Chemical patterns of colony membership and mother-offspring similarity in Antarctic fur seals are reproducible over time

Jonas Tebbe[1], Emily Humble[1,2,3], Martin A. Stoffel[1,4], Lisa J. Tewes[5], Caroline Müller[5*], Jaume Forcada[3], Barbara Caspers[6*] & Joseph I. Hoffman[1,3*]

1 Department of Animal Behaviour, Bielefeld University, 33501 Bielefeld, Germany
2 Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, EH25 9RG, UK
3 British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 OET, UK
4 Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3FL, UK
5 Department of Chemical Ecology, Bielefeld University, 33501 Bielefeld, Germany
6 Department of Behavioural Ecology, Bielefeld University 33501 Bielefeld, Germany
* Jointly supervised this research

Corresponding Author:
Joseph I. Hoffman[1]
Department of Animal Behaviour, Bielefeld University, 33501 Bielefeld, Germany
Email address: joseph.hoffman@uni-bielefeld.de

Abstract

Replication studies are essential for assessing the validity of previous research findings and for probing their generality. However, it has proven challenging to reproduce the results of ecological and evolutionary studies, partly because of the complexity and lability of many of the phenomena being investigated, but also due to small sample sizes, low statistical power and publication bias. Additionally, replication is often considered too difficult in field settings where many factors are beyond the investigator's control and where spatial and temporal dependencies may be strong. We investigated the feasibility of reproducing original research findings in the field of chemical ecology by attempting to replicate a previous study by our team on Antarctic fur seals (*Arctocephalus gazella*). In the original study, skin swabs from 41 mother-offspring pairs from two adjacent breeding colonies on Bird Island, South Georgia, were analysed using gas chromatography-mass spectrometry. Seals from the two colonies differed significantly in their chemical fingerprints, suggesting that colony membership may be chemically encoded, and mothers were also chemically similar to their pups, implying that phenotype matching may be involved in mother-offspring recognition. Here, we generated and analysed comparable chemical data from a non-overlapping sample of 50 mother-offspring pairs from the same two colonies five years later. The original results were corroborated in both hypothesis testing and estimation contexts, with *p*-values remaining highly significant and effect sizes, standardized between studies by bootstrapping the chemical data over individuals, being of comparable magnitude. We furthermore expanded the geographic coverage of our study to include pups from a total of six colonies around Bird Island. Significant chemical differences were observed in the majority of pairwise comparisons, indicating not only that patterns of colony membership persist over time, but also that chemical signatures are colony-specific in general. Our study systematically confirms and extends our previous findings, while also implying that temporal and spatial heterogeneity need not necessarily negate the reproduction and generalization of ecological research findings.

Introduction

Replication studies are fundamental to the scientific process as they are essential for evaluating the correctness of scientific claims and the conclusions of other scientists (Schmidt, 2009). Indeed, Fisher (1974) recommended that a null hypothesis should always be rejected more than once because "no isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon". Nevertheless, replication studies are still "troublingly rare", particularly in fields such as ecology and evolutionary biology (Nakagawa & Parker, 2015). Palmer (2000) argued that we ignore reproducibility at our peril because this perpetuates a "contract of error" that undermines our understanding of important ecological and evolutionary phenomena.

There has also been debate and confusion over exactly what constitutes reproducible research (Mendoza & Garcia, 2017). Goodman and colleagues (2016) recognized three basic concepts, (i) "methods reproducibility", which requires that the methodology of a given study be provided in sufficient detail to allow it to be repeated; (ii) "results reproducibility", often known as "replication", which is the ability to corroborate previous results using the same experimental methods in a new study; and (iii) "inferential reproducibility", which relates to whether or not qualitatively similar conclusions are reached on the basis of either an independent replication of a study or a re-analysis of the original data. Furthermore, replication studies can be "exact", meaning that they show a high degree of fidelity to the original experiment, "partial", which involves some procedural or methodological changes, or "conceptual", where the same questions are investigated but using different approaches (Kelly, 2006). The latter two categories include "quasi-replication" studies, which extend the scope of the original study beyond the specific system or species in question (Palmer, 2000). In general, the closer the replication attempt is to the original study, the more valuable are the results for assessing the validity of the original claims (Nakagawa & Parker, 2015). However, quasi and conceptual replications can also be valuable because they can shed light on the robustness and generality of the biological phenomena under investigation (Goodman, Fanelli & Ioannidis, 2016; Dirnagl, 2019; Piper et al., 2019; although see Kelly, 2006)

Another conceptual difficulty relates to the basis on which replication success is judged. Although there is no single standard for evaluating replication outcomes, most replication attempts are deemed successful if a null hypothesis that was rejected in the original study is again rejected (Rosenthal, 1991; Kelly, 2006). However, due to the dependence of *p*-values on sample sizes, success or failure in attaining significance may not always provide a good measure of replication success (Kelly, 2006). As explained by Goodman (2016), two studies reporting identical effect sizes would produce different *p*-values if their sample sizes were substantially different. If one of these *p*-values was highly significant and the other failed to reach significance, the authors of these studies might reach opposite conclusions, despite the results themselves being mutually corroborative. Consequently, several authors have advocated the reporting of effect sizes and associated measures of precision, as these allow replication outcomes to be gauged in a continuous manner rather than on the basis of binary significance outcomes (Kelly, 2006; Goodman, Fanelli & Ioannidis, 2016; Piper et al., 2019).

In recent years, high-profile failures to reproduce a significant proportion of studies in the medical and social sciences (e.g. Begley & Ellis, 2012; Open science collaboration, 2015, reviewed by Kelly, 2019) have led to a crisis of confidence (Baker, 2016). The generally low success of replication studies has been attributed to a "publish or perish" culture that incentivizes dubious research practices such as selectively reporting significant results, *p*-value hacking and establishing hypotheses after the results of a study are known, or HARKing (Fidler et al., 2017; Fraser et al., 2018). All of these practices increase the risk of false positives and contribute towards publication bias (Jennions & Møller, 2002), which undermines the robustness of the scientific literature. Further issues include poor study design, low statistical power, variability in reagents or the use of specialized techniques that are difficult to repeat, lack of scientific oversight, inadequate reporting of data, methods and results, and insufficient incentives for sharing data and code (Baker, 2016; Fidler et al., 2017; Piper et al., 2019).

Despite growing awareness of these issues not being specific to any particular scientific field, ecological and evolutionary studies are seldom replicated, with only around 0.02% of studies having been self-reported as exact replications (Kelly, 2019). One reason for this may be the general perception that research in these fields can be difficult to replicate, partly due to the complexity and lability of many of the phenomena under investigation, but also because in many field situations replication may be unfeasible or even unethical (Kelly, 2006; Nakagawa & Parker, 2015; Fidler et al., 2017). Furthermore, numerous factors cannot be controlled for in natural settings and environmental variation in particular may confound attempts to reproduce previous results (Kelly, 2006). However, these are not valid reasons to neglect replication studies as it is important to understand the extent to which scientific outcomes hinge upon these and other factors.

The field of chemical ecology provides an interesting case in point. Increasing numbers of studies are using approaches like gas chromatography-mass spectrometry (GC-MS) to characterize the chemical composition of biological samples such as skin swabs or urine. The resulting "chemical fingerprints", otherwise commonly referred to as "chemical profiles", "scent profiles" or "odour profiles" (Hurst & Beynon, 2010), comprise multiple peaks that are separated according to their retention times or mass spectrograms and which represent different substances. Studies of both captive and wild animal populations have shown that these chemical fingerprints can convey information about species identity (Caspers et al., 2009; Fratini et al., 2012; Krause et al., 2014), population membership (Schneeberger et al., 2016; Wierucka et al., 2019), sex, age and reproductive state (Caspers et al., 2011; Kean, Müller & Chadwick, 2011; Vogt et al., 2016), family membership (Sun & Müller-Schwarze, 1998; Müller & Müller, 2016), individual identity (Kean, Chadwick & Mueller, 2015; Kohlwey et al., 2016), social status (Burgener et al., 2009) and genotype (Yamazaki et al., 1990; Charpentier, Boulet & Drea, 2008; Setchell et al., 2011). However, concerns have been raised over the small sample sizes of many studies, which afford little statistical power and may ultimately lead to effect sizes being overestimated of effect sizes (Wyatt, 2015). Furthermore, GC-MS data are inherently noisy and chemical fingerprints are complex and multidimensional, being influenced by a multitude of factors (Hurst & Beynon, 2010; Stoffel et al., 2015) including both intrinsic (e.g. genes, hormones and metabolic status) and extrinsic (e.g. environmental variation and diet) variables. Consequently, it remains unclear to what extent many chemical patterns will be repeatable, particularly under natural and often highly heterogeneous conditions.

Pinnipeds provide interesting model systems for studying chemical communication as they possess large repertoires of functional olfactory receptor genes (Kishida et al., 2007) and are sensitive to even the faintest of smells (Kowalewsky et al., 2006). Many pinnipeds have a strong musky smell (Hamilton, 1956), which has been attributed to facial glands that show hypertrophy during the breeding season (Ling & K, 1974; Hardy et al., 1991), suggesting an important role of olfactory communication during the peak reproductive period. Olfaction may be particularly crucial for mother-offspring recognition because females of many pinniped species accept or reject pups after naso-nasal inspection (Kovacs, 1995; Dobson & Jouventin, 2003; Phillips, 2003). For example, a study of Australian sea lions showed that mothers are capable of discriminating their own pups from nonfilial conspecifics based on odour alone (Pitcher et al., 2011). This discovery motivated our team to perform a study of Antarctic fur seals (*Arctocephalus gazella*), in which chemical fingerprints were characterized from skin swabs taken from 41 mother-offspring pairs at two breeding colonies–the special study beach (SSB) and freshwater beach (FWB)–at Bird Island, South Georgia (Stoffel et al., 2015). Despite being separated by less than 200 m, animals from these two colonies exhibited highly significant chemical differences, while mothers showed greater chemical similarity to their pups than was expected by chance. This finding suggests that mother-offspring recognition in Antarctic fur seals could potentially involve self-referent phenotype matching, a conceptually simple mechanism whereby an individual's own phenotype is used as a template for the recognition of close relatives (Blaustein, 1983).

Here, we attempted to replicate the chemical patterns of colony membership and mother-offspring similarity reported by Stoffel et al. (2015). We returned to the same two breeding colonies five years later, collecting and analysing chemical samples from 50 new mother-offspring pairs using virtually identical methodology. Because these two studies were carried out five years apart, none of the individuals overlapped, precluding analysis of the reproducibility of chemical patterns within individuals. Instead, we use the term "reproducibility" to refer to the extent to which broad chemical patterns, i.e. differences between colonies and similarities between mothers and their offspring, can be replicated with non-overlapping samples from different time points.

In addition, we wanted to know whether chemical differences between animals from SSB and FWB are specific to this particular experimental setting, or whether chemical signatures are colony-specific in general. We therefore collected additional chemical samples from 60 pups originating from a further four colonies around Bird Island to test for the generality of the colony membership pattern, by which we mean the extent to which chemical differences are more generally present among animals from different colonies. We hypothesised that (i) the originally reported patterns of colony membership and mother-offspring similarity would be repeatable; and (ii) that animals from different breeding colonies would differ chemically from one another in general.

Materials & Methods

Study site and fieldwork

Chemical samples were taken from six Antarctic fur seal breeding colonies on Bird Island, South Georgia (54° 00′ S, 38° 02′ W) during the peak of the 2016 breeding season (November to December; the previous study was conducted during the peak of the 2011 breeding season). A total of 50 mother-offspring pairs (including one pair of twins) were sampled from SSB and FWB as part of annual routine procedures of the long-term monitoring and survey program of the British Antarctic Survey (BAS). Additional samples were collected from a total of 60 pups from four colonies (15 samples each from Johnson Cove, Main Bay, Landing Beach and Natural Arch, Figure 1). Here, pups were opportunistically sampled from areas of the beach that were easily accessible. Adult females and pups were captured and restrained on land using standard methodology (Gentry & Holt, 1982). Chemical samples were obtained by rubbing the cheek underneath the eye and behind the snout with sterile cotton wool swabs, which were stored individually at -20°C in glass vials containing approximately 10 mL of 60%/40% (vol/vol) ethanol/water. All of the chemical samples were collected immediately after capture by the same team of experienced field scientists.

GC-MS profiling and data alignment

We first took 2 mL of each sample and allowed the ethanol to evaporate at room temperature for a maximum of 12 hours before resuspending in 2 mL dichloromethane (DCM). After a further evaporation step, in which the DCM was reduced to a final volume of approximately 100 µL, the samples were analysed on a GC with a VF5-MS column (30 m x 0.25 mm inner diameter, 10 m guard column; Agilent Technologies, Santa Clara, USA) connected to a mass spectrometer (GCMS-QP2020, Shimadzu, Kyoto, Japan). One µL of each sample was injected into a deactivated glass-wool-packed liner with an inlet temperature of 225°C. A split ratio of 3.2 was used and the carrier gas (Helium) flow rate was held constant at 1.2 mL/min. The GC run started with three minutes at 60 °C and then ramped up in increments of 10 °C/min to reach a final temperature of 280 °C, which was maintained for 30 minutes. Mass spectra were taken in electron ionization mode with five scans per second in full scan mode (50-600 m/z). The resulting GC-MS data were then processed using OpenChrom (Wenig & Odermatt, 2010) for detection and correction of split peaks. Afterwards, we used GCalignR in R (Ottensmann et al., 2018; R Core Team, 2019) to align the resulting chromatograms by correcting minor shifts in retention times among samples and maximizing the number of shared components.

Data visualisation and statistical analysis

Prior to data analyses, we excluded any compounds that were only observed in a single seal sample. We then used non-metric multidimensional scaling (NMDS) to visualize the chemical data. This approach reduces dimensionality so that each individual data point can be placed in a 2D scatterplot where ranked between-individual distances are preserved and individuals that are chemically more similar are closer together. NMDS was performed on a log(x+1) transformed relative abundance matrix comprising pairwise Bray-Curtis similarity values. We tested for differences among and between *a priori* defined groups (i.e. the breeding colonies and mother-offspring pairs) using a non-parametric permutational multivariate analysis of variance (PERMANOVA). PERMANOVA tests whether the centroids of pre-defined groups differ statistically for a chosen distance measure. It compares within-group to among-group variance components and assigns statistical significance based on random permutations of objects within groups. Each PERMANOVA was based on 99,999 permutations. To determine whether differences between our pre-defined groups were purely attributable to compositional differences between groups rather than compositional differences within groups, we used the ‘betadisper’ function in the vegan package in R to analyse the multivariate homogeneity of group dispersions (Oksanen et al., 2019).

Quantification of the explained variance

To facilitate a comparison of our effect sizes with those reported by Stoffel et al. (2015), we quantified the proportion of the total chemical variance attributable to colony membership and family ID in both studies. We did not re-align the dataframe of Stoffel et al. (2015) as this would have resulted in deviations from the effect sizes of the original study. To standardise effect size estimates between the studies, both chemical datasets were bootstrapped over individuals to generate 5,000 datasets per study, each comprising 15 mother-offspring pairs from SSB and 15 pairs from FWB (i.e. a total of 60 individuals). PERMANOVA was then implemented separately for each dataset and the resulting R^2 values were extracted for each of the predefined groups.

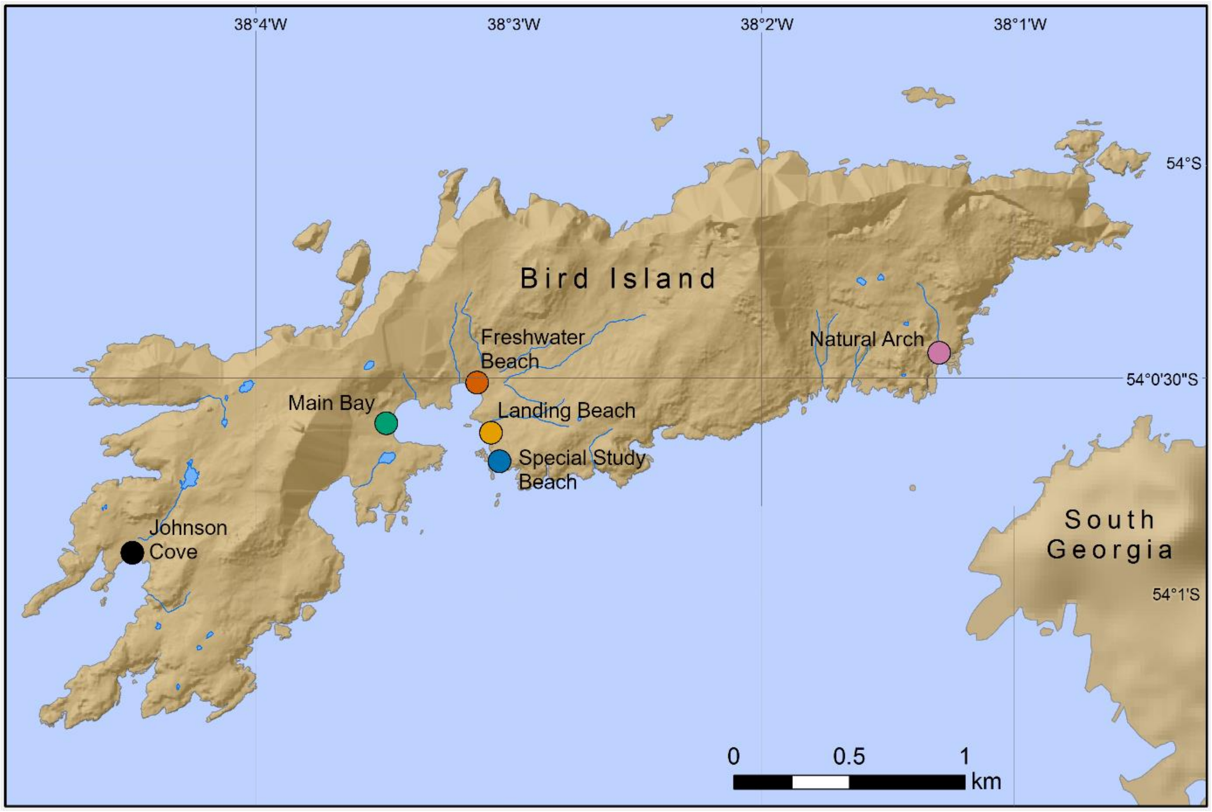
Data availability

The raw chemical data generated during this study are available via Github and the data of Stoffel et al. (2015) can be downloaded from https://github.com/mastoffel/seal_chemical_fingerprints. All of the code used to analyse the raw data are available as a PDF file written in Rmarkdown (see supplementary information). The full documented data analysis pipeline can be downloaded from our GitHub repository at https://github.com/tebbej/SealScent2020/.

Ethical statement

Samples were collected as part of the Polar Science for Planet Earth program of the British Antarctic Survey under the authorization of the Senior Executive and the Environment Officers of the Government of South Georgia and the South Sandwich Islands. Samples were collected and retained under Scientific Research Permits for the British Antarctic Survey field activities on South Georgia, and in accordance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora. All field procedures were approved by the British Antarctic Survey Animal Welfare and Ethics Review Body (reference no. PEA6).

Figure 1: Locations of six Antarctic fur seal breeding colonies on Bird Island, South Georgia, where chemical samples were taken. Mother-offspring pairs were sampled from the special study beach (SSB) and freshwater beach (FWB), whereas only pups were sampled from the other four colonies.



Results

In order to investigate the reproducibility of chemical patterns of colony membership and mother-offspring similarity in Antarctic fur seals, we analysed chemical data from mother-offspring pairs from SSB and FWB as well as pups from an additional four breeding colonies around Bird Island (Figure 1). An average of 42 ± 15 s.d. chemical compounds was detected per sample. No significant differences were found in the number of chemicals between mothers and offspring (unpaired t-test, $t = 0.8403$, $p = 0.403$) or among pups from the six breeding colonies (ANOVA, $F_{5,104}=0.001$, $p = 0.98$).

Reproducibility of chemical patterns

Multivariate statistical analysis of the relative proportions of each substance revealed highly significant differences between animals from SSB and FWB (PERMANOVA, $p < 0.0001$, Figure 2a, Table 1a). We also found a significant effect of mother-pup pair ID nested within colony (PERMANOVA, $p < 0.0001$, Figure 2b, Table 1a), indicating that mothers and their pups are chemically more similar to one another than expected by chance. A test for multivariate homogeneity of group variances uncovered marginally significant differences among the groups ($p = 0.026$, Table 1a), which could potentially indicate the involvement of additional explanatory factors that were not accounted for in the model. However, *post hoc* calculation of Tukey's honest significant differences showed the adjusted p-value to be non-significant.

As p-values cannot be directly compared between studies with different sample sizes, we used the PERMANOVA framework to estimate the effect sizes of colony membership and mother-offspring similarity in both studies. To facilitate direct comparisons while also incorporating uncertainty due to chemical variation among individuals, both datasets were bootstrapped over individuals as described in the Materials and methods. We found that effect size estimates for colony membership and mother-offspring similarity (maximum density R^2 values) differed by only few percent between the two studies (Figure 3) and consistently fell within the range of $0.08 < R^2 < 0.15$.

Figure 2: Two-dimensional non-metric multidimensional scaling (NMDS) plots of chemical data from skin swabs of Antarctic fur seal mother-offspring pairs from SSB and FWB. NDMS was performed using Bray-Curtis similarity values calculated from log(x+1) transformed chemical abundance data. The scales of the two axes are arbitrary and the closer two points appear in the plot, the more similar they are chemically. Individual data points in panel A are colour-coded by colony (SSB = blue, FWB = red) and age (mother = filled, offspring = empty). Panel B shows mother offspring pairs, which are depicted by unique combinations of symbols and colours. The three blue triangles correspond to a mother on SSB with twin pups.

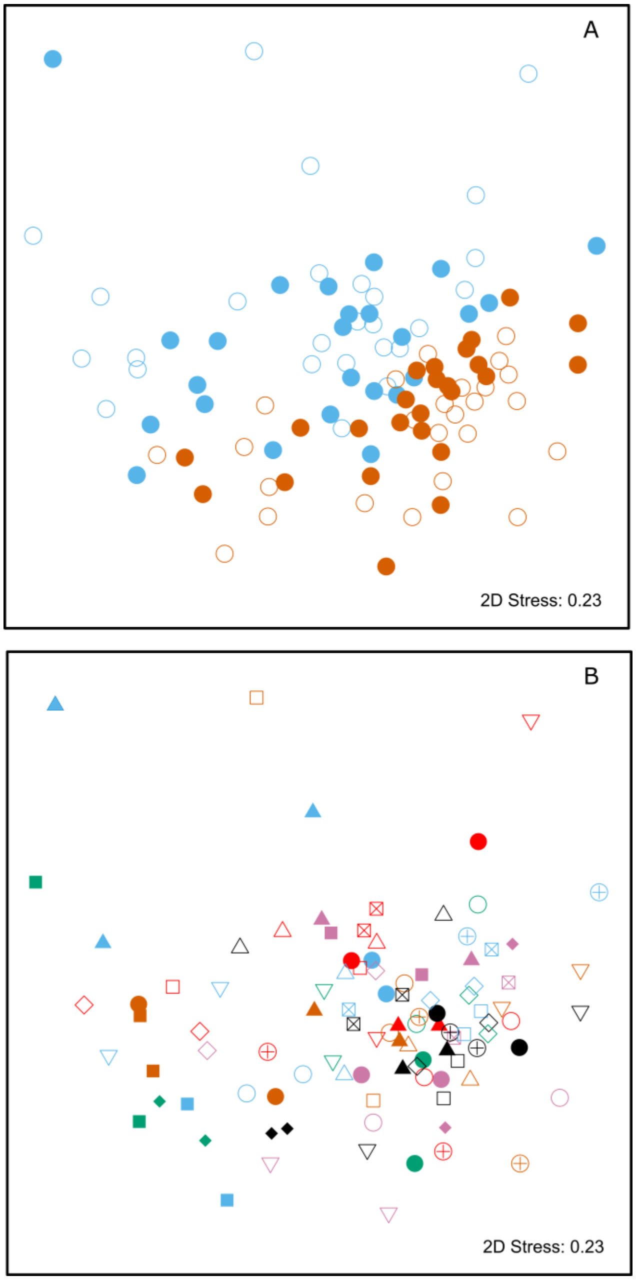
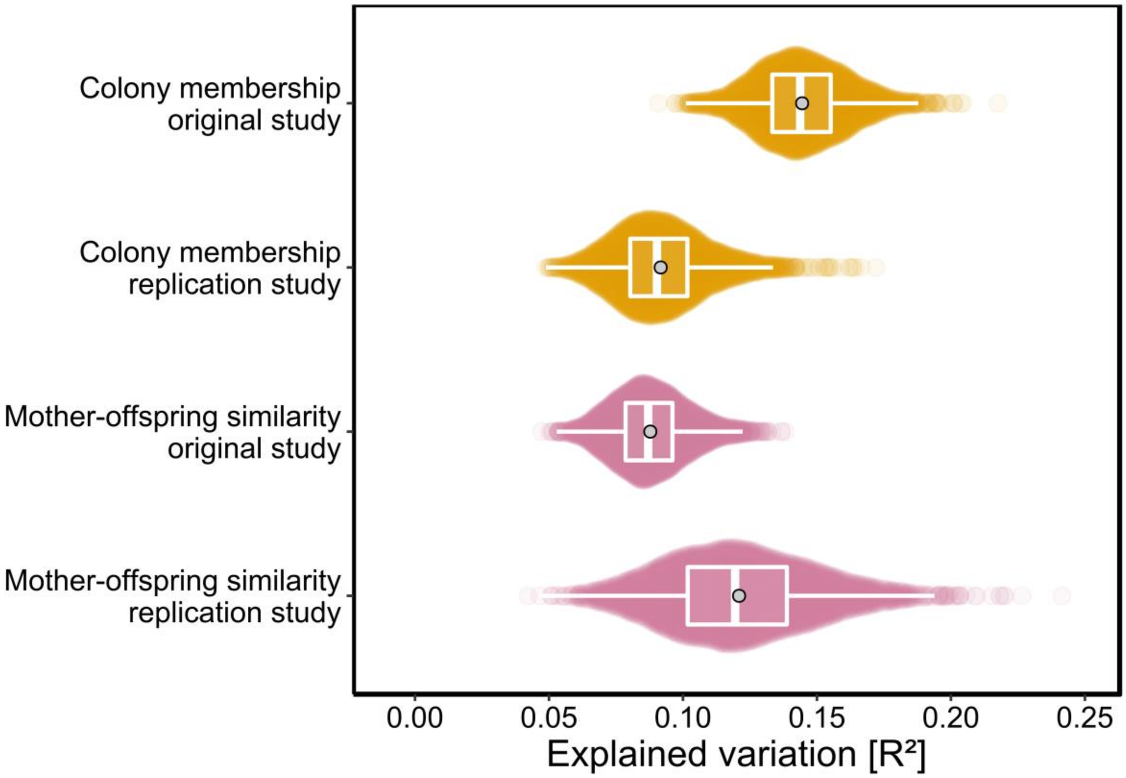


Table 1. Results of PERMANOVAs of (a) 50 mother-offspring pairs from two colonies (SSB and FWB); and (b) 110 pups from six colonies. For comparison, a re-analysis of the chemical data of Stoffel et al. (2015) is shown in part (c). See Materials and methods for details.

(a) PERMANOVA of mothers and offspring from two colonies	F	R^2	***p*-value**
Age	2.65	0.023	0.004
Colony membership	9.07	0.076	<0.0001
Family ID nested within colony membership	9.02	0.153	<0.0001
Test for homogeneity of variance	5.14	–	0.026
(b) PERMANOVA of pups from six colonies	F	R^2	***p*-value**
Colony membership	5.17	0.191	<0.0001
Test for homogeneity of variance	0.50	–	0.778
(c) PERMANOVA of mothers and offspring from two colonies (data from Stoffel et al. 2015).	F	R^2	***p*-value**
Age	0.98	0.010	0.461
Colony membership	12.35	0.128	<0.0001
Family ID nested within colony membership	3.13	0.065	<0.0001
Test for homogeneity of variance	0.22		0.639

Figure 3: Effect sizes of colony membership and mother-offspring similarity in the original study (Stoffel et al. 2015) and in this replication study. To quantify the amount of explained variance, we bootstrapped both datasets over individuals and extracted the corresponding R^2 values for each of the predefined groups in separate PERMANOVAs (see Materials and methods for details). The data are presented as sinaplots with overlaid boxplots (centre line = median, bounds of box = 25th and 75th percentiles, upper and lower whiskers = largest and lowest value but no further than 1.5 * inter-quartile range from the hinge) and the grey points represent effect sizes based on the full datasets.



Generality of chemical patterns

To investigate whether chemical signatures are colony-specific in general, we analysed chemical data from pups sampled from a total of six colonies around Bird Island. PERMANOVA uncovered chemical differences not only between SSB and FWB, but also more generally among colonies (Supplementary figure 1). These differences were significant both overall ($p < 0.0001$, Table 1b) and for the majority of pairwise comparisons after FDR correction (Table 2).

Table 2. Results of PERMANOVAs of colony membership showing the magnitude and significance of chemical differences in all possible pairwise combinations of pups from six breeding colonies. See Materials and methods for details.

Pairs	***F***	***R^2***	***p*-value**	**Corrected *p*-value**
SSB versus FWB	6.08	0.110	<0.0001	<0.001
SSB versus Landing Beach	3.54	0.083	<0.001	0.012
SSB versus Main Bay	6.18	0.137	<0.0001	<0.0001
SSB versus Natural Arch	4.17	0.097	<0.001	0.002
SSB versus Johnson Cove	4.40	0.104	<0.0001	<0.001
FWB versus Landing Beach	4.16	0.099	<0.0001	<0.001
FWB versus Main Bay	7.29	0.161	<0.0001	<0.001
FWB versus Natural Arch	7.91	0.172	<0.0001	<0.001
FWB versus Johnson Cove	7.15	0.162	<0.0001	<0.001
Landing Beach versus Main Bay	3.32	0.106	0.002	0.030
Landing Beach versus Natural Arch	3.00	0.097	0.004	0.063
Landing Beach versus Johnson Cove	2.70	0.091	0.012	0.177
Main Bay versus Natural Arch	5.92	0.175	<0.0001	<0.001
Main Bay versus Johnson Cove	3.14	0.104	<0.001	0.005
Natural Arch versus Johnson Cove	2.43	0.083	0.016	0.245

Discussion

A major obstacle to reproducible research in ecology and evolution is the perceived difficulty of replicating original research findings in natural settings where many variables cannot be controlled for and where spatial and temporal dependencies may confound faithful replication attempts (Nakagawa & Parker, 2015; Fidler et al., 2017). Although the inherent variability of natural systems undoubtedly poses a major challenge to replication studies, our research suggests that, at least under some circumstances, previously reported chemical patterns may be replicable. Specifically, we found that the effect sizes of patterns of colony membership and mother-offspring similarity in Stoffel et al. (2015) were of similar magnitude in a new sample of mother-offspring pairs separated by five years. By expanding the geographical scope of our sampling, we could furthermore show that chemical signatures are colony-specific in general. Our results lend further support to the conclusion that colony membership and mother-offspring similarity are chemically encoded in Antarctic fur seals.

Motivation and study design

A number of factors motivated the current replication attempt. First, the results of Stoffel et al. (2015) were based on a fairly modest sample of Antarctic fur seal mother-offspring pairs sampled in a single season. We therefore wanted to safeguard against type I error while also testing for the repeatability of chemical patterns over time. Second, chance results can become highly influential (Kelly, 2006) and our original study already appears to have motivated comparable investigations in other pinniped species. For example, a recent study of Australian sea lions using a very similar experimental design also reported chemical differences between two breeding colonies, but chemical similarities were not found between mothers and their pups (Wierucka et al., 2019). Although it is not unreasonable to assume that different species might vary in how chemical information is encoded and used in mother-offspring recognition, this point of difference nevertheless encouraged us to revisit our original findings. Finally, being able to confirm and extend our original results strengthens the case for follow-up studies and reduces the risk of time and resources being wasted on chasing up false positives.

Although we acknowledge that no study of a wild population can ever be perfectly replicated (Nakagawa & Parker, 2015; Fidler et al., 2017), we believe that our replication study of chemical patterns in Antarctic fur seals is sufficiently close to that of Stoffel et al. (2015) in terms of both the experimental design and implementation to be considered an exact replication. In practice, there were a handful of minor differences between the two studies, but these were mainly a consequence of incremental improvements to our methodology and are unlikely to have had a major influence on the final outcome. For example, because replication studies often produce weaker effect sizes than original studies (Simonsohn, 2015; Open science collaboration, 2015), we attempted to enlarge our sample size of mother-offspring pairs as far as was practicable. We also improved the standardization and reproducibility of our chemical analysis pipeline by performing peak detection with open source software and by integrating the alignment algorithm of Stoffel et al. (2015) into an R package (Ottensmann et al., 2018). However, these minor modifications appear to have been of little consequence as the effect sizes of colony membership and mother-offspring similarity did not differ systematically between the two studies.

Two further methodological differences were beyond our control. First, owing to the fact that the original and replication studies were conducted five years apart, the sampling was conducted by different teams of field biologists. However, we used carefully standardized field protocols in order to minimize any inadvertent experimental variation. Second, the GC-MS machine used by Stoffel et al. (2015) was subsequently replaced by a newer and more sensitive model. One might have expected this to result in more chemicals being detected in the replication study, which would be expected to provide greater power to detect chemical patterns. If anything, however, fewer chemicals in total were detected in the current study, possibly because of differences in the concentrations of samples or because we used different peak calling software and manually curated the resulting dataset to remove redundant split peaks. Regardless of the exact explanation, the overall similarity of the results of the two studies suggests that patterns of colony membership and mother-offspring similarity in Antarctic fur seals are robust to these sources of experimental variation. This robustness would be expected if chemical patterns are influenced by large numbers of compounds and therefore persist independently of minor methodological differences that may influence which subsets of peaks are detected and retained for analysis.

Replication outcomes

Successful replication can be defined either in the context of statistical significance (Rosenthal, 1991) or on the basis of a comparison of effect sizes (Goodman, Fanelli & Ioannidis, 2016; Piper et al., 2019). Consequently, we not only tested for significance but also developed an approach based on PERMANOVA to evaluate the effect sizes of colony membership and mother-offspring similarity in both datasets. Specifically, we extracted R^2 values for the terms in question after bootstrapping both chemical datasets over individuals. This approach controlled for differences in sample size between the two studies while also providing a visual representation of the magnitude of uncertainty associated with the R^2 estimates. We not only found that the patterns reported by Stoffel et al. (2015) remained highly significant, but also that the effect size estimates of colony membership and mother-offspring similarity in the two studies were more or less similar, varying by at most a few percent. Elsewhere, in a study that attempted to replicate a hundred psychological studies (Open science collaboration, 2015), variation in the strength of the original evidence, such as *p*-values, was more predictive of replication success than other characteristics such as the experience or expertise of the original and replication teams. This is consistent with the outcome of the current replication exercise given the high significance ($p < 0.0001$) of the patterns originally reported by Stoffel et al. (2015).

Temporal reproducibility of chemical patterns

Our results suggest that chemical patterns of colony membership and mother-offspring similarity are repeatable over a timescale of several years, despite the individuals that we sampled being themselves different. This conclusion is not only supported by our chemical data from two discrete time points (2011 and 2016) but also by another study from an intermediate time point (2014) that reported differences in skin microbial communities between SSB and FWB as well as similarities between mothers and their pups (Grosser et al., 2019). These parallel patterns are consistent with the fermentation hypothesis of chemical communication, which argues that microbes metabolize host secretions into odorous compounds that may be used as olfactory cues for individual host recognition (Albone et al., 1974; Gorman, 1976). Nevertheless, further studies will be needed to establish a direct link between host-associated bacterial communities and individual chemical phenotypes.

Broader patterns of colony membership

Although exact replication studies are essential for determining the validity of scientific discoveries, they are only capable of showing whether a given effect can be replicated in a specific setting (Goodman, Fanelli & Ioannidis, 2016; Dirnagl, 2019; Piper et al., 2019). We therefore wanted to know whether chemical differences between SSB and FWB were specific to these two colonies, or whether chemical signatures are colony-specific in general. Unfortunately, it was not possible to sample mothers from locations other than SSB and FWB due to the difficulty of capturing adult females farther away from the BAS field station where fieldwork is rarely if ever conducted. However, the relative ease of capturing pups enabled us to gather a more representative collection of chemical samples from multiple breeding sites around Bird Island. After controlling for the false discovery rate, statistically significant chemical differences were detected in all but two out of 15 pairwise comparisons between colonies. This suggests not only that chemical patterns of colony membership are repeatable over time, but also that they can be generalized over space. Both are prerequisites for olfactory site recognition, which several authors have speculated may play a role in pinniped orientation (Sergeant, 1970; Burton, Anderson & Summers, 1975).

Mechanisms encoding chemical information

Relatively little is currently known about the mechanisms by which colony membership and mother-offspring are chemically encoded in Antarctic fur seals. We know that animals from SSB and FWB exhibit chemical differences despite a lack of genetic differentiation (Stoffel et al 2015), which implies that environmental drivers may be disproportionately important. However, it remains unclear exactly what these drivers might be. Food is unlikely to be an important determinant of colony-specific chemical patterns because females from across Bird Island use the same foraging grounds and feed predominantly on Antarctic krill (Boyd, Staniland & Martin, 2002). The underlying substrate is also relatively homogenous, with the vast majority of animals occupying cobblestone breeding beaches that show little in the way of obvious differences to the human eye. It is therefore more likely that colony-specific chemical phenotypes are influenced by differences in local conditions such as temperature, wind or solar radiation, either directly or via alterations to the skin microbiota (Grosser et al 2019). A further possibility could be that chemical differences between colonies reflect differences in microbial communities shaped by social stress. For example, stressful conditions such as high densities of conspecifics can suppress microbial diversity (Bailey et al., 2011; Stothart et al., 2016; Noguera et al., 2018; Partrick et al., 2018; Zha et al., 2018). This is consistent with our data, as breeding females on SSB are present at higher density and have chronically elevated levels of the stress hormone cortisol (Meise et al., 2016), while skin microbial diversity is also lower in this colony (Grosser et al 2019). Investigating the potential linkages between social stress, cortisol, microbial community structure and chemical phenotypes therefore represents a promising avenue for future research.

Conclusions

Our study set out to test two hypotheses, namely that chemical patterns of colony membership and mother-offspring similarity in Antarctic fur seals are reproducible over time, and that chemical differences will be present not only between SSB and FWB, but also more generally among breeding colonies. Both hypotheses were supported by our data. The overall robustness of chemical patterns of colony membership and mother-offspring similarity in Antarctic fur seals is consistent with the notion that chemical information could be important for both orientation and social communication in pinnipeds, and lays a solid foundation for future studies of the mechanisms responsible for chemical variation among individuals. Finally, as a lack of access to raw data, code and software has been identified as a fundamental obstacle to replication (Fidler et al., 2017), we have made the data from both studies as well as the code used to analyze them freely available, while also using open access software for peak detection and alignment.

Acknowledgements

We are grateful to the field assistants on Bird Island who contributed toward animal handling and tissue sampling. We would also like to thank John Dickens who helped collect samples from additional beaches on Bird Island. This work contributes to the Ecosystems project of the British Antarctic Survey, Natural Environmental Research Council, and is part of the Polar Science for Planet Earth Programme.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of a Sonderforschungsbereich (project numbers 316099922 and 396774617–TRR 212) and the priority programme “Antarctic Research with Comparative

Investigations in Arctic Ice Areas" (SPP 1158, project number 424119118). It was also supported by core funding from the Natural Environment Research Council to the British Antarctic Survey's Ecosystems Program. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University

References

Albone ES, Eglinton G, Walker JM, Ware GC. 1974. The anal sac secretion of the red fox (*Vulpes vulpes*); its chemistry and microbiology. A comparison with the anal sac secretion of the lion (*Panthera leo*). *Life Sciences* 14:387–400. DOI: 10.1016/0024-3205(74)90069-1.

Bailey MT, Dowd SE, Galley JD, Hufnagle AR, Allen RG, Lyte M. 2011. Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation. *Brain, Behavior, and Immunity* 25:397–407. DOI: 10.1016/j.bbi.2010.10.023.

Baker M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533:452–454. DOI: 10.1038/533452a.

Begley CG, Ellis LM. 2012. Raise standards for preclinical cancer research. *Nature* 483:531–533. DOI: 10.1038/483531a.

Blaustein AR. 1983. Kin Recognition Mechanisms: Phenotypic Matching or Recognition Alleles? *The American Naturalist* 121:749–754. DOI: 10.1086/284101.

Boyd I, Staniland I, Martin A. 2002. Distribution of foraging by female Antarctic fur seals. *Marine Ecology Progress Series* 242:285–294. DOI: 10.3354/meps242285.

Burgener N, Dehnhard M, Hofer H, East ML. 2009. Does anal gland scent signal identity in the spotted hyaena? *Animal Behaviour* 77:707–715. DOI: 10.1016/j.anbehav.2008.11.022.

Burton RW, Anderson SS, Summers CF. 1975. Perinatal activities in the Grey seal (*Halichoerus grypus*). *Journal of Zoology* 177:197–201. DOI: 10.1111/j.1469-7998.1975.tb05978.x.

Caspers BA, Schroeder FC, Franke S, Streich WJ, Voigt CC. 2009. Odour-based species recognition in two sympatric species of sac-winged bats (*Saccopteryx bilineata*, *S. leptura*): combining chemical analyses, behavioural observations and odour preference tests. *Behavioral Ecology and Sociobiology* 63:741–749. DOI: 10.1007/s00265-009-0708-7.

Caspers BA, Schroeder FC, Franke S, Voigt CC. 2011. Scents of Adolescence: The Maturation of the olfactory phenotype in a free-ranging Mammal. *PLoS ONE* 6:e21162. DOI: 10.1371/journal.pone.0021162.

Charpentier MJE, Boulet M, Drea CM. 2008. Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology* 17:3225–3233.

Dirnagl U. 2019. Rethinking research reproducibility. *The EMBO Journal* 38:2018–2020. DOI: 10.15252/embj.2018101117.

Dobson FS, Jouventin P. 2003. How mothers find their pups in a colony of Antarctic fur seals. *Behavioural Processes* 61:77–85. DOI: 10.1016/S0376-6357(02)00164-X.

Fidler F, En Chee Y, Wintle BC, Burgman MA, McCarthy MA, Gordon A. 2017. Metaresearch for evaluating reproducibility in ecology and evolution. *BioScience* 67:282–289. DOI: 10.1093/biosci/biw159.

Fisher RA. 1974. *The Design of Experiments*. New York: Hafner Press.

Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. 2018. Questionable research practices in ecology and evolution. *PLOS ONE* 13:e0200303. DOI: 10.1371/journal.pone.0200303.

Fratini G, Lois S, Pazos M, Parisi G, Medina I. 2012. Volatile profile of Atlantic shellfish species by HS-SPME GC/MS. *Food Research International* 48:856–865. DOI: 10.1016/j.foodres.2012.06.033.

Goodman SN, Fanelli D, Ioannidis JPA. 2016. What does research reproducibility mean? *Science translational medicine* 8.

Gorman ML. 1976. A mechanism for individual recognition by odour in *Herpestes auropunctatus* (Carnivora: Viverridae). *Animal Behaviour* 24:141–145. DOI: 10.1016/S0003-3472(76)80107-8.

Grosser S, Sauer J, Paijmans AJ, Caspers BA, Forcada J, Wolf JBW, Hoffman JI. 2019. Fur seal microbiota are shaped by the social and physical environment, show mother–offspring similarities and are associated with host genetic quality. *Molecular Ecology*:mec.15070. DOI: 10.1111/mec.15070.

Hamilton JE. 1956. Scent of otariids. *Nature* 177:900.

Hardy MH, Roff E, Smith TG, Ryg M. 1991. Facial skin glands of ringed and grey seals, and their possible function as odoriferous organs. *Canadian Journal of Zoology* 69:189–200. DOI: 10.1139/z91-029.

Hurst JL, Beynon RJ. 2010. Making progress in genetic kin recognition among vertebrates. *Journal of biology* 9:13.

Jennions MD, Møller AP. 2002. Publication bias in ecology and evolution: an empirical assessment using the ‘trim and fill’method. *Biological Reviews* 77:211–222.

Kean EF, Chadwick EA, Mueller CT. 2015. Scent signals individual identity and country of origin in otters. *Mammalian Biology* 80:99–105. DOI: 10.1016/j.mambio.2014.12.004.

Kean EF, Müller CT, Chadwick EA. 2011. Otter Scent Signals Age, Sex, and Reproductive

Status. *Chemical Senses* 36:555–564. DOI: 10.1093/chemse/bjr025.

Kelly CD. 2006. Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. *The Quarterly review of biology* 81:221–236.

Kelly CD. 2019. Rate and success of study replication in ecology and evolution. *PeerJ* 7:e7654. DOI: 10.7717/peerj.7654.

Kishida T, Kubota S, Shirayama Y, Fukami H. 2007. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biology Letters* 3:428–430. DOI: 10.1098/rsbl.2007.0191.

Kohlwey S, Krause ET, Baier MC, Müller C, Caspers BA. 2016. Chemical Analyses Reveal Family-Specific Nest Odor Profiles in Zebra Finches (*Taeniopygia guttata*): A Pilot Study. In: *Chemical Signals in Vertebrates 13*. Cham: Springer International Publishing, 167–175. DOI: 10.1007/978-3-319-22026-0_12.

Kovacs KM. 1995. Mother-pup reunions in harp seals, *Phoca groenlandica* - Cues for the relocation of pups. *Canadian Journal of Zoology -Revue Canadienne de Zoologie* 73:843–849. DOI: 10.1139/z95-099.

Kowalewsky S, Dambach M, Mauck B, Dehnhardt G. 2006. High olfactory sensitivity for dimethyl sulphide in harbour seals. *Biology Letters* 2:106–109. DOI: 10.1098/rsbl.2005.0380.

Krause ET, Brummel C, Kohlwey S, Baier MC, Müller C, Bonadonna F, Caspers BA. 2014. Differences in olfactory species recognition in the females of two Australian songbird species. *Behavioral Ecology and Sociobiology* 68:1819–1827. DOI: 10.1007/s00265-014-1791-y.

Ling J, K. 1974. The integument of marine mammals. *Functional anatomy of marine mammals* 2:1–44.

Meise K, von Engelhardt N, Forcada J, Hoffman JI. 2016. Offspring Hormones Reflect the Maternal Prenatal Social Environment: Potential for Foetal Programming? *PLOS ONE* 11:e0145352. DOI: 10.1371/journal.pone.0145352.

Mendoza D, Garcia CA. 2017. Defining Research Reproducibility: What Do You Mean? *Clinical Chemistry* 63:1777.

Müller T, Müller C. 2016. Consequences of mating with siblings and nonsiblings on the reproductive success in a leaf beetle. *Ecology and Evolution* 6:3185–3197. DOI:

10.1002/ece3.2103.

Nakagawa S, Parker TH. 2015. Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology* 13:1–6. DOI: 10.1186/s12915-015-0196-3.

Noguera JC, Aira M, Pérez-Losada M, Domínguez J, Velando A. 2018. Glucocorticoids modulate gastrointestinal microbiome in a wild bird. *Royal Society Open Science* 5:171743. DOI: 10.1098/rsos.171743.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: Community Ecology Package.

Open science collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349:aac4716–aac4716. DOI: 10.1126/science.aac4716.

Ottensmann M, Stoffel MA, Nichols HJ, Hoffman JI. 2018. GCalignR: An R package for aligning gas-chromatography data for ecological and evolutionary studies. *PLoS ONE* 13:1–20. DOI: 10.1371/journal.pone.0198311.

Palmer AR. 2000. Quasi-Replication and the Contract of Error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics* 31:441–480. DOI: 10.1146/annurev.ecolsys.31.1.441.

Partrick KA, Chassaing B, Beach LQ, McCann KE, Gewirtz AT, Huhman KL. 2018. Acute and repeated exposure to social stress reduces gut microbiota diversity in Syrian hamsters. *Behavioural Brain Research* 345:39–48. DOI: 10.1016/j.bbr.2018.02.005.

Phillips A V. 2003. Behavioral cues used in reunions between mother and pup south american fur seals (Arctocephalus australis). *Journal of Mammalogy* 84:524–535. DOI: 10.1644/1545-1542(2003)084<0524:BCUIRB>2.0.CO;2.

Piper SK, Grittner U, Rex A, Riedel N, Fischer F, Nadon R, Siegerink B, Dirnagl U. 2019. Exact replication: Foundation of science or game of chance? *PLoS Biology* 17:1–9. DOI: 10.1371/journal.pbio.3000188.

Pitcher BJ, Harcourt RG, Schaal B, Charrier I. 2011. Social olfaction in marine mammals: wild female Australian sea lions can identify their pup's scent. *BIOLOGY LETTERS* 7:60–62. DOI: 10.1098/rsbl.2010.0569.

R Core Team. 2019. R: A Language and Environment for Statistical Computing.

Rosenthal R. 1991. Replication in Behavioural Research. In: Neuliep JW ed. *Replication research in the social sciences*. Newbury Park [u.a.]: Sage Publ., 1–31.

Schmidt S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of general psychology* 13:90–100.

Schneeberger K, Voigt CC, Müller C, Caspers BA. 2016. Multidimensionality of Chemical Information in Male Greater Sac-Winged Bats (*Saccopteryx bilineata*). *Frontiers in Ecology and Evolution* 4. DOI: 10.3389/fevo.2016.00083.

Sergeant DE. 1970. Migration and orientation in harp seals. In: *Proc. Seventh Ann. Conf. Biol. Sonar Diving Mamm Menlo Park, CA: Stanford Research Institute Biological Sonar Laboratories*. 123–131.

Setchell JM, Vaglio S, Abbott KM, Moggi-Cecchi J, Boscaro F, Pieraccini G, Knapp LA. 2011. Odour signals major histocompatibility complex genotype in an Old World monkey. *Proceedings of the Royal Society B: Biological Sciences* 278:274–280. DOI: 10.1098/rspb.2010.0571.

Simonsohn U. 2015. Small Telescopes. *Psychological Science* 26:559–569. DOI: 10.1177/0956797614567341.

Stoffel MA, Caspers BA, Forcada J, Giannakara A, Baier M, Eberhart-Phillips L, Müller C, Hoffman JI. 2015. Chemical fingerprints encode mother–offspring similarity, colony membership, relatedness, and genetic quality in fur seals. *Proceedings of the National Academy of Sciences* 112:E5005–E5012. DOI: 10.1073/pnas.1506076112.

Stothart MR, Bobbie CB, Schulte-Hostedde AI, Boonstra R, Palme R, Mykytczuk NCS, Newman AEM. 2016. Stress and the microbiome: linking glucocorticoids to bacterial community dynamics in wild red squirrels. *Biology Letters* 12:20150875. DOI: 10.1098/rsbl.2015.0875.

Sun L, Müller-Schwarze D. 1998. Anal gland secretion codes for family membership in the beaver. *Behavioral Ecology and Sociobiology* 44:199–208.

Vogt K, Boos S, Breitenmoser U, Kolliker M. 2016. Chemical composition of Eurasian lynx urine conveys information on reproductive state, individual identity, and urine age. *Chemoecology* 26:205–217. DOI: 10.1007/s00049-016-0220-2.

Wierucka K, Barthes N, Harcourt R, Schaal B, Charrier I, Pitcher BJ. 2019. Chemical fingerprints suggest direct familiarisation rather than phenotype matching during olfactory recognition in Australian sea lions (*Neophoca cinere*a). *Journal of Experimental Marine Biology and Ecology* 517:49–53. DOI: 10.1016/j.jembe.2019.06.001.

Wyatt TD. 2015. The search for human pheromones: the lost decades and the necessity of returning to first principles. *Proceedings of the Royal Society B: Biological Sciences* 282:20142994. DOI: 10.1098/rspb.2014.2994.

Yamazaki K, Beauchamp GK, Imai Y, Bard J, Phelan SP, Thomas L, Boyse EA. 1990. Odortypes determined by the major histocompatibility complex in germfree mice. *Proceedings of the National Academy of Sciences of the United States of America* 87:8413–8416. DOI: 10.1073/pnas.87.21.8413.

Zha Y, Eiler A, Johansson F, Svanbäck R. 2018. Effects of predation stress and food ration on perch gut microbiota. *Microbiome* 6:28. DOI: 10.1186/s40168-018-0400-0.