Section 1. Conceptualizing the estimation of indirect fitness contributions of pedagogical investments

William Hamilton's (1966) "force of selection" describes the sensitivity of fitness r (where r is the population growth rate) due to an infinitesimal perturbation in a vital rate [1]. It relies on Lotka's renewal equation reflecting the relative fitness contributions of individuals of different ages x (x = 0, 1, 2, ..., T), assuming stable population growth (eq. S1):

$$1 = \sum_{x=0}^{T} e^{-rx} \, l_x \, m_x \tag{S1}$$

One common way to estimate the force of selection is to take the partial derivative of the population growth rate (λ) with respect to perturbations in age-specific survival p_x from age x to x+1 (eq. S2):

$$\frac{d\lambda}{dp_x} = \frac{\lambda}{p_x} \frac{\sum_{y=x+1}^{T} e^{-ry} \, l_y \, m_y}{\sum_{y=0}^{T} y \, e^{-ry} \, l_y \, m_y}. \tag{S2}$$

Estimating the fitness contributions of pedagogical investments follows the effects of pedagogy on lifetime production by individuals age y through surplus production (P_y) and its effects on the nutritionally-dependent vital rates underlying population growth via transfers to individuals age x (F_x). Through serial sensitivities and the chain rule, we can estimate the indirect fitness contributions of surplus production transfers ($\partial \lambda / \partial P_y$) using the summed product of four sensitivities:

(1) sensitivities of fitness to fertility ($\partial \lambda / \partial m_x$) and to survivorship ($\partial \lambda / \partial p_x$)
(2) the sensitivity of vital rates (m_x, p_x) to food ratio E_x, where E_x refers to the proportion of optimal caloric consumption necessary to maximize fertility or survivorship (e.g. $\partial m_x / \partial E_x$). Vital rates are at a maximum when E_x = 1, following the approach of Lee & Tuljapurkar [2]
(3) sensitivity of food ratio E_x to receiving a food transfer (F_x) at age x by individuals age y, based on caloric demand across the age structure
(4) sensitivity of food transfers to an increase in food production, e.g. rules governing the extent of kin-based production transfers depending on age-specific net production returns P_y ($\partial F_x / \partial P_y$):

$$\frac{\partial \lambda}{\partial P_y} = \sum_{x=0}^{T} \frac{\partial \lambda}{\partial m_x} \frac{\partial m_x}{\partial E_x} \frac{\partial E_x}{\partial F_x} \frac{\partial F_x}{\partial P_y} + \sum_{x=0}^{T} \frac{\partial \lambda}{\partial p_x} \frac{\partial p_x}{\partial E_x} \frac{\partial E_x}{\partial F_x} \frac{\partial F_x}{\partial P_y}. \tag{S5}$$

In addition to relying on assumptions about food sharing based on surplus production ($\partial F_x / \partial P_y$), age-profiles of unmet caloric demand ($\partial E_x / \partial F_x$) and kin-preference ($\partial F_x / \partial P_y$), this sensitivity to production transfers ($\partial \lambda / \partial P_y$) depends on assumptions about how food consumption (nutritional deficit and surplus) affects fertility and survivorship ($\partial \lambda / \partial m_x$ and $\lambda / \partial p_x$), which are poorly quantified in humans, complicating empirical estimation. It is not unreasonable to assume that the norms of reaction for survival and fertility with respect to caloric consumption are concave functions, where minor deprivation has small effects and severe deprivation has increasingly larger impacts on vital rates.

Once we have the fitness sensitivities of production transfers, we can additionally estimate the fitness sensitivities of pedagogy by modifying eq. S5. What is needed is to include how a unit of pedagogy (for simplicity here represented as time allocation at age z, (t_z) impacts surplus food production (P_y) by increasing knowledge (K_y). The modification of eq. S5 therefore applies the response of production P_y to skill level K_y (after Gurven and Kaplan [3]) and incorporates the effects of teacher's pedagogical time-allocation at age z (t_z) on pupil skill K_y at each age y ($\partial K_y / \partial t_z$):

$$\frac{\partial \lambda}{\partial t_z} = \sum_{x=0}^{T} \sum_{y=0}^{T} \frac{\partial \lambda}{\partial p_x} \frac{\partial p_x}{\partial E_x} \frac{\partial E_x}{\partial F_x} \frac{\partial F_x}{\partial P_x} \frac{\partial P_x}{\partial K_y} \frac{\partial K_y}{\partial t_z} + \sum_{x=0}^{T} \sum_{y=0}^{T} \frac{\partial \lambda}{\partial m_x} \frac{\partial m_x}{\partial E_x} \frac{\partial E_x}{\partial F_x} \frac{\partial F_x}{\partial P_x} \frac{\partial P_x}{\partial K_y} \frac{\partial K_y}{\partial t_z},$$ (S6)

Where ($\partial \lambda / \partial t_z$) is Hamilton's force of selection expressed as the change in population growth rate, or fitness, due to one year of pedagogical or informational "transfer" at age z (t_z). This is a product of five terms: ($\partial \lambda / \partial p_x$ and $\partial \lambda / \partial m_x$) are the traditional vital rate sensitivities reflecting fitness responses to changes in fertility or survivorship at age x; ($\partial m_x / \partial E_x$ and $\partial p_x / \partial E_x$) describe nutrition effects reflected in fertility and survivorship changes through caloric consumption; ($\partial E_x / \partial F_x$) addresses the change in food ratio E_x as a result of food transfers F_x to individuals age x stemming from net surplus production by individuals age y (P_y); ($\partial F_x / \partial P_y$) incorporates food sharing strategies for distributing transfers across age classes; ($\partial P_y / \partial K_y$) reflects knowledge- or skill-effects on production; ($\partial K_y / \partial t_z$) refers to how subsistence knowledge K_y at age y changes with information transfers t_z by individuals age z.

In our framework, information transfers alter the traditional fitness contributions of mortality and fertility to reflect the impacts of teaching on food production across the life cycle and their concomitant effects on population fitness through the nutritional dependence of vital rates. The estimation of fitness contributions from pedagogy in human and other social animals is currently under investigation. The current paper contributes to this broader goal by helping to understand and operationalize the last two elements of eq. S6: changes in production due to increased knowledge or skills ($\partial P_y / \partial K_y$), and changes in subsistence knowledge due to information transfer ($\partial K_y / \partial t_z$).

Section 2. Additional methodological detail on teacher cost estimation in human hunting.

The extensive cross-cultural hunting database recently compiled by Jeremy Koster and colleagues includes >23,000 records of 1,800 individuals from 40 locations [4]. As noted in the main text, for our analyses we used two separate methods to identify cases of potential teaching in this database. The difference between these methods comes from an explicit distinction made by the primary architects of the data set between primary foragers and assistants on hunting expeditions. Specifically, primary foragers were denoted as individual or cooperative hunters who sometimes pooled harvests and were active in decision-making and direct pursuit/capture of prey, whereas an additional category, termed "assistants," was defined as individuals on hunting trips "to whom harvested biomass cannot be credited." In most cases assistants were young or old individuals who accompanied foraging trips without technology or implements that would allow active participation in hunting. For each of these categories, we generated a binary variable of *potential* teaching and conducted analysis as follows.

For the first case (cooperative hunts between primary foragers), we focused only on cases of cooperative hunts (more than 1 primary forager on a hunting trip) and distinguished cases in which at least one forager of age <20 years old (other than the focal forager) was part of a hunting group. Given this focus, we first excluded any societies from the data set for which no cooperative hunting instances were observed (we expect that this indicates that cooperative hunting was not recorded or coded, or if truly absent indicates that a given society is not relevant to the analysis), and removed any cases of solo hunting. Each row of the data set was therefore comprised of a person-hunt observation for cooperative hunts, with separate rows for each forager in the hunting group and outcomes noted at either the individual or pooled level. If at least one other forager in the group was <20 years old, then a focal hunter would have 1 assigned for potential teaching, and a 0 if not.

For the second case (hunts with or without the presence of assistants), we started by reducing the overall data set to only those societies for which the explicit presence of assistants was noted at least one time (due to the fact that several of the original ethnographers did not include such observations in their records). For each row, which represents a single person-hunt observation, we then calculated the number of individual assistants reported with ages of <20 years old (to remove cases of older individuals who likely joined hunts as assistants for non-pedagogical reasons). Because so few cases included more than one potential teaching assistant, we then reduced this to a binary variable indicating the presence of at least one potential teaching assistant. Prior to analysis, we further excluded observations of groups with party size >5 and individuals >75 years old (very few data were reported in this age range and thus estimation of smoothing splines performed poorly past this point).

Hunting data often has a high prevalence of zeros and a high degree of right-skew among non-zero outcomes, and thus such data are amenable to the use of zero-adjusted gamma (or alternatively lognormal) models [5]. Here we used Bayesian estimation in R with the packages *brms* and *rstan* to fit a multilevel lognormal hurdle model of hunting returns as a function of each teaching variable (presence of young primary foragers in group or presence of young assistants), age (modeled with a thin-plate smoothing spline), and adjusting for a pooled harvest, sex, the use of guns, foraging party size, hunting trip duration, use of dogs, and an interaction between pooled harvest and party size. Individual and population were modeled as with random intercepts to account for hierarchical clustering at these levels. If *y* is an indicator variable for a non-zero hunting return, and *h* is the continuous value of non-zero harvests, then the model is of the following form:

$$y_i \sim Bernoulli(p_i)$$

$$h_i \sim LogNormal(\mu_i, \sigma)$$

$$\text{logit}(p_i) = \alpha_{individual[i]} + \gamma_{society[i]} + s_1(age_i) + \beta_1 teaching_i + \beta_2 sex_i + \beta_3 trip\_duration_i$$
$$+ \beta_4 dogs_i + \beta_5 guns_i + \beta_6 pooled_i + \beta_7 party\_size_i + \beta_8 pooled_i * party\_size_i$$

$$\mu_i = \alpha_{individual[i]} + \gamma_{society[i]} + s_2(age_i) + \beta_9 teaching_i + \beta_{10} sex_i + \beta_{11} trip\_duration_i + \beta_{12} dogs_i$$
$$+ \beta_{13} guns_i + \beta_{14} pooled_i + \beta_{15} party\_size_i + \beta_{16} pooled_i * party\_size_i$$

Where *s1* and *s2* are thin-plate regression splines formulated in the *mgcv* package, "teaching" is a binary variable associated with the presence of young assistants or young group members in cooperative foraging, "pooled" is a binary indicator variable of whether or not the outcome reflects a pooled harvest among multiple individuals, "trip_duration" is the length of foraging trip in hours, "dogs" is a binary variable indicating the presence of hunting dogs, "guns" is a binary variable indicating the use of firearms by the focal individual, "sex" indicates male versus female hunters, and "party_size" is a numeric variable indicating the number of primary foragers in the group. Non- or weakly-informative priors were used for all parameters.

Because a small number of observations were missing data on the use of dogs, guns, or trip duration, we employed multiple imputation using the *mice* package in R and ran the same fitted model on n=5 imputed data sets. Bayesian methods are convenient because posterior values can be combined from all models to produce a single posterior that accounts for variation from multiple imputation. We assessed chain convergence and mixing by visual inspection of trace plots, as well as the Gelman-Rubin diagnostic $\hat{R}$ and effective sample size (calculated by *brms* and *rstan*). All parameters exhibited $\hat{R}$ values less than 1.01, with acceptable effective sample sizes and good mixing. Here we present only a single model, but note that other model formulas, such as those that allowed for an interaction between our teaching variables and age or a third-order polynomial for age, were considered and produced qualitatively equivalent results. Full reproducible R code that loads data from the original source [4] is available at https://osf.io/38624/.

Section 3. Additional methods detail on human hunting and age of teaching.

The hunting database from [4] also conveniently provides information on the age profiles of hunting as it pertains to different group compositions. We conducted two separate analyses of group composition. For the first, we constructed a binary outcome variable indicating whether a person-hunt observation was an independent solo hunt (1) or part of a cooperative group hunt (0). For the second, we used the same variable described above indicating whether a cooperative hunt included a young (<20 years old) group member.

For each outcome variable we specified a multilevel model with a Bernoulli distribution (binomial distribution with n_{trials} = 1 for each observation) for the response and a logit link function, estimating the response as a function of a third-order (orthogonal) polynomial of age, the use of dogs, and the use of guns, with random intercepts for individual forager and society:

$$y_i \sim Bernoulli(p_i)$$
$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \alpha_{individual[i]} + \gamma_{society[i]} + \beta_0 age_i + \beta_1 age_i^2 + \beta_2 age_i^3 + \beta_3 dogs_i + \beta_4 guns_i$$
$$\beta_0 \sim Normal(0,10)$$
$$\beta_1 \sim Normal(0,50)$$
$$\beta_2 \sim Normal(0,50)$$
$$\beta_3 \sim Normal(0,10)$$
$$\beta_4 \sim Normal(0,10)$$
$$\alpha_j \sim Normal(\bar{\alpha}, \sigma_\alpha)$$
$$\gamma_j \sim Normal(0, \sigma_\gamma)$$
$$\bar{\alpha} \sim StudentT(3, 0, 10)$$
$$\sigma_\alpha \sim StudentT(3, 0, 10)$$
$$\sigma_\gamma \sim StudentT(3, 0, 10)$$

Where *y* is a vector of binary outcomes indicating whether an observation is a solo hunt (first model) or whether or not a cooperative hunting observation was part of a group where at least one other individual was young (second model). For the model of young group members on cooperative hunts, we additionally included a variable for the number of primary foragers to account for the fact that a larger group is more likely to include at least one young member. Multiple imputation was performed and model diagnostics and convergence were assessed as described previously. Code is available at https://osf.io/38624/.

Section 4. Additional methods detail on ages of teacher-student dyads among Tsimane Amerindians.

To assess the age profiles of potential teacher-student dyads, we used a scan sampling database on Tsimane forager-horticulturalists comprised of 70,745 observations. For a relatively small number of these observations (n=3551), the activity of individuals was recorded in the behavioral category of "accompaniment," in which the focal participated as part of the activity or social group of others engaged in productive activities. This behavioral category is therefore essentially equivalent to the "assistant" category described above for the data in [4]. For each instance of accompaniment, we identified all of the individuals engaged directly in the primary activity as well as all of the individuals noted to have accompanied them. Accompanying individuals were further filtered to include only those of age <20 years old in order to identify individual assistants who were potential targets of pedagogy. We then used this information to construct a database of all observed teacher-student (forager-assistant) dyads. Because repeated behavioral scans were conducted over the course of single days, we filtered observations to include only a single instance of an activity group on each day (e.g. if person A hunted with assistant B on day 1, and there were 6 observations of this activity group between 1 pm and 3 pm, this group would appear only as a single row indicating one dyad observed on day 1). Each row was also assigned a macro-activity code (fishing, agriculture, other foraging (fruit/honey collection), hunting) based on the activity of the primary forager of each dyad. To produce Figure 7, we calculated the unweighted mean age of students and teachers for each macro-activity category and calculated bootstrapped 95% confidence intervals using the *boot* package in R.

In addition to investigating the age profiles of dyads involving foragers and assistants, we also examined the age profiles of cooperative foraging dyads. The key difference is that in the latter case multiple individuals were part of the same working group and all individuals were listed as participating directly in the main activity (rather than some individuals having received the behavioral code for

"accompaniment"). Thus, these dyads represent cooperative partnerships in which both individuals can be considered as primary foragers.

After identifying all pairwise cooperative foraging dyads, we repeated the procedures described above to reduce observations to include only a single instance of an activity group on each day, and calculated means and bootstrapped 95% confidence intervals of ego and alter ages for the following: 1) macro-activity categories (fishing, gardening (horticulture), hunting, other foraging (fruit/honey gathering); 2) sub-categories of fishing (bow and arrow, hook/line, net, poison); 3) sub-categories of horticulture (tree-chopping, clearing (with machete), harvesting, planting). These relationships are plotted in Fig. 7. Instead of focusing on only mean ages of teachers and pupils (which itself is not independent of the age structure of the population), Figure S7 compares the distribution of teacher and pupil ages for horticulture and hunting using a kernel density plot. Though showing similar mean ages for teacher and pupil ages in Figure 7a, horticultural activities involve a higher density of individuals at young ages (<15 yo) while hunting includes an extended range of egos and alters.

Table S1. Examples of animal teaching from Thornton & Raihani [6]. Life history traits for animals with teaching as evidenced by teachers altering behavior in the presence of pupils. For each species examined by [6], we provide common name, Latin name, the type of behavior reported, whether behavior was altered (evidence of teaching), evidence of apparent costs or benefits to teaching. To these we add life history traits: life expectancy (e_0), age at maturity/first birth (*AFB*), the proportion of the lifespan after maturity [Pr(mat) = (e_0 - *AFB*) / e_0]. Except for chimpanzees, where life history traits are taken from [41], lifespan and maturity measures are taken from https://animaldiversity.org/.

Common Name	Latin Name	Behavior Studies	Behavior?	Cost?	Benefit?	C/B?	e_0 (y)	*AFB* (d)	Pr(mat)	Life History Source
tiger	*Panthera tigris*	foraging	YS	N	N	0	9.0	1268	61%	https://animaldiversity.org/accounts/Panthera_tigris/
Canadian otter	*Lontra canadensis*	foraging	YS	N	N	0	8.5	913	71%	https://animaldiversity.org/accounts/Lontra_canadensis/
peregrine falcon	*Falco pergrinus*	foraging	YS	N	N	0	13.0	1095	77%	https://animaldiversity.org/accounts/Falco_peregrinus/
gorilla	*Gorilla gorilla*	motor	YS	N	N	0	41.0	2829	81%	https://animaldiversity.org/accounts/Gorilla_gorilla/
california sealion	*Zalophus californianus*	motor	YW	N	N	0	17.0	1095	82%	https://animaldiversity.org/accounts/Zalophus_californianus/
lion	*Panthera leo*	foraging	YW	N	N	0	22.0	1095	86%	https://animaldiversity.org/accounts/Panthera_leo/
killer whale	*Orcinus orca*	foraging	YS	N	N	0	63.0	2920	87%	https://animaldiversity.org/accounts/Orcinus_orca/
Dwarf mongoose	*Helogale parvula*	foraging	YS	N	N	0	10.0	456	88%	https://animaldiversity.org/accounts/Helogale_parvula/
hamadryas baboon	*Papio hamadryas*	motor	YS	N	N	0	37.0	1514	89%	https://animaldiversity.org/accounts/Papio_hamadryas/
black-footed cat	*Felis nigripes*	foraging	YS	N	N	0	13.0	444	91%	https://animaldiversity.org/accounts/Felis_nigripes/
european sparrowhawk	*Accipiter nisus*	foraging	YS	N	N	0	30.0	730	93%	https://animaldiversity.org/accounts/Accipiter_nisus/
kestrel	*Falco tinnunculus*	foraging	YS	N	N	0	15.5	365	94%	https://animaldiversity.org/accounts/Falco_tinnunculus/
bat-eared fox	*Otocyon egalotis*	foraging	YS	N	N	0	14.0	255	95%	https://animaldiversity.org/accounts/Otocyon_megalotis/
olive baboon	*Papio anubis*	motor	YS	N	YW	1	25.2	2738	70%	https://animaldiversity.org/accounts/Papio_anubis/
barbary macaque	*Macaca sylvana*	motor	YS	N	YW	1	22.0	1380	83%	https://animaldiversity.org/accounts/Macaca_sylvanus/
chimpanzee	*Pan troglodytes*	communication	YS	N	YW	1	59.4	3650	83%	Gurven and Davison 2019
European fox	*Vulpes vulpes*	foraging	YS	N	YW	1	5.0	304	83%	https://animaldiversity.org/accounts/Vulpes_vulpes/
norway rat	*Rattus norvegicus*	foraging	NC	YS	N	1	2.0	105	86%	https://animaldiversity.org/accounts/Rattus_norvegicus/
marmoset	*Callithrix jacchus*	foraging	NC	N	YW	1	11.0	477	88%	https://animaldiversity.org/accounts/Callithrix_jacchus/
pigtailed macaque	*Macaca nemestrina*	motor	YS	N	YW	1	28.0	1186	88%	https://animaldiversity.org/accounts/Macaca_nemestrina/
osprey	*Pandion haliaetus*	foraging	YS	N	YW	1	26.2	1095	89%	https://animaldiversity.org/accounts/Pandion_haliaetus/
meerkat	*Suricatta suricatta*	foraging	YS	YS	YS	1	10.0	365	90%	https://animaldiversity.org/accounts/Suricata_suricatta/
golden lion tamarin	*Leontopithecus rosalia*	foraging	YS	N	YW	1	15.0	547	90%	https://animaldiversity.org/accounts/Leontopithecus_rosalia/
vervet monkey	*Chlorocebus aethiops*	communication	YS	N	YW	1	31.0	1034	91%	https://animaldiversity.org/accounts/Chlorocebus_aethiops/
Pied babbler	*Turdoides bicolor*	communication	YS	YS	YS	1	13.5	365	93%	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6065330/
cheetah	*Acinonyx jubatus*	foraging	YS	YS	N	1	17.0	435	93%	https://animaldiversity.org/accounts/Acinonyx_jubatus/
cowbird	*Molothrus ater*	communication	YS	N/NC	YS	1	16.9	365	94%	https://animaldiversity.org/accounts/Molothrus_ater/
domestic cat	*Felis catus*	foraging	YS	N	YS	1	14.0	180	96%	https://animaldiversity.org/accounts/Felis_catus/

Table S2. Model variable definitions.

Symbol	Variable	Units	Description
S	Strength	% max	Increases rapidly with age, declines early
K	Skill	% max	Increases slowly with age, declines late
α	Strength Effect	% max	Exponential production ~ strength
β	Skill Effect	% max	Exponential production ~ skill
x	Consumption Age	y	Age of consumption (vital rate effects)
y	Production Age	y	Age of production (pupil boost)
z	Teacher Age	y	Age of teaching (teacher handicap)
θ	Pupil Boost	% base	Increase in pupil skill-acquisistion rate
φ	Teacher Handicap	% base	Decrease in teacher production
z	Pupil Onset Age	y	Age of pupil when instruction begins
b	Teacher Onset Age	y	Age of teacher when instruction begins
q	Age Difference	y	Teacher age - pupil age
t	Duration of Instruction	y	Years of instruction

Table S3. Life cycle for model construction.

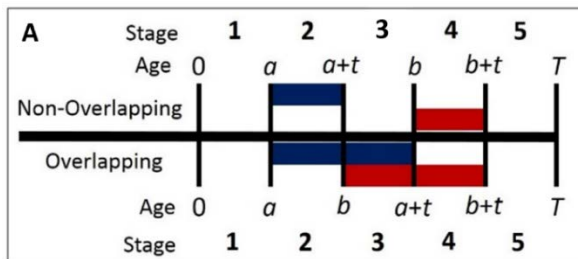| Stage	Start (x)	End (x)	Stages	θ	ϕ	Phases	θ	ϕ
			No Overlap			**Overlap**		
1	0	a-1	Pre-Pupil	Current	None	Pre-Pupil	Current	None
2	a	b-1	Pupil	Residual	None	Pupil	Current	None
3	b	a+t	Post-Pupil	Residual	None	Pupil/Teacher	Residual	Current
4	a+t+1	b+t	Teacher	Residual	Current	Teacher	Residual	Current
5	b+t+1	T	Post-Teacher	Residual	None	Post-Teacher	Residual	None

Fig. S1. Strength and skill requirements of subsistence activities. (a) Stylized ontogenies for strength (S_x) and skill (K_x) by age [sensu 3]. (b) Life cycle production ($P_x = S_x^\alpha K_x^\beta$) of four stylized subsistence activities combining high (H) or low (L) requirements for strength (S) and skill (K) (e.g. LS/LK for low-strength/low-skill), with production determined by the exponents α and β [3]. For low/high strength: α = 0.1 / 0.7; low/high skill: β = 0.1 / 0.7. (c) Stable population production contributions ($C_x = P_x l_x$) made at each age x are discounted by survivorship (l_x, blue line).
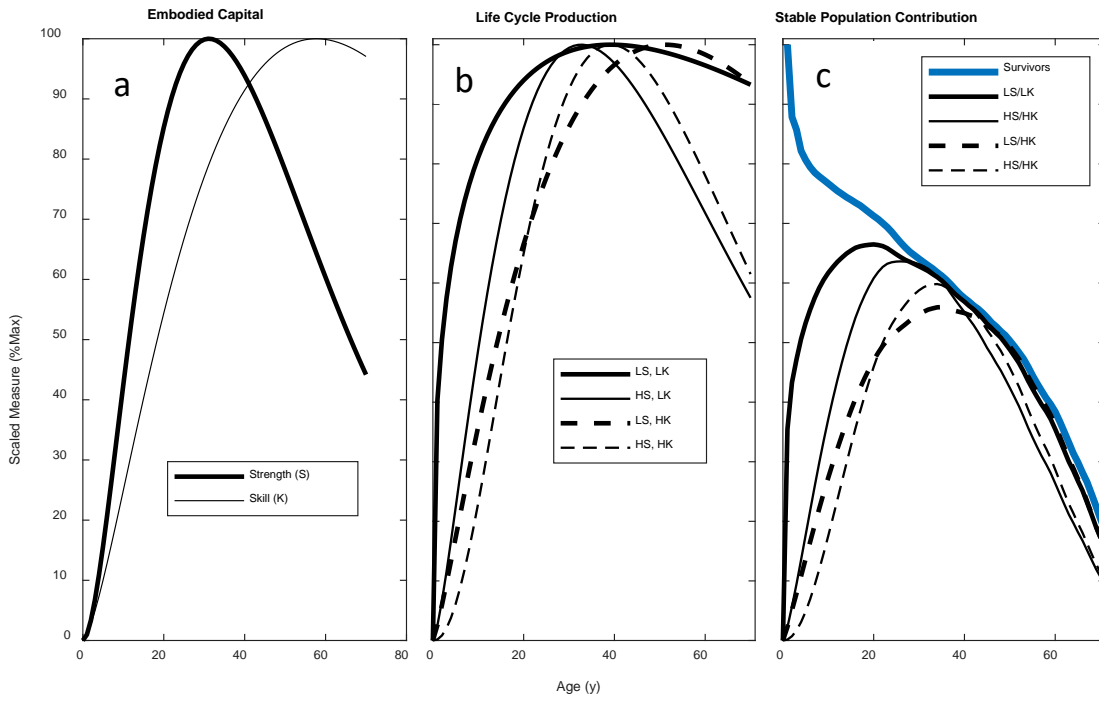
Fig. S2. Age profiles of production as a function of strength and skill requirements. The top row (a-c) shows production when skill requirements are fixed (low or high) but strength requirements vary ($0 < \alpha < 1$); bottom row (d-f) shows production when strength requirements are fixed but skill requirements vary ($0 < \beta < 1$). Bold black lines show the highest strength ($\alpha = 1$) or skill ($\beta = 1$) requirement in the series. Rightmost column (c, f) shows the effects of skill (or strength) by subtracting the low skill (or strength) production from the high skill (strength) production. Production deficits in high- vs. low-skill contexts are concentrated at younger ages (<30, where skill is low) but at older ages (>30) high-skill production is higher; deficits in high- vs. low-strength context are evident at both young ages (<20) and older ages (>40) and it is prime-age adults (ages 20-40) who produce more via high-strength activities.
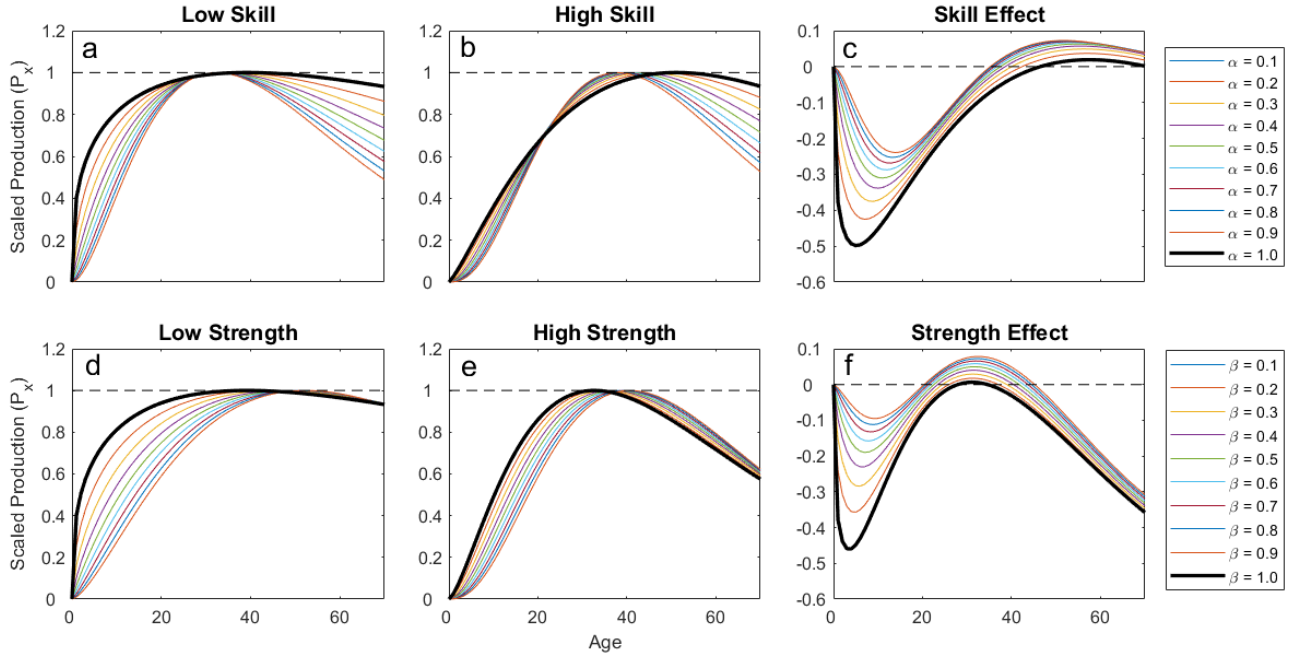
Figure S3. Effects of pedagogy on age at skill mastery. Age at mastery (x_{K*}, z-contours) for a given pedagogical context (here, θ = 25%-100%, t = 10y) responds to age at pupil onset (a, x-axis) and age at teacher onset (b, y-axis) (here, shown for high strength, high skill activities). Optimal age of teacher ($b*$) is between ages 30 to 40, whereas optimal age of pupil depends on any constraints on student learning. With no constraints, learning should occur early when the rate of skill acquisition is fastest at age 7 (top row). If pupils benefit more from teaching when they already have some skills, then the optimal age of pupil onset occurs at later ages (age 21, bottom row). Note: in these examples, skill mastery occurs at age 59 with no pedagogy (baseline/maximum x_{K*}), and we examine only pairings in which the teacher is older than the pupil (note the absence of contours below and to the right of the bold black age-parity line). Optimal ages of teacher and student are similar across columns, despite varying θ, which decreases the age at mastery, but optimal pupil ages are older when some pupil skill is necessary to benefit from pedagogy.
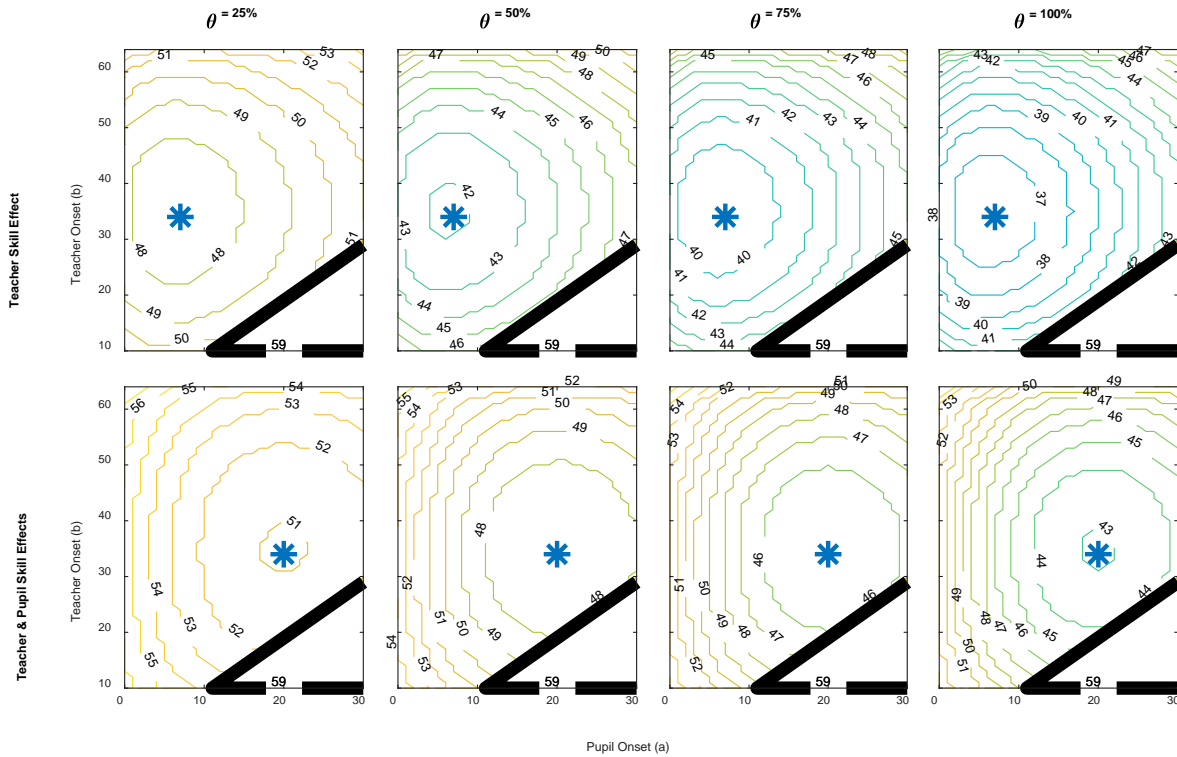

11

Figure S4. Effects of teacher age on age at mastery (based on Fig. S3). Mastery is attained at the youngest ages (x_{K*}) when the pupil boost (θ) is high, but responds nonlinearly to age of teacher onset (b). When teachers begin instruction early (e.g. b=20, left column), age of pupil mastery is later because teachers are not yet very skilled (note that we do not examine conditions where teachers are younger than pupils). When teachers begin instruction late (e.g. b=60, right column), mastery also occurs later, because of mortality attrition limiting availability of teachers at those ages. If pedagogical effects on skills ontogeny only depend on teacher age (b), mastery is attained earliest when pupils begin instruction at the age where the rate of learning is highest (a* = 7) but if skills ontogeny also depends on pupil skill levels during instruction (reflecting pupil preparedness), optimal pupil ages are later (a*= 21).
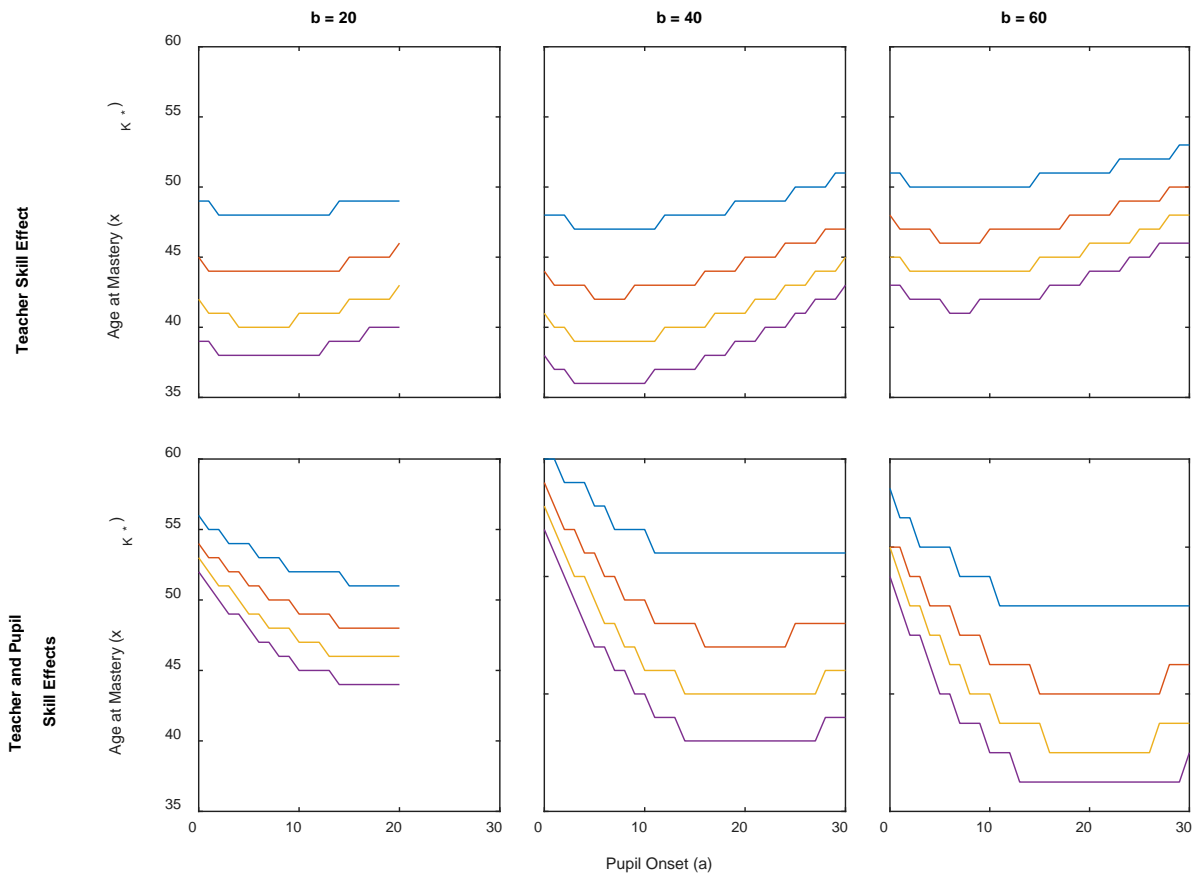
Figure S5. Production effects of pedagogy. For different teacher handicaps (φ = 1%, 5%, 10% panel columns), production responds to pupil onset (*a*, *x*-axes), teacher onset (*b*, *y*-axes), skill requirements (β) and strength (α) requirements (rows). Production (*z*-contours) is scaled relative to baseline production without pedagogy (bold contour at P_{T*} / P_T = 1.0) and we do not examine pairings where teachers are younger than students (*b*>*a*, bottom right triangle). For ease of comparison all scenarios apply maximal 10 years of instruction (*t* = 10) with equally strong pupil boost (θ = 100%).

13
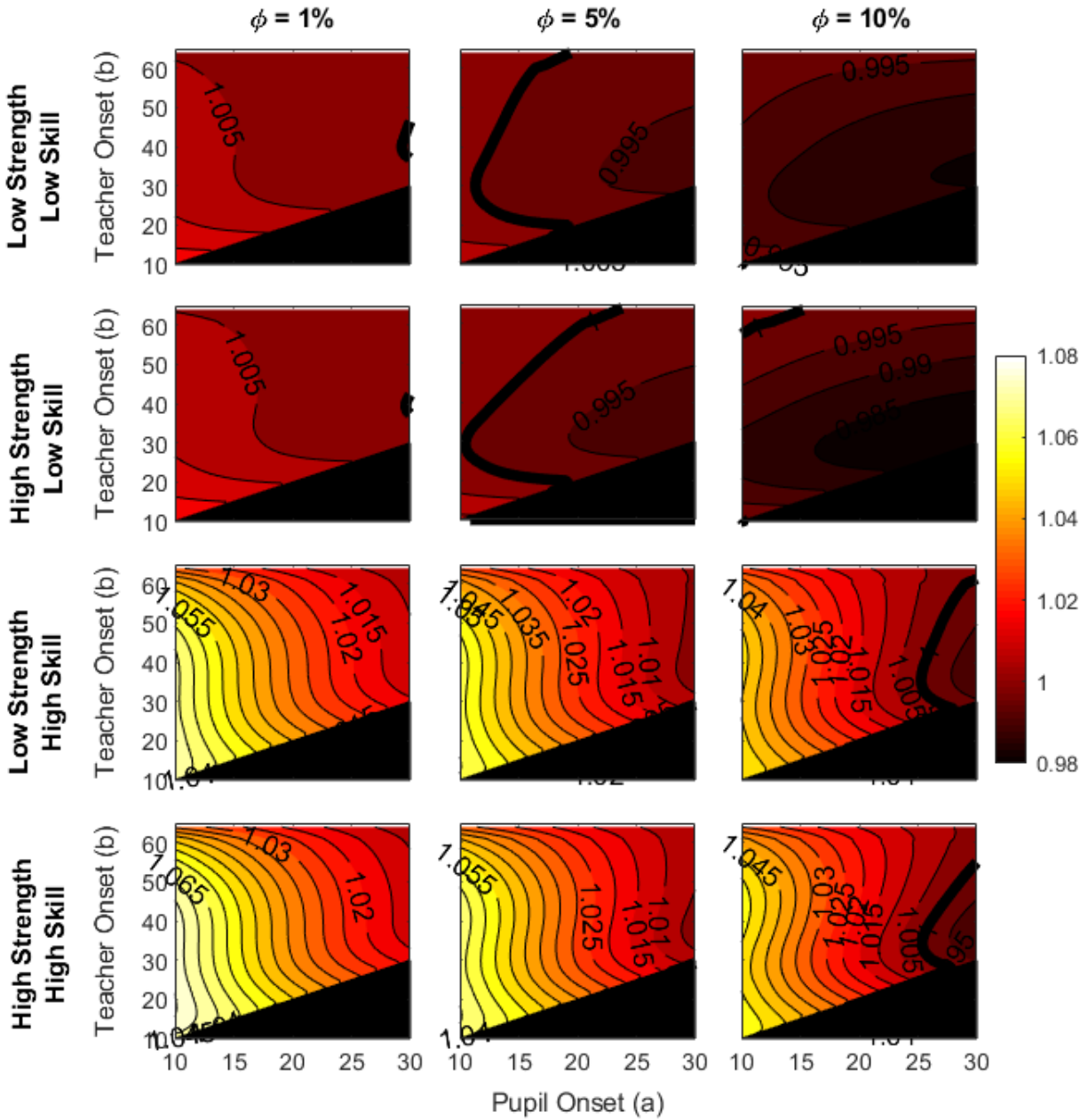
Interpretation of Production effects shown in Fig. S5:

1. Low-skill activities favor peer-tutoring when teaching costs are low and elder instruction when teaching is costly. High strength requirements also make elder teaching more favorable because strength-related production declines, thereby reducing opportunity costs.

2. Pupils learning high-skill activities benefit from pedagogy across the life cycle, and by teachers of a broader range of ages, but under a more stringent set of conditions. When teaching costs are relatively high, prime-age producers should never teach. Whereas low-skill activities favor peer tutoring or elder instruction, high-skill activities favor teaching by elder peers (5-10 years older than pupils) who still haven't reached peak production and by adults just past prime production. As with low-skill activities, high strength requirements favor older teachers, especially when handicaps are strong.

3. Peer tutoring and social learning of low-strength/low-skills activities is consistent with chimpanzee-human comparisons showing little teaching in chimpanzee subsistence and costly teaching in humans.

Figure S6. Optimal teacher age. Given pupil age at onset (a, x-axes), lines show optimal teacher age at onset (b^*, y-axes) for different teacher handicaps (ϕ, lines in legend) assuming students double their learning rate (pupil boost θ = 100%) over maximal ten years of instruction (t = 10). Thresholds indicate peak age of production (solid line), baseline at mastery (peak skill, dashed line) and the maximum age we allowed for teachers (dotted line).
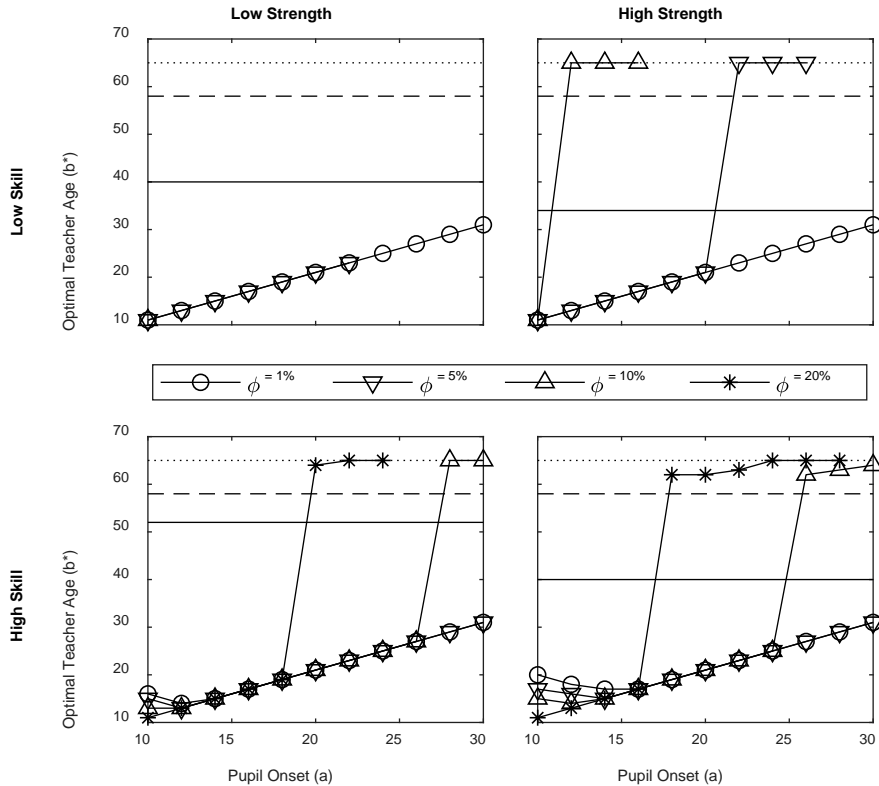
Interpretation of Figure S6:

a) **Low Skill/Low Strength** - Peer instruction is optimal if teaching costs are low but under higher costs, peer instruction is optimal for only young pupils below a threshold age

b) **Low Skill/High Strength** – If teacher costs are low, peer instruction is optimal for young pupils below a threshold age; if costs are moderate, peer tutoring is optimal up to a certain age of pupil after which (oldest) elder instruction is optimal; if costs are high, teaching is not favored

c) **High Skill/Low Strength** – Peer tutoring of young pupils is favored even if teacher costs are high. Low costs favor elder peers (~5y older than pupil), moderate costs favor switching to (oldest) elder instruction, high costs favor switching to elder instruction for younger pupils.

d) **High Strength/High Skill** – Peer tutoring of young pupils is favored under high teacher costs; if costs are low, elder peer-tutors (5-10y older than pupil) are favored for older pupils than for L/H activities. Higher costs favor switching to elder instruction for older pupils but not the oldest elders (except for older pupils), showing how in a H/H economy elders are eligible for teaching at earlier ages due to their high skill but production declines from strength.

Figure S7. Kernel density plot of cooperative foraging partnerships for hunting and horticulture. These activities have similar mean values for ego and alter age (shown in Fig. 7a), but differ in overall distributions, with horticultural activities involving a higher density of individuals at young ages (<15 yo) and hunting including an extended range of egos and alters.
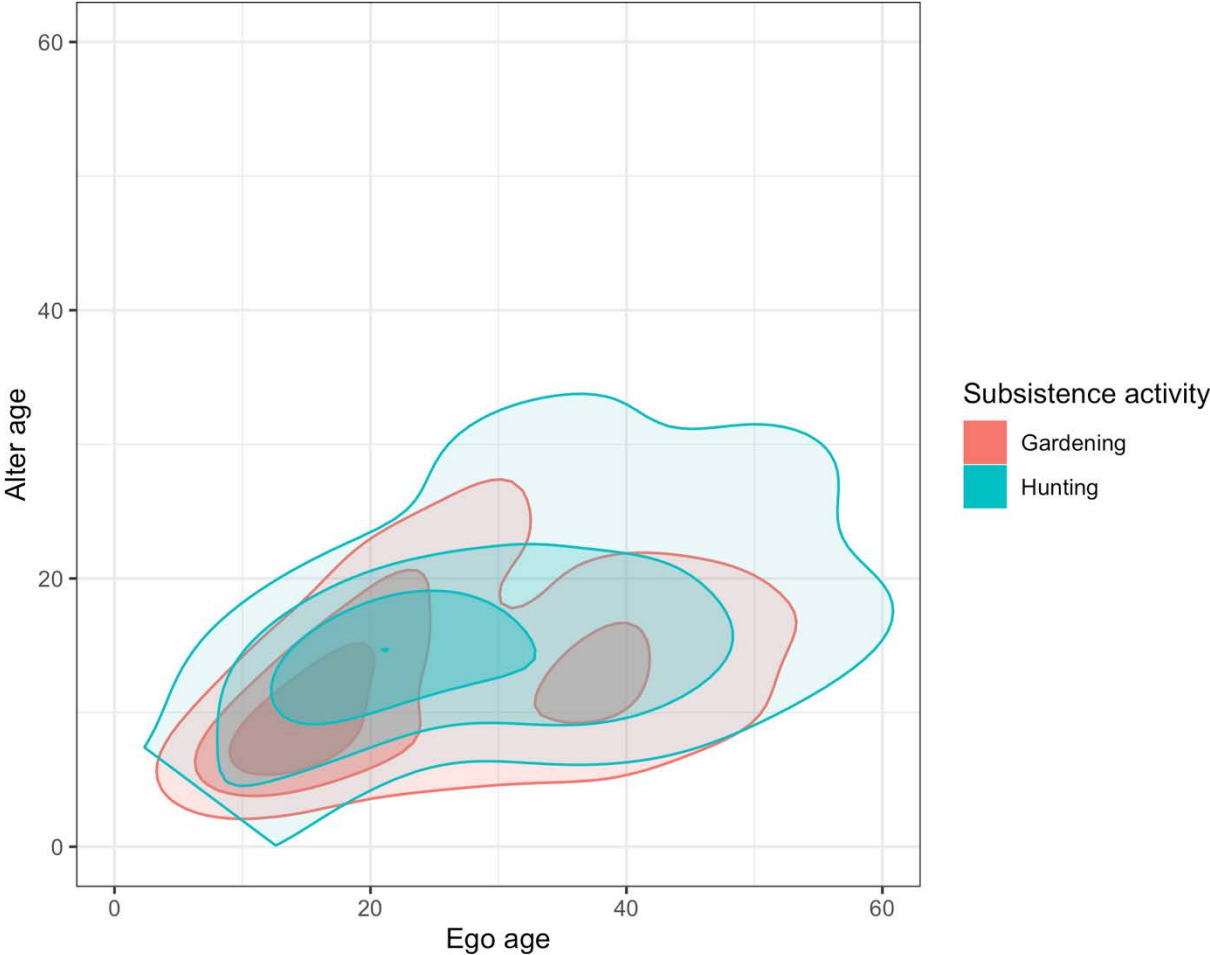
References

1.	Hamilton W.D. 1966 The molding of senescence by natural selection. *Journal of Theoretical Biology* **12**, 12-45.
2.	Lee C.T., Tuljapurkar S. 2008 Population and prehistory I: Food-dependent population growth in constant environments. *Theoretical population biology* **73**(4), 473-482.
3.	Gurven M.D., Kaplan H.S. 2006 Determinants of time allocation to production across the lifespan among the Machiguenga and Piro Indians of Peru. *Human Nature* **17**(1), 1-49.
4.	Koster J., Mcelreath R., Hill K., Yu D., Shepard G., Vliet N.V., Gurven M., Kaplan H., Trumble B., Bird R.B., et al. 2019 The Life History Of Human Foraging: Cross-Cultural And Individual Variation. *bioRxiv*, 574483. (doi:10.1101/574483).
5.	McElreath R., Koster J. 2014 Using multilevel models to estimate variation in foraging returns. *Human nature* **25**(1), 100-120.
6.	Thornton A., Raihani N.J. 2008 The evolution of teaching. *Animal behaviour* **75**(6), 1823-1836.