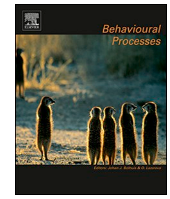
Behavioural Processes 169 (2019) 103961

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc




# Evaluating the stability of individual variation in social and nonsocial behavioural types using prairie voles (*Microtus ochrogaster*)

Todd H. Ahern[a,*], Ophir Alexander[b], Burn David[c]

[a] Center for Behavioural Neuroscience, Department of Psychology, Quinnipiac University, 275 Mount Carmel Ave., BC-SCI, Hamden, CT, USA
[b] Department of Psychology, Cornell University, Ithaca, NY, 14853, USA
[c] Department of Mathematics and Statistics, Quinnipiac University, 275 Mount Carmel Ave., BC-SCI, Hamden, CT, USA

ARTICLE INFO

*Keywords:*
Social Behaviour
Prai
rie vole
Reproducibility
Replicability
Reciprocal interaction
Animal personality

ABSTRACT

Prairie voles (*Microtus ochrogaster*) exhibit remarkable individual variation in social behaviour, suggesting differences in behavioural types. To date, however, there has been little assessment of whether these behavioural types are stable across test sessions, nor to what extent internal states and external contexts (domains) drive individual differences. Here we examined the individual consistency of social (*huddling*) and non-social (*distance moved*) behaviour across repeated, long-duration tests, in same-sex cagemate (SS-CM), same-sex stranger (SS-S), opposite-sex stranger (OS-S), and standard partner preference test (PPT) contexts. The SS-CM and SS-S tests were repeated multiple times (SS-CM 1–2; SS-S 1–5) to assess state-dependent variation. A second cohort was used to determine the replicability of findings. Overall, there was a general lack of stability in huddling behavior. It was inconsistent across repeated sessions of the same test type and between types of tests, suggesting a strong contribution of state-dependent variation. Non-social behaviour was more consistent and appeared more domain-dependent and less state-dependent than huddling. Translational and comparative studies of individual variation would likely benefit from testing across multiple contexts and employing repetitive testing paradigms to account for state-dependent variation.

## 1. Introduction

The process of natural selection is built on the principle that individuals within a population vary. The variation can take several forms—behavioural, morphological, physiological, etc.—but ultimately the sources of variance must be rooted in genetic heritability and the local environment.

Social behaviors, or behaviors that result from the interaction of two or more individuals, are among the most variable phenotypes that exist. Animals must react to dynamic social environments, in which interactions depend on important internal state and external contextual (i.e., domain) factors—e.g., reproductive vs non-reproductive contexts (Anacker and Beery, 2013). Examples of typical (e.g., introversion) and atypical (e.g., autism spectrum) variations in human social behaviour abound, and the underlying causes of these differences are the focus of intense study (Depue and Collins, 1999; Guastella and Hickie, 2016; Rapin and Tuchman, 2008). It has been challenging to make sense of natural individual variation because it is often (perhaps mistakenly) disregarded as 'noise' (e.g., Drummond and Gordon, 1979). Biomedical studies generally use highly inbred lines of animals to reduce individual variation (Kelly and Ophir, 2015; Stevenson et al., 2018; Taborsky et al., 2015), and behavioural ecologists historically viewed behavioural responses as a function of optimality; they assumed mean responses were under selection pressure as approximations of the best possible outcome under a set of circumstances (Sloan Wilson et al., 1994). In both fields, variation was traditionally expected to be low and idiosyncratic.

This view has undergone a change in recent years, and scientists in behavioural biology and neuroscience are beginning to develop a deeper appreciation for behavioural variation within and between individuals (Budaev, 1997; Sih et al., 2004a, 2004b; Snell-Rood, 2013; Wolf and Weissing, 2010). Identification of behavioural types, from strain-bred (e.g., Koolhaas et al., 2010, 2007) and natural populations (Bell, 2007; Bell and Stamps, 2004), has opened the door to studying the importance of behavioural variation and consistency between and within individuals. This is an important advancement because individual variation can impact many overarching elements of sociality, including social behavior, social organization, and group dynamics (e.g., Kappeler, 2019; Lott, 1984; Pinter-Wollman et al., 2016; Schradin et al., 2012).

* Corresponding author.
*E-mail addresses:* todd.ahern@quinnipiac.edu (T.H. Ahern), ophir@cornell.edu (A. Ophir).

https://doi.org/10.1016/j.beproc.2019.103961
Received 3 April 2019; Received in revised form 7 September 2019; Accepted 10 September 2019
Available online 11 September 2019
0376-6357/ 

Understanding individual variation in behaviour can be approached in at least two ways. First, the behavioural domain (i.e., the context) can drive the response activation of an animal (Eysenck, 1967; Fahrenberg, 1977). In other words, behavioural variation can be domain-dependent. Second, the intrinsic state of the animal can influence the likelihood of a behavioural response within a domain, and across domains (Eysenck, 1967; Fahrenberg, 1977). In this case, behavioural variation is state-dependent. These two factors need not be mutually exclusive, but the primary driver of behavioural variation is likely attributable to one or the other source. As a result, a gross distinction can be made. In the case of domain-specific variation, an individual should show behavioural consistency when examined repeatedly in the same context, but exhibit behavioural variation across different contexts. In the case of state-specific variation, an individual's behavioural responses should vary when tested repeatedly in a single context and may or may not also vary across contexts (or domains). Thus, predictions about when behavioural consistency is expected can be tied to the source of the variation. This can provide important insights into the mechanistic nature of individual variability.

Prairie voles (*Microtus ochrogaster*) are one promising model for testing predictions about individual consistency of social behaviour. Prairie voles are small social rodents that are used in ecology and evolution and in translational neuroscience. Not only do they display individual differences in a range of interesting social behaviours (Carter et al., 1995; Getz and McGuire, 1993; Getz et al., 1981), but these differences have been linked to variation in neurochemistry and genetics (King et al., 2016; McGraw and Young, 2009; Okhovat et al., 2015; Phelps et al., 2009; Young and Wang, 2004). Despite this compelling research, many laboratory studies characterize behavior using only one test session. It is unclear whether behavioral types are stable and to what extent behaviour might be driven by domain- and state-dependent factors.

In the current study, we hypothesized that stable differences in behavioral type are the primary driver of individual variation, so social behaviour should be relatively consistent across domains (i.e., low domain-dependent variation) and test sessions (i.e., low state-dependent variation).

To test this hypothesis, we quantified male and female social (huddling) and nonsocial (distance moved) behaviour in three types of reciprocal interaction tests: same-sex cagemate (SS-CM), same-sex stranger (SS-S), and opposite-sex stranger (OS-S). We also used the partner preference test (PPT; Ahern et al., 2009; Williams et al., 1992). Each test type approximated a different social domain. The PPT is the gold standard for assessing affiliative social behavior in prairie voles and huddling is a key metric that exhibits individual variation (Amadei et al., 2017; Carter et al., 1995; Young and Wang, 2004). Reciprocal interaction tests have also been used to characterize rodent social behavior (Chang et al., 2017; Defensor et al., 2011). By adding multiple SS-CM and SS-S test sessions to our design, we aimed to characterize behavior not only across test types but within them. We also examined replicability by testing males and females in two independent cohorts (12 males, 12 females in each).

Overall, we predicted that prairie voles would exhibit significant correlations in huddling behavior between test types and test sessions. As a corollary, we predicted that high and low huddlers would remain high and low huddlers across test domains and test sessions. We predicted a similar pattern for locomotor behavior. We also predicted that a sample of 12 animals would contain enough individual variation to identify statistically distinct behavioral types, and findings made in one cohort could be replicated in a second cohort. Importantly, by analyzing the sexes and the cohorts separately, our statistical modeling aimed to avoid identifying effects in atypically large sample sizes of approximately 24 or 48, which would be hard to replicate in other laboratories.

## 2. Methods

### 2.1. Animals and cohorts

Prairie voles (*Microtus ochrogaster*), originally derived from Illinois, were bred and reared in standard rat cages (43 cm x 20 cm x 20 cm). At 21 d of age, pups were weaned into same-sex pairs. Males and females were weaned into separate cages and rooms. All weanlings were housed in standard mouse cages (25 cm x 15 cm x 13 cm). Breeding and home cages contained 1/4″ corn cob bedding (Teklad, 7097), cotton nestlets (Ancare, NES3600QTY 59CS), and *ad libitum* access to water and food (LabDiet, 5263-4). Ambient temperature and humidity were maintained at ~20-25 C and 35–60%, respectively, with a 14:10 light:dark cycle. This study was carried out in accordance with the recommendations of the Guide for the Care and Use of Laboratory Animals, NIH. The protocol was approved by the Institutional Animal Care and Use Committee of Quinnipiac University.

Two cohorts of adult male and female prairie voles (70–150 days old) were tested: Cohort 1 (n males = 12, mean age = 99.5 ± 3.4 days old; n females = 12, mean age = 99 ± 3.2 days old) and Cohort 2 (n males = 12, mean age = 124 ± 4.9 days old, n females = 12, mean age = 124 ± 7.0 days old). These two cohorts were reared and tested in different months but followed the same testing procedures.

Approximately 2 days prior to the first test session (see below), one animal from each weanling pair was marked; the rump was shaved to reveal the lighter skin underneath (~2 in. anterior-to-posterior by ~1–2 in. right-to-left) and the neck was loosely collared with a bright yellow zip-tie (Commercial Electric, 826 843). Each cohort had a male group consisting of 6 marked animals and 6 unmarked animals and a female group consisting of 6 marked animals and 6 unmarked animals.

### 2.2. Reciprocal interaction test design

Within each cohort, males and females underwent seven same-sex 3-h reciprocal interaction tests spread over 2–3 weeks using a partial round-robin design. For example, in cohort 1 there were 6 marked males and 6 unmarked males. Each marked male was tested with its unmarked cagemate twice (SS-CM tests 1–2) and with each unmarked stranger once (5 strangers = SS-S tests 1–5); thus, every SS-S session was with a novel stranger. The design was only a partial round-robin, however, because marked animals were never tested with other marked animals and unmarked animals were never tested with other unmarked animals. The order of the SS-CM and SS-S tests was pseudo-randomized to avoid temporal position effects.

We acknowledge that including an equal number of SS-CM and SS-S tests might have been better. At the outset, however, we expected limited variability in behavior across test sessions. We primarily included the extra three SS-S test sessions because they fit well with the partial round-robin design and would allow us to better characterize the range of potential variability for different stranger dyads. Importantly, adding three additional SS-CM tests would have pushed us beyond the time-frame available for testing and we did not believe we would need as many SS-CM tests to estimate variation between familiar animals.

After all SS-CM and SS-S tests were completed for a cohort, subjects underwent an eighth reciprocal interaction test with an opposite-sex stranger (OS-S). Marked males were paired with unmarked females and vice versa. Male-female sibling pairs were avoided.

Other than the identity of the stimulus animal, each reciprocal interaction test was conducted identically.

### 2.3. Reciprocal interaction test setup

For reciprocal interaction testing, we used our standard PPT boxes (75 × 20 × 30 cm), which have been described in detail elsewhere

(Ahern et al., 2009). Each box consists of 3 chambers (each measuring 25 cm × 20 cm × 30 cm), with an open passage down the middle, allowing animals easy access to each other while providing the option of moving into a separate chamber. Water and food were provided *ad libitum* during each 3 h test.

The testing room contained 2 digital cameras mounted over 4 test boxes. The cameras were connected to a surveillance system (Q-See® 8 Camera Complete Surveillance System Slim 10.2" Integrated LCD-DVR, 5832791) located in an adjacent room, which digitally recorded the tests.

The reciprocal interaction tests consisted of placing two animals (one marked, one unmarked) in the test box and allowing them to roam freely for 3 h. Four reciprocal interaction pairs were tested simultaneously: four pairings in an early session, four new pairings a later session. Thus testing occurred throughout the day. In addition to pseudo-randomizing the same-sex test types (SS-S vs SS-CM), we also pseudo-randomized in which box and during which time of day each animal was tested to avoid territoriality and temporal positioning confounds.

### 2.4. Reciprocal interaction test video-tracking

A desktop computer (Dell, Windows 7, Intel Core 2 Duo processor) running SocialScan 2.0 (Clever Sys Inc., Reston, VA) with the Social Basic and AggressionScan modules was used to automatically score social and nonsocial behaviours from the video recordings. SocialScan can consistently and accurately quantify prairie vole huddling behaviour (Ahern and Young, 2009; Barrett et al., 2014; Keebaugh and Young, 2011; King et al., 2016). For this study, we used the "mark tracking" feature to allow tracking of two free-roaming subjects.

### 2.5. Reciprocal interaction test behaviours

In the reciprocal interaction test, we measured *huddling* and *distance moved* as our primary measures of social and nonsocial behavior, respectively. We only focused on huddling and locomotor behavior because they are the standard measures of social behavior in the prairie vole PPT and they are the best validated using the automated behavior scoring software (Ahern et al., 2009). In the automated scoring, huddling is scored when animals were in contact with each other and the immobility measure was below a threshold of 0.05 (Ahern et al., 2009). Distance moved was calibrated to the length and width of each test box and corresponds to the path length traced by the center of each animal. Human vs computer scoring revealed high inter-rater reliability (r = 0.98) for huddling, with a near 1:1 ratio (Slope = 1.07, N = 54 video clips).

### 2.6. Partner preference testing

Directly after the 3 h OS-S test, each male-female pair was moved to a standard mouse cage and allowed to cohabitate for 21 additional hours, for a total of 24 h of cohabitation. Males were then assessed in the PPT.

Shortly before the PPT began, cohabitated females were loosely tethered and positioned at the ends of each test box, according to standard practice (Ahern et al., 2009; Williams et al., 1992b). The test male was then introduced and allowed to roam freely for 3 h. Four males were tested in an early session; four different males were tested in a later session. In each session, half the females served as partners (i.e., they had cohabitated with the test male), the other half as strangers (i.e., they had cohabitated with a different male). The locations of the partner and stranger were counter-balanced. *Ad libitum* food and water were provided.

After the PPT, male-female pairs were returned to their home-cages after the females had been untethered. They continued to cohabitate until the next day (~48 h total cohabitation), when the female PPT occurred. Males were tethered, with half serving as partners and half serving as strangers in each test session. Female PPT behaviour was digitally recorded for 3 h and processed by SocialScan.

As in the reciprocal interaction tests, test sessions were digitally recorded (Q-See) and SocialScan scored *huddling* (i.e., immobile-social-contact) with the partner and stranger, as well as *distance moved* in the nonsocial zone. Huddling was quantified using the same parameters as in the reciprocal interaction tests.

### 2.7. Data analysis

All data were exported from SocialScan and aligned and decoded in MS Excel before being imported into SPSS (v. 23, IBM). General Linear Modeling (GLM) with the Greenhouse-Geisser adjustment was used to test within-subjects factors such as *test type*. Univariate GLM was used to distinguish individual differences. In these tests, the average behavior of the stimulus animal was included as a covariate. All *post hoc* comparisons represent Tukey-corrected p-values, unless otherwise noted. PPT huddling behaviour was examined using the non-parametric related-samples Wilcoxon signed rank tests to account for nonnormality.

Pearson's correlations, intraclass correlations (ICC), and data plots were used to assess behavioural consistency of individuals, both between and within test types. The strength of the relationship between behaviors and/or different tests was assessed by comparing the absolute value of each Pearson's correlation according to guidelines from Hinckle and colleagues: 0.0-0.3 = little to no correlation; 0.3-0.5 = low; 0.5-0.7 = moderate; 0.7-0.9 = high; and 0.9–1.0 = very high (Hinkle et al., 2003). ICCs, however, are arguably better at measuring behavioural consistency if mean differences are not expected (Bell and Stamps, 2004; Hayes and Jenkins, 1997), as in the case of repeated test sessions of the same test type (e.g., SS-S1–5). We report ICC reliabilities as described in Koo & Li (2016): 0.0-0.5 = poor; 0.5-0.75 = moderate; 0.75–0.9 = good; and 0.9–1.0 = excellent. ICCs and their 95% confidence intervals were calculated using a mean-rating (SS-CM: $k = 2$; SS-S: $k = 5$), consistency, two-way random effects model, because not all test animals were tested with all other animals (e.g., males were not tested with females; marked males were not tested with unmarked males).

Statistical significance was determined relative to an alpha level of 0.05.

## 3. Results

### 3.1. Effect of test type on behaviour in the first exposure to each test

Our first repeated-measures GLM examined the effect of *test type* (SS-CM 1, SS-S 1, OS-S, and PPT) on *huddling behaviour* expressed in the first test session of each type. The PPT huddling consisted of partner plus stranger total huddling. For males in cohort 1, test type did not significantly affect huddling (F1.57,15.72 = 1.53, p = 0.24, partial $\eta^2 = 0.13$ Fig. 1A), whereas test type was significant for females (F2.39,26.26 = 5.13, p = 0.01, partial $\eta^2 = 0.32$; Fig. 1C). PPT huddling was significantly higher than the huddling expressed in the SS-S 1 and OS-S tests ($p < 0.05$, d = 0.97 and $p < 0.01$, d = 1.40, respectively; Fig. 1C).

To test the replicability of these findings, we conducted an independent analysis using cohort 2. Here, males showed a significant effect of test type (F1.96,21.59 = 9.16, p = 0.001, partial $\eta^2 = 0.45$), with SS-CM 1 huddling significantly higher than huddling in the SS-S 1 and OS-S tests ($p < 0.01$, d = 1.40 and $p < 0.05$, d = 1.00, respectively; Fig. 1B) and PPT huddling was significantly higher than S-SS 1 huddling ($p < 0.05$, d = 0.89). There was also a significant effect of test type for females (F2.49,27.48 = 4.52, p = 0.01, partial $\eta^2 = 0.29$; Fig. 1D). Here OS-S huddling was significantly lower than SS-CM 1 and PPT huddling ($p < 0.05$, d = 1.15 and $p < 0.05$, d = 1.01, respectively). Notably, the only finding that replicated across cohorts is that

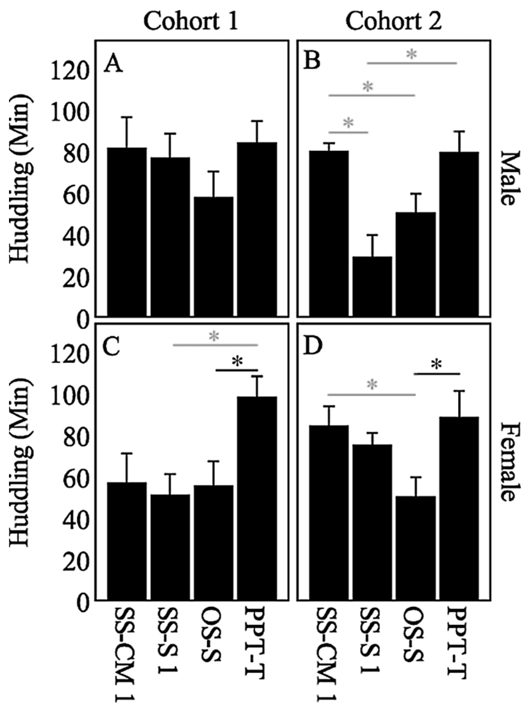



**Fig. 1.** The total *huddling* duration for males (**A** and **B**) and females (**C** and **D**) during the first exposure to each test type in cohort 1 (**A** and **C**) and cohort 2 (**B** and **D**): SS-CM 1 = first same-sex cagemate reciprocal interaction test; SS-S 1 = first same-sex stranger reciprocal interaction test; OS-S = first and only opposite-sex stranger reciprocal interaction test; PPT = first and only partner-preference test. In the PPT, huddling represents the combined partner and stranger huddling. Gray significance symbols represent results that appear in only one cohort. Black significant symbols represent the one result that repeated across cohorts: females huddled more in the PPT than in the OS-S test. Bars represent means + SEM. * $p < 0.05$.

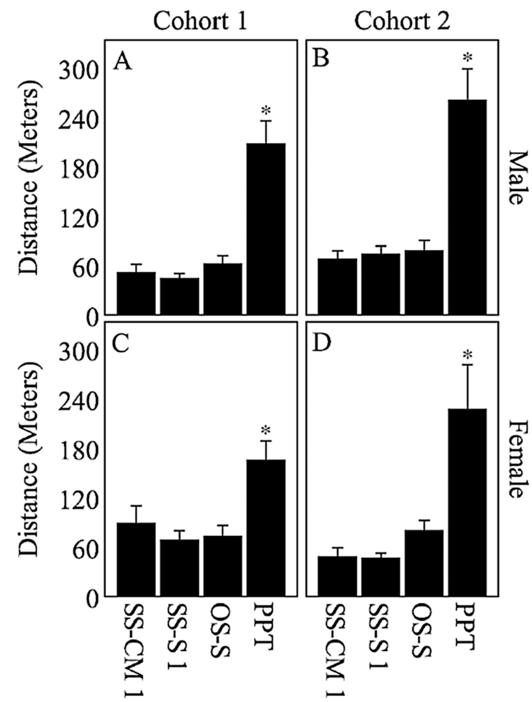



**Fig. 2.** The total *distance moved* by males (**A** and **B**) and females (**C** and **D**) during the first experience with each test type in cohort 1 (**A** and **C**) and in cohort 2 (**B** and **D**): SS-CM 1 = $1^{st}$ same-sex cagemate reciprocal interaction test; SS-S 1 = $1^{st}$ same-sex stranger reciprocal interaction test; OS-S = $1^{st}$ and only opposite-sex stranger reciprocal interaction test; PPT = $1^{st}$ and only partner-preference test. For the PPT, distance moved was only quantified in the nonsocial zone (i.e., outside the zones in which the partner and stranger were tethered). Movement was significantly higher in the PPT than in any of the three reciprocal interaction tests. Bars represent means + SEM. * $p < 0.05$.

females huddled more in the PPT than during OS-S test.

We conducted an identical analysis for the nonsocial behaviour *distance moved*. Here, test type significantly affected movement for males ($F1.37,13.73 = 31.27$, $p = 0.00003$, partial $\eta^2 = 0.76$; Fig. 2A) and females ($F2.15,23.66 = 12.63$, $p = 0.0001$, partial $\eta^2 = 0.54$; Fig. 2C). *Post hoc* comparisons within each sex revealed that the distance moved in the PPT was significantly higher than in all other test types for both sexes (all comparisons with PPT were $p < 0.05$, with cohen's d effect sizes > 1.00). The three reciprocal interaction tests (SS-CM 1, SS-S 1, and OS-S) were not significantly different from one another ($p > 0.05$).

We tested the replicability of these findings using cohort 2 and found that both sexes moved a greater distance in the PPT than in the three reciprocal interaction tests (Males: $F1.22,13.42 = 29.08$, $p = 0.00006$, partial $\eta^2 = 0.758$, all *post hoc* comparisons $p < 0.01$, $d > 1.52$; Fig. 2B; Females: $F1.11,12.21 = 12.46$, $p = 0.003$, partial $\eta^2 = 0.53$, all *post hoc* comparisons $p < 0.05$, $d > 0.91$; Fig. 2D). Cohort 2 essentially replicated cohort 1.

### *3.2. Consistency of behaviour across the first exposure to each test type*

We next examined whether *huddling* behaviour was statistically correlated across test types. We again split by sex and cohort and focused only the first experience with each test. We then calculated Pearson's correlations. Cohort 1 revealed significant, positive correlations in male huddling behaviour between SS-CM 1 and SS-S 1 (Table 1) and between SS-S 1 and the PPT, whereas female huddling exhibited a positive correlation between OS-S and PPT.

Our assessment of repeatability using cohort 2, however, revealed no significant correlations in huddling behaviour between tests in males and a different significant correlation was observed in females, this time between SS-CM 1 and the PPT huddling (Table 1).

**Table 1**
Pearson's Correlations Coefficients for Social Huddling Behaviour.

| C 1 | Test Type | SS-CM 1 | SS-S 1 | OS-S | C 2 | Test Type | SS-CM 1 | SS-S 1 | OS-S |
|---|---|---|---|---|---|---|---|---|---|
| M | SS-CM 1 | | | | M | SS-CM 1 | | | |
| | SS-S 1 | 0.68* | | | | SS-S 1 | −0.15 | | |
| | OS-S | −0.15 | 0.29 | | | OS-S | 0.37 | 0.21 | |
| | PPT | 0.46 | 0.59* | $0.55^{\#}$ | | PPT | 0.32 | −0.36 | 0.26 |
| F | SS-CM 1 | | | | F | SS-CM 1 | | | |
| | SS-S 1 | 0.31 | | | | SS-S 1 | −0.26 | | |
| | OS-S | 0.27 | 0.34 | | | OS-S | $0.55^{\#}$ | −0.01 | |
| | PPT | 0.16 | −0.02 | 0.67* | | PPT | 0.83* | −0.17 | 0.45 |

C 1 = Cohort 1; C 2 = Cohort 2.
M = Male; F = Female.
* $p < 0.05$, # $p < 0.1$.
Significant correlations replicated in both cohorts are bracketed.

We conducted a similar correlational analysis for our nonsocial measure *distance moved*. Cohort 1 males had a significant correlation only between OS-S and PPT, whereas females exhibited a significant correlation between SS-CM 1 and SS-S 1, as well as the SS-S 1 and OS-S tests (Table 2).

Cohort 2 saw a replication of the male correlation in *distance moved* between OS-S and PPT and this same correlation appeared for cohort 2 females. The other female correlations did not appear in cohort 2, although cohort 2 females did show a correlation between SS-CM 1 and PPT (Table 2).

**Table 2**
Pearson's Correlations Coefficients for Nonsocial Distance Moved Behaviour.

| C 1 | Test Type | SS-CM 1 | SS-S 1 | OS-S | C 2 | Test Type | SS-CM 1 | SS-S 1 | OS-S |
|---|---|---|---|---|---|---|---|---|---|
| M | SS-CM 1 | | | | M | SS-CM 1 | | | |
| | SS-S 1 | −0.18 | | | | SS-S 1 | 0.39 | | |
| | OS-S | −0.58 | −0.12 | | | OS-S | 0.35 | 0.57# | |
| | PPT | 0.06# | −0.26 | [0.59*] | | PPT | 0.41 | 0.36 | [0.84*] |
| F | SS-CM 1 | | | | F | SS-CM 1 | | | |
| | SS-S 1 | 0.67* | | | | SS-S 1 | 0.55 | | |
| | OS-S | 0.48 | 0.80* | | | OS-S | 0.44 | 0.42 | |
| | PPT | 0.48 | −0.02 | 0.44 | | PPT | 0.80* | 0.53# | 0.66* |

C 1 = Cohort 1; C 2 = Cohort 2.
M = Male; F = Female.
* $p < 0.05$, # $p < 0.1$.
Significant correlations replicated in both cohorts are bracketed.

### 3.3. Consistency of behaviour within the same-sex cagemate test types

We then looked at consistency within test types. We first assessed the consistency of *huddling* behaviour over the two cagemate reciprocal interaction testing sessions (SS-CM 1 and SS-CM 2). We again split by sex and cohort, and then calculated Pearson's correlations. In cohort 1 males and females, there was a strong correlation in huddling behaviour between SS-CM 1 and SS-CM 2 (Males: r = 0.99, p = 0.00005; Females: r = 0.71, p = 0.01). In cohort 2, the SS-CM 1 and SS-CM 2 huddling correlation was repeated in females (r = 0.66, p = 0.01), but not in males (r = 0.39, p = 0.21).

A similar result was found using the ICC analysis (Table 3, top row). For *huddling* in the SS-CM context, males and females in cohort 1 exhibited significant, moderate to excellent reliability, whereas only females demonstrated significant reliability in cohort 2.

For our *distance moved* measure, the Pearson's correlations for cohort 1 revealed a significant correlation between SS-CM 1 and SS-CM 2 for males (r = 0.888, p = 0.018), but not for females (r = 0.352, p = 0.261). The correlation for males was repeated in cohort 2 (r = 0.633, p = 0.027) and was found for females as well (r = 0.947, p = 0.000003). The ICC analysis for *distance moved* revealed consistency. Over the two cohorts and both sexes, reliability was mostly significant, with the correlation coefficients ranging from moderate to excellent (Table 4, top row).

### 3.4. Consistency of behaviour in the first two same-sex stranger tests

In order to parallel the analysis for SS-CM in the previous section for SS-S test sessions, we initially focused our attention on the relationship between SS-S 1 and SS-S 2. In cohort 1, there was not a significant correlation for males (r = 0.367, p = 0.241), but there was for females (r = 0.663, p = 0.019). This significant finding did not hold for cohort 2: neither males (r = 0.332, p = 0.292) nor females (r = -0.098, p = 0.762) exhibited a significant correlation between SS-S 1 and SS-S 2 in huddling behaviour. The lack of consistency in huddling is also seen in the ICC analysis for reliability (Table 3, middle row).

For *distance moved*, the cohort 1 correlational analysis revealed SS-S 1 and SS-S 2 had a nearly significant correlation for females (r = 0.574, p = 0.051), but not for males (r = 0.143, p = 0.658). In cohort 2, it was significant for females (r = 0.619, p = 0.032), but not males (r = 0.386, p = 0.215). The ICC analysis found a lack of reliability for males, but moderate reliability for females (Table 4, middle row).

### 3.5. Intra-individual variation in the same-sex stranger tests

To more fully explore consistency of behavior, we used the data from all five SS-S tests to better characterize intra-individual variation. First, we calculated a full correlation matrix to examine the relationship for *huddling* between all SS-S tests in cohort 1. For males, there were no significant correlations between any pair of SS-S tests (R-values range from -0.269 to 0.512, with p-values ranging from 0.088 to 0.995). Essentially the same was found for females (R-values range from -0.168 to 0.503, with P-values ranging from 0.095 to 0.888), with two exceptions: cohort 1 female huddling in SS-S 1 correlated with SS-S 2 (r = 0.663, p = 0.019) and with SS-S 3 (r = 0.741, p = 0.006). In cohort 2, however, neither males nor females exhibited significant correlations between SS-S sessions (data not shown; all P-values > 0.121).

Again, we examined the reliability of behaviour using the ICC analysis. Looking at *huddling* across all five SS-S tests, there was poor to no consistency in reliability, regardless of sex and cohort (Table 3, bottom row).

We also examined intra-individual consistency in huddling behaviour by calculating intra-individual *ranges* and *95% confidence intervals*. Including both cohorts and both sexes, the average intra-individual range for SS-S huddling was nearly 85 min (47% of a 180 min test session, N = 48 animals), with some intra-individual ranges as high as 120 min (68% of a 180 min test session). 95% confidence intervals are plotted in Fig. 3. On average, 95% confidence interval widths averaged approximately 90 min (50% of a 180 min test session), with average lower and upper bounds equaling 6.6 min and 97.6 min, respectively.

**Table 3**
Intra-Class Correlations Assessing Repeatability of Huddling Behaviour.

| Cohort 1 | Sex | ICC | LB, UB | Cohort 2 | Sex | ICC | LB, UB |
|---|---|---|---|---|---|---|---|
| SS-CM 1-2 | M | 0.99* | 0.95, 1.0 | SS-CM 1-2 | M | 0.25 | −0.35, 0.71 |
| | F | [0.70*] | 0.24, 0.90 | | F | [0.68*] | 0.21, 0.90 |
| SS-S 1-2 | M | 0.36 | −0.25, 0.76 | SS-S 1-2 | M | 0.33 | −0.27, 0.75 |
| | F | 0.66* | 0.16, 0.89 | | F | −0.08 | −0.61, 0.50 |
| SS-S 1-5 | M | 0.14 | −0.05, 0.50 | SS-S 1-5 | M | 0.11 | −0.06, 0.44 |
| | F | 0.25* | 0.03, 0.59 | | F | 0.05 | −0.01, 0.37 |

ICC: Intra-Class Correlation Coefficient.
LB, UB: 95% Confidence Interval, Lower and Upper Bound.
* $p < 0.05$.
Significant correlations replicated in both cohorts are bracketed.

**Table 4**
Intra-Class Correlations Assessing Repeatability of Distance Moved Behaviour.

| Cohort 1 | Sex | ICC | LB, UB | Cohort 2 | Sex | ICC | LB, UB |
|---|---|---|---|---|---|---|---|
| SS-CM 1-2 | M | [0.87*] | 0.34, 0.98 | SS-CM 1-2 | M | [0.60*] | 0.06, 0.86 |
| | F | 0.32 | −0.28, 0.74 | | F | 0.91* | 0.71, 0.97 |
| SS-S 1-2 | M | 0.12 | −0.46, 0.63 | SS-S 1-2 | M | 0.37 | −0.27, 0.75 |
| | F | [0.57*] | 0.03, 0.86 | | F | [0.54*] | −0.02, 0.84 |
| SS-S 1-5 | M | [0.30*] | 0.06, 0.65 | SS-S 1-5 | M | [0.62*] | 0.37, 0.84 |
| | F | [0.49*] | 0.23, 0.77 | | F | [0.57*] | 0.32, 0.82 |

ICC: Intra-Class Correlation Coefficient.
LB, UB: 95% Confidence Interval, Lower and Upper Bound.
* $p < 0.05$.
Significant correlations replicated in both cohorts are bracketed.

After this triple analysis of huddling variation, we then turned to an examination of nonsocial *distance moved* behaviour. As with huddling, we split by sex and cohort and calculated Pearson's correlation matrices. Of the ten correlation coefficients calculated for cohort 1 males, only one was significant (SS-S 4 v SS-S 5: $r = 0.83$, $p = 0.002$); the other nine were non-significant ($p > 0.16$). For cohort 1 females, SS-S distance moved was significantly correlated for five of the ten bivariate analyses. A subsequent analysis of cohort 2 found eight of the ten bivariate analyses significantly correlated ($p < 0.05$) for males, and six of the ten bivariate analyses significantly correlated for females ($p < 0.05$; data not shown).

In general, intra-individual ranges and 95% confidence intervals for *distance moved* behaviour were narrower than for *huddling* (see Fig. 4). The ICC analysis also revealed much higher reliability across SS-S tests (Table 4, bottom row). While ICCs were only poor to moderate, distance moved exhibited significant reliability in both sexes in both cohorts.

### 3.6. Individual differences in behaviour

We then calculated the average *huddling* time exhibited by each individual and graphed it along with its 95% confidence interval. To facilitate visual comparisons between individuals, we ordered subjects by their average huddling score within each sex and cohort group (Fig. 3). We then examined whether there were significant differences in average huddling behaviour between individuals using a GLM with *individual animal* as the fixed factor and *average* SS—S *huddling behaviour of the stimulus animal* as a covariate. We recognize that using *individual animal* as a factor is unusual. But given that we have multiple measures per individual, we can use this procedure as an imperfect method for statistically assessing if there are individuals who significantly differ from one another (e.g., "high" vs "low" huddlers) or if all individuals look statistically similar.

For cohort 1, huddling varied significantly by individual animal for males ($F_{11,49} = 3.736$, $p = 0.001$, partial $\eta^2 = 0.45$) and females ($F_{11,49} = 10.37$, $p = 0.002$, partial $\eta^2 = 0.42$). In cohort 2 the intra-individual variability was so high that the GLM did not find significant differences between individual animals (Males: $F_{11,47} = 1.428$, $p = 0.19$, partial $\eta^2 = 0.25$; Females: $F_{11,47} = 1.60$, $p = 0.130$, partial $\eta^2 = 0.27$). The covariate was significant in both sexes in cohort 1 ($p < 0.01$, $d > 0.184$), but was not significant in cohort 2 in either sex ($p > 0.05$)

Due to the high intra-individual variability, we lacked the power to identify statistically significant differences between individuals if we Tukey-corrected for all possible comparisons. With the overall cohort 1 male and female GLMs significant, we examined whether the average huddling behaviour of individual animals could be statistically distinguished from one another by first splitting the data into 8 subgroups based on *cohort* (1, 2), *sex* (male, female), and *marking* (marked, unmarked). We then ran the same GLM analysis as above but included *post hoc* tests. Of the 8 subgroups, the factor *individual animal* was significant in one ($p < 0.05$, $d = 0.43$), with two others nearly significant ($p < 0.06$, $d = 0.347$ and 0.348); the other five subgroups contained no two individuals that were statistically distinguishable when using the Tukey correction. The overlap in huddling behavior between individual animals can be seen visually by comparing overlapping 95% confidence intervals in Fig. 3.

We conducted a similar analysis of individual differences for *distance moved* with *average SS-S distance moved behavior of the stimulus animal* as a covariate. Significant individual differences in distance moved were found for cohort 1 males ($F_{11,50} = 3.80$, $p = 0.001$, partial $\eta^2 = 0.48$) and females ($F_{11,50} = 5.33$, $p = 0.00002$, partial $\eta^2 = 0.56$). The covariate was significant in males ($p < 0.01$, partial $\eta^2 = 0.14$), but not females ($p = 0.54$). The effect of *individual animal* was repeated in cohort 2 (Males: $F_{11,48} = 13.419$, $p < 0.000001$,

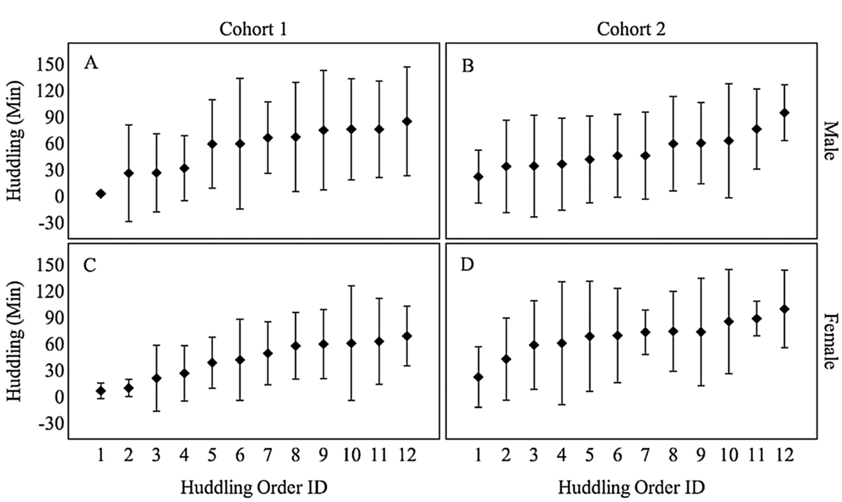



**Fig. 3.** Mean *huddling* behaviour for each test subject averaged over five SS—S tests. Interval bars represent 95% confidence intervals. Animals are ordered from lowest average huddling score to highest within each cohort-by-sex subgroup. Illustrates the remarkable inter- and intra-individual variation for huddling behavior.

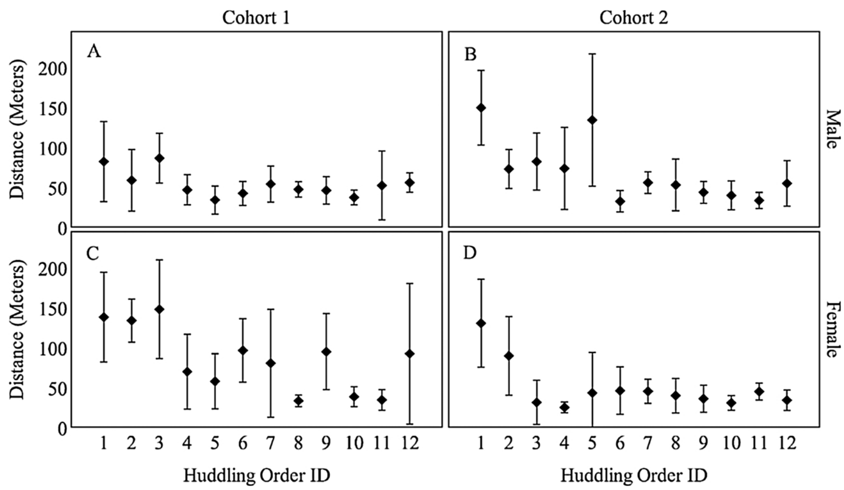



**Fig. 4.** Average *distance moved* behaviour for each test subject averaged over five SS—S tests. Interval bars represent 95% confidence intervals. Animals are ordered from lowest average huddling score to highest within each cohort-by-sex subgroup (same order as Fig. 3).

partial $\eta^2 = 0.66$; Females: $F1{,}48 = 13.42$, $p < 0.000001$, partial $\eta^2 = 0.76$). Here, the covariate was significant in both sexes ($p < 0.01$, partial $\eta^2 > 0.11$). See Fig. 4.

### 3.7. Effect of test type on individual averages of behaviour

In Section 3.1, we presented data examining the effect of different test types on *huddling behaviour* during the first experience of each test. Section 3.5, however, indicates that individual animals may exhibit widely different huddling behaviour from one test session to the next, particularly in the SS-S test type. We therefore used the intra-individual averages calculated in Section 3.6 to examine the effect of test type on average SS-CM and average SS-S huddling behaviour (Fig. 5).

In cohort 1, males exhibited significantly more average huddling behaviour in the SS-CM testing context than the SS-S testing context ($F1{,}11 = 5.70$, $p = 0.04$, partial $\eta^2 = 0.34$); females exhibited a non-significant trend in the same direction ($F1{,}11 = 4.14$, $p = 0.07$, partial $\eta^2 = 0.27$). These differences were replicated in cohort 2, with females

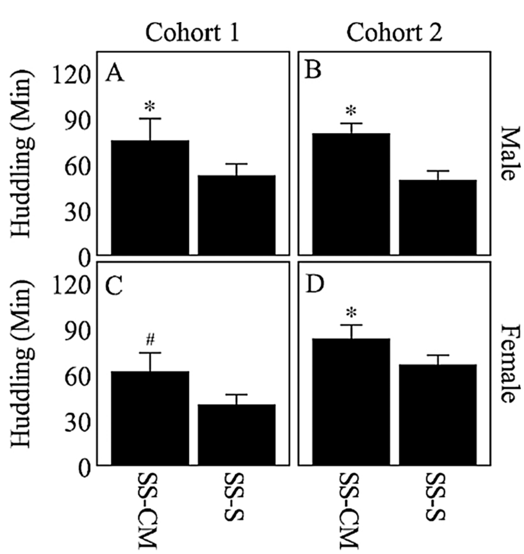



**Fig. 5.** The total *huddling* duration for males (**A** and **B**) and females (**C** and **D**) in cohort 1 (**A** and **C**) and cohort 2 (**B** and **D**). Unlike Fig. 1, which presents data from only the first exposure to each test type, these data account for each individual's average behavior. By calculating average SS—CM (sessions 1–2) and SS—S (sessions 1—–5) huddling for each subject first and then calculating group averages, we see consistently less huddling in the SS—S domain than in the SS—CM. Bars represent means ± SEM. * $p < 0.05$; # $p = 0.07$.

being significant this time (Males: $F1{,}11 = 22.22$, $p = 0.001$, partial $\eta^2 = 0.67$; Females: $F1{,}11 = 5.72$, $p = 0.04$, partial $\eta^2 = 0.34$). Fig. 5 illustrates the consistency of the difference between SS-CM and SS-S huddling between sexes and cohorts.

### 3.8. Consistency of behaviour for individual averages of behaviour

With intra-individual averages revealing previously unseen effects (Section 3.7), we tested if intra-individual average behaviour in one test type (SS-CM 1–2) would consistently correlate with intra-individual average behaviour in another (SS-S 1–5). Cohort 1 Pearson's correlations revealed significant, positive relationships for cohort 1 male huddling averages between SS-CM 1–2 and SS-S 1–5, whereas females showed a non-significant relationship (Table 5). The intra-individual averages did not significantly correlate with OS-S or PPT behavior.

In cohort 2, there were no significant correlations between test types when using intra-individual average huddling as the measured variable (Table 5).

We conducted a similar correlational analysis for *distance moved*. Here, the Pearson's analyses revealed a significant positive correlation between intra-individual average behaviour in the SS-CM 1–2 and SS-S 1–5 in males and females in cohort 1 (Table 6). In males, this was the only significant correlation. In females, huddling in each test type significantly correlated with every other test type.

Cohort 2 replicated the significant correlation for distance moved between SS-CM 1–2 and SS-S 1–5 in both sexes. In females, SS-CM 1–2 and SS-S 1–5 significantly correlated with PPT locomotor activity in both cohorts (Table 6).

**Table 5**
Pearson's Correlation Coefficients for Social Huddling Behaviour.

| C 1 | Test Type | SS-CM 1-2 | SS-S 1-5 | C 2 | Test Type | SS-CM 1-2 | SS-S 1-5 |
|---|---|---|---|---|---|---|---|
| M | SS-CM 1-2 | | | M | SS-CM 1-2 | | |
| | SS-S 1-5 | 0.76* | | | SS-S 1-5 | 0.32 | |
| | OS-S | −0.11 | 0.36 | | OS-S | 0.03 | 0.40 |
| | PPT | 0.54# | 0.56# | | PPT | 0.40 | −0.04 |
| F | SS-CM 1-2 | | | F | SS-CM 1-2 | | |
| | SS-S 1-5 | 0.52# | | | SS-S 1-5 | 0.43 | |
| | OS-S | −0.01 | 0.43 | | OS-S | 0.34 | 0.50 |
| | PPT | 0.10 | 0.04 | | PPT | 0.51# | 0.31 |

C 1 = Cohort 1; C 2 = Cohort 2.
M = Male; F = Female.
* $p < 0.05$, # $p < 0.1$.
Significant correlations replicated in both cohorts are bracketed.

**Table 6**
Pearson's Correlation Coefficients for Nonsocial Distance Moved Behaviour.

| C 1 | Test Type | SS-CM 1-2 | SS-S 1-5 | C 2 | Test Type | SS-CM 1-2 | SS-S 1-5 |
|---|---|---|---|---|---|---|---|
| M | SS-CM 1-2 | | | M | SS-CM 1-2 | | |
| | SS-S 1-5 | [0.73*] | | | SS-S 1-5 | [0.67*] | |
| | OS-S | 0.54# | 0.20 | | OS-S | 0.54# | 0.88* |
| | PPT | 0.37 | −.0.01 | | PPT | 0.56# | 0.79* |
| F | SS-CM 1-2 | | | F | SS-CM 1-2 | | |
| | SS-S 1-5 | [0.80*] | | | SS-S 1-5 | [0.90*] | |
| | OS-S | 0.71* | 0.85* | | OS-S | 0.47 | 0.45 |
| | PPT | [0.67*] | [0.72*] | | PPT | [0.82*] | [0.88*] |

C 1 = Cohort 1; C 2 = Cohort 2.
M = Male; F = Female.
* $p < 0.05$, # $p < 0.1$.
Significant correlations replicated in both cohorts are bracketed.

### *3.9. Partner preference test behaviour*

Finally, we assessed stranger-vs-partner huddling preference in the PPT. In cohort 1, related-samples Wilcoxon signed-rank tests indicated that males and females huddled with partners more than strangers (Males: Partner Median = 60.23, Stranger Median = 9.65, Z = -2.51, r = 0.51, p = 0.01; Females: Partner Median = 87.14, Stranger Median = 0.83, Z = -2.59, r = 0.53, p = 0.01; Fig. 6A). We observed the same pattern of partner preference in cohort 2 (Males: Partner Median = 61.16, Stranger Median = 9.96, Z = -2.04, r = 0.42, p = 0.04; Females: Partner Median = 85.90, Stranger Median = 1.27, Z = -3.06, r = 0.62 p = 0.002; Fig. 6B).

## 4. Discussion

To our knowledge, this is the first study to systematically assess the stability of individual differences in prairie vole huddling behaviour across multiple types of long-duration social tests and over multiple test sessions within test types. Importantly, this assessment included testing males and females and a second cohort to test replicability. We predicted that prairie voles would exhibit a stable behavioral type such that there would be significant correlations in huddling behavior across test types and repeated test sessions. In most respects, our data did not support these predictions (Table 7). Instead, variation in huddling appears to be driven more by domain- and state-factors, which resulted in

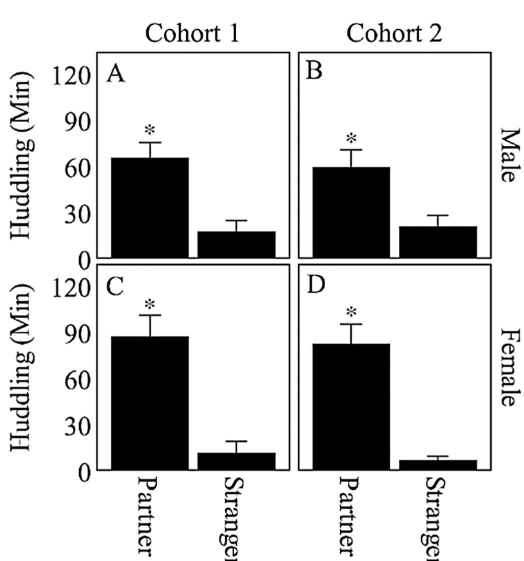



**Fig. 6.** In cohorts 1 (**A**) and 2 (**B**), males and females huddled significantly more with the partner than the stranger during the 3 -h PPT ($p < 0.05$ for all). Bars represent means ± SEM. * $p < 0.05$.

relatively inconsistent huddling behavior across sessions (e.g., SS-S 1–5), test types (SS-CM 1, SS-S 1, OS-S, PPT), and cohorts (1 and 2).

Importantly, distance moved behavior was more consistent. Intra-individual variation was generally narrower (Fig. 4), subjects were generally more reliable even in the presence of different strangers (Table 4), and context-driven differences in locomotion were replicable across cohorts and sexes (Fig. 2).

These lines of analysis indicate that, unlike a nonsocial behavior like locomotor activity, a single test session to measure huddling behavior may not be sufficient to characterize a stable sociobehavioral type of an individual animal. Both the type of test (domain) and the state induced by the stimulus animal seem to drive variation in huddling duration more than individual differences in behavioral type.

### *4.1. The influence of testing domain on behaviour*

Our findings on the domain-dependence of huddling duration aligns with several recent studies indicating that prairie vole social behavior is context dependent. For example, Beery and colleagues compared prairie vole social behavior in the standard PPT versus in the short social investigation tasks typically used with mice (Beery et al., 2018). There was no relationship: behavior in the short tasks did not correspond with PPT behavior. The same group also found no relationship in behavior between the PPT and the socially conditioned place preference test (Goodwin et al., 2018). Both studies suggest there is a strong domain-driven component to the expression of social behavior. One counterargument is that, in comparison to the PPT, these tests do not allow animals to fully interact, the testing duration is different, and the non PPT tests are more olfactory dependent. It could be that these tests are measuring different aspects of social behavior rather than measuring the same social behavior in different domains.

Our study addresses these concerns. Our four testing domains were comparable in terms of duration and the ability to fully interact. Initially, our results indicated that if social behaviour is measured the same way (huddling duration) for the same duration (3 h) in all tests, domain exerts only a modest effect in females and seemingly none in males. Cohort 1 males averaged similar huddling durations across test-types, and females appeared consistent in huddling duration across the reciprocal interaction tests (SS-CM, SS-S, and OS-S), only increasing significantly in the PPT (Fig. 1). Considered together, we seemed to be able to draw two conclusions: (i) with a single stimulus animal and given equal testing durations, huddling behaviour is domain-general; and (ii) females, but not males, might be sensitive to a shift in the number of social stimuli (compare Fig. 1A vs C).

Further analysis, however, undermines this conclusion. Subjects in cohort 2 exhibited a strikingly different pattern of huddling behaviour across test types (Fig. 1B and D). If we only had these data, we would have concluded that huddling is strongly driven by context and is thus domain-specific.

A similar discrepancy occurs in Table 1, where we examined the first exposure to each test type. Cohort 1 suggests that measuring male huddling in the SS-CM context would allow us to predict behaviour in the SS-S context, and the SS-S context would allow prediction of PPT huddling. But these cross-domain correlations did not replicate in cohort 2. Cohort 2 would again argue for distinct behavioural profiles in the different domains.

### *4.2. The importance of direct replication*

This discrepancy between cohort 1 and 2 deserves greater attention. Our use of two cohorts raises the larger issues of replicability and generalizability—issues that have gained increasing attention in several fields (Collaboration, 2015; Higginson and Munafò, 2016; Nosek and Errington, 2017), including social neuroscience (Walum et al., 2016). As seen throughout this study, several significant findings made with cohort 1 did not repeat in cohort 2.

**Table 7**
Summary Overview.

**Hypothesis:** Stable differences in behavioral type are the primary driver of individual variation, so social behaviour should be relatively consistent across domains (i.e., low domain-dependent variation) and test sessions (i.e., low state-dependent variation) in males and females, and this consistency should be replicable.

| Prediction | Test | Results | Interpretation |
|---|---|---|---|
| Huddling would be consistent across test types | Compare huddling durations across SS-CM 1, SS-S 1, OS-S, and PPT tests | Generally, huddling varied by test type (cohort 2 males, both cohorts of females) | Huddling behavior appears to be domain-dependent |
| Huddling in one test type would predict huddling in another (i.e., "high" huddlers would exhibit high huddling durations and "low" huddlers, low) | Pearson's correlations of huddling duration across SS-CM 1, SS-S 1, OS-S, and PPT tests | Generally, there were few correlations, and correlations that did occur typically did not replicate | Huddling in one test type does not predict huddling in another. Single test sessions may not adequately characterize "high" and "low" huddlers |
| Huddling in one test session would correlate with huddling in subsequent test sessions; intra-individual variation in huddling duration would be low | ICC tests of reliability of huddling durations in SS-CM 1-2 and then in the SS-S 1-5 tests; comparison of intra-individual 95% confidence intervals | Generally, huddling with the same cagemate was reliable; huddling duration with different strangers was not reliable due to high intra-individual variation | Individuals express different levels of huddling behavior depending on the stimulus animal; i.e., huddling appears to be state-dependent |
| Significant results obtained from ~12 prairie voles would replicate in a second cohort of ~12 subjects. | Statistical tests conducted using cohort 1 or cohort 2 data | Generally, only a limited set of significant results occurred in both cohorts. High intra-individual variation occurred in all cohorts. | High variability in huddling behavior due to state- and domain-dependent factors can affect replicability of seemingly significant findings |

One explanation for the discrepancy is inconsistent testing. While this explanation cannot be fully excluded, the data presented in Fig. 2 argue against it. In both sexes and both cohorts, the pattern for distance moved is consistent between all four sex-by-cohort groups. We suspect prairie vole activity is dependent on domain factors such as the number and configuration of social stimuli, rather than the specific type of social stimulus (familiar vs unfamiliar, same-sex vs opposite sex). Given the replicability of locomotor activity and the use of unbiased automated behaviour scoring of huddling, the inter-cohort discrepancies in huddling duration might be due to domain- and state-dependent factors. This study also emphasizes the need for direct replication and repetitive testing in the literature: sample sizes of ~12 subjects per group, which is common in prairie vole studies of social behaviour, may reduce animal usage, but it can also lead to spurious and potentially contradictory findings (e.g., Fig. 1).

### *4.3. The contribution of state-dependent variation*

In a sense, these same concerns of repeatability arise for individual animals. We began this study expecting subjects to exhibit relatively stable huddling behaviour across test sessions. By conducting repetitive testing, we found that state-dependent variation should not be overlooked.

In the two SS-CM tests, huddling was relatively consistent in females and possibly males (Section 3.3 and Table 3). This suggests (i) stochastic differences in internal state likely drive some of the differences in huddling observed between test session 1 and 2, and (ii) being tested with a familiar cagemate seems to induce a similar enough state to produce moderately repeatable huddling behaviour (Table 3).

This contrasts with the SS-S tests. In response to a different unfamiliar stranger in each test session, we observed not only remarkable variation between individuals but also remarkable intra-individual variation (i.e., a lack of individual consistency; Table 4). Lack of familiarity with the stimulus animal appears to amplify the stochastic differences in state-dependent variation on a session-to-session basis, leading to high intra-individual variability across test sessions. Interestingly, the lack of familiarity did not amplify state-driven variation enough to obscure individual differences in general activity. Distance moved was relatively stable for most individuals (Fig. 4 and Table 4), a stability that has been noted in other vole species and testing contexts (e.g., Eilam, 2003).

Overall, our data suggest that a single test session to measure huddling as a proxy for social behaviour may not be a reliable characterization of an animal's sociobehavioral type. Moreover, our results suggest that the measure of social behaviour in the PPT may not correspond to social behaviour in other contexts even if the same metric (huddling) is used. What, then, does that mean for using huddling in the PPT and reciprocal interaction tests as proxy measures of social behaviour?

It is important to acknowledge that reciprocal interaction tests are not the same as PPTs. In the former, both animals are free roaming; in the latter the test animal is free, but the stimulus animals are confined to specific locations (Ahern et al., 2009; Williams et al., 1992). In the PPT, huddling behaviour might exhibit less intra-individual variation across test sessions. Unfortunately, it is unknown because only a few studies have attempted repeat PPT testing (Ahern and Young, 2009; Blocker and Ophir, 2016; Perkeybile et al., 2013; Sun et al., 2014) and they tended to focus only on the consistency of group differences or a different experimental question in each test. For example, Blocker and Ophir (2016) tested prairie voles in multiple PPTs, but in each, the subject was given a different choice of stimulus animals. The use of reciprocal interaction tests may allow for repetitive testing, but our data comparing test types suggest the results may not correlate with what is being measured in the PPT.

Another difference between the reciprocal interaction and PPT tests is the state of the experimental and stimulus animals. From early (for review, Carter et al., 1995) to more recent (Amadei et al., 2017) high impact PPT studies, females are often estradiol-primed to encourage mating and facilitate bonding during cohabitation. This procedure likely narrows the range of states the test and stimulus animals can adopt, which increases consistency but likely narrows generalizability. The test may characterize only one narrowly defined aspect of social beahvior.

Both issues may explain why studies focused on manipulating a one-time measure of PPT huddling behavior sometimes do not translate or repeat. For example, laboratory PPT studies identified the importance of variation in the arginine vasopressin system in the ventral pallidum for partner preference formation as measured by huddling duration, but the importance of this variation did not translate to field studies of mating strategy nor mating success (Okhovat et al., 2015; Ophir et al., 2008). Similarly, a recent study found little evidence for the heritability of huddling duration within the PPT (Vogel et al., 2018). The nature of huddling variability we observed here highlights the need to consider the many potential sources of variation in explaining behavioural outcomes. These include developmental factors, social and ecological contexts, genetic predispositions (and gene x environment interactions), and the forces of evolution that act upon these and other important factors (Dingemanse and Wolf, 2013; Dochtermann et al., 2015; Foster, 2013; Hayes and Jenkins, 1997; Schradin et al., 2012; Snell-Rood, 2013; Zimmermann et al., 2017). Any one or more of these factors working in combination could provide mechanistic explanations for behavioural variation in huddling. For instance, early life social environment can impact the latency or ability to form pair bonds in prairie voles (Ahern and Young, 2009; Bales et al., 2007). Furthermore, allelic variation and epigenetic modification of that variation, and population structure and dynamics within a natural community appear to account for differences in predispositions to establish monogamous pair bonds (Okhovat et al., 2015). Each of these factors has the potential to interact with the context and the state of the animal at the time of being tested. Study designs that account for differences in domain and state are necessary to understand how these factors shape behavior.

Together, our data and a growing literature suggest that one-time tests of laboratory PPT huddling, particularly using estradiol-primed females, might not reveal behavioural differences that are stable or meaningful enough to characterize how the range of factors described above affect multiple dimensions of sociality. This becomes an even greater challenge when translating to human behaviour and behavioural disorders, such as autism spectrum. Much human social behaviour occurs in nonreproductive, free-roaming contexts (domains) with people who can induce different internal states. It is important therefore to leverage comparable animal models and testing paradigms (Goodwin et al., 2018; Kelly and Ophir, 2015). Likewise, researchers interested in the stability of typical and atypical behaviour might benefit from pre-clinical repetitive testing designs to better distinguish behavioural types and account for domain- and state-specific variation.

### *4.4. Re-evaluating behavioural types*

Considering the extent to which state- (and possibly domain-) specific variation influenced huddling behaviour in our data, we wondered if we could identify highly social versus less social subjects at all – e.g., high huddlers vs low huddlers. We wondered the same about high versus low activity individuals. Repeated observations provided a way to tease this out. Indeed, repeat observations of parenting behaviour is common practice when distinguishing high versus low licking and grooming rats or the amount of care received by biparentally and damonly reared prairie vole offspring (Ahern et al., 2011; Caldji et al., 1998; Francis et al., 1999; Hammock and Young, 2005; Meaney, 2001; Perkeybile et al., 2013).

For activity, it was possible to distinguish high versus low activity subjects (non-overlapping 95% CIs in Fig. 4). Identifying high versus low huddlers, however, was less clear (Fig. 3). There were animals with higher average huddling durations than others, but there was also extensive overlap: one subject might be counted as a high huddler in one test session, but a low huddler in the next. In short, any given test session gave, at best, a highly uncertain measure of an animal's behavioural type. Taken together, this view of the data would argue for a lack of distinct behavioural types for huddling.

But that would ignore the extreme ends of each panel in Fig. 3. While imperfect, combining our analysis in Section 3.6 with a visual examination of Fig. 3 provides evidence for significant differences in huddling behaviour between individuals; at least a few prairie voles consistently huddle at a high level, whereas a few consistently huddle only a little (e.g., compare the right versus left side in each panel). Viewed this way, the findings argue that at least some animals have a relatively stable high or low (social) huddling type.

Future research will need to assess whether these average differences in huddling are meaningful. It might be that huddling is, in general, too state-dependent to distinguish individual differences in sociality on a consistent basis or average behaviour is not relevant. Alternatively, using average huddling behaviour as a proxy for behavioural type might provide insight into the contribution of known (and unknown) genetic and neurobiological factors to individual differences in sociality that might better translate to other species. It is also possible that intra-individual huddling behaviour might co-vary with other social and non-social behaviours in a way that a single measure of huddling might not. Thus average huddling might contribute to an overarching behavioural syndrome. In rodents and even voles, behavioural syndromes (or correlated suites of behaviour in which consistent behaviour in one context spills over into another to produce 'personality types'; Sih et al., 2004b) have been identified, such as aggressiveness (Crews et al., 2009; Koolhaas et al., 1999; Korpela et al., 2011), anxiousness and emotionality (Auclair et al., 2013; Korpela et al., 2011), and riskiness, boldness, and exploration (Eilam, 2003; Herde and Eccard, 2013; Korpela et al., 2011); although here too, the robustness and stability of these traits is not always clear (Gracceva et al., 2011; Herde and Eccard, 2013).

## 5. Conclusion

Overall, the remarkable variation in prairie vole social behaviour—and likely that of other outbred species—arises in part due to differences in genetics and neurobiology. The data presented here and elsewhere, however, is painting a more complex picture than has been previously appreciated. Domain- and state-dependent variation also appears to exert a significant influence and they might mask differences in behavioural type and obscure the contributions of specific genes, neurobiological variants, gene-by-rearing interactions, etc. Exploring multiple testing paradigms and utilizing repetitive testing will be important for gaining a fuller understanding of social behaviour and individual variation.

## Author contributions

TA conceived, organized, and conducted all parts of the study. TA acquired the data, TA and DB performed the statistical analyses, TA and AO drafted the manuscript, TA, AO, and DB contributed substantial revisions.

## Funding

Funding: This work was supported by Quinnipiac University's College of Arts & Sciences (TA) and a NARSAD Young Investigator Grant from the Brain & Behaviour Research Foundation (TA).

## Declaration of Competing Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Acknowledgements

I thank Holli Zampano, Lena Grahm, and Emily Lucibella for their excellent animal care, and Melissa Boucher for her work in helping me explore different ways to conduct and quantify reciprocal interaction tests. I also thank the referees.

## References

Ahern, T.H., Hammock, E.A.D., Young, L.J., 2011. Parental division of labor, coordination, and the effects of family structure on parenting in Monogamous Prairie Voles (Microtus ochrogaster). Dev. Psychobiol. 53, 118–131. https://doi.org/10.1002/dev.20498.

Ahern, T.H., Modi, M.E., Burkett, J.P., Young, L.J., 2009. Evaluation of two automated metrics for analyzing partner preference tests. J. Neurosci. Methods. https://doi.org/10.1016/j.jneumeth.2009.06.010.

Ahern, T.H., Young, L.J., 2009. The impact of early life family structure on adult social attachment, alloparental behavior, and the neuropeptide systems regulating affiliative behaviors in the monogamous prairie vole (microtus ochrogaster). Front. Behav. Neurosci. 3, 17. https://doi.org/10.3389/neuro.08.017.2009.

Amadei, E.A., Johnson, Z.V., Kwon, Y.J., Shpiner, A.C., Saravanan, V., Mays, W.D., Ryan, S.J., Walum, H., Rainnie, D.G., Young, L.J., Liu, R.C., 2017. Dynamic corticostriatal activity biases social bonding in monogamous female prairie voles. Nature 546, 297–301. https://doi.org/10.1038/nature22381.

Anacker, A.M.J., Beery, A.K., 2013. Life in groups: the roles of oxytocin in mammalian sociality. Front. Behav. Neurosci. 7. https://doi.org/10.3389/fnbeh.2013.00185.

Auclair, Y., König, B., Lindholm, A.K., 2013. A selfish genetic element influencing

longevity correlates with reactive behavioural traits in female house mice (Mus domesticus). PLoS One 8. https://doi.org/10.1371/journal.pone.0067130.
Bales, K.L., Lewis-Reese, A.D., Pfeifer, L.A., Kramer, K.M., Carter, C.S., 2007. Early experience affects the traits of monogamy in a sexually dimorphic manner. Dev. Psychobiol. 49, 335–342. https://doi.org/10.1002/dev.20216.
Beery, A.K., Christensen, J.D., Lee, N.S., Blandino, K.L., 2018. Specificity in sociality: mice and prairie voles exhibit different patterns of peer affiliation. Front. Behav. Neurosci. 12. https://doi.org/10.3389/fnbeh.2018.00050.
Bell, A.M., 2007. Future directions in behavioural syndromes research. Proc. Biol. Sci. 274, 755–761. https://doi.org/10.1098/rspb.2006.0199.
Bell, A.M., Stamps, J.A., 2004. Development of behavioural differences between individuals and populations of sticklebacks, Gasterosteus aculeatus. Anim. Behav. 68, 1339–1348. https://doi.org/10.1016/j.anbehav.2004.05.007.
Blocker, T.D., Ophir, A.G., 2016. A preference to bond? Male prairie voles form pair bonds even in the presence of multiple receptive females. Anim. Behav. 122, 89–97. https://doi.org/10.1016/j.anbehav.2016.10.007.
Budaev, S.V., 1997. “Personality” in the guppy (Poecilia reticulata): a correlational study of exploratory behavior and social tendency. J. Comp. Psychol. 111, 399–411. https://doi.org/10.1037/0735-7036.111.4.399.
Caldji, C., Tannenbaum, B., Sharma, S., Francis, D., Plotsky, P.M., Meaney, M.J., 1998. Maternal care during infancy regulates the development of neural systems mediating the expression of fearfulness in the rat. Proc. Natl. Acad. Sci. U. S. A. 95, 5335–5340.
Carter, C.S., DeVries, A.C., Getz, L.L., 1995. Physiological substrates of mammalian monogamy: the prairie vole model. Neurosci. Biobehav. Rev. 19, 303–314 https://doi.org/7630584.
Chang, Y.-C., Cole, T.B., Costa, L.G., 2017. Behavioral phenotyping for autism spectrum disorders in mice. Curr. Protoc. Toxicol. 72https://doi.org/10.1002/cptx.19. 11.22. 1-11.22.21.
Collaboration, O.S., 2015. Estimating the reproducibility of psychological science. Science 349, aac4716. https://doi.org/10.1126/science.aac4716.
Crews, D., Rushworth, D., Gonzalez-Lima, F., Ogawa, S., 2009. Litter environment affects behavior and brain metabolic activity of adult knockout mice. Front. Behav. Neurosci. 3. https://doi.org/10.3389/neuro.08.012.2009.
Defensor, E.B., Pearson, B.L., Pobbe, R.L.H., Bolivar, V.J., Blanchard, D.C., Blanchard, R.J., 2011. A novel social proximity test suggests patterns of social avoidance and gaze aversion-like behavior in BTBR T+ tf/J mice. Behav. Brain Res. 217, 302–308. https://doi.org/10.1016/j.bbr.2010.10.033.
Depue, R.A., Collins, P.F., 1999. Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. Behav. Brain Sci. 22, 491–517 discussion 518-569.
Dingemanse, N.J., Wolf, M., 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. Anim. Behav. Including Special Sect.: Behav. Plasticity Evol. 85, 1031–1039. https://doi.org/10.1016/j.anbehav.2012.12.032.
Dochtermann, N.A., Schwab, T., Sih, A., 2015. The contribution of additive genetic variation to personality variation: heritability of personality. Proc. Biol. Sci. 282. https://doi.org/10.1098/rspb.2014.2201.
Drummond, H., Gordon, E.R., 1979. Luring in the neonate alligator snapping turtle (Macroclemys temminckii): description and experimental analysis. Z. Tierpsychol. 50, 136–152. https://doi.org/10.1111/j.1439-0310.1979.tb01022.x.
Eilam, D., 2003. Open-field behavior withstands drastic changes in arena size. Behav. Brain Res. 142, 53–62. https://doi.org/10.1016/S0166-4328(02)00382-0.
Eysenck, H.J., 1967. The Biological Basis of Personality, 1967. Charles C Thomas.
Fahrenberg, J., 1977. Physiological concepts in personality research. Handbook of Modern Personality Theory. Hemisphere, pp. 585–611.
Foster, S.A., 2013. Evolution of behavioural phenotypes: influences of ancestry and expression. Anim. Behav. Including Special Sect.: Behav. Plasticity Evol. 85, 1061–1075. https://doi.org/10.1016/j.anbehav.2013.02.008.
Francis, D.D., Caldji, C., Champagne, F., Plotsky, P.M., Meaney, M.J., 1999. The role of corticotropin-releasing factor-norepinephrine systems in mediating the effects of early experience on the development of behavioral and endocrine responses to stress. Biol. Psychiatry 46, 1153–1166. https://doi.org/10.1016/S0006-3223(99)00237-1.
Getz, L., McGuire, B., 1993. A comparison of living singly and in male-female pairs in the prairie vole, Microtus ochrogaster. Ethology 94, 265–278.
Getz, L.L., Carter, C.S., Gavish, L., 1981. The mating system of the prairie vole, Microtus ochrogaster: field and laboratory evidence for pair-bonding. Behav. Ecol. Sociobiol. 8, 189–194.
Goodwin, N.L., Lopez, S.A., Lee, N.S., Beery, A.K., 2018. Comparative role of reward in long-term peer and mate relationships in voles. Horm. Behav. https://doi.org/10.1016/j.yhbeh.2018.10.012.
Gracceva, G., Koolhaas, J.M., Groothuis, T.G.G., 2011. Does the early social environment affect structure and consistency of personality in wild-type male rats? Dev. Psychobiol. 53, 614–623. https://doi.org/10.1002/dev.20586.
Guastella, A.J., Hickie, I.B., 2016. Oxytocin treatment, circuitry, and autism: a critical review of the literature placing oxytocin into the autism context. Biol. Psychiatry 79, 234–242. https://doi.org/10.1016/j.biopsych.2015.06.028.
Hammock, E.A.D., Young, L.J., 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science 308, 1630–1634 https://doi.org/308/5728/1630.
Hayes, J.P., Jenkins, S.H., 1997. Individual variation in mammals. J. Mammal. 78, 274–293. https://doi.org/10.2307/1382882.
Herde, A., Eccard, J.A., 2013. Consistency in boldness, activity and exploration at different stages of life. BMC Ecol. 13, 49. https://doi.org/10.1186/1472-6785-13-49.
Higginson, A.D., Munafò, M.R., 2016. Current incentives for scientists lead to underpowered studies with erroneous conclusions. PLoS Biol. 14. https://doi.org/10.1371/journal.pbio.2000995.
Hinkle, D.E., Wiersma, W., Jurs, S.G., 2003. Applied Statistics for the Behavioral Sciences. Houghton Mifflin.
Kappeler, P.M., 2019. A framework for studying social complexity. Behav. Ecol. Sociobiol. 73, 13. https://doi.org/10.1007/s00265-018-2601-8.
Kelly, A.M., Ophir, A.G., 2015. Compared to what: what can we say about nonapeptide function and social behavior without a frame of reference? Curr. Opin. Behav. Sci. Integr. Study Anim. Behav. 6, 97–103. https://doi.org/10.1016/j.cobeha.2015.10.010.
King, L.B., Walum, H., Inoue, K., Eyrich, N.W., Young, L.J., 2016. Variation in the oxytocin receptor gene predicts brain region specific expression and social attachment. Biol. Psychiatry 80, 160–169. https://doi.org/10.1016/j.biopsych.2015.12.008.
Koolhaas, J.M., de Boer, S.F., Buwalda, B., van Reenen, K., 2007. Individual variation in coping with stress: a multidimensional approach of ultimate and proximate mechanisms. Brain Behav. Evol. 70, 218–226. https://doi.org/10.1159/000105485.
Koolhaas, J.M., de Boer, S.F., Coppens, C.M., Buwalda, B., 2010. Neuroendocrinology of coping styles: towards understanding the biology of individual variation. Front. Neuroendocrinol. 31, 307–321. https://doi.org/10.1016/j.yfrne.2010.04.001.
Koolhaas, J.M., Korte, S.M., De Boer, S.F., Van Der Vegt, B.J., Van Reenen, C.G., Hopster, H., De Jong, I.C., Ruis, M.A.W., Blokhuis, H.J., 1999. Coping styles in animals: current status in behavior and stress-physiology. Neurosci. Biobehav. Rev. 23, 925–935. https://doi.org/10.1016/S0149-7634(99)00026-3.
Korpela, K., Sundell, J., Ylönen, H., 2011. Does personality in small rodents vary depending on population density? Oecologia 165, 67–77. https://doi.org/10.1007/s00442-010-1810-2.
Lott, D.F., 1984. Intraspecific variation in the social systems of wild vertebrates. Behaviour 88, 266–325.
McGraw, L.A., Young, L.J., 2009. The prairie vole: an emerging model organism for understanding the social brain. Trends Neurosci. https://doi.org/10.1016/j.tins.2009.11.006.
Meaney, M.J., 2001. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. Annu. Rev. Neurosci. 24, 1161–1192 https://doi.org/11520931.
Nosek, B.A., Errington, T.M., 2017. Making sense of replications. eLife 6. https://doi.org/10.7554/eLife.23383.
Okhovat, M., Berrio, A., Wallace, G., Ophir, A.G., Phelps, S.M., 2015. Sexual fidelity trade-offs promote regulatory variation in the prairie vole brain. Science 350, 1371–1374. https://doi.org/10.1126/science.aac5791.
Ophir, A.G., Campbell, P., Hanna, K., Phelps, S.M., 2008. Field tests of cis-regulatory variation at the prairie vole avpr1a locus: association with V1aR abundance but not sexual or social fidelity. Horm. Behav. 54, 694–702. https://doi.org/10.1016/j.yhbeh.2008.07.009.
Perkeybile, A.M., Griffin, L.L., Bales, K.L., 2013. Natural variation in early parental care correlates with social behaviors in adolescent prairie voles (Microtus ochrogaster). Front. Behav. Neurosci. 7. https://doi.org/10.3389/fnbeh.2013.00021.
Phelps, S.M., Campbell, P., Zheng, D.-J., Ophir, A.G., 2009. Beating the boojum: comparative approaches to the neurobiology of social behavior. Neuropharmacology. https://doi.org/10.1016/j.neuropharm.2009.06.043.
Pinter-Wollman, N., Keiser, C.N., Wollman, R., Pruitt, J.N., 2016. The effect of keystone individuals on collective outcomes can Be mediated through interactions or behavioral persistence. Am. Nat. 188, 240–252. https://doi.org/10.1086/687235.
Rapin, I., Tuchman, R.F., 2008. What is new in autism? Curr. Opin. Neurol. 21, 143–149. https://doi.org/10.1097/WCO.0b013e3282f49579.
Schradin, C., Lindholm, A.K., Johannesen, J., Schoepf, I., Yuen, C.-H., König, B., Pillay, N., 2012. Social flexibility and social evolution in mammals: a case study of the African striped mouse (Rhabdomys pumilio). Mol. Ecol. 21, 541–553. https://doi.org/10.1111/j.1365-294X.2011.05256.x.
Sih, A., Bell, A., Johnson, J.C., 2004a. Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol. Evol. (Amst.) 19, 372–378. https://doi.org/10.1016/j.tree.2004.04.009.
Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004b. Behavioral syndromes: an intergrative overiew. Q. Rev. Biol. 79, 241–277.
Sloan Wilson, D., Clark, A.B., Coleman, K., Dearstyne, T., 1994. Shyness and boldness in humans and other animals. Trends Ecol. Evol. (Amst.) 9, 442–446. https://doi.org/10.1016/0169-5347(94)90134-1.
Snell-Rood, E.C., 2013. An overview of the evolutionary causes and consequences of behavioural plasticity. Anim. Behav. Including Special Sect.: Behav. Plasticity Evol. 85, 1004–1011. https://doi.org/10.1016/j.anbehav.2012.12.031.
Stevenson, T.J., Alward, B.A., Ebling, F.J.P., Fernald, R.D., Kelly, A., Ophir, A.G., 2018. The value of comparative animal research: Krogh’s principle facilitates scientific discoveries. Policy Insights Behav. Brain Sci. 5, 118–125. https://doi.org/10.1177/2372732217745097.
Sun, P., Smith, A., Lei, K., Liu, Y., Wang, Z., 2014. Breaking bonds in male prairie vole: long-term effects on emotional and social behavior, physiology, and neurochemistry. Behav. Brain Res. 265, 22–31. https://doi.org/10.1016/j.bbr.2014.02.016.
Taborsky, M., Hofmann, H.A., Beery, A.K., Blumstein, D.T., Hayes, L.D., Lacey, E.A., Martins, E.P., Phelps, S.M., Solomon, N.G., Rubenstein, D.R., 2015. Taxon matters: promoting integrative studies of social behavior: NESCent working group on integrative models of vertebrate sociality: evolution, mechanisms, and emergent properties. Trends Neurosci. 38, 189–191. https://doi.org/10.1016/j.tins.2015.01.004.
Vogel, A.R., Patisaul, H.B., Arambula, S.E., Tiezzi, F., McGraw, L.A., 2018. Individual variation in social behaviours of male lab-reared prairie voles (Microtus ochrogaster) is non-heritable and weakly associated with V1aR density. Sci. Rep. 8, 1396. https://doi.org/10.1038/s41598-018-19737-9.
Walum, H., Waldman, I.D., Young, L.J., 2016. Statistical and methodological considerations for the interpretation of intranasal oxytocin studies. Biol. Psychiatry 79,

251–257. https://doi.org/10.1016/j.biopsych.2015.06.016.
Williams, J.R., Carter, C.S., Insel, T., 1992. Partner preference development in female prairie voles is facilitated by mating or the central infusion of oxytocin. Ann. N. Y. Acad. Sci. 652, 487–489.
Wolf, M., Weissing, F.J., 2010. An explanatory framework for adaptive personality differences. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365, 3959–3968. https://doi.org/10.1098/rstb.2010.0215.
Young, L.J., Wang, Z., 2004. The neurobiology of pair bonding. Nat. Neurosci. 7, 1048–1054 https://doi.org/15452576.
Zimmermann, T.D., Kaiser, S., Hennessy, M.B., Sachser, N., 2017. Adaptive shaping of the behavioural and neuroendocrine phenotype during adolescence. Proc. Biol. Sci. 284. https://doi.org/10.1098/rspb.2016.2784.