+No. days+No. observations	6	-4.33	6.30	0.02
Mean corrected body mass*Sex+No. days+No. observations	7	-2.26	8.37	0.01

Within-individual variation in docility for a given individual was measured as the standard deviation of the standardized handling score. To control for the number of observations (No. observations) and the time elapsed between measurements (No. days) (as our baseline model), these two factors were added as fixed effects. The selected model is in bold.

far. Indeed, most previous studies of variation in repeatability of behavioural traits have focused on a single life stage (Bell et al., 2009; Stamps & Groothuis, 2010). We found that docility was a repeatable behaviour in all age–sex classes in roe deer. This result is in line with findings in yellow-bellied marmots, in which docility was repeatable irrespective of age, as well as strongly correlated across ages (Petelle et al., 2013), in firebugs, *Pyrrhocoris apterus* (Gyuris, Feró, & Barta, 2012) and lake frogs, *Rana ridibunda* (Wilson & Krause, 2012) in which behaviours were repeatable within life stages. However, we found that the level of repeatability for docility in roe deer was slightly higher in juvenile males than in other sex and age class categories, providing no support for the hypothesis that maturity shapes behavioural consistency. Bell et al. (2009) reported no overall difference in behavioural repeatability between juveniles and adults, but some studies they reviewed also indicated that behaviour was consistently more repeatable in juveniles than in adults. More generally, age-related variation in repeatability seems to depend on the behavioural trait and the species considered.

Sex differences in repeatability generally occur in adults (Bell et al., 2009) so that, when excluding mate preference behaviours, females express more consistent behaviours than males (but see Schuett & Dall, 2009 on zebra finches, and Jenkins, 2011 on kangaroo rats). However, this pattern relies on those few studies that provided sex-specific estimates of repeatability, and among those, the pattern is very variable. Among juveniles, the cost of deviating from a developmental trajectory is higher for males, potentially linking consistency to sexual selection (Biro & Stamps, 2008). The juvenile stage is a key phase in the development of individual behaviour. For instance, a strong correlation of docility across ages, with docility developing mainly during the juvenile stage and then remaining stable, was reported in yellow-bellied marmots (Petelle et al., 2013). In sexually dimorphic species, young males are known to differ from females in both their growth rate and behaviour early in life (Andersson, 1994), potentially leading to the observed tendency for a higher consistency in young roe deer males.

Individuals may differ substantially in their behavioural consistency (Bell et al., 2009) such that certain individuals may be able to adjust their behaviour to varying environmental conditions better than others (Dingemanse, Kazem, Réale, & Wright, 2010). In this study, we found that within-individual variation in docility was unrelated to individual body mass, a reliable indicator of phenotypic quality in most vertebrates, including roe deer (Toïgo et al., 2006). Although it has repeatedly been suggested that behavioural flexibility should entail costs (DeWitt et al., 1998; Niemela et al., 2013), these costs have only rarely been documented in wild populations (Auld, Agrawal, & Relyea, 2010; Lind & Johansson, 2009), and, indeed, some studies failed to identify any costs (Scheiner & Berrigan, 1998; Van Buskirk & Steiner, 2009; Van Kleunen, Fischer, & Schmid, 2000). In this context, our results suggest that there is some plasticity in docility in roe deer, but that the degree of this plasticity is unrelated to individual phenotypic quality, indicating either particularly low (with all individuals being similarly flexible) or high (with all individuals being similarly consistent) costs of behavioural flexibility.

The factors shaping between-individual variation in consistency have received little attention to date (Rodríguez-Prieto, Martin, & Fernandez-Juricic, 2011) and require further investigation (Dingemanse et al., 2010; Sih et al., 2004). Notably, in the context of rapid environmental modifications due to human disturbance and global change, the ability of individuals of a given population to adapt and flourish may depend on their ability to cope with

fluctuations of their environment through the expression of behavioural flexibility. Hence, as individual differences in behavioural flexibility might have consequences for ecological and evolutionary processes (Dingemanse & Wolf, 2013), we need more studies to evaluate how predictable different behaviours are and which factors influence their level of consistency.

## Acknowledgments

C.V. and P.K. are both last coauthors: C.V. obtained the grant that financed L.D. and J.F.L. and contributed substantially to the research project, P.K. designed the study and organized the long-term data collection. C.V., L.D. and J.F.L. were funded by the PATCH RPDO ANR project (ANR-12-PDOC-0017-01) awarded to C.V. from the French National Research Agency. This study was supported by the PATCH RPDO ANR project. The field study was supported by grants from the Swedish Association for Hunting and Wildlife Management (P.K.), Marie-Claire Cronstedts Stiftelse and the Swedish Environmental Protection Agency (U.A.B., P.K.). We thank G. Sjöo and L. Höglund for their assistance during roe deer capture. We thank the editor and referees for constructive comments on the manuscript.

## References

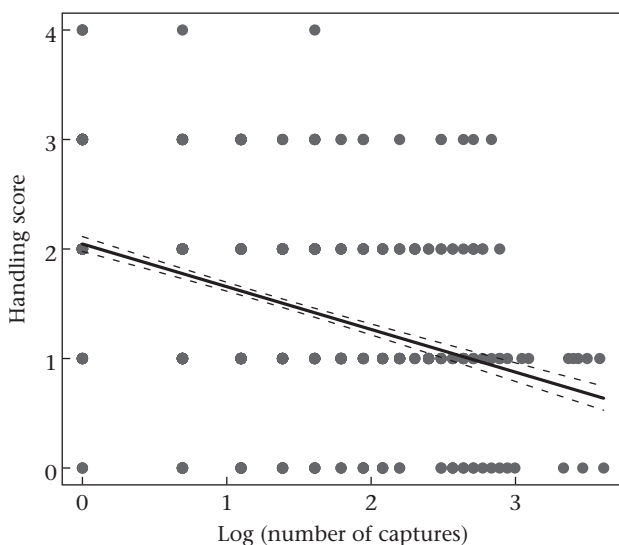
- Andersen, R., Duncan, P., & Linnell, J. D. C. (1998). *The European roe deer: The biology of success*. Oslo, Norway: Scandinavian University Press.
- Andersson, M. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.
- Auld, J. R., Agrawal, A. A., & Relyea, R. A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B: Biological Sciences*, 277(1681), 503–511.
- Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, 274(1611), 755–761. <http://dx.doi.org/10.1098/rspb.2006.0199>.
- Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, 77(4), 771–783.
- Benhajali, H., Boivin, X., Sapa, J., Pellegrini, P., Boulesteix, P., Lajudie, P., et al. (2010). Assessment of different on-farm measures of beef cattle temperament for use in genetic evaluation. *Journal of Animal Science*, 88(11), 3529–3537. <http://dx.doi.org/10.2527/jas.2010-3132>.
- Biro, P. A., & Stamps, J. A. (2008). Are animal personality traits linked to life-history productivity? *Trends in Ecology & Evolution*, 23(7), 361–368. <http://dx.doi.org/10.1016/j.tree.2008.04.003>.
- Briffa, M. (2013). Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, 9(5). <http://dx.doi.org/10.1098/rsbl.2013.0592>, 20130592.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference—A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.
- Careau, V., Thomas, D., Humphries, M. M., & Réale, D. (2008). Energy metabolism and animal personality. *Oikos*, 117(5), 641–653.
- Cederlund, G., Kjellander, P., & Stålfelt, F. (1991). *Age determination of roe deer by tooth wear and cementum layers—tests with known age material*. (pp. 540–545). Proceedings of the XXth congress of the International Union of Game Biologists, Gödöllő, Hungary, 21–26 August.
- Charmantier, A., Perrins, C., McCleery, R. H., & Sheldon, B. C. (2006). Age-dependent genetic variance in a life-history trait in the mute swan. *Proceedings of the Royal Society B: Biological Sciences*, 273(1583), 225–232. <http://dx.doi.org/10.1098/rspb.2005.3294>.
- Christensen, R. H. B. (2015). *Ordinal—regression models for ordinal data*. R Package Version 2015.1-21 <http://www.cran.r-project.org/package=ordinal/>.
- Cote, J., Dreiss, A., & Clobert, J. (2008). Social personality trait and fitness. *Proceedings of the Royal Society B: Biological Sciences*, 275(1653), 2851–2858. <http://dx.doi.org/10.1098/rspb.2008.0783>.
- David, M., Auclair, Y., & Cezilly, F. (2012). Assessing short- and long-term repeatability and stability of personality in captive zebra finches using longitudinal data. *Ethology*, 118(10), 932–942. <http://dx.doi.org/10.1111/j.1439-0310.2012.02085.x>.
- David, M., & Giraldeau, L.-A. (2012). Zebra finches in poor condition produce more and consume more food in a producer-scrounger game. *Behavioral Ecology*, 23(1), 174–180. <http://dx.doi.org/10.1093/beheco/arr171>.
- David, M., Le Hö, M., Laskowski, K., Salignon, M., Gillingham, M. A. F., & Giraldeau, L.-A. (2014). Individual differences in behavioral consistency are related to sequential access to resources and body condition in a producer-scrounger game. *Frontiers in Ecology and Evolution*, 2(19). <http://dx.doi.org/10.3389/fevo.2014.00019>.



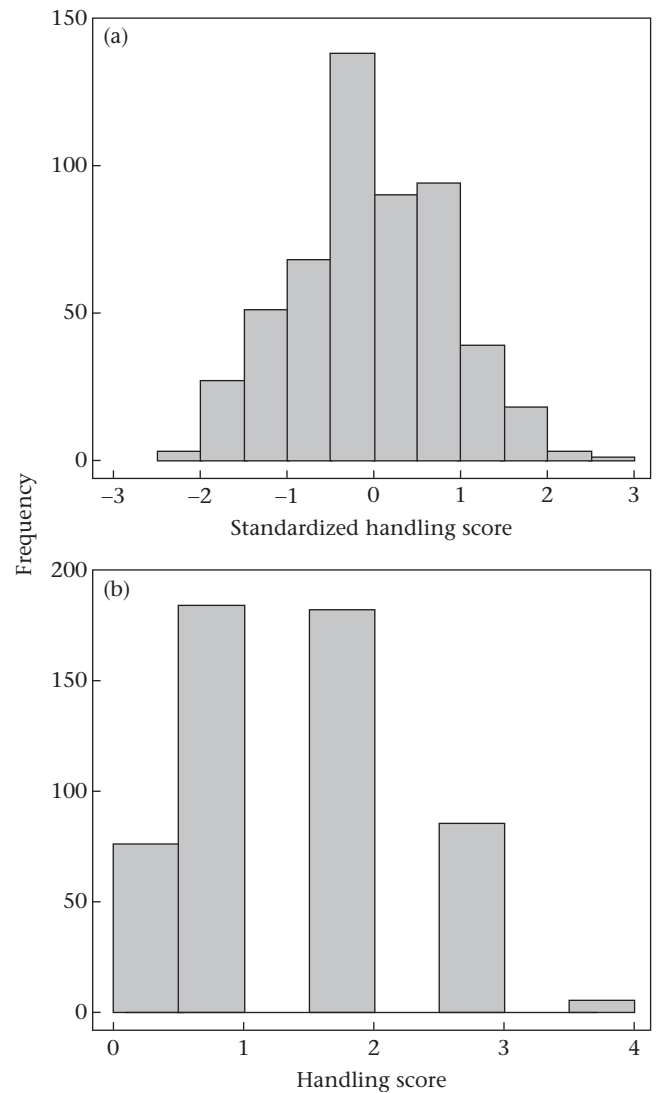
- Debeffe, L., Morellet, N., Bonnot, N., Gaillard, J. M., Cargnelutti, B., Verheyden-Tixier, H., et al. (2014). The link between behavioural type and natal dispersal propensity reveals a dispersal syndrome in a large herbivore. *Proceedings of the Royal Society B: Biological Sciences*, 281(1790). <http://dx.doi.org/10.1098/rspb.2014.0873>, 20140873.
- De Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., & Vehrencamp, S. L. (2009). Trill consistency is an age-related assessment signal in banded wrens. *Proceedings of the Royal Society B: Biological Sciences*, 276(1665), 2315–2321. <http://dx.doi.org/10.1098/rspb.2009.0127>.
- DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, 13(2), 77–81. [http://dx.doi.org/10.1016/S0169-5347\(97\)01274-3](http://dx.doi.org/10.1016/S0169-5347(97)01274-3).
- Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, 271(1541), 847–852. <http://dx.doi.org/10.1098/rspb.2004.2680>.
- Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, 25(2), 81–89.
- Dingemanse, N. J., & Réale, D. (2005). Natural selection and animal personality. *Behaviour*, 142(9–10), 1159–1184. <http://dx.doi.org/10.1163/156853905774539445>.
- Dingemanse, N. J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour*, 85(5), 1031–1039. <http://dx.doi.org/10.1016/j.anbehav.2012.12.032>.
- Ferrari, C., Pasquaretta, C., Carere, C., Cavallone, E., von Hardenberg, A., & Réale, D. (2013). Testing for the presence of coping styles in a wild mammal. *Animal Behaviour*, 85(6), 1385–1396. <http://dx.doi.org/10.1016/j.anbehav.2013.03.030>.
- Garamszegi, L. Z., Eens, M., & Janos, T. (2009). Behavioural syndromes and trappability in free-living collared flycatchers, *Ficedula albicollis*. *Animal Behaviour*, 77(4), 803–812. <http://dx.doi.org/10.1016/j.anbehav.2008.12.012>.
- Gifford, M. E., Clay, T. A., & Careau, V. (2014). Individual (co) variation in standard metabolic rate, feeding rate, and exploratory behavior in wild-caught semi-aquatic salamanders. *Physiological and Biochemical Zoology*, 87(3), 384–396. <http://dx.doi.org/10.1086/675974>.
- Gyuris, E., Feró, O., & Barta, Z. (2012). Personality traits across ontogeny in firebugs, *Pyrrochoris apterus*. *Animal Behaviour*, 84(1), 103–109. <http://dx.doi.org/10.1016/j.anbehav.2012.04.014>.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22.
- Hayes, J. P., & Jenkins, S. H. (1997). Individual variation in mammals. *Journal of Mammalogy*, 78(2), 274–293. <http://dx.doi.org/10.2307/1382882>.
- Hewison, A. J. M., Gaillard, J. M., Delorme, D., van Laere, G., Amblard, T., & Klein, F. (2011). Reproductive constraints, not environmental conditions, shape the ontogeny of sex-specific mass-size allometry in roe deer. *Oikos*, 120(8), 1217–1226.
- Hewison, A. J. M., Vincent, J. P., Joachim, J., Angibault, J. M., Cargnelutti, B., & Cibien, C. (2001). The effects of woodland fragmentation and human activity on roe deer distribution in agricultural landscapes. *Canadian Journal of Zoology*, 79, 679–689.
- Jenkins, S. H. (2011). Sex differences in repeatability of food-hoarding behaviour of kangaroo rats. *Animal Behaviour*, 81(6), 1155–1162. <http://dx.doi.org/10.1016/j.anbehav.2011.02.021>.
- Jepsen, J. U., & Topping, C. J. (2004). Modelling roe deer (*Capreolus capreolus*) in a gradient of forest fragmentation: behavioural plasticity and choice of cover. *Canadian Journal of Zoology*, 82(9), 1528–1541.
- Kjellander, P., Gaillard, J. M., & Hewison, A. J. M. (2006). Density-dependent responses of fawn cohort body mass in two contrasting roe deer populations. *Oecologia*, 146(4), 521–530. <http://dx.doi.org/10.1007/s00442-005-0188-z>.
- Le Neindre, P., Boivin, X., & Boissy, A. (1996). Handling of extensively kept animals. *Behaviour and Welfare of Extensively Farmed Animals*, 49(1), 73–81. [http://dx.doi.org/10.1016/0168-1591\(95\)00669-9](http://dx.doi.org/10.1016/0168-1591(95)00669-9).
- Le Neindre, P., Trillat, G., Sapa, J., Menissier, F., Bonnet, J., & Chupin, J. (1995). Individual-differences in docility in limousin cattle. *Journal of Animal Science*, 73(8), 2249–2253.
- Liberg, O., Johansson, A., Andersen, R., & Linnell, J. D. C. (1998). Mating system, mating tactics and the function of male territory in roe deer. In R. Andersen, P. Duncan, & J. Linnell (Eds.), *The European roe deer: The biology of success* (pp. 221–256). Oslo, Norway: Scandinavian University Press.
- Lind, M. I., & Johansson, F. (2009). Costs and limits of phenotypic plasticity in island populations of the common frog *Rana temporaria* under divergent selection pressures. *Evolution*, 63(6), 1508–1518. <http://dx.doi.org/10.1111/j.1558-5646.2009.00647.x>.
- Martin, J. G. A., & Réale, D. (2008a). Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour*, 75, 309–318. <http://dx.doi.org/10.1016/j.anbehav.2007.05.026>.
- Martin, J. G. A., & Réale, D. (2008b). Animal temperament and human disturbance: Implications for the response of wildlife to tourism. *Behavioural Processes*, 77(1), 66–72.
- Mazurek, M., McGee, M., Crowe, M. A., Prendiville, D. J., Boivin, X., & Earley, B. (2011). Consistency and stability of behavioural fear responses of heifers to different fear-eliciting situations involving humans. *Applied Animal Behaviour Science*, 131(1–2), 21–28. <http://dx.doi.org/10.1016/j.applanim.2011.01.004>.
- Montiglio, P.-O., Garant, D., Pelletier, F., & Réale, D. (2012). Personality differences are related to long-term stress reactivity in a population of wild eastern chipmunks, *Tamias striatus*. *Animal Behaviour*, 84(4), 1071–1079. <http://dx.doi.org/10.1016/j.anbehav.2012.08.010>.
- Montiglio, P.-O., Garant, D., Thomas, D., & Réale, D. (2010). Individual variation in temporal activity patterns in open-field tests. *Animal Behaviour*, 80(5), 905–912. <http://dx.doi.org/10.1016/j.anbehav.2010.08.014>.
- Nakagawa, S., Gillespie, D. O. S., Hatchwell, B. J., & Burke, T. (2007). Predictable males and unpredictable females: sex difference in repeatability of parental care in a wild bird population. *Journal of Evolutionary Biology*, 20(5), 1674–1681. <http://dx.doi.org/10.1111/j.1420-9101.2007.01403.x>.
- Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85(4), 935–956. <http://dx.doi.org/10.1111/j.1469-185X.2010.00141.x>.
- Niemela, P. T., Vainikka, A., Forsman, J. T., Loukola, O. J., & Kortet, R. (2013). How does variation in the environment and individual cognition explain the existence of consistent behavioural differences? *Ecology and Evolution*, 3(2), 457–464. <http://dx.doi.org/10.1002/ece3.451>.
- Norris, D., Ngambi, J. W., Mabelebele, M., Alabi, O. J., & Benyi, K. (2014). Genetic selection for docility: a review. *Journal of Animal and Plant Sciences*, 24(2), 374–379.
- Petelle, M. B., McCoy, D. E., Alejandro, V., Martin, J. G. A., & Blumstein, D. T. (2013). Development of boldness and docility in yellow-bellied marmots. *Animal Behaviour*, 86(6), 1147–1154. <http://dx.doi.org/10.1016/j.anbehav.2013.09.016>.
- Poissant, J., Réale, D., Martin, J. G. A., Festa-Bianchet, M., & Coltman, D. W. (2013). A quantitative trait locus analysis of personality in wild bighorn sheep. *Ecology and Evolution*, 3(3), 474–481. <http://dx.doi.org/10.1002/ece3.468>.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Réale, D., & Festa-Bianchet, M. (2003). Predator-induced natural selection on temperament in bighorn ewes. *Animal Behaviour*, 65, 463–470. <http://dx.doi.org/10.1006/anbe.2003.2100>.
- Réale, D., Gallant, B. Y., Leblanc, M., & Festa-Bianchet, M. (2000). Consistency of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour*, 60, 589–597. <http://dx.doi.org/10.1006/anbe.2000.1530>.
- Réale, D., Martin, J., Coltman, D. W., Poissant, J., & Festa-Bianchet, M. (2009). Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology*, 22(8), 1599–1607.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, 82(2), 291–318. <http://dx.doi.org/10.1111/j.1469-185X.2007.00010.x>.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3–25. <http://dx.doi.org/10.1037/0033-2909.126.1.3>.
- Rodriguez-Prieto, I., Martin, J., & Fernandez-Juricic, E. (2011). Individual variation in behavioural plasticity: direct and indirect effects of boldness, exploration and sociability on habituation to predators in lizards. *Proceedings of the Royal Society B: Biological Sciences*, 278(1703), 266–273.
- Scheiner, S. M., & Berrigan, D. (1998). The genetics of phenotypic plasticity. VIII. The cost of plasticity in *Daphnia pulex*. *Evolution*, 52(2), 368–378. <http://dx.doi.org/10.2307/2411074>.
- Schuett, W., & Dall, S. R. X. (2009). Sex differences, social context and personality in zebra finches, *Taeniopygia guttata*. *Animal Behaviour*, 77(5), 1041–1050.
- Schuett, W., Godin, J.-G. J., & Dall, S. R. X. (2011). Do female zebra finches, *Taeniopygia guttata*, choose their mates based on their "personality"? *Ethology*, 117(10), 908–917. <http://dx.doi.org/10.1111/j.1439-0310.2011.01945.x>.
- Schuett, W., Tregenza, T., & Dall, S. R. X. (2010). Sexual selection and animal personality. *Biological Reviews*, 85(2), 217–246. <http://dx.doi.org/10.1111/j.1469-185X.2009.00101.x>.
- Sempéré, A. J., Mauget, R., & Mauget, C. (1998). Reproductive physiology of roe deer. In R. Andersen, P. Duncan, & J. Linnell (Eds.), *The European roe deer: The biology of success* (pp. 161–188). Oslo, Norway: Scandinavian University Press.
- Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, 19(7), 372–378. <http://dx.doi.org/10.1016/j.tree.2004.04.009>.
- Sih, A., Cote, J., Evans, M., Fogarty, S., & Pruitt, J. (2012). Ecological implications of behavioural syndromes. *Ecology Letters*, 15(3), 278–289. <http://dx.doi.org/10.1111/j.1461-0248.2011.01731.x>.
- Smith, B. R., & Blumstein, D. T. (2008). Fitness consequences of personality: a meta-analysis. *Behavioral Ecology*, 19(2), 448–455.
- Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, 83(6), 1325–1334. <http://dx.doi.org/10.1016/j.anbehav.2012.02.017>.
- Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*, 85(2), 301–325. <http://dx.doi.org/10.1111/j.1469-185X.2009.00103.x>.
- Taylor, R. W., Boon, A. K., Dantzer, B., Réale, D., Humphries, M. M., Boutin, S., et al. (2012). Low heritabilities, but genetic and maternal correlations between red

- squirrel behaviours. *Journal of Evolutionary Biology*, 25(4), 614–624. <http://dx.doi.org/10.1111/j.1420-9101.2012.02456.x>.
- Toigo, C., Gaillard, J. M., Van Laere, G., Hewison, M., & Morellet, N. (2006). How does environmental variation influence body mass, body size, and body condition? Roe deer as a case study. *Ecography*, 29(3), 301–308.
- Turner, S. P., Jack, M. C., & Lawrence, A. B. (2013). Precalving temperament and maternal defensiveness are independent traits but precalving fear may impact calf growth. *Journal of Animal Science*, 91(9), 4417–4425. <http://dx.doi.org/10.2527/jas2012-5707>.
- Van Buskirk, J., & Steiner, U. K. (2009). The fitness costs of developmental canalization and plasticity. *Journal of Evolutionary Biology*, 22(4), 852–860. <http://dx.doi.org/10.1111/j.1420-9101.2009.01685.x>.
- Van Kleunen, M., Fischer, M., & Schmid, B. (2000). Costs of plasticity in foraging characteristics of the clonal plant *Ranunculus reptans*. *Evolution*, 54(6), 1947–1955. [http://doi.org/10.1554/0014-3820\(2000\)054\[1947:COPIFC\]2.0.CO;2](http://doi.org/10.1554/0014-3820(2000)054[1947:COPIFC]2.0.CO;2).
- Van Oers, K., Klunder, M., & Drent, P. J. (2005). Context dependence of personalities: risk-taking behavior in a social and a nonsocial situation. *Behavioral Ecology*, 16(4), 716–723.
- Veiberg, V., Myrsetrud, A., Gaillard, J.-M., Delorme, D., Laere, G. V., & Klein, F. (2007). Bigger teeth for longer life? Longevity and molar height in two roe deer populations. *Biology Letters*, 3(3), 268–270. <http://dx.doi.org/10.1098/rsbl.2006.0610>.
- Wilson, A. D. M., & Krause, J. (2012). Personality and metamorphosis: is behavioral variation consistent across ontogenetic niche shifts? *Behavioral Ecology*, 23(6), 1316–1323. <http://dx.doi.org/10.1093/beheco/ars123>.

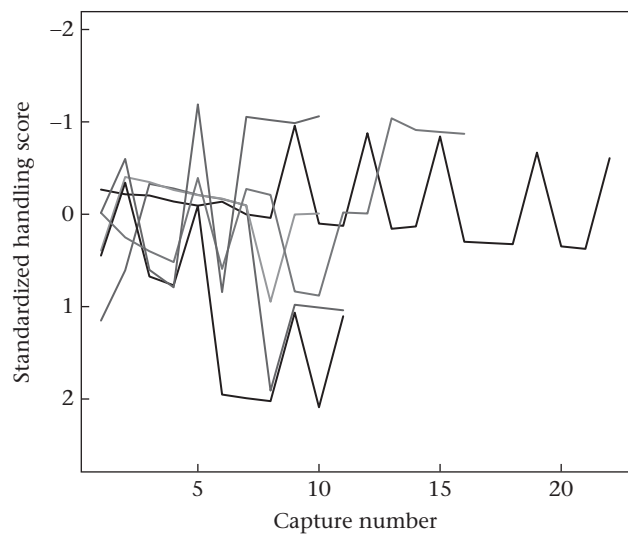
## Appendix



**Figure A1.** Linear regression between handling score (indexing docility, where low values indicate high docility) and the log-transformed number of captures using all available individuals ( $N_{\text{captures}} = 532$ ,  $N_{\text{individuals}} = 130$ ). The line represents the estimated slope of the relationship which was used to calculate the standardized handling score, and the dashed lines its associated standard error.



**Figure A2.** (a) Distribution of the handling score (to index docility) corrected for habituation, i.e. a negative link between reaction to handling during capture and the number of captures previously experienced by the individual. Note that the distribution is not normal: normality test of Shapiro–Wilk:  $W = 0.99$ ,  $P = 0.004$ . (b) Distribution of the raw handling score (normality test of Shapiro–Wilk:  $W = 0.89$ ;  $P < 0.001$ ). Note that low values indicate high docility.



**Figure A3.** Individual trajectories for estimates of docility (i.e. standardized handling score) across capture events for the six roe deer individuals that were captured at least 10 times. Each line corresponds to a given individual. Note that low values indicate high docility.