
#1

The Social Brain Hypothesis Thirty Years On: Some Philosophical Pitfalls of Deconstructing Dunbar's Number

Robin Dunbar

Abstract

The social brain hypothesis was proposed 30 years ago as an explanation for the fact that primates have much larger brains than all other animals. The claim was that primates live in unusually complex societies, and hence need a large 'computer' to manage the relationships involved. The core evidence subsequently provided in support of this claim was a simple statistical relationship between the social group size characteristic of a species and the size of its brain, with humans fitting into this pattern. However, testing evolutionary hypotheses raises some challenging philosophical and statistical issues that are often overlooked, and great care is needed to ensure that we test the hypothesis we think we are testing. Here, I examine some of these challenges and illustrate the traps they can create for the unwary.

1. Introduction

The social brain hypothesis, and its use to predict a natural grouping size for humans (Dunbar's Number), was established through a series of empirical and theoretical studies beginning thirty years ago (Dunbar 1992, 1993, 1998). Its origin lies in an attempt to understand why primate brains are so much larger than those of all other animals (Jerison 1973). The hypothesis itself is based on the fact that primates live in much more complex societies than other animals, and hence need a bigger 'computer' (i.e. brain) to handle the relationships involved (Byrne 1996; Byrne & Whiten 1988).

Over the decades, it has been established that: (1) there exists a statistical relationship between the typical size of a species' social group and the size of its neocortex, ostensibly derivative of selection for specialised cognition required for group-living (the social brain hypothesis) (Dunbar 1992, 1998; Shultz & Dunbar 2022), (2) the quantitative form of this relationship applies only to primates (in most other mammals and birds, the hypothesis takes the much simpler form of a qualitative switch between pairbonded and non-pairbonded species) (Shultz & Dunbar 2007, 2010), (3) the relationship actually consists of a set of four (possibly five) grades arranged in a fractal series that explains the multilevel structure of primate (and human) social systems (Dunbar 1993, 1998; Kudo & Dunbar 2001; Hill & Dunbar 2003; Zhou et al. 2005; Hill et al. 2009; Sutcliffe et al. 2012; Dunbar & Shultz 2021a), (4) the grades differ in group size, brain size, social complexity, cognitive competences, and ecological context (Dunbar & Shultz 2021a), (5) the regression equation for the social brain relationship predicts a value of ~150 as the core group size for modern humans (Dunbar 1993), (6) there is now considerable empirical evidence that both the size of personal social networks and the size of natural social groups for humans is indeed ~150, and that this is nested within a fractal series of social layers (Dunbar 2020; Wang et al. 2016, 2021; Bird et al. 2019), (7) 150 is a stable value (an 'attractor') because it turns out to be a criticality in the efficiency of information flow in networks with the layers around it forming harmonics (West et al. 2020, 2023), and (8) at least in humans, the fractal structure is the product of a trade-off between the time costs required to maintain different kinds of relationship and the benefits these provide (Sutcliffe et al. 2016; Tamarit et al. 2018, 2022).

The social brain hypothesis was established through a series of empirical and theoretical studies beginning thirty years ago.

Most of these claims are uncontroversial in that they are simply empirical facts,

though some are based on theoretical modelling from first principles. Two, however, have proved to be more contentious in that contradictory findings have sometimes been reported. One is the question of whether the evolution of large brain size in the primate lineage was driven by sociality or some other more strictly ecological (usually dietary) benefit; the other focusses on whether or not the primate social brain relationship, even if true, allows us to predict social group size in modern humans (otherwise known as Dunbar's Number). More interestingly, however, these two questions raise some deep philosophical issues, both about how we test evolutionary hypotheses and about the statistical methods we use. There are hidden traps for the unwary here that apply widely throughout the sciences and the humanities. These traps are both methodologically illuminating and at the same time provide us with novel insights into the processes of evolution. To see why, we need to look at these two claims more closely.

2. How not to test evolutionary hypotheses

There has been a longstanding debate as to whether primate brain evolution has been the outcome of selection for ecologically relevant traits (principally foraging) or selection for the social environment. More specifically, the contrast lies between selection for the capacity to live in stable social groups so as to provide group-level benefits versus selection for the capacity to make individual-level optimal foraging decisions that maximise survival (Dunbar & Shultz 2017). The difference between these two (and there are many subdivisions of each) lies mainly in whether the selection pressure to increase brain size has been due to the need to solve the ecological problems of survival and successful reproduction *socially* (i.e. as a group, with novel forms of cognition needed to enable group cohesion as an intervening step) or by individual trial-and-error learning (with group-living a cognitively costless irrelevance of limited functional significance) (Dunbar & Shultz 2017). The majority of studies so far come down strongly in favour of the social explanation: brain size is correlated with group size rather than ecological variables like diet (Shultz & Dunbar 2022). However, a handful of recent studies (notably DeCasien et al. 2017; Powell et al. 2017) have come to the opposite conclusion, claiming that better data and new statistical methods have made the difference. How is this possible? The answer doesn't, in fact, lie in either the data or the statistical methods, since these don't actually differ at all from the previous studies. The answer lies in several traps for the unwary that lie at the heart of the biological world.

First, there has been a surprising tendency to approach the problem of testing between evolutionary hypotheses with a psychologist's mechanistic frame of mind rather than a biologist's systems-oriented mindset. This causes the alternative hypotheses to be seen as mutually exclusive: one must be right and hence the other, by definition, wrong. Doing so makes it natural to test between the hypotheses using multiple regression analysis. Unfortunately, biological phe-

nomena do not fit well into this format, especially when the data are based on naturalistic observations. Biology is a systems-based discipline, and most organismic phenomena consist of a mixture of causes and their resulting mechanistic effects (outcomes), constraints (things that need to be changed in order to make the effect possible, but do not select for it as such) and consequences (windows of evolutionary opportunity that emerge as added benefits for the main effect, but, again, do not select for it directly). Failing to appreciate the difference leads to a category error (a well known form of logical fallacy) in which different levels of Tinbergen's (1963) "Four Why's" are confounded. The Four Why's refer to the four different kinds of questions a biologist might ask: why (function, or purpose), how (mechanisms), what (ontogeny, or development) and when (phylogeny or evolutionary history). Since these identify different, but equally necessary, parts of a biological system, they are not mutually exclusive in the way a conventional regression analysis assumes. The question we end up asking is not which of two selection factors drove the evolution of large brains, but whether a functional explanation (group size) is more (or less) important than a mechanisms explanation (diet). As Tinbergen reminded us, such a contrast is meaningless, since both must in fact be simultaneously true: everything has a biological function, and every function needs a mechanism to allow it to happen. But – and here is the key point – we cannot equate the two types of explanation as logically equivalent alternatives in a multiple regression analysis.

A second trap is set by Dobzhansky's Dictum (Dobzhansky 1973). Dobzhansky reminded us that when biologists test an hypothesis about evolutionary adaptation, they can do so in either of two equally legitimate ways: by testing for goodness of fit ('being adapted': the eye is designed to allow an animal to see) or by testing for the *process* of adaptation ('becoming adapted' – the 'baby-counting' method familiar from behavioural ecology). The difference lies mainly in the timescale. For cases where the fitness benefit of a trait can be directly observed (e.g. optimal foraging or mate choice decisions where the outcome is immediate), testing the process of adaptation is appropriate because we can watch selection in action (do some individuals gain more fitness-related benefits than others do because of the trait they possess?). In this case, the outcome of selection is, in effect, instantaneous. But for hypotheses that involve the consequences of selection *in the past* (on a time scale of millions of years), or for those where the fitness gain can only be directly measured in the distant future (e.g. at the end of a lifetime), then we have to use a goodness of fit approach. Here, the consequences of selection take a long time to become apparent, or at least longer than the average scientist or funding agency is prepared to wait before seeing a return on their investment. In this case, the hypothesis we test is one about constraints: in the here-and-now, does the historical effect (or outcome of selection) impose a constraint on (i.e. statistically determine) the historical cause (putative selection variable). This harks back to the reason why the selection pressure originally occurred: a constraint on an outcome variable represents the cost an organism has to pay to move a trait in a given direction. In other words, the constraint is the resistance against which selection had to push the organism, and

this will always continue to exist into the future. A species will always resist increasing its brain size beyond what is immediately necessary because doing so incurs energetic and other costs (Dunbar & Shultz 2017, 2021b). Remove the benefit of having a large brain, and there will be selection pressure (proportionate to the energetic costs of neural tissue) to reduce brain size. This kind of reversal seems to have happened several times in the course ungulate and carnivore evolution, but not at all during the course of primate evolution (Pérez-Barbería et al. 2007). In primates, it seems that the two variables are in a very tight evolutionary ratchet.

A species will always resist increasing its brain size beyond what is immediately necessary because doing so incurs energetic and other costs.

Although the underlying hypothesis is actually the same in the two versions of Dobzhansky's Maxim (a cause has selected for an effect), the way we formulate it for testing is reversed in the two cases because of the limitations imposed by the timescales involved. From a process of adaptation ('becoming adapted') perspective, the prediction we actually test is that group size (or diet) selects for (i.e. determines) brain size, but from a goodness of fit ('being adapted') perspective the prediction is that brain size (or diet) constrains (i.e. determines) group size. In other words, the X and Y variables in the regression equation are reversed. This may not matter in a bivariate statistical model, but it can make a great deal of difference in a multivariate statistical model – so much so that it can in fact reverse the conclusion we draw.

The third trap reflects just this and is a consequence of a widely unrecognised limitation in the design of multivariate regression. Most statistical tests were designed with the kind of psychological causality discussed above in mind: we have two variables that we assume are both *bona fide* candidates for being the causal determinant of some effect, and we wish to know which is the more important. The problem is that a regression model will only allow us to have one dependent, or outcome, variable. This means that we need to think very carefully about which variable we allocate to this position: different ways of combining our variables will result in very different questions being asked. If we identify brain size as the dependent variable, we are asking which variables constrain brain size; if we place group size in pole position, we are asking which variables constrain group size. It should be obvious that the answers are likely to be very different. Wartel et al. (2019) confirmed that this is indeed so: run the regression a different way, and you get a completely different answer. Unfortunately, not being biologists, Wartel et al. failed to appreciate the implications of their findings, and instead assumed it was

either an inadequacy of the data or a weakness of the theory. It is neither: it is the consequence of a failure to understand statistics.

So, let's look again at the Social Brain Hypothesis in the light of this. The hypothesis is conventionally tested by regressing group size on brain size (because, in the here-and-now, brain size will be a constraint on group size if group size selected for increases in brain size in the past). Most analyses have done exactly this. However, because DeCasien et al. (2017) and Powell et al. (2017) wanted to test simultaneously between alternative selection drivers, they inverted the causal structure and used brain size as the dependent variable so as to be able to have both group size and diet as drivers in a multiple regression model. Doing so unwittingly commits a category mistake by confounding different types of explanations, or Tinbergen "Why's". In the here-and-now, it is perfectly plausible (and indeed correct) for diet (or foraging competence) to be a determinant of brain size because, ultimately, the size of brain you can grow as an individual is limited by nutrient throughput. It is not at all plausible to suggest that group size (or social skills, for that matter) determines brain size: brain growth is completed (effectively at birth) long before most animals, including humans, achieve full competence in their social skills (which does not occur until subadulthood or even adulthood), never mind arrive at their natural adult group size. In effect, by structuring the statistical model in the way they did, DeCasien et al. (2017) and Powell et al. (2017) assume that causes can act backwards in time – a philosophically interesting proposition for sure, but one that, if true, would have the unfortunate implication that the entire structure of modern physics (and hence all of science) is flawed. Science is based on the assumption that causes *must* precede their effects (as David Hume, the founding father of modern empiricist philosophy, trenchantly reminded us).

In fact, the only correct way to approach problems of this kind is by using path analysis (or, if there are only three variables, mediation analysis). Path analyses using several different datasets (see Dunbar & Shultz 2007, 2017; Navarette et al. 2016; Shultz & Dunbar 2022) yield results that are in close agreement with each other but are diametrically opposite to the ones obtained by DeCasien et al. (2017) and Powell et al. (2017). The reason is obvious: both DeCasien et al. and Powell et al. thought they were testing a selection hypothesis when in fact they were testing a constraints hypothesis. The correct explanatory model turns out to have the form A→B→C, not, as is assumed in their analyses, {(A→B) ⊽ (C→B)} (where, in standard symbolic logic, the logical operator ⊽ stands for disjunction ['either/or but not both']). Diet determines (or more correctly, constrains) brain size (as they, in fact, correctly found), but brain size determines (or constrains) group size. This reflects a selection logic that runs in the reverse direction: had we been able to observe the evolutionary process in action, we would have observed that the need to increase group size imposed a selection pressure favouring an increase in brain size (A→B) so as to enable the animals to maintain social cohesion (in order to solve some ecological problem), and the need to increase brain size in turn imposed a selection

pressure favouring ways of improving nutrient throughput (by changing diet or improving foraging efficiency: (B→C) so as to provide the spare energy needed for the additional brain growth.

At this point, the unwary run the risk of falling for another well-known logical error: the fallacy of the illicit affirmative. Whether or not group size is the main factor selecting for brain evolution is irrelevant to the question of whether brain size imposes a constraint on (i.e. predicts) the size of group that animals can live in. The fact that A is known to cause B (and C does *not* cause B) tells us nothing at all about whether B causes C. Thus, even if it was the case that primate brain evolution was simply a consequence of, say, the fact that animals with larger bodies have larger brains (i.e. increases in brain size are an unintended consequence of selection for large body mass), it may still be the case that brain size constrains group size because that is a derivative psychological constraint, not a selection effect. Whether we can predict group size from brain size is an empirical question about cognitive constraints and cannot necessarily be inferred from the reasons why large brains have evolved. Tinbergen's Four Why's again: questions at different explanatory levels are completely independent of each other.

One last point is worth noting. It has been suggested that these tests of the social brain hypothesis may be subject to a great deal of measurement error because they invariably use mean values. Most species have group sizes that vary widely, especially those that live in large groups. Using the variance in group size might yield very different conclusions about how group size relates to brain size than if we use the mean. In fact, this claim rests on a poor understanding of primate group size dynamics and a poor understanding of statistical theory. Primate group sizes take the form of a nonlinear oscillator that varies round a target mean value set by the local habitat's predation risk and other environmental constraints (see Dunbar et al. 2009; Dunbar et al. 2018). Over time, a group's size varies over a range limited by predation risk and the stresses of group-living, which results in group size having a Poisson distribution, usually with a characteristic peak defining the mean and a very long tail to the right (Dunbar et al. 2009, 2018a,b; Dunbar & MacCarron 2019; Dunbar 2019). In other words, they are almost always Poisson-distributed rather than being normal in form. In a Poisson distribution, the variance is always equal to the mean and so, in fact, it won't matter which statistical moment is used in an analysis. Sandel et al. (2016) confirm that the results of the primate social brain analysis are the same irrespective of whether you use the mean or some measure of the variance.

3. Reconstructing Dunbar's Number

Notwithstanding the points made in the previous section, Lindenfors et al. (2021) have claimed that it is not possible to predict human group sizes (Dunbar's Number) from the equation for the primate social brain hypothesis – or rather, to interpret their actual results correctly, you *can* predict human group size, but the confidence intervals (CIs) are so wide that almost any number would fall within

them. Apparently, they agree that the social brain hypothesis is true (their analysis does in fact confirm that there is a statistically significant relationship between group size regressed on brain size for primates using several datasets and several different regression methods). They presumably also agree that humans do have a natural (or "typical") group size, otherwise there would be no point at all to their analysis. They also agree that this observed human group size must be in the tens or hundreds rather than the thousands, since the CIs they give do not have an infinite range: the 95% CIs on their various estimates range from 2 to 520. If the typical human group size is >520 in size, that would clearly disprove the prediction, so it is not the case that the prediction cannot be tested at all. In fact, the only substantive claim they make is that human group size cannot be predicted *with precision* from the primate social brain equation.

Luckily, their conclusion is not actually true. This is because their analysis simultaneously manages to combine several statistical indiscretions and a fallacy. These indiscretions are, however, themselves rather interesting and have important implications for how we do comparative evolutionary analyses.

First, the fallacy. We have known since the 1930s that, when there are grades present in a dataset, conventional OLS (ordinary least squares) regression underestimates the true slope (a point that was much discussed during the 1980s in the context of comparative analyses of brain size evolution: Mace et al. 1981; Harvey & Mace 1982; Aiello 1992). This is known as Simpson's Paradox (or the Yule-Simpson Effect) and is a version of the ecological fallacy. One indication that there are grades present in a dataset is that the data have a bivariate uniform distribution (i.e. the distribution has the form of a tube) rather than being bivariate normal. OLS regression requires the data to be bivariate normal, otherwise it will always underestimate the slope of the regression. It was pointed out very early on (Dunbar 1993, 1998), and later confirmed by Dunbar & Shultz (2021a), that the primate social brain dataset consists of a set of grades whose regression equations differ in their intercepts but not in their slopes. (Indeed, the presence of grades in the vertebrate brain size data was, in fact, first noted by Jerison [1973].) Failure to take grades into account when using OLS regression (as Lindenfors et al. did) results in a flattened slope and wide confidence intervals (just as they found). This is a consequence of several more subtle statistical issues that they seemed to be unaware of.

One issue is that OLS regression assumes that the X-axis values are measured without error. This is because OLS regression was developed for use in experimental designs where the X-axis values are specified exactly in advance and thus have little or no measurement error (as, for example, in a dose-response experimental design). OLS regression methods exploit this assumption to simplify the process of estimating the statistical moments (mean and variance) of the slope. When there is error variance on both axes, OLS regression inevitably underestimates the slope (resulting, once again, in wide CIs) (Kendall & Stuart 1979; Rayner 1985). Grades in the data add to the problem by increasing the error variance on both variables. For this reason, OLS regression methods are normally recommended only when $r^2>0.95$ (in other

words, when the data approximate a straight line). This problem was widely known as an issue for comparative analyses in the 1970s and 1980s (Mace et al. 1981; Harvey & Mace 1982; Martin 1990), and was specifically identified as a problem in the context of the social brain hypothesis by Dunbar (1992, 1993).

The conventional solution is to use reduced major axis (or RMA) regression since this places the regression line up the centre of the distribution of the data rather than across it. The only problem with RMA regression is that it is still not possible to specify the variance associated with the regression fit: this is because the regression minimises the deviation on both the X and Y axes simultaneously (not just on the Y-axis as OLS regression does) and this makes estimating the variance conceptually tricky – so tricky in fact, that after more than a century it has proved impossible to find a satisfactory solution. As a result, the significance of an RMA relationship cannot be determined. However, since an RMA regression cannot be less significant than the equivalent OLS regression, most statistical packages simply provide only the OLS regression as a default conservative proxy: if the OLS regression is significant, the RMA regression will certainly be.

Our more immediate concern, however, is with what happens when we use an OLS equation to predict a value for an unknown species.

There is, however, a second problem. OLS regression is perfectly fine for determining whether there is a correlation between two variables, since at worst it acts conservatively by underestimating the significance. However, this comes at a cost: while using the OLS obviously minimises Type I errors (you are less likely to reject a false hypothesis), it necessarily increases Type II errors (you are more likely to reject a true hypothesis). This is because Type I and Type II errors are reciprocally related: if you decrease one, you necessarily increase the other, and vice versa. This may be acceptable for hypothesis-testing purposes (all I want to know is how confident I can be that the two variables are correlated), but it is not acceptable if you want to use the regression equation to predict new values. And herein lies a hidden pitfall for multiple regression analysis: the residuals for any variable for which an OLS regression is a poor estimator will be much larger than those for any variable whose slope is accurately estimated. As a result, the effect size for the first variable will be underestimated relative to that for the second variable. This can result in the apparent impact of the two predictor variables being reversed. In extreme cases, the first variable may appear to make no contribution at all.

Our more immediate concern, however, is with what happens when we use an OLS equation to predict a value for an unknown species. Because the slope of the regression is reduced (i.e. flattened), OLS will always overpredict on the left

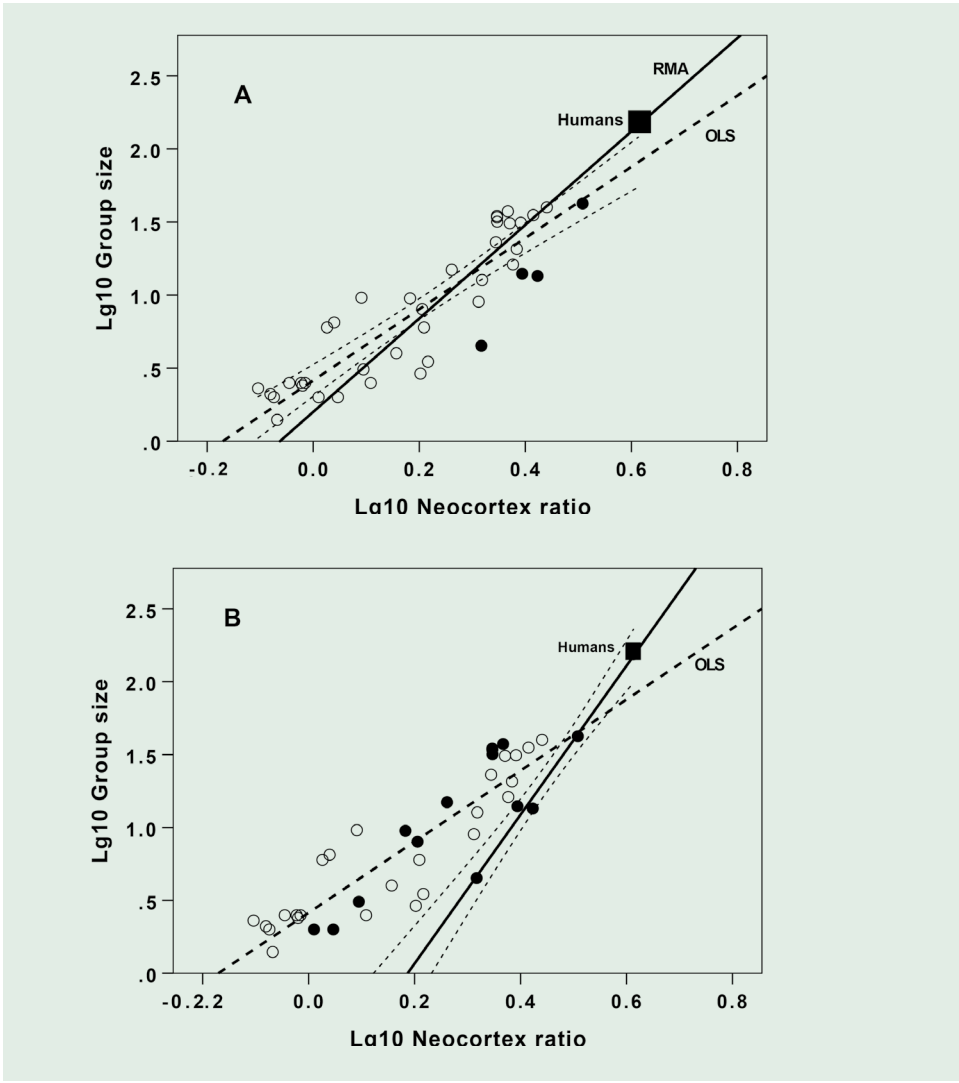

Figure 1. Mean genus group sizes for primates plotted against neocortex ratio for the Stephan brain dataset. The hominoids are indicated as filled symbols (from left to right: *Hylobates, Gorilla, Pongo* and *Pan).* The square symbol indicates the observed mean social group size for humans (154: Dunbar 2020). (a) The heavy dashed line is the OLS regression line (with 95% CIs for the prediction interval as light dashes); the heavy solid line is the RMA regression line. For the OLS regression, $r^2=0.729$; it is not possible to specify statistical moments for RMA regressions, but the fit cannot be less than that for the OLS regression. (b) Grades (indicated by alternating black and white symbols) identified using a *k*-means clustering (see Dunbar & Shultz 2021a). The solid regression line is the OLS regression (with 95% CIs as light dashed lines) for the hominoids-only grade. OLS regression is used because $r^2\geq0.95$ for the individual grades. For the full dataset, $r^2=0.978$ taking grades into account; for the hominoids-only grade, $r^2=0.989$.

Table 1.

Regression equations for the different regression equations set to the data in Fig. 1.

Regression model	Equation	r^2
All data: OLS	$Lg_{10}(Group) = 0.42 + 2.44*Lg_{10}(NeoRatio†)$	0.729
All data: RMA	$Lg_{10}(Group) = 0.22 + 3.11*Lg_{10}(NeoRatio)$	n/a
Mean for 4 grades: OLS	$Lg_{10}(Group) = 0.09 + 3.88*Lg_{10}(NeoRatio)$	0.978
Hominoids grade only: OLS	$Lg_{10}(Group) = -0.95 + 5.11*Lg_{10}(NeoRatio)$	0.989
Apes only: OLS	$Lg_{10}(Group) = -0.90 + 4.97*Lg_{10}(NeoRatio)$	0.958

† Neocortex ratio (neocortex volume divided by volume of rest of brain)

hand side of the graph and underpredict on the right hand side (for data with a positive relationship; the converse will obviously hold if the underlying relationship is negative). The difference is illustrated in Fig. 1a which compares the OLS and RMA regressions set through the classic social brain data. Notice, in particular, the predictions that the two equations would make for humans, compared to the observed value for human groups (~154, range 72-250 for N=24 samples: Dunbar 2020). The 95% CIs for the OLS regression are indicated by the dotted lines. Notice that the datapoint for humans lies well outside the 95% CIs for the OLS regression, whereas the RMA regression line runs right through the human datapoint (the large square symbol).

Fig. 1b plots the same data, but with the consensus grades (from Dunbar & Shultz 2021b) indicated. An OLS regression (acceptable here because $r^2>0.92$ on all four grades) is shown only for the hominoid grade (the grade that includes all the apes and humans). The regression equations, and their respective r^2 values, are given in Table 1. Notice how dramatically the regression slope steepens: the conventional overall OLS regression through the entire dataset has a slope (b=2.44), which is, of course, highly significant ($p<0.001$). It is, however, much shallower than that for the equivalent RMA regression (b=3.11), and both are considerably shallower than the OLS slopes for the individual grades (averaged across the four grades: b=3.88; for the hominoid-only grade: b=5.11; for apes only: b=4.97). The goodness of fit for the conventional overall OLS regression is a respectable $r^2=0.729$, which would normally be considered very acceptable by most standards; the overall goodness of fit for the whole dataset taking the grades into account (Fig. 1b) is $r^2=0.978$, a *very* significant improvement. The goodness-of-fit for the hominoid-only grade on its own is $r^2=0.989$, that for the apes alone is $r^2=0.958$.

These slope differences result in very different predictions for human group size. The conventional OLS regression for the full dataset predicts a value of 82.4 (just as Lindenfors et al. found), and the observed value clearly lies well outside 95% CIs on this estimate (56.2 - 121.6) (Fig. 1a). The

RMA regression does better, with a prediction of 132.0. The hominoid-only regression, however, predicts a value of 152.2 which is indistinguishable from the observed value (with 95% CIs of 100.0-227.5 – very close to those originally estimated in Dunbar [1993], and, in fact, close to the 95% CIs for the actual data). The lesson here is that various regression methods differ in the assumptions they make about the data, and we need to be very careful that we apply the right method for our situation. Failure to do so can lead us woefully astray.

Perhaps fortunately, Lindenfors et al. (2020) rest their claim not on the predictions themselves but on the fact that the confidence intervals for predicted human group size are so wide that almost any value would confirm the prediction. However, in doing so, they compound their statistical problems. The clue lies in the fact that their confidence intervals (2-520) are almost twice as wide as those given by the regression equations in Table 1 (100-227). How can this be? The reason is straightforward. Lindenfors et al. chose to give confidence intervals instead of prediction intervals. Though both are often referred to as confidence intervals, the two are, in fact, conceptually quite different: one is based on the scatter in all the data and hence estimates the range within which all observed values (known and as yet unknown) will lie; the other gives the range within which the slope parameter is likely to vary, and hence the range within which predictions for a mean value should lie. The difference this makes is illustrated by the values for the overall OLS regression line in Fig. 1a: the 95% confidence interval round the prediction is {25.7-245.5} whereas the prediction interval is {56.2-121.6} (it is the second that is plotted in Fig. 1). We are only interested in the second because we are concerned with predicting the *mean* value for humans, not the likely range of values for all possible samples of human populations. The latter isn't especially interesting in this context, any more than the range of group size for individual primate species is especially interesting. The prediction interval radically reduces the range of acceptable values, making for a much more robust test (fewer values will satisfy the criterion).

In reality, the question Lindenfors et al. should have been asking is not how wide the CIs are around the predicted value for human group size, but how close the observed value is to the predicted value. This is a Bayesian question: it asks how good a fit the data are to a predicted value, not whether or not the data differ from the null hypothesis – a contrast reflected in the statistical p-values that indicate significance ($p \geq 0.95$ in the first case, $p < 0.05$ in the second). The observed value of 154 is clearly a very good fit indeed to the value predicted by the social brain equations (for the hominids-only regression: the Bayesian $p_{posterior}=0.95$ 2-tailed; for the ape-only regression, $p=0.95$). To have confused the mean with the entire range of values is careless. Less partisan refereeing and less casual editing should have picked this up.

If Lindenfors et al. had really wanted to debunk Dunbar's Number as a concept (as the tone of their article strongly suggests), then what they really should have done is an empirical analysis to show that there is no characteristic human group size. Had they done so, they would have discovered that a value of ~150 is widely characteristic of human egocentric

(personal) social networks as well as human social organisations in a wide range of contexts (Dunbar 2020). They would have found that many of these samples are extremely large. Of the 20 or so samples that establish this, no less than half have samples >10,000 individuals. One based on a sample of 61,000,000 individuals gave a mean personal network size of exactly 149 (Bond et al. 2012). The robustness of this value may be because 150 constitutes a criticality in the structure of networks: information flow is optimised at this particular network size and falls away rapidly either side of it (West et al. 2020).

One test of the validity of a concept in science is whether it carries sufficient weight to be adopted and applied in the real world.

One test of the validity of a concept in science is whether it carries sufficient weight to be adopted and applied in the real world. As opposed to the world of the academic ivory tower, ideas have to work in the real world of business: there are hardnosed financial consequences if they do not. Dunbar's Number has been applied in a number of interesting real world contexts. The Swedish government tax collection agency, for example, re-organised itself a few years ago so that each tax officer dealt with only 150 clients (a small enough number for them to know each one personally). Twitter is reported to be developing a new social networking facility, with the placeholder name *Flock*, that allows users to limit their list to 150 favoured individuals (Moon 2022) – a design that had underpinned an earlier social networking site *Path* and had been deliberately based on Dunbar's Number. It has also been widely implemented as an ideal team size by business consultancies over the last decade or so. More intriguingly, Dunbar's Number has been used as the basis for a very successful bot detection algorithm (Berry et al. 2019).

It is well to remember that the value of ~150 is just one of a fractally structured series of layers in human social groups and personal social networks, the layers of which have very characteristic values (Zhou et al. 2005; Hamilton et al. 2007; Hill et all. 2008; Arnaboldi et al. 2015; Dunbar et al. 2018; Dunbar 2020; Wang et al. 2020). This fractal structure, or Dunbar graph (Acherjee et al. 2020), extends across a series of layers that range in size from 1.5 to 5000 individuals with a scaling ratio of approximately 3 (each layer is three times the size of the layer inside it), reflecting the complex multilevel structure of all human social groupings (Dunbar 2020). These layers are characteristic of the way Facebook and Twitter personal networks are organised (Arnaboldi et al. 2015; Dunbar et al. 2015) as well as telephone call patterns (MacCarron et al. 2016), the structure of informal business organisations (Webber & Dunbar 2020), alliance patterns in the online gaming world (Fuchs et al. 2014) and the structure of all modern armies (Dunbar 2011). We find

exactly the same layer sizes in the social groups of monkeys and apes (Dunbar et al. 2018; Dunbar & Shultz 2021b; Escribano et al. 2022). In short, the way we organise our businesses and our personal social worlds, and the way the social groups of our monkey and ape cousins are structured, all turn out to follow exactly the same rule and appear to be a consequence of criticalities in the way information flow is optimised in social networks (West et al. 2020, 2023).

4. Conclusions

There are two salutary philosophical lessons here that apply widely throughout the sciences and the humanities. First, the world we live in is complex, and we forget this at our peril. We need to be sure that when we compare between alternative explanations for some effect that we are comparing like with like. The statistical tests we use do not guard against poorly thought out hypotheses: they simply mechanically calculate what they are told to calculate. Failure to think through the theoretical structure of our explanations can easily result in our misinterpreting the statistical results. We need to make much greater efforts to set the particular problem that interests us into the bigger real-world picture of which it is a part. Second, we need to think much more carefully about the way we test hypotheses, and especially about the assumptions that underlie the statistical tests available to us. Too often we just press buttons in software programs without actually knowing what the statistical algorithms actually do, or why the original programmer included the particular version of a test that they did. When we fail to understand the assumptions that lie behind the statistics, we risk sliding into the GIGO (Garbage In, Garbage Out) mode of science.

The Author

Robin Dunbar

Robin Dunbar is a Professor of Evolutionary Psychology at the University of Oxford.

External member of the Finnish Academy of Science and Letters

Photo: Colin McPherson

References

Acharjee, S., Bora, B., Dunbar, R.I.M. (2020). On M-polynomials of Dunbar graphs in social networks. *Symmetry*, 12, 932.

Aiello, L.C. (1992). Size and shape in human evolution. *Journal of Human Evolution,* 22, 127-147.

Arnaboldi, V., Passarella, A., Conti, M., Dunbar, R.I.M. (2015). *Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs*. Elsevier. https://doi.org/10.1016/C2014-0-02436-7

Berry, J.W., Phillips, C.A., Saia, J. (2019). Making social networks more human: A topological approach. *Statistical Analysis and Data Mining*, 12, 449-464.

Bird, D. W., Bird, R. B., Codding, B. F., Zeanah, D. W. (2019). Variability in the organization and size of hunter-gatherer groups: Foragers do not live in small-scale societies. *Journal of Human Evolution*, 131, 96-108.

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D., Marlow, C., Settle, J.E., Fowler, J.H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489, 295-298.

Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology* 5: 172-180.

Byrne, R. W., Whiten, A. (1988). Machiavellian intelligence. In: A. Whiten & R. W. Byrne (eds.) *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, pp. 1-23. Oxford: Oxford University Press.

DeCasien, A.R., Williams, S.A., Higham, J.P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology and Evolution*, 1, 0112.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35, 125-129.

Dunbar, R.I.M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22, 469–493.

Dunbar, R.I.M. (1993). Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16, 681-735.

Dunbar, R.I.M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 6, 178-190.

Dunbar, R.I.M. (2011). Constraints on the evolution of social institutions and their implications for information flow. *Journal of Institutional Economics*, 7, 345-371.

Dunbar, R.I.M. (2019). Fertility as a constraint on group size in African great apes. *Biological Journal of the Linnaean Society*, 129, 1-13.

Dunbar, R.I.M. (2020). Structure and function in human and primate social networks: Implications for diffusion, network stability and health. *Proceedings of the Royal Society, London*, 476A, 20200446.

Dunbar, R.I.M. (2021). Virtual touch and the human social world. *Current Opinion in Behavioral Science*, 44, 14-19.

Dunbar, R.I.M., MacCarron, P. (2019). Group size as a trade-off between fertility and predation risk: implications for social evolution. *Journal of Zoology*, 308, 9-15.

Dunbar, R.I.M., Shultz, S. (2007). Understanding primate brain evolution. *Philosophical Transactions of the Royal Society*, London, 362B, 649-658.

Dunbar, R.I.M., Shultz, S. (2017). Why are there so many explanations for primate brain evolution? *Philosophical Transactions of the Royal Society, London*, 372B, 201602244.

Dunbar, R.I.M., Shultz, S. (2021a). Social complexity and the fractal structure of social groups in primate social evolution. *Biological Reviews*, 96, 1889-1906.

Dunbar, R.I.M., Shultz, S. (2021b). The infertility trap: the fertility costs of group-living in mammalian social evolution. *Frontiers in Ecology and Evolution*, 9, 634664.

Dunbar, R.I.M., Korstjens, A.H., Lehmann, J. (2009). Time as an ecological constraint. *Biological Reviews*, 84, 413-429.

Dunbar, R.I.M., Arnaboldi, V., Conti, M., Passarella, A. (2015). The structure of online social networks mirrors those in the offline world. *Social Networks*, 43, 39-47.

Dunbar RIM, MacCarron, P., Robertson, C. (2018a). Tradeoff between fertility and predation risk drives a geometric sequence in the pattern of group sizes in baboons. *Biology Letters*, 14, 20170700.

Dunbar, R., MacCarron, P., Shultz, S. (2018b). Primate social group sizes exhibit a regular scaling pattern with natural attractors. *Biology Letters,* 14, 20170490.

Escribano, D., Doldan-Martelli, V., Cronin, K.A., Haun, D.B.M., van Leeuwen, E.J.C., Cuesta, J.A., Sanchez, A. (2022). Chimpanzees organize their social relationships like humans. *BioRxiv* https://doi.org/10.1101/2022.03.16.484664

Fuchs, B., Sornette, D., Thurner, S. (2014). Fractal multi-level organisation of human groups in a virtual world. *Scientific Reports*, 4, 1-6.

Hamilton, M.J., Milne, B.T., Walker, R.S., Burger, O., Brown, J.H. (2007). The complex structure of hunter-gatherer social networks. *Proceedings of the Royal Society*, London, 274B, 2195-2202.

Harvey, P.H., Mace, G.M. (1982). Comparisons between taxa and adaptive trends: problems of methodology. In Bertram, B.C.R., Clutton-Brock, T.H., Dunbar, R.I.M., Rubenstein, D.R. & Wrangham, R.W. (Eds.) *Current Problems in Sociobiology*, pp. 343-361. Cambridge: Cambridge University Press.

Hill. R.A., Dunbar, R.I.M. (2003). Social network size in humans. *Human Nature*, 14, 53-72.

Hill, R.A., Bentley, A., Dunbar, R.I.M. (2008). Network scaling reveals consistent fractal pattern in hierarchical mammalian societies. *Biology Letters*, 4, 748-751.

Jerison, H. (1973). *Evolution of the Brain and Intelligence*. London: Academic Press.

Kamilar, J.M., Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society, London,* 368B, 20120341.

Kendall, M. G., Stuart, A. (1979). *The Advanced Theory of Statistics. Vol. 2. Inference and Relationships*. London: Griffin.

Kudo, H., Dunbar, R.I.M. (2001). Neocortex size and social network size in primates. *Animal Behaviour*, 62, 711-722.

Lindenfors, P., Wartel, P., Lind, J. (2021). Deconstructing Dunbar's Number. *Biology Letters*, 17, 20210158.

MacCarron, P., Kaski, K., Dunbar, R.I.M. (2016). Calling Dunbar's numbers. *Social Networks*, 47,151-155.

Mace, G.M., Harvey, P.H., Clutton-Brock, T.H. (1981). Brain size and ecology in small mammals. *Journal of Zoology*, 193, 333-354.

Martin, R. D. (1990). *Primate Origins and Evolution.* London: Chapman and Hall.

Moon, M. (2022). Twitter's experimental 'Flock' feature will let you share tweets with your closest friends. *Engadget* 25 January 2022.

Navarrete, A.F., Reader, S.M., Street, S.E., Whalen, A., Laland, K.N. (2016). The coevolution of innovation and technical intelligence in primates. *Philosophical Transactions of the Royal Society, London*, 371B, 20150186.

Pérez-Barbería, J., Shultz, S., Dunbar, R.I.M. (2007). Evidence for intense coevolution of sociality and brain size in three orders of mammals. *Evolution*, 61, 2811-2821.

Powell, L.E., Isler, K., Barton, R.A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings of the Royal Society, London*, 284B, 2017.1765.

Rayner, J. M. V. (1985). Linear relations in biomechanics: the statistics of scaling functions. *Journal of Zoology*, 286, 415-439.

Sandel, A.A., Miller, J.A., Mitani, J.C., Nunn, C.L., Patterson, S.K., Garamszegi, L.Z. (2016). Assessing sources of error in comparative analyses of primate behavior: Intraspecific variation in group size and the social brain hypothesis. *Journal of Human Evolution*, 94, 126-133.

Shultz, S., Dunbar, R.I.M. (2007). The evolution of the social brain: Anthropoid primates contrast with other vertebrates. *Proceedings of the Royal Society, London*, 274B, 2429-2436.

Shultz, S., Dunbar, R.I.M. (2010). Encephalisation is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences, USA,* 107, 21582-21586.

Shultz, S., Dunbar, R.I.M. (2022). Socioecological complexity in primate groups and its cognitive correlates. *Philosophical Transactions of the Royal Society, London*, 377B, 20210296.

Sutcliffe, A.J., Dunbar, R.I.M., Wang, D. (2016). Modelling the evolution of social structure. *PLoS One*, 11, e0158605.

Sutcliffe, A.J., Dunbar, R.I.M., Binder, J., Arrow, H. (2012). Relationships and the social brain: integrating psychological and evolutionary perspectives. *British Journal of Psychology*, 103, 149-168.

Tamarit, I., Cuesta, J.A., Dunbar, R.I.M., Sánchez, A. (2018). Cognitive resource allocation determines the organisation of personal networks. *Proceedings of the National Academy of Sciences, USA*, 115, 1719233115.

Tamarit, I., Sánchez, A., Cuesta, J.A. (2022). Beyond Dunbar circles: a continuous description of social relationships and resource allocation. *Scientific Reports*, 12, 1-11.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410-433.

Wang, Q., Gao, J., Zhou, T., Hu, Z., Tian, H. (2016). Critical size of ego communication networks. *Europhysics Letters*, 114, 58004.

Wang P., Ma, J.C., Jiang, Z.Q., Zhou, W.-X., Sornette, D. (2020). Comparative analysis of layered structures in empirical investor networks and cellphone communication networks. *EPJ Data*, 9, 11.

Wartel, A., Lindenfors, P., Lind, J. (2019). Whatever you want: Inconsistent results are the rule, not the exception, in the study of primate brain evolution. *PloS One*, 14, e0218655.

Webber, E., Dunbar, R.I.M. (2020). The fractal structure of communities of practice: implications for business organization. *PLoS One*, 15, e0232204.

West, B., Massari, G.F., Culbreth, G., Failla, R., Bologna, M., Dunbar, R.I.M., Grigolini, P. (2020). Relating size and functionality in human social networks through complexity. *Proceedings of the National Academy of Sciences, USA*, 117, 18355-18358.

West, B., Culbreth, G., Dunbar, R. I. M., Griglioni, P. (2023). Fractal structure of human and primate social networks provide optimum information flow. *Proceedings of the Royal Society, London, A*.

Zhou, W.-X., Sornette, D., Hill, R.A., Dunbar, R.I.M. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society, London*, 272B, 439-444.