NTNU
Norwegian University of
Science and Technology

Personality and pace-of-life behavioral syndromes in a model species, the House Sparrow (*Passer domesticus*)

Mette Helene Finnøen

MSc in Biology
Submission date: May 2016
Supervisor: Jonathan Wright, IBI

Norwegian University of Science and Technology
Department of Biology

Acknowledgements
First and most importantly, I would like to thank my supervisor, Professor Jonathan Wright, for being a huge pain in my ass, and of course for being there every step of the way. I would also like to thank Mirja Olsen for keeping me moderately sane, and John Hammerås for his help when 'R' was being impossible. A big thanks to Bernt Rønning for surviving five weeks in the field with us girls, and to all the sparrow guys at CBD for their help with the data collection, and analysis. I'd also like to thank Monica Berdal for all her help with structural equation modelling, and Henrik Jensen for constructive comments. And lastly, thank you Øystein B. Svendsen for all your support!

Sammendrag
Konsekvente forskjeller i atferd mellom individer ("animal personality") kan være ett resultat av naturlig seleksjon. Adaptive forskjeller i livshistorie- og fysiologi-parametere, som utgjør et 'pace-of-life' atferdssyndrom (PoLS), kan drive disse konsekvente forskjellene. Vi har testet denne hypotesen samt muligheten for andre atferdssyndromer ved bruk av individuelle atferdstester av 198 gråspurv (*Passer domesticus*) i fangenskap. Individuelle forskjeller i de ulike atferdene (aktivitet i nye omgivelser og i nærheten av ett nytt objekt og en ny potensiell matressurs) ble sammenliknet med morfologiske data og basalmetabolisme (BMR). Variasjon blant individene i de målte atferdene og BMR ble testet ved bruk av 'univariate tests', og kovariasjon mellom atferdene, BMR og morfologiske variabler ble testet ved bruk av 'structural equation modeling'. Aktivitetsnivå var konsekvent over tid og på tvers av sammenhenger, og all variasjon i atferd var drevet av ett aktivitets-syndrom. Dette kan bety at våre atferdsassay ikke målte noen betydningsfull variasjon i 'neophilia' og 'foraging innovation'. Det var ingen konsekvente forskjeller mellom individer i distanse til nytt objekt eller ny matressurs, eller i BMR. Det var heller ingen kjønnsforskjeller i atferd, og det var ingen link mellom metabolisme og 'personality' i aktivitet. Vi kan ikke konkludere at det ikke er ett PoLS i gråspurv da disse populasjonene ble forsket på på vinteren, og på grunn av at BMR-baserte eksperimenter har ført til lite pålitelige informasjon angåeende deres livshistorieparametere (f.eks. reproduksjonsrate). Men de fleste fenotypiske kovariasjoner som antas i PoLS konseptet var ikke å finne i disse gråspurvpopulasjonene. Det var ett status-syndrom i hannfuglene som bestod av kovariasjon mellom fjærdrakt-trekk og alder som var forbundet med individuell hann-kvalitet. Her var det også overraskende få linker mellom status, individuell aktivitetsnivå og/eller BMR. Dette indikerer at den individuelle variasjonen i atferd og BMR kan forklares av "within-individual plasticity", i stedet for "personality" eller ett PoLS.

Abstract
Consistent individual differences in behaviour (“animal personality”) may be favoured by natural selection because they are driven by adaptive differences in life-history and physiology traits, constituting a pace-of-life syndrome (PoLS). Here we test this hypothesis as well as other behavioural syndrome structures using individual behavioural assays of 198 house sparrows (*Passer domesticus*) in captivity, and connecting individual levels of behaviour (activity in a novel environment and with a novel object and novel food) with morphological and basal metabolic rate (BMR) data. Variation between individuals in the measured behaviours and BMR, and the covariance pattern of behaviours, BMR and morphological measures were quantified using univariate tests and structural equation modelling, respectively. Activity level was the only consistent behaviour across time and contexts, and all behavioural variation was driven by an 'activity' syndrome, suggesting that our assays failed to capture any meaningful variation in neophilia or foraging innovation. There was no individual consistency in distance to the novel object or food, or in BMR. There were no sex differences in behaviour, and there was no obvious link between metabolism and personality in 'activity'. We cannot necessarily conclude from this that there is no PoLS in house sparrows, because these populations were studied only in the winter and due to BMR-based experiments there is limited reliable information on the details of their life-history traits (i.e., rates of reproduction). However, many of the most obvious phenotypic covariances hypothesised by the PoLS do not seem to be present here. There was a male-specific 'status' syndrome involving positive covariation of plumage traits and age that were related to individual quality. Again, there were surprisingly few links between 'status', individual male 'activity' behaviour and/or BMR. This suggests that the individual variation in behaviour (and BMR) measured in our sparrows is largely explained by within-individual plasticity, rather than personality or a wider pace-of-life syndrome structure.

Table of Contents

Introduction

Individual phenotypic variation is a common observation in many animal species. However it is only recently that consistent individual differences in behaviour have become a focus of research interest (see Reale et al. 2010a). Various terms have been used to describe this variation, such as 'animal personality', 'behavioural syndromes', 'coping styles' and 'temperament'. Here, I will use 'animal personality' to refer to consistent individual differences in a single behaviour within a population. Thus, animal personality, and behavioural syndromes, are features of a population, not individuals. A 'behavioural type' (e.g. 'shy' or 'bold') is a feature of an individual, as it refers to the suite of behaviours that an individual can express (Bell 2007). A 'behavioural syndrome' is used here for when one behaviour correlates across different contexts, and/or when it correlates with a different behaviour. Two well-studied examples are the correlation between aggression and boldness (Dingemanse et al. 2007), and the behavioural syndrome that involves superficial exploration, boldness and activity (Sih et al. 2004, Sih and Bell 2008, Reale et al. 2009, Reale et al. 2010a). A common misconception is that personality excludes plasticity (Dingemanse et al. 2010b). However, individuals can be consistently different in their average level of a behaviour, and at the same time be plastic across different contexts (Sih et al. 2004).

Studying and understanding animal personality is important given that ecological processes, such as niche expansion, dispersal and social organization are likely affected by it (Reale et al. 2007). It is difficult to understand why we see such consistent individual differences in behaviour when, as behavioural ecologists, we expect that any behaviour should have evolved to an adaptive optimum over time (Davies et al. 2012). Behavioural syndromes exist in a variety of taxa, and recent studies suggest that they may be the product of adaptive evolution. This predicts that behavioural syndromes are a result of natural selection favouring optimal trait combinations (Careau and Garland 2012), as was in fact shown in stickleback populations in response to variation in predation threat by Dingemanse et al. (2007). Despite such apparently adaptive patterns, we currently lack a general evolutionary explanation for animal personality and behavioural syndromes (Reale et al. 2010a).

One concept that has emerged in the last couple of decades, which is associated with animal personalities, is the pace-of-life syndrome (PoLS). PoLS is conceptually similar to MacArthur and Wilson’s (1967) r- and k-selection as applied to

different species (Careau and Garland 2012). PoLSs are essentially an expansion of the fast-slow life history continuum, which includes life history traits, such as life span, age at maturation and reproductive rates, as well as behaviours such as aggression and boldness under predation threat (Reale et al. 2010b). This continuum is based on the observation of covariation between life-history traits observed in most species, and is a result of the classic trade-offs between current and future reproduction. Along this continuum we find (1) fast-living individuals who have a short life span, mature early and have a high reproductive rate, (2) slow-living individuals who have long life spans, mature late and have low reproductive rates, and (3) individuals that are somewhere between the two extremes (Reale et al. 2010b). The concept was extended when physiological (morphological) traits, such as basal metabolic rate (BMR) and body size, were added to the concept (Hennemann III 1983, Jones et al. 2008). A fast-living individual is then expected to have a high metabolic rate, while a slow-living individual should have a low metabolic rate. Finally, Reale et al. (2010b) proposed that behaviour is best studied using an integrative approach, such as PoLS. Furthermore, they suggested that consistent individual differences in animal behaviour are highly relevant as an outcome of the evolution of PoLSs (Reale et al. 2010b). Using the current PoLS concept, consisting of behaviour, physiological/morphological and life history traits, then a fast-living individual is expected to be bold, aggressive, superficially exploring, highly active and not very social, and conversely, a slow-living individual is expected to be the complete opposite (Reale et al. 2010b).

In this project, I am interested in whether a pace-of-life syndrome (PoLS) or other behavioural syndromes exist in house sparrow (*Passer domesticus*) populations on the islands of Vikna, Lauvøya and Leka in Northern Norway. This was investigated using individual behavioural assays in captivity, and relating individual levels of behaviour to morphological and BMR data.

Because activity, exploration, and aggressiveness are energetically costly behaviours, personality and metabolism should be correlated and physiological constraints may underlie behavioural syndromes (Careau et al. 2008). I also expect BMR to correlate with body mass, as this relationship has been repeatedly shown in the literature (Careau et al. 2008, Wiersma et al. 2007). Activity may to be positively correlated with BMR if house sparrows follow the predictions of the performance model (Fig. 1a), which suggests that an individual has a certain metabolic rate that

determines the amount of available energy, thus a high metabolic rate will equal a high level of activity. However, BMR may correlate negatively with activity if the sparrows follow the predictions of the allocation model (Fig. 1b), in which case the individual has a certain amount of energy available that has to be allocated to both its metabolism and activity, thus a high metabolic rate will equal a low level of activity. Albeit fig. 1 depicts a relationship between activity and resting metabolic rate (RMR) not BMR, the difference between RMR and BMR is so small that the two terms are used interchangeably (Careau et al. 2008).

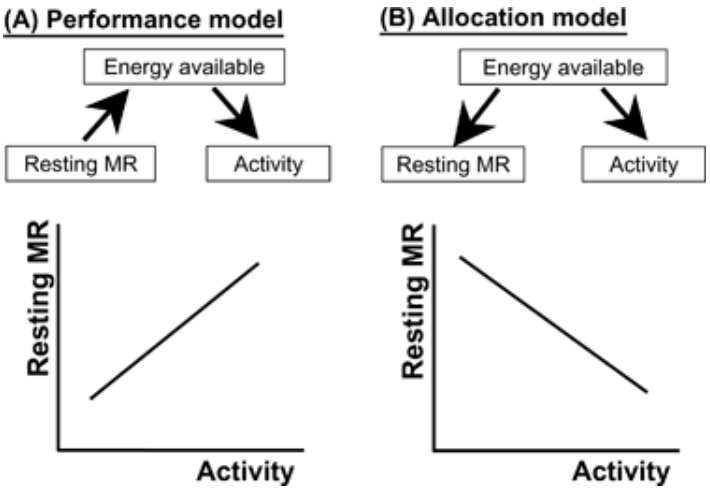

Fig. 1. The hypothetical relationships possible between metabolism and activity, adapted from Careau et al. (2008), where the performance model (A) predicts a positive relationship, and the allocation model (B) predicts a negative relationship.

Despite the fact that house sparrow badge size (the area of black feathers on the breast of males) has been extensively studied, no consensus has emerged regarding its precise function and evolution, and many research findings appear contradictory (Anderson 2006). Badge size is therefore one of the morphological variables that needs to be considered in a separate pace-of-life syndrome focused on males. Specifically, the relationship between badge size and BMR was explored. This is based on studies that have determined a positive relationship between testosterone and BMR (Buchanan et al. 2001), and between testosterone and badge size (Evans et al. 2000). One problem regarding this line of logic is that other studies have found either a negative relationship between testosterone and BMR (Wikelski et

al. 1999), or no relationship between testosterone and BMR (Buttemer et al. 2008). It was therefore of interest to explore such a relationship between BMR and badge size, and the sign of any correlation. If BMR is positively related to testosterone, then perhaps it is also related to aggression. A positive correlation between testosterone and aggression has been found in many bird species (see Soma 2006). Higher levels of testosterone may lead to larger badge sizes, in which case a larger badge will indicate an individual's propensity to fight (Johnstone and Norris 1993). This is in contrast to the common idea that badge size acts as a reliable signal for resource holding potential (RHP) in aggressive encounters with other males (Liker & Barta 2001) and that it thus signals fighting ability (Nakagawa et al. 2007). Trainor et al. (2003) showed that injecting male mice with testosterone after winning an aggressive encounter led to more aggressive mice the following day. However, Blue Tits (Cyanistes caeruleus) crown coloration was shown to affect the outcome of a fight only if individuals had not fought before (Vedder et al. 2010), and Laucht et al. (2010) failed to find a correlation between badge size and testosterone level in house sparrows. Laucht et al. (2010) suggested that perhaps testosterone level and badge size only correlate at the time of badge development (i.e. during the autumn moult), and not at the time of year in that study. Furthermore, since all the individuals in this study were familiar with each other, it is possible that badge size is only partially or not at all related to aggression. Solberg and Ringsby (1997) found that badge size explained a small amount of the variance in dominance rank in three house sparrow populations, suggesting that the importance of badge size decreases as individuals become more acquainted. Thus, the relationship between testosterone and aggression seems dependent on a multitude of environmental factors, such as time of year (breeding vs. non-breeding season) and population stability and/or density. Given the various conflicting results in the literature, it seems important for this study to investigate the correlations between badge size, badge category, beak coloration and age (all status/condition symbols). It is possible that such covariance will be part of the sparrow PoLS in our case, because Jensen et al. (2004) working on nearby sparrow populations in Norway found that badge size increased with age, and that it correlated positively with lifetime reproductive success. It is of great interest to document the extent of the phenotypic covariances between these different morphological, physiological and behavioural traits.

This project was part of an extensive long-term study of the house sparrow by members of the Centre of Biodiversity Dynamics (CBD), involving over 20 years of data collection in the field and genetic parentage analyses in the laboratories. Previous studies have focused on individual morphology in fledglings (Ringsby et al. 2002) and adults (Jensen et al. 2003, Jensen et al. 2004), effective population sizes (Engen et al. 2007, Baalsrud et al. 2014), survival (Holand et al. 2014), extinction (Ringsby et al. 2006) and dispersal (Altwegg et al. 2000, Pärn et al. 2009, Pärn et al. 2012). The aim of this study is to explore variation in personality (i.e. individually consistent behavioural variation) and behavioural syndromes (i.e. covariation between behaviours) by examining individual variation in exploration, neophilic behaviour, aggression, badge size, and basal metabolic rate, and thereby investigate a potential link between these (i.e. a pace-of-life syndrome). Because this study is largely descriptive, the advantage of carrying out this work as part of such a well-established project, is that it involves sub-populations located on three different islands, which allows me explore whether any of the aforementioned features covary with other measured traits of individuals. I can also explore whether the average behaviours and covariances differ across sub-populations (i.e. different farms) in a way that might suggest something adaptive is going on at the level of the social group or flock (see Dingemanse et al. 2007).

Methods

Study species and locations

The house sparrow is a great model species as it is easy to observe in the wild and in captivity, specifics regarding much of its life history are known, and it is a highly successful, adaptive and innovative species (Anderson 2006). House sparrows are highly social, breeding in loose colonies, and almost exclusively feed and move around in flocks. Although adults usually eat seeds and feed insects to their chicks, this species is regularly described as an opportunistic forager that is innovative in its choice of food, and as such it can feed on a variety of items, for example animal feed (pellets) (Anderson 2006). The study populations consist of the house sparrow meta-populations located on three Islands off of the coast of Norway: Leka, Lauvøya, and Vikna. These archipelago meta-populations have been part of a House Sparrow research project run by the Centre of Biodiversity Dynamics (CBD) and Department

of Biology (NTNU) since 2002. Populations are closely associated with dairy farms, with some natal dispersal and negligible breeding dispersal between them (Pärn et al. 2012). Low BMR was selected for at Leka 2012-2014, where by individuals of high BMR were removed. And in 2012, the Lauvøya population was cleared out and replaced with 70 birds of high (from Leka) and 70 birds of low BMR (from Vega) at a sex-ratio of approximately 50:50. Data were collected in 2015 at Leka February 6th-12th, at Lauvøya February 19th-23rd, and at Vikna February 28th-March 5th. The number of birds that were assayed at these islands were 57 (36 males & 21 females), 60 (30 males & 30 females), and 81 (45 males & 36 females), respectively.

Basal metabolic rate & morphological measurements

Birds were assayed in groups ranging in size from 4-16 individuals. These clusters of birds were caught on the same day at the same barn, or in the area around the same barn using mist nets. These clusters are thought to represent natural social groups or flocks. There was one holding barn per island, where all the captured birds were kept for a two-week period. During this period, the birds were ringed with a numbered metal ring (distributed by Stavanger Museum) and three coloured plastic rings, which allowed for identification of each individual throughout the study. Additionally, morphological and physiological features were measured/gathered: body mass, mask length, throat patch length and width, blackness of throat patch (1-5), and beak coloration (1-5; 1 equals light yellow and 5 equals black). Body mass was measured to nearest 0.1g with a pesola spring balance. All other morphological features were measured to nearest 0.1mm using slide calipers (see Jensen et al. 2003, 2006). Several fieldworkers were involved in the measurements of morphology. Following a period of training, each fieldworker measured a minimum of 30 individuals together with T.H. Ringsby or another experienced fieldworker. Differences in linear measurements were then tested using paired t-tests. Significant ($p<0.05$) measurements were adjusted to the T.H. Ringsby-equivalent measurements by adding mean differences. Total badge size and visible badge size (equation 1) were calculated following Møller (1987).

$$Badge\ Size = 166.7 + (0.45 \times \text{length (mm)} \times \text{width (mm)}) \qquad \text{equation 1}$$

See Solberg & Ringsby (1997) for a detailed description of badge size measurements. It is common practice to square root transform badge size to standardize its variance relative to the other phenotypic traits (Whiting et al. 2003, Ringsby et al. 2009). However, this led to a highly leptokurtic distribution when performed on badge size, and was therefore disregarded.

Prior to the individual and communal assays, all birds had their basal metabolic rate (BMR) measured. They spent an 8-hour period in a respiratory chamber where BMR was measured as oxygen consumption rates using an open flow system. A Servomex type 4100 two channel oxygen analyser (Crowborough, England) was used to measure oxygen concentration in the dried effluent air. An automatic valve-system switched between the eight chambers (1.1 L metal boxes), allowing four chambers to be measured simultaneously for 26min and fresh air to be pumped through the system for four minutes between the switches. The measuring protocol is described in detail elsewhere (Rønning et al. 2015). The rate of oxygen consumption (ml O_2 h^{-1}) was calculated following Withers (1977), using a respiratory quotient of 0.71. The lowest 10 min running average VO_2 was used to represent BMR. Each bird was measured once, and to maximize the capacity eight birds were measured in the evening between 16:00 and 22:30 local time (period 1), while eight birds were measured at night from 23:00 to 08:00 (period 2). Wintertime in northern Norway is characterized by a short day length, and the average daily light cycle during the measurement period was 10L:14D (light: 07:30 - 17:30 local time). Consequently, the birds were measured during their normal resting phase irrespectively of whether being measured during the evening or during the night. The basal metabolic rate measurements were adjusted for chamber-differences within each period.

Once the birds had spent ~8 hours in the BMR chambers, each social group or flock was kept in a room where the temperature was in the range ~11-19°C. They were given *ad libitum* access to food and water for a 15-24 hour settling period. The food was then removed at ~23:00, which was nine hours before the start of individual assays. Every bird in each flock was therefore exposed to identical conditions for a minimum of 24 hours prior to the individual assays, which in theory left all birds in a similar state (equally hungry and equally habituated to captivity). To account for any possible differences in condition, repeated body mass measurements were conducted 1) upon capture (mass1), 2) before BMR measurements (mass2), 3) after

BMR measurements (mass3), 4) before individual assay (pre-assay mass), and 5) after communal feeder assay (post-assay mass). Pre-assay body mass was used as one measure of state/condition, and a relative mass was calculated (see equation 2) and used as a second measure of state/condition.

$$\text{Relative mass} = \frac{\text{pre-assay mass}}{(\text{mass1} + \text{mass2} + \text{mass3} + \text{post-assay mass}) \div 4} \qquad \text{equation 2}$$

Individual assays

In January 2014, two master's degree students at NTNU, Sindre Lysfjord Sommerli and John Hammerås, built eight cages with two different perch arrangements that were used for individual behavioural assays. These eight cages accommodated eight visually confined birds at once (see Fig. 2). Each cage had a settling box with a 'piston' attached to the outside of the cage (for encouraging the bird to exit at the start of the trial without the need for handling), with a removable piece of cardboard blocking the entrance to the cage. Each cage had eight wooden perches attached to one wall of the cage arranged either as four and four forming two parallel lines (type 1), or as two above and two below the two parallel lines (type 2) – see Fig. 2.

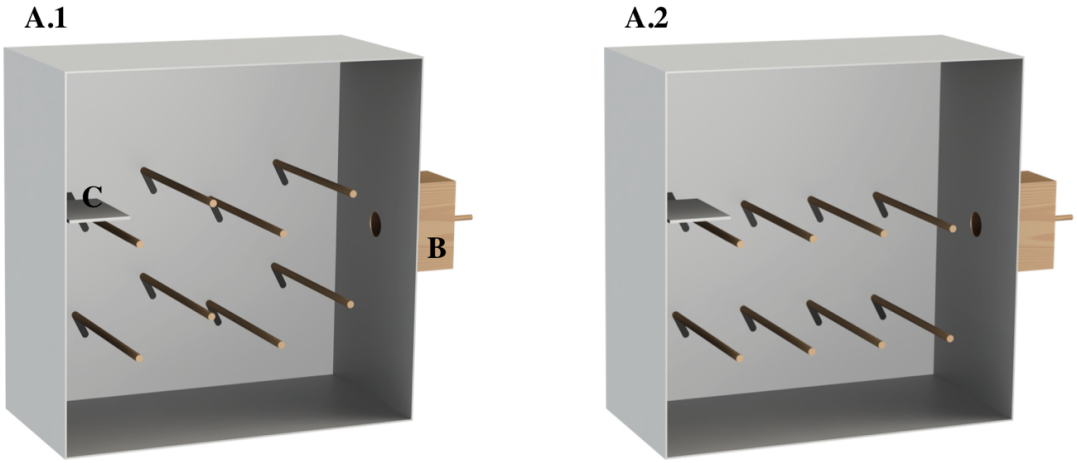

Fig. 2. Illustration of the two types of individual assay cages: type 1 (A.1) had eight perches arranged in four sets of two, and type 2 (A.2) had eight perches arranged in two sets of four. The cage had a settling box with a piston to encourage exits (B), and a metal shelf with a paper tray (C) with a small separate door that allowed experimenters to introduce/swap over a novel object and novel food. For illustrative purposes the front of the cage is not shown, and the cage walls are modelled as solid even though they consisted of chicken wire, and the walls were covered with thick, white paper on the inside. (Illustration provided by designer Vegard Bakke Svendsen)

On the opposite side from the settling box there was a metal shelf with a paper tray, which had a small door next to it to allow the introduction of a novel object and novel food by experimenters with minimal disturbance to the birds. The inside of each cage was lined with thick, white paper in order to enhance detection of the bird during video analysis. Visual isolation between cages was achieved by separating the cages using black plastic sheets.

As the individual assay cages held a maximum of eight individuals at once, every social group (flock) containing more than eight individuals had to be split into two individual assay groups varying in size from 2 to 8. The first individual assay group started between 08:19-10:36, while the second individual assay group started between 08:44-10:52. Each individual assay round lasted for approximately one hour, starting with each bird spending a 5-minute period in its respective settling box. This allowed the birds to calm down subsequent to capture, handling and transfer from the flock aviary. The cardboard between the box and cage was then removed and the bird was encouraged to move out of the box and into the cage using the 'piston'. The bird was then free to explore and move around the cage for 20 minutes. After the bird had experienced the novel cage environment, a novel object was introduced onto the metal shelf (see Fig. 2). There were two alternative novel objects: (1) a small rubber duck with a semi-inflated, dotted balloon covering its head, and (2) a lykketroll (see Fig. 3). After 20 minutes the novel object was replaced with novel food, which was left in the cage for 20 minutes, marking the end of the assay (i.e. three trials of 20 mins each). In some cases, the novel object fell down from the metal shelf, and had to be left in the cage for the rest of the assay. There were two alternative types of novel food: (1) canned dog food ("Bestevenn" from Rema 1000), and (2) a boiled egg cut in half (boiled for ~ 10min). A ruler was attached to the front of one of the cages for calibration of distance during video analyses.

Using two types of novel object, novel food, and individual assay cage perch arrangements allowed for repeatability assays to be run on one flock on each island: 16 individuals at Vikna, 8 at Leka, and 12 at Lauvøya. The second round of assays on this sub-set of 36 individuals started 48 hours after the first round. During the repeatability assays individuals belonging to the same flock were assayed a second time using a different cage type, novel object, and novel food than they were introduced to the first time. This made it possible to repeat the unfamiliarity/novelty of the situation with regard to the cage, novel object and food, whilst also taking into

account the types of novel object and food, and their order of presentation. Each repeatability group therefore spent two more days in the separate aviary in the barn with *ad libitum* access to food and water until ~23:00 the evening before their second round of individual and communal feeder assays.

Fig. 3. The novel objects: on the left is the lykketroll, and on the right is the rubber duck with a semi-inflated, dotted balloon covering its head.

Individual movements were recorded in two-dimensions using a Sony CCD camera (model HDR-AS20), and were analysed using the automatic data collection program EthoVision Software (Noldus Inc., Spink and Tegelenbosch 2001). Each cage was divided into four quadrants, and video analyses were conducted for the three trials separately: exploration (0-20 minutes), novel object (21-41 minutes), and novel food (42-62 minutes). Four variables were measured: (1) total distance moved (mm), (2) minimum distance to novel object achieved (mm), (3) minimum distance to novel food achieved (mm), and lastly (4) the latency to visiting each of the four quadrants (s). Some individuals did not visit one or more of the quadrants, in which case each missing value for these quadrants was replaced with the maximum length of the trial (1200 s). The latencies to visit all quadrants were pooled to represent an individual's time to explore cage behaviour during the first 'exploration' trial. Neophobia and neophilia can be evaluated by noting how quickly the individual approached the object, how much time it spent around the object and how often it approached the object. Thus, 'minimum distance to novel object' and 'minimum distance to the novel food' were the individual's minimum distance to novel object/neophilia towards the novel object, and the novel food, respectively. 'Total

distance moved' during each trial represented the individual's general level of activity: activity1 for the initial exploration trial, activity2 for the novel object trial, and activity3 for the final novel food trial.

Communal feeder assays

After the individual assays, all individuals were released back into an aviary with a communal feeder containing food. The aim of the communal feeder assay was to assess individual foraging (i.e. producer-scrounger) behaviour in the presence of conspecifics from the same social group. See Mirja Carola Olsen's masters' thesis for details regarding the communal feeder assays.

Post Assays

Upon completion of the individual and communal feeder assays, the whole flock was captured and released into a larger aviary within the barn. This aviary contained all other birds captured on the island, except those yet to be assayed. In the release-aviary birds had *ad libitum* access to food and water until the release date. Early on release-day, all birds were captured using mist nets, and then released back into the wild at their respective sites of capture around noon.

Statistical Analyses

Statistical analyses were performed in R 3.2.2 (R Core Team 2013) using the lmerTest (Kuznetsova et al. 2015) package's lmer() function for mixed-effect models, the lavaan (Rosseel 2012) package's sem() function for structural equation modelling, and the MCMCglmm's (Jarrod Hadfield 2010) MCMCglmm() function as well as coda's (Plummer et al. 2006) HDPinterval() function for estimating repeatability. The variables that represent time to explore cage, minimum distance to novel object and minimum distance to novel food were all log-transformed due to non-normality, which also has the advantage of standardizing variances across the different variables (Houle et al. 2011).

Univariate tests

Linear mixed-models were used to investigate the effects of sex and body mass on all six behaviours separately. The two masses, pre-assay body mass and relative mass per individual, were run in independent models because they were positively

correlated. It is possible that body mass has different effects on the two sexes, and so interactions were included between mass and sex. Additionally, to account for expected variation in behaviour among the social groups, 'flock' was used as a random factor in every model. Linear mixed models were also used to investigate the effects of 'island' and 'flock' on BMR in independent models. Sex may have different effects on the three islands or the 17 flocks, and so interactions were included between flock and sex, and island and sex. To account for possible variation in BMR between evening and night measurements, 'measurement period' was used as a random factor in all BMR models. Model simplification was performed based on Akaike information criterion and significance values (Forstmeier and Schielzeth 2011). The maximum relative mass was 1.05 (μ ± S.E. = 0.975±0.002), which is an individual weighing 5% more than its average self, i.e. 1.00. Conversely the minimum relative mass was 0.92, which is an individual weighing 8% less than its average self.

Repeatability

Repeatability, which is a pre-requisite for any personality (Bell et al. 2009), was estimated following Dingemanse and Dochtermann's (2013) mixed-effects model method. Repeatability was estimated for all behaviours using 'individual' as a random factor, and 'assay order' as fixed effect. A dummy variable for 'novel object' was added as a fixed effect for 'minimum distance to novel object' and 'activity2', and a dummy variable for 'novel food' was added as a fixed effect for 'minimum distance to novel food' and 'activity3'. Repeatability was also estimated for basal metabolic rate using measurements for 29 individuals that had their BMR measured in 2014 as well (provided by Bernt Rønning). BMR repeatability was estimated using 'individual' as a random factor, and 'measurement period' and 'order' (year 2014 vs. 2015) as fixed effects.

Correlations

The Pearson product-moment correlation coefficient was used to determine the strength and direction of associations between (1) all six behavioural variables: activity1, activity2, activity3, time to explore cage, minimum distance to novel object and minimum distance to novel food, (2) between all structural equation modelling variables, which were the six behaviours, BMR, pre-assay mass and relative mass, and (3) for males only: between the six behaviours, BMR, pre-assay mass, relative

mass, visible badge size, mask length, beak coloration, badge category, and minimum age. The strength of these associations were assessed prior to SEMs to indicate whether or not the hypothesized models were likely to produce reasonable results. Due to 'measurement period' differences in the raw BMR data (see univariate test results), I used the residuals of a linear regression with 'raw BMR' as the response variable and 'measurement period' as a fixed effect (i.e., controlling for 'measurement period').

Structural Equation Modelling

Structural equation modelling (SEM) was used to further study the pattern of covariances between the different behavioural measures, both separately and in combination with the physiological variables (BMR, mass). Due to 'measurement period' differences in the raw BMR data (see univariate test results), I used the residuals of a linear regression with 'raw BMR' as the response variable and 'measurement period' as a fixed effect. All behaviours, pre-assay mass, relative mass and BMR had leptokurtic distributions, and minimum distance to novel object, activity1, activity2 and activity3 were also asymmetrical. Multivariate normal distribution is important in structural equation modelling, as non-symmetry and kurtosis increases type 1 error (Nachtigall et al. 2003). Thus, the maximum likelihood-based χ^2 test statistics were rescaled by a value that reflects the degree of kurtosis, which is called the Satorra-Bentler scaled test statistic (Satorra and Bentler 1988, 1994). To assess the fit of a model apart from its relative fit (AIC), the comparative fit indices (CFI), which compares the model to the null model (worst fitting model) is used. The 'worst fitting model' is always the model that assumes an absence of covariances. The higher the CFI the better, and an acceptable model would lie in the range of 0.90 and 0.95 (Brown 2013).

Seven different behaviour models (198 individuals) were run: H0, H1, H2A, H2B, H3, H4 and H5 (see Figure 5 for an overview). These models imply a structure for the covariances between the observed variables, and assume that the covariances between our observed variables can be explained by one or more underlying behaviours (latent factors). Model H0 assumes an absence of covariances, and H1 assumes covariance around a single latent variable – possibly the live-fast-die-young pace-of-life syndrome (PoLS). H2 assumes an activity and exploration syndrome (with covariance between these two latent variables [H2b] or

not [H2a]). H3 assumes that activity is the only underlying behaviour, explaining correlations among the activity variables. H4 assumes that exploration is the only underlying behaviour, explaining correlations among the other three behaviours. The last model H5 assumes that personality is activity-driven, with all other covariances being specific to trial-specific levels of activity. As a second step in the SEM analyses, the data set was divided in an attempt to check whether or not any of these syndrome structures were affected by sex, state or flock identity (Dingemanse et al. 2010a). However, this was not always possible due to the small sample sizes and/or lack of covariances (i.e. negative estimated variances were obtained for some of the behaviours when the sexes were separated). Therefore, as an alternative approach, the same behaviour models were re-run using the residuals of linear models where all variables had been regressed on (1) sex, (2) social group/flock, and (3) relative mass separately. The purpose of this was to remove the (co-)variance and therefore test if the SEM could be explained by sex, social group/flock, and relative mass separately.

The physiological variables involving relative mass, pre-assay mass and BMR were added to the best behaviour model H5 (190 individuals), resulting in five physiology plus behaviour syndrome models (see Figure 6 for an overview). The null model here is the behavioural model (H5) that was used to construct these physiology and behaviour models. The H5_1 model is the live-fast-die-young pace-of-life (PoLS) model where everything can be explained by one underlying latent variable. BMR and pre-assay mass are allowed to covary in all models, seeing as we know that they correlate. In model H5_2 relative mass is removed from the syndrome, and in H5_3 both mass variables are removed. Models were rerun using residuals of flock, and sex.

The behaviour and behaviour-and-physiology models were run for males only (102 individuals). Then, the physiological variables relative mass, pre-assay mass, BMR, mask length, badge size, badge category, age, and beak coloration were added to the best behaviour model H5, resulting in six male-specific behaviour-and-physiology syndrome models (see Figure 7 for an overview). The null model (M0) here is the behaviour-and-physiology model (H5_4) that was used to construct these male-specific physiology and behaviour models. Model M1 assumes a syndrome in which all morphological variables (except rel-mass) covary with personality (activity variables) and BMR. In the two M2 models (M2A and M2B) everything can be

explained by two underlying latent variables: activity and status. Activity explains the covariation between the activity variables, and Status explains the covariation between BMR, age, mask length, badge size and category, and beak coloration. Models M3A and M3B were identical, except that BMR was part of Activity. Models were rerun using residuals of flock.

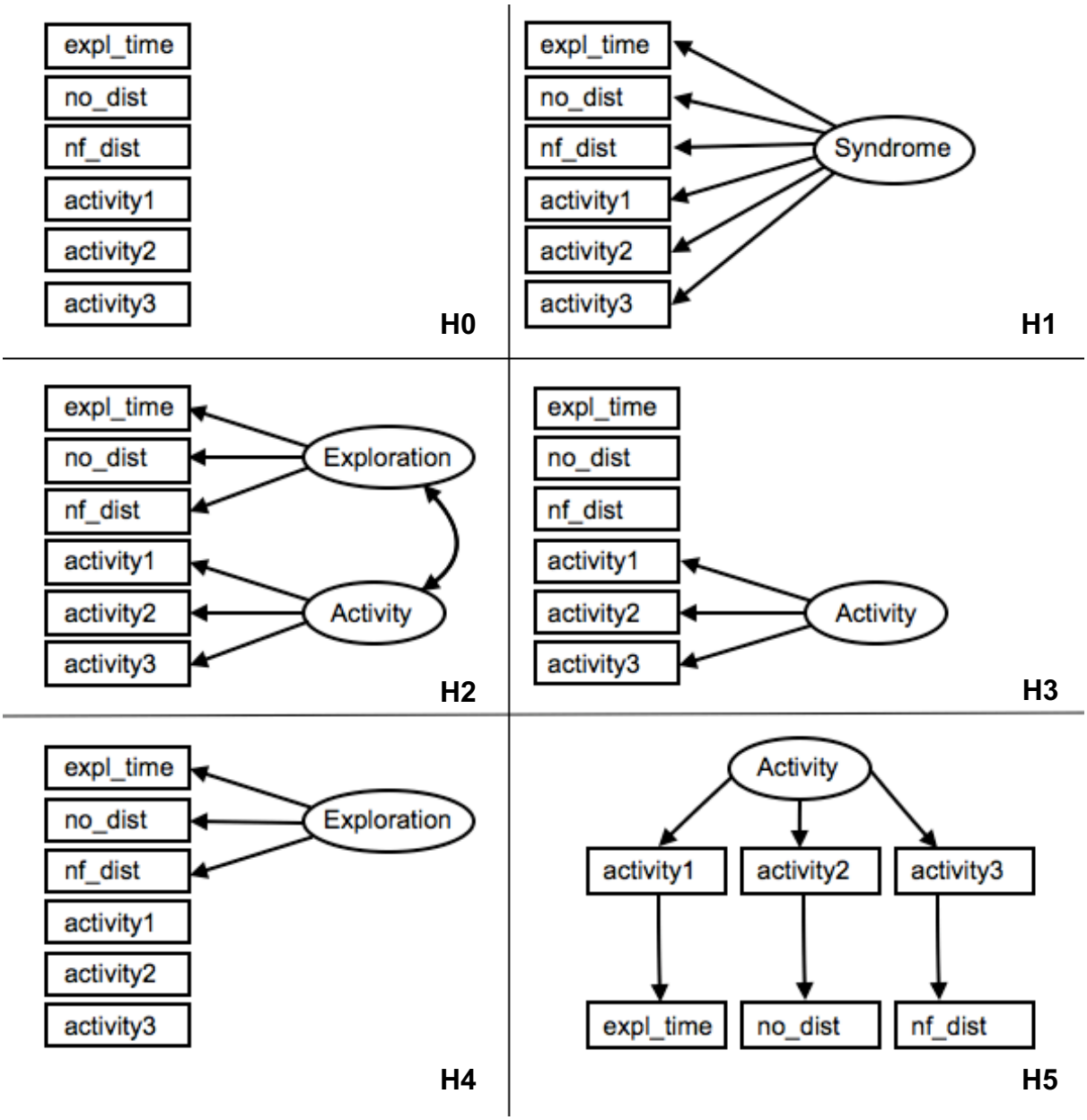

Fig. 5. Illustration of the seven hypothesized behavioural models that were tested using structural equation modelling (SEM). Model H0 assumes an absence of covariances. 'expl_time' is time to explore, 'no_dist' is minimum distance to novel object, and 'nf_dist' is minimum distance to novel food. H1 assumes covariance around a single latent variable – possibly the live-fast-die-young pace-of-life syndrome (PoLS). H2 assumes an activity and exploration syndrome (with covariance between these two latent variables [H2b] or not [H2a]). H3 assumes that activity is the only underlying behaviour that covaries. H4 assumes that exploration is the only underlying behaviour. The last model H5 assumes that personality is activity-driven, with all other covariances being specific to trial-specific levels of activity. All

observed variables are illustrated using squares, single-headed arrows are regressions, and double-headed arrows are co-variances when connecting two variables and residual variances when connecting a variable to itself.

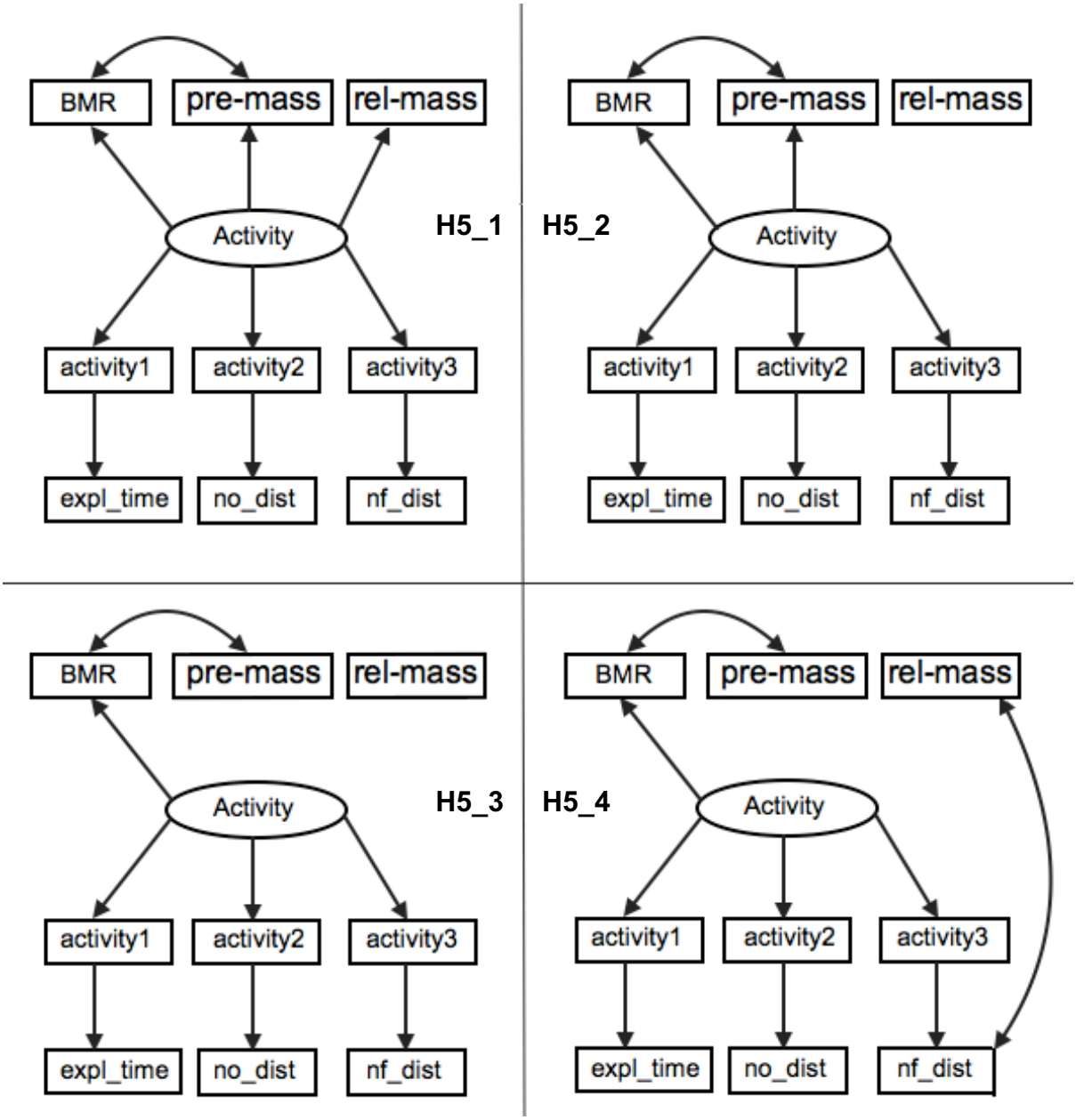

Fig. 6. Illustration of the main types of physiology/morphology plus behaviour models that are expansions of the activity-driven personality behaviour model, H5 from Fig. 5. BMR is basal metabolic rate, 'expl_time' is time to explore, 'no_dist' is minimum distance to novel object, 'nf_dist' is minimum distance to novel food, and Activity (circle) is the hypothesized underlying behaviour (latent variable). Model H5_0 (not illustrated) includes pre-assay mass, relative mass and BMR, however, it assumes an absence of physiology in the syndrome, and so no links with any behavioural variables. H5_1 assumes the live-fast-die-young pace-of-life syndrome (PoLS). H5_2 assumes a PoLS in which relative body mass is removed from the syndrome, and H5_3 assumes a PoLS in which both mass variables are removed from the syndrome. Model H5_4 assumes a PoLS in which pre-assay mass is removed from the syndrome, whilst relative mass co-varies with 'minimum distance to novel food'. All observed variables are illustrated using squares, single-headed arrows are regressions, and double-headed arrows are co-variances.

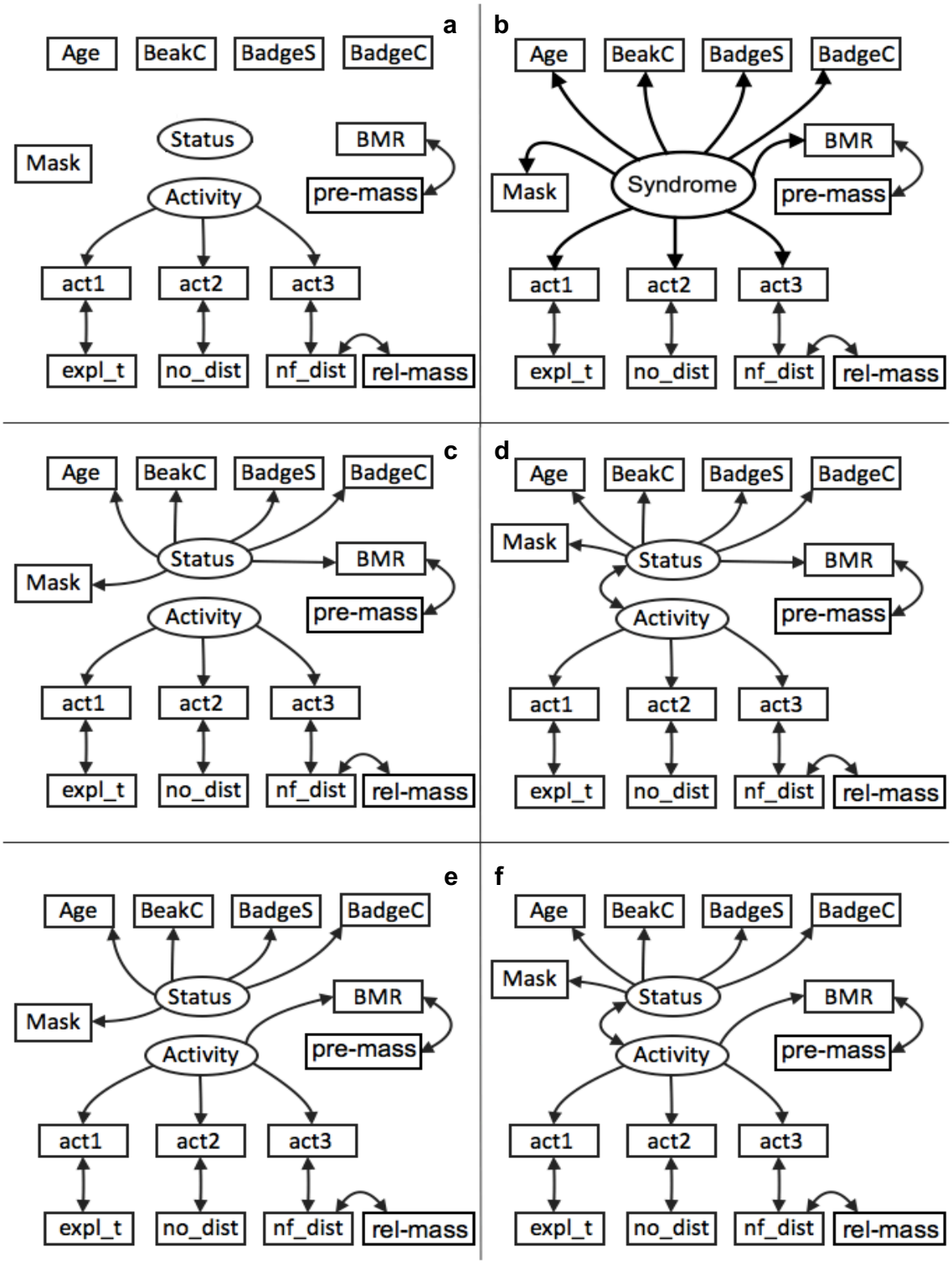

Fig. 7. Illustration of the main types of male-specific physiology/morphology plus behaviour models that are expansions of the activity-driven personality behaviour model, H5 from Fig. 5. BMR is basal metabolic rate, mask is mask length, BeakC is beak coloration, BadgeS is visible badge size, BadgeC is badge category, 'expl_time' is time to explore, 'no_dist' is minimum distance to novel object, 'nf_dist' is minimum distance to novel food, and Status, Activity, and Syndrome (circles) are the three hypothesized underlying behaviours (latent variables). Model M0 (a) assumes an absence of physiology. M1 (a) assumes a syndrome in which all morphological variables (except pre-mass and rel-mass) covary with personality

(activity variables) and BMR. M2 (c and d) assumes an activity and status syndrome where BMR correlates with a male's status signals (with covariance between these two latent variables [M2b] or not [M2a]). M3 (e and f) assumes assumes an activity and status syndrome where BMR correlates with a male's activity level (with covariance between these two latent variables [M3b] or not [M3a]). All observed variables are illustrated using squares, single-headed arrows are regressions, and double-headed arrows are co-variances.

Results

Univariate Tests

As illustrated in Fig.8, there was a positive relationship between BMR and pre-assay mass ($r = 0.50$, $n = 188$, $p<0.001$). This well-documented correlation is why only one of them (i.e., pre-assay mass) could be used as explanatory variable in the univariate tests for all six behaviours. Likewise, there was a positive relationship between pre-assay mass and relative mass ($r = 0.37$, $n = 196$, $p<0.001$), and univariate tests for all six behaviours were therefore split into separate tests for the two measures of mass.

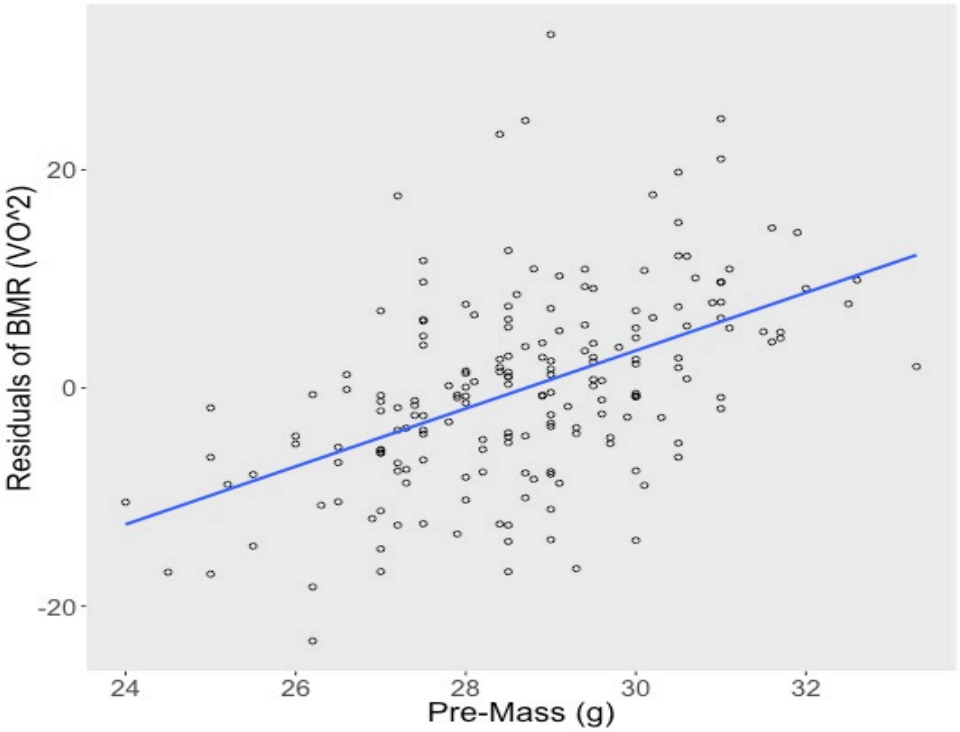

Fig. 8. The positive linear relationship between basal metabolic rate (residual rate of oxygen consumption (VO^2), controlling for the effect of 'measurement period') and pre-assay body mass (g), with the fitted line y = -76.27 + 2.66*(pre-mass).

When testing the effects of sex, pre-assay mass and relative mass on the variation in each of the six behaviours, there were very few significant effects (Table 1 and 2). One exception was that relative-mass and sex had a significant effect on activity in the presence of novel food, where male activity decreased with relative mass, whilst female activity increased with relative mass. The other exception was that minimum distance to novel food increased significantly with an increase in relative mass (β = 6.87±2.34, $F_{1,194}$ = 8.63, p = 0.00371). There was a strong effect of 'flock' on basal metabolic rate (Table 3), which explained 26% of the variance (Fig. 9). There was an effect of island on BMR ($F_{2,185.01}$ = 4.120, p= 0.018), where BMR was higher at Lauvøya (β= 81.309±2.298) than at Leka (β= 76.636±2.320), whilst Vikna (β= 78.998±2.223) did not differ from either island. Measurement period seemed to account for ~9-10% of the variance in basal metabolic rate, and island accounted for ~5%.

Table 1. The effect of sex and pre-assay mass on the six behaviour variables. The highly significant effects (p < 0.01) are shown in bold, while the various moderately significant effects (p < 0.05) are in italics. The effect degrees of freedom are 1 for all models, and the error degrees of freedom are provided. Variance and standard deviation (S.D.) for the random effect 'Flock' and the residuals are also provided.

	expl_time			***no_distance***			***nf_distance***		
	df	F	P	df	F	P	df	F	P
Sex	193.90	1.46	0.228	193.63	0.972	0.325	194	0.107	0.744
Pre-mass	176.04	0.003	0.957	182.08	3.68	0.057	194	0.634	0.427
Sex: Pre-mass	191.92	0.676	0.412	191.57	0.872	0.352	193	1.55	0.215
	σ^2	S.D.		σ^2	S.D.		σ^2	S.D.	
Flock	0.017	0.132		0.008	0.090		<0.001	<0.001	
Residuals	0.525	0.725		0.180	0.425		0.575	0.758	
	activity1			**activity2**			**activity3**		
	df	F	P	df	F	P	df	F	P
Sex	192.35	0.331	0.566	193.59	0.659	0.418	200.73	0.810	0.369
Pre-mass	177.54	0.297	0.587	193.59	0.549	0.460	200.73	0.607	0.437
Sex: Pre-mass	187.09	0.002	0.965	200.76	0.005	0.945	185.56	0.070	0.791
	σ^2	S.D.		σ^2	S.D.		σ^2	S.D.	
Flock	7032	83.85		<0.001	<0.001		<0.001	<0.001	
Residuals	232750	482.44		270300	519.9		235423	485.2	

Table 2. The effect of sex and relative mass on the six behaviour variables. The highly significant effects ($p < 0.01$) are shown in bold, while the various moderately significant effects ($p < 0.05$) are in italics. The effect degrees of freedom are 1 for all models, and the error degrees of freedom are provided. Variance and standard deviation (S.D.) for the random effect 'Flock' and the residuals are also provided.

	expl_time			***no_distance***			***nf_distance***		
	df	F	P	df	F	P	df	F	P
Sex	188.03	5.260	**0.002**	192.69	0.467	0.495	194	0.168	0.682
Rel-mass	186.23	0.842	0.360	188.19	1.420	0.235	194	8.630	**0.004**
Sex: Rel-mass	188.12	5.130	*0.025*	188.24	0.224	0.636	193	0.032	0.859
	σ^2	S.D.		σ^2	S.D.		σ^2	S.D.	
Flock	0.017	0.129		0.007	0.083		<0.001	<0.001	
Residuals	0.514	0.717		0.183	0.428		0.552	0.743	
	activity1			**activity2**			**activity3**		
	df	F	P	df	F	P	df	F	P
Sex	188.52	0.272	0.602	189.91	0.250	0.618	191.67	5.270	*0.023*
Rel-mass	183.37	0.441	0.508	189.91	4.340	*0.039*	191.67	0.035	0.852
Sex: Rel-mass	189.96	2.340	0.128	188.47	2.490	0.116	191.67	5.350	*0.022*
	σ^2	S.D.		σ^2	S.D.		σ^2	S.D.	
Flock	6989	83.6		<0.001	<0.001		<0.001	<0.001	
Residuals	232604	482.3		265100	514.9		230955	480.6	

Table 3. The effect of sex, flock, and their interactions on basal metabolic rate. Highly significant effects ($p < 0.01$) are shown in bold, while the various moderately significant effects ($p < 0.05$) are in italics. The effect degrees of freedom and the error degrees of freedom are provided. Variance and standard deviation (S.D.) for the random effect 'measurement period' (M. Period) and the residuals are also provided.

Flock and Sex Models				
	Error d.f.	Effect d.f.	F	P
Sex	171.01	1	1.064	0.304
Flock	171.10	16	3.793	**< 0.001**
Sex:Flock	155.19	16	0.972	0.489
	σ^2	S.D.		
M. Period	7.373	2.715		
Residuals	64.388	8.024		
Island and Sex Models				
	Error d.f.	Effect d.f.	F	P
Sex	185.03	1	0.627	0.430
Island	185.01	2	4.120	*0.018*
Sex:Island	183.09	2	1.293	0.277
	σ^2	S.D.		
M. Period	8.017	2.831		
Residuals	77.164	8.784		

These results suggest that it was an individual's state relative to itself that affects the time it took to explore the cage, minimum distance to novel food, activity in the presence of a novel object, and activity in the presence of novel food. However, most of these significance effects are >0.01, and so they should perhaps

be viewed with caution given the number of tests that were run, because they might represent false-positives (type 1 errors). Although the large sample size here (n=198) greatly increases the statistical power, thus reducing the probability of false-negatives (type 2 errors), such a large sample size also means that I can detect very small effects of limited biological interest. The aforementioned effects are not particularly small, but their significance levels are less than convincing and the possibility that these effects are statistical artefacts should be considered.

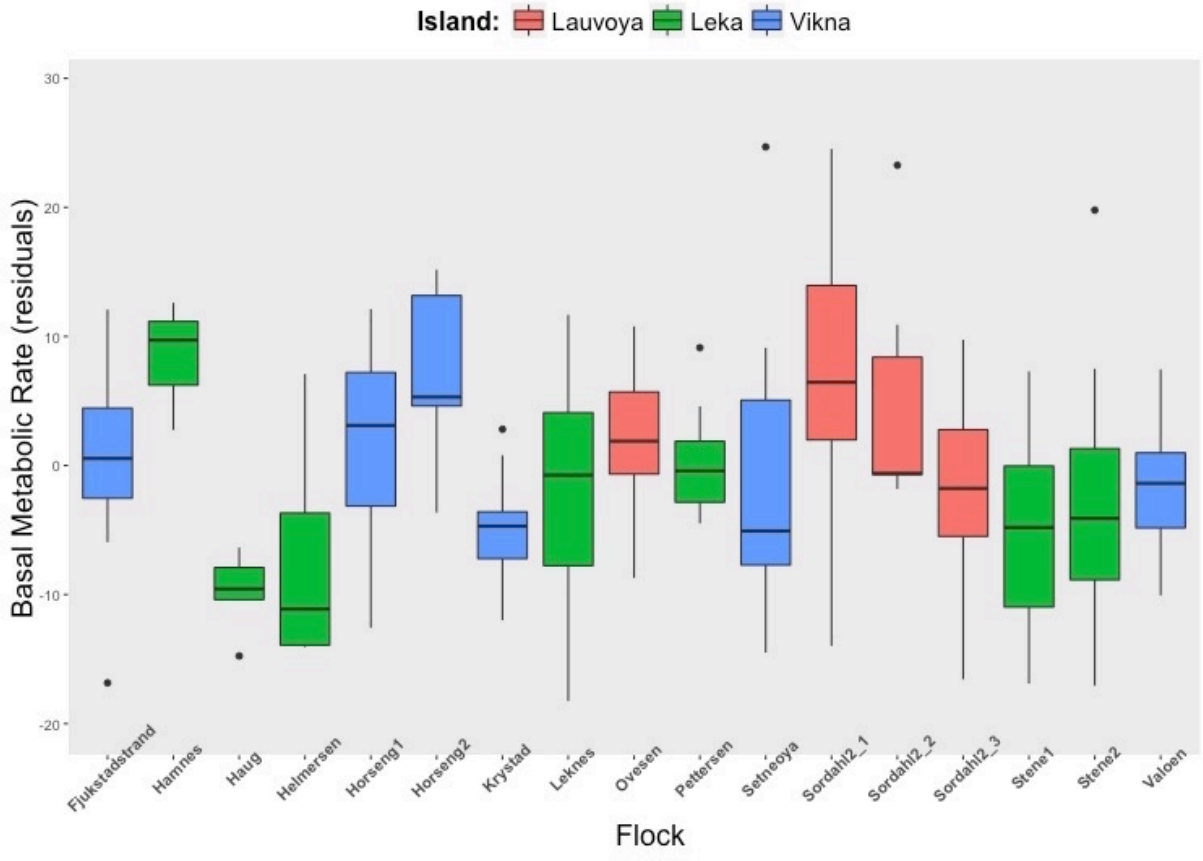

Fig. 9. The effect of flock identity on basal metabolic rate (residual rate of oxygen consumption (VO^2), controlling for the effect of 'measurement period') between 17 flocks on the three islands Leka (green), Vikna (blue) and Lauvøya (red). Horizontal lines represent the median, each box depicts a flock, whiskers show the greatest value excluding outliers, which are represented with dots.

Repeatability

Neither BMR, time to explore the cage, minimum distance to novel object nor minimum distance to novel food showed any repeatability (Table 4). In contrast, all three activity variables showed high repeatability. There was a significant effect of novel object type on 'minimum distance to novel object' (see Appendix A: p=0.045),

where individuals came closer to the 'duck' (β=0.599±0.094) than the 'lykketroll' (β=0.794±0.090). There were no other effects on the non-repeatable behaviours, which was unexpected. However, all three activity variables were positively affected by order, i.e. individuals were more active in the second assay compared to the first (see Appendix A). The average level of activity was 24.7% to 32.2% higher for the repeatability individuals in the second assay compared to the first assay (Fig. 10). Additionally, lower repeatability estimates when these fixed effects were not accounted for supports the notion that these factors also increased the within-individual variance.

Table 4. Repeatability estimates and associated 95% confidence intervals (C.I.) for BMR (n = 58), and the six behaviours time to explore cage, minimum distance to novel object, minimum distance to novel food, activity1, activity2 and activity3 (n = 72). Behaviour models with fixed effects included order, and novel object type as a factor for minimum distance to novel object and activity2, and novel food type for minimum distance to novel food and activity3. BMR model with fixed effects included measurement period and flock.

	Without Fixed Effects		**With Fixed Effects**	
	Repeatability	95% C.I.	Repeatability	95% C.I.
expl_time	< 0.001	< 0.001 – 0.312	< 0.001	< 0.001 – 0.284
no_distance	< 0.001	< 0.001 – 0.323	< 0.001	< 0.001 – 0.413
nf_distance	< 0.001	< 0.001 – 0.183	< 0.001	< 0.001 – 0.237
Activity1	0.601	0.268 – 0.756	0.659	0.346 – 0.775
Activity2	0.476	0.104 – 0.678	0.541	0.187 – 0.692
Activity3	0.678	0.485 – 0.840	0.712	0.507 – 0.849
BMR	< 0.001	< 0.001 – 0.556	< 0.001	< 0.001 – 0.408

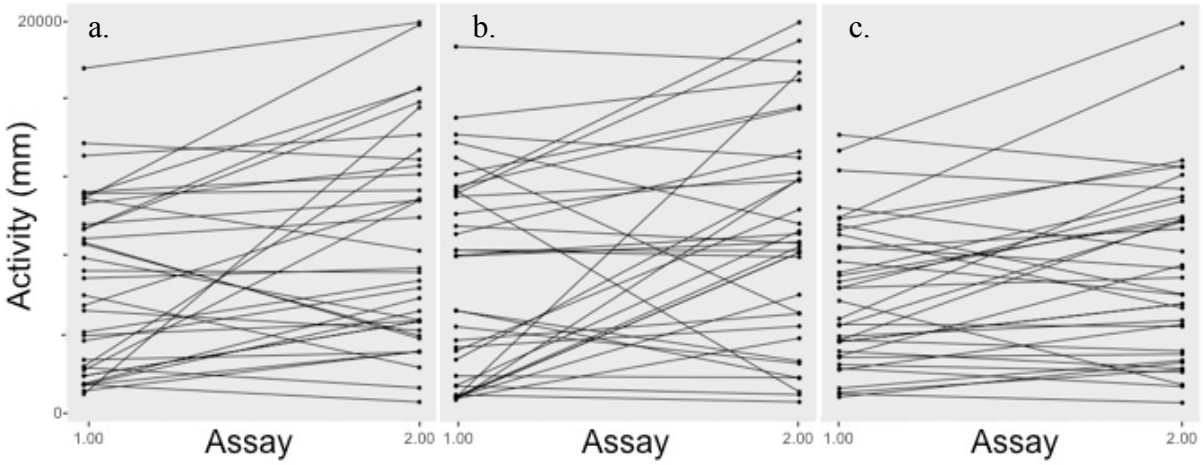

Fig. 10. The positive effect of order on the activity (total millimetres moved) variables illustrated using individual reaction norms for all repeat individuals (n=36) in each of the three trials (a) exploration, (b) novel object, and (c) novel food. Assay '1' represents the first behavioural assay, and '2' the second. The average level of activity1 increased by 31% from the first to the second assay, activity2 increased by 32%, and activity3 increased by 25%.

Correlations

There were weak negative trial-specific correlations between (1) activity1-time to explore cage, (2) activity2-minimum distance to novel object, and (3) activity3-minimum distance to novel food (highlighted with squares in Table 5). These correlations in combination with personality in activity indicate that if there is a behavioural syndrome in these house sparrows, it is likely to be an activity-driven one, where the three activity variables correlate with each other, and each activity variable co-varies with the other trial-specific behaviour. There were no distinct differences between correlations within all individuals and only within males (see Appendix B). A correlation matrix is difficult to interpret as a whole, which is why the covariance pattern of the behaviours and physiological measures were then examined more rigorously using structural equation modelling.

Table 5. Pair-wise correlations between the six behaviours: time to explore cage (expl_time), minimum distance to novel object (no_dist), minimum distance to novel food (nf_dist), activity1, activity2 and activity3), basal metabolic rate (BMR), pre-assay mass and relative mass. Correlations between behaviours derived from the same trial are marked with squares, and all significant correlations are marked in bold.

	pre-mass	rel-mass	expl_time	no_dist	nf_dist	act1	act2	act3
BMR	**r=0.50 p<0.001**	r=0.02 p=0.80	r=0.08 p=0.30	r=0.04 p=0.56	r=0.10 p=0.16	r= -0.02 p=0.81	r=0.05 p=0.53	r= -0.01 p=0.85
pre-mass		**r=0.37 p<0.001**	r< -0.01 p=0.92	r=0.12 p=0.09	r=0.05 p=0.46	r= -0.03 p=0.71	r=0.04 p=0.56	r=0.04 p=0.55
rel-mass			r= -0.05 p=0.45	r=0.09 p=0.20	**r=0.20 p=0.004**	r= -0.04 p=0.56	**r= -0.16 p=0.03**	r= -0.01 p=0.84
expl_time				r=0.05 p=0.41	r=0.09 p=0.17	**r= -0.36 p<0.001**	r= -0.08 p=0.25	**r= -0.16 p=0.03**
no_dist					**r=0.40 p<0.001**	r= -0.02 p=0.74	**r= -0.18 p=0.01**	r= -0.08 p=0.29
nf_dist						r= -0.01 p=0.85	r= -0.04 p=0.62	r= -0.12 p=0.09
act1							**r=0.59 p<0.001**	**r=0.57 p<0.001**
act2								**r=0.54 p<0.001**

Structural Equation Modelling

Behavioural Models

The SEM representing an absence of correlations model (H0) had the worst fit (Table 6), which indicates that there is a syndrome of sorts involving these behavioural variables. The best model was the activity-driven personality model, H5. This syndrome explains the correlations among the activity variables quite well (Fig. 11), which suggests that individual activity level was consistent regardless of context (i.e., 'animal personality'). The correlation between activity1 and time to explore the cage indicates that highly active individuals 'explored' the cage faster. Correlations between minimum distance to novel object and activity2, and activity3 and minimum distance to novel food were weaker. All correlations in this syndrome were in agreement with bivariate correlations (Table 5). For both sex- and flock-residual models, the activity-driven model remained the best, and there were no noteworthy changes in model CFIs (see Appendix A). Thus, the syndrome structure in H5 was not caused by simple differences between the two sexes, or between the 17 flocks. However, there was insufficient data to test whether H5 fitted sub-sets of the data within each sex or farm equally.

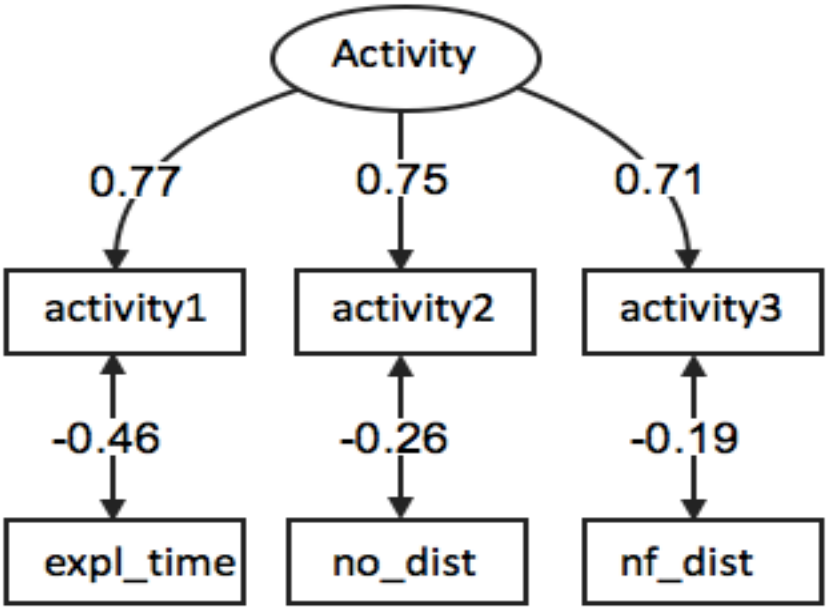

Fig. 11. Path diagram illustrating the behavioural syndrome H5, where behaviour co-varies with activity (n = 197). The activity variables act1, act2 and act3 load onto a latent variable (circle) called Activity. Time to explore cage (expl_time) co-varies with activity1, minimum distance to novel object (no_dist) with activity2, and minimum distance to novel food (nf_dist) with activity3. The numbers on the arrows indicate the standardized path coefficients.

Table 6. The resulting model ranking of behaviour (n = 197) models H0, H1, H2A, H2B, H3, H4 and H5 based on Akaike's Information Criterion (AIC). See Methods for detailed model descriptions.

Model	Degrees of Freedom	ΔAIC	CFI
H5 – Activity Driven	18	0.00	0.865
H2B – Activity & Exploration (correlation between latent variables)	19	6.97	0.848
H2A – Activity & Exploration	18	7.39	0.847
H1 – Single Syndrome	18	25.95	0.758
H3 – Activity	15	37.16	0.702
H4 – Exploration	18	183.38	0.164
H0 – Null Model	12	213.15	0.000

Behaviour plus Physiology Models

The best behaviour-and-physiology model was 'behaviour with relative mass' (H5_4), where behaviour was activity-driven, relative mass covary with minimum distance to novel food, and BMR and pre-assay mass were allowed to covary separately from Activity (Table 7). The correlations between the activity variables and the Activity latent variable were similar to that of the original activity-driven model, whilst the within-trial covariances were identical (see Fig. 11 versus 12). As expected and in agreement with univariate test results, BMR and pre-assay mass correlated with each other (Fig 12). The difference between the 'behaviour with relative mass' model and the 'behaviour only' model was statistically significant, which indicates that relative mass did affect the syndrome structure in some way. Relative mass and minimum distance to novel food correlated positively, which meant that lighter individuals moved closer to the novel food. This suggests that what we think of as 'innovation' in these birds, i.e. utilizing a new food source, may partially be affected by hunger or state.

For both sex- and flock-residual models, H5_4 remained the best and there were no noteworthy changes in model CFIs (see Appendix A). Given the large effect of flock on variation in BMR (Table 3, Fig. 9), it was surprising that the only detectable change in model H5_4 was a very small increase in BMR's correlation with the activity variables. This suggests that there were no differences in syndrome structure between the two sexes, or between the 17 flocks.

Table 7. The resulting model ranking of behaviour and physiology (n = 190) models H5_0, H5_1, H5_2, H5_3, and H5_4 based on Akaike's Information Criterion (AIC). See Methods for detailed model descriptions.

Model	d.f.	ΔAIC	CFI
H5_4 – Behaviour with relative mass	27	0.00	0.773
H5_0 – Behaviour only ("null model")	25	6.17	0.747
H5_3 – PoLS without relative and pre-assay mass	26	8.15	0.748
H5_2 – PoLS without relative mass	27	9.95	0.746
H5_1 – Live-fast-die-young PoLS	28	10.56	0.749

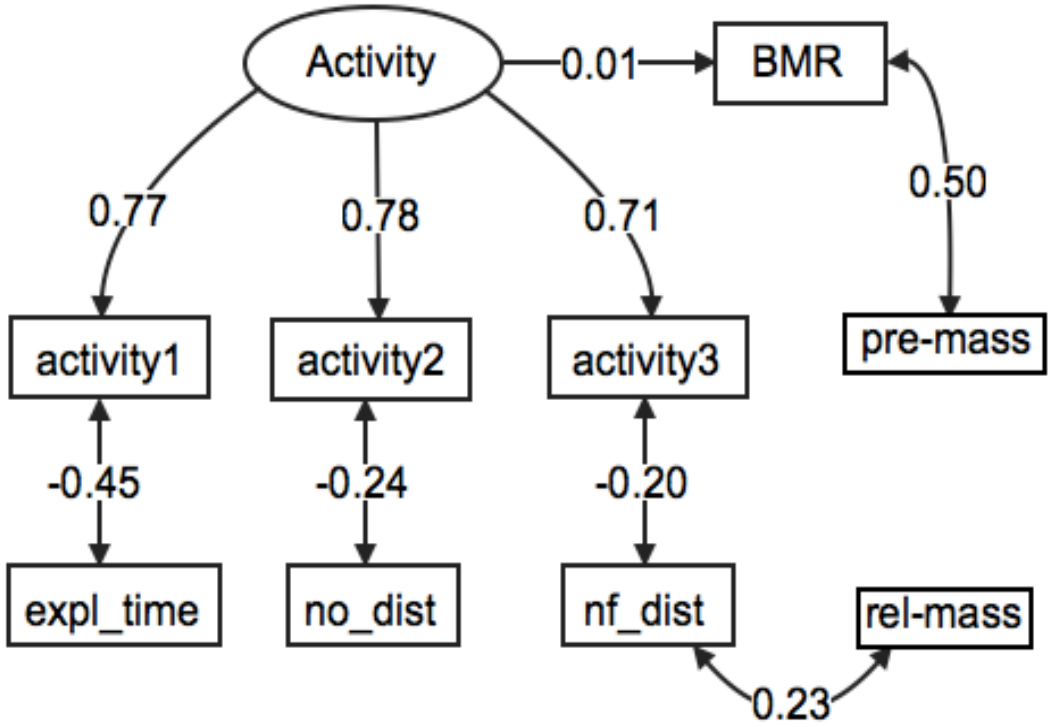

Fig. 12. Path diagram illustrating the behaviour-and-physiology model H5_4, which was identical to the activity-driven behavioural model H5, except that relative mass and minimum distance to novel food (nf_dist) were allowed to covary, and basal metabolic rate (BMR) and pre-assay mass as well (n = 190). The activity variables act1, act2 and act3 load onto a latent variable (circle) called Activity. Time to explore cage (expl_time) co-varies with activity1, minimum distance to novel object (no_dist) with activity2, minimum distance to novel food with activity3, and BMR with pre-assay mass. The numbers on the arrows indicate the standardized path coefficients.

Male-Specific Behaviour, and Behaviour plus Physiology Models

Behaviour models H2A, H2B and H4 would not converge using males only, perhaps suggesting insufficient covariation in these patterns with the smaller sample sizes. However, the remaining models did converge, of which model H5 was still the best based on the relative AIC values (see Appendix A). There were some changes in the strengths of the syndrome correlations, such as stronger correlations between the latent variable and the activities, and between activity3 and 'minimum distance to novel food', whilst there was a weaker correlation between activity1 and 'time to explore'. For flock-residual models, H5 remained the best, which suggests again that differences between the 17 flocks did not drive the male syndrome structure.

All behaviour-and-physiology models converged using males only. However, models H5_0, H5_1 and H5_4 were statistically indistinguishable (see Appendix A). This indicates that there might be subtle differences in the behavioural-and-physiology syndrome structure between the sexes even though there was no change in 'best model' with regard to behaviour-and-physiology models, or any changes in the syndrome structure of the best model (H5_4) when using residuals controlling for sex. There was no behaviour-BMR-mass syndrome in male house sparrows, seeing as the behaviour-only model (H5_0) described the covariance pattern just as well as other models. Removing the effect(s) of flock did not change model rankings, thus the male behaviour-and-physiology syndrome was not the result of flock-differences.

Lastly, a male-specific set of behaviour-and-physiology models were tested, and models M2A, M2B and M3A were statistically indistinguishable (see Table 8). The only difference between M2A and M2B is a correlation between Status and Activity, which was not significant. The difference between M2A and M3A was whether BMR correlated with the Activity variables (M3A) or the Status variables (M2A). There was no correlation between BMR and Activity (β= -0.02, p=0.78), or BMR and Status (β= -0.07, p=0.46). Both correlations were weak and non-significant, resulting in two statistically equal models. This was surprising, as I expected metabolism to correlate with either personality or energy costly/testosterone related coloration. However, the strength of correlations within each latent variable remained fairly constant regardless of where BMR was "placed" in these models (Figure 13). For flock-residual models there were no noteworthy changes in model ranking, nor in model CFIs (see Appendix A). Thus, these male syndrome structures were not driven by between-flock differences. With regard to the best behaviour-only, behaviour-and-

physiology, and male-specific behaviour-and-physiology models, removing the effect of 'flock' increased the fit (CFI) of most models, indicating that there were some flock-differences obscuring the strength of the between-individual differences tested in these SEMs.

Table 8. The resulting model ranking of male-specific (n = 102) extended behaviour-and-physiology models M0, M1, M2A, M2B, M3A, and M3B based on Akaike's Information Criterion (AIC). See Methods for detailed model descriptions.

		Raw Data		'Flock'	
Model	**d.f.**	**ΔAIC**	**CFI**	**ΔAIC**	**CFI**
M2A	42	0.00	0.788	0.03	0.823
M3A	42	0.38	0.789	0.00	0.825
M2B	43	1.99	0.784	2.01	0.818
M3B	43	2.37	0.785	1.98	0.819
M0	37	24.00	0.679	40.38	0.660
M1	43	54.68	0.589	68.59	0.569

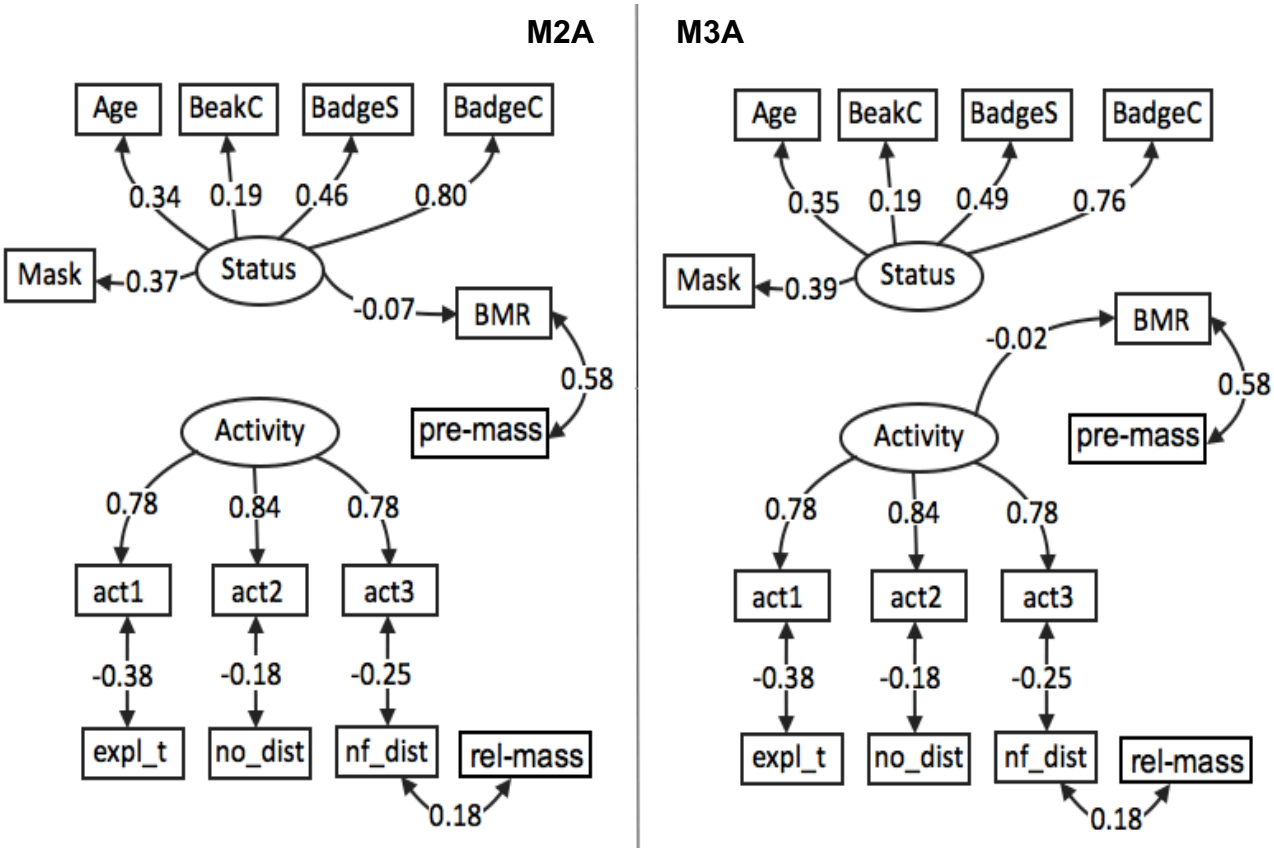

Fig. 13. Path diagram illustrating the male specific behaviour-and-physiology models M2A and M3A. M2A was identical to the activity-driven behavioural-and physiology model H5_4, except that age, mask size, beak coloration (BeakC), badge size (BadgeS) and badge category (BadgeC) load onto an additional latent variable (circle) 'Status' (n = 102). Model M3A is identical to M2A, except that BMR loads onto 'Status' rather than 'Activity'. The numbers on the arrows indicate the standardized path coefficients.

Discussion

I assayed the behaviour of nearly 200 house sparrows on 17 farms on three islands off the coast of northern Norway to test if there was animal personality, and evidence for behavioural and/or pace-of-life syndromes present in these populations. I am careful in how I label the measured behaviours, as there is a lack of consensus as to what exactly is a measure of 'exploration', 'neophobia' and 'innovation' in a social bird species, such as the house sparrow. For example, both Dingemanse et al. (2003) and Verbeek et al. (1994) quantified exploration of a novel environment in great tits as the time it took an individual to visit the last novel tree in a room (five in total), which may not be comparable to our 'exploration' measure, and their measures may contain a lot of variation due to individual activity rather than exploration *per se*. On the other hand, they quantified 'reaction to novel object' as the minimum distance (amongst other measures), which is the same behaviour as I recorded. Whereas Van Oers et al. (2004) described exploratory behaviour in great tits using an open field test without clarifying which dependent variable was measured (i.e., distance covered per unit time, latency to leave start area, time spent without movement etc., Walsh & Cummins 1976). Thus, I have refrained from referring to 'time to explore the cage' as 'exploration', 'minimum distance to novel object' as 'neophobia' and 'minimum distance to novel food' as 'innovation'. It is possible that I was unable to detect repeatability in these three behaviours because of (1) small sample size (n=36), and (2) not enough repeated measures per individual. Dingemanse and Dochtermann (2013) demonstrated that two measures per individual and a small sample size (~25) was enough to estimate repeatability exceeding 0.5. However, with sample sizes of ≤100 and a repeatability below 0.5, a minimum of four repeated measures per individual is needed to estimate repeatability accurately. Thus, I would have been able to detect it if these behaviours had been highly repeatable.

The activity-driven model explained the correlations among the repeatable activity variables very well, where strong correlations suggest that activity is consistent regardless of environmental context in these house sparrows (i.e., 'animal personality'). There was a fairly strong correlation between 'time to explore cage' and an individual's activity level at that time. Whilst an assumption of novel environment tests is that the amount of movement mirrors an individual's level of exploration (Russell 1983), it has been suggested that it may also reflect a behaviour unrelated

to exploration (Renner 1990, Hughes 1997). Thus, this behaviour may be a result of active individuals being superficial explorers, or it may be an outcome of activity level, where active individuals were more likely to visit all parts of the cage faster because they moved more, and not necessarily because they were exploring the novel environment. However, if activity is the underlying cause of 'time to explore the cage', then the latter behaviour should also have been repeatable to some degree. Although this behaviour may exhibit an undetectable repeatability of up to 0.5 (see previous paragraph), the lack of repeatability in 'time to explore cage' could also be a result of the behaviour encompassing both activity and exploration. I found that individuals were more active during the second assay than the first, which in combination with an increase in exploration from one assay to another (see Dingemanse et al. 2002) would render 'time to explore cage' non-repeatable. The correlation between 'minimum distance to novel object' and activity level at that time was weak, indicating that perhaps 'minimum distance to novel object' is also a measure of multiple behaviours, such as activity-level in combination with neophobia (avoidance of unfamiliar objects, Barnett 1958). Thus, a combination of individuals responding differently to the two novel objects and activity increasing from one assay to another may have rendered 'minimum distance to novel object' non-repeatable. However, the two novel food types did not affect 'minimum distance to novel food', which in combination with the weak correlation with activity-level at the time suggest that this design of assay for sparrows did not access individual differences in foraging innovation and the use of novel foods (Reader and Laland 2003). Although the intra-trial covariances were weaker, especially within the 'novel object'- and 'novel food'-trials, the activity-driven model (H5) was significantly better than the activity-only model (H3), suggesting that little other than individual differences in activity were driving the behavioural syndrome here. Therefore, it seems likely that the assays used here only really tested activity, and did not reveal true individual variation that might exist in house sparrow exploration, neophilia or innovation. The small cages, the novel objects and the novel food may not have been 'novel' enough for these types of organisms. The type of behavioural assay that was used in this study is a common method for examining behaviour such as exploration, neophobia/neophilia and innovation. And it has worked in species such as Japanese Quail (Richard et al. 2008) and the Guppy (Berdal 2015). However, testing the presence of animal personality and behavioural syndromes has never been done in this system before,

and our results suggest that these behavioural assays are not appropriate. Perhaps different results would be achieved using tests that more suitably represent novel environments, objects and food sources that these birds perceive as worth exploring, avoiding and attempting to consume.

Careau et al. (2008) argued that physiological constraints may underlie behavioural syndromes, seeing as behaviours such as activity should be energetically costly. I expected a negative or a positive correlation between activity and basal metabolic rate (BMR) depending on whether house sparrows follow the allocation model or the performance model, respectively (Careau et al. 2008). However, there was no behaviour-and-physiology syndrome involving a link between metabolism and personality in these house sparrow populations. This was unexpected based on current theory (Careau et al. 2008), but there are no known studies that connect basal metabolic rate in adult female and male birds with activity-level to compare with. The allocation model (Fig.1) could still be the rule here if the amount of available energy varies between individuals. The BMR measurement process, followed by equal access to food *ad libitum* and then starvation over night before the behavioural assays were thought to standardize the state of all individuals. Thus, the amount of available energy should have been equal for all individuals. However, how much and at what time each individual ate the day before the behavioural assays is unknown, hence some may have been starved more than others. Without knowing the amount of energy available to each individual, this relationship effectively becomes undetectable in studies such as this.

There was an increase in BMR with an increase in absolute mass, which was expected as this relationship has been found repeatedly (Careau et al. 2008, Wiersma et al. 2007). Although there was no metabolism-personality link in these house sparrows, the physiology-and-behaviour model (H5_4) indicates that 'minimum distance to novel food' was influenced by relative mass. The lighter individuals (relative to themselves) came closer to the novel food than heavier individuals. Given this effect of relative mass, and the fact that 'minimum distance to novel food' was not repeatable, what we think of as 'innovation' (i.e., utilizing novel food source, Reader and Laland 2003) in these house sparrows may be a behaviour partially driven by hunger-level/state (i.e., not 'animal personality'). There were no noteworthy changes in SEM model rankings or model parameters with the use of flock- and sex-residuals, indicating that the syndrome structures are driven by true

differences at the individual level and not due to differences between the sexes, or flocks. However, the AIC based behaviour-and-physiology model ranking changed for males only, where model H5_4 was no better than the null model. These results appear contradictory, but may be a result of residuals removing by chance some crucial amount of the otherwise quite weak between-individual effects, and/or of the small sample size of males only (n=102). The reduction in sample size may have led to the destabilization of our behaviour-and-physiology models. Although ~100 is considered a large sample size in terms of more common statistical analyses (i.e., ANOVAs), many researchers suggest a minimum sample size of 200 to prevent unstable SEM results (Kline 1998). The fact that the null model was equal to the previous best physiology-and-behaviour model, means that there is no demonstrable behaviour-and-physiology syndrome involving the physiological features mass and BMR in males.

Once male-specific behaviour-and-physiology syndromes were tested alongside the morphological variables, two equally good models (M2A or M3A) became significantly better than the null model, thus suggesting that there is a syndrome in male house sparrows. These models consist of activity-driven behaviours (i.e., H5_4) and status signals. The latent variable 'status' captured the positive allometric correlations one might expect among mask length, badge size and category, and beak coloration (Pélabon et al. 2014). Conversely, Bókony et al. (2006) found that house sparrow body size was unrelated to badge and wingbar coloration. Even though it was unclear whether it was only the coloration, or the size of the morphological traits that were unrelated to body size as well, this supports that at least badge category is independent of body size. Jensen et al. (2004) found that badge size increases with age, and that it is an important determinant of life time reproductive success in male house sparrows (extra-pair young included). After an extensive meta-analysis on house sparrow badge size, Nakagawa et al. (2007) concluded that badge size first and foremost signals dominance, but perhaps also age and body condition. The age-badge size effect is reflected in our syndrome, where the 'status' of a male is positively correlated with age, badge size, and badge category. There is also a positive correlation with mask length, and a weaker correlation with beak category. Although most studies that involve badge of status only consider the throat patch of a house sparrow (Møller 1987, Veiga 1993, Gonzalez et al. 1999, Evans et al. 2000, Liker and Barta 2001, McGraw et al. 2003,

Nakagawa et al. 2007), it may be more appropriate to consider variation in multiple plumage patches (i.e., throat patch *and* mask length). Kingma et al. (2008) found that mask size in male Eurasian penduline tits (*Remiz pendulinus*) signals male quality as it increased with size and age. Our 'status'-syndrome may indicate that house sparrow "Badge of Status" consists of multiple phenotypic traits simultaneously. Multiple plumage characteristics may function as multiple messages, either working together to reflect individual quality, or as backup signals that facilitate accurate assessments of the actual quality trait (Candolin 2003). Thus, badge size along with one or more of the morphological house sparrow traits mask length, beak coloration and badge category could convey male quality. They could be part of a "multiple badges of status", or may facilitate the detection and/or assessment of badge quality/size (Bòkony et al. 2006).

There was no indication of any correlation between BMR and Activity, or between BMR and Status. Both correlations were weak and non-significant within the SEMs, resulting in two statistically equal 'best models'. It was surprising that BMR did not correlate with activity, nor with energy-costly male coloration. The latter is in line with Laucht et al. (2010), who found no relationship between testosterone and badge size in house sparrows, suggesting that this relationship is only be detectable at the time of autumn moult. Likewise, Evans et al. (2000) found a relationship between testosterone level and badge size during breeding, but not post-breeding. Thus it may be of considerable interest to investigate the metabolism-status correlation in house sparrows during the fall, rather than in February-March when the present study was conducted. Apart from the positive correlation between body mass and BMR, the only other interesting effect on BMR was flock, accounting for as much as 26% of the variation in BMR. I expected larger differences between islands than between flocks, given past selection experiments on Leka, and a common-garden experiment on Lauvøya. There was selection for low BMR at Leka, which explains why the average BMR was lower at Leka than at Vikna. However, BMR at Leka did not differ from Lauvøya. The collected data that reveals which of the 140 individuals of high and low BMR that were released at Lauvøya in 2012 survived and reproduced has not yet been analysed. However, there was a larger range of BMR at Lauvøya (53.8-105.6) than at Leka (58.7-100.9) and Vikna (60.1-101.7). This suggests that the hybrid-population that has resulted from a mixture of extreme phenotypes has maintained the full range of BMR phenotypes. Thus, the average

BMR at Lauvøya does not differ from the other islands. The between flock differences within each island (Fig.9) was surprising, and cannot be explained by island-specific experiments. These between-flock difference in BMR cannot be a result of genetic drift creating local differences between these small sub-populations, because there was no individual repeatability in BMR from 2014 to 2015 (Table 4). This is based on the notion that repeatability often gives an indication of a trait's heritability (Falconer and Mackay 1997). Long-term repeatability in BMR in captive zebra finches (*Taeniopygia guttata*) has been shown (Rønning et al. 2005). However, there may be environmental conditions that affect an individual's metabolism, either adaptively or otherwise, that are not detected in a laboratory setting (e.g., an infection, Chappell et al. 1996). Fig. 9 shows fewer differences between 'flocks' on Lauvøya, as compared to the between-flock differences on Vikna and Leka, despite the fact that the Lauvøya population was composed of a mix of birds from Leka and another island. The different 'flocks' on Lauvøya were actually part of one large flock resident on the same farm, which may give some indication as to why the major differences in BMR were between flocks on Leka and Lauvøya. The metabolic rate of great tits (*Parus major*) has been shown to be locally adapted to respond to environmental conditions (Broggi et al. 2004, Broggi et al. 2005). Perhaps the between-farm differences in BMR on Vikna and Leka were due to flock-specific environmental effects on individual physiological phenotypes, such as differences in flock density, food, infections and/or capture stress.

A pace-of-life syndrome (PoLS) assumes covariation of life history, physiology, and behavioural traits (Reale et al. 2010b). I did not test a true pace-of-life hypothesis in this study, because I did not have access to natural life history trait variation for these birds, such as reproductive success (number of recruits) and lifespans. Therefore, I may have tested only part of the traits in the PoLS concept, without examining its foundation in life history variation. Blood was sampled from all captured birds on these island, thus data regarding reproductive success can perhaps be examined in future studies. However, the natural population structures have been disrupted for most of the birds (or their parents), either by the BMR selection experiment on Leka, or by translocation to Lauvøya. Therefore, these life history data may not reveal the sorts of stable conditions expected of the PoLS, but rather populations in a state of ecological and evolutionary flux.

As this was largely a pilot study of these methods on a new species, the results of this study would suggest that future behavioural studies on house sparrows should consider: (1) measure basal metabolic rate and male plumage (badge size and mask length) to compare with individual behavioural differences (and their hormonal drivers) during the fall; (2) collect a larger sample size for the estimation of repeatability, and have a minimum of four measures per individual; (3) design more natural behavioural assays that measure behaviours that are better proxies of an individual's real levels of exploration, neophobia and innovation in the wild (i.e., leave the captive open field test behind); (4) collect a larger sample size in total to allow more comprehensive structural equation modelling of the sexes separately without risking unstable results; and (5) investigate if there is a "multiple badges of status" in house sparrows that affects a male's life-time reproductive success (LRS).

In conclusion, our findings indicate that there is activity animal personality in house sparrow populations on three islands off of the northern coast of Norway. I failed to identify additional behaviours associated with exploration, neophilia and innovation in a standard small-cage behaviour assay test, suggesting that more sophisticated methods are needed for birds that live in close association with humans. There was no link between Activity driven individual behavioural variation and metabolism as predicted by the pace-of-life-syndrome. Despite not being individually repeatable, basal metabolic rate did vary with body mass and between social groups, suggesting some meaningful phenotypic plasticity in this physiological trait. There was a male-specific Status syndrome, perhaps conveying male quality in the sense of the well-studied badge-of-status individual variation in house sparrows. However, there were no links to physiological or behavioural traits, as expected from a male PoLS. Therefore, it appears that there was substantial within-individual plasticity in many of the traits studied here, and that this needs to be studied further and understood before we can explain many of the individual differences that might constitute any PoLS in house sparrows.

References

Altwegg, R., Ringsby, T.H., Sæther, B.E. Phenotypic correlates and consequences of dispersal in a metapopulation of house sparrows Passer domesticus. *J. Anim. Ecol.*, **2000**, 69(5): 762-770.

Anderson, T.R. Chapter 5: Plumages and Molt: Badge Size in Males. Biology of the Ubiquitous House Sparrow: From Genes to Populations. Oxford University press, **2006**, pp 224-445.

Baalsrud, H.T., Saether, B.E., Hagen, I.J., Myhre, A.M., Ringsby, T.H., Parn, H., Jensen, H. Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation. *Mol. Ecol*., **2014**, 23(11): 2653-2668.

Barnett, S.A. Experiments on 'neophobia' in wild and laboratory rats*. Brit. J. Psychol.*, **1958**, 49(3): 195-201.

Bell, A.M. Future directions in behavioural syndromes research. *Proc. Biol. Sci.,* **2007**, 274(1611): 755–761.

Bell, A.M., Hankison, S.J., Laskowski, K.L. The repeatability of behaviour: a meta-analysis. *Anim. Behav.*, **2009**, 77(4): 771-783.

Berdal, M.A. Distinguishing animal personality and innovation in the guppy-the effect of sex and body size. MSc Thesis, Norwegian University of Science and Technology, **2015**.

Bókony, V., Lendvai, Á.Z., Liker, A. Multiple cues in status signalling: the role of wingbars in aggressive interactions of male house sparrows. *Ethol*., **2006**, 112(10): 947-954.

Broggi, J., Hohtola, E., Orell, M. Nilsson, J.Å. Local adaptation to winter conditions in a passerine spreading north: a common-garden approach. *Evol*., **2005**, 59(7): 1600-1603.

Broggi, J., Orell, M., Hohtola, E. Nilsson, J.Å. Metabolic response to temperature variation in the great tit: an interpopulation comparison. *J. Anim. Ecol*., **2004**, 73(5): 967-972.

Brown, T.A. Confirmatory factor analysis for applied research. Guilford Publications, **2013**, pp 462.

Buchanan, K.L., Evans, M.R., Goldsmith, A.R., Bryant, D.M., Rowe, L.V. Testosterone influences basal metabolic rate in male house sparrows: a new cost of dominance signaling? *Proc. R. Soc. Lond. Biol*., **2001**, 268: 1337 – 1344.

Buttemer, W.A., Warne, S., Bech, C., Astheimer, L.B. Testosterone effects on avian basal metabolic rate and aerobic performance: Facts and artefacts. Comp. Biochem. *Physiol., Part A: Mol. Integr. Physiol.,* **2008**, 150(2): 204-210.

Candolin, U., The use of multiple cues in mate choice. *Biol. Rev.*, **2003**, 78(4): 575-595.

Careau, V., Garland, T. Performance, personality, and energetics: correlation, causation, and mechanism. *Physiol. Biochem. Zool*., **2012**, 85(6): 543–571.

Careau, V., Thomas, D., Humpries, M.M., Reale, D. Energy metabolism and animal personality. *Oikos,* **2008**, 117: 641-653.

Chappell, M.A., Zuk, M. and Johnsen, T.S. Repeatability of aerobic performance in red junglefowl: effects of ontogeny and nematode infection. *Funct. Ecol.*, **1996**: 578-585.

Dingemanse, N.J., Both, C., Drent, P.J., Van Oers, K., Van Noordwijk, A.J. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Anim. Behav.*, **2002**, 64(6): 929-938.

Dingemanse, N.J., Both, C., Van Noordwijk, A.J., Rutten, A.L., Drent, P.J. Natal dispersal and personalities in great tits (*Parus major*). *Proc. Roy. Soc. Lond. B. Biol.*, **2003**, 270(1516): 741-747.
Dingemanse, N.J., Dochtermann, N.A. Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J. Anim. Ecol.*, **2013**, 82(1), pp.39-54.
Dingemanse, N.J., Dochtermann, N., Wright, J. A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Anim. Behav.*, **2010a**, 79(2): 439-450.
Dingemanse, N.J., Kazem, A.J., Réale, D., Wright, J. Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol. Evol.*, **2010b**, 25(2): 81-89.
Dingemanse, N., Wright, J., Kazem, A.J.N., Thomas, D.K., Hickling, R., Dawnay, N. Behavioural syndromes differ predictably between 12 populations of three-spined sticklebacks*. J. Anim. Ecol.,* **2007**, 76: 1128-1138.
Engen, S., Ringsby, T.H., Saether, B.E., Lande, R., Jensen, H., Lillegard M., Ellegren, H. Effective size of fluctuating populations with two sexes and overlapping generations. *Evol.*, **2007**, 61(8): 1873-1885.
Evans, M.R., Goldsmith, A.R., Norris, S.R.A. The effects of testosterone on antibody production and plumage coloration in male house sparrows (*Passer domesticus*). *Behav. Ecol. Siciobiol*., **2000**, 47 (3): 156-163.
Falconer, D. S., Mackay, T.F.C. Introduction to Quantitative Genetics (4th Ed.). Edinburgh: Longman. **1997**, pp 480.
Gonzalez, G., Sorci, G., de Lope, F. Seasonal variation in the relationship between cellular immune response and badge size in male house sparrows (*Passer domesticus*). *Behav. Ecol. Sociobiol.*, **1999**, 46(2): 117-122.
Hennemann III, W.W. Relationship among body mass, metabolic rate, and the intrinsic rate of natural increase in mammals. *Oecol.*, **1983**, 56: 104-108.
Holand, H., Jensen, H., Tufto, J., Soliman, M., Pärn, H., Saether, B.E., Ringsby, T.H. Lower survival probability of house sparrows severely infected by the gapeworm parasite. *J. Avian Biol*., **2014**, 45(4): 365-373.
Hughes, R.N. Intrinsic exploration in animals: motives and measurement. *Behav. Processes*, **1997**, 41(3): 213-226.
Hadfield, J.D. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J. Stat. Softw.*, **2010**, 33(2), 1-22.
Jensen, H., Saether, B.E., Ringsby, T.H., Tufto, J., Griffith, S.C., Ellegren, H. Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (*Passer domesticus*). *J. Evol. Biol*., **2003**, 16(6): 1296-1307.
Jensen, H., Saether, B.E., Ringsby, T.H., Tufto, J., Griffith, S.C., Ellegren, H. Lifetime Reproductive Success in Relation to Morphology in the House Sparrow (*Passer domesticus*). *J. Anim. Ecol*., **2004**, 73(4): 599-611.
Jensen, H., Svorkmo-Lundberg, T., Harald Ringsby, T., Sæther, B.E. Environmental influence and cohort effects in a sexual ornament in the house sparrow, Passer domesticus. *Oikos*, **2006**, 114(2): 212-224.
Johnstone, R.A., Norris, K. Badges of status and the cost of aggression. *Behav. Ecol. Sociobio.*, **1993**, 32(2): 127-134.
Jones, O.R., Gaillard, J.M., Tuljapurkar, S., Alho, J.S., Armitage, K.B., Becker, P.H., Bize, P., Brommer, J., Charmantier, A., Charpentier, M., Clutton-Brock, T. Senescence rates are determined by ranking on the fast-slow life-history continuum. *Ecol. Lett*., **2008**, 11: 664 – 673.

Kingma, S.A., Szentirmai, I., Székely, T., Bókony, V., Bleeker, M., Liker, A., Komdeur, J. Sexual selection and the function of a melanin-based plumage ornament in polygamous penduline tits (*Remiz pendulinus*). *Behav. Ecol. Sociobio.*, **2008**, 62(8): 1277-1288.
Kline, R.B. Principles and practice of structural equation modeling. New York: Guilford, **1998**, pp 354.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. **2015**. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. http://CRAN.R-project.org/package=lmerTest
Laucht, S., Kempenaers, B., Dale, J. Bill color, not badge size, indicates testosterone-related information in House Sparrows. *Behav. Ecol. Sociobiol.,* **2010**, 64(9): 1461-1471.
Liker, A., Barta, Z. Male Badge Size Predicts Dominance Against Females in House Sparrows. *Condor*, **2001**, 103: 151-157.
MacArthur, R.H., Wilson, E.O. The Theory of Island Biogeography. Princeton Univeristy Press, **1967**.
McGraw, K.J., Dale, J., Mackillop, E.A. Social environment during molt and the expression of melanin-based plumage pigmentation in male house sparrows (*Passer domesticus*). *Behav. Ecol. Sociobio.*, **2003**, 53(2): 116-122.
Møller, A.P. Variation in badge size in male house sparrows Passer domesticus: evidence for status signalling. *Anim. Behav.*, **1987**, 35(6): 1637-1644.
Nakagawa, S., Ockendon, N., Gillespie, D.O.S., Hatchwell, B.J., Burke, T. Assessing the function of house sparrows' bib size using a flexible meta-analysis method. *Behav. Ecol.*, **2007**, 18(5): 831-840.
Noldus, L.P.J.J., Spink A.J., and Tegelenbosch R.A.J. 2001. EthoVision: A versatile video tracking system for automation of behavioural experiments. *Behav. Res. Methods Instrum. Comput.*, **2001**, 33 (3): 398-414.
Pärn, H., Jensen, H., Ringsby, T.H., Sæther, B.E. Sex-specific fitness correlates of dispersal in a house sparrow metapopulation. *J. Anim. Ecol.*, **2009**, 78(6): 1216-1225.
Pärn, H., Ringsby, T.H., Jensen, H., Sæther, B.E. Spatial heterogeneity in the effects of climate and density-dependence on dispersal in a house sparrow metapopulation. *Proc. Roy. Soc. Lond. B. Biol.*, **2012**, 279(1726): 144-152.
Pélabon, C., Firmat, C., Bolstad, G.H., Voje, K.L., Houle, D., Cassara, J., Rouzic, A.L., Hansen, T.F. Evolution of morphological allometry. *Ann. N.Y. Acad. Sci.*, **2014**, 1320(1): 58-75.
Plummer, M., Best, N., Cowles, K., Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News*, **2006**, 6: 7-11.
R Core Team. **2013**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
Reale, D., Dingemanse, N.J., Kazem, A.J.N., Wright, J. Evolutionary and ecological approaches to the study of personality. *Phil. Trans. R. Soc. B.*, **2010a**, 365: 3937-3946.
Reale, D., Garant, D., Humphries, M.M., Bergeron, P., Careau, V., Montiglio, P-O. Personality and the emergence of the pace-of-life syndrome concept at the population level. *Phil. Trans. R. Soc. B*., **2010b**, 365: 4051-4063.
Reale, D., Martin, J., Coltman, D.W., Poissant, J. Festa-Bianchet, M. Male personality, life-history strategies and reproductive success in a promiscuous mammal. *J. Evol. Biol*., **2009**, 22: 1599–1607.

Reale, D., Reader, S.M., Sol, D., McDougall, P.T., Dingemanse, N.J. Integrating animal temperament within ecology and evolution. *Biol. Rev.*, **2007**, 82: 291–318.
Renner, M.J. Neglected aspects of exploratory and investigatory behavior. *Psychobiol.*, **1990**, 18(1): 16-22.
Richard, S., Wacrenier-Ceré, N., Hazard, D., Saint-Dizier, H., Arnould, C., Faure, J.M. Behavioural and endocrine fear responses in Japanese quail upon presentation of a novel object in the home cage. *Behav. Process.*, **2008**, 77(3): 313-319.
Ringsby, T.H., Berge, T., Saether, B.E., Jensen, H. Reproductive success and individual variation in feeding frequency of House Sparrows (*Passer domesticus*). *J. Ornithol.*, **2009**, 150(2): 469-481.
Ringsby, T.H., Saether, B.E., Jensen, H., Engen, S. Demographic characteristics of extinction in a small, insular population of house sparrows in northern Norway. *Conserv. Biol.*, **2006**, 20(6): 1761-1767.
Ringsby, T.H., Saether, B.E., Tufto, J., Jensen, H., Solberg, E.J. Asynchronous spatiotemporal demography of a house sparrow metapopulation in a correlated environment. *Ecol.*, **2002**, 83(2): 561-569.
Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.*, **2012**, 48(2): 1-36.
Rønning, B., Broggi, J., Bech, C., Moe, B., Ringsby, T.H., Pärn, H., Hagen, I.J., Sæther, B.E., Jensen, H. Is basal metabolic rate associated with recruit production and survival in free-living house sparrows? *Funct. Ecol.*, **2015**, 11(1).
Rønning, B., Moe, B., Bech, C. Long-term repeatability makes basal metabolic rate a likely heritable trait in the zebra finch (*Taeniopygia guttata*). *J. Exp. Biol.*, **2005**, 208(24): 4663-4669.
Satorra, A., Bentler, P.M. Scaling corrections for chi-square statistics in covariance structure analysis. Proceedings of the Business and Economic Statistics Section (Vol. 1). American Statistical Analysis, **1988**, pp 308-313.
Satorra, A., Bentler, P.M. Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis. Latent Variables Analysis: Applications for Developmental Research. Sage, **1994**, pp. 399–419.
Sih, A., Bell, A.M. Insights for behavioural ecology from behavioural syndromes. *Adv. Stud. Behav.*, **2008**, 38: 227– 281.
Sih, A., Bell, A.M., Johnson, J.C. Ziemba, R.E. Behavioural syndromes: an integrative overview. *Q. Rev. Biol.*, **2004**, 79: 241–277.
Solberg, E.J., Ringsby, T.H. Does male badge size signal status in small island populations of house sparrows, *Passer domesticus*?. *Ethol.*, **1997**, 103(3): 177-186.
Soma, K.K. Testosterone and aggression: Berthold, birds and beyond. *J. neuroendocrinol.*, **2006**, 18(7): 543-551.
Trainor, B.C., Bird, I.M., Marler, C.A. Opposing hormonal mechanisms of aggression revealed through short-lived testosterone manipulations and multiple winning experiences. *Horm. Behav.*, **2004**, 45(2): 115-121.
Van Oers, K., Drent, P.J., De Goede, P., Van Noordwijk, A.J. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proc. Roy. Soc. Lond. B. Biol.,* **2004**, 271(1534): 65-73.
Vedder, O., Schut, E., Magrath, M.J.L., Komdeur, J. Ultraviolet crown coloration affects contest outcomes among blue tits, but only in the absence of prior encounters. *Funct. Ecol.*, **2010**, 2: 417-425.

Veiga, J.P. Badge size, phenotypic quality, and reproductive success in the house sparrow: a study on honest advertisement. *Evol*., **1993**: 1161-1170.
Verbeek, M.E., Drent, P.J., Wiepkema, P.R. Consistent individual differences in early exploratory behaviour of male great tits. *Anim. Behav*., **1994**, 48(5): 1113-1121.
Walsh, R.N., Cummins, R.A. The open-field test: A critical review. *Psychol. Bull*., **1976**, 83(3): 482-504.
Wiersma, P., Munoz-Garcia, A., Walker, A., Williams, J.B. Tropical birds have a slow pace of life. *PNAS*, **2007**, 104(22): 9340-9345.
Wikelski, M., Lynn, S., Breuner, C., Wingfield, J.C. Energy metabolism, testosterone, and corticosterone in white-crowned sparrows. *J. Comp. Physiol. A Sens. Neural. Behav. Physiol*., **1999**, 185: 463-470.
Withers, P.C. Measurements of VO2, VCO2 and evaporative water loss with a flowthrough mask. *J. Appl. Physiol*., **1977**, 42: 120-123.
Whiting, M.J., Nagy, K.A., Bateman, P.W. Evolution and maintenance of social status-signaling badges. In Fox et al. (Ed.): Lizard social behaviour. The John Hopkins University Press, **2003**, pp. 47-82.

Appendix A

Table 9. The effects of assay order, novel object type and novel food type on the six behaviours. The effect degrees of freedom are 1 for all models, and the error degrees of freedom are provided.

	expl_time			**no_distance**			**nf_distance**		
	df	F	P	df	F	P	df	F	P
Order	35	0.515	0.478	34	0.137	0.713	34	2.65	0.113
Object Type	-	-	-	34	4.320	**0.045**	-	-	-
Food Type	-	-	-	-	-	-	34	0.293	0.592
Order:Object Type	-	-	-	34	0.214	0.646	-	-	-
Order:Food Type	-	-	-	-	-	-	34	0.505	0.482
	activity1			**activity2**			**activity3**		
	df	F	P	df	F	P	df	F	P
Order	45.92	8.46	**0.006**	35.52	5.727	**0.023**	33.78	7.15	**0.011**
Object Type	-	-	-	35.52	0.110	0.742	-	-	-
Food Type	-	-	-	-	-	-	33.78	0.014	0.906
Order:Object Type	-	-	-	34	1.662	0.206	-	-	-
Order:Food Type	-	-	-	-	-	-	34	0.592	0.447

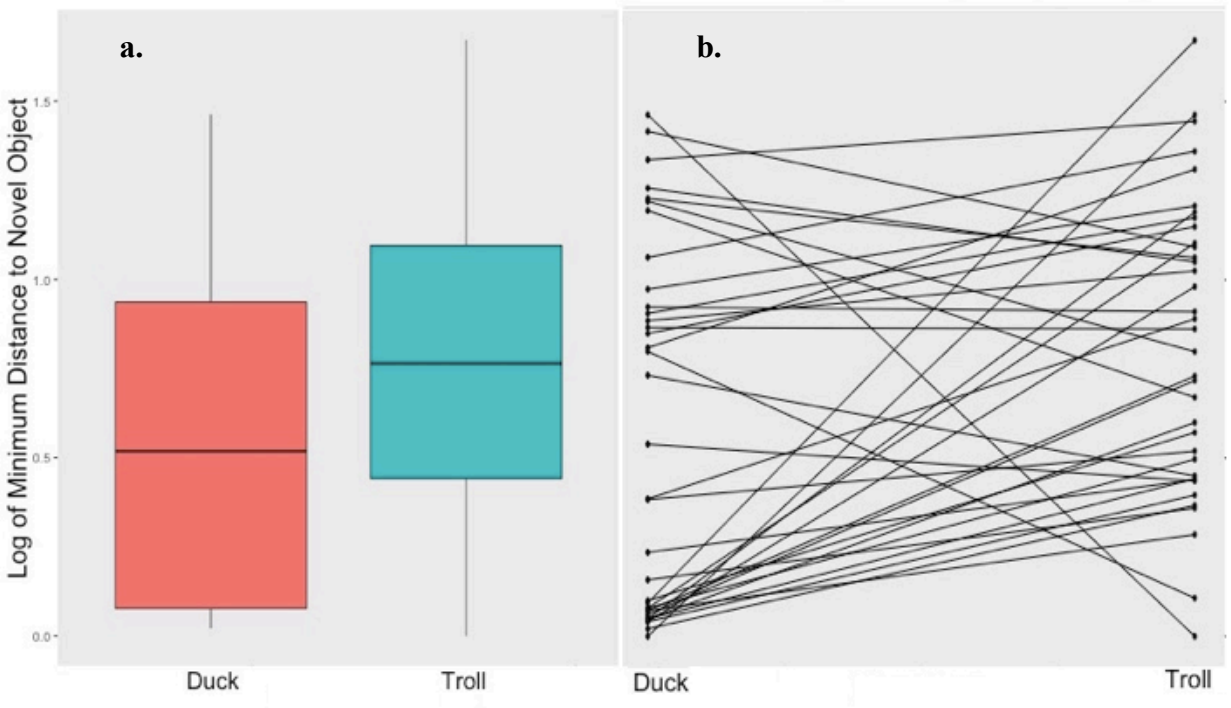

Fig. 14. The effect of novel item type on the behaviour 'minimum distance to novel object' (p=0.045) illustrated using (a) boxplots, and (b) individual reaction norms. Individuals came closer to the 'duck' (β=0.599±0.094) than the 'troll (β=0.794±0.090).

Table 10. The resulting model ranking of sex- and flock-residual behaviour models H0, H1, H2A, H2B, H3, H4 and H5 (n = 190) based on Akaike's Information Criterion (AIC). Degrees of freedom (d.f.) for each model is provided.

		'Sex'		'Flock'	
Model	**d.f**	**ΔAIC**	**CFI**	**ΔAIC**	**CFI**
H5	18	0.00	0.864	0.00	0.884
H2B	19	6.59	0.850	6.84	0.868
H2A	18	6.98	0.849	6.97	0.867
H1	18	25.42	0.763	24.29	0.795
H3	15	36.75	0.705	36.76	0.741
H4	15	182.97	0.164	189.99	0.144
H0	12	212.74	0.000	219.77	0.000

Table 11. The resulting model ranking of sex- and flock-residual behaviour-and-physiology models H5_0, H5_1, H5_2, H5_3, and H5_4 (n = 190) based on Akaike's Information Criterion (AIC). Degrees of freedom (d.f.) for each model is provided.

		'Sex'		'Flock'	
Model	**d.f**	**ΔAIC**	**CFI**	**ΔAIC**	**CFI**
H5_4	27	0.00	0.768	0.00	0.788
H5_0	25	6.35	0.742	3.83	0.768
H5_2	27	10.04	0.741	7.02	0.769
H5_1	28	10.79	0.743	6.98	0.773
H5_3	26	57.80	0.595	35.28	0.681

Table 12. The resulting model ranking of male-specific behaviour models H0, H1, H3 and H5 (n = 110) based on Akaike's Information Criterion (AIC). See Methods for detailed model descriptions.

		Raw Data		'Flock'	
Model	**d.f.**	**ΔAIC**	**CFI**	**ΔAIC**	**CFI**
H5 – Activity Driven	18	0.00	0.883	0.00	0.901
H3 – Activity	15	16.98	0.758	20.59	0.777
H1 – one behaviour syndrome	18	17.39	0.779	19.64	0.814
H0 – Null Model	12	134.00	0.000	152.83	0.000

Table 13. The resulting model ranking of male-specific behaviour-and-physiology models H5_0, H5_1, H5_2, H5_3, and H5_4 (n = 102) based on Akaike's Information Criterion (AIC). See Methods for detailed model descriptions.

		Raw Data		'Flock'	
Model	**d.f.**	**ΔAIC**	**CFI**	**ΔAIC**	**CFI**
H5_4	27	0.00	0.790	0.53	0.789
H5_0	25	1.59	0.775	0.01	0.787
H5_1	28	1.84	0.798	0.00	0.807
H5_3	26	3.47	0.775	2.00	0.786
H5_2	27	4.80	0.776	3.67	0.785

Appendix B

Table 14. Male-specific pair-wise correlations between the six behaviours: time to explore cage (expl_t), minimum distance to novel object (no_dist), minimum distance to novel food (nf_dist), activity1, activity2 and activity3, and the morphological/physiological measures visible badge size (badgeS), badge category (badgeC), beak coloration (beakC), mask length, minimum age, basal metabolic rate (BMR), pre-assay mass, and relative mass. Correlations between behaviours derived from the same trial are marked with squares, and all significant correlations are marked in bold.

	BMR	**pre-mass**	**rel-mass**	**expl_t**	**no_dist**	**nf_dist**	**activity1**
BMR		**r=0.58**	r<0.01	r< -0.01	r=0.08	r=0.08	r<0.01
		p<0.001	p=0.97	p=0.96	p=0.42	p=0.40	p=0.98
pre-mass			**r=0.37**	r=0.07	**r=0.20**	r=0.14	r= -0.04
			p<0.001	p=0.48	**p=0.04**	p=0.14	p=0.90
rel-mass				r=0.09	r=0.06	**r=0.20**	r= -0.15
				p=0.34	p=0.54	**p=0.03**	p=0.12
expl_t					r= -0.03	**r=0.20**	**r=0.33**
					p=0.75	**p=0.04**	**p<0.001**
no_dist						**r=0.36**	r=0.08
						p<0.001	p=0.40
nf_dist							r=0.03
							p=0.73
activity1							

	activity2	**activity3**	**badgeS**	**badgeC**	**mask**	**age**	**beakC**
BMR	r=0.01	r=0.01	r=0.15	r= -0.08	r=0.05	r=0.05	r=0.01
	p=0.88	p=0.89	p=0.14	p=0.45	p=0.61	p=0.58	p=0.88
pre-mass	r=0.06	r=0.08	r=0.13	r=0.04	r=0.07	r= 0.17	r=0.05
	p=0.54	p=0.43	p=0.17	p=0.69	p=0.47	p=0.09	p=0.64
rel-mass	**r= -0.25**	r=-0.16	r=0.04	r=0.01	r= -0.05	r=0.04	r=0.10
	p=0.007	p=0.09	p=0.68	p=0.92	p=0.64	p=0.69	p=0.33
expl_t	r= -0.07	r=-0.14	r=0.17	r= 0.17	r= 0.14	**r=0.25**	r=0.13
	p=0.45	p=0.14	p=0.19	p=0.08	p=0.16	**p=0.01**	p=0.18
no_dist	r= -0.07	r=0.07	r=0.17	r= -0.07	r= 0.04	r=0.03	r= -0.04
	p=0.45	p=0.49	p=0.09	p=0.51	p=0.69	p=0.76	p=0.72
nf_dist	r=0.02	r= -0.11	r=0.16	r= 0.15	r= 0.06	r=0.09	r= 0.05
	p=0.84	p= 0.27	p=0.09	p=0.14	p=0.54	p=0.37	p=0.65
activity1	**r=0.61**	**r=0.62**	r= -0.05	r= 0.03	r= -0.07	r=0.06	r= 0.01
	p<0.001	**p<0.001**	p=0.60	p=0.74	p=0.51	p=0.52	p=0.89
activity2		**r=0.61**	r= -0.02	r= 0.04	r= 0.02	r= -0.14	r= -0.03
		p<0.001	p=0.81	p=0.71	p=0.82	p=0.17	p=0.76
activity3			r<0.01	r= -0.08	r= -0.02	r= -0.09	r= 0.09
			p=0.99	p=0.41	p=0.81	p=0.37	p=0.37
badgeS				**r= 0.37**	**r= 0.23**	r= 0.12	r= 0.03
				p<0.001	**p=0.02**	p=0.22	p=0.74
badgeC					**r= 0.27**	**r= 0.26**	r= 0.07
					p=0.005	**p=0.007**	p=0.49
mask						r= 0.16	r= 0.18
						p=0.11	p=0.10
age							r= 0.17
							p=0.08