

## A general intelligence factor in dogs

Rosalind Arden<sup>a,\*</sup>, Mark James Adams<sup>b</sup>

<sup>a</sup> Centre for Philosophy of Natural & Social Science, Lakatos Building, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

<sup>b</sup> Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, United Kingdom



### ARTICLE INFO

#### Article history:

Received 10 May 2015

Received in revised form 19 January 2016

Accepted 19 January 2016

Available online xxxx

#### Keywords:

Dog

Intelligence

Canine

Cognitive abilities

Evolution

### ABSTRACT

Hundreds of studies have shown that, in people, cognitive abilities overlap yielding an underlying 'g' factor, which explains much of the variance. We assessed individual differences in cognitive abilities in 68 border collies to determine the structure of intelligence in dogs. We administered four configurations of a detour test and repeated trials of two choice tasks (point-following and quantity-discrimination). We used confirmatory factor analysis to test alternative models explaining test performance. The best-fitting model was a hierarchical model with three lower-order factors for the detour time, choice time, and choice score and a higher order factor; these accounted jointly for 68% of the variance in task scores. The higher order factor alone accounted for 17% of the variance. Dogs that quickly completed the detour tasks also tended to score highly on the choice tasks; this could be explained by a general intelligence factor. Learning about g in non human species is an essential component of developing a complete theory of g; this is feasible because testing cognitive abilities in other species does not depend on ecologically relevant tests. Discovering the place of g among fitness-bearing traits in other species will constitute a major advance in understanding the evolution of intelligence.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

In humans cognitive abilities such as navigating through space, understanding written language and number skills correlate positively; a person who is above average at one task is likely to be good at others (Deary, Penke, & Johnson, 2010; Deary, 2013). Hundreds of empirical phenotypic studies show that the structure of human abilities can be represented as a hierarchy with observed manifest measures or tests (such as verbal comprehension or arithmetic) at the bottom level, latent group factors (such as spatial or verbal skills) at the second level and a third factor at the apex (Carroll, 1993). This third factor, called *g* or Spearman's *g* after its discoverer Charles Spearman (Spearman, 1927), is a major focus of psychometric studies in the human behavioural sciences (Jensen, 1998; Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Spinath, Ronald, Harlaar, Price, & Plomin, 2003).

Quantitative genetic methods developed in the 1970s and applied to data from adoption and twin studies have established the existence of genetic *g*; that is, abilities are correlated at the genetic as well as the phenotypic level (Bouchard & McGue, 1981; Deary, Spinath, & Bates, 2006; Loehlin, Horn, & Willerman, 1997; Pedersen, Plomin, Nesselroade, &

McClearn, 1992). More recently, evidence from molecular genetic studies using DNA from large samples of unrelated people show that *g* is highly polygenic (Davies et al., 2011). Research on *g* is motivated partly because it is phenotypically associated with many important life outcomes including health (Batty, Deary, & Gottfredson, 2007; Luciano et al., 2010; Möttus, Luciano, Starr, & Deary, 2013; Schou, Østergaard, Rasmussen, Rydahl-Hansen, & Phanareth, 2012), physical attractiveness (Langlois et al., 2000; Zebrowitz, Hall, Murphy, & Rhodes, 2002), brain resilience (Santarnecchi, Rossi, & Rossi, 2015), and life-expectancy (Batty et al., 2009; Batty et al., 2007; Whalley & Deary, 2001). The phrase *cognitive epidemiology* was coined to characterise research into the association between measured intelligence and traits such as health and life-expectancy in people (Deary & Der, 2005). It would be useful to learn whether the pattern of findings linking higher *g* with better health outcomes (Gottfredson, 2004) is particular to people or common among animals. Links between intelligence and health in non human animals would be especially interesting to probe because other animals neither smoke nor drink alcohol (habits that are lifestyle confounders in human studies). But as the legendary recipe prescribes, 'first catch your hare'; in this case, evidence concerning the structure of cognitive abilities in other species. This 'hare' is an essential first step in probing a link between intelligence and health in other species.

There is some evidence of *g* in non human animals (reviewed in Chabris, 2007; Galsworthy, Arden, & Chabris, 2013; Matzel, Sauce, & Wass, 2013). Yet evidence of the distribution, structure (phenotypic and genetic correlations among cognitive abilities), and the consequences of those differences in other species is exiguous: relatively

\* Corresponding author at: London School of Economics, CPNSS, Lakatos Building, Houghton Street, London WC2A 2AE, United Kingdom.

E-mail addresses: arden.rosalind@gmail.com (R. Arden), m@differentialist.info (M.J. Adams).

<sup>1</sup> MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, United Kingdom.

few studies on general intelligence have been conducted in non human animals since 1920 (one review comprised 21 studies (Chabris, 2007), another comprised 24 studies (Galsworthy et al., 2013)). In order to test whether cognitive abilities are correlated or not, individual-level data on task performance need to be collected, in a sample of reasonable size. This has been done in mice (Galsworthy, Paya-Cano, Monleon, & Plomin, 2002; Locurto, Fortin, & Sullivan, 2002; Matzel et al., 2003; Wass et al., 2012), where a *g* factor was found, and in chimpanzees (Banerjee et al., 2009; Herrmann & Call, 2012; Hopkins, Russell, & Schaeffer, 2014) where a *g* factor was found in two out of three studies.

We tested the structure of measured cognitive abilities in dogs. Dogs and dog breeds are good models for within- and between-species spectra of cognitive abilities. The reasons are plural. Dogs are tractable; they enjoy interacting with people and can visit testing facilities, while living in their own homes. Dogs are not subject to confounding arising from lifestyles that may contribute to causal differences such as smoking, alcohol and drug use. Individual differences in dogs' cognitive abilities are not causally confounded with variability in socio-economic status. It is more feasible, cheaper and less intrusive to conduct repeated behavioural testing with dogs. Following phenotypic studies, dogs will be useful in genetic studies; genes associated with complex traits are easier to find in dogs than people because of their longer haplotype structure in dogs (Lequarré et al., 2011; Ostrander, Giger, & Lindblad-Toh, 2006). A consequence of their haplotype structure is that sample sizes needed for genomic analyses are much smaller in dogs than people. Some behavioural adaptations are breed-specific (pointing, herding); these involve both innate propensities and learning. Some traits are typical across all breeds, such as a tendency to affiliate with humans (see for review Benksy, Sinn, & Gosling, 2013; Miklosi, 2007; Shipman, 2010).

Our underlying assumption was that cognitive abilities would vary among dogs. This is implied by existing data in the animal behaviour literature but variance is rarely the focus of the work. For example, many animal cognition studies are framed as 'can species X do the Y task?' yet the results usually include animals that did, and did not, pass the test. Behavioural variability is the rule not the exception; since variance supplies evolution with its traction, it is a worthwhile object of study.

The present empirical study owes an intellectual debt to the work of John Paul Scott and John L Fuller (Scott & Fuller, 1965). We examined individual differences on a set of cognitive tasks (four increasingly complex versions of a detour task first designed in 1927 by the German psychologist, Wolfgang Kohler (1887–1967) (Frank & Frank, 1982; Scott & Fuller, 1965), a quantity-discrimination task (Bonanni, Natoli, Cafazzo, & Valsecchi, 2011; Macpherson & Roberts, 2013; Prato-Previde, Marshall-Pescini, & Valsecchi, 2008; Ward & Smuts, 2006) and a point-following task (Elgier, Jakovcevic, Mustaca, & Bentosela, 2012; Ittyerah & Gaunet, 2009; Kaminski & Nitzschner, 2013; Lakatos, Gácsi, Topál, & Miklósi, 2012; Miklosi, Soproni, Miklósi, & Soproni, 2006). These tasks were administered to one breed of dog (border collies) selected from similar rearing and living environments. We administered six tasks (of which four were related) to the dogs and, guided by the human psychometrics literature, tested the fit of four basic models against the data.

## 2. Methods

### 2.1. Sample

We recruited 68 farm-living border collies from Wales. We chose a single breed to avoid confounds arising from differential selection. Scores from a basset hound tested against a whippet would be uninterpretable (Udell, Ewald, Dorey, & Wynne, 2014) This is because dogs have been selected by people for different behaviours, and they are the most polymorphic species on earth, varying greatly in leg length and other traits relevant to task performance. We selected farm border collies for several reasons. First, we wanted the dogs' backgrounds to be similar (in contrast with pet or companion animals, because variation in

level of enrichment could contribute to cognitive differences). Although border collies have been subject to artificial selection its focus has been on behaviour more than appearance; border collies remain morphologically variable with a reported moderate inbreeding coefficient of around 2.8% (Hoffman, Hamann, & Distl, 2002) but unknown empirically in our sample. Our sample comprised 68 dogs, (males 34, females 34) ranging in age from 1 to 12 years. We chose Wales as our recruitment centre because it is rural and enriched for border collies, having many hill farms where dogs work stock.

The animals in our sample differ from companion animals in background and behaviour that may be relevant to the study. They are kennelled outdoors and, although socialised to respond to their owner in a farmyard setting, they are unaccustomed to games, indoor behaviour and food treats.

### 2.2. Testing facility

All testing was conducted in a purpose-built barn (see supplementary materials) with a concrete floor that was washed down after each dog's session to eliminate or minimise scents. The same two colleagues administered all the testing. Testers wore plain clothing in all sessions to maintain consistency across dogs, and to reduce distraction in this breed, which is sensitive to visual markers including dress. All equipment was the same for each dog. All start points, and set-up points were marked with tape on the floor. All timings were recorded with a stopwatch.

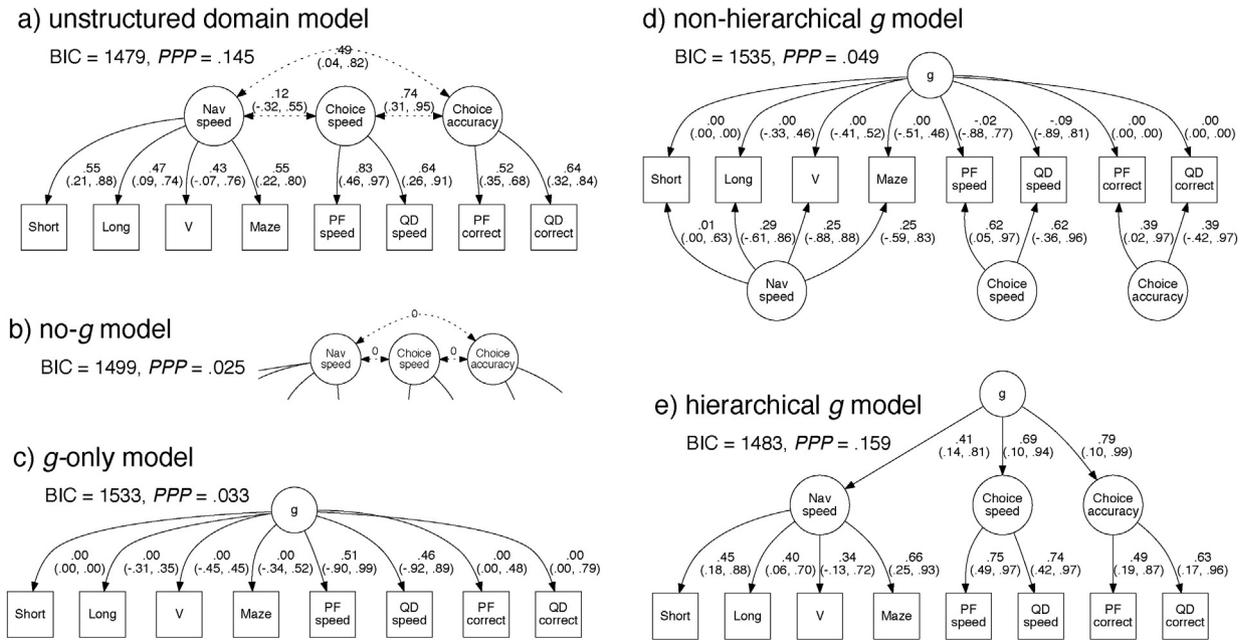
### 2.3. Behavioural tests

On entering the barn, each dog was released to wander freely for three minutes among toys and treats scattered on the floor. This allowed the dogs to relax and adjust to the experimental setting. Diagrams of the set up for each test are given in the electronic supplementary materials.

Problem-solving tests were selected and adapted from the literature. All were appetitive—each problem was motivated by a food treat. We administered 4 versions of a detour test that was designed to measure the underlying construct of insight, navigation, and spatial ability. In each detour test a food treat is placed behind a see-through barrier in 4 configurations (short, long, V-shaped and maze-shaped). The test was to gain the food from a start point. We recorded how much time elapsed between the tester releasing the dog from the start point and the dog reaching the food reward.

The next test, point-following, was designed to measure how well each dog would make a behavioural inference from a visual cue (a human pointing towards a beaker). The point-following task was administered by a tester who stood equidistant between two inverted beakers (one was baited) set on the floor. The tester looked straight ahead, and pointed one arm towards an inverted beaker (the pointed beaker was not baited). On release by the second tester, the dog could choose one (or none) of two beakers to probe. Each dog was administered 10 consecutive trials. The pointed side was pseudo-randomised. We measured how many times the dog went to the pointed beaker as well as how quickly the dog went to the beaker. The baited beaker was not strongly olfactory (as far as we could tell), but served the purpose of focusing the challenge on what seems to be an evolved propensity in dogs—to follow human spatial directive points (Riedel, Schumann, Kaminski, Call, & Tomasello, 2008).

Next we tested individual differences in dogs' discrimination between two quantities by counting how many times each dog went to the larger of two presented food treats. Plates were prepared: a circle was drawn on each plate (diameters were: 2.5 cm, 3.5 cm, 4 cm, 6 cm, and 8.5 cm). Wet dog food mixed with tuna was spread inside the drawn circles. In each trial the dog's attention was directed to a tester holding two plates with circles of different diameters of food. The plates were shown to the dog, and then placed on the floor, in front of the tester. When the dog was released (by the second tester), he (or she) could go to a plate and eat. As soon as the dog went to a plate the other plate



**Fig. 1.** Confirmatory factor analysis of performance on cognitive tasks. Note: Narrower observed variables (squares) are influenced by latent variables (circles). Solid lines represent factor loadings and dashed lines represent correlations among factors. Standardized factor loadings with 95% credible intervals in brackets are placed on the right of each factor loading. Error variances are omitted for clarity. (a) Unstructured domain model posits three first-order factors and allows the correlations between latent factors to be estimated freely. (b) no-g model that fixes the correlations among latent factors to zero. (c) g-only model specifies a single latent factor explaining the covariance among all observed variables. (d) non-hierarchical g model specifies a single general factor in addition to the separate domain factors. (e) hierarchical g model posits a higher-order factor explaining the variance shared among the first order domain factors. Model fit specified by Bayesian Information Criterion (BIC), which gives a relative measure of fit (lower indicates better fit), and posterior predictive p-value, which gives an absolute measure of fit ( $PPP < 0.5$  indicates good model fit,  $PPP < 0.05$  indicates poor model fit).

was removed. This was repeated with different pairings of circle diameters for 8 trials. We measured the number of times each dog chose the plate with the larger diameter, as well as the latency in seconds from release to choice. We aimed to test individual differences in quantity estimation with this test. Further details concerning test administration, diagrams and scoring are given in the Supplementary Materials.

**2.4. Analysis**

We first estimated intra-individual correlations for task completion times to evaluate how consistent a dog was in being either fast or slow. Repeatability in performance is the basis for determining whether there are between-dog differences in ability that could underlie a g-factor. We calculated individual consistency in time to complete the navigation tasks (that is, adjusting for average completion time for all dogs on each barrier configuration). We calculated repeatability of time to complete the point-following task and time to decide in the quantity discrimination task. We estimated consistency and reliability using log completion/decision times in a linear mixed model with dog as a random effect. Age and sex were fit as fixed effects in all analyses. We estimated repeatability as the proportion of variance accounted for by differences between dogs, or  $R = \sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \sigma_{\epsilon}^2)$  where  $\sigma_{\alpha}^2$  is the dog variance and  $\sigma_{\epsilon}^2$  is the residual variance (Nakagawa & Schielzeth, 2010; Sokal & Rohlf, 1995). We then estimated how repeatable the average of performance across multiple trials would be to determine whether mean trial performance would make a more suitable for assessing individual differences in ability. We calculated the consistency of the average navigation performance on four trials and the repeatability of average performance on eight point-following trials and 10 quantity discrimination trials. We estimated extrapolated repeatability (McGraw & Wong, 1996) of mean performance as  $R_{n_0} = \sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \frac{1}{n_0} \sigma_{\epsilon}^2)$  where  $n_0$  is the adjusted number of trials per dog (Gelman & Hill, 2007; Nakagawa & Schielzeth, 2010).

We used confirmatory factor analysis (CFA) to posit latent ability factors that explained the inter-correlations among the observed

(manifest) variables. We log transformed and reversed times so that the observed variables entered into the model represented task speed. We fit five models (Fig. 1): a) an unstructured domain model that posited a latent factor for each type of task; b) a no-g model that constrained the covariances among the latent variables to zero, which specifies that performance within each task domain was correlated but performance is specialised and does not generalize to other task domains; c) one-factor, g-only model specifying that the shared variance between all tasks could be explained by a single latent factor; d) a bifactor, non-hierarchical-g model specifying task-specific factors and a separate task-general factor; and e) a hierarchical-g structure with a single higher-order factor explaining the correlations among the first-order factors. From the hierarchical-g model, we calculated the variance attributable to the general and the first-order factors (“omega”) and attributable to the general factor alone (“omega hierarchical”) (Brunner, Nagy, & Wilhelm, 2012).

We fit the CFA models in Mplus (Muthén & Muthén, 2012) and used a Bayesian estimator that is more sensitive to rejecting models with poor fit when sample size is small, as ours was (B. Muthén & Asparouhov, 2012). Instead of the traditional maximum likelihood fit criteria (e.g.,  $\chi^2$  and RMSEA), this Bayesian estimator reports an absolute measure of model fit, the posterior predictive p-value (PPP), which we used to reject models if  $PPP < 0.05$ ; and a relative measure of fit, the Bayesian Information Criterion (BIC), which we used to compare models that could not be rejected.

**3. Results**

**3.1. Descriptive statistics**

The dogs in our sample demonstrated inter-individual variability. Table 1 shows the raw means, modes, standard deviations and ranges of each test score. There were no significant mean test score differences between the sexes.

**Table 1**  
Descriptive statistics for task performance.

Task	N	Trials	Mean	Mode	SD	Range
Detour short	63	1	11.4 s	4.7 s	20.9 s	1.8–133.0 s
Detour long	64	1	20.7 s	6.4 s	21.8 s	2.3–113.0 s
Detour V	56	1	17.0 s	12.9 s	15.5 s	4.7–74.1 s
Detour maze	59	1	17.6 s	4.5 s	26.7 s	2.9–118.0 s
Point-following time	43	10	4.3 s	1.9 s	5.4 s	0.85–43.0 s
Point-following accuracy	65		11.7	12.0	3.7	2–20
Quantity time	57	8	2.4 s	1.5 s	3.3 s	0.85–38.4 s
Quantity accuracy	65		14.1	12.0	3.9	6–21

### 3.2. Intra-individual variability

We first estimated how much within-dog variability there was on task performance. The consistency of performance was low for navigation ( $R = 0.26$ , 95% credible interval [CI] = 0.11, 0.42) and repeatability was low for the point-following ( $R = 0.35$ , CI = 0.22, 0.50) and moderate for quantity discrimination ( $R = 0.51$ , CI = 0.40, 0.63). Consistency on mean navigation completion time was moderate ( $R_{n0} = 0.58$ , CI = 0.35, 0.74). Repeatability of mean completion time on point-following was high ( $R_{n0} = 0.77$ , CI = 0.63, 0.87) and of average completion time on quantity discrimination was also high ( $R_{n0} = 0.88$ , CI = 0.83, 0.92).

### 3.3. Confirmatory factor analysis and correlations

Because each dog was presented once with each of the four navigation tasks (Short, long, V-shaped and Maze-shaped), we modelled speed to complete each maze as a separate observed variable. Speed on the point-following and quantity discrimination tasks were averaged across trials because these were repeated assessments of the same task configurations and because mean performance on these tasks showed repeatability. We calculated accuracy on the point-following and quantity discrimination tasks as the number of correct responses. The correlations among time on the detour task, time on the point-following and quantity discrimination tasks, and accuracy on the point-following and quantity discrimination tasks are listed in Table S1.

We started with a CFA model, the unstructured domain model, that posited three first-order factors for each cognitive ability domain (i) speed on the navigation tasks, (ii) average speed on the choice tasks (point-following and quantity discrimination), and (iii) accuracy on the choice tasks. The first-order factors represent latent variables that directly influence the narrower observed variables and also account for measurement error in each domain while allowing correlations among the latent variables to be freely estimated. Consequently, this model specifies that similar variables (such as the four navigation tasks) will be influenced by the same latent factor. Different kinds of variables, such as choice speed or accuracy variables, will be influenced by the different latent factors, and the variables in one set will correlate with the those in other sets only to the extent that their latent factors intercorrelate. In this model and in successive models the residual covariances among the observed variables were fixed to zero. The unstructured domain model had acceptable fit (Fig. 1a) and showed that navigation and choice speed factors were positively, though weakly, correlated with each other and both were strongly positively correlated with choice accuracy (Fig. 1a). Thus dogs that were faster also had more correct responses, which showed that there was not a speed/accuracy tradeoff (Fig. 1a). We could reject a no-g model (Fig. 1b). Rejecting the no-g model is not proof of a general intelligence factor but is suggestive of one.

The next models attempted to determine the nature of this g factor. We rejected both a one-factor, g-only (Fig. 1c) and a bifactor, non-hierarchical-g model (Fig. 1d) of general intelligence. Thus intelligence was not described by a general factor directly influencing performance

on each task, as in the g-only model, or after removing task specific variance as in the non-hierarchical-g model. Instead, like human intelligence, cognitive ability in dogs could be described by a hierarchical-g structure (Fig. 1e). Because our models only contained three first-order factors, the unstructured domain model and the hierarchical g-model were equivalent. We were thus not able to assess whether hierarchical g model had a better fit than a correlated factor model. However, the hierarchical model was still useful for calculating how much variation in ability was captured by a g-factor. The g-factor and the first-order factors together captured 68% of the variance in task performance but only 17% of the variance in task performance could be attributed to the g-factor.

We also examined the relationship of age with scores on the domain factors and the hierarchical g-factor. Age was not related to average speed on the detour tasks ( $r = 0.01$ , CI =  $-0.03$  to  $-0.05$ ,  $p = 0.78$ ) or average speed on the pointing and quantity tasks ( $r = 0.15$ , CI =  $-0.13$  to  $0.46$ ,  $p = 0.39$ ). However, older dogs were more accurate on the pointing and quantity tasks ( $r = 0.35$ , CI = 0.11 to 0.72,  $p = 0.004$ ). Finally, age was not related to g ( $r = -0.10$ , CI =  $-0.55$ , 0.53,  $p = 0.61$ ).

## 4. Discussion

Our results indicate that even within one breed of dog, where the sample was designed to have a relatively homogeneous background, there is variability in test scores. The phenotypic structure of cognitive abilities in dogs is similar to that found in people; a dog that is fast and accurate at one task has a propensity to be fast and accurate at another. It may seem obvious that once a detour task (finding the treat behind a barrier) has been solved in one form, the solution to the other forms will follow naturally, but dogs are not people. Experiments have shown that dogs' problem-solving skills do not transfer readily from one problem to a different form of the same problem as ours do (Osthaus, Marlow, & Ducat, 2010). The g factor we report is consistent with the prediction made by the many experts in the 'dog world' (trainers, veterinarians, members of dog societies, and farmers) who were consulted in the early stages of this study. Those experts said that in their experience some dogs were more likely to catch-on, learn and solve problems more quickly than others. Our results show structural similarities between canine and human intelligence. Individual tests have some test-specific variance, tests are influenced by a group-level factor, and the group-level factor is influenced by a g factor. We tested models without the g factor, without the group-level factors and with uncorrelated group-level factors; models positing correlated group-level factors (the unstructured model and the hierarchical g model) fit the data. We emphasize the hierarchical g model because the poor fit of the no-g model rules out uncorrelated first-order factors; the hierarchical g allows us to examine how those correlations arise.

Although we cannot calculate empirically the impact of range-restriction (of intelligence) on our results we surmise that our sample of farm dogs is somewhat analogous to a human university student population because farm dogs at the low tail of the intelligence distribution are more likely to be given away as companion animals. Range restriction attenuates correlations (Alexander, Carson, Alliger, & Barrett, 1984; Wells & Fruchter, 1970) so we cautiously interpret the g factor we found as being a low estimate of commonality. A plot showing the possible impact on our results given various estimates of range restriction is given in the Supplementary Information together with the zero-order correlation matrix for all test scores.

Noise may arise from variation in appetite for treats. We assume that dogs vary in their appetitive motivation—and that differential interest in food treats may be confounded with test scores. Our finding that speed and accuracy are positively correlated suggests that this has not been a major concern, yet we expect that performance on a problem-solving test is affected by more than just 'smarts'. Affective traits such as motivation, persistence, and so on likely influence performance on cognitive

tasks, but if they contribute to covariance among tasks, it may be hard to distinguish these aspects from  $g$ ; there is no a priori reason why  $g$  should not have an affective component. The crucial point is that our study investigates the covariance, the structure, among test scores. In humans where  $g$  has been most studied,  $g$  arises among mathematical and vocabulary tests even though students often have different preferences and motivation to do these kinds of tasks. If  $g$  tapped motivation heavily, we would expect to see covariance among measures of motivation across different kinds of test; in humans we do not see this (Loken, 2004).

#### 4.1. Limitations

Our sample size is reasonably large among animal studies where individual-level data on cognitive scores are less often collected prospectively on one species, but our data are limited. We cannot present measures of test re-test reliability, nor is our sample size as large as in a typical human study. A larger sample would give more power to rule out alternative models and, by setting part of the data aside, allow for exploratory analysis. For example, because we were interested in testing for a trade-off between speed and accuracy, we posited a latent variable that grouped choice speed measures together and another that grouped choice accuracy together. An alternative approach would have been to group the observed variables by task (point-following and quantity determination) rather than by domain (choice speed and choice accuracy). An exploratory modelling step could be used to generate hypotheses about this sort of structure in the data before fitting confirmatory models to a separate set of data. The models we tested were not a flawless fit with the data. While our data were inconsistent with there being no shared ability across the tasks, the magnitude of overlap in task performance was much smaller than that found in human data. We would thus like to see further work on cognitive abilities in dogs with larger samples and tests that capture a broader range of problem-solving abilities. These tests, and this study, are not the 'last word' in mental ability testing in dogs. We hope that the study of individual differences in non human animals' intelligence has a bright future involving more scholars, improved methods and larger samples. We found that dogs did not always perform consistently between trials, nor does the correlation matrix show pairwise significant associations among all tasks. A strong  $g$  factor may have appeared in the one-factor or bi-factor model. However, the mean performance across a larger number of trials could serve as the basis for an intelligence battery that would show high test-retest reliability. A larger variety of tests would also allow us to test models that include a two or more higher order factors that load on overlapping subsets of first-order domain factors, which is a reasonable model given that the estimation of the correlation between two of the first order factors (navigation speed and choice speed in model a) overlapped with zero. A final limitation of our data is that because we assessed intelligence in only one breed, we do not know whether the same structure of intelligence generalizes to other breeds. This study is presented as evidence suggestive of  $g$  in dogs, and as proof of the concept that individual-level data on canine cognitive abilities are useful and collectible.

#### 4.2. The nature of $g$

That  $g$  is a latent variable, the covariance among a set of tasks administered to a group of individuals, leads inescapably to  $g$  being a statistical construct. This does not imply that  $g$  is fugitive, delicately poised on a precarious Jenga of tests. Quite the reverse; it is probably the most robust construct in the whole of the human sciences, evidenced by hundreds of studies (Carroll, 1993). But because  $g$  is a latent variable that emerges from manifest measures (such as solving a detour test, or discriminating between quantities in the dog study above),  $g$  resists between-species comparisons unless the same tasks are administered to each individual. By analogy,  $g$  is like a distillation 'boiled off' from

observed components or scores on specific tests. The distillate varies according to the observed components, but research in the human literature shows that  $g$  factors correlate highly and positively with each other (Johnson et al., 2004).

We suggest that  $g$ , the covariance among manifest abilities, emerges because it taps lower-level biological traits such as neuronal density, myelination, glucose transport in the brain, that vary among individuals. In this way  $g$  may integrate across all brain and central nervous system functions that subserve our capacity to perceive, discriminate, and respond to the problems posed by the environment. To the extent that species inhabit similar cognitive niches and share greater phylogeny, cognitive architecture and brain homologies, the 'distillate',  $g$ , will be more closely comparable. Brains across vertebrate species, for example, have evolved raw materials some of which are held in common across species (such as the hippocampus among mammals) and some that vary, such as the cortex and pallium that evolved differently between mammals and birds.

There is evidence of convergent evolution across taxa in specific functional abilities (such as quantity discrimination and navigation, which is found in mammals, birds and cephalopods) but these may depend on different biological mechanisms. Specific abilities, such as navigation, emerge in turn from lower level processes (such as perception and attention) as higher-level solutions to recurrent problems faced by organisms. These problems may include finding food, shelter, and mates, raising offspring, avoiding harm from predation, inclement weather, gravity and toxins. The taxonomy of cognitive challenges presented by the environment varies among species. Genomes pick out species-relevant problems: the salience of an owl differs between mouse and man (dangerous predator versus wise muse). The set of ancestrally-typical recurrent problems, faced by any individual, varies by species (and sometimes by sex within a species when males and females have faced nuanced differences in the problems posed by their environments); thus the component abilities from which  $g$  arises must vary somewhat between species. Yet the structure of cognitive abilities may be common across species. In order to find out, studies guided by the extant psychometric work on humans are essential. If all animal species with a complex nervous system show general intelligence, this would tell us that variation in intelligence is a universal property of developing brains. If it is more common for cognitive abilities to be uncorrelated, and general intelligence is found only in rare cases, we would learn under what conditions such a dense manifold of traits such as intelligence evolves.

If  $g$  arises from pleiotropy, we would expect to find that  $g$  is a widespread phenomenon across taxa, because the causes are general evolutionary genetic processes rather than particular to people. We would also expect a weak, but detectable, phenotype-wide manifold in most animal species. Since selection acts on behaviour, and behaviour is shaped by species-typical 'problems', we would expect that the cognitive architecture (the matrix of correlational relationships among cognitive traits) would have partly overlapping commonalities across species, as well as some species specificity. The extent of commonality would be determined by the extent to which species and or populations share the same problems. For example we would speculate that the capacity to read others' minds is more highly correlated with bone density (presumably a fitness-bearing trait) in social species such as Pinyon jays (*Gymnorhinus cyanocephalus*) than among the less social Clark's nutcracker (*Nucifraga columbiana*) where reading conspecifics' minds may confer a lesser benefit.

#### 4.3. Measurement of animal intelligence

In discovering  $g$ , Spearman coined the phrase "the indifference of the indicator"; he meant that  $g$  saturates any cognitive task, even the simplest, such as reaction time tasks (Deary, Der, & Ford, 2001). This was an important finding because it shows that one can learn about  $g$  without knowing what selection pressures shaped cognitive abilities, or how

to parse *intelligence* in an evolutionarily-informed manner. The very indifference of the indicator makes it possible to test cognitive abilities in other species without depending on ecologically relevant tasks (such as finding and choosing a mate, avoiding predators, discriminating between nest sites). Tests used routinely on humans are reliable and valid while bearing no relation to the ecological reality of being a person (such as making and retaining friends, finding a place to live, getting a job). They have little connection with the myriad recurrent problems faced by our ancestors and many tests are valid even when administered to people who have little experience with their content or format (Rushton, Čvorović, & Bons, 2007).

Learning about individual differences in animal intelligence is a first step in understanding how cognitive abilities fit into the fitness landscape. It will provide crucial information on the relationship between intelligence and health, ageing and mortality. Data from non human animals are essential if we are to develop a complete understanding of intelligence, one of the most important traits in the entire animal kingdom.

### Conflict of interest

The authors declare no competing interests.

### Acknowledgements

The authors wish to thank Dr. Angela Driscoll and Mrs. Paula Handoll who conducted the experimental work and Alexander Weiss for aid in running the models.

Permission to conduct this study was given by the Animal Welfare and Ethical Review body at King's College London, Denmark Hill campus.

RA conceived, designed and supervised the study. MJA conducted the statistical analyses. Both authors wrote the manuscript and gave final approval for publication.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.intell.2016.01.008>.

### References

- Alexander, R. a., Carson, K. P., Alliger, G. M., & Barrett, G. V. (1984). Correction for restriction of range when both X and Y are truncated. *Applied Psychological Measurement*, 8(2), 231–241. <http://dx.doi.org/10.1177/014662168400800212>.
- Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: Individual differences across cognitive task performance in a new world monkey (*Saguinus oedipus*). *PLoS ONE*, 4(6), e5883. <http://dx.doi.org/10.1371/journal.pone.0005883.1001>.
- Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, 17(4), 278–288. <http://dx.doi.org/10.1016/j.annepidem.2006.07.010>.
- Batty, G. D., Wennerstad, K. M., Smith, G. D., Gunnell, D., Deary, I. J., Tynelius, P., & Rasmussen, F. (2009). IQ in early adulthood and mortality by middle age: Cohort study of 1 million Swedish men. *Epidemiology*, 20(1), 100–109. <http://dx.doi.org/10.1097/EDE.0b013e31818ba076>.
- Benksy, M., Sinn, D. L., & Gosling, S. D. (2013). The World from a dog's point of view: A review and synthesis of dog cognition research. *Advances in the Study of Behaviour*, 209–406. <http://dx.doi.org/10.1016/B978-0-12-407186-5.00005-7>.
- Bonanni, R., Natoli, E., Cafazzo, S., & Valsecchi, P. (2011). Free-ranging dogs assess the quantity of opponents in intergroup conflicts. *Animal Cognition*, 14(1), 103–115. <http://dx.doi.org/10.1007/s10071-010-0348-3>.
- Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, 212(4498), 1055–1059. <http://dx.doi.org/10.1126/science.7195071>.
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796–846. <http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x>.
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. *Educational Researcher*, Vol. 23, Cambridge; New York: Cambridge University Press. <http://dx.doi.org/10.2307/1177226>.
- Chabris, C. F. (2007). Cognitive and neurobiological mechanisms of the law of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 449–491). Hove UK: Psychology Press.
- Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, 16(10), 996–1005. <http://dx.doi.org/10.1038/mp.2011.85>.
- Deary, I. J. (2013). Intelligence. *Current Biology*, 23(16), R673–R676. <http://dx.doi.org/10.1016/j.cub.2013.07.021>.
- Deary, I. J., & Der, G. (2005). Reaction time explains IQ's association with death. *Psychological Science*, 16, 64–69. <http://dx.doi.org/10.1111/j.0956-7976.2005.00781.x>.
- Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, 29, 389–399.
- Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews. Neuroscience*, 11(3), 201–211. <http://dx.doi.org/10.1038/nrn2793>.
- Deary, I. J., Spinath, F. M., & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics: EJHG*, 14(6), 690–700. <http://dx.doi.org/10.1038/sj.ejhg.5201588>.
- Elgier, A. M., Jakovcovic, A., Mustaca, A. E., & Bentosela, M. (2012). Pointing following in dogs: Are simple or complex cognitive mechanisms involved? *Animal Cognition*, 15(6), 1111–1119. <http://dx.doi.org/10.1007/s10071-012-0534-6>.
- Frank, H., & Frank, M. G. (1982). Comparison of problem-solving performance in six-week-old wolves and dogs. *Animal Behaviour*. [http://dx.doi.org/10.1016/S0003-3472\(82\)80241-8](http://dx.doi.org/10.1016/S0003-3472(82)80241-8).
- Galsworthy, M. J., Arden, R., & Chabris, C. F. (2013). Animal models of general cognitive ability for genetic research into cognitive functioning. In C. Reynolds, & D. Finkel (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 257–277). New York: Springer.
- Galsworthy, M. J., Paya-Cano, J. L., Monleon, S., & Plomin, R. (2002). Evidence for general cognitive ability (g) in heterogeneous stock mice and an analysis of potential confounds. *Genes, Brain, and Behavior*, 1(2), 88–95. <http://dx.doi.org/10.1034/j.1601-183X.2002.10204.x>.
- Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Data analysis using regression and multilevel/hierarchical models*.
- Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' & elusive & fundamental cause & of social class inequalities in health? *Journal of Personality and Social Psychology*, 86(1), 174–199. <http://dx.doi.org/10.1037/0022-3514.86.1.174>.
- Herrmann, E., & Call, J. (2012). Are there geniuses among the apes? *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1603), 2753–2761. <http://dx.doi.org/10.1098/rstb.2012.0191>.
- Hoffman, U., Hamann, H., & Distl, O. (2002). Genetic analysis of traits of the working test for border collies. *Zuchungskunde*, 74(2), 213–231.
- Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology*, 24(14), 1649–1652. <http://dx.doi.org/10.1016/j.cub.2014.05.076>.
- Ittyerah, M., & Gaunet, F. (2009). The response of guide dogs and pet dogs (*Canis familiaris*) to cues of human referential communication (pointing and gaze). *Animal Cognition*, 12(2), 257–265. <http://dx.doi.org/10.1007/s10071-008-0188-6>.
- Jensen, A. R. (1998). *The g factor*. Westport, Connecticut: Praeger.
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32, 95–107. [http://dx.doi.org/10.1016/S0160-2896\(03\)00062-X](http://dx.doi.org/10.1016/S0160-2896(03)00062-X).
- Kaminski, J., & Nitzschner, M. (2013). Do dogs get the point? A review of dog-human communication ability. *Learning and Motivation*, 44(4), 294–302. <http://dx.doi.org/10.1016/j.lmot.2013.05.001>.
- Lakatos, G., Gácsi, M., Topál, J., & Miklósi, A. (2012). Comprehension and utilisation of pointing gestures and gazing in dog-human communication in relatively complex situations. *Animal Cognition*, 15(2), 201–213. <http://dx.doi.org/10.1007/s10071-011-0446-x>.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423. <http://dx.doi.org/10.1037/0033-2909.126.3.390>.
- Lequarré, A.-S., Andersson, L., André, C., Fredholm, M., Hitte, C., Leeb, T., ... Georges, M. (2011). LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. *The Veterinary Journal*, 189(2), 155–159. <http://dx.doi.org/10.1016/j.tvjl.2011.06.013>.
- Locurto, C., Fortin, E., & Sullivan, R. (2002). The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. *Genes, Brain, and Behavior*, 2(1), 1–16. <http://dx.doi.org/10.1034/j.1601-183X.2003.00006.x>.
- Loehlin, J. C., Horn, J. M., & Willerman, L. (1997). Heredity, environment, and IQ in the Texas Adoption Project. *Intelligence, heredity, and environment* (pp. 105–125). Cambridge: Cambridge University Press.
- Loken, E. (2004). Online study behavior of 100,000 students preparing for the SAT, ACT, and GRE. *Journal of Educational Computing Research*, 30(3), 255–262.
- Luciano, M., Batty, G. D., McGilchrist, M., Linksted, P., Fitzpatrick, B., Jackson, C., ... Smith, B. H. (2010). Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. *Intelligence*, 38(3), 304–313. <http://dx.doi.org/10.1016/j.intell.2010.03.002>.
- Macpherson, K., & Roberts, W. A. (2013). Can dogs count? *Learning and Motivation*, 44(4), 241–251. <http://dx.doi.org/10.1016/j.lmot.2013.04.002>.
- Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., ... Gandhi, C. C. (2003). Individual differences in the expression of a "general" learning ability in mice. *The Journal of Neuroscience*, 23, 6423–6433.
- Matzel, L. D., Sauce, B., & Wass, C. (2013). The architecture of intelligence: Converging evidence from studies of humans and animals. *Current Directions in Psychological Science*, 22(5), 342–348. <http://dx.doi.org/10.1177/0963721413491764>.

- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <http://dx.doi.org/10.1037/1082-989X.1.1.30>.
- Miklósi, A. (2007). *Dog behaviour, evolution, and cognition*. Oxford: Oxford University Press.
- Miklósi, A., & Soproni, K. (2006). A comparative analysis of the animals' understanding of the human pointing gesture. *Animal Cognition*, 9(2), 81–93. <http://dx.doi.org/10.1007/s10071-005-0008-1>.
- Möttus, R., Luciano, M., Starr, J. M., & Deary, I. J. (2013). Diabetes and life-long cognitive ability. *Journal of Psychosomatic Research*, 75(3), 275–278. <http://dx.doi.org/10.1016/j.jpsychores.2013.06.032>.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <http://dx.doi.org/10.1037/a0026802>.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles CA: Muthén & Muthén.
- Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, 85(4), 935–956. <http://dx.doi.org/10.1111/j.1469-185X.2010.00141.x>.
- Osthaus, B., Marlow, D., & Ducat, P. (2010). Minding the gap: Spatial perseveration error in dogs. *Animal Cognition*, 13(6), 881–885. <http://dx.doi.org/10.1007/s10071-010-0331-z>.
- Ostrander, E. A., Giger, U., & Lindblad-Toh, K. (2006). *The dog and its genome*. Woodbury, New York: Cold Spring Harbor Laboratory Press.
- Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, 3, 346–353. <http://dx.doi.org/10.1111/j.1467-9280.1992.tb00045.x>.
- Prato-Previde, E., Marshall-Pescini, S., & Valsecchi, P. (2008). Is your choice my choice? The owners' effect on pet dogs' (*Canis lupus familiaris*) performance in a food choice task. *Animal Cognition*, 11(1), 167–174. <http://dx.doi.org/10.1007/s10071-007-0102-7>.
- Riedel, J., Schumann, K., Kaminski, J., Call, J., & Tomasello, M. (2008). The early ontogeny of human–dog communication. *Animal Behaviour*, 75(3), 1003–1014. <http://dx.doi.org/10.1016/j.anbehav.2007.08.010>.
- Rushton, J. P., Čvorović, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence*, 35(1), 1–12. <http://dx.doi.org/10.1016/j.intell.2006.09.002>.
- Santaronechi, E., Rossi, S., & Rossi, A. (2015). The smarter, the stronger: Intelligence level correlates with brain resilience to systematic insults. *Cortex*, 64, 293–309. <http://dx.doi.org/10.1016/j.cortex.2014.11.005>.
- Schou, L., Østergaard, B., Rasmussen, L. S., Rydahl-Hansen, S., & Phanareth, K. (2012). Cognitive dysfunction in patients with chronic obstructive pulmonary disease — A systematic review. *Respiratory Medicine*, 106(8), 1071–1081. <http://dx.doi.org/10.1016/j.rmed.2012.03.013>.
- Scott, J. P., & Fuller, J. L. (1965). *Genetics and the social behavior of the dog*. Chicago: University of Chicago Press.
- Shipman, P. (2010). The animal connection and human evolution. *Current Anthropology*, 51(4), 519–538. <http://dx.doi.org/10.1086/653816>.
- Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practices of statistics in biological research [hardcover]*. W. H. Freeman.
- Spearman, C. (1927). *The abilities of man: their nature and measurement*. New York: Macmillan.
- Spinath, F., Ronald, A., Harlaar, N., Price, T., & Plomin, R. (2003). Phenotypic g early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2–4 years. *Intelligence*, 31, 195–210.
- Udell, M. A. R., Ewald, M., Dorey, N. R., & Wynne, C. D. L. (2014). Exploring breed differences in dogs (*Canis familiaris*): Does exaggeration or inhibition of predatory response predict performance on human-guided tasks? *Animal Behaviour*, 89, 99–105. <http://dx.doi.org/10.1016/j.anbehav.2013.12.012>.
- Ward, C., & Smuts, B. B. (2006). Quantity-based judgments in the domestic dog (*Canis lupus familiaris*). *Animal Cognition*, 10(1), 71–80. <http://dx.doi.org/10.1007/s10071-006-0042-7>.
- Wass, C., Denman-Brice, A., Rios, C., Light, K. R., Kolata, S., Smith, A. M., & Matzel, L. D. (2012). Covariation of learning and “reasoning” abilities in mice: Evolutionary conservation of the operations of intelligence. *Journal of Experimental Psychology. Animal Behavior Processes*, 38(2), 109–124. <http://dx.doi.org/10.1037/a0027355>.
- Wells, D. G., & Fruchter, B. (1970). Correcting the correlation coefficient for explicit restriction on both variables. *Educational and Psychological Measurement*, 30(4), 925–934. <http://dx.doi.org/10.1177/001316447003000412>.
- Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, 322(7290), 819–822. <http://dx.doi.org/10.1136/bmj.322.7290.819>.
- Zebrowitz, L., Hall, J., Murphy, N., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin*, 28, 238–249. <http://dx.doi.org/10.1177/0146167202282009>.