In the format provided by the authors and unedited.

Inference of ecological and social drivers of human brain-size evolution

Mauricio González-Forero[1] & Andy Gardner[1]*

1School of Biology, University of St Andrews, St Andrews, UK. *e-mail: mgf3@st-andrews.ac.uk

Supplementary Information for:

Inference of ecological and social drivers

of human brain-size evolution

Mauricio González-Forero & Andy Gardner

Contents

1 Model overview

Our model introduces social interactions into a previous non-social model[1]. We consider a clonal, well-mixed, female population of large and constant size, where the environment is constant, generations are overlapping, individual's age is measured in continuous time, and the population size is regulated via density-dependent competition through fecundity (common demographic assumptions in life history models[2]). We partition the body of each individual into brain tissue, reproductive tissue, and the rest, which we call somatic tissue. Alternative partitions of body mass are possible, for example to study trade-offs of digestive[3] or adipose[4] tissues with brain tissue, but here we consider this minimal partition as a simpler starting point (Supplementary Information section 9). At each time, each individual has genetic control of the fraction of growth metabolic rate allocated to each tissue, which reflects the amount of energy that is allocated to the production of each tissue. We call the allocation profile throughout life the growth strategy, which is the evolving trait in the model. We assume that some of the energetic consumption of the brain is due to skill production and maintenance, and we consider skills that allow the individual to overcome challenges of energy extraction of four types: ecological (me vs nature), cooperative (us vs nature), between-individual competitive (me vs you), or between-group competitive (us vs them) challenges. If the individual fails to overcome the challenge, the individual can extract energy from an environment facilitated by her mother if the individual is young enough. Natural selection can then mould how much an individual grows her brain given brain metabolic costs and that brain supports skills that allow the individual to obtain energy by overcoming the different types of energy extraction challenges.

To study the evolution of the growth strategy, we perform a standard invasion analysis. Specifically, we seek a growth strategy that constitutes a monomorphic evolutionary equilibrium (which is the relevant case for brain size as we consider only females so sex dimorphism is not a concern), called an uninvadable strategy[5]. Hence, we assume that almost all individuals in the population have a resident growth strategy v and we assess whether or not the growth strategy u of rare mutant individuals can invade (increase in frequency in) the resident population. We thus seek an uninvadable growth strategy u^* so that, when it is a resident, it resists invasion by any mutant strategy taken from the set U of feasible growth strategies. From these assumptions and those stated in the previous paragraph, an uninvadable strategy u^* satisfies[6]

$$u^* \in \arg\max_{u \in U} R_0(u, u^*) \tag{1}$$

where

$$R_0(u, v) = \int_0^T \ell(t) m(t) \mathrm{d}t \tag{2}$$

is the expected lifetime number of offspring of the mutant strategy u given that residents have strategy v. Here, T is an age after which the individual no longer reproduces or is dead, $\ell(t)$ is the probability that a mutant survives from birth to age t (survival), and $m(t)$ is a mutant's rate of offspring production at age t with density dependence (effective fecundity). Hence, Eq. (1) implies that an uninvadable strategy u^* is a best response to itself in the sense that no other rare strategy has a higher expected lifetime number of offspring than u^* when u^* is resident. Thus, below we search for an uninvadable strategy by identifying a strategy that is a best response to itself (similar to a Nash equilibrium, but different because here only the mutant's payoff is maximized whereas in a Nash equilibrium the payoffs of both players are maximized).

Finding a u^* satisfying Eq. (1) poses an "evolutionary differential game" problem fully specified in section 3 below. To solve it, we proceed like this: given a resident, we find a best response to the resident by solving an optimal control problem; then, we set this best response as the new resident and iterate until we achieve a situation where the resident and its best response are indistinguishable to a chosen extent. To facilitate such convergence, we constrain the best response to be similar to a certain degree to the resident; this technique is known as damping[7] and is an assumption of weak selection (specifically, of phenotypic or δ-weak selection[8]).

As a first approximation, we assume that the mortality rate μ is independent of age and of the evolving traits (this assumption can be relaxed in future extensions of the model; section 9). Consequently, survival reduces to $\ell(t) = \exp(-\mu t)$. Also, we equate reproductive tissue with preovulatory ovarian follicles for three reasons: (i) so that reproductive tissue is not involved in offspring maintenance, (ii) so that the metabolic costs of reproductive tissue are representative of the metabolic costs of the germline, and (iii) so that reproductive tissue directly relates to fecundity (the latter two reasons are explained in paragraph B_r in section 4.3). Since reproductive tissue thus defined is not involved in offspring maintenance, we can write effective fecundity at age t as being proportional to the mass $x_r(t)$ of reproductive tissue at that age[1]: that is, $m(t) \approx C(\nu) f_0 x_r(t)$, where $C(\nu)$ is a positive value that depends on the resident strategy and regulates population size, and f_0 is a proportionality constant that measures the number of offspring produced per unit time per mass unit of reproductive tissue in the absence of density dependence.

Hence, an uninvadable growth strategy u^* generates an "optimal" mass $x_i^*(t)$ of tissue i at every age t (for $i \in \{b, r, s\}$ denoting brain, reproductive, and somatic tissue, respectively) and an "optimal" skill level $x_k^*(t)$ at every age. These values are optimal in the sense that they result from a growth strategy that is a best response to itself. Additionally, an uninvadable growth strategy u^* generates the timing of the life history. In particular, with the parameter values used, the resulting uninvadable growth strategy typically produces a life history with four critical ages where the growth strategy changes suddenly (called switching times in optimal control terminology): the age of brain growth onset t_{b0}, which is when allocation to brain growth starts; the age of brain growth arrest t_b, when allocation to brain growth stops; the age at maturity t_m, when allocation to growth of reproductive tissue starts; and the age at adulthood t_a, when allocation to growth of non-reproductive tissues stops. These four ages are an output, not parameters, of the model.

2   Expressions for energy extraction efficiency

In the model, each individual extracts energy depending on her own skills and on the skills of her social partners. Specifically, we assume that at each age the individual faces an energy extraction challenge that can be of one of four types and if the individual fails at the challenge, then she can extract energy from an environment facilitated by the mother if the individual is young enough. The four challenge types are: (1) ecological, where the individual acts alone against a constant environment (me vs nature); (2) cooperative, where the individual acts in concert with a cooperating partner of the same age against the constant environment (us vs nature); (3) between-individual competitive, where the individual acts alone against a competitor of the same age (me vs you); and (4) between-group competitive, where the individual acts with a cooperating partner against two cooperating competitors, all of the same age (us vs them). We index these challenge types, respectively by

$j \in \mathbb{C} = \{1, 2, 3, 4\}$. As explained below, we let P_j be the proportion of energy-extraction challenges of type $j \in \mathbb{C}$ that the individual faces during any vanishingly small time interval $[t, t + \Delta t]$, so that $\sum_{j \in \mathbb{C}} P_j = 1$. We assume that P_j is constant with respect to time for all $j \in \mathbb{C}$. We allow the contribution of cooperating partners to be either additive, multiplicative, or submultiplicative (i.e., the geometric average of their skill level).

2.1   Derivation of energy extraction efficiency with maternal provisioning

In this section, we derive an expression for $e(t)$ that implements our consideration of various challenge types and our assumption that if the individual fails to overcome a challenge, then she can extract energy from an environment facilitated by her mother if the individual is young enough.

Let $E(t)$ be the rate of energy extraction at age t of the focal mutant individual and let $E_{\max}(t)$ be an individual's rate of energy extraction at age t if she is maximally successful at energy extraction (in this section, arguments highlight time dependence, although variables may have other dependencies). We define $e(t)$, the energy extraction efficiency (EEE) of the focal mutant individual at age t, as $e(t) = E(t)/E_{\max}(t)$. Let $E_j(t)$ be the rate of energy extraction at age t of the focal mutant when she faces a challenge of type $j \in \mathbb{C}$. We then define the energy extraction efficiency when facing a type-j challenge as $e_j(t) = E_j(t)/E_{\max}(t)$.

From these definitions, we have that the amount of energy that the mutant extracts during a time interval $[t, t + \Delta t]$ is approximately $E(t)\Delta t$, where the approximation is good when $E(t)$ varies little over the time interval and becomes exact as Δt tends to zero. Similarly, the amount of energy that the mutant extracts during a time interval $[t, t + \Delta t]$ given that it is facing a type-j challenge is approximately $E_j(t)\Delta t$. We assume that the individual faces an energy extraction challenge at every time and let the proportion of challenges of type j during the time interval $[t, t + \Delta t]$ be $P_j(t, \Delta t)$, which in general depends on t and Δt. Hence, the amount of energy that the mutant extracts during the time interval $[t, t + \Delta t]$ is approximately $\sum_{j \in \mathbb{C}} P_j(t, \Delta t) E_j(t)\Delta t$. That is,

$$E(t)\Delta t \approx \sum_{j \in \mathbb{C}} P_j(t, \Delta t) E_j(t)\Delta t, \tag{3}$$

where the approximation becomes exact when Δt approaches zero. Hence, dividing Eq. 3 by $E_{\max}(t)\Delta t$ and taking the limit $\Delta t \to 0$, we obtain

$$e(t) = \sum_{j \in \mathbb{C}} P_j(t) e_j(t), \tag{4}$$

where $P_j(t)$ is no longer dependent on Δt, provided the limit converges. For simplicity, we assume that the proportions $P_j(t)$ are constant with respect to t for all $t \in [0, T]$. In practice, the meaning of P_j can be seen as follows. Provided that the unit of t is years, that Δt is comparatively smaller than this unit (say, one week), and that the rates of energy extraction $E(t)$ and $E_j(t)$ are fairly constant during weekly time intervals, then $P_j(t)$ are the proportions of type-j challenges faced during a week starting at age t; our assumption of constant $P_j(t)$ with respect to t then means that the proportion of type-j challenges faced during any week is approximately the same throughout (reproductive) life. This assumption of constant P_j can be relaxed in future extensions of the model (section 9).

We now seek expressions for the energy extraction efficiency at age t when facing type-j challenges, $e_j(t)$. Suppose that when the individual faces only energy extraction challenges of type j during a time interval $[t, t + \Delta t]$, the individual succeeds at these challenges during a proportion $S_j(t, \Delta t)$ of the time interval. When she

succeeds, the individual extracts energy at a rate $E_{\max}(t)$, but when she fails at energy extraction, she extracts energy at a rate $\Phi(t)$ from maternal provisioning, which decreases with age t. Hence, we have

$$E_j(t)\Delta t \approx S_j(t, \Delta t)E_{\max}(t)\Delta t + [1 - S_j(t, \Delta t)]\Phi(t)\Delta t, \tag{5}$$

where the approximation becomes exact when Δt approaches zero. Dividing Eq. (5) by $E_{\max}(t)\Delta t$ and taking the limit $\Delta t \to 0$, we obtain

$$e_j(t) = S_j(t) + [1 - S_j(t)]\varphi(t), \tag{6}$$

where $\varphi(t) = \Phi(t)/E_{\max}(t)$ is the energy extraction efficiency from maternal provisioning and $S_j(t)$ is no longer dependent on Δt, again provided the limit converges.

We do not assume that $S_j(t)$ is constant with respect to time. Instead, we let the success proportion at energy extraction challenges of type j take the form of a contest success function[9,10]:

$$S_j(t) = \frac{c_j(t)}{c_j(t) + d_j(t)}, \tag{7}$$

where $c_j(t)$ measures the individual's "competence" at energy extraction in a type-j challenge at age t, and $d_j(t)$ measures the "difficulty" of a type-j challenge at that age. Using Eqs. (6) and (7) in (4) yields the individual's energy extraction efficiency

$$e(t) = \sum_{j \in \mathbb{C}} P_j \left[\frac{c_j(t)}{c_j(t) + d_j(t)} + \frac{d_j(t)}{c_j(t) + d_j(t)}\varphi(t) \right]. \tag{8}$$

Setting $P_1 = 1$ yields the EEE in the purely ecological setting:

$$e(t) = \frac{c_1(t) + d_1(t)\varphi}{c_1(t) + d_1(t)}. \tag{9}$$

Eq. (9) has a different form from that used in the non-social model we build upon[1] (i.e., compared to their Eq. 32, where $e = c/[c + \alpha(1 - \varphi)]$), which did not have a derivation for incorporating maternal provisioning (φ). This difference still allows for comparison with that work as our improved expression [Eq. (9)] yields the same predictions with the same benchmark parameter values after adjusting the parameter value controlling maternal provisioning at birth (Extended Data Fig. 2g).

2.2 Equations for competence, challenge difficulty, and maternal provisioning

In this section, we define the expressions for the energy extraction efficiency e_j when facing a challenge of type j (in this section, arguments highlight skill dependencies rather than time dependencies). We assume that some of the brain metabolic rate is due to production and maintenance of domain-general skills that can be used to overcome the four types of challenges. Let $x_k(t)$ and $y_k(t)$ be, respectively, the domain-general skill level of a mutant and a resident of age t. An individual's competence at a challenge depends on whether she engages in the challenge alone or with a cooperating partner. So, we let the mutant's competence at challenge of type j be $c_j = c_j(G_j(x_k, y_k))$, where $G_j(x_k, y_k)$ is a production function describing how the skills of the cooperating partners interact (e.g., in an additive, multiplicative, or submultiplicative way).

Using our assumption of domain-general skills, we let the competence function be independent of the challenge type, $c_j(G_j(x_k, y_k)) = c(G_j(x_k, y_k))$, although its argument depends on the challenge type. We con-

sider two forms for the competence function:

$$c(G_j(x_k, y_k)) = \begin{cases} G_j^\gamma(x_k, y_k) & \text{with power competence} \\ \exp[G_j(x_k, y_k)]^\gamma & \text{with exponential competence,} \end{cases} \tag{10a}$$

where γ measures the effectiveness (decidability) of skills at the challenge. The production function $G_j(x_k, y_k)$ is

$$G_j(x_k, y_k) = \begin{cases} x_k & \text{for } j \in \{1,3\} \\ x_k + y_k & \text{for } j \in \{2,4\} \text{ with additive cooperation} \\ x_k y_k & \text{for } j \in \{2,4\} \text{ with multiplicative cooperation} \\ \sqrt{x_k y_k} & \text{for } j \in \{2,4\} \text{ with sub-multiplicative cooperation} . \end{cases} \tag{10b}$$

The difficulty d_j of a challenge depends on the challenge type. For an ecological or a cooperative challenge, the challenge difficulty is α which depends on the "ecological" environment, which we assume constant (this assumption can be relaxed in future extensions; see section 9). In turn, the difficulty of a competitive challenge depends on the skill of the individual's competitors. Since the mutant is rare, a mutant's competitors are residents, so the difficulty of a competitive challenge is the competence of the resident, $c(G_j(y_k, y_k))$. In general, the difficulty of a type-j challenge is

$$d_j(y_k) = \begin{cases} \alpha & \text{for } j \in \{1,2\} \\ c(G_j(y_k, y_k)) & \text{for } j \in \{3,4\}. \end{cases} \tag{10c}$$

In turn, we let the energy extraction efficiency from maternal provisioning when the individual is of age t be

$$\varphi(t) = \varphi_0 \exp(-\varphi_r t), \tag{10d}$$

where φ_0 is the energy extraction efficiency from maternal provisioning at birth and φ_r measures the rate of decrease of maternal provisioning.

2.3 Energy extraction efficiency given each challenge type

In this section, we write explicitly the expressions for the energy extraction efficiency given a challenge, e_j. Since we consider two forms for the competence function (power or exponential) and three forms for cooperation (additive, multiplicative, or submultiplicative), we have six cases for the energy extraction efficiency. Using Eqs. (8), (6), and (10), we obtain the following expressions for e_j.

2.3.1 Power competence with additive cooperation (PC-AC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{x_k^\gamma(t) + \alpha\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + \alpha} \tag{11a}$$

$$e_2(t, x_k, y_k) = \frac{[x_k(t) + y_k(t)]^\gamma + \alpha\varphi_0 \exp(-\varphi_r t)}{[x_k(t) + y_k(t)]^\gamma + \alpha} \tag{11b}$$

$$e_3(t, x_k, y_k) = \frac{x_k^\gamma(t) + y_k^\gamma(t)\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + y_k^\gamma(t)} \tag{11c}$$

$$e_4(t, x_k, y_k) = \frac{[x_k(t) + y_k(t)]^\gamma + [y_k(t) + y_k(t)]^\gamma \varphi_0 \exp(-\varphi_r t)}{[x_k(t) + y_k(t)]^\gamma + [y_k(t) + y_k(t)]^\gamma}. \tag{11d}$$

2.3.2 Power competence with multiplicative cooperation (PC-MC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{x_k^\gamma(t) + \alpha\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + \alpha} \tag{12a}$$

$$e_2(t, x_k, y_k) = \frac{[x_k(t) y_k(t)]^\gamma + \alpha\varphi_0 \exp(-\varphi_r t)}{[x_k(t) y_k(t)]^\gamma + \alpha} \tag{12b}$$

$$e_3(t, x_k, y_k) = \frac{x_k^\gamma(t) + y_k^\gamma(t)\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + y_k^\gamma(t)} \tag{12c}$$

$$
\begin{aligned}
e_4(t, x_k, y_k) &= \frac{[x_k(t) y_k(t)]^\gamma + [y_k(t) y_k(t)]^\gamma \varphi_0 \exp(-\varphi_r t)}{[x_k(t) y_k(t)]^\gamma + [y_k(t) y_k(t)]^\gamma} \\
&= \frac{x_k^\gamma(t) + y_k^\gamma(t)\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + y_k^\gamma(t)} \\
&= e_3(t, x_k, y_k). \tag{12d}
\end{aligned}
$$

2.3.3 Power competence with sub-multiplicative cooperation (PC-SC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{x_k^\gamma(t) + \alpha\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + \alpha} \tag{13a}$$

$$e_2(t, x_k, y_k) = \frac{[x_k(t) y_k(t)]^{\gamma/2} + \alpha\varphi_0 \exp(-\varphi_r t)}{[x_k(t) y_k(t)]^{\gamma/2} + \alpha} \tag{13b}$$

$$e_3(t, x_k, y_k) = \frac{x_k^\gamma(t) + y_k^\gamma(t)\varphi_0 \exp(-\varphi_r t)}{x_k^\gamma(t) + y_k^\gamma(t)} \tag{13c}$$

$$
\begin{aligned}
e_4(t, x_k, y_k) &= \frac{[x_k(t) y_k(t)]^{\gamma/2} + [y_k(t) y_k(t)]^{\gamma/2}\varphi_0 \exp(-\varphi_r t)}{[x_k(t) y_k(t)]^{\gamma/2} + [y_k(t) y_k(t)]^{\gamma/2}} \\
&= \frac{x_k^{\gamma/2}(t) + y_k^{\gamma/2}(t)\varphi_0 \exp(-\varphi_r t)}{x_k^{\gamma/2}(t) + y_k^{\gamma/2}(t)}. \tag{13d}
\end{aligned}
$$

2.3.4 Exponential competence with additive cooperation (EC-AC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{\exp[\gamma x_k(t)] + \alpha \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \alpha} \tag{14a}$$

$$e_2(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) + y_k(t)]\} + \alpha \varphi_0 \exp(-\varphi_r t)}{\exp\{\gamma[x_k(t) + y_k(t)]\} + \alpha} \tag{14b}$$

$$e_3(t, x_k, y_k) = \frac{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)] \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)]} \tag{14c}$$

$$e_4(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) + y_k(t)]\} + \exp\{\gamma[y_k(t) + y_k(t)]\} \varphi_0 \exp(-\varphi_r t)}{\exp\{\gamma[x_k(t) + y_k(t)]\} + \exp\{\gamma[y_k(t) + y_k(t)]\}}$$

$$= \frac{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)] \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)]}$$

$$= e_3(t, x_k, y_k). \tag{14d}$$

2.3.5 Exponential competence with multiplicative cooperation (EC-MC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{\exp[\gamma x_k(t)] + \alpha \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \alpha} \tag{15a}$$

$$e_2(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) y_k(t)]\} + \alpha \varphi_0 \exp(-\varphi_r t)}{\alpha + \exp\{\gamma[x_k(t) y_k(t)]\}} \tag{15b}$$

$$e_3(t, x_k, y_k) = \frac{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)] \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)]} \tag{15c}$$

$$e_4(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) y_k(t)]\} + \exp\{\gamma[y_k(t) y_k(t)]\} \varphi_0 \exp(-\varphi_r t)}{\exp\{\gamma[x_k(t) y_k(t)]\} + \exp\{\gamma[y_k(t) y_k(t)]\}}. \tag{15d}$$

2.3.6 Exponential competence with sub-multiplicative cooperation (EC-SC)

For this case, the energy extraction efficiency for each challenge type is

$$e_1(t, x_k) = \frac{\exp[\gamma x_k(t)] + \alpha \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \alpha} \tag{16a}$$

$$e_2(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) y_k(t)]^{1/2}\} + \alpha \varphi_0 \exp(-\varphi_r t)}{\alpha + \exp\{\gamma[x_k(t) y_k(t)]^{1/2}\}} \tag{16b}$$

$$e_3(t, x_k, y_k) = \frac{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)] \varphi_0 \exp(-\varphi_r t)}{\exp[\gamma x_k(t)] + \exp[\gamma y_k(t)]} \tag{16c}$$

$$e_4(t, x_k, y_k) = \frac{\exp\{\gamma[x_k(t) y_k(t)]^{1/2}\} + \exp\{\gamma[y_k(t) y_k(t)]^{1/2}\} \varphi_0 \exp(-\varphi_r t)}{\exp\{\gamma[x_k(t) y_k(t)]^{1/2}\} + \exp\{\gamma[y_k(t) y_k(t)]^{1/2}\}}. \tag{16d}$$

3 Evolutionary differential game

Here we state the evolutionary differential game problem[11] posed by Eq. (1), which we solve to obtain an uninvadable growth strategy. The problem is a "game" between mutant and resident because the mutant's payoff $[R_0(u, v)]$ depends on the resident strategy, it is "differential" because the problem depends on differential equations, and "evolutionary" rather than a typical differential game because only the mutant strategy is subject to maximization so that an evolutionary equilibrium is attained when given a resident the "best mutants" coincide with the resident. As stated above, we find such strategy by iterating optimal control problems until convergence to a strategy that is a best response to itself. The definition of quantities is summarized in Supplementary Table. 1.

3.1 Optimal control problem (OCP) to iterate

From our assumptions above and in previous work[1], the expected lifetime number of offspring of a mutant $u = \{\mathbf{u}(t)\}_{t=0}^{T}$ in the context of resident $v = \{\mathbf{v}(t)\}_{t=0}^{T}$ is

$$R_0(u, v) = C(v) J(u, v), \tag{17}$$

for some density-dependent scalar $C(v) > 0$ and objective functional $J(u, v)$. In the terminology of optimal control theory, we seek a mutant's control (growth strategy)

$$\mathbf{u}(t) = (u_{\mathrm{b}}(t), u_{\mathrm{r}}(t), u_{\mathrm{s}}(t)) \quad \text{for all} \ \ t \in [0, T], \tag{18a}$$

given the resident's control

$$\mathbf{v}(t) = (v_{\mathrm{b}}(t), v_{\mathrm{r}}(t), v_{\mathrm{s}}(t)) \quad \text{for all} \ \ t \in [0, T], \tag{18b}$$

so as to maximize the mutant's objective functional

$$J(u, v) = f_0 \int_0^T \exp(-\mu t) x_{\mathrm{r}}(t) \mathrm{d}t, \tag{18c}$$

subject to the control constraint

$$\mathbf{u}(t) \in [0, 1]^3 \quad \text{where} \ \ u_{\mathrm{b}}(t) + u_{\mathrm{r}}(t) + u_{\mathrm{s}}(t) = 1, \ \ \text{for all} \ \ t \in [0, T], \tag{18d}$$

such that the mutant's state variable

$$\mathbf{x}(t) = (x_{\mathrm{b}}(t), x_{\mathrm{r}}(t), x_{\mathrm{s}}(t), x_{\mathrm{k}}(t)) \geq 0, \ \ \text{for all} \ \ t \in [0, T] \tag{18e}$$

given the resident's state variable

$$\mathbf{y}(t) = (y_{\mathrm{b}}(t), y_{\mathrm{r}}(t), y_{\mathrm{s}}(t), y_{\mathrm{k}}(t)) \geq 0, \ \ \text{for all} \ \ t \in [0, T] \tag{18f}$$

satisfies the dynamic constraints

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{u}, \mathbf{x}, y_{\mathrm{k}}) \tag{18g}$$

with

$$g_i(t, \mathbf{u}, \mathbf{x}, y_{\mathrm{k}}) = a_i u_i B_{\mathrm{syn}}(t, \mathbf{x}, y_{\mathrm{k}}) \quad \text{for} \ i \in \{\mathrm{b}, \mathrm{r}, \mathrm{s}\} \tag{18h}$$

$$g_{\mathrm{k}}(t, \mathbf{u}, \mathbf{x}, y_{\mathrm{k}}) = b_1 \left[x_{\mathrm{b}} B_{\mathrm{b}} + u_{\mathrm{b}} B_{\mathrm{syn}}(t, \mathbf{x}, y_{\mathrm{k}}) \right] - b_2 x_{\mathrm{k}}, \tag{18i}$$

where $a_i = 1/E_i$, $b_1 = s_k/E_k$, and $b_2 = B_k/E_k$. The growth metabolic rate is

$$B_{\text{syn}}(t, \mathbf{x}, y_k) = Ke(t, x_k, y_k)x_B^\beta - B_b x_b - B_r x_r - B_s x_s, \tag{18j}$$

where the mutant's body mass is

$$x_B = x_b + x_r + x_s, \tag{18k}$$

and the mutant's expected energy extraction efficiency at age t is

$$e(t, x_k, y_k) = \sum_{j \in \mathbb{C}} P_j e_j(t, x_k, y_k), \tag{18l}$$

where $\mathbb{C} = \{1, 2, 3, 4\}$ and $\sum_{j \in \mathbb{C}} P_j = 1$ for all $t \in [0, T)$. The energy extraction efficiency e_j for each type of challenge and all cases considered are given by Eqs. (11)-(16). Finally, the initial conditions of (18g) are

$$x_i(0) = x_{i0} \quad \text{for all } i \in \{b, r, s, k\} \tag{18m}$$

and we do not consider any terminal conditions for (18g). Since $\mathbf{v}(t)$ and $\mathbf{y}(t)$ are given for all $t \in [0, T)$, they constitute dynamic parameters (i.e., they are given quantities that vary with t).

3.2   Iteration of OCP to obtain uninvadable strategy

As explained above, to facilitate convergence to a strategy that is a best response to itself, we impose the additional path constraint that an optimal control $\widetilde{\mathbf{u}}^*$ solving problem (18) must also satisfy $|\widetilde{\mathbf{u}}^*(t) - \mathbf{v}(t)| < \delta$ for all $t \in [0, T)$ and for some $\delta > 0$; this constraint is known as damping[7] and is an assumption of δ-weak selection[8]. Once a solution $(\widetilde{\mathbf{u}}^*, \widetilde{\mathbf{x}}^*)$ for the optimal control problem (18) is found, this solution $(\widetilde{\mathbf{u}}^*, \widetilde{\mathbf{x}}^*)$ is set as the resident $(\mathbf{v}, \mathbf{y})$ and the problem is solved again. Ideally, this procedure is iterated until the optimal control $\widetilde{\mathbf{u}}^*$ of the current iteration satisfies $|\widetilde{\mathbf{u}}^*(t) - \mathbf{v}(t)| < \epsilon$ for all $t \in [0, T)$ and for some ϵ such that $\delta > \epsilon > 0$ so that such $\widetilde{\mathbf{u}}^* = \mathbf{u}^*$ is a best response to itself to ϵ degree. However, in numerical solutions, controls sustain numerical jitter that prevents the application of this ϵ criterion, so convergence is declared by visual comparison of $\widetilde{\mathbf{u}}^*$ and $\mathbf{v}$ (for an illustration, see Extended Data Fig. 2a,f). Note that the tilde in $(\widetilde{\mathbf{u}}^*, \widetilde{\mathbf{x}}^*)$ denotes the optimal control and associated optimal state for a given iteration, while $(\mathbf{u}^*, \mathbf{x}^*)$ denotes the best response to itself solving problem (1) to ϵ degree.

	Control variables	
$u_i(t)$	Mutant's fraction of growth metabolic rate allocated to tissue i at age t with $i \in \{b, r, s\}$ (for brain, reproductive, and somatic tissue, respectively)	
	State variables	
$x_i(t)$	Mutant's mass of of tissue i at age t with $i \in \{b, r, s\}$ or skill level at age t with $i \in \{k\}$	
	Switching times	
t_l	Ages† of brain growth onset, brain growth arrest, maturity, and adulthood with $l \in \{b0, b, m, a\}$	
	Dynamic parameters	
$v_i(t)$	Resident's fraction of growth metabolic rate allocated to tissue i at age t with $i \in \{b, r, s\}$	
$y_k(t)$	Resident's skill level at age t	
	Parameters	
x_{i0}	Mass of tissue i at birth with $i \in \{b, r, s\}$ or skill level at birth with $i \in \{k\}$	
B_i	Rate of heat release by an average mass unit of tissue i of the resting individual due to tissue maintenance with $i \in \{b, r, s\}$ (i.e., each tissue's mass-specific maintenance metabolic rate); or rate of heat release by the resting individual's brain for maintenance of an average skill unit with $i \in \{k\}$ (i.e., skill-specific memory metabolic rate)	
E_i	Rate of heat release by the resting individual due to the production of an average mass unit of tissue i with $i \in \{b, r, s\}$ (i.e., each tissue's mass-specific growth metabolic rate); or rate of heat release by the resting individual's brain for production of an average skill unit with $i \in \{k\}$ (i.e., skill-specific learning metabolic rate)	
K	Resting metabolic rate with a body mass of one unit and full energy extraction efficiency	
β	Coefficient scaling resting metabolic rate with body mass under full energy extraction efficiency	
μ	Mortality rate for individuals of age t	
T	Age of death or of reproductive cessation	
s_k	Fraction of the resting individual's brain metabolic rate allocated to energy extraction skills	
α	Difficulty of environmental (non-social) challenges	
γ	Effectiveness (decidability) of skills at overcoming challenges	
φ_0	Energy extraction efficiency of the newborn from maternal provisioning	
φ_r	Rate at which maternal provisioning decreases as the individual grows	
P_j	Proportion of energy-extraction challenges of type j faced during any $[t, t + \Delta t]$ with $j \in \{1, 2, 3, 4\}$ and $\sum_j P_j = 1$ as Δt tends to zero	
f_0	Number of offspring produced per unit time and per mass unit of reproductive tissue without density dependence	

Supplementary Table 1: Summary of definitions. †The switching times are an output of the model.

4 Parameter values

Here we summarize the parameter values used. We classify parameters into $\mathbf{P}$ (4 of them), $\mathbf{Q}$ (13 of them), and $\mathbf{R}$ (9 of them) parameters. $\mathbf{P}$-parameters control challenge proportion and we exhaustively explore their values to study their effects. In section 4.1, we give the values for the $\mathbf{Q}$-parameters used, which measure newborn mass, tissue metabolism, and demography and were previously[1] estimated for human females using published data; in particular, $\mathbf{Q}$-parameters include estimates for brain metabolic costs. In section 4.2, we list the values of the $\mathbf{R}$-parameters used, which measure skill metabolism, energy extraction, maternal provisioning, and mutation effect and we use reasonable values given the available data; in particular, $\mathbf{R}$-parameters include the metabolic cost of memory, whose values fall within a previously empirically estimated range for stored motor patterns in cerebellum Purkinje cells in rats. In section 4.3 we discuss the estimates of a subset of the $\mathbf{Q}$-parameters, specifically the metabolic maintanence costs, B_i. In Extended Data Fig. 3g,h we show that the values used for the $\mathbf{R}$-parameters yield a high ontogenetic fit with *H. sapiens* while generating adult body and brain sizes of ancient human scale in the purely ecological scenario[1]. The values used for $\mathbf{R}$ involve intermediate difficulty of environmental challenges, moderately effective skills, and metabolically expensive memory (although in the low end of the empirically estimated range). The values used here for $\mathbf{Q}$ and $\mathbf{R}$ are the same as those used previously[1], except for φ_0, which is adjusted given our improved implementation of maternal provisioning.

Since the parameter f_0 only displaces the objective J vertically and thus has no effect on the solution, we choose its value to scale the objective and facilitate numerical solutions (Supplementary Table 2). The $\mathbf{R}$-parameter values are different between the cases of power (Supplementary Table 3) and exponential competence (Supplementary Table 4). For power competence, the units of α are different when cooperation is multiplicative when compared to the other cases; in contrast, for exponential competence, the units of γ are different if cooperation is multiplicative (keeping track of units is important to properly rescale state variables for numerical solutions; see section 5).

4.1 Q-parameter values

The values for $\mathbf{Q}$-parameters are listed in Supplementary Table 2.

Newborn mass			Tissue metabolism				Demography	
		K	$132.7281\,\frac{\text{MJ}}{\text{y}}\text{kg}^{-\beta}$	β	0.7378			
$x_{\text{s}0}$	$2.0628\ \text{kg}$	B_{s}	$29.6891\,\frac{\text{MJ}}{\text{kg}\times\text{y}}$	E_{s}	$12.4594\,\frac{\text{MJ}}{\text{kg}}$	f_0	$0.5\,\frac{\#\text{offspring}}{\text{kg}\times\text{y}}$	
$x_{\text{b}0}$	$0.3372\ \text{kg}$	B_{b}	$313.0962\,\frac{\text{MJ}}{\text{kg}\times\text{y}}$	E_{b}	$123.7584\,\frac{\text{MJ}}{\text{kg}}$	μ	$0.034\,\frac{1}{\text{y}}$	
$x_{\text{r}0}$	$0\ \text{kg}$	B_{r}	$2697.1179\,\frac{\text{MJ}}{\text{kg}\times\text{y}}$	E_{r}	$190.8196\,\frac{\text{MJ}}{\text{kg}}$	T	$47\ \text{y}$	

Supplementary Table 2: $\mathbf{Q}$-parameter values. Estimated parameter values and f_0, which is set to an arbitrary value. In section 4.3 we confirm the value of B_{b} with three independent data sets, check that the value of B_{s} remains after a minor correction, and explain that there is uncertainty in the value of B_{r} but show that this has virtually no effect on predicted adult brain and body mass provided B_{r} is at least within 70 and 2700 MJ/(kg y).

4.2 R-parameter values

The **R**-parameter values are listed in Supplementary Table 3 for the power competence case and in Supplementary Table 4 for the exponential competence case. The specific values used yield a high ontogenetic fit with *H. sapiens* as shown in Extended Data Fig. 3g,h.

The value used for φ_r yields maternal provisioning for ≈ 20 years, which is similar to the age where negative caloric production ends in Tsimane forager-horticulturalists[12].

4.2.1 For power competence

Skill metabolism		Energy extraction		Maternal provisioning		Mutation effect	
s_k	0.5	α	$1\,\mathrm{TB}^\gamma$†	φ_0	**0.4**	δ	0.1
B_k	$36\frac{\mathrm{MJ}}{\mathrm{TB}\times\mathrm{y}}$	γ	1.4	φ_r	$0.2\frac{1}{\mathrm{y}}$		
E_k	$370\frac{\mathrm{MJ}}{\mathrm{TB}}$	x_{k0}	$1\,\mathrm{TB}$				

Supplementary Table 3: **R**-parameter values with power competence. The value of φ_0 (0.4, in bold face) is smaller than that used previously[1] (0.6) because of our improved implementation of maternal provisioning φ; as shown in Extended Data Fig. 2g, this change has virtually no effect on the resulting growth strategy as well as on brain and body sizes but allows to incorporate maternal provisioning for the various challenge types. †For the expression of e_2, the unit is $\mathrm{TB}^{2\gamma}$ with multiplicative cooperation but TB^γ with other forms of cooperation.

4.2.2 For exponential competence

Skill metabolism		Energy extraction		Maternal provisioning		Mutation effect	
s_k	0.5	α	1.15	φ_0	**0.6**	δ	0.1
B_k	$50\frac{\mathrm{MJ}}{\mathrm{TB}\times\mathrm{y}}$	γ	$0.6\,\mathrm{TB}^{-1}$†	φ_r	$0.2\frac{1}{\mathrm{y}}$		
E_k	$250\frac{\mathrm{MJ}}{\mathrm{TB}}$	x_{k0}	$0\,\mathrm{TB}$				

Supplementary Table 4: **R**-parameter values with exponential competence. The value of φ_0 (0.6, in bold face) is smaller than that used previously[1] (0.8) because of our improved implementation of maternal provisioning φ; as shown in Extended Data Fig. 2g, this change has virtually no effect on the resulting growth strategy. †For the expressions of e_2 and e_4, the unit is TB^{-2} with multiplicative cooperation but TB^{-1} with the other forms of cooperation.

4.3 Discussion of estimates of the maintenance costs (B_i for $i \in \{b, r, s, k\}$)

In Extended Data Fig. 3a,b we show that the predicted adult mass of tissue i tends to decrease with the mass-specific maintenance cost of tissue i, B_i, and similarly the adult skill level decreases with the memory cost B_k. Thus, the predicted adult values of tissue mass and skill level depend on the accuracy of the corresponding B_i. In this section we discuss the estimates of B_i.

As discussed previously[1], we assume as a first approximation that B_i and E_i are constant with respect to age, although in reality they may vary with age as is known to be the case for the mass-specific maintenance cost of brain tissue[13]. So, our approach is to estimate these parameter values around the ages where the parameter is expected to have the strongest effects on growth dynamics. From the shape of the ontogenetic dynamic equations (Eqs. 1 and 2 in Methods), the growth dynamics are more likely to be driven by maintenance costs B_i later in life and by acquisition costs E_i at points in life where the tissue is growing the fastest[14]. So, to obtain the values of B_i and E_i at the ages that are presumably most affected by them, González-Forero et al.[1] estimated B_i from data for adults, E_b from data for newborns, E_r from data for fifteen year old females, and E_s for newborns.

B_b: Here we corroborate the value of B_b with three independent data sets. The value of B_b (Supplementary Table 2) was previously[1] calculated using the ratio of glucose uptake by the brain per unit time in adult human females (rate of glucose uptake by the brain divided by that of the body) as reported by Kuzawa et al.[15] who used data from Chugani et al.[16] and González-Forero et al.[1] corrected for non-oxidative glucose metabolism using data in Goyal et al.[17] [hence, the sexed data in Kuzawa et al.[15] yields an estimate of B_b of 313.0962 for females and of 312.7776 MJ/(kg y) for males (sample size: 2 females and 7 males; no confidence intervals provided)]. A second estimate for mixed sex data is the following. Herculano-Houzel[18] used estimates from data in Karbowski[19], who in turn used the data in Clarke and Sokoloff[20]. Clarke and Sokoloff[13] (p. 650), who used data compiled by Sokoloff[21] for mixed sexes from earlier (1940-1950's) studies, reported an oxygen consumption of 3.5 ml/min per 100 g of brain tissue in a normal young adult human (no sample size or confidence intervals provided). Oxygen consumption can be transformed[22] into power units by multiplying by 20.1 kJ/l O_2. Hence, with these values, the mass-specific cost of brain maintenance is $3.5 \times \frac{\text{ml}\,O_2}{\text{min}\,100\,\text{g}} \times \frac{1\,\text{l}}{1000\,\text{ml}} \times 20.1 \frac{\text{kJ}}{\text{l}\,O_2} \times \frac{1\,\text{MJ}}{1000\,\text{kJ}} \times \frac{1\,\text{l}}{1000\,\text{ml}} \times \frac{60\,\text{min}}{1\,\text{h}} \times \frac{24\,\text{h}}{1\,\text{d}} \times \frac{365\,\text{d}}{1\,\text{y}} \times \frac{1000\,\text{g}}{1\,\text{kg}} = 369.7596 \frac{\text{MJ}}{\text{kg} \times \text{y}}$, which approximates the value of B_b estimated before[1] (Supplementary Table 2). Most recently, a third data set with mixed sexes yields a similar estimate. On page 202, McKenna et al.[23], who used data from Madsen et al.[24] for sexes combined, reported an oxygen consumption for whole brain in humans of 1.425 ± 0.189 μmol/(g min) (mean $\pm$ SD; sample size: 8). In ideal gas conditions, a liter of oxygen has 22.7 moles, and hence this results in 341.7371 ± 45.3251 MJ/(kg y), which includes the two estimates above. The discrepancies in these three estimates of B_b yield a difference of up to ≈ 250 g in the predicted adult brain size in the purely ecological scenario with exponential competence (see legend of Extended Data Figs. 3c,d). As the ontogenetic data in Kuzawa et al.[15] allowed González-Forero et al.[1] to estimate most of the **Q** parameters [$x_b(0)$, $x_s(0)$, B_b, B_s, E_b, E_r, E_s, K, and β], to preserve data homogeneity in our parameter estimates, which is important to obtain a meaningful model outcome, we continue to use the estimate of B_b based on the data of Kuzawa et al.[15] (for a similar aim of preserving data homogeneity see van der Veer et al.[25]). To be sure, the discrepancy between the estimate for

B_b that we use [313.0962 MJ/(kg y)] [15] and the most recent estimate [341.7371 MJ/(kg y)] [24] yields a difference of up to $\approx$ 140 g in the predicted adult brain size in the purely ecological scenario with exponential competence (see legend of Extended Data Figs. 3c,d). This discrepancy approaches the range of variation of $\approx$ 130 g in healthy adult female brain size of the original data set for adult brain size we use [15,26], suggesting the choice of the value we use for B_b does not substantially increase our prediction error.

B_r: Here we explain that there is uncertainty in the value of B_r and justify the value we use. In short, there is uncertainty in this value because we are unaware of data that allows to estimate it for preovulatory ovarian follicles, but the value we use yields predictions that closely match observed patterns.

We identify reproductive tissue with preovulatory ovarian follicles for three reasons. First, with this restricted notion, reproductive tissue is not involved in offspring maintenance and thus allows for a simpler expression of fecundity [1] (see section 1). Second, identifying reproductive tissue with the ovarian follicle rather than the oocyte takes into account that much of the oocyte-related metabolism is carried out by other cells in the follicle, and so to capture the maintenance and production cost of the oocyte, the metabolism of these accompanying cells should in principle be considered [27,28]. Indeed, the follicle is regarded as the functional unit of the ovary [27]. Third, we identify reproductive tissue with preovulatory follicles rather than follicles at earlier stages of development to account for the fact that most follicles at earlier stages are lost and do not contribute to fertility, while preovulatory follicles directly do so [27,28]. Therefore, the parameter B_r would ideally be estimated for preovulatory ovarian follicles.

However, we have been unable to find reports on the metabolic rate of preovulatory follicles. To gain an idea of the value of B_r, we previously [1] used the metabolic rate of oocytes as a proxy, which thus lacks information on the overall maintenance cost of reproductive tissue as we define it. Hence, González-Forero et al. [1] used the oxygen consumption of a human oocyte (in mixed developmental stages) estimated [29] to be 0.53×10^{-9} l O_2/(h oocyte), and assumed that the dry mass of a human oocyte is similar to that of a mouse oocyte. This introduced two additional sources of error in the estimate of B_r: first, the volume of a human oocyte is an order of magnitude larger [30] than that of a mouse oocyte [31]; and second, by using dry mass, the predicted mass of reproductive tissue is that of dry mass rather than overall mass, which is the relevant one for the model (particularly because we take body mass to be $x_B = x_b + x_r + x_s$).

Correcting these two sources of error in the calculation of B_r, although still using the metabolic rate of oocytes, the estimate of the maintenance cost of reproductive tissue is the following. Since oocytes in preovulatory follicles become arrested in metaphase II, it would be more suitable for our purposes to use the metabolic rate of metaphase-II oocytes (which may have low metabolic rate because of their arrested state [32]). The oxygen consumption for human oocytes at stages between metaphase I and II has been estimated [32] to be 0.573×10^{-9} l O_2/(h oocyte) (corresponding estimates for exclusively metaphase-II oocytes in the same study were suggested to be artificially low as these oocytes were preserved for one day in vitro after extraction [32]). As oxidative metabolism takes place inside rather than outside the cell membrane, it is likely more appropriate to divide this oocyte metabolic rate by the mass of the ooplasm (i.e., the cytoplasm of the oocyte) and intracellular structures rather than the mass of the whole oocyte (which includes external structures such as the periviteline space and zona pelucida). The volume of the ooplasm of a metaphase-II human oocyte has been estimated [30]

to be $1.43 \times 10^6 \mu\text{m}^3$. This volume was corroborated by later studies[33] which also find that the diameter of the ooplasm does not change between metaphase I and II. Taking this volume as being of water density, these values yield the estimate: $B_r = 0.573 \times 10^{-9} \frac{1\,O_2}{\text{h} \times \text{oocyte}} \times 20.1 \frac{\text{kJ}}{1\,O_2} \times \frac{1\,\text{oocyte}}{1430\,\text{ng}} \times \frac{24\,\text{h}}{1\,\text{d}} \times \frac{365\,\text{d}}{1\,\text{y}} \times \frac{10^9\,\text{ng}}{1\,\text{g}} \times \frac{1000\,\text{g}}{1\,\text{kg}} \times \frac{1\,\text{MJ}}{1000\,\text{kJ}} = 70.5535$ MJ/(kg year), which is two orders of magnitude smaller than the previous estimate[1] [Supplementary Table 2; alternatively, if rather than using the volume of only the ooplasm, one uses the volume of the whole oocyte, which is estimated[30] to be $3.52 \times 10^6 \mu\text{m}^3$, the estimate for B_r becomes 28.6624 MJ/(kg year)]. Recall that this estimate of B_r is for oocytes rather than the more desirable estimate for preovulatory follicles.

In Extended Data Figs. 3a–f we show that when B_r falls within 70 and 2700 MJ/(kg y), B_r affects primarily the predictions of adult mass of reproductive tissue but has virtually no effect on adult brain and body mass. If $B_r < 70$ MJ/(kg y), the effects on adult brain and body mass become pronounced (Extended Data Figs. 3e,f). The reason that brain and body mass are largely unaffected when B_r falls within 70 and 2700 MJ/(kg y) is that B_r affects the allocation to reproductive growth during adolescence, which occurs in the life history after all of brain growth has occurred, and after most of body growth has taken place (Extended Data Fig. 3i). Values of B_r approaching the low limit of the interval 70 and 2700 MJ/(kg y) shrink the length of adolescence, yielding a poorer prediction of the human life history than larger values (Extended Data Fig. 3i). Additionally, values of B_r toward the lower end of 70 and 2700 MJ/(kg y) yield a very large adult mass of reproductive tissue, while the observed adult mass of preovulatory follicles is very small (≈ 10 g; see section B_s below). Finally, if $B_r < 70$ MJ/(kg y), we find that the life history is severely disrupted, yielding for example brain growth throughout life and no non-reproductive childhood which is inconsistent with human life history (Extended Data Fig. 3i).

Hence, given the uncertainty in the value of B_r as data to estimate it for preovulatory follicles remains unavailable to our knowledge, the better life history predictions with larger B_r, and the better predictions of adult reprodutive mass with larger B_r, the model predicts that estimates of B_r for preovulatory follicles would fall toward the high end of 70 and 2700 MJ/(kg y). Therefore, given the little effect on adult brain and body mass as B_r changes within 70 and 2700 MJ/(kg y), we use the large value of B_r used previously[1], particularly as there appears to be little reason for it to be of concern for the purposes of this paper as we are concerned with brain and body mass rather than reproductive mass.

B_s: Here we confirm the value of B_s after a minor correction. Since the previous[1] estimate of B_s considered the dry mass of reproductive tissue, a minor correction for the calculation of B_s is also in order although its value is virtually unchanged. Adult human females have on average about 2 preovulatory follicles at any given age[34]. A preovulatory follicle has an average diameter of 21.1 mm[35]. Denote by $X_i(\tau_a)$ the observed mass of tissue i (or of body mass if $i = B$) at the observed age at adulthood τ_a [so[15], $X_B(\tau_a) = 51.1$ kg and $X_b(\tau_a) = 1.31$ kg]. Approximating the follicle mass by the mass of a spherical cell of water with such diameter, then the adult mass of reproductive tissue is $X_r(\tau_a) = 2$ follicles $\times \frac{4}{3}\pi \left(\frac{21.1\,\text{mm}}{2}\right)^3 \times \frac{1\,\text{kg}}{10^6\,\text{mm}^3\text{H}_2\text{O}} = 9.8373 \times 10^{-3}$ kg rather than the previously[1] estimated value of 3.9349×10^{-3} kg. Hence, $X_s(\tau_a) = X_B(\tau_a) - X_b(\tau_a) - X_r(\tau_a) = 49.7802$ kg which is virtually the same to the previously[1] estimated 49.7861 kg.

Since at human adulthood there is no growth, it must be the case that the adult observed metabolic rate is $B_{\text{rest}}(\tau_a) = B_{\text{maint}}(\tau_a) = \sum_{i \in \{b,r,s\}} X_i(\tau_a) B_i$. Because we have[15] that $B_{\text{rest}}(\tau_a) = 1898.8707$ MJ/y, it follows that $B_s = [B_{\text{rest}}(\tau_a) - B_b X_b(\tau_a) - B_r X_r(\tau_a)]/X_s(\tau_a) = 29.8918$ MJ/(kg y) which is also virtually the same to the previ-

ously[1] estimated 29.6891 MJ/(kg y).

B_k: Here we describe estimates of the metabolic memory cost B_k and a rough calculation of information storage capacity in the human neocortex.

We use values for B_k that fall within an estimated range for stored motor patterns in cerebellum Purkinje cells of rats. Specifically, Purkinje cells in the rat cerebellum are estimated to have a resting energy consumption of between 1 and 400 mW/GB of stored motor patterns[36]. Changing units, this range equals 32 to 12614 MJ/(TB y), which includes the values we use for B_k. The value used for the metabolic learning cost E_k is of a similar order of magnitude but empirical estimates of the learning cost remain unavailable to our knowledge.

As the value for the metabolic cost of memory B_k is in information units, the model makes predictions for skill level in such units. In particular, the resulting predicted skill level for the best fitting scenario for *H. sapiens* in Fig. 4b in the main text is $x_k^*(\tau_a) = 3.92$ TB (Extended Data Fig. 10e). For comparison, we contrast this value with available, preliminary estimates of the human brain's information storage capacity. Neuropil in the rat hippocampus is estimated to sustain 4.7 bits of information per synapse[37] and the human neocortex is estimated to have 0.15 quadrillion synapses[38], which would very roughly suggest 587.5 TB of storage in the human neocortex.

5 Numerical implementation

We obtained approximate numerical solutions using the software GPOPS 2.3[39] in MATLAB R2015b in a 32-CPU Linux machine running Debian 8.9 (Jessie). GPOPS is based on a pseudospectral method which converts the optimal control problem into a finite-dimensional nonlinear program. GPOPS adapts the underlying time mesh (partitions of the optimization horizon, or time interval) until the error tolerance is met or until the maximum number of iterations is reached. We discarded solutions that did not meet the error tolerance or where the solver failed [e.g., stalled in a zero objective (i.e., a minimum)].

The GPOPS setup implemented previously[1] produced unstable costate estimation in the newer GPOPS release used here. Hence, we used the following GPOPS setup:

```
mesh.method = 'hp-LiuRao';

mesh.tolerance = 1e-6;

mesh.maxiterations = 45;

setup.method = 'RPM-Differentiation';
```

To facilitate convergence of dynamic optimization algorithms, non-negative state variables should be scaled so that they fall roughly between 0 and 1. We rescaled somatic mass to Mg (megagrams; 10^6 grams), brain mass to hundreds of kg (10^5 grams), and reproductive mass to tens of kg (10^4 grams). We rescaled skill units to tens of TB (10^1 terabytes) for the cases with power competence, and did not rescale skill units (10^0 terabytes) for the cases with exponential competence. These scalings were found by trial and error and were chosen because they prevent solver failure and minimize running time.

For each of the six cases considered (PC-AC, PC-MC, PC-SC, EC-AC, EC-MC, and EC-SC), we performed a parameter sweep taking $P_j \in \{0, 0.1, 0.2, 0.3, \ldots, 0.9, 1\}$ for all $j \in \mathbb{C}$. Since $\sum_{j \in \mathbb{C}} P_j = 1$, there are only three free parameters which we took as being P_1, P_2, P_3, and so $P_4 = 1 - P_1 - P_2 - P_3$. To allow the solver to find an optimal solution, it is important to feed it with an initial guess that is sufficiently close to the solution. Hence, we implemented the parameter sweep so that the solution for one parameter combination was used as both the initial guess and the resident for "next" parameter combinations as explained as follows.

Specifically, each of the six cases involved: an initialization and the parameter sweep, as follows.

Initialization

1. Solve the purely ecological optimal control problem of previous work[1] with the new GPOPS settings, new scalings, and the initial guess given by:

$$x_b(0) = x_{b0} \quad x_b(T) = 1.3 \text{ kg} \tag{19a}$$

$$x_r(0) = x_{r0} \quad x_r(T) = 0.2 \text{ kg} \tag{19b}$$

$$x_s(0) = x_{s0} \quad x_s(T) = 50 \text{ kg} \tag{19c}$$

$$x_k(0) = x_{k0} \quad x_k(T) = 40 \text{ TB} \tag{19d}$$

$$u_b(0) = 0.2 \quad u_b(T) = 0 \tag{19e}$$

$$u_s(0) = 0.8 \quad u_s(T) = 0. \tag{19f}$$

(So, here there is no damping or iterations of optimal control problems to reach a best response to itself).

2. Solve again using the solution of step 1 as the initial guess. This was done as an attempt to obtain a solution with a coarser mesh partition to minimize running time in subsequent runs. (Again, without damping or iterations for a best response to itself).

Parameter sweep

3. Solve the optimal control problem (18), now with damping and iterations for a best response to itself, for the purely ecological setting with the solution of step 2 as initial guess and resident. The parameter combination here is thus $\mathbf{P}_0 = (P_{1,0}, P_{2,0}, P_{3,0}, P_{4,0}) = (1, 0, 0, 0)$.

4. Iterate parameter combinations as follows. For a given $\mathbf{P}_l = (P_{1,l}, P_{2,l}, P_{3,l}, P_{4,l})$, define $\mathbf{P}_{l+1} = (P_{1,l+1}, P_{2,l+1}, P_{3,l+1}, P_{4,l+1})$ such that $P_{1,l+1} = P_{1,l} - 0.1$ and $P_{j,l+1} = P_{j,l} + 0.1$ for only one of the other $j \in \mathbb{C}$ (that is, to obtain $\mathbf{P}_{l+1}$, reduce P_1 in $\mathbf{P}_l$ by 0.1 and increase one of the other P_j by 0.1). Solve (18) for all such $\mathbf{P}_{l+1}$ using the solution for $\mathbf{P}_l$ as initial guess and resident. Stop when $P_{1,l} = 0$.

Since the solution of a run is needed to initialize subsequent ones, runs over combinations of P_j cannot be run all in parallel and so we implemented them partly in series and partly in parallel (i.e., in a tree manner). Solutions were discarded if a best response to itself was not achieved, which happened because of cycling solutions, solver failure, or solver warnings indicating problem infeasibility (i.e., that the problem cannot be solved without violating constraints; Extended Data Fig. 5b,c,e). Warnings of problem infeasibility need not mean that the problem is indeed infeasible, but the solver slows down dramatically. Determining if a problem for a given parameter combination is indeed infeasible is of major difficulty so this was not attempted. To streamline the parameter sweeps, the run of a given parameter combination $\mathbf{P}_l$ was preceded by a test with the GPOPS setting `mesh.maxiterations = 0` so GPOPS only attempts one mesh partition rather than up to 46. The run was discarded if infeasibility warnings arose in the test; otherwise the run proceeded.

When the solution for a parameter combination $\mathbf{P}_l$ is discarded, the solution for $\mathbf{P}_{l+1}$ cannot be initialized as indicated in step 4 and so the parameter space along that trajectory was not explored. Consequently, the parameter sweep over P_j is exhaustive within these practical limits.

Because of Eqs. (12d) and (14d), some parameter combinations are equivalent: for instance, with PC-MC, the parameter combination $(P_1, P_2, P_3, P_4) = (0.9, 0, 0.1, 0)$ is equivalent to $(0.9, 0, 0, 0.1)$. Similarly, some parameter combinations that do not involve cooperation are equivalent between cases: for instance, the parameter combination $(0.9, 0, 0.1, 0)$ with PC-MC is equivalent to $(0.9, 0, 0.1, 0)$ with PC-AC. Hence, for a set of equivalent parameter combinations, we ran only one combination and took its results as being the same for the equivalent combinations.

For a given parameter combination, a typical run converging to a best response to itself took from 2 to 4 days to complete (compared to 1 to 20 minutes in the undamped, uniterated non-social case[1]). The total running time for the parameter sweep over the P_j over the six cases was approximately 2.5 months.

6 Measures of fit

Here we define the quantities we used to measure fit. Fit could be measured in multiple ways (Extended Data Fig. 2h). We measure fit as an "age-wise distance" of the prediction from the observation. By observation we do not mean raw observations but averaged data: specifically, average brain and body mass at a given age. Averaged data are themselves predictions from statistical models that fit the raw data, so the problem of measuring goodness of fit here is different from the inferential statistics problem of inferring averaged data from raw data: we seek to measure the distance of predictions from averaged data rather than the ability of the predictions to describe the raw data. Given this different nature of the problem, rather than using some of the common measures of fit of predictions and raw data (e.g., likelihood and associated quantities), we use the following measures of fit.

Let $X_b(\tau)$ and $X_B(\tau)$ be the observed brain and body mass at the observed age τ, respectively. We seek to measure how far the prediction is from these quantities. We define the normalized residuals at age τ as

$$r_b(\tau) = \frac{x_b^*(\tau) - X_b(\tau)}{X_b(\tau)} \tag{20a}$$

$$r_B(\tau) = \frac{x_B^*(\tau) - X_B(\tau)}{X_B(\tau)}. \tag{20b}$$

We normalize these residuals to prevent body size from obscuring deviations in brain size since body size is typically much larger than brain size. Similarly, we normalize with the values observed at age τ rather than at another age (say, the observed age at adulthood τ_a) to prevent deviations at earlier ages from being obscured by deviations at later ages where values are typically larger.

We measure the deviation of predicted brain and body mass at age τ as the magnitude of the residuals vector $(r_b(\tau), r_B(\tau))$ given by its Euclidean distance to the origin:

$$D(\tau) = \sqrt{r_b^2(\tau) + r_B^2(\tau)}. \tag{21}$$

Thus, a deviation of $D(\tau) = 0$ occurs if and only if there is a perfect fit between prediction and observation in both brain and body mass at age τ. An increasing deviation $D(\tau)$ thus means a poorer fit.

Adult fit: We denote the observed age at adulthood as τ_a, which we take to be $\tau_a = 25$ y. As previously stated, there are multiple ways to measure adult fit (Extended Data Fig. 2h). We measure the adult deviation with $D(\tau_a)$, which corresponds to arrow (2) in Extended Data Fig. 2h. Thus, we measure adult fit as $-D(\tau_a)$. This measure of adult fit is zero if and only if there is a perfect fit between predicted and observed adult brain and body mass at τ_a ($= 25$) y of age. Decreasing (i.e., towards negative infinity) adult fit means poorer fit.

Ontogenetic fit: The average deviation throughout ontogeny is $E[D(\tau)] = \frac{1}{n}\sum_\tau^n D(\tau)$, where E denotes expectation over τ and n is the discrete number of ages sampled. We measure ontogenetic fit as $-E[D(\tau)]$. This measure of ontogenetic fit is zero if and only if there is a perfect fit between predicted and observed brain and body mass throughout ontogeny. Decreasing ontogenetic fit (up to negative infinity) means poorer fit.

Our exploration of the challenge parameters $\mathbf{P}$ allows to identify which combination $\mathbf{P}^*$ yields a best fit with data. However, it is also useful to gain an idea of how sharply fit decreases as the challenge parameters $\mathbf{P}$

depart from the best fitting combination $\mathbf{P}^*$. With this aim, in Extended Data Fig. 9 we obtain high fit intervals around the scenarios yielding best adult fits for the various *Homo* species where the best adult fit is higher than $-D(\tau_a) = -0.05$. This threshold $-D(\tau_a) = -0.05$ is arbitrary, but facilitates seeing how flat adult fit is around its maximum value. We obtain high fit intervals by plotting adult fit vs P_j around the corresponding best fitting scenario, interpolating adult fit to obtain a continuous line, and obtaining the interval where this interpolated curve is greater than -0.05. This shows that the adult fit at the best fitting scenarios is generally a steep peak for the proportion of competitive challenges (P_3 and P_4), but is flatter for the proportions of ecological and cooperative challenges (P_1 and P_2). Consequently, the proportions of ecological and cooperative challenges in the $\mathbf{P}^*$ across *Homo* can typically be varied to some extent without drastically decreasing adult fit (i.e., the high fit intervals for P_1 and P_2 are generally wider than those for P_3 and P_4), but the proportions of competitive challenges cannot be varied much without strongly diminishing adult fit.

7   Observed values of brain and body sizes in hominins

Here we list the observed (or empirically estimated) values used for adult brain and body mass in hominins. Since the model considers females only, we use only female values when available. We used observed adult values of brain and body sizes for *H. sapiens* [body (51.1 kg) and brain (1.31 kg) for females [15]], *H. neanderthalensis* [body (66.4 kg) for females, brain (1.442 kg) averaged over sexes [40,41]], *H. heidelbergensis* [body (54.24 kg) and brain (1.16 kg) averaged over sexes [42]], *H. erectus* [body (55 kg) for females, brain (0.98 kg) averaged over sexes [43]], *H. ergaster* [body (56 kg) for females, brain (0.849 kg) averaged over sexes [44]], *H. habilis* [body (32 kg) for females, brain (0.601 kg) averaged over sexes [44]], *H. floresiensis* [body (25 kg) and brain (0.4 kg) averaged over sexes [45,46]], *H. naledi* [body (37.4 kg) and brain (0.5 kg) averaged over sexes [47]], and *Australopithecus afarensis* [body (29 kg) and brain (0.434 kg) averaged over sexes [44]]. For *H. floresiensis*, *H. naledi*, and *A. afarensis*, we find adult fits that are more than ten times poorer than those reported in Fig. 4a of the main text (Extended Data Fig. 8), suggesting that different $\mathbf{Q}$ and $\mathbf{R}$ parameters are needed for a similarly good account of these species.

8   Adjustment of the energy extraction efficiency from maternal provisioning at birth, φ_0

Here we describe how we identified the value of maternal provisioning at birth (φ_0) that locally maximized ontogenetic fit in Fig. 4b of the main text after having maximized adult fit. In our exhaustive search across the $\mathbf{P}$-parameters, we identified the $\mathbf{P}^*$-parameter combination that maximized adult fit $-D(\tau_a)$ for *H. sapiens*. Previous work [1] showed that varying φ_0 affects body and brain mass early in ontogeny but has little effect on their adult values. We then took the ten $\mathbf{P}$-parameter combinations that yield the best adult fit with *H. sapiens* across the six cases considered. All such 10 combinations occurred in the case of exponential competence with submultiplicative cooperation (in which case, $\varphi_0 = 0.6$). For each of these 10 combinations, we obtained uninvadable growth strategies with $\varphi_0 \in \{0.4, 0.45, 0.5\}$ to find a φ_0^* and a combination of $\mathbf{P}$ that yielded a solution maximizing ontogenetic fit $-\mathrm{E}[D(\tau)]$. The resulting best ontogenetic fit occurred for the same parameter combination $\mathbf{P}^*$ but with the value $\varphi_0^* = 0.5$ (Fig. 4b in the main text).

9   Possible extensions of the model

We intend our model to serve as a foundation upon which future work may be built to gain a causal understanding of brain-size evolution and related evolutionary phenomena. This approach will benefit from further exploration of the large parameter space and from progress on parameter estimates. Moreover, here we mention some lines of study that may be approached, although many will require major extensions of the model or substantial progress in overcoming the computational challenges posed by the evolutionary study of function-valued traits. These possible extensions are the following. (1) Partitioning body to distinguish other expensive tissues such as digestive[3] and adipose[4]. This would allow to assess the expensive tissue hypothesis[3] proposing a trade-off between brain size and the size of other expensive tissues by studying how the size of those tissues changes with changes in brain size. (2) Letting challenge proportions and environmental difficulty vary with age or evolutionary time[48] or with the population skill level. The dependence of environmental difficulty on evolutionary time would allow assessing hypotheses of environmental variability as drivers of human brain expansion[49]. (3) Allowing mortality rate to depend on skill level[50] or on tissues involved in bodily maintenance such as immune system components[51]. The latter would contribute to a quantitative understanding of the evolution of senescence[52,53]. (4) Studying the evolution of allocation to tissue maintenance rather than to tissue growth. This would allow for evolutionary study of tissue-specific (e.g., brain) senescence[54]. (5) Considering multiple types of skills with different functions. This would allow addressing hypotheses concerning the evolution of domain-general or domain-specific cognition[55]. (6) Partitioning brain tissue into regions with specific functions. This would permit to study how the size of some brain components is affected by the size of other components or of total brain size and viceversa, thus addressing the mosaic theory of brain evolution[56]. (7) Considering the evolutionary dynamics rather than only evolutionary equilibria, which would allow assessing evolutionary branching in brain size[57]. (8) Explicitly incorporating social learning and culture, as well as studying their effects on EEE, which would allow assessing if these factors interact with ecological challenges to drive human brain expansion, and would allow for further evaluation of links between social complexity and brain size[58–63].

References

1. González-Forero, M., Faulwasser, T. & Lehmann, L. A model for brain life history evolution. *PLoS Comput. Biol.* **13**, e1005380 (2017).

2. Metz, J. A. J., Staňková, K. & Johanson, J. The canonical equation of adaptive dynamics for life histories: from fitness-returns to selection gradients and Pontryagin's maximum principle. *J. Math. Biol.* **72**, 1125–1152 (2016).

3. Aiello, L. C. & Wheeler, P. The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Curr. Anthropol.* **36**, 199–221 (1995).

4. Navarrete, A., van Schaik, C. P. & Isler, K. Energetics and the evolution of human brain size. *Nature* **480**, 91–93 (2011).

5. Lehmann, L., Mullon, C., Akçay, E. & Van Cleve, J. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution* **70**, 1689–1702 (2016).

6. Mylius, S. D. & Diekmann, O. On evolutionarily stable life histories, optimization and the need to be specific about density dependence. *Oikos* **74**, 218–224 (1995).

7. Houston, A. I. & McNamara, J. M. *Models of Adaptive Behaviour* (Cambridge Univ. Press, 1999).

8. Wild, G. & Traulsen, A. The different limits of weak selection and the evolutionary dynamics of finite populations. *J. Theor. Biol.* **247**, 382–390 (2007).

9. Hirshleifer, J. Theorizing about conflict. In Hartley, K. & Sandler, T. (eds.) *Handbook of Defense and Economics*, vol. 1, chap. 7, 165–189 (Elsevier, 1995).

10. Jia, H., Skaperdas, S. & Vaidya, S. Contest functions: theoretical foundations and issues in estimation. *Int. J. Ind. Organ.* **31**, 211–222 (2013).

11. Bressan, A. Noncooperative differential games. A tutorial. Lecture notes (2010). URL https://www.math.psu.edu/bressan/PSPDF/game-lnew.pdf.

12. Gurven, M., Stieglitz, J., Hooper, P. L., Gomes, C. & Kaplan, H. From womb to the tomb: The role of transfers in shaping the evolved human life history. *Exp. Gerontol.* **47**, 807–813 (2012).

13. Clarke, D. D. & Sokoloff, L. Circulation and energy metabolism of the brain. In Spiegel, G. J., Agranoff, W., Albers, R. W., Fisher, S. K. & Uhler, M. D. (eds.) *Basic Neurochemistry*, chap. 31, 637–669 (Lippincott-Raven Publishers, 1999), 6th edn.

14. West, G. B., Brown, J. H. & Enquist, B. J. A general model for ontogenetic growth. *Nature* **413**, 628–631 (2001).

15. Kuzawa, C. W. *et al.* Metabolic costs and evolutionary implications of human brain development. *Proc. Nat. Acad. Sci. USA* **111**, 13010–13015 (2014).

16. Chugani, H. T., Phelps, M. E. & Mazziotta, J. C. Positron emission tomography study of human brain functional development. *Ann. Neurol.* **22**, 487–497 (1987).

17. Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. & Raichle, M. E. Aerobic glycolysis in the human brain is associated with development and neotenous gene expression. *Cell Metab.* **19**, 49–57 (2014).

18. Herculano-Houzel, S. Scaling of brain metabolism with a fixed energy budget per neuron: Implications for neuronal activity, plasticity and evolution. *PLoS ONE* **6**, e17514 (2011).

19. Karbowski, J. Global and regional brain metabolic scaling and its functional consequences. *BMC Biology* **5**, 18 (2007).

20. Clarke, D. D. & Sokoloff, L. Circulation and energy metabolism of the brain. In Siegel, G. J., Agranoff, B. W., Albers, R. W. & Molinoff, P. B. (eds.) *Basic Neurochemistry*, 645–680 (Raven Press, 1994), 5th edn.

21. Sokoloff, L. The metabolism of the central nervous system *in vivo*. In Field, J., Magoun, H. W. & Hall, V. E. (eds.) *Handbook of Physiology-Neurophysiology*, vol. 3, 1843–1864 (American Physiological Society, 1960).

22. Blaxter, K. *Energy Metabolism in Animals and Man* (Cambridge Univ. Press, Cambridge, UK, 1989).

23. McKenna, M. C., Dienel, G. A., Sonnewald, U., Waagepettersen, H. S. & Schousbe, A. Energy metabolism of the brain. In Braddy, S. T., Siegel, G. J., Albers, R. W. & Price, D. L. (eds.) *Basic Neurochemistry*, chap. 11, 200–231 (Academic Press, 2012), 8th edn.

24. Madsen, P. L. *et al.* Persistent resetting of the cerebral oxygen/glucose uptake ratio by brain activation: evidence obtained with the kety-schimdt technique. *J. Cereb. Blood Flow Metab.* **15**, 485–491 (1995).

25. van der Veer, H. W., Cardoso, J. F. M. F. & van der Meer, J. The estimation of DEB parameters for various Northeast Atlantic bivalve species. *J. Sea Research* **56**, 107–124 (2006).

26. Dekaban, A. S. & Sadowsky, D. Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights. *Ann. Neurol.* **4**, 345–356 (1978).

27. Tilly, J. L. & Sinclair, D. A. Germline energetics, aging, and female infertility. *Cell Metab.* **17**, 838–850 (2013).

28. Gougeon, A. Regulation of ovarian follicular development in primates: facts and hypotheses. *Endocr. Rev.* **17**, 121–155 (1996).

29. Magnusson, C., Hillensjö, T., Hamberger, L. & Nilsson, L. Oxygen consumption by human oocytes and blastocysts grown in vitro. *Hum. Reprod.* **1**, 183–184 (1986).

30. Goyanes, V. J., Ron-Corzo, A., Costas, E. & Maneiro, E. Morphometric categorization of the human oocyte and early conceptus. *Hum. Reprod.* **5**, 613–618 (1990).

31. Abramczuk, J. & Sawicki, W. Variation in dry mass and volume of nonfertilized oocytes and blastomeres of 1-, 2- and 4-celled mouse embryos. *J. Exp. Zool.* **188**, 25–34 (1974).

32. Scott, L. *et al.* Human oocyte respiration-rate measurement–potential to improve oocyte and embryo selection? *Reprod. Biomed. Online* **17**, 461–469 (2008).

33. Wolf, J. P., Bulwa, S., Rodrigues, D. & Jouannet, P. Human oocyte cytometry and fertilisation rate after subzonal insemination. *Zygote* **3**, 101–109 (1995).

34. Dickey, R. P. *et al.* Effect of diagnosis, age, sperm quality, and number of preovulatory follicles on the outcome of multiple cycles of clomiphene citrate-intrauterine insemination. *Reproductive Endocrinology* **78**, 1088–1095 (2002).

35. O'Herlihy, C. *et al.* Preovulatory follicular size: a comparison of ultrasound and laparoscopic measurements. *Fertil. Steril.* **34**, 24–26 (1980).

36. Howarth, C., Peppiatt-Wildman, C. M. & Attwell, D. The energy use associated with neural computation in the cerebellum. *J. Cereb. Blood Flow Metab.* **30**, 403–414 (2010).

37. Bartol Jr, T. M. *et al.* Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife* **4**, e10778 (2015).

38. Pakkenberg, B. *et al.* Aging and the human neocortex. *Exp. Gerontol.* **38**, 95–99 (2003).

39. Patterson, M. A. & Rao, A. V. GPOPS-II: A MATLAB Software for solving multiple-phase optimal control problems using hp-adaptive Gaussian quadrature collocation methods and sparse nonlinear programming. *ACM Trans. Math. Softw.* **41**, 1–37 (2014).

40. Froehle, A. W. & Churchill, S. E. Energetic competition between Neandertals and anatomically modern humans. *PaleoAnthropology* 96–116 (2009).

41. Ruff, C. B., Trinkaus, E. & Holliday, T. W. Body mass and encephalization in Pleistocene *Homo*. *Nature* **387**, 173–176 (1997).

42. Rightmire, G. P. Brain size and encephalization in early to mid-pleistocene *Homo*. *Am. J. Phys. Anthropol.* **124**, 109–123 (2004).

43. McHenry, H. M. Tempo and mode in human evolution. *Proc. Natl. Acad. Sci. USA* **91**, 6780–6786 (1994).

44. McHenry, H. M. & Coffing, K. *Australopithecus* to *Homo*: transformations in body and mind. *Annu. Rev. Anthropol.* **29**, 125–146 (2000).

45. Brown, P. *et al.* A new small-bodied hominin from the late pleistocene of flores, indonesia. *Nature* **431**, 1055–1061 (2004).

46. Kubo, D., Kono, R. T. & Kaifu, Y. Brain size of *Homo floresiensis* and its evolutionary implications. *Proc. R. Soc. B* **280**, 20130338 (2013).

47. Garvin, H. M. *et al.* Body size, brain size, and sexual dimorphism in homo naledi from the Dinaledi Chamber. *J. Hum. Evol.* **111**, 119–138 (2017).

48. Sæther, B.-E. & Engen, S. The concept of fitness in fluctuating environments. *Trends Ecol. Evol.* **30**, 273–281 (2015).

49. Potts, R. Variability selection in hominid evolution. *Evol. Anthropol.* **7**, 81–96 (1998).

50. Gurven, M. & Kaplan, H. Longevity among hunter-gatherers: a cross-cultural examination. *Popul. Dev. Rev.* **33**, 321–365 (2007).

51. McDade, T. W. Life history theory and the immune system: steps toward a human ecological immunology. *Am. J. Phys. Anthropol.* **122**, 100–125 (2003).

52. Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* **12**, 12–45 (1966).

53. Kaplan, H. & Robson, A. J. The emergence of humans: the coevolution of intelligence and longevity with intergenerational transfers. *Proc. Natl. Acad. Sci. USA* **99**, 10221–10226 (2002).

54. Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U. & Bäckman, L. Memory aging and brain mainte-nance. *Trends Cogn. Sci.* **16**, 292–305 (2012).

55. Cosmides, L., Barrett, C. & Tooby, J. Adaptive specializations, social exchange, and the evolution of human intelligence. *Proc. Natl. Acad. Sci. USA* **107**, 9007–9014 (2010). Supplement 2.

56. Barton, R. A. & Harvey, P. H. Mosaic evolution of brain structure in mammals. *Nature* **2000**, 1055–1058 (2000).

57. Dieckmann, U., Heino, M. & Parvinen, K. The adaptive dynamics of function-valued traits. *J. Theor. Biol.* **241**, 370–389 (2006).

58. Kaplan, H., Hill, K., Lancaster, J. & Hurtado, A. M. A theory of human life history evolution: diet, intelli-gence, and longevity. *Evol. Anthropol.* **9**, 156–185 (2000).

59. Moll, H. & Tomasello, M. Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Phil. Trans. R. Soc. B* **362**, 639–648 (2007).

60. Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B. & Tomasello, M. Humans have evolved special-ized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**, 1360–1366 (2007).

61. van Schaik, C. P. & Burkart, J. M. Social learning and evolution: the cultural intelligence hypothesis. *Phil. Trans. R. Soc. B* **366**, 1008–1016 (2011).

62. Henrich, J. *The Secret of our Success* (Princeton Univ. Press, 2016).

63. Laland, K. N. *Darwin's Unfinished Symphony* (Princeton Univ. Press, 2017).