PROCEEDINGS OF THE ROYAL SOCIETY B

Research

CrossMark
click for updates

Author for correspondence:
T. Dávid-Barrett
e-mail: tamas.david-barrett@psy.ox.ac.uk

Royal Society Publishing
Informing the science of the future

Processing power limits social group size: computational evidence for the cognitive costs of sociality

T. Dávid-Barrett and R. I. M. Dunbar

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

Sociality is primarily a coordination problem. However, the social (or communication) complexity hypothesis suggests that the kinds of information that can be acquired and processed may limit the size and/or complexity of social groups that a species can maintain. We use an agent-based model to test the hypothesis that the complexity of information processed influences the computational demands involved. We show that successive increases in the kinds of information processed allow organisms to break through the glass ceilings that otherwise limit the size of social groups: larger groups can only be achieved at the cost of more sophisticated kinds of information processing that are disadvantageous when optimal group size is small. These results simultaneously support both the social brain and the social complexity hypotheses.

1. Introduction

Social living can take one of two forms: animals can form loose aggregations (exemplified by insect swarms and antelope herds, which are typically based on short-term advantages and whose persistence depends on immediate costs and benefits) or they can form congregations (exemplified by the bonded social groups of primates and some other mammals, including cetaceans, elephants and equids among others) whose advantages derive from long-term association, and whose persistence mainly depends on the trade-off between short- and long-term benefits and costs [1]. It has been suggested that this second kind of sociality is a kind of large-scale coordination problem that depends on bonding processes underpinned by more sophisticated cognitive mechanisms (associated with larger brains [2]). Indeed, the social brain hypothesis [2–7] implies that the size of group that can coordinate its behaviour is limited by the cognitive capacity (essentially, the neural processing capacity) that organisms can bring to bear on the problem.

Although there is considerable neuroanatomical evidence (at the individual as well as the species level) to support the social brain hypothesis in both primates [8] and specifically humans [9–12], the role that cognition plays in this remains unspecified. The recent resurgence of interest in the relationship between communication complexity and social complexity (the social complexity hypothesis) [13,14] offers a possible mechanism by postulating that (i) the complexity of information processing limits social group size and (ii) these communication competences depend on computationally expensive cognitive capacities.

We here use an agent-based model to show that the cognitive demands of coordination impose an upper limit on the size of social group that a species can maintain, but that costly increases in information processing, while disadvantageous when optimal group size is small, can allow the evolution of much larger groups providing there is sufficient benefit from doing so. In our model, the objective function is the size of group that can achieve effective social coordination, and we use a central processing unit's time required to achieve this objective function as an estimate of the cognitive demands of different information-processing strategies. In this, we assume that the size of the brain affects the speed and volume, and hence complexity, of decisions that can be made.

We use a novel coordination task for this. A pure coordination task differs from the conventional public good tasks in that there is no optimal solution, the sole point being to converge on some common behaviour, with the payoff simply reflecting the extent to which the members of the group converge. Conventional public good games involve what amounts to a trading relationship, whereas a coordination task of the kind we use simply requires agents to converge on some common solution. In this respect, it mirrors the common human case in which individuals adopt a set of shared cultural values, where the cultural icon or marker may be arbitrary and the value of the icon itself secondary to the fact that individuals are bound together in a cultural community, which in turn allows them to solve collective action problems more effectively [15,16]. Coordination problems of this kind may have been especially important in the context of the evolution of human sociality [16]. One important consequence of this conception is that, since agents do not pay or withhold some of their capital, there is no payoff for free-riding.

While coordination tasks can, of course, have a direct functional outcome (e.g. agreeing direction of travel to some desirable resource [17]), many coordination tasks in humans simply involve agreeing on cultural markers (or 'tags') to identify group members so as to enable collective action in the future [15,16]. In many of these cases, the coordination task can be quite arbitrary (agreeing on a common cultural icon or belief about the world [18–20], or even a common dialect [21,22]). Cultural convergence of this kind is not a functional end in itself, but rather provides the psychological basis for subsequently achieving a functional end (in some cases—but not always—reinforced through costly signalling [23,24] or costly punishment [25]). Signing up to the same cultural marker may signal acceptance of a set of cultural or moral values that acts as a cue of trustworthiness and willingness to reciprocate.

An important feature of our model is that agents are not panmictic, but rather, as in the real world, are constrained in the number of individuals with whom they can interact by the structure of the social network in which they are embedded [26–28]. The social networks of these species are characterized by dyadic relationships that require expensive maintenance (for instance, in grooming time) leading to long-term edge stability. Although panmixia has often been assumed in models of the evolution of behaviour [29,30], natural human populations are invariably structured and network dynamics are radically different in structured populations [31–33]. Our models map agents onto a network, and so limit the range of agents with whom they can interact.

2. Material and methods

The basic structure of our approach focuses on a group of n agents that face a coordination problem which requires behavioural synchrony, such that each agent has to do her part at the right time and in the right way for the group to be able to act as one (for a general overview of the model, see [31]). We use synchronization on a dial to capture this problem. This is a simple but widely used [31,34] device that stands for almost any coordination problem, though it most obviously relates to how a group decides on direction of travel. (We stress that agreeing a direction of travel is only one of many possible coordination tasks, many of which—as in agreeing on an

arbitrary cultural icon as a marker of group identity—are in themselves only indirectly functional. Agreeing the direction of travel has the advantage of being particularly easy to model.) Each agent is first assigned an initial value between $0°$ and $360°$, and, for convenience, we refer to this as its *information value*. One of these vectors is defined as the 'true information' and is the property of just one agent, while all the other agents are assigned a randomly distributed value. The agents are arranged in a random n-node network in which each agent is linked to k others. The agents' task is to synchronize their information values, which they do by comparing their respective values and then finding a consensus. Synchronization takes place via a set of random dyadic meetings, in which the agents exchange and update their information values. (For detailed mathematical definitions and derivation, see appendices.) After each meeting (and exchange of information), each agent calculates a weighted average of three bits of information: her original information, the partner's information and the information the partner received in her previous meeting. These meetings are repeated until, on average, each agent has taken part in τ meetings. At this point, the average distance between the agents' individual information and the true information is

$$d(n, w, \omega) = \frac{1}{n} \sum_{i=1}^{n} |\phi_{T,i} - \phi_{TI}|,$$

where w and ω are weight matrices for the weights the agent uses for the partners' and the third party information, respectively, while $\phi_{T,i}$ is the information held by agent i at the end of the synchronization episode, ϕ_{TI} is the true information and T is the total number of meetings in the group: $T = n\tau/2$.

Agents are assumed to be trying to get as close to the true information as possible, though they are not necessarily aware of what this value actually is. To do this, each of them estimates an 'optimal' set of weights using a memory of past information exchanges and a simple least-squares optimization function (which is similar to the way humans optimize [35]). Note that requiring the agents to converge on a single value (the true information) is not in itself a defining feature of the model: it is simply a heuristic device to force coordination while at the same time minimizing computational demand. Allowing the agents to find their own equilibrium leads to convergence in just the same way [31], but it invariably takes longer. Our concern is with the constraints that different communication and cognitive strategies place on how fast convergence (synchronization) occurs, and the limits that this imposes on the size of social groups.

In our analysis, agents have three possible ways of estimating the optimal weights, which correspond to increasing levels of cognitive demand. Model $F = 1$ is the simplest: agents ignore both the third party information and the differences among their partners. In model $F = 2$, the agents ignore third party information, but recognize that there are differences (e.g. in reliability) among their partners. In model $F = 3$, the agents use both types of information. In calculating these weights, we varied the size of memory sample that the agents could use in their optimization. We used the processor time associated with each optimization act as an index of the cognitive demand of a strategy, and used this as a proxy for the amount of brain tissue needed in managing a strategy. This allowed us to introduce an implicit distance function

$$\delta(F, n, \tilde{c})|_{\tau,k} = d(n, \hat{w}, \hat{\omega})|_{\tau,k},$$

where $\tilde{c}$ is the measured processor time, F is the index of the method used by the agents, and $\hat{w}$ and $\hat{\omega}$ are the 'optimal' weight matrices as estimated by the agents.

We assume that there is some threshold of synchronization efficiency, above which the group is deemed unable to perform the communal action, and below which the group is in sufficient behavioural synchrony to be able to act as one. Using this
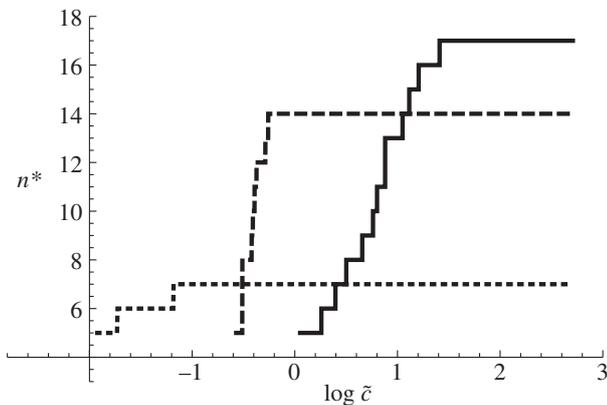
Figure 1. Limiting values for social group size n^* as a function of the processor time costs $\tilde{c}$ required to achieve synchrony in an ecological objective, for three different cognitive strategies in an agent-based model with a structured network. $F = 1$ (dotted line): agents rely only on current information about agents they interact with. $F = 2$ (dashed line): agents take note of individual differences in the quality of information other agents offer. $F = 3$ (solid line): agents take note of individual differences between other agents and rely on third party information about each other received from other agents. Increasing calculation capacity allows larger groups in each of the three strategies for evaluating the quality of potential collaborators. However, each strategy has an upper limit (glass ceiling), and if groups need larger group size they have to switch to more sophisticated methods of information acquisition, and that necessitates an increase in the computational costs (i.e. brain size). (Parameters used: $k = 4$, $\tau = 20$, $\lambda = 11$. The results are robust to these parameters.)

threshold, we can define the largest group that can be in synchrony as follows:

$$n^*(F, \tilde{c})|_{\tau,k,\lambda} = \max n \text{ s.t. } \delta(F, n, \tilde{c})|_{\tau,k} \leq \lambda,$$

where λ is the synchrony threshold.

3. Results

As might be expected, our simulations show that the maximum group size increases as calculation capacity increases for all the three models (figure 1). More importantly, however, figure 1 shows that for both models $F = 2$ and $F = 3$ there is a $c^*(F)$ such that $n^*(F - 1, c) > n^*(F, c)$, for $c < c^*$ and $n^*(F - 1, c) < n^*(F, c)$, for $c > c^*$. In other words, the simplest model with the lowest computational demand allows groups to form, but these are constrained to relatively small sizes. There is some possibility of increasing group size by increasing computational capacity, but this option is capped. To achieve a significant further increase in group size, the agents must switch to a more complex information-processing strategy (model $F = 2$) that allows them to differentiate among their partners. Note that while this method is uneconomical for small groups, it yields an increase in group size if the calculation capacity is large enough. Just as in the simple model, a further increase in calculation capacity allows a further increase in group size, but it too hits an upper limit (albeit at a higher level). To move beyond this limit, the agents not only need larger computational capacities, but also have to add an additional information stream (model $F = 3$). This is the least economical method for small or middle-sized groups, but, by permitting third party information to be exchanged, it allows the group to

cut through the glass ceiling imposed by the other two models and significantly raises the limit on group size.

4. Discussion

It is important to note that the more complex strategies are highly disadvantageous when group size is small: indeed, the more complex the strategy, the more disadvantageous it is. Thus, the evolution of communicative and cognitive complexity is explicitly dependent on an ecological demand for large social groups: it is only when there is a need for large groups that the selection pressure will be sufficient to warrant the costs involved. This is driven by the demands of social coordination. If there is no requirement for coordination (in terms of the present model, optimal group size $n^* \approx 1$ because organisms do not need to cooperate), then there is insufficient selection pressure to motivate either complex communication or investment in the large brains required to support the requisite cognitive abilities.

This suggests that complex communicative abilities, large social groups that involve social coordination and large brains will all be equally rare, as indeed seems to be the case [13]. Although we have used a very simple (and computationally easy to implement) objective function in the model (achieving synchrony in compass direction), it is important to be clear that our model is not limited to this particular context. Rather, our device stands as an abstraction for any behaviour that requires synchrony or coordination in order to maximize biological fitness, whether the fitness payoff is a direct or an indirect consequence of coordination along stable, expansive dyadic network edges. Direct fitness payoffs may arise from coordinated foraging or hunting (as in some social carnivores [36]), cooperative defence against predators or rival conspecifics (e.g. group territorial defence; as in many primates) or any other ecologically relevant behaviour that requires group members to synchronize or coordinate their behaviour in some way. However, in species characterized by multi-level social systems (e.g. elephants, some cetaceans, most primates [37] and humans [27,28]) where the higher level of organization functions to solve a collective action problem, the payoff may be indirect and mechanisms are needed to facilitate cooperation at group level. In these cases, solving a collective action problem is a two-step process: willingness to collaborate is established before the need to collaborate [38], and is often (but not necessarily always) signalled by some marker (or 'tag') of group membership.

Our model not only shows how cognition (processing power) could limit group size, but also sheds light on why some other species (for instance, herding mammals with panmictic group structure) use entirely different—and much simpler—coordination strategies. This, of course, does not preclude the possibility that bonded species such as humans might use simple heuristics to solve coordination tasks when the task demands are simple [39,40]. However, our analysis does raise the question as to why some species choose simple coordination methods associated with panmictic structure, while others resort to using more complex and costly cognition.

In sum, our analysis provides a formal mechanism for both the social brain hypothesis [3] and the social (or communicative) complexity hypothesis [13] by demonstrating that greater information-processing demands are reflected in

3

greater cognitive (computational) costs, but that bringing these on stream can allow organisms to break through glass ceilings to significantly increase social group size. If size matters (large groups offer greater protection against predators [41], are more efficient for foraging [42] or win more territorial fights [43,44]), then there will be selection pressure to pay these costs. But the significant finding is that these costs are considerable and, when optimal group size is small, make the costs prohibitive. In these circumstances, simpler cognitive strategies are more profitable. In the limiting case, when there is no requirement for bonded relationships to ensure group stability through time, it is not worth paying the costs of complex cognition. This suggests that these kinds of more complex sociality will be relatively rare. Broadly speaking, this is what we see in the natural world [2].

Appendix A. The formal model

Let us define a group as a set of n agents that are interlinked in a connected network with each agent having a degree of k. Let $G(n)$ denote the set of all possible such networks:

$$G(n) = \{g(n) | \#e_i = k \forall i, \rho(i,j) < \infty \forall i, j\},$$

where e_i denotes the set of agents that agent i is connected to, and hence $\#e_i$ denotes the length of this set and $\rho(i,j)$ denotes the network distance between nodes i and j.

Let f denote the basic information-updating function, defined in the following way:

$$f(A,B,x) = \begin{cases} f_1 & \text{if } 0° \leq f_1 \leq 360° \\ f_1 - 360° & \text{if } 360° < f_1 \\ f_1 + 360° & \text{if } f_1 \leq 0° \end{cases}$$

$$f_1(A,B,x) = \begin{cases} A - \dfrac{x}{1+x}(A-B) & \text{if } 0° \leq |A-B| \leq 180° \\ A - \dfrac{x}{1+x}(A-B-360°) & \text{if } 180° < A-B \\ A - \dfrac{x}{1+x}(A-B+360°) & \text{if } A-B < -180°. \end{cases}$$

Thus, the function f calculates a weighted average of two values on a dial, A and B, with the weights being, respectively, 1 and x. Thus, this function is a simple weighted average, and the only reason for the complication above is that it is on a dial. (It is possible to define f using trigonometric functions; however, the form is less intuitive that way.)

Let h denote the information-updating function with two information sources, defined as a nested f function:

$$h(A,B,C,x,y) = f(f(A,B,x), f(A,C,y), y/x).$$

That is, the way the function h works is that first the function f calculates the weighted average of dial values A and B using the weight x, and dial values A and C using the weight y; and then the weight of the resulting two values is calculated using the weight y/x. (Note that due to the construction of the y/x weight, the parameters x and y do not have symmetric effect. This reflects the fact that B and C play different roles in the model, with B being the information the agent's meeting partner holds, and C being the third party information.)

Let $\phi_{t,i}$ denote the value of the information variable held by agent i at time t. Then, let $\text{TI} \sim U\{1, 2, \ldots, n\}$ denote a randomly selected agent that receives the true information, $\phi_{\text{TI}} \sim U(0°, 360°)$. The initial information values are defined

the following way:

$$\phi_{0,i} = \begin{cases} \phi_{\text{TI}} & \text{if } i = \text{TI} \\ \sim U(0°, 360°) & \text{if } i \neq \text{TI}. \end{cases}$$

That is, all agents receive a random initial value on a dial, except for the agent that was chosen as TI, which receives the true information, ϕ_{TI}.

Let $\pi_{t,i}$ denote the third party information, and let $p_{t,i}$ denote the index of the agent from which the third party information originates, both held by agent i at time t. The initial values for these two variables are set the following way:

$$\pi_{0,i} = \phi_{0,i} \forall i$$

and $\quad p_{0,i} = i \forall i.$

That is, the first third party information the agent receives from herself.

Let us define a synchronization event on a network $g(n) \in G(n)$ as a series of T meetings among the agents, where a 'meeting' between two agents, a and b, is defined in the following way:

$$\{a,b\} \sim U\{1, 2, \ldots, n\}^2 \text{ s.t. } a \neq b, a \in e_b, b \in e_a.$$

That is, first, the agents a and b are chosen such that they are linked to each other. Second, the agents exchange and update their respective information, third party information, and third party agent index the following way:

$$\phi_{t+1,a} = h(\phi_{t,a}, \phi_{t,b}, \pi_{t,b}, w_{a,b}, \omega_{a,p_{t,b}}),$$
$$\phi_{t+1,b} = h(\phi_{t,b}, \phi_{t,a}, \pi_{t,a}, w_{b,a}, \omega_{b,p_{t,a}}),$$
$$p_{t+1,a} = b,$$
$$p_{t+1,b} = a,$$
$$\pi_{t+1,a} = \phi_{t,b}$$

and $\quad \pi_{t+1,b} = \phi_{t,a},$

where $w_{a,b}$ and $\omega_{a,p_{t,b}}$ are the weights associated, respectively, with the information agent a receives from agent b and the third party information she receives from agent b.

These random meetings among the agents are repeated until $t = T$, where $T = \tau n/2$. That is, on average each agent takes part in τ meetings.

Given the definition of the synchronization as a series of information exchanges on a graph, we can define a function that measures the average distance of the agent from the true information at the end of the synchronization process. Let d denote this measure in the following way:

$$d(n, w, \omega) = \frac{1}{n} \sum_{i=1}^{n} |\phi_{T,i} - \phi_{\text{TI}}|.$$

where $w = \{w_{i,j}\}$ and $\omega = \{\omega_{i,j}\}$ are the respective weight matrices.

(Note that $d(n,w,\omega)$ is independent of the arbitrary parameter ϕ_{TI}. And thus, this structure is why we employ the cumbersome use of a dial, rather than a one-dimensional range for the information variable. At the same time, if the weight parameters of the agents and the network structure are well behaved, the average distance, d, converges to zero as τ goes to infinity.)

Let us assume that a set of n agents, linked up in the network $g(n) \in G(n)$, goes through a series of S synchronization events in a such a way that both the agent selected to be TI

and the true information stays the same throughout, while the information variables of all non-TI agents acquire a new random initial value in each of the synchronization events. (All elements of the synchronization event are as defined earlier.) Then, let $r_{s,t,i}$ denote a memory record that agent i collects at meeting number t in the synchronization event s, defined the following way:

$$r_{s,t,i} = \begin{cases} \{a_{s,t}, b_{s,t}, p_{s,t,b}, \phi_{s,t,a}, \phi_{s,t,a}, \pi_{s,t,p_{s,t,b}}\} & \text{if } i = a \\ \{b_{s,t}, a_{s,t}, p_{s,t,a}, \phi_{s,t,b}, \phi_{s,t,a}, \pi_{s,t,p_{s,t,a}}\} & \text{if } i = b \\ \varnothing & \text{otherwise.} \end{cases}$$

That is, if an agent is party to a meeting, she records all the relevant information in her memory.

Then let R_i denote the entire memory of agent i, containing all her records for all of T meetings in all of S synchronization events:

$$R_i = \left\{ \{r_{s,t,i}\}_{t=1}^{T} \right\}_{s=1}^{S}.$$

(For the initial values of $w_{i,j}$ and $\omega_{i,j}$ used in the memory build-up, see below.)

The agents will select a sample from their memory, and choose a weight that will minimize their distance from the true information. Before we describe this process, however, we introduce a generic weight-optimization function. Let $z(Q)$ denote the optimal weight given the sample Q, defined the following way:

$$z(Q) = \arg\min_{\beta} E\left[\left(\phi_{\text{TI}} - \frac{A + \beta B}{1 + \beta} \right)^2 \middle| \{A, B\} \in Q \right], \quad \text{(A 1)}$$

where E is the expectations operator. That is, z is the coefficient that minimizes the distance of the weighted value from the true information. (We used optimization on a linear range rather than on a dial, as the latter results in impractically long calculation time. Robustness checks showed, however, that there is no significant difference between the two measures in practice.)

Given the definition of optimization in (A 1) let us consider three different models of how the agents sample their memory, and calculate the optimal weights. Let F serve as an index of these models.

In the first model, $F = 1$, the agents ignore all third party information and do not differentiate among the other agents they encounter. Formally,

$$Q_i \sim U\{\{A, B\} | \{a, b, c, d, A, B, C, D, \} \in R_i, a = i, b \neq i\}^q,$$

and

$$\hat{w}_{i,j} = \begin{cases} z(Q_i) & \text{if } j \in e_i \\ 0 & \text{if } j \notin e_i \end{cases}$$

and

$$\hat{\omega}_{i,k} = 0 \forall k = 1, 2, \dots, n,$$

where q denotes the size of the sample. That is, the agents only focus on primary information, and only differentiate self from not-self.

In the second model, $F = 2$, the agents still ignore third party information, but they do differentiate among their partners. Formally,

$$Q_{i,j} \sim U\{\{A, B\} | \{a, b, c, d, A, B, C, D, \} \in R_i, a = i, b = j\}^q,$$

and

$$\hat{w}_{i,j} = \begin{cases} z(Q_{i,j}) & \text{if } j \in e_i \\ 0 & \text{if } j \notin e_i \end{cases}$$

and (as before)

$$\hat{\omega}_{i,k} = 0 \forall k = 1, 2, \dots, n.$$

That is, the agents still only focus on the primary information, but also use the information that lies in the differentiation between their partners. That is, by recognizing that their partners are different agents, they allow the weights they assign to their partners to reflect the difference in the quality of the information.

In the third model, $F = 3$, the agents use the third party information and differentiate among the other agents. Formally,

$$Q_{i,j} \sim U\{\{A, B\} | \{a, b, c, d, A, B, C, D, \} \in R_i, a = i, b = j\}^q,$$

$$Q_{i,k} \sim U\{\{A, B\} | \{a, b, c, d, A, B, C, D, \} \in R_i, a = i, c = k\}^q,$$

and (as before)

$$\hat{w}_{i,j} = \begin{cases} z(Q_{i,j}) & \text{if } j \in e_i \\ 0 & \text{if } j \notin e_i \end{cases}$$

and

$$\hat{\omega}_{i,j} = \begin{cases} z(Q_{i,k}) & \text{if } Q_{i,k} \neq \varnothing \\ 0 & \text{if } Q_{i,k} = \varnothing. \end{cases}$$

Let $c(F, q, n, g(n))$ denote the processor time, which is the time it takes for the simulation computer to perform the optimization defined in (A 1). The processor time varies with the model index F, the sample size q, the group size n and the graph (network) that connects the agents.

Using this concept of processor time, let us define an implicit distance function δ the following way:

$$\delta(F, n, \tilde{c})|_{\tau,k} = d(n, \hat{w}, \hat{\omega})|_{\tau,k}, \quad \text{(A 2)}$$

where $\tilde{c}$ is the average measured processor time for all $g(n) \in G(n)$. (That is, $\tilde{c}$ is a parameter that is measured during the simulation process.) Note that $\#e_i = k \forall i$ and $g(n)$ allows the comparative measure, and thus the definition of the implicit distance function in (A 2). Also note that due to the construction of (A 2) the processor time c scales with the sample size q (although not linearly), allowing for a wider interpretation of computational complexity.

Using the δ function, we can define a maximum group size, denoted by n^*, as follows:

$$n^*(F, \tilde{c})|_{\tau,k,\lambda} = \max n \text{ s.t. } \delta(F, n, \tilde{c})|_{\tau,k} \leq \lambda,$$

where λ is an arbitrary synchrony threshold. That is, n^* is the largest number of agents that can perform a group synchronization (on an average k-degree network) in a way that the group reaches a threshold of synchrony, given the calculation capacity and method of optimization of the agents.

Proposition A.1. $(\partial n^*(F, \tilde{c}) / \partial \tilde{c}) > 0$ for all $F = 1, 2, 3$.

Proposition A.2. For both $F = 2,3$ there is a $c^*(F)$ such that $n^*(F - 1, c) > n^*(F, c)$ for $c < c^*$ and $n^*(F - 1, c) < n^*(F, c)$ for $c > c^*$.

There is no algebraic solution to this problem, to our knowledge. However, a simulation result is obtained with $\tau = 20$, $k = 4$ and $\lambda = 11°$. (Further computational parameters used were estimation limits in (A 2): $-L \leq \beta \leq L$, where we used $L = 1000$, and the number of synchronization events in the
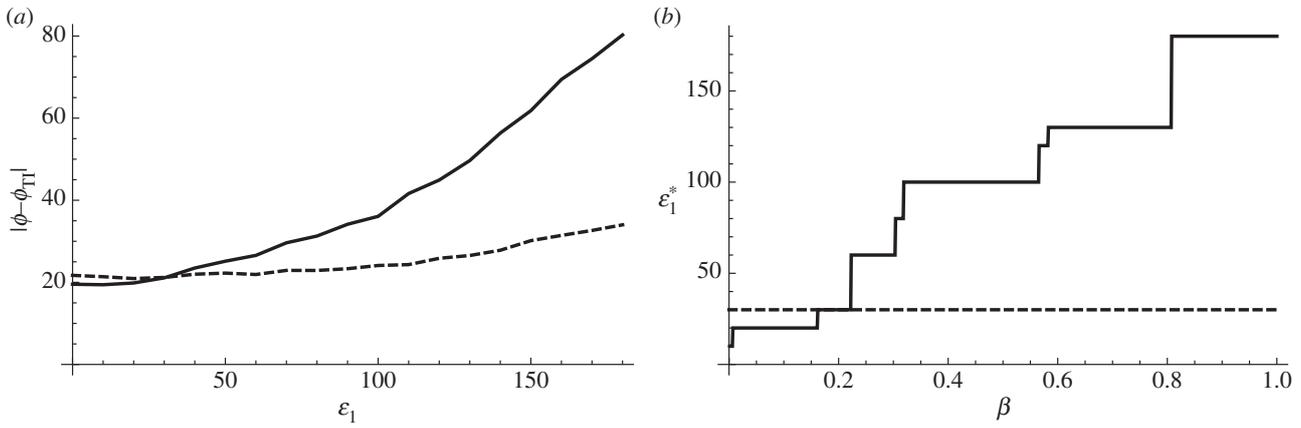
Figure 2. The public good problem can be bypassed in the behavioural synchrony framework. (*a*) How the error level of agent 1 affects agent 1's and the other agents' respective distances from the true information. Solid line, agent 1; dashed line, mean for all the other agents. (*b*) A positive relationship between the cost parameter (β) and agent 1's optimal error level. Solid line, optimal error level of agent 1; dashed line, mean error level of all other agents. Note that if the cost is sufficiently low, then agent 1 would choose to have an error level lower than that of the rest of the group.

calculated memory $S = 100$. The initial values for the simulations' memory build up were $w_{i,j} = 1$ and $\omega_i k = 0$ for $F = 1,2$, and $w_{i,j} = 1$ and $\omega_i k = 1$ for $F = 3$, for all i, j, and k. For robustness checks, see the electronic supplementary material.)

For the calculated evidence for both propositions A.1 and A.2, see figure 1. (Note that due to the nature of numerical simulation, the inequalities of propositions A.1 and A.2 hold only weakly for some parameter ranges.)

Appendix B. A note on evolutionary dynamics

The model we present in this paper shows the group size limit of any network using the particular mode of information updating. One of the important questions concerning our model is whether there is a plausible evolutionary mechanism that would favour higher processing capacity on the individual level. Although our model is primarily concerned with constraints on the group size that would serve as limiting factors of any evolutionary process, we recognize that there might be a collective action problem (or public good problem) underlying the phenomenon we are modelling [45]. If the increased processing power is costly for the individual while aiding the group-level synchronization efficiency only marginally, then there might be an incentive for the individual to 'cheat' on the others by reducing its processing power, and thus free-ride on the others' higher capacities. If that was evolutionarily beneficial, there would be no group-level action at all [46].

Let us consider the set of graphs $G(10)$ with $k = 5$, a group of 10 agents linked up in a random graph in such a way that each of them is connected to five others. Let us, for simplicity, assume that they are going through a series of synchronization events without any optimization, without differentiating among their partners and without third party information (i.e. $w = 1$ and $\omega = 0$).

However, let us assume that the information reception of the agents contains noise in the following way:

$$\phi_{t+1,i} = f(\phi_{t,i}, \hat{\phi}_{t,j}),$$

where

$$\hat{\phi}_{t,j} \sim U(\phi_{t,j} - \varepsilon, \phi_{t,j} + \varepsilon)$$

and where

$$\varepsilon = \begin{cases} \varepsilon_1 & \text{if } i = 1 \\ 30 & \text{if } i \neq 1 \end{cases}.$$

That is, all agents apart from agent 1 receive the information from others within a bandwidth of ± 30. For agent 1, the error level can vary in the range $\varepsilon \in (0,180)$.

Let us assume that, just as in the main model of the paper, there is true information received by one randomly selected agent. After $\tau = 30$ information exchanges per agent, we measure the average distance of the true information from agent 1, as well as from all the other agents. We found that both agent 1's distance and the other agents' distance from the true information increases with the error level of agent 1; however, the former relationship is more steep (figure 2*a*).

Now let us consider a payoff function for agent 1 that has the following form:

$$p_1 = a - \frac{1}{n}\sum_{i=1}^{n} |\phi_{T,i} - \phi_{\text{TI}}| - \beta \cdot (180 - \varepsilon_1) - |\phi_{T,1} - \phi_{\text{TI}}|,$$

(B 1)

where p is the total payoff, a is the payoff that agent 1 receives as a result of the collective action, β is a cost parameter and $n = 10$. Thus, the payoff function is of the following form: (payoff to focal individual) = (benefit of collective action) − (average group deviance) − (cost of cognitive investment) − (focal individual's deviation).

That is, the total payoff that the agent receives is dependent on the group being able to perform the collective action with efficiency, where the agent faces a cost for reducing her information's noise level, and she pays a penalty for being far from the true information. (Note that this is a particular example for a payoff function, where all the functional forms are linear, and the cost associated with group and individual efficiency is 1 in both cases. Trivially, the payoff from the collective action, a, can be any value, but unless a is big enough, p will not be positive and hence no action takes place.)

This formulation of the payoff function allows us to search for the optimal error level given the cost parameter:

$$\varepsilon_1^*(\beta) = \arg\max_{\varepsilon_1 \in (0,180)} p_1.$$

Simulation results show that—as expected—the distance from the true information increases with the error level of

the focus agent for both the focus agent and the rest of the group, albeit with a different slope (figure 2a). This results in the effect that agent 1's optimal level of error increases as the cost of noise reduction increases (see figure 2b). Importantly, in some parameter ranges, the optimal error is smaller than that of the rest of the group. This suggests that if the focus agent's error level was the same as the rest of the group initially, reducing the error level would be beneficial for this individual, providing the route to evolutionary dynamics.

Note that there are two reasons why the public good problem does not arise in the context of a coordination game, in line with the observation that public good problems are essentially two-trait in nature [47]. First, the public good in a coordination game is non-rival, which extends the range of the cost parameter (β), under which the public goods

problem does not emerge. Second, and most importantly, investment in the public good and exploitation of the public good are one and the same thing. This means that the focus agent must invest in the public good (increase coordination efficiency) in order to gain from it at all (increase the agent's alignment). Hence, there is no possibility of acting as a free-rider: there is no benefit to withdrawing from the public good, and the value of the public good is diminished as a direct consequence.

Although the functional form of the payoff function (equation (B 1)) is entirely arbitrary, this simple example illustrates an interesting property of the behavioural synchrony model: it can bypass (or perhaps even resolve) the public good problem, adding to the existing literature on the relationship between cognition and cooperation [48,49].

References

1. Dunbar RIM, Shultz S. 2010 Bondedness and sociality. *Behaviour* **147**, 775–803. (doi:10.1163/000579510X501151)

2. Shultz S, Dunbar R. 2010 Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proc. Natl Acad. Sci. USA* **107**, 21 582–21 586. (doi:10.1073/pnas.1005246107)

3. Dunbar RIM. 1992 Neocortex size as a constraint on group-size in primates. *J. Hum. Evol.* **22**, 469–493. (doi:10.1016/0047-2484(92)90081-J)

4. Dunbar RIM. 1998 The social brain hypothesis. *Evol. Anthropol.* **6**, 178–190. (doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8)

5. Barton RA. 1996 Neocortex size and behavioural ecology in primates. *Proc. R. Soc. Lond. B* **263**, 173–177. (doi:10.1098/rspb.1996.0028)

6. Byrne RB, Whiten A (eds) 1988 *A Machiavellian intelligence: social complexity and the evolution of intellect in monkeys, apes and humans*. Oxford, UK: Oxford University Press.

7. Barton RA, Dunbar RIM. 1997 Evolution of the social brain. In *Machiavellian intelligence. II. Extensions and evaluations* (eds A Whiten, RB Byrne), pp. 240–263. Cambridge, UK: Cambridge University Press.

8. Sallet J et al. 2011 Social network size affects neural circuits in macaques. *Science* **334**, 697–700. (doi:10.1126/science.1210027)

9. Lewis PA, Rezaie R, Brown R, Roberts N, Dunbar RI. 2011 Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage* **57**, 1624–1629. (doi:10.1016/j.neuroimage.2011.05.030)

10. Powell J, Lewis PA, Roberts N, Garcia-Finana M, Dunbar RIM. 2012 Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans. *Proc. R. Soc. B* **279**, 2157–2162. (doi:10.1098/rspb.2011.2574)

11. Bickart KC, Wright CI, Dautoff RJ, Dickerson BC, Barrett LF. 2011 Amygdala volume and social network size in humans. *Nat. Neurosci.* **14**, 163–164. (doi:10.1038/nn.2724)

12. Kanai R, Bahrami B, Roylance R, Rees G. 2012 Online social network size is reflected in human brain structure. *Proc. R. Soc. B* **279**, 1327–1334. (doi:10.1098/rspb.2011.1959)

13. Freeberg TM, Dunbar RI, Ord TJ. 2012 Social complexity as a proximate and ultimate factor in communicative complexity. *Phil. Trans. R. Soc. B* **367**, 1785–1801. (doi:10.1098/rstb.2011.0213)

14. Dunbar RIM. 2012 Bridging the bonding gap: the transition from primates to humans. *Phil. Trans. R. Soc. B* **367**, 1837–1846. (doi:10.1098/rstb.2011.0217)

15. Dunbar RIM. 2008 Mind the gap: or why humans aren't just great apes. *Proc. Br. Acad.* **154**, 403–423.

16. Ihara Y. 2011 Evolution of culture-dependent discriminate sociality: a gene-culture coevolutionary model. *Phil. Trans. R. Soc. B* **366**, 889–900. (doi:10.1098/rstb.2010.0247)

17. Sigg H, Stolba A. 1981 Home range and daily march in a hamadryas baboon troop. *Folia Primatol. (Basel)* **36**, 40–75. (doi:10.1159/000156008)

18. Tajfel H, Billig MG, Bundy RP, Flament C. 1971 Social categorization and intergroup behavior. *Eur. J. Soc. Psychol.* **1**, 149–177. (doi:10.1002/ejsp.2420010202)

19. Riolo RL, Cohen MD, Axelrod R. 2001 Evolution of cooperation without reciprocity. *Nature* **414**, 441–443. (doi:10.1038/35106555)

20. Antal T, Ohtsuki H, Wakeley J, Taylor PD, Nowak MA. 2009 Evolution of cooperation by phenotypic similarity. *Proc. Natl Acad. Sci. USA* **106**, 8597–8600. (doi:10.1073/pnas.0902528106)

21. Nettle D, Dunbar RIM. 1997 Social markers and the evolution of reciprocal exchange. *Curr. Anthropol.* **38**, 93–99. (doi:10.1086/204588)

22. Cohen E. 2012 The evolution of tag-based cooperation in humans the case for accent. *Curr. Anthropol.* **53**, 588–616. (doi:10.1086/667654)

23. Sosis R, Alcorta C. 2003 Signaling, solidarity, and the sacred: the evolution of religious behavior. *Evol. Anthropol.* **12**, 264–274. (doi:10.1002/evan.10120)

24. Sosis R, Kress HC, Boster JS. 2007 Scars for war: evaluating alternative signaling explanations for cross-cultural variance in ritual costs. *Evol. Hum. Behav.* **28**, 234–247. (doi:10.1016/j.evolhumbehav.2007.02.007)

25. Boyd R, Gintis H, Bowles S, Richerson PJ. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535. (doi:10.1073/pnas.0630443100)

26. Granovetter M. 1973 The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380. (doi:10.1086/225469)

27. Hamilton MJ, Milne BT, Walker RS, Burger O, Brown JH. 2007 The complex structure of hunter–gatherer social networks. *Proc. R. Soc. B* **274**, 2195–2202. (doi:10.1098/rspb.2007.0564)

28. Zhou WX, Sornette D, Hill RA, Dunbar RIM. 2005 Discrete hierarchical organization of social group sizes. *Proc. R. Soc. B* **272**, 439–444. (doi:10.1098/rspb.2004.2970)

29. Boyd R, Richerson PJ. 1985 *Culture and the evolutionary process*, pp. viii, 331. Chicago, IL: University of Chicago Press.

30. Sachs JL, Mueller UG, Wilcox TP, Bull JJ. 2004 The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160. (doi:10.1086/383541)

31. David-Barrett T, Dunbar RI. 2012 Cooperation, behavioural synchrony and status in social networks. *J. Theor. Biol.* **308**, 88–95. (doi:10.1016/j.jtbi.2012.05.007)

32. Dunbar RIM. 2011 Constraints on the evolution of social institutions and their implications for information flow. *J. Institutional Econ.* **7**, 345–371. (doi:10.1017/S1744137410000366)

33. Gomez-Gardenes J, Reinares I, Arenas A, Floria LM. 2012 Evolution of cooperation in multiplex networks. *Sci. Rep.* **2**, 620. (doi:10.1038/srep00620)

34. Couzin ID, Krause J, Franks NR, Levin SA. 2005 Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516. (doi:10.1038/nature03236)

35. Griffiths TL, Tenenbaum JB. 2006 Optimal predictions in everyday cognition. *Psychol. Sci.* **17**, 767–773. (doi:10.1111/j.1467-9280.2006.01780.x)

36. Kruuk H. 1972 *The spotted hyena: a study of predation and social behavior*, pp. xvi, 335. Chicago, IL: University of Chicago Press.

37. Hill RA, Bentley RA, Dunbar RIM. 2008 Network scaling reveals consistent fractal pattern in hierarchical mammalian societies. *Biol. Lett.* **4**, 748–751. (doi:10.1098/rsbl.2008.0393)

38. Harcourt AH. 1992 Coalitions and alliances: are primates more complex than non-primates? In *Coalitions and alliances in humans and other animals* (eds AH Harcourt, FBM de Waal), pp. 445–472. Oxford, UK: Oxford University Press.

39. Dyer JR, Ioannou CC, Morrell LJ, Croft DP, Couzin ID, Waters DA, Krause J. 2008 Consensus decision making in human crowds. *Anim. Behav.* **75**, 461–470. (doi:10.1016/j.anbehav.2007.05.010)

40. Kearns M, Suri S, Montfort N. 2006 An experimental study of the coloring problem on human subject networks. *Science* **313**, 824–827. (doi:10.1126/science.1127207)

41. Shultz S, Noe R, McGraw WS, Dunbar RIM. 2004 A community-level evaluation of the impact of prey behavioural and ecological characteristics on predator diet composition. *Proc. R. Soc. Lond. B* **271**, 725–732. (doi:10.1098/rspb.2003.2626)

42. Reader SM, Hager Y, Laland KN. 2011 The evolution of primate general and cultural intelligence. *Phil. Trans. R. Soc. B* **366**, 1017–1027. (doi:10.1098/rstb.2010.0342)

43. Adams ES. 1990 Boundary disputes in the territorial ant *Azteca trigona*: effects of asymmetries in colony size. *Anim. Behav.* **39**, 321–328. (doi:10.1016/S0003-3472(05)80877-2)

44. Wilson ML, Wrangham RW. 2003 Intergroup relations in chimpanzees. *Annu. Rev. Anthropol.* **32**, 363–392. (doi:10.1146/annurev.anthro.32.061002.120046)

45. Torney CJ, Levin SA, Couzin ID. 2010 Specialization and evolutionary branching within migratory populations. *Proc. Natl Acad. Sci. USA* **107**, 20 394–20 399. (doi:10.1073/pnas.1014316107)

46. West SA, Griffin AS, Gardner A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672. (doi:10.1016/j.cub.2007.06.004)

47. Brown SP, Taylor PD. 2010 Joint evolution of multiple social traits: a kin selection analysis. *Proc. R. Soc. B* **277**, 415–422. (doi:10.1098/rspb.2009.1480)

48. Brosnan SF, Salwiczek L, Bshary R. 2010 The interplay of cognition and cooperation. *Phil. Trans. R. Soc. B* **365**, 2699–2710. (doi:10.1098/rstb.2010.0154)

49. McNally L, Brown SP, Jackson AL. 2012 Cooperation and the evolution of intelligence. *Proc. R. Soc. B* **279**, 3027–3034. (doi:10.1098/rspb.2012.0206)