Research

Author for correspondence:
Conor Goold
e-mail: conor.goold@nmbu.no

Using network analysis to study behavioural phenotypes: an example using domestic dogs

Conor Goold[1], Judit Vas[1], Christine Olsen[2] and Ruth C. Newberry[1]

[1]Department of Animal and Aquacultural Sciences, and [2]Section of Public Health, Department of Landscape, Architecture and Spatial Planning, Norwegian University of Life Sciences, Ås, Norway

CG, 0000-0002-9198-0889; JV, 0000-0001-8195-8293; CO, 0000-0002-9630-1099; RCN, 0000-0002-5238-6959

Phenotypic integration describes the complex inter-relationships between organismal traits, traditionally focusing on morphology. Recently, research has sought to represent behavioural phenotypes as composed of quasi-independent latent traits. Concurrently, psychologists have opposed latent variable interpretations of human behaviour, proposing instead a network perspective envisaging interrelationships between behaviours as emerging from causal dependencies. Network analysis could also be applied to understand integrated behavioural phenotypes in animals. Here, we assimilate this cross-disciplinary progression of ideas by demonstrating the use of network analysis on survey data collected on behavioural and motivational characteristics of police patrol and detection dogs (*Canis lupus familiaris*). Networks of conditional independence relationships illustrated a number of functional connections between descriptors, which varied between dog types. The most central descriptors denoted desirable characteristics in both patrol and detection dog networks, with 'Playful' being widely correlated and possessing mediating relationships between descriptors. Bootstrap analyses revealed the stability of network results. We discuss the results in relation to previous research on dog personality, and benefits of using network analysis to study behavioural phenotypes. We conclude that a network perspective offers widespread opportunities for advancing the understanding of phenotypic integration in animal behaviour.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Understanding the biological organization of complex phenotypes is a mainstay of evolutionary biology [1–3]. Phenotypic integration describes the 'pattern of functional, developmental and/or genetic correlation...among different traits in a given organism' [4, p. 266]. Most commonly, phenotypic integration has been concerned with morphological traits (e.g. beak length and size in Darwin's finches [5]; sexual traits [6]). Recently, organization of the behavioural phenotype has also been cast in terms of phenotypic integration. Araya-Ajoy & Dingemanse [7], inspired by research in human psychology, propose that behavioural phenotypes consist of a collection of latent variables (behavioural *characters*) that play a causal role in producing correlated responses in patterns of behaviour, both within and between individuals. They discuss how this conceptualization could be applied to a number of topical themes in the study of animal behaviour including personality (consistent between-individual differences in behaviour) [8] and behavioural plasticity (between-individual differences in behavioural change) [9].

Phenotypic integration of biological traits is increasingly envisaged as interactions (whether physical or correlational) between modules played out on complex networks [10]. For example, Perez *et al.* [11] demonstrate how landmarks of the mammalian mandible can be represented as a network of nodes and correlational edges. Moreover, Wilkins *et al.* [6] advocate a 'phenotype network' approach for understanding correlations between sexual traits in the North American barn swallow (*Hirundo rustica erythrogaster*). The merit of a network perspective is that it naturally incorporates interactions within the components of, and between different, functional traits ('trait complexes') [12] and provides novel analytical insights (e.g. global and local network metrics). This is commensurate with studying organisms as 'developmentally, functionally and phenotypically integrated complex units' [13, p. 279], which numerous researchers argue is integral to improving our knowledge of the phenotype [3,14–16]. It follows that the organization of the behavioural phenotype can also benefit from being represented as an integrated network.

A network perspective has recently emerged in human psychology [17]. Psychological phenomena, such as personality dimensions (e.g. the Five Factor Model) [18], have traditionally been represented as latent variables and analysed with principal components analysis or varieties of factor analysis, respectively. However, this latent variable formulation has been contested (e.g. see commentaries in [19]), based on long-standing concerns that latent variable approaches can be conceptually, statistically and empirically ambiguous [20–24]. The central criticisms are that: (i) latent variables are often represented as fixed entities, failing to portray the dynamics of individual patterns of behaviours and the variability or lack of unidimensionality in psychological variables [23,25,26], (ii) observed behaviours are treated as passive and exchangeable indicators of the particular latent state [27,28], (iii) finding realizations of latent variables in biological organization (e.g. intelligence) [22] is challenging and, more conceptually, (iv) latent variables are unobservable by definition [28,29], promoting circularity in definitions of psychological phenomena ('verbal magic') [21] and leading to the fallacy of misplaced concreteness [30].

The network approach expounded by Cramer *et al.* [17,19] (see also [31–33]) presents personality and psychopathological phenomena as networks of autonomous and causally related cognitive, affective and behavioural components. These components possess conditional independence relationships, such that variation in one component can result in variation in another component conditional on all other measured components [34,35]. Given this assumption, components are more likely to have causal relationships when they possess a functional relationship, and when multiple components form close connections, functional clusters may emerge. For instance, networks of symptoms (e.g. 'loss of energy' and 'weight/appetite change') in long-term patients with major depression disorder were more densely connected (i.e. had greater network connectivity) than those of remitted patients [27]. Van der Mass *et al.* [22] further show how the positive manifold of general intelligence, defined as the observed correlations between cognitive skills related to intelligence, can be explained (and predicted) by direct mutualistic feedback relationships between those cognitive skills. While relationships between network components are influenced by underlying biological mechanisms (e.g. developmental pathways or genetic covariance [36,37]), the network approach aims to understand the behavioural phenotype as its own causal network of self-organizing components, rather than being comprised of passive indicators of 'common cause' latent variables [17].

In this paper, we synthesize the themes introduced above by exploring direct relationships among different behavioural and motivational characteristics in domestic dogs (*Canis lupus familiaris*). Dogs are useful in this respect because it is possible to gather information efficiently about multiple variables in a range of contexts using surveys directed at dog owners, who interact with their dog on a regular basis

and are thus qualified to answer questions about their dog's typical behaviour. Such surveys have been shown to be reproducible and corroborate behavioural observations (e.g. clinical reports) [38]. Until now, multivariate data on dog behaviour (e.g. from surveys or direct behavioural assessments) have usually been analysed using latent variable methods to reduce dimensionality and extract latent behavioural traits, or dimensions of dog personality, that explain the correlations between measured variables. This approach has resulted in the identification of a wide number of possible traits [38–40]. Alas, these putative traits often lack strong predictive validity [41–43], a practical concern when recruitment of suitable dogs for specific human uses depends upon reliable predictions. One possible reason is that predictive power is diminished when traits are overestimated as stable, dissociated constructs rather than components of dynamic integrated phenotypes. Further, after conducting a meta-analysis on behavioural consistency across numerous traits, Fratkin *et al.* [40] emphasized that personality dimensions in dogs may still be changeable in adults and sensitive to environmental and social perturbations. Thus, network analysis may be particularly beneficial when applied to the study of dog behaviour because it takes a bottom-up perspective to analysing direct functional relationships between behavioural components rather than decomposing the phenotype into latent variables.

Below, we apply network analysis to survey data collected from police dog handlers on desirable and undesirable behavioural and motivational descriptors of police patrol and detection dogs. Patrol dogs are selected and trained for diverse tasks, such as patrolling areas, controlling crowds, and tracking and detaining suspects, whereas detection dogs search for contraband, commonly drugs and money. Although studies have explored differences between working and non-working dogs on broad behavioural dimensions [44,45], few have compared different types of working dogs. A better understanding of how police dog behaviour is organized is of practical relevance to dog recruitment and training for specialized duties. Rather than focusing on deriving assumed latent traits as a basis for predicting future performance, we elucidate network structures that represent the behavioural phenotypes of patrol and detection dogs. Although our analyses are primarily exploratory, we expected to find some differences between patrol and detection dog networks due to differences in working duties. To our knowledge, this is the first application of network analysis to understand the behavioural phenotypes of animals.

2. Material and methods

2.1. Subjects

This study was carried out in collaboration with members of the Norwegian Police University College in Kongsvinger, Norway who oversee dog selection and training for the Norwegian police force. Professional police dog handlers ($N = 227$) across Norway were invited to complete an online survey in Norwegian investigating the personality and performance of police dogs. Handlers were requested to fill out one survey for each adult dog they had worked with as a handler. A total of 174 surveys were submitted. Three were removed for pertaining to more than one dog. The remaining responses concerned 171 dogs from 117 handlers (mean ± s.d. survey response per person: $1.46 ± 0.65$), including 117 patrol dogs (91 German shepherd dogs; 22 Belgian malinois; 1 rottweiler; 1 giant schnauzer; 1 Belgian tervueren; 1 unrecorded breed) and 54 detection dogs (17 labradors; 12 flat coated retrievers; 8 German shepherd dogs; 8 springer spaniels; 2 Belgian malinois; 2 Welsh springer spaniels; 1 German shepherd dog × Belgian shepherd dog; 1 labrador × German pointer; 1 cocker spaniel; 1 Nova Scotia duck-tolling retriever; 1 unrecorded breed). Breed differences were not explored due to the limiting sample sizes. Dogs were mostly entire ($n = 117$) and male ($n = 149$). Responses were received from 79 male and 17 female handlers (21 did not disclose their sex), aged between 28 and 57 (28–37 years: $n = 18$; 38–47 years: $n = 50$; 48–57 years: $n = 28$; undisclosed: $n = 21$). Handlers had between 1 and 30 years of experience as police dog handlers, and on average had 3.75 (s.d. = 4.64) previous dogs (including pet and working dogs).

2.2. Survey development

Survey questions and instructions were constructed in English, translated to Norwegian and back-translated to English to confirm intended meanings. The 'personality section' of the survey included 43 situational and adjective-based descriptors of police dog behavioural characteristics (electronic supplementary material, table S1). The list of descriptors was developed through (i) discussion with members of the Norwegian Police University College to include desirable and undesirable behavioural characteristics of relevance to police dog handlers, (ii) incorporation of characteristics evaluated in

standardized assessments of Norwegian police dog behaviour, and (iii) refinement following pilot tests for comprehensibility. Dog handlers rated how well they agreed with the descriptors as portrayals of their dog's typical behaviour, which ranged from 1 = 'Strongly disagree' to 5 = 'Strongly agree', where 3 = 'Neutral'. Participants could also choose 0 = 'Not relevant/I do not know'. All participants were familiar with the terminology used as descriptions of police dog behaviour.

2.3. Data preparation

All data handling and analysis was conducted using R v. 3.2.3 [46] (see the electronic supplementary files for the R script). The raw data for each descriptor contained a mean ± s.d. of 1.23 ± 1.38 (0.72 ± 0.81%) truly missing responses and 4.16 ± 7.64 (2.43 ± 4.47%) zero responses ('Not relevant/I do not know'). Zero responses were particularly prevalent for certain descriptors and five descriptors with at least 10% of zero responses were removed (electronic supplementary material, table S1). The remaining 38 descriptors were all of relevance to both dog types. One handler's responses for a patrol dog were removed as 18.5% were coded as zero (after removal of the five descriptors above), whereas the mean ± s.d. of the percentage of zero responses per dog was 1 ± 2.6%. The remaining zero responses were converted to missing values (as these were not comparable to other responses on the 1–5 scale).

2.4. Multiple imputation

Subsequently, a multiple imputation procedure (using *Amelia*) [47] was used to impute missing scores, rather than applying listwise deletion or mean substitution [48–50]. To ensure its robustness, we investigated any further biases in the data. We first considered whether the pattern of missingness in the data was dependent on dog type (i.e. patrol dogs and detection dogs), or on handlers for those submitting multiple surveys on different dogs (see §1 of the electronic supplementary material for statistical details). There were fewer missing values in the patrol than detection dog responses, and differences in the number of missing values varied between handlers. Thus, we included dog type and numerical handler ID as relevant conditioning variables for the multiple imputation procedure. Secondly, we investigated whether any descriptors had too many missing values to impute. The proportion of missing responses advisable for multiple imputation procedures is variable [51], although 5% or less is commonly considered unproblematic whereas greater than 5% [52] or 10% [53] have been reported to bias results. We chose to remove four descriptors with greater than 5% of missing responses (electronic supplementary material, table S1). Finally, we identified five pairs of variables that were theoretically similar and had high correlations relative to the data as a whole (polychoric correlations > |0.8|; see §2 of the electronic supplementary material for details), indicating redundancy. Therefore, we removed one descriptor from each pair (retaining the more specific one where evident, on the presumption that it was answered more reliably; electronic supplementary material, table S1). The resulting 29 descriptors had a mean of 1.2 ± 1.03% missing responses.

Subsequently, 15 multiply imputed datasets were generated. We averaged the datasets and rounded any non-integers to integers to produce a single dataset of ordinal responses. We examined the independence of responses to each question by the 44 handlers who filled out surveys for more than one dog. For eight descriptors, a high ratio of between- to within-handler variation indicated that repeated responses by the same handler lacked independence (see §3 of the electronic supplementary material for methods). Therefore, these eight descriptors were removed (electronic supplementary material, tables S1 and S2). Because the descriptor 'Good at catching a ball' had a particularly low variation ratio (defined as the proportion of responses not the mode) relative to other descriptors (mode = 5; variation ratio = 0.124), it was also removed. The final 20 descriptors used for the network analyses are presented in table 1, along with their modes, variation ratios and abbreviations used in the figures below.

2.5. Network analysis

2.5.1. Network construction

Networks were constructed and analysed using the *qgraph* package [54]. To construct networks that represented conditional independence relationships, we used Gaussian graphical models (GGM; see [55,56] for an overview). GGMs have been applied successfully to understand personality and psychopathology symptomatic networks (e.g. [27,57]). We used GGMs employing L_1 lasso penalties (i.e. least absolute shrinkage and selection operator), where the inverse covariance matrix (i.e. the matrix

Table 1. Descriptors used in the network analysis, including their abbreviations, modes and variation ratios (whole sample statistics shown outside parentheses; patrol and detection dog statistics, respectively, shown within parentheses). Descriptors are placed in alphabetical order (see electronic supplementary material, table S1, for ordering used in the survey).

abbreviation	descriptor name	mode	variation ratio
ACT[a]	active and nimble	5 (5; 5)	0.247 (0.284; 0.167)
ADP[a]	adapts to new situations quickly	5 (5; 5)	0.406 (0.414; 0.389)
CUR[a]	curious	5 (5; 5)	0.229 (0.224; 0.241)
DA[b]	aggressive towards other dogs ('Dog aggressive')[c]	4 (4; 1)	0.735 (0.707; 0.685)
FDA[b]	guards food ('Food aggressive')	1 (1; 1)	0.418 (0.414; 0.426)
FIT[a]	physically fit	5 (5; 5)	0.247 (0.293; 0.148)
FL[a]	fearless	5 (5; 5)	0.482 (0.422; 0.611)
FoH[b]	fear of heights	1 (1; 1)	0.461 (0.457; 0.500)
FSH[a]	able to stay focused during searches	5 (5; 5)	0.324 (0.371; 0.222)
GUS[b]	gives up searches quickly	1 (1; 1)	0.553 (0.586; 0.481)
GWL[b]	strong tendency to growl at strangers	1 (1; 1)	0.476 (0.483; 0.463)
PLA[a]	playful	5 (5; 5)	0.200 (0.207; 0.185)
PS[a]	solves problems on own ('Problem solving')	5 (5; 5)	0.353 (0.345; 0.370)
PSV[a]	persevering	5 (5; 5)	0.265 (0.267; 0.259)
REC[a]	comes when called ('Recalls')	5 (5; 5)	0.424 (0.466; 0.333)
SLP[a]	good at walking on slippery surfaces	5 (5; 5)	0.265 (0.302; 0.185)
SOC[a]	socially attached to you	5 (5; 5)	0.200 (0.224; 0.148)
STR[b]	nervous and tense when startled	1 (1; 1)	0.606 (0.552; 0.722)
TOY[a]	willing to give you a toy	5 (5; 5)	0.424 (0.457; 0.352)
WIL[a]	desires to make you happy ('Willing to please')	5 (5; 5)	0.353 (0.397; 0.259)

[a]Desirable descriptor.
[b]Undesirable descriptor.
[c]Brief descriptions used to form some abbreviations are shown in parentheses.

of partial correlations) was subject to regularization through penalized maximum-likelihood estimation. This resulted in a sparse graph with credibly non-zero partial correlations, with partial correlations near zero being shrunk to zero. Regularization was controlled by a parameter $\lambda \in [0, 1]$ [58]. The optimal value of λ was chosen according to the graph with the lowest Extended Bayesian Information Criterion (EBIC) following Foygel & Drton [59] (see also [56]) and implemented in the 'EBICglasso' function in the *qgraph* package. The EBIC criterion was in turn tuned by a parameter $\gamma \in [0, 1]$ that performs best for positive values of γ [59]. We explored the networks over the entire range of γ (by 0.05 increments) and chose the most conservative value of $\gamma = 0.65$, where values above this resulted in empty graphs for the detection dog network. This method optimized specificity in network estimation (i.e. prioritized the elimination of truly non-existent edges) [60]. Because our data were ordinal, we conducted GGM construction and selection using the matrix of polychoric correlations (see the R script file in the electronic supplementary material), which provided the correlations between ordinal variables assumed to have latent continuous distributions.

2.5.2. Centrality analysis

We explored and compared the structures of patrol and detection dog networks using node-level centrality metrics because nodes that are more central are more important for influencing network structure than peripheral nodes. We chose the metrics *betweenness* and *strength* centrality (defined formally for weighted networks in electronic supplementary material, table S3), where node betweenness represents how many shortest paths (i.e. with minimum distance between two nodes) run through a given node and node strength indicates how strongly each node is connected to other nodes [61,62].

Nodes with high betweenness values acted as mediators between indirectly connected nodes, and nodes with high strength values had stronger correlations with other descriptors.

2.5.3. Network comparison and stability

To compare descriptor centrality between patrol and detection dog networks, 2000 non-parametric bootstrap samples for each network were computed (R package: *bootnet*) [63]. Each bootstrap constructed a network of randomly sampled dogs, with replacement. From these bootstrap samples, we calculated the mean centrality of each descriptor (the overall mean of descriptors' mean betweenness and strength values) and these means were compared with Cliff's delta (δ; R package: '*effsize*') [64], a non-parametric effect size ranging between -1 and $+1$ (see [27]). To explore network stability, we computed bootstrap samples of the networks 2000 times from networks of 3 to 19 nodes (node-wise bootstrapping), and 2000 times from 25% to 95% (at approximately 8% increments) of the original sample sizes (subject-wise bootstrapping; see [65]). This allowed investigating the rank-order consistency of descriptor centrality values and the correlation between centrality values in the bootstrapped networks with the original networks. Confidence intervals on bootstrapped parameters are not reported due to known biases in their estimation [65].

3. Results

3.1. Descriptive network structures

The patrol dog network (figure 1*a*; see association matrix in the electronic supplementary files) had 55 edges (28.95% of possible edges). 'Curious' had strong positive correlations with 'Playful', 'Problem solving' and 'Fearless'. Additional salient positive correlations appeared between: 'Socially attached to you', 'Recalls' and 'Willing to please'; 'Strong tendency to growl at strangers' and 'Food aggressive'; 'Good at walking on slippery surfaces' and 'Physically fit'; 'Active and nimble' and 'Physically fit'; and 'Fearless' and 'Adapts to new situations quickly'. Negative correlations were evident between: 'Fearless' and 'Nervous and tense when startled'; 'Fear of heights' and 'Good at walking on slippery surfaces'; 'Dog aggressive' and 'Willing to please'; 'Food aggressive' and 'Playful'; and 'Gives up searches quickly' with 'Able to stay focused during searches' and 'Willing to please'.

The detection dog network (figure 1*b*; see association matrix in the electronic supplementary files) had 70 edges (36.84% of possible edges). 'Playful' shared salient positive correlations with 'Curious', 'Persevering', 'Adapts to new situations quickly' and 'Problem solving', and was most negatively correlated with 'Gives up searches quickly'. 'Able to stay focused during searches' shared salient positive correlations with 'Socially attached to you', 'Willing to please', 'Adapts to new situations quickly', 'Willing to give you a toy' and 'Active and nimble'. Strong positive correlations were also evident between: 'Fearless' and 'Curious'; 'Fearless' and 'Problem solving'; 'Good at walking on slippery surfaces' with 'Problem solving' and 'Adapts to new situations quickly'; 'Persevering' and 'Physically fit'; 'Willing to please' and 'Socially attached to you'; 'Gives up searches quickly' and 'Food aggressive'; and 'Strong tendency to growl at strangers' with 'Food aggressive' and 'Dog aggressive'. A strong negative correlation was present between 'Curious' and 'Nervous and tense when startled'.

3.2. Network centrality

Most of the desirable descriptors (table 1) had higher observed centrality values compared with undesirable descriptors (figure 2; see electronic supplementary material, table S4, for raw values). In the patrol dog network, 'Playful' had the highest betweenness centrality and 'Curious' the highest strength centrality, whereas 'Playful' had both the highest betweenness and highest strength centrality values in the detection dog network. Across both networks, 'Active and nimble', 'Curious', 'Physically fit', 'Recalls' and 'Good at walking on slippery surfaces' had higher betweenness and strength values in the patrol dog compared to detection dog network (figure 2). In the detection dog network, 'Dog aggressive', 'Able to stay focused during searches', 'Gives up searches quickly', 'Strong tendency to growl at strangers', 'Problem solving', 'Persevering', 'Nervous and tense when startled' and 'Willing to give you a toy' had higher betweenness and strength values than in the patrol dog network.

Across non-parametric bootstrap samples, only certain descriptor centrality differences had strong effect sizes (figure 3; raw values provided in electronic supplementary material, table S5). Mean centrality differences in 'Curious' ($\delta = 0.452$), 'Good at walking on slippery surfaces' ($\delta = 0.290$) and 'Active and
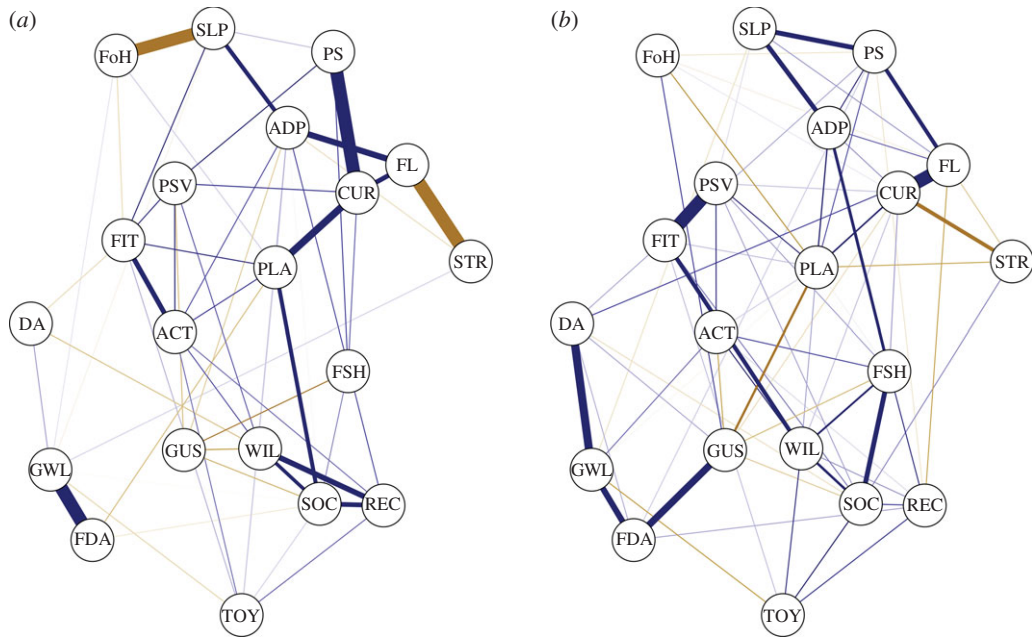
Figure 1. Gaussian graphical models of patrol (*a*) and detection (*b*) dogs. Blue edges show positive correlations, gold edges negative correlations; stronger correlations have thicker edges. See table 1 for descriptor abbreviations.
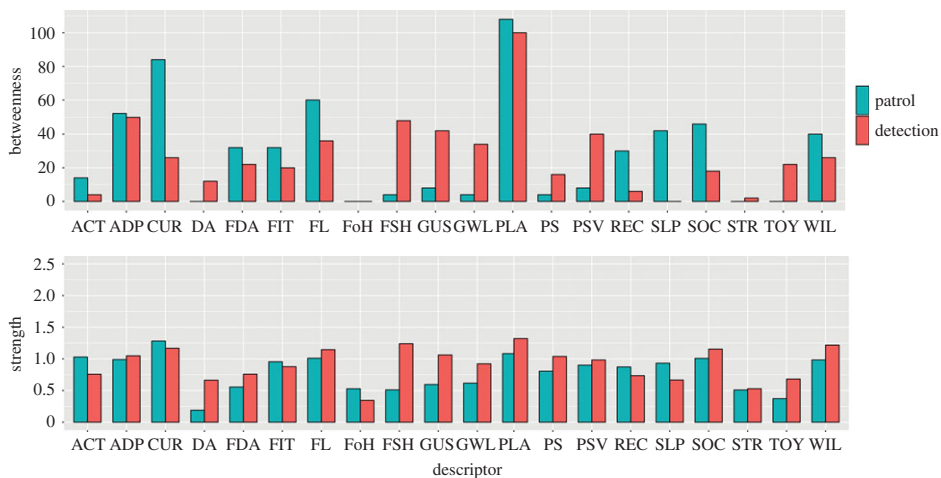

Figure 2. Observed betweenness and strength centrality values (bar heights) for patrol and detection dog networks. See table 1 for descriptor abbreviations and electronic supplementary material, table S4, for raw values.

nimble' ($\delta = 0.282$) had the largest effect sizes in favour of the patrol dog network. 'Able to stay focused during searches' ($\delta = -0.614$), 'Dog aggressive' ($\delta = -0.609$), 'Gives up searches quickly' ($\delta = -0.582$), 'Willing to give you a toy' ($\delta = -0.465$), 'Strong tendency to growl at strangers' ($\delta = -0.310$) and 'Food aggressive' ($\delta = -0.302$) had the largest effect sizes in favour of the detection dog network.

3.3. Network stability

The standard deviation of the number of edges in the patrol dog network across non-parametric bootstrap samples was 12.82, and 29.75 for the detection dog network. Node-wise bootstrapping demonstrated reasonable stability of the original network structures: centrality values from the bootstrapped networks were positively correlated with centrality values in the original networks (figure 4*a,b*), even for networks of only three nodes, although the patrol dog network was more stable than the detection dog network (see electronic supplementary material, figure S1, for the rank-order stability of individual descriptors). Network structure was more sensitive under subject-wise bootstrapping. For the patrol dog network, sampled networks of around 60 dogs or less (approximately
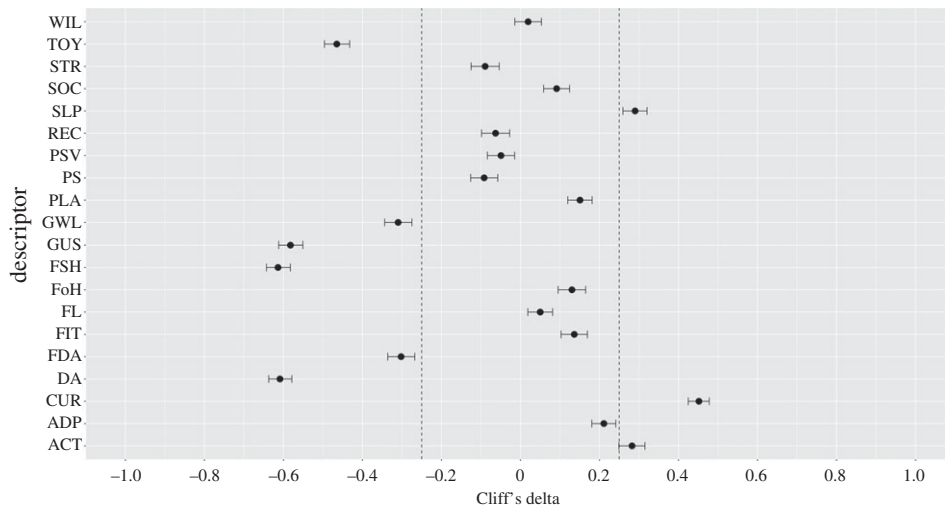
Figure 3. Cliff's delta effect sizes (and 95% CIs) for differences between patrol and detection dog centrality values (average of betweenness and strength) calculated from non-parametric bootstrap samples. Positive values indicate a larger mean for patrol dogs and negative values a larger mean for detection dogs. Values lying within the dashed lines at ±0.25 indicate negligible effect sizes. See table 1 for definitions of descriptor abbreviations and electronic supplementary material, table S5, for raw values.
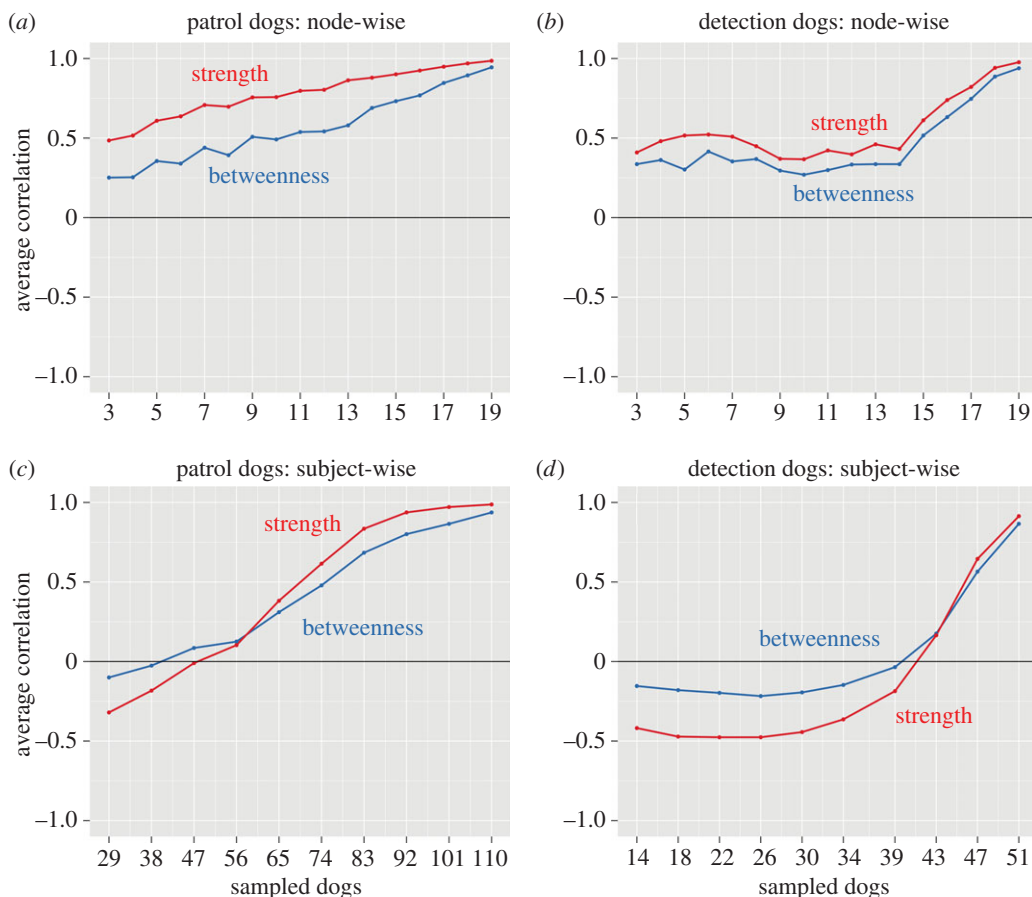

Figure 4. Stability of betweenness and strength centrality values in node-wise (*a*,*b*) and subject-wise (*c*,*d*) bootstrapping. The centrality values in each bootstrapped network were correlated with values in the original networks. Panels (*a*–*d*) show the average correlation across descriptors for each node-wise and subject-wise bootstrap sampling level, respectively.

50% of the original sample size) showed little correlation with the original network values (figure 4*c*). For the detection dog network, networks less than around 40 dogs (approximately 70% of the original sample size) had low to negative correlations with the original network (figure 4*d*; see electronic supplementary material, figure S2, for the rank-order stability of individual descriptors).

4. Discussion

There has been much interest in biology about phenotypic integration of morphological traits, particularly their genetic and developmental bases [2,12,36]. Recent work has extended these notions to conceiving of the behavioural phenotype as composed of quasi-independent latent behavioural traits that form an integrated unit [7]. Network analysis offers benefits for understanding phenotypic integration [6,10,11] and has emerged in human psychology as an efficacious theoretical and analytical framework to understand human behaviour as a causally connected unit [17,32,33,66]. In this regard, it assimilates the study of behavioural phenotypes with research on a number of other complex systems showing how structure can emerge from self-organizing interactions between component parts (e.g. social groups [67], genetic/physiological networks [15,68] and evolutionary processes [14]). In this paper, we have assimilated this cross-disciplinary progression of ideas by using network analysis to understand relationships among behavioural and motivational characteristics in police patrol and detection dogs.

Our analyses revealed numerous direct correlations between functionally related descriptors in both patrol and detection dog networks (figure 1). For instance, behaviours related to aggression ('Dog aggressive', 'Strong tendency to growl at strangers' and 'Food aggressive') were positively correlated, especially in the detection dog network, as were descriptors indicating levels of sociability and/or trainability ('Socially attached to you', 'Willing to please', 'Recalls', 'Willing to give you a toy'). Moreover, various positive correlations involving 'Playful', 'Curious', 'Fearless' and 'Socially attached to you' are partly consistent with Svartberg & Forkman's [39] interrelated factors 'playfulness', 'curiosity/fearlessness' and 'sociability'. Together with 'chase proneness', these factors formed a super-trait referred to as 'boldness' that was related to working dog performance [44]. Svartberg & Forkman's [39] results were based on first- and second-order exploratory factor analyses of pairwise correlations, positing boldness as a higher-order latent variable causing covariation between boldness-related behaviours. Our findings extend these results by disentangling potential causal, mutually reinforcing relationships between behaviours. For instance, despite 'Curious' and 'Socially attached to you' sharing positive pairwise correlations (0.41 and 0.40 for patrol and detection dogs, respectively; see the R script file in the electronic supplementary files for their calculation), they were not directly related in either network of conditional independence relationships, suggesting that their pairwise correlation was due to common mediating variables. For the assessment of dog behaviour, low predictive values of behavioural and personality tests [41–43] may arise from over-estimating the homogeneity of behavioural traits from pairwise correlations when, in fact, trait compositions could be dynamic through time and across contexts. Distinguishing causal relationships from pairwise correlations could refine behavioural assessments through identifying behavioural variables that cause widespread changes in behavioural phenotypes.

Network analysis provides a number of unique metrics to understand patterns of relationships in multivariate data, such as the estimation of network centrality, indicating the relative importance individual components have across network topologies. In particular, the descriptor 'Playful' held a central position across networks (figure 2) both in its number of direct behavioural correlations (i.e. strength centrality) with other descriptors, but also in its mediating role between other relationships across the network (i.e. betweenness centrality). Playfulness is postulated to have a positive influence on the success or trainability of working dogs [44,45], comprising part of Svartberg & Forkman's boldness dimension [39], and has been assayed in working dog assessments by rating a dog's attentiveness and intensity when engaging in tug-type games with a toy [39,69]. Play also represents a heterogeneous category of behaviour that includes object-related, locomotory and social components [70], and constitutes an important method of reinforcement in training protocols. Thus, from a network viewpoint, playful behaviour may have important causal connections to a wide range of behaviours. In the patrol dog network (figure 1*a*), 'Playful' connected additional central descriptors (figure 2), such as between 'Socially attached to you' and 'Curious' or 'Fearless', respectively. In the detection dog network (figure 1*b*), 'Playful' had a strong negative relationship with 'Gives up searches quickly', the latter being particularly undesirable for detection dogs. As Bradshaw *et al.* [71] review, play in dogs correlates with

a number of variables indicating positive well-being, including obedience indicative of close social bonds with owners. Therefore, the centrality of the 'Playful' descriptor in our network analyses holds an interesting organizational position in the behavioural phenotype of police dogs. This organizational role could be further examined in a network framework by quantifying how different forms of playful behaviour relate to other behaviours through time, or between breeds or types of dogs differing systematically in playfulness (e.g. working and pet dogs) [45].

Other descriptors differed in relative centralities between patrol and detection dog networks. In particular, 'Curious' had larger betweenness and strength centrality values in the patrol dog compared with the detection dog network (figure 2), which was also borne out in the non-parametric bootstrap analyses (figure 3). Moreover, 'Good at walking on slippery surfaces' and 'Active and nimble' had larger mean centrality values across bootstrap samples in the patrol dog compared to detection dog networks. By contrast, task-specific descriptors such as 'Able to stay focused during searches' and 'Gives up searches quickly' (which was negatively correlated with desirable descriptors such as 'Playful'; figure 1b) were more central in the detection dog network than the patrol dog network, as was 'Willing to give you a toy', which may reflect the tendency for detection dogs to be trained to hold objects gently in their mouths and relinquish objects easily. Descriptors related to aggression were more frequently and strongly negatively correlated with desirable descriptors and positively correlated with undesirable descriptors compared to the patrol dog network. At the same time, weak positive correlations appeared between desirable and undesirable descriptors, such as between 'Dog aggressive' and 'Fearless', 'Recalls' and 'Food aggressive' or 'Socially attached to you' and 'Nervous and tense when startled', which were not present in the patrol dog network. These findings may indicate less stringent behavioural selection criteria for detection dogs compared with patrol dogs, conditional on detection dogs being good at searching. Consequently, successful detection dogs may, on average, be more likely to show correlations between undesirable and desirable behaviours than successful patrol dogs, as long as they show good performance during search tasks.

Nonetheless, our results also demonstrate uncertainty in network structures. Across non-parametric bootstrap samples, the detection dog network had a large standard deviation of estimated edges, probably due to the smaller detection dog sample size. Both networks were relatively stable in response to node-wise bootstrapping (figure 4a,b; electronic supplementary material, figure S1), but their stability was more sensitive in the subject-wise bootstrapping (figure 4c,d; electronic supplementary material, figure S2), and so may differ at larger samples sizes. As highlighted by Epskamp *et al.* [65], it is important that network analyses are checked for stability, and that uncertainty in parameter estimates is reported to gauge the predictive accuracy of network models. This is particularly important in dog personality studies employing exploratory analyses of multivariate datasets.

4.1. Limitations and future directions

There are potential limitations to the example presented here. First, the survey descriptors analysed include general behavioural and motivational characteristics (e.g. 'Fearless') that integrate a number of possible behaviour patterns. Thus, this lexical rating approach differs from the quantitative behavioural assays common in, for instance, behavioural ecology research. Nonetheless, rating approaches may be comparable or more beneficial than direct behavioural observations (e.g. in dogs: [72–74]), particularly in cases where raters are highly familiar with the individual animals (see also the discussion in [75]). However, while the survey here was completed by knowledgeable participants and explicated the network approach, no checks of reliability or validity were conducted. Instead, we employed a rigorous data cleaning process, removing 23 of the original 43 descriptors and employing multiple imputation of missing data. Checks of validity have not been fully developed under a network approach [76]. Validity theory attempts to answer whether an indicator measures what it is intended to measure (e.g. whether 'Strong tendency to growl at strangers' measures aggression) and is motivated by a 'reflective' latent variable conceptualization of scientific constructs [77]. However, the network approach does not view indicators, such as the behavioural descriptors analysed here, as measures of latent traits. Instead, the relationship between constructs and indicators is mereological [32,78], such that 'the observables [i.e. indicators] do not measure the construct, but are part of it' [32, p. 5]. Although validity in a network framework is currently in its infancy, exploring how the network approach can refine the predictive validities of current personality tests in dogs would be a fruitful avenue of research.

Secondly, the network analysis reported here was based on one survey per dog. Although handlers responded regarding dogs' typical behaviours, there are advantages to gathering repeated measurements

to directly estimate variation between and within individuals. Network analysis can also be applied to this end (e.g. see [79] for a multilevel time-series network model).

Finally, there is a natural relationship between integration of behavioural phenotypes and the study of animal personality and, relatedly, behavioural syndromes. Animal personality is defined by repeatable between-individual differences in behaviour reflecting personality traits [8,80,81]. As in studies of human personality, investigations into animal personality have used latent variable approaches (e.g. exploratory factor analysis [38] or structural equation modelling [72,82] in dogs) to extract relevant traits. However, the conceptualization of personality traits has been a point of confusion in animal behaviour [40,75,83] and psychologists have related a similar ambiguity in human research directly to latent variable interpretations [20,21,23,24]. Combining the network perspective established in human psychology and the more general biological concept of phenotypic integration may improve the clarity of personality definitions. That is, the behavioural phenotype becomes organized through causal connections between its components. By virtue of this organization, consistent behavioural expression is maintained through principles of network stability [84]. In this way, traits are emergent properties of clustering between functionally related behaviours [17,32]. In psychology, dynamic systems approaches to behaviour have a long history [85], supporting the process of behavioural integration as a self-organizing system [86,87]. In evolutionary biology, morphological trait complexes have been elucidated as emergent properties ('evolutionarily stable configurations') [12] and, more recently, Watson *et al.* [14] use principles of supervised and unsupervised learning to outline how phenotypic correlations can become causal connections over evolutionary timescales, highlighting the role of self-organization in the evolution of phenotypic integration.

5. Conclusion

Network analysis provides a novel approach to conceptualizing and analysing the behavioural phenotype, in both humans and animals. Following recent work across the biological study of phenotypic integration and human psychology, network analysis can be used to conceive of the behavioural repertoire of individuals as a connected system of causally dependent components. We have demonstrated how network analysis can be applied using police patrol and detection dogs as an example, elucidating commonalities and differences between networks in the interrelationships between behavioural and motivational descriptors. Moreover, we have demonstrated how analyses can be carried out to ascertain the stability of the results. We conclude that a network approach offers widespread opportunities for advancing the understanding of phenotypic integration in animal behaviour.

References

1. Wagner GP. 1996 Homologues, natural kinds and the evolution of modularity. *Am. Zool.* **36**, 36–43. (doi:10.1093/icb/36.1.36)

2. Wagner GP, Laubichler MD. 2000 Character identification in evolutionary biology: the role of the organism. *Theory Biosci.* **119**, 20–40. (doi:10.1007/s12064-000-0003-7)

3. Murren CJ. 2012 The integrated phenotype. *Integr. Comp. Biol.* **52**, 64–76. (doi:10.1093/icb/ics043)

4. Pigliucci M. 2003 Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecol. Lett.* **6**, 265–272. (doi:10.1046/j.1461-0248.2003.00428.x)

5. Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. 2006 The calmodulin pathway and

evolution of elongated beak morphology in Darwin's finches. *Nature* **442**, 563–567. (doi:10.1038/nature04843)

6. Wilkins MR, Shizuka D, Joseph MB, Hubbard JK, Safran RJ. 2015 Multimodal signalling in the North American barn swallow: a phenotype network approach. *Proc. R. Soc. B* **282**, 20151574. (doi:10.1098/rspb.2015.1574)

7. Araya-Ajoy YG, Dingemanse NJ. 2014 Characterizing behavioural 'characters': an evolutionary framework. *Proc. R. Soc. B* **281**, 20132645. (doi:10.1098/rspb.2013.2645)

8. Sih A, Bell AM, Johnson JC, Ziemba RE. 2004 Behavioral syndromes: an integrative overview. *Q. Rev. Biol.* **79**, 241–277. (doi:10.1086/422893)

9. Dingemanse NJ, Wolf M. 2013 Between-individual differences in behavioural plasticity within populations: causes and consequences. *Anim. Behav.* **85**, 1031–1039. (doi:10.1016/j.anbehav.2012.12.032)

10. Wagner GP, Pavlicev M, Cheverud JM. 2007 The road to modularity. *Nat. Rev.* **8**, 921–931. (doi:10.1038/nrg2267)

11. Perez SI, de Aguiar MAM, Guimarães Jr PR, dos Reis SF. 2009 Searching for modular structure in complex phenotypes: inferences from network analysis. *Evol. Biol.* **36**, 416–422. (doi:10.1007/s11692-009-9074-7)

12. Wagner GP, Schwenk K. 2001 Function and the evolution of phenotypic stability: connecting pattern with process. *Am. Zool.* **41**, 552–563.

13. Forsman A. 2015 Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity* **115**, 276–284. (doi:10.1038/hdy.2014.92)

14. Watson RA *et al.* 2015 Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evol. Biol.* 1–29. (doi:10.1007/s11692-015-9358-z)

15. Davila-Velderrain J, Martinez-Garcia JC, Alvarez-Buylla ER. 2015 Modelling the epigenetic attractors landscape: toward a post-genomic mechanistic understanding of development. *Front. Genet.* **6**, 1–14. (doi:10.3389/fgene.2015.00160)

16. Ellers J, Liefting M. 2015 Extending the integrated phenotype: covariance and correlation in plasticity of behavioural traits. *Curr. Opin. Insect Sci.* **9**, 31–35. (doi:10.1016/j.cois.2015.05.013)

17. Cramer AOJ, van der Sluis S, Noordhof A, Wichers M, Geschwind N, Aggen SH, Kendler KS, Borsboom D. 2012 Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people. *Euro. J. Pers.* **26**, 414–431. (doi:10.1002/per.1866)

18. McCrae RR, Costa PT. 1995 Trait explanations in personality psychology. *Euro. J. Pers.* **9**, 231–252. (doi:10.1002/per.2410090402)

19. Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. 2010 Comorbidity: a network perspective. *Behav. Brain Sci.* **33**, 137–150. (doi:10.1017/S0140525X09991567)

20. Borsboom D. 2005 *Measuring the mind*. Cambridge, UK: Cambridge University Press.

21. Boag S. 2011 Explanation in personality psychology: 'verbal magic' and the five-factor model. *Phil. Psychol.* **24**, 223–243. (doi:10.1080/09515089.2010.548319)

22. van der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers EJ. 2006 A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861. (doi:10.1037/0033-295X.113.4.842)

23. Cervone D. 2005 Personality architecture: within-person structures and processes. *Annu. Rev. Psychol.* **56**, 423–452. (doi:10.1146/annurev.psych.56.091103.070133)

24. Lamiell JT. 2013 Statisticism in personality psychologists' use of trait constructs: what is it? How was it contracted? Is there a cure? *New Ideas Psychol.* **31**, 65–71. (doi:10.1016/j.newideapsych.2011.02.009)

25. Adolf J, Schuurman NK, Borkenau P, Borsboom D, Dolan CV. 2014 Measurement invariance within and between individuals: a distinct problem in testing the equivalence of intra- and inter-individual model structures. *Front. Psychol.* **5**, 9–22. (doi:10.3389/fpsyg.2014.00883)

26. Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D. 2016 Measuring depression over time . . . or not? Lack of unidimensionality and longitudinal measurement invariance in four common ratings scales of depression. *Psychol. Assess.* 1–15. (doi:10.1037/pas0000275)

27. Van Borkulo C, Boschloo L, Borsboom D, Phennix BWJH, Waldorp LJ, Schoevers RA. 2015 Association of symptom network structure with the course of longitudinal depression. *JAMA Psych.* **72**, 1219–1226. (doi:10.1001/jamapsychiatry.2015.2079)

28. Bollen KA. 2002 Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.* **53**, 605–634. (doi:10.1146/annurev.psych.53.100901.135239)

29. Borsboom D. 2008 Psychometric perspectives on diagnostic systems. *J. Clin. Psychol.* **64**, 1089–1108. (doi:10.1002/jclp.20503)

30. Flynn M. 1997 The concept of intelligence in psychology as a fallacy of misplaced concreteness. *Interchange* **28**, 231–244. (doi:10.1023/A:1007317410814)

31. Borsboom D, Cramer AOJ. 2013 Network analysis: an integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.* **9**, 91–121. (doi:10.1146/annurev-clinpsy-050212-185608)

32. Schmittmann VD, Cramer AOJ, Waldorp LJ, Epskamp S, Kievet RA, Borsboom D. 2013 Deconstructing the construct: a network perspective on psychological phenomenon. *New Ideas Psychol.* **31**, 43–53. (doi:10.1016/j.newideapsych.2011.02.007)

33. Goekoop R, Goekoop JG, Scholte HS. 2012 The network structure of human personality according to the NEO-PI-R: matching network community structure to factor structure. *PLoS ONE* **7**, e51558 (doi:10.1371/journal.pone.0051558)

34. Costantini G, Epskamp S, Borsboom D, Perugini M, Mõttus R, Waldorp LJ, Cramer AOJ. 2015 State of the aRt personality research: a tutorial on network analysis of personality data. *J. Res. Pers.* **54**, 13–29. (doi:10.1016/j.jrp.2014.07.003)

35. Pearl J. 2009 Causal inference in statistics: an overview. *Stat. Surv.* **3**, 96–146. (doi:10.1214/09-SS057)

36. Gonzalez PN, Hallgrimsson B, Oyheart EE. 2011 Developmental plasticity in covariance structure of the skull: effects of prenatal stress. *J. Anat.* **218**, 243–257. (doi:10.1111/j.1469-7580.2010.01326.x)

37. Santos JC, Cannatella DC. 2011 Phenotypic integration emerges from aposematism and scale in poison frogs. *Proc. Natl Acad. Sci. USA* **108**, 6175–6180. (doi:10.1073/pnas.1010952108)

38. Hsu Y, Serpell JA. 2003 Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *J. Am. Vet. Med. Assoc.* **9**, 1293–1300. (doi:10.2460/javma.2003.223.1293)

39. Svartberg K, Forkman B. 2002 Personality traits in the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* **79**, 133–155. (doi:10.1016/S0168-1591(02)00121-1)

40. Fratkin JL, Sinn DL, Patall EA, Gosling SD. 2013 Personality consistency in dogs: a meta-analysis. *PLoS ONE* **8**, e54907. (doi:10.1371/journal.pone.0054907)

41. Mornement KM, Toukhsati S, Coleman GJ, Bennett P. 2015 Evaluation of the predictive validity of the Behavioural Assessment for Re-homing K9's (B.A.R.K.) protocol and owner satisfaction with adopted dogs. *Appl. Anim. Behav. Sci.* **167**, 35–42. (doi:10.1016/j.applanim.2015.03.013)

42. Reimer S, Müller C, Virányi Z, Huber L, Range F. 2014 The predictive value of early behavioural assessments in pet dogs—a longitudinal study from neonates to adults. *PLoS ONE* **9**, e101237. (doi:10.1371/journal.pone.0101237)

43. Sinn DL, Gosling SD, Hillard S. 2010 Personality and performance in military working dogs: reliability and predictive validity of behavioral tests. *Appl. Anim. Behav. Sci.* **127**, 51–65. (doi:10.1016/j.applanim.2010.08.007)

44. Svartberg K. 2002 Shyness–boldness predicts performance in working dogs. *Appl. Anim. Behav. Sci.* **79**, 157–174. (doi:10.1016/S0168-1591(02)00120-X)

45. Asp HE, Fikse WF, Nilsson K, Strandberg E. 2015 Breed differences in everyday behaviour of dogs. *Appl. Anim. Behav. Sci.* **169**, 69–77. (doi:10.1016/j.applanim.2015.04.010)

46. R Core Team. 2015 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.

47. Honaker J, King G, Blackwell M. 2011 Amelia II: a program for missing data. *J. Stat. Softw.* **45**, 1–47. (doi:10.18637/jss.v045.i07)

48. Schlomer GL, Bauman S, Card NA. 2010 Best practices for missing data management in counseling psychology. *J. Couns. Psychol.* **57**, 1–10. (doi:10.1037/a0018082)

49. Osborne JW. 2013 *Best practices in data cleaning*. London, UK: Sage Publications, Inc.

50. Rubin DB. 1976 Inference and missing data. *Biometrika* **63**, 581–592. (doi:10.1093/biomet/63.3.581)

51. Dong Y, Peng CJ. 2013 Principled missing data methods for researchers. *SpringerPlus* **2**, 222. (doi:10.1186/2193-1801-2-222)

52. Schafer JL. 1999 Multiple imputation: a primer. *Stat. Methods Med. Res.* **8**, 3–15. (doi:10.1191/096228099671525676)

53. Bennett DA. 2001 How can I deal with missing data in my study? *Aust. N. Z. J. Public Health* **25**, 464–469. (doi:10.1111/j.1467-842X.2001.tb00294.x)

54. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. 2012 qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* **48**, 1–18. (doi:10.18637/jss.v048.i04)

55. Kolaczyk ER, Csárdi G. 2014 *Statistical analysis of network data with R*. New York, NY: Springer.

56. Yuan M, Lin Y. 2007 Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19–35. (doi:10.1093/biomet/asm018)

57. Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, Borsboom D. 2016 What are 'good' depression symptoms? Comparing the centrality of DSM and

non-DSM symptoms of depression in a network analysis. *J. Affec. Disord.* **189**, 314–320. (doi:10.1016/j.jad.2015.09.005)

58. Zhao T, Liu H, Roeder K, Lafferty J, Wasserman L. 2012 The huge package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.* **13**, 1059–1069.

59. Foygel R, Drton M. 2010 Extended Bayesian Information Criterion for Gaussian graphical models. *Adv. Neural Inf. Process Syst.* **23**, 2020–2028.

60. Epskamp S. 2016 Regularized Gaussian psychological networks: brief report on the performance of extended BIC model selection. See https://arxiv.org/pdf/1606.05771.

61. Brandes U. 2001 A faster algorithm for betweenness centrality. *J. Math. Sociol.* **25**, 163–177. (doi:10.1080/0022250X.2001.9990249)

62. Newman MEJ. 2001 Scientific collaboration II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 1–7. (doi:10.1103/PhysRevE.64.016132)

63. Epskamp S. 2015 bootnet: bootstrap methods for various network estimation routines. R package version 0.3. See https://github.com/SachaEpskamp/bootnet.

64. Torchiano M. 2015 effsize: efficient effect size computation. R package version 0.5.4. See http://www.cran.r-project.org/package=effsize/.

65. Epskamp S, Borsboom D, Fried EI. 2016 Estimating psychological networks and their stability: a tutorial paper. See https://arxiv.org/abs/1604.08462.

66. Robinaugh DJ, LeBlanc NJ, Vuletich HA, McNally RJ. 2014 Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *J. Abnorm. Psychol.* **123**, 510–522. (doi:10.1037/abn0000002)

67. Sumpter DJT. 2010 *Collective animal behavior.* Princeton, NJ: Princeton University Press.

68. Cohen AA, Martin LB, Wingfield JC, McWilliams SR, Dunne JA. 2012 Physiological regulatory networks: ecological roles and evolutionary constraints. *Trends Ecol. Evol.* **27**, 428–435. (doi:10.1016/j.tree.2012.04.008)

69. Foyer P, Wilsson E, Wright D, Jensen P. 2013 Early experiences modulate stress coping in a population of German shepherd dogs. *Appl. Anim. Behav. Sci.* **146**, 79–87. (doi:10.1016/j.applanim.2013.03.013)

70. Špinka M, Newberry RC, Bekoff M. 2001 Mammalian play: training for the unexpected. *Q. Rev. Biol.* **76**, 141–168. (doi:10.1086/393866)

71. Bradshaw JWS, Pullen AJ, Rooney NJ. 2015 Why do adult dogs 'play'? *Behav. Process.* **110**, 82–87. (doi:10.1016/j.beproc.2014.09.023)

72. Barnard S, Marshall-Pescini S, Passalacqua C, Beghelli A, Capra A, Normando S, Pelosi A, Valsecchi P. 2016 Does subjective rating reflect behavioural coding? Personality in 2 month-old dog puppies: an open-field test and adjective-based questionnaire. *PLoS ONE* **11**, e0149831. (doi:10.1371/journal.pone.0149831)

73. Wilsson E, Sinn DL. 2012 Are there differences between behavioral measurement methods? A comparison of the predictive validity of two rating methods in a working dog program. *Appl. Anim. Behav. Sci.* **141**, 158–172. (doi:10.1016/j.applanim.2012.08.012)

74. Mirkó E, Dóka A, Miklósi Á. 2013 Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* **149**, 45–54. (doi:10.1016/j.applanim.2013.10.003)

75. Nettle D, Penke L. 2010 Personality: bridging the literatures from human psychology and behavioural ecology. *Phil. Trans R. Soc. B* **365**, 4043–4050. (doi:10.1098/rstb.2010.0061)

76. Cramer AOJ. 2012 Why the item '23 + 1' is not in a depression questionnaire: validity from a network perspective. *Measurement Interdiscip. Res. Perspect.* **10**, 50–54. (doi:10.1080/15366367.2012.681973)

77. Borsboom D, Mellenbergh GJ, van Heerden J. 2004 The concept of validity. *Am. Psychol. Rev.* **111**, 1061–1071. (doi:10.1037/0033-295X.111.4.1061)

78. Borsboom D, Cramer AOJ, Kievet RA, Scholten AZ, Franic S. 2009 The end of construct validity. In *The concept of validity* (ed. RW Lissitz), pp. 135–170. Charlotte, NC: Information Age Publishing.

79. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, Borsboom D, Tuerlinckx F. 2013 A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS ONE* **8**, e60188. (doi:10.1371/journal.pone.0060188)

80. Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R. 2013 Animal personality: what are behavioral ecologists measuring? *Biol. Rev.* **88**, 465–475. (doi:10.1111/brv.12007)

81. Koski SE. 2014 Broader horizons for animal personality research. *Front. Ecol. Evol.* **2**, 1–6. (doi:10.3389/fevo.2014.00070)

82. Ley JM, Bennett PC, Coleman GJ. 2009 A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). *Appl. Anim. Behav. Sci.* **116**, 220–227. (doi:10.1016/j.applanim.2008.09.009)

83. David M, Dall SRX. 2016 Unravelling the philosophies underlying 'animal personality' studies: a brief re-appraisal of the field. *Ethology* **122**, 1–9. (doi:10.1111/eth.12445)

84. Kauffman SA. 1993 *The origins of order: self-organization and selection in evolution.* Oxford, UK: Oxford University Press.

85. Thelen E, Smith LB. 1994 *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: Massachusetts Institute of Technology Press.

86. Smith LB, Thelen E. 2003 Development as a dynamic system. *Trends Cogn. Sci.* **7**, 343–348. (doi:10.1016/S1364-6613(03)00156-6)

87. Barrett L. 2011 *Beyond the brain: how body and environment shape animal and human minds.* Princeton, NJ: Princeton University Press.

88. Goold C, Vas C, Olsen C, Newberry RC. 2016 Data from: Using network analysis to study behavioural phenotypes: an example using domestic dogs. Dryad Digital Repository. (doi:10.5061/dryad.81k11)