9 Supplementary information

9.1 A word on Knowledge

Knowledge is interpreted here as a latent trait of individuals which underlies the probability of answering the questions in the sample correctly. Henceforth, statements such 'ecological knowledge is influenced by' can be read as 'the probability of correctly answering the questions in the sample is influenced by'. Despite this practical approach, we did our best to improve the construct validity of our study, by both developing the questionnaire and managing the resulting data in collaboration with informants sharing language and culture with the interviewees. We then expect knowledge, as measured by our model, to reflect the general ecological knowledge possessed by individuals, but do not deal with general epistemological considerations on the connection between the two.

9.2 Sampling bias

As most observational studies, ours is a potential victim of sampling bias. The use of selection diagrams can help deal with sampling biases that do not interest the outcome variable. For example, controlling for the variables on which selection is happening can help balance samples (Deffner et al., 2021). Unfortunately, as a result of our opportunistic sampling strategy, our study is at risk of having sampled individuals because of their knowledge (if, for example more knowledgeable individuals were more likely to participate to the interviews). If this is the case, selection would have happened on the outcome node (see DAG in figure 3) and no statistical procedure can correct the resulting bias. But, luckily, we can estimate the presence of a sampling bias.

As a result of the opportunistic sampling, the sample contains almost all individuals between 5 and 20 years old living close to the research station where the researchers resided. This constitutes a–almost–complete sample for the sub-village where the research station is located. Individuals residing farther from the research station had lower rate of participation in the study, and hence we would expect selection bias to cause a difference in knowledge between individuals residing close and further from the research station.

To test for selection bias, we first added distance (of households from the research station) as a linear predictor of knowledge (the analysis was conducted on the 84 individuals for whom the GPS coordinate for household were available). The coefficient for distance was estimated to be at least partially positive, indicating some effect of a selection bias. In comprehensible terms, an individual living 450m away from the research station would reach the same knowledge of a 20 years old individual living 50m from the station 3.2 years earlier, on average. It is indeed possible that our sample includes more knowledgeable individuals than expected by chance.

We then run our main analysis describing the variation in knowledge by age and sex (model 1 in the main text) including only the 39 individuals residing in the sub-village, for which we have an almost complete sample. The results are shown in Figure S1 and are qualitatively comparable to the results for the entire sample.

To conclude, although we cannot exclude that our sample was biased–in fact we estimate that at least part of it was biased in a non-recoverable way–we argue that the qualitative results presented in the paper hold despite this problem, as the unbiased results are comparable.

9.3 Demographic interviews. Notes on age and activity variables.

9.3.1 Notes on reported and other ages

The present analysis focuses on age of children in a population where individuals do not celebrate birthdays and only few have or conserve birth certificates. A word on how age of individuals was estimated is hence needed. The age for each individual was assigned individually after cross-validating four sources of information: ages reported by parents or guardians during the census interviews; ages reported by the children themselves during the knowledge interviews; when available, ages reported in the registry of the local elementary school, where most of the individuals in the sample were, at some point, enrolled; focus interview with Amina Mussa Hamadi, mother of some of

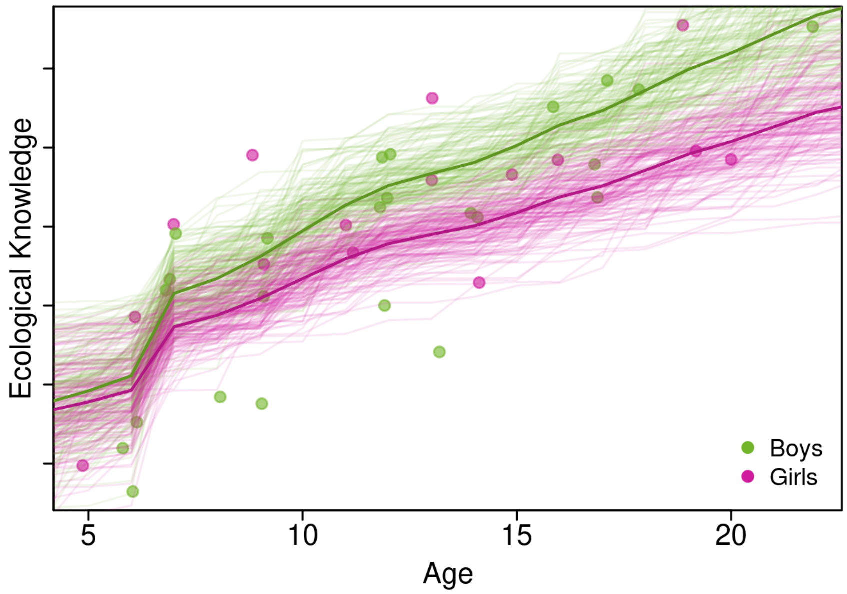

Figure S1: Knowledge estimated at different ages for the subset of individuals living close to the research station. In this sub-village, almost all individuals between 5 and 20 years old have been sampled, making this sample most likely exempt from sampling bias.

the interviewees for whom a birth certificate is available, to establish relative age of other children (e.g. child A is 8-10 months older than child B). The most supported age has then been assigned manually. With this process, we believe to have substantially reduced error around individual ages, but any year specific effect is still to be distributed over similar ages (e.g. in figure S7, higher effect of ages 7 and 11 is probably spread over the whole period 6-12).

9.3.2 Activities details

As complementary information of both the census and knowledge interviews, parents and children were asked whether each child in the sample practiced one of 10 different activities. Here follows an ethnographic description of what each activity entails.

- **Seashell collection:** The coastline close to the village is a very good place to collect seashells. This activity is practiced by girls and women of all ages, although adults participate less frequently, and by young boys, who follow their mothers or older sisters. It mainly involves spotting the position of shells under the sand and extracting them bare-handed or with the help of a knife. Different species of crabs are also pursued, as well as other occasional preys. This activity is temporally limited by the height of the tides, as most of the sandy bottom remains covered and inaccessible during high tide. The least exploited areas, that emerge from the water only when the tide is particularly low, are the most productive for foragers, so that when a tidal minimum happens on a weekend, people arrive from other villages to collect seashells close to the village, but normally only few people can be seen walking along the coast at low tide.

- **Livestock handling:** Boys and young man are sometimes responsible for the family livestock. This mostly entails moving a couple of tethered cows or a few goats from one patch of grass to the next few times a day, and bringing them to the river to drink.

- **Fishing:** Most adult men in the village practice some sort of fishing at least occasionally. Boys start to accompany their fathers and uncles sometimes as early as 12 or 14, but often don't start participating until they reach 16 or 18. Different types of fishing are practiced in the area, from diving alone with a mask in shallow water to capture small fish named *ngogo*, to participating in larger fishing parties together with men from other villages on motor boats. But most boys and young men in our sample practice small scale fishing, from dugout canoes and small boats.

- **Bird hunting:** This activity is practiced by many young and older boys with a variety of techniques. These include roaming the *shambas*, the cultivated fields around the village, with a slingshot trying to bring down little birds; using glue extracted from a local vine to which birds get stuck, once sticky sticks are placed on frequently visited branches; and also placing snares in the forest to capture terrestrial birds such as ibises.

- **Hunting with dogs:** Older boys and young men sometimes leave the village in noisy hunting parties accompanied by a variable number of dogs. These trips are usually aiming to kill monkeys or civets that attack the village chicken, and in general animals that are considered a pest.

- **Diving:** Some individuals practice some form of spearfishing, with or without oxygen tanks. These are usually older boys and young men, who bring fishing to an advanced stage.

- **Algae farming:** Many women and girls farm algae destined to the Asian food market. This is a very tiring activity, but one of the few commercial activities that women practice with some frequency. It does not require much knowledge, but spreading, collecting and drying the algae takes a lot of time.

- **Cloves picking:** During harvest season, most young boys and girls are hired by the families owning clove trees to pick the little buds these plants produce. This activity is limited to younger, lighter individuals who do not break the fragile branches of the trees, and is quite dangerous. But it is one of the main sources of cash for many families and most children engage with it for many years.

- **Household chores:** With this expression we include all sort of household-related tasks, such as washing clothes or pots and pans at the stream, or cooking. These tasks are mostly practiced by girls, although some boys contribute with some smaller tasks, such as bringing water from the well.

- **Agriculture:** Farming seems to be the main occupation of most families: tilling the rice field, planting stalks of cassava, weeding and harvesting peanuts are rotating as main activities depending on the season. In these families, most people of all ages contribute to agricultural tasks, as farming is not automatized and hence it is very labor intensive.

9.4 Freelist, questionnaire and image recognition. Collection and differences in results.

All interviews were conducted by IP in the village of Bandarikuu, Pemba. Children who volunteered for the interviews either showed up by themselves at the research station (hence the almost complete sampling of children living in the neighboring houses) or were encouraged to participate by friends and siblings (who received no remuneration for this). Participating children received a pen/pencil and a candy after the end of the interview, but such small gifts were handed out by the researchers regularly to all children, so participation did not guarantee any special treatment. Interviews were carried out in a (relatively) protected location to reduce probability of interviewees being influenced by other participants' answers. Audio recordings of (almost) all interviews are also available.

9.4.1 Freelist, details

Interviewees were asked to freely list all living creatures they knew with the question 'Could you tell me all the living creatures you know which live in the fields and village/forest/sea. All the animals, birds, fish, critters, trees and plants you can think of (Swahili: *"Unaweza kuniambia viumbe vyote vinaoishi shambani na kijijini / msituni / pwani na baharini. Wanyama wote, na ndege, samaki, wadudu, miti na mimea wewe unaweza kukumbuka'*). The question was repeated at least three times -once referring to fields and village, once to forest and once to the sea- to ensure that interviewees would mentally explore all environment types. The same question was also repeated if the interviewee seemed confused or paused for a very long period of time. Listing all types of creatures (animals, birds etc) also aimed at encouraging the interviewee to search among all living creatures. The words listed are culturally relevant for categorizing animals and plants: *wanyama* refers to land mammals and other ground animals, *ndege* are all birds and bats, *samaki* includes fishes and other marine animals such as whales and dolphins, *wadudu* is a word for pests, and includes most insects, but also often other animals percieved as pests such as rats and snakes, *miti* and *mimea* are respectively trees and shrubs. The freelilst section of the interviews was terminated when the interviewee stated he/she couldn't remember any more creatures or if the interviewee did not add any new item even after being primed multiple times by repeating the question.

The possible presence of a ceiling effect, i.e. interviewees reach a maximum number of items listed because of other reasons than knowledge, was a concern when designing and carrying on the interviews. A ceiling effect would be problematic especially if it correlates with age, e.g. younger individuals could get bored earlier and stop listing creatures even if they actually know more of them. We do not believe this to be the case, as the total number of items listed by individuals does not seem to differ among age groups (see Table S1). The maximum number if items listed is around 300 in all age groups, the highest number of items has been listed by a 12 years old individual. In general, most individuals kept listing items even when they had exhausted all their knowledge on the natural environment, by repeating items already listed or by listing objects and other non living things.

To deal with the repetitions and non coherent answers, and more generally to make sure the dataset was correctly representing knowledge of individuals, each item listed has been assigned to one of the following categories *wanyama* (W), *ndege* (N), *samaki* (S), *wadudu* (D), *miti/mimea* (M). Lists of items can be found in the GitHub repository, including those considered 'not_a_creature' and hence removed from the dataset. The list has been additionally reviewed by three separate local specialists (Massoud Bakar Massoud–Head of Planning and Administration at DFNRNR Pemba, Bakar Makame Khamsis–field assistant, and Haji Masoud Hamad–DFNRNR) and as well as by Dr. Martin Walsh (Adjunct Professor in the School of Business Studies and Humanities, Nelson Mandela African Institution of Science and Technology (NM-AIST)), expert of Swahili culture and ecological lexicon.

After cleaning, the list comprises 708 items, of which 201 named by a single individual. The item named by most individual is the mango (tree and fruit associated), named by 82 individuals. Table S2 reports the number of items listed by participants of different age groups after cleaning of the data.

9.4.2 Questionnaire, details

The 50 questions in the questionnaire have been developed in collaboration with the personnel of the DFNRNR in Pemba and through focus groups involving several adults in the village of Bandarikuu. The full list of questions, in English translations, is available in table S4, while the original Swahili version can be found in the GitHub repository. Table S3 shows some descriptive statistics divided by decade relative to the questionnaire.

	All ages	Aged <10	Aged 10-20	Aged >20
n	94	27	64	8
Max n items listed	342	300	342	275
Min n items listed	7	7	18	59
Mean n items listed	132.8	110.5	133,1	196.1
SD n items listed	78.0	86.5	69.3	74.2

Table S1: Total number of items listed in freelist for the whole sample and by age group (including repetitions and non coherent items).

	All ages	Aged <10	Aged 10-20	Aged >20
n	93	27	64	7
max n items listed	231	110	195	231
min n items listed	3	3	9	51
mean n items listed	85.77	51.93	91.84	150.62
sd n items listed	47.1	30.56	39.82	60.24

Table S2: Sample sizes and descriptive statistics for freelist data

	All ages	Aged <10	Aged 10-20	Aged >20
n	93	27	64	7
max % questions correct	0.82	0.64	0.82	0.76
min % questions correct	0.16	0.2	0.16	0.62
mean % questions correct	0.56	0.5	0.56	0.69
sd % questions correct	0.12	0.11	0.11	0.05

Table S3: Sample sizes and descriptive statistics for questionnaire data

QN number	English translation	Right answer	Question type
1	Do bats make eggs(A) or give birth(B)?	B	forest
2	Do hyraxes stay on the floor only(A) or climb trees as well(B)?	B	forest
3	Helmet shells(A) or mud whelks(B) live in between the mangroves?	B	shell
4	Can catfishes swim in the sea(A) or only in the river(B)?	B	forest
5	Do vultures eat meat(A) or fruits(B)?	A	forest
6	Mbaazi(A) or beans(B) is a vine?	B	farming
7	Do kingfishers eat insects(A) or fish(B)?	A	forest
8	When scared, the squid releases ink. True(A) or false(B).	A	fishing
9	Per single event, do sharks give birth to about five(A) or fifty(B) offspring?	A	forest
10	Green spotted snakes are dangerous. True (A) or false(B).	B	forest
11	Which season is good for cultivating seaweed, long dry season(A) or long rainy season(B)?	A	farming
12	Do sea turtles return to their beach every six months(A) or one-two years(B)?	B	fishing
13	Who eats snails, the civet(A) or the mongoose(B)?	B	fishing
14	Civets eat the meat of oil nuts. True(A) or false(B).	A	forest
15	Bushbabies sleep at night(A) or during the day(B)?	B	forest
16	Mkuu wa usiko is used as a medicine for the head(A) or the stomach(B)?	B	medicine
17	The territorial vervet monkey male lives with his grandchildren(A) or by himself(B)?	B	forest
18	Cone shells have venom. True(A) or false(B).	A	shell
19	Bushbabies, when eating bananas, eat one at a time(A) or they ruin the whole batch(B)?	A	farming
20	Which bat uses its ears to fly, flying fox(A) or banana bat(B)?	B	forest
21	Bats produce excrements out of their mouth. True(A) or false(B)?	B	forest
22	Groupers have many(A) or no(B) teeth?	A	fishing
23	Which banana is good for cooking, mkono wa tembo(A) or kukusa(B)?	A	farming
24	Rays sleep between the corals(A) or in the muddy patches(B)?	B	fishing

QN number	English translation	Right answer	Question type
25	Whales travel to Pemba to give birth(A) or to rest(B)?	A	fishing
26	The mgulele tree makes fruits that resemble java plums(A) or custard apples(B)?	A	farming
27	Chaza(A) or tondoo(B) is hard to crack open?	B	shell
28	Hyraxes have no fingers(A) or have three fingers(B)?	B	forest
29	The mti ya ulaya tree grows in the forest border(A) or in its center(B)?	A	forest
30	Which plant is a good medicine to kill fly larvae, msindu(A) or forest tonga(B)?	B	medicine
31	Shrimps can swim in the sea. True(A) or false(B).	A	shell
32	Which wood is good to build boats, mgulele(A) or mkarati(B)?	B	building
33	Geckos can be which color, white(A) or green(B)?	B	forest
34	Tunas reach the size of a cow(A) or of a child(B)?	A	fishing
35	After planting sweet potatoes, two(A) or four months(B) are needed before harvesting?	B	farming
36	The octopus is fished with fishing spears(A) or with nets(B)?	A	fishing
37	Kichachuli is the offspring of the vervet monkey(A) or of the civet(B)?	A	forest
38	Sea turtles travel in the cold(A) or hot season(B)?	A	fishing
39	Which shark is dangerous, nyambrani(A) or charawanzi(B)?	B	fishing
40	When the tide is low, can you find cowrie shells(A) or razor shells(B) on the beach?	B	shell
41	Swifts travel to Europe. True(A) or false(B)?	A	forest
42	Ducks lay eggs in the short(A) or long(B) rainy season?	B	forest
43	Which fruit is produced once a year, mkorosho(A) or java plums(B)?	A	farming
44	The offspring of dolphins drink milk. True(A) or false(B).	A	fishing
45	Which coconut lives a longer life, the cultivation one(A) or the normal one(B)?	B	farming
46	The caterpillar is the offsping of the butterfly. True(A) or not true(B).	A	forest
47	Lion fishes are dangerous. True(A) or false(B).	A	fishing
48	Frogfishes swim(A) or walk on the bottom of the sea(B)?	B	fishing

QN number	English translation	Right answer	Question type
49	Of the plant mjafari, the roots(A) or the leaves(B) are used as medicine?	A	medicine
50	Which period is good for panga: beginning(A) or mid(B) moon?	A	shell

Table S4: The English translation of the 50 questions asked in the questionnaire section of the interviews, together with correct answer and type of question.

Figure S2: First image shown to interviewees during the image recognition task. It contains three creatures to recognize: a snake, an ant on the branch and the plant itself.

9.4.3 Image recognition, details

The image recognition task was the last task of each interview. The interviewees were shown 27 pictures on a 9.7" screen, always in the same sequence, as we were interested in variation among individuals, more than in characteristics of the picture themselves. Each picture shows from 1 to 7 species of plants and animals. The interviewees were asked to point at and name all the living creatures in the picture. For example, in Figure S2, there are three organisms that interviewees can name: a snake, an ant on the branch and the plant itself. Any answer was recorded and, in a second stage, all the answer have been reviewed in collaboration with adult volunteers in the village. All possible answers were divided in acceptable (even if not strictly correct) vs. non acceptable. For example, the word *nyoka*, meaning 'snake', was not considered acceptable for the snake in the picture, as it is too generic. Instead, *gangawia* was considered an acceptable alternative to the correct *ukukwi* as the two species are similar. One plant was excluded from the analysis as no interviewee was able to recognize it. Table S5 shows some descriptive statistics divided by decade relative to the image recognition task.

	All ages	**Aged <10**	**Aged 10-20**	**Aged >20**
n	93	27	64	7
max % images recognized	0.68	0.38	0.61	0.68
min % images recognized	0.11	0.11	0.11	0.35
mean % images recognized	0.34	0.22	0.36	0.49
sd % images recognized	0.12	0.07	0.1	0.11

Table S5: Sample sizes and descriptive statistics for image recognition data

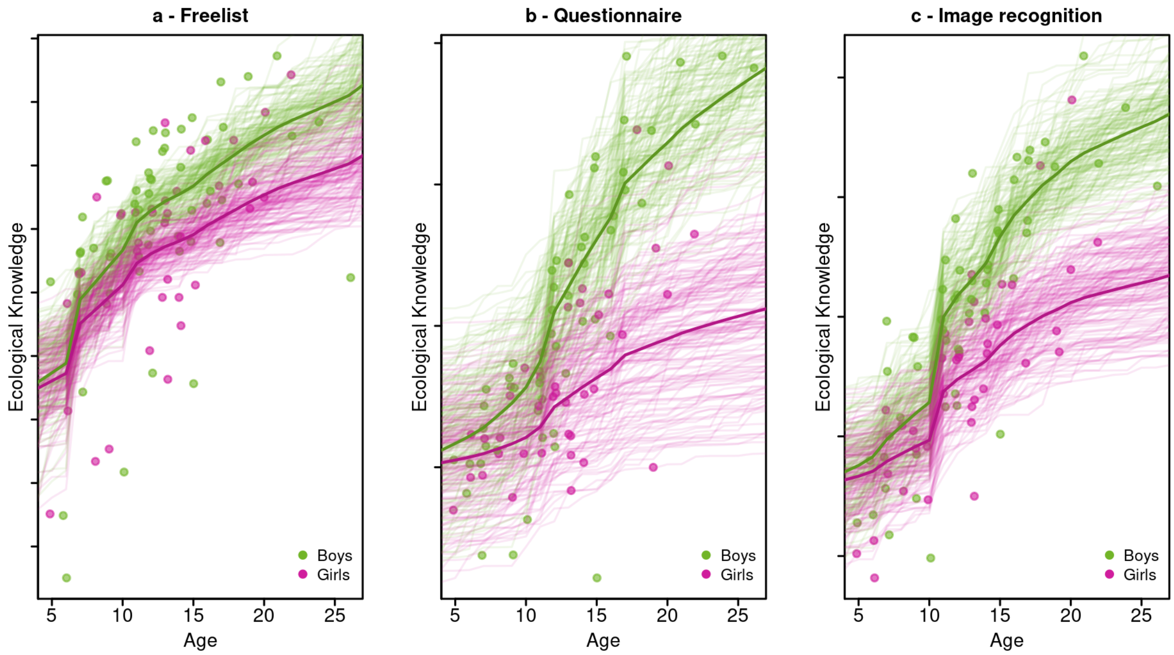

Figure S3: Individual knowledge and predictions per age and sex groups estimated from single types of data: panel *a* is estimated from freelist data, panel *b* from the answers to the questionnaire only and panel *c* from the image recognition task.

9.4.4 Notes on methods. Comparisons of results from different question types.

The three types of data collected during the interviews have been combined in the main analysis to yield a total estimated knowledge measure for each individual. Here, we present the separate results from the three types of data and discuss the pro and cons of each in terms of methodology.

The freelist represent the majority of the data points. As each of the 708 items was treated as a question that could have been answered correctly–i.e. that item was named–by each individuals, freelist items alone represent 85.4% of all data for each individual. This means that freelist data provide the majority of information for the main model in the paper. The estimates from freelist data are shown in Figure S3 panel *a* for one and S4 for three dimensions. Panels *b* and *c* in figure S3 show knowledge estimated from the questionnaire and image recognition data only, respectively. As expected, they contain approximately the same information as the full model, to which each contributes: knowledge increases with age and there are differences between the sexes. Both the questionnaire and the image recognition, though, estimate much higher knowledge for boys than girls. This could be due to a bias in the way the questionnaire and image recognition were constructed, given that, compared to the freelist, the agency of the researcher was much more important in these two types of question. In general, knowledge of individuals as estimated with each of the three question types alone positively correlates (see figure S5, suggesting that they describe overall the same individual characteristic. Moreover, if looking at the dimension analysis, we can find the three dimensions of knowledge described in the main text also in the freelist data (Figure S4) and in the questionnaire and image recognition, although for these two types of questions the number of data points is not sufficient to estimate clear age and sex-specific trajectories.

IRT models allow to evaluate difficulty and discrimination for each item (creature named, question in questionnaire and creature in images). We can hence compare how effective are the types of question in helping us estimate knowledge. We can see the curves describing difficulty and discrimination of items in freelist, questionnaire and image recognition in figure S6, left, middle and right

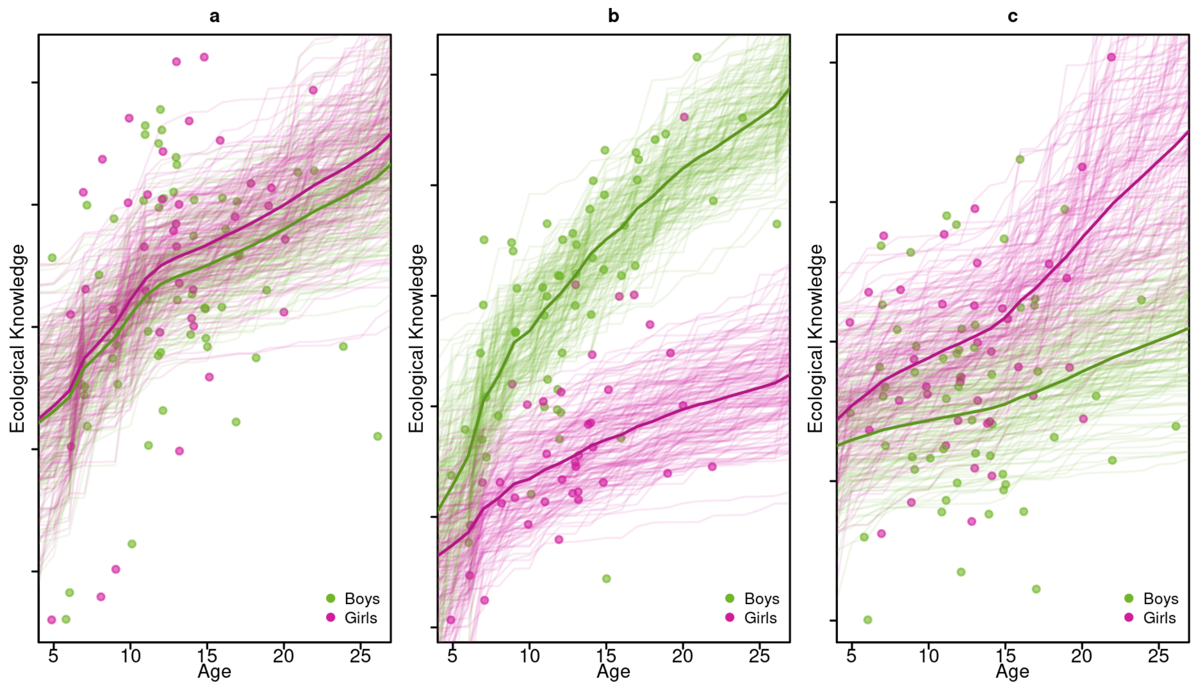

Figure S4: Three dimensions of knowledge as estimated from the freelist data only.

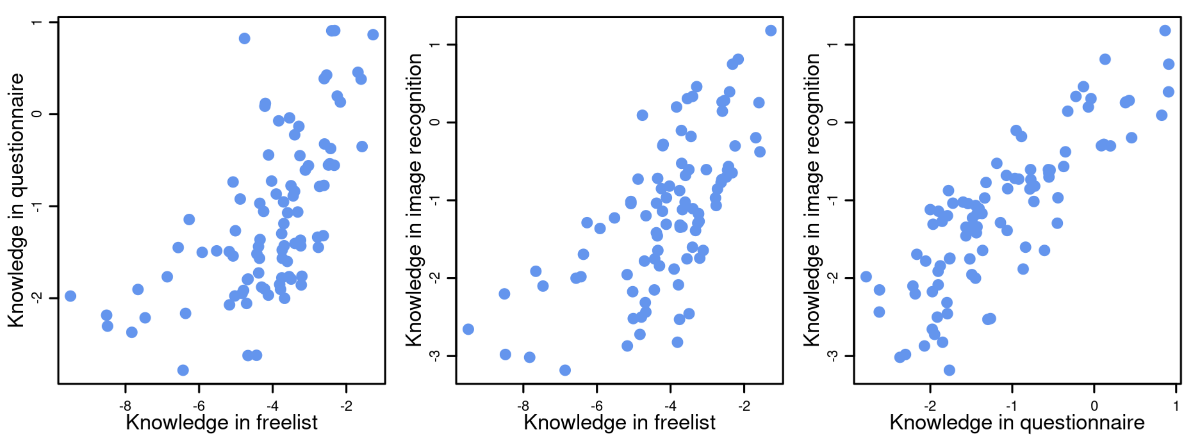

Figure S5: Knowledge estimated by each question type alone.

panel respectively. The position of the center of each curve on the x axis represents the difficulty of a question, while the inclination represents the discrimination. We can see that the three types of question do a decent job of helping us estimate knowledge. Most questions are quite discriminatory and vary in difficulty so that each type of question alone could potentially be enough for estimating knowledge, at least in one dimension (note that these plots describe parameters from Model 1 described in the main text, not from the models applied to each type of question separately).

From these observations we can draw some general conclusions. First, all the types of data collected are a good description of knowledge of individuals. Different types of questions produce comparable results, so we expect to be observing some individual characteristics that we could define ecological knowledge. Second, despite this, the type of data are not interchangeable and the decisions of the researcher might affect knowledge estimation. We observe this by looking at how the results from the freelist differ from those of the questionnaire and image recognition (in figure S3), which are more impacted by data collection strategies (devising questions, choosing images). Third, the freelist appears to be the most effective way of collecting data, as it allows to collect many more data points with less effort (the design of questionnaire and the choice of images are very time consuming). Maybe more interestingly, freelists are less conditioned by the choices of the researcher, potentially yielding results that better describe reality. Post collection treatment of data has larger impact in the case of freelisting, but it is less likely to be gender biased, as well as easier to track and evaluate post-hoc.

9.5 DAG: Factors influencing Knowledge

Explicitly defined causal relationships described below are at the basis of both data collection and the choices guiding the analysis, including simulated data. Here follows a description of the factors we expect are exerting an influence on ecological knowledge, with their causal implications, as shown in figure 3:

Age Knowledge changes as individual get older. We expect to observe an increase with age, with individual differences that can emerge as a result of several factors. Age's total influence on knowledge includes several pathways. Age stands for increasing cognitive abilities of human brains that allow to store and manage information, and also for experiences, observations and instructions individuals receive as time passes. These unmeasured factors mediate the effect of age and are represented by the circled U in the DAG in figure 3 (Age → U → Knowledge). These remain unmeasured factors explaining variation of Ecological Knowledge with age. Other, measured, factors mediate the effect of age on Ecological Knowledge: as individuals age, the time they spend performing specific activities varies (Age → Activities), they start or stop going to school (Age → Schooling).

Sex We do not expect a direct effect of sex of individuals on knowledge, as we do not believe there are innate differences in the ability to learn. Rather, we think of gender differences as influencing both the probability at which activities are performed, some tasks being typically done by girls and other by boys, as well as, potentially, school attendance (Sex → Activities and Sex → Schooling). Through these indirect pathways, differences in ecological knowledge can originate because of differential engagement in gendered activities.

Activites In our expectations, the activities children perform more frequently have a strong influence on knowledge (Activities), both because of the exposure to the relevant information while performing the activity and of the increased returns derived from learning these information. When not in school, children in Bandarikuu spend their time doing domestic or farming chores, hunting, collecting seashells, fishing or playing, at different probabilities as they get older. Some of these activities, such as hunting, are expected to have a positive effect on the knowledge of the natural environment.

Family Family context, such as the presence of parents or older siblings, is thought to influence what and how much children know (Family → Knowledge). Engaging in activities with parents, can have an impact on the acquisition of knowledge related to those activities. Also, the simple presence of adults in the households can have a similar effect just by exposing children to conversations about subjects. Domestic situations can have large effect on the activities children perform (Family →

Activities). Older same sex siblings can have positive effect on knowledge by representing a model and introducing individuals to specific activities (imagine an older brother teaching to shoot with a slingshot) or have negative effect if they fulfill a specific role in the household (only one person is sufficient to take care of the cattle of a family). Younger siblings could change the expected time allocation into activities (for example reducing the time one can spend roaming the forest in exchange for time spent taking care of them). Family context and attitudes can also influences the effort children are allowed, or encourage, to devote to formal education (Family → Schooling).

School The time and effort children invest in formal education can have different effects on knowledge of the natural environment. On the one hand, it can increase the the amount of information individual can manage. Ont the other hand, though, it can imply an opportunity cost by reducing the contact with natural environment (Schooling → Knowledge).

Summarizing, we expect knowledge to increase with age, to vary in accordance to which activities are performed, as certain activities favor the learning of ecological knowledge; by family, as access to knowledge depends on the access to older individuals able to transmit it, for example; and probably by access to schools, which provide certain types of knowledge but not others.

9.6 Model details

As mentioned in the main text, the models used are composed of two main parts: an IRT part that estimates knowledge from the answers to questions, and a generalized multilevel linear model section that estimates the effect of various predictors on individual knowledge measures. Here we will describe more in detail the functioning and parameters of the two parts.

The Item Response Theory models estimate a measure of knowledge for individuals in a latent scale (i.e. the absolute values have no meaning, what matters are the measures in relation to each other) in association to question parameters. It works by estimating an S shaped logistic curve for each question which describes the probability of answering the question correctly over the latent knowledge space (see figure S6). An individual's position on this axis is estimated according to which questions were answered correctly. Questions vary in difficulty b and discrimination a, which is a strength of IRT models in the fact that they do not assume all the questions to be equivalent. Difficulty is represented by the position of the curve for each question on the x-axis–harder questions are placed to the right while easier questions are on the left. The inclination of the curve indicates how discriminatory each question is, i.e. how much it can help distinguish knowledgeable people from not-so-knowledgeable ones. Most questions have quite good discrimination–their curve raises fast, drawing a sharp S–but some are not very good at discriminating. For example, a very smooth S shape is very evident for a couple of items in the image recognition plot in the right panel (figure S6). In addition to difficulty and discrimination, questionnaire items have a third parameter: 'pseudo-guessing' c. These items represent a choice between two options, which means that individuals could have 50% probability of answering correctly just by answering randomly. But in practice individuals do not answer at random, and baseline chance of answering correctly varies from question to question. To deal with this problem, a 'pseudo-guessing' parameter estimates the probability of answering the question correctly by chance. The curves in the middle panel (figure S6), indeed, do not start from zero, but rather from the estimated c value, which means that even individuals with very low knowledge have a probability greater than zero of answering correctly.

$$
\begin{aligned}
Y_{i,j} &\sim \text{Bernoulli}(logit(p_{i,j})) \\
p_{i,j} &= c_j + \frac{1 - c_j}{1 + \exp^{-a_j(K_i - b_j)}}
\end{aligned}
$$

Knowledge of individuals is represented in figure S6 by the position on the x axis of the blue dots superimposed to the curves (the position at y = 0.5 is arbitrary). The model estimates it simultaneously to the question parameter, so that each observation (the answer of the ith individual to the jth question) contributes to all these parameters. Knowledge is the same in all three panels, as the three data types contribute to the estimation of a single measure of knowledge.

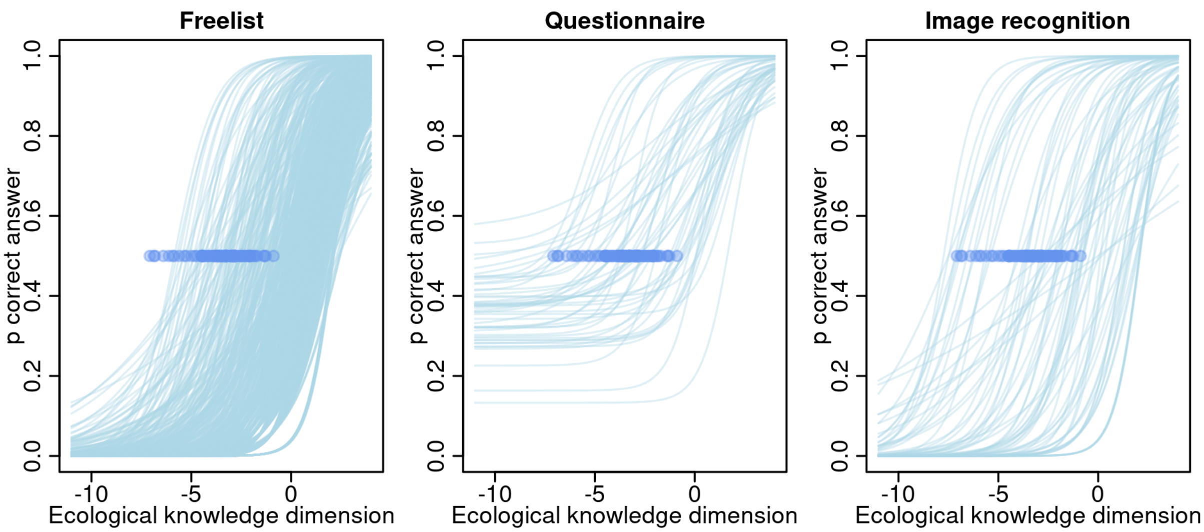

Figure S6: Logistic curves describing difficulty and discrimination of items in freelist, questionnaire and image recognition tasks respectively. Estimated knowledge of individuals has been superimposed as dots placed at the arbitrary y value of 0.5. The x axis describes the latent knowledge dimension.

The values on the x axis are those of the latent dimension of knowledge, which is scale-free, so that they have no absolute meaning. In our model, most of the knowledge estimates are negative numbers, but this does not mean that individuals have negative knowledge. Rather, the whole latent space is pushed into the negative area of the plot by the fact that there are many freelist items with a high difficulty. These are all the items named only once, for which the model cannot estimate different difficulty–because they are all named once–but which clearly the majority of individuals could not name. These are visible crowding the area around 0 in panel (a).

Once knowledge is estimated in the IRT part of the models, the effect of individual and social characteristics are estimated in the second part. Each year contributes a proportion of the total age effect. Figure S7 shows the proportion of the total age effect relative to each year. The estimates are relative to the δ_y parameters, with $\delta_{y=1}$ having value of zero. The actual effect of age for an individual of age y is the proportion of β equal to the sum of all δ_y parameters for the younger ages. Because the sum of all δ_y values is equal to one, individuals with maximum age have the full effect of β. Note here that variation for posterior values for δ_y decreases after age 20, probably due to a reduction in sample size for those ages. A detailed description of the statistical approach to ordered categorical variables used here can be found in Chapter 12.4–pages 391 to 396–of *Statistical Rethinking* (McElreath, 2020).

In addition to the models described in the main text, we present here the models estimating the effect of birth order and of school participation. In both cases, the effect of the factor is modeled as an ordered categorical, as in the case of age. Notice that schooling is modeled as the effect of not participating to school (each class not completed gives an increasing proportion of the total effect of not participating), but other ways of describing school attendance yielded the same results.

$$K_i = \alpha_i\sigma_1 + \eta_h\sigma_2 + \beta_s \sum_{y=1}^{Age_i} \delta_y + \lambda \sum_{Sib=1}^{Sib_i} \delta_{Sib} \tag{1}$$

$$K_i = \alpha_i\sigma_1 + \eta_h\sigma_2 + \beta_s \sum_{y=1}^{Age_i} \delta_y + \zeta_s \sum_{Sch=1}^{Sch_i} \delta_{Sch} \tag{2}$$

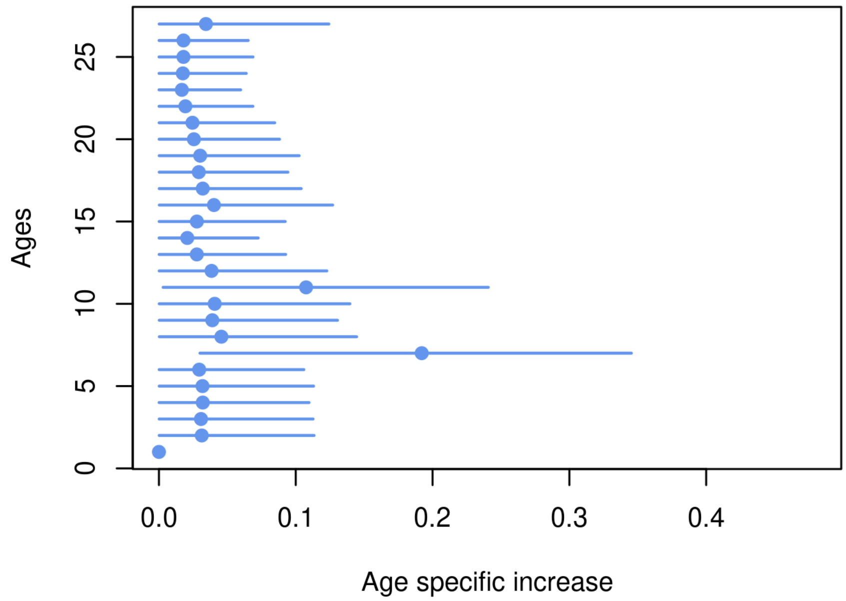

Figure S7: Proportion of the total effect of age relative to each year. The bars represent 89% of the posterior estimates around the mean of δ_y parameters, shown by the central point.

All models are coded including a different parameter per each dimension.

$$Y_{i,j} \sim \text{Bernoulli}(logit(p_{i,j})) \tag{3}$$

$$p_{i,j} = \sum_{d=1}^{d=D} p_{i,j,d} \tag{4}$$

Moreover, we need to mention that some adjustments were necessary to deal with the fact that all knowledge measures resulting from the IRT part of the model were in the negative space. In order to allow the function describing the categorical effect of age and the other parameters to move in this space, a global intercept ω was included in all models. This parameter moves the baseline of the functions below zero, to allow the other parameters to be positive. The prior for this parameter was negative (mean = -5), but very weak (sd = 3). The posterior values for ω are around -6 in Model 1 with one dimension, but vary in the models with more dimensions from -2.1 to -3.7. These measures have no biological meaning, but allow the other parameters to be comparable.

The full Model 1 used in the analysis looks like this:

$$
\begin{aligned}
&Y_{i,j} \sim Bernoulli(logit(p_{i,j}))\\
&p_{i,j} = \sum_{d=1}^{d=D} p_{i,j,d}\\
&p_{i,j,d},_d = a_{j,d}(K_{i,d} - b_{j,d})\\
&K_{i,d} = \omega + \alpha_{i,d}\sigma_1 + \eta_{h,d}\sigma_2 + \beta_{s,d}\sum_{y=1}^{Age_i} \delta_{y,d}
\end{aligned}
$$

9.6.1 Choice of priors

Priors for the parameters in the models were chosen in order to guarantee that the polarity of the latent axis would place people with more knowledge as having higher values than people with less knowledge. Simulated logistic curves describing difficulty and discrimination of question parameters with the priors chosen in for the models presented in the main text are shown in figure S8. These are

$$
\begin{aligned}
a_j &\sim \text{Half-Normal}(0,1)\\
b_j &\sim \text{Normal}(0,2)
\end{aligned}
$$

The same model has been run changing the priors for b to a Normal distribution with mean zero but standard deviation of 1 or 3. The resulting values of K and curves for question parameters can be seen in figure S9, where the curves and knowledge values in panel a are the product of Model 1 (equation 3) with prior $b \sim$ Normal (0,2), whereas panel b and c are respectively from the same model but with priors $b \sim$ Normal (0,1) and $b \sim$ Normal (0,3) respectively.

Finally, Figure S10 represents the priors for the curves that relate age and sex to knowledge. The priors are very vague and allow many possible relations in the space. In green and purple are shown the average values of the posteriors for these curves as estimated by Model 1.

9.6.2 Label switching

One problem introduced by the use of a Multidimensional IRT model is label switching. This happens when parameters are not uniquely defined and two or more parameters can 'exchange values'. In the case of Multidimensional IRT models, this happens because there is no order in the dimensions and parameters can assume in each dimension any of the possible values for that parameters. So, running the model over multiple chains could mean that the chains could settle on different values: chain 1 assigns the value 2 to the first dimension and 4 to the second, while chain 2 does the contrary. Each chain samples fine and clearly distinguishes between the dimensions, but the final result, that averages between chains, will show no difference between the dimensions. Luckily this behavior is easily recognizable by simply visually inspecting the chains, and can be avoided by using a single chain. Some label switching can remain when running a highly dimensional model (traceplots show some switching in the model with 5 dimensions, for example), but up to three dimensions no label switching has been observed in our models. These have been run on longer chains to ensure comparable results to those of multiple chains.

9.7 Simulation

Preliminary to the data collection, we simulated data in silico to test the models and inform data collection procedure. The simulation code is available in the GitHub repository. Several functional correlation between age and knowledge have been simulated, and the model used in the analysis– which includes age as a ordinal categorical predictor of knowledge with monotonically increasing

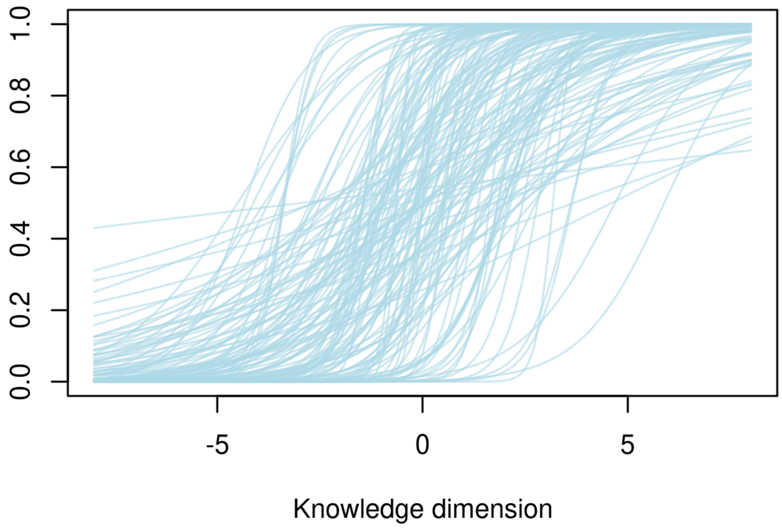

Figure S8: Simulated priors for question parameters in the IRT section of the model.

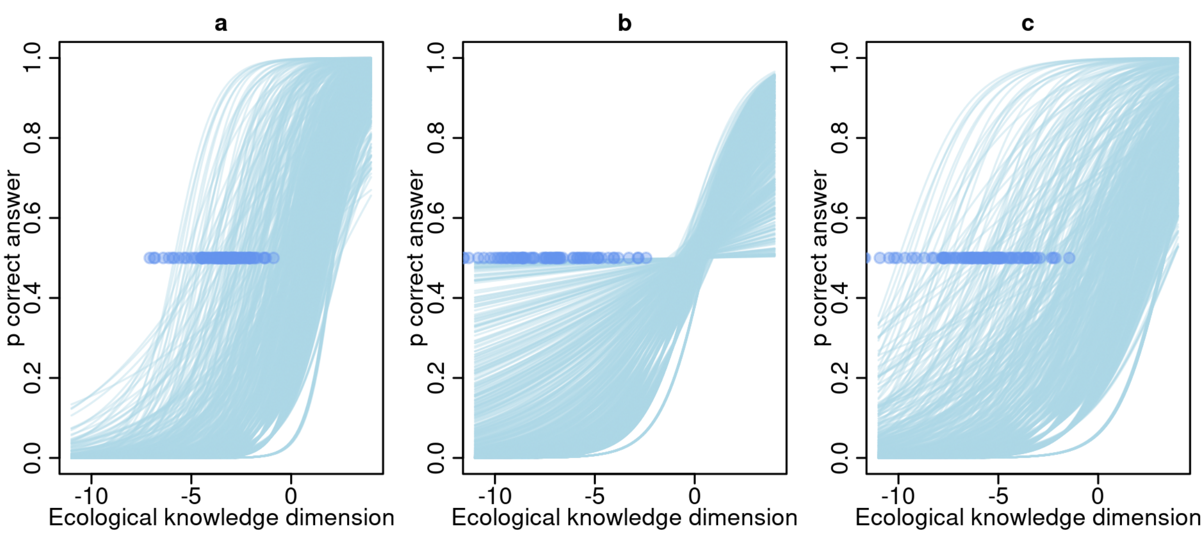

Figure S9: Knowledge of individuals and curves describing question parameters from the posterior distributions for three models fit with different priors for difficulty of questions. Panel a refers to the model fit with $b \sim$ Normal (0,2), which is the priors used in the models described in the main text. Panel b used $b \sim$ Normal (0,1) and panel c $b \sim$ Normal (0,3).

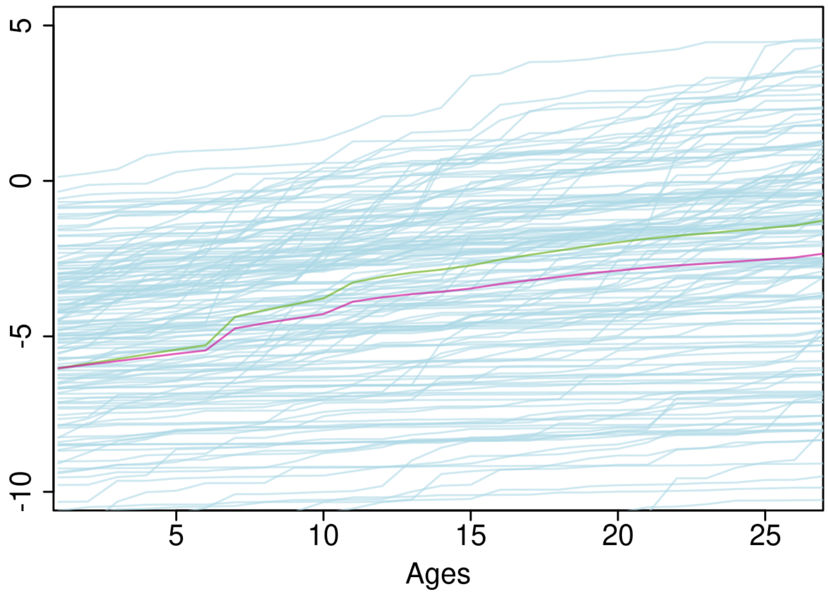

Figure S10: Simulated priors for individual parameters in the second section of the model. In green and purple are the averages of the posterior parameters depicted in figure 4, for comparison.

effect–has been able to recover the different shapes. Causal effect of activities, family composition and schooling have been simulated and tested.

The simulated data have been used–albeit in a previous version–to estimate the minimum number of interviewees necessary to recover the parameter values. If individuals were to name a maximum of 300 items in the freelist, 50 interviewees would have been sufficient to obtain reliable estimates of the parameters. Given that data collection in vivo is much less regular and less controllable than in silico, we roughly doubled the number of interviewees and that of questions.

9.8 Dimension analysis

Multidimensional IRT analyses can be used to assess the dimensionality or underlying latent variable structure of a measurement. Its results are comparable to those of a Factor Analysis, but IRT have several advantages, in particular they allow the latent individual trait and the question parameters to vary independently, whereas in Factor Analysis item difficulty is assessed as a function of the abilities of the sample, and the abilities of respondents are assessed as a function of item difficulty (Osteen, 2010). To test for dimensionality in our knowledge measure, we run the model described in the main text and above (equations S3, S4 and equation 3 in main text) forcing it to separate knowledge into multiple dimensions, from 1 to 5. We then calculated the WAIC values of each of these model fits to estimate how well they fit the data and compared them. WAIC values measure how accurately a model can predict new data, and are widely applied for model comparison. A model with lower WAIC value can make better estimates out of sample. When we compare WAIC values for models fit to different number of dimensions, we can expect an improvement in the WAIC score if adding a new dimension helps describe structure in the latent variable, i.e. helps explain variation. In figure S11 are shown WAIC values for models with one to five dimensions, divided by type of question (freelists, questionnaire and image recognition). This is because the structure of knowledge as described by the three different data types is different. The results from the questionnaire, in the middle panel, indicate no underlying structure, as the model with one dimension ('age_1') performs best: it has lower WAIC score and none of the models including more dimensions perform similarly

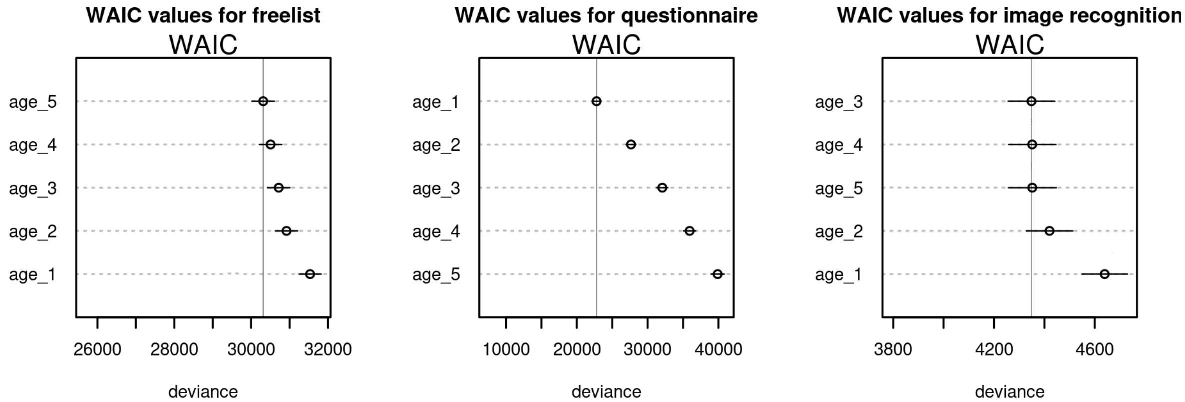

Figure S11: Comparison of WAIC values between models with variable number of dimensions, from 1 to 5. The first panel to the left is relative to freelist questions only, the middle refers to questionnaire results and the right panel is for image recognition. The vertical line marks the deviance value of the best model (top line), so that the other models can be easily compared.

well. Results from freelist and image recognition, instead, show that including more dimensions helps to better describe variation in knowledge. In particular, a model with three dimensions ('age_3' in the right panel) seems to best describe the image recognition data, and four or five dimensions perform similarly well. For freelists, increasing the number of dimensions keeps improving the fit of the model ('age_5' with five dimensions in the left panel has the lowest WAIC score, and models with four or three dimensions follow). This suggests that knowledge, the latent variable in our analysis, is structured in multiple dimensions, and that the different types of data are not equally good at allowing this structure to emerge (Supplementary section 9.4.4 offers some more insight on how data collection procedures can influence the results, especially in terms of dimensionality). It is not necessarily clear, though, which number of dimensions better describes the data. We hence approached the problem mainly descriptively and compared the results from the different models with up to 5 dimensions of knowledge.

Figures S12, S13 and S14 show the patterns of knowledge variation when subsetting knowledge in two, four and five dimensions respectively, and can be compared with figure 5 in the main text that shows results from a three dimensional model. From a visual inspection, we can see that three main patterns tend to be repeated, as described in the main text: a dimension in which both sexes learn at similar speed (General Knowledge dimension), one in which boys learn more than girls (Male-specific dimension) and some remaining variation (Other Knowledge dimension). If more than three dimensions are present, some of these patterns are repeated, so that for example, in four dimensions, there are two different dimensions that describe remaining variation. Based on these results we decided to describe three dimensions as the most helpful in order to tackle the subject of knowledge specialization.

In this process, though, interpreting the real life implications of these dimensions would be very important. In the main text, we offer an interpretation of the dimensions based on the difficulty parameters b_j for freelist items. This procedure is not standard for IRT models, and can only offer an insight on what these dimensions represent. To better understand how these dimensions correlate with different areas of knowledge, we should run analyses that can let differences between hypothesized groups of similar content emerge from the data (i.e. including predictors for question level parameters).

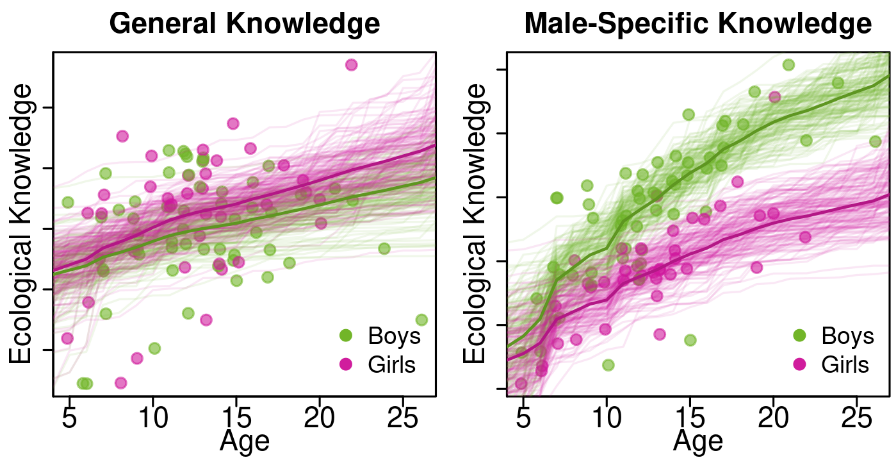

Figure S12: Individual knowledge K_i and predicted values by age and sex in two dimensions.

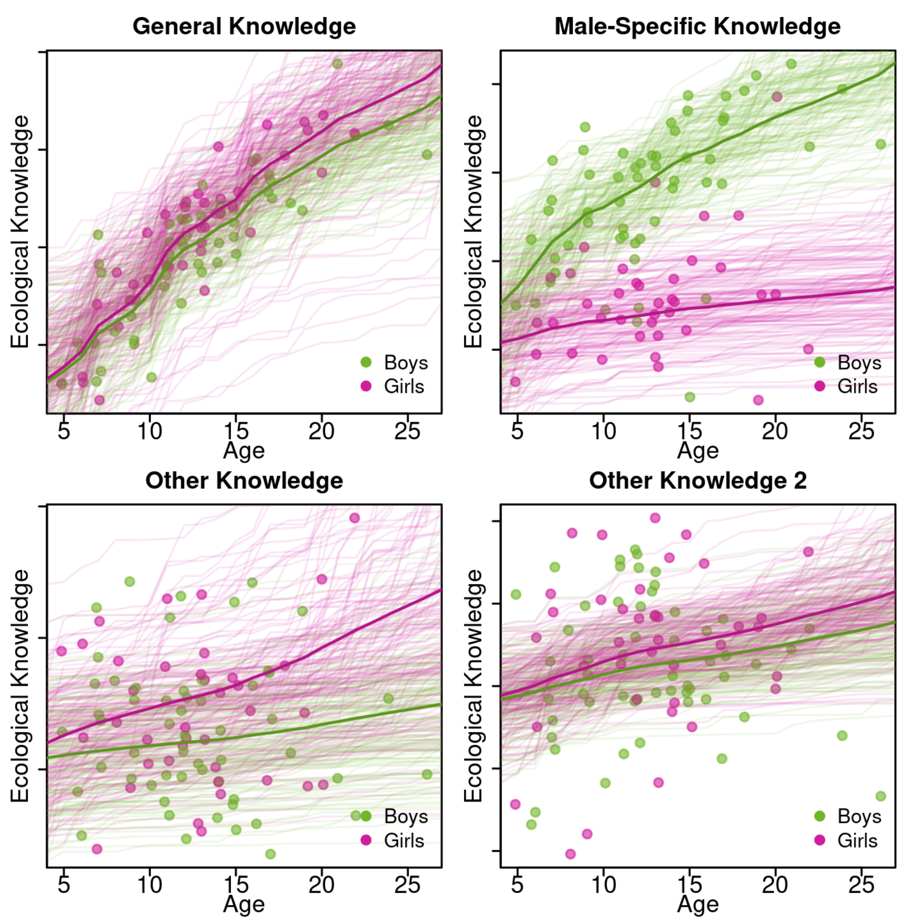

Figure S13: Individual knowledge K_i and predicted values by age and sex in four dimensions.

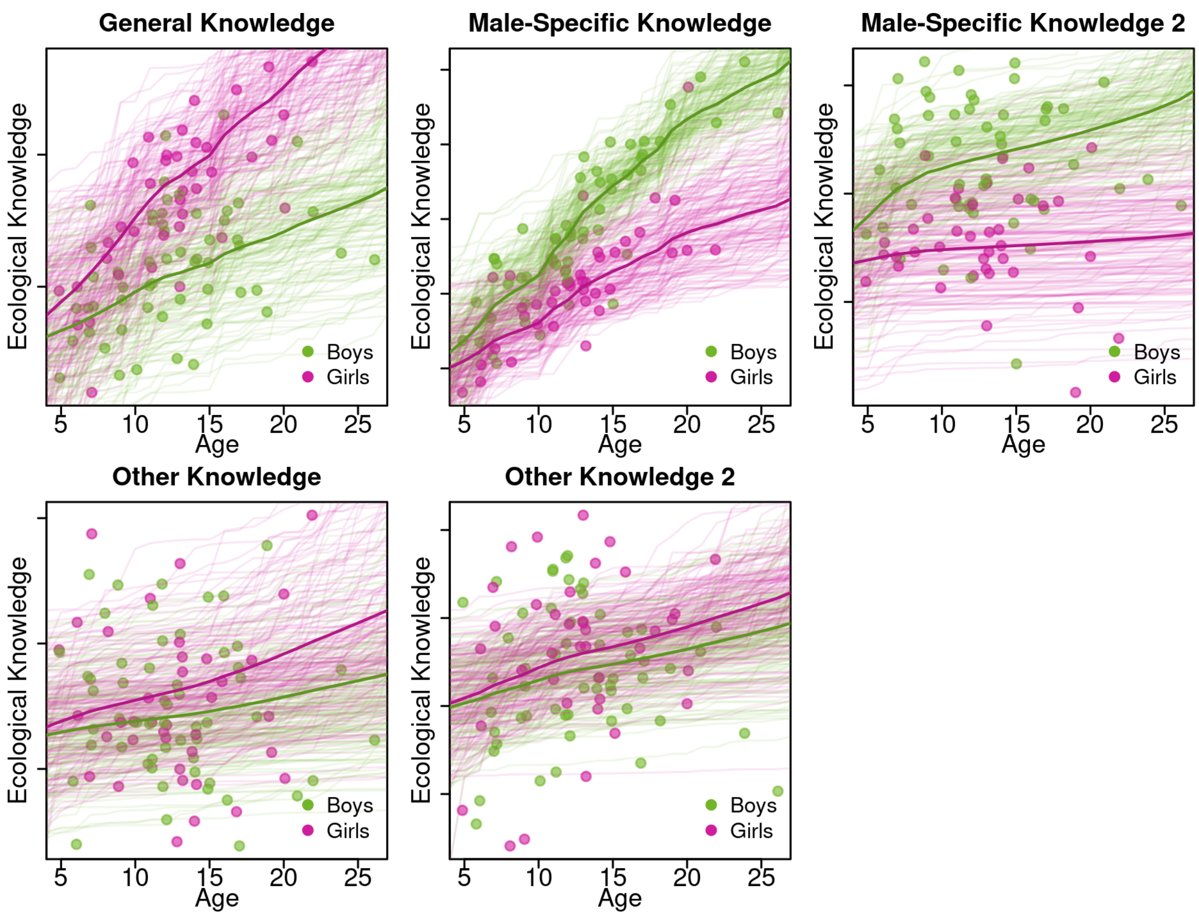

Figure S14: Individual knowledge K_i and predicted values by age and sex in five dimensions.

9.9 Other factors influencing knowledge

In the main text of the article we presented some results on the causal effects of some factors influencing knowledge. Here we will describe these results more in detail and present some plots that help discuss them.

9.9.1 Activities

Activities practiced are an important element for the development of knowledge. In the main text we saw that practicing an activity can shift the expected age for an individual of a certain knowledge of even more than one decade (see figure 6). Figure S15 shows the posterior distribution of the effects of each activity. We also mentioned that, once controlled by activities practiced, the difference in knowledge between sexes seems to disappear. Figure S16 shows estimated knowledge by age and sex, controlling for activities. Note that the distribution of the lines is mostly overlapping and the contrast, reported in table S6, show no difference between the sexes.

	mean difference	5% quantile	95% quantile
Model 1 unidimensional, fig 4	1.07	0.54	1.60
Model 1, 3 dim, general knowledge, fig 5a	-0.76	-1.31	-0.23
Model 1, 3 dim, male knowledge, fig 5b	2.71	2.09	3.41
Model 1, 3 dim, other knowledge, fig 5c	-0.57	-1.58	0.20
Model 1, freelist only, fig S3a	1.11	0.40	1.86
Model 1, questionnaire only, fig S3b	1.70	1.04	2.43
Model 1, image only, fig S3c	1.35	0.91	1.85
Model 1, freelist only, 3 dim, general knowledge, fig S4a	-0.26	-1.00	0.39
Model 1, freelist only, 3 dim, male knowledge, fig S4b	2.60	1.96	3.35
Model 1, freelist only, 3 dim, other knowledge, fig S4c	-1.42	-2.63	-0.43
Model 1, 2 dim, general knowledge, fig S12a	-0.54	-1.36	0.20
Model 1, 2 dim, male knowledge, fig S12b	1.87	1.38	2.41
Model 1, 4 dim, general knowledge, fig S13a	-0.31	-1.28	1.36
Model 1, 4 dim, male knowledge, fig S13b	2.54	1.55	3.37
Model 1, 4 dim, other knowledge, fig S13c	-0.87	-2.30	0.08
Model 1, 4 dim, other knowledge2, fig S13d	-0.37	-1.24	0.38
Model 1, 5 dim, general knowledge, fig S14a	-1.49	-2.43	-0.61
Model 1, 5 dim, male knowledge, fig S14b	1.85	1.03	2.68
Model 1, 5 dim, male knowledge2, fig S14c	1.32	0.57	2.05
Model 1, 5 dim, other knowledge, fig S14d	-0.51	-1.62	0.32
Model 1, 5 dim, other knowledge2, fig S14e	-0.35	-1.23	0.41
Model 2 unidimensional, fig S16	0.27	-0.57	1.09

Table S6: Mean difference between β_s values for males and females, plus lower and upper 5th percentiles. Values above zero indicate that male values are larger, below zero mean that female values are larger. Model results are described referring to figures in the main text and Supplementary Information.

	Mean difference in years	5% quantile	95% quantile
agriculture	-6.42	-17.52	1.15
household	-4.77	-13.38	0.23
cloves	-3.95	-13.70	1.87
diving	-2.36	-11.22	4.68
algae	-2.14	-8.46	2.56
game	-1.59	-8.98	3.94
birds	3.58	0.40	7.70
fishing	3.60	-2.50	9.79
livestock	3.72	-0.04	8.80
seashells	4.18	0.48	8.54

Table S7: Mean estimated difference between individuals non-practicing and practicing a certain activity, also plotted in figure S15.

9.9.2 Family

As discussed in the main text, no clear theory exist on which aspects of the social environment are the most important for ecological knowledge.

Looking at the random effects for household (figure S17), there doesn't appear to be any strong difference between families. Some households might be slightly less knowledgeable, on average, but the effects are not strong. Something more specific could be more important, as families are concerned.

Availability of models and sources of information for vertical and/or horizontal transfer is probably one of the most important aspects of it. In line with this expectation, presence of both parents in the household seems to have positive effect on knowledge of boys, but not of girls (see figure S18), who benefit from co-residing with mothers but seem to be penalized by the presence of fathers.

Other aspects of the family life seem to be less important. Birth order does not seem to have any effect (the total effect for birth order does not differ from zero and there not seem to be any direction in the effect). Also, firstborns do not seem to have any advantage/disadvantage, which could be the case if the firstborn had differential access to the parents, or if a firstborn had responsibilities that could prevent from engaging in activities such as hunting birds.

But further analyses looking at shared knowledge within an household or along friends networks could give more interesting results.

9.9.3 Schooling

Schooling does not seem to have a strong effect on knowledge of the natural environment. Figure S20 shows how many years earlier (or later) would an individual who does not go to school reach the same knowledge as a 20 years old individual who attended the whole cycle. No strong effect seems to be present, although girls who do not attend school might have higher knowledge of those who do. This could mean that increasing schooling in rural areas would not impact local ecological knowledge acquisition. As an alternative explanation,the effect of schooling might not be visible because we did not subset to locally relevant ecological knowledge, especially in the freelist, but rather accepted any answer relative to animals and plants. So, for example, animals they could have learned of at school were also accepted, and might have replaced local ecological knowledge,

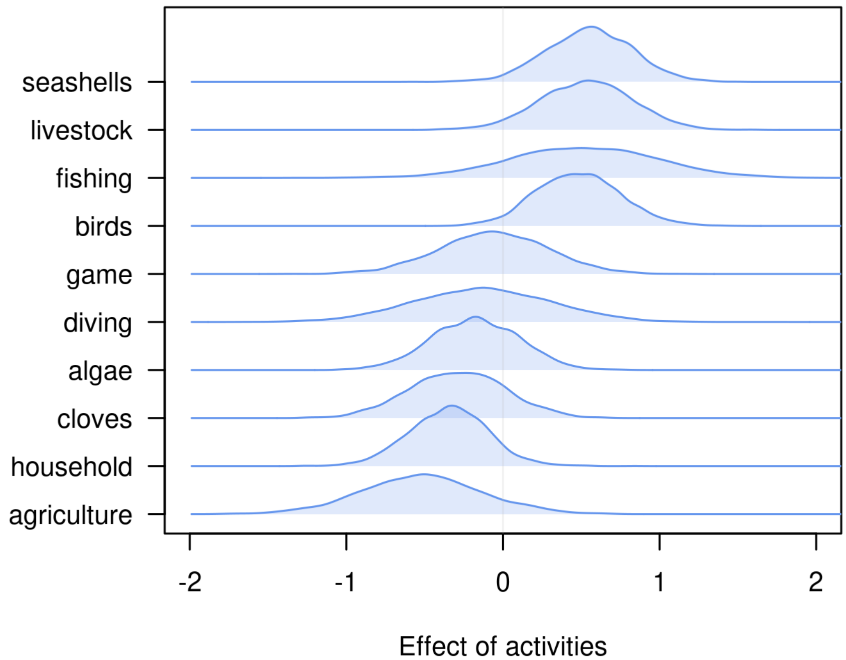

Figure S15: Posterior distribution of effect of each activity.

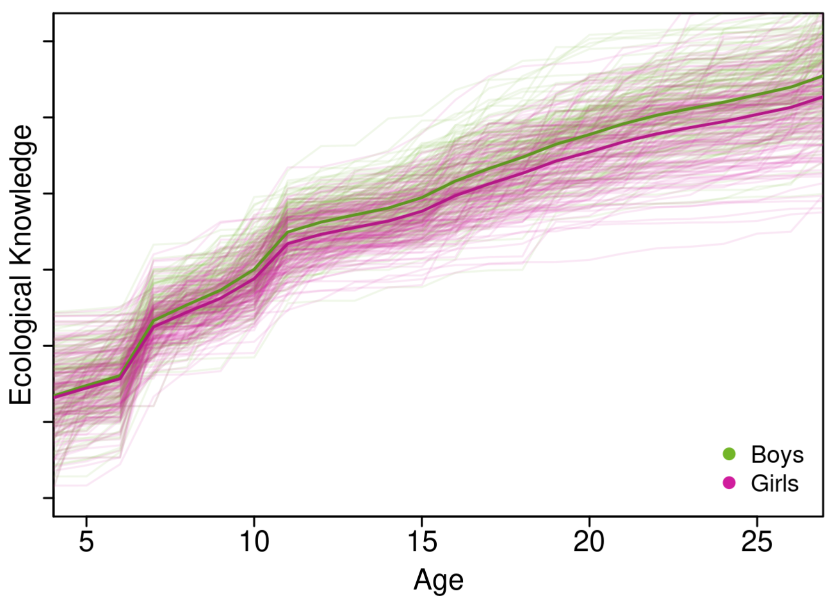

Figure S16: Knowledge estimated by age and sex, controlling for activities.

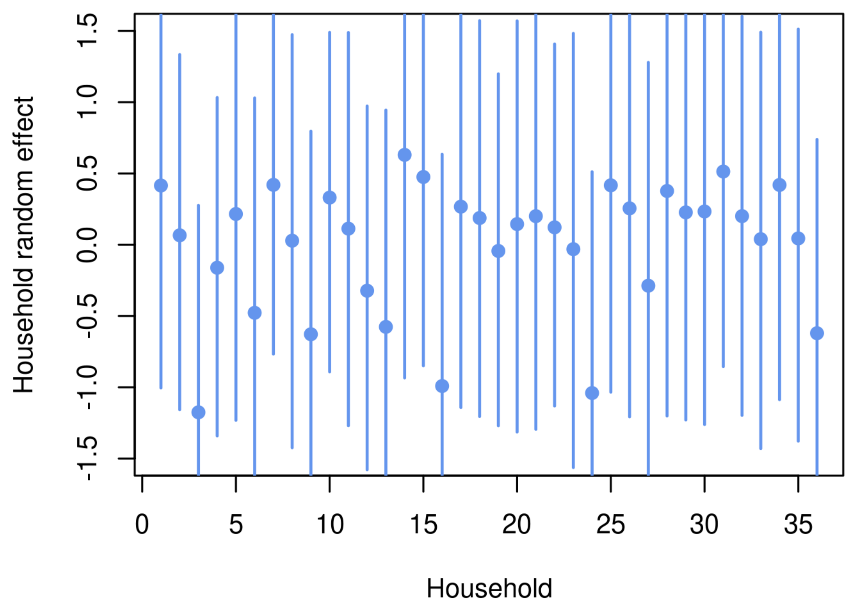

Figure S17: Random effects for families, from model 2, which includes age, sex and activities as other predictors

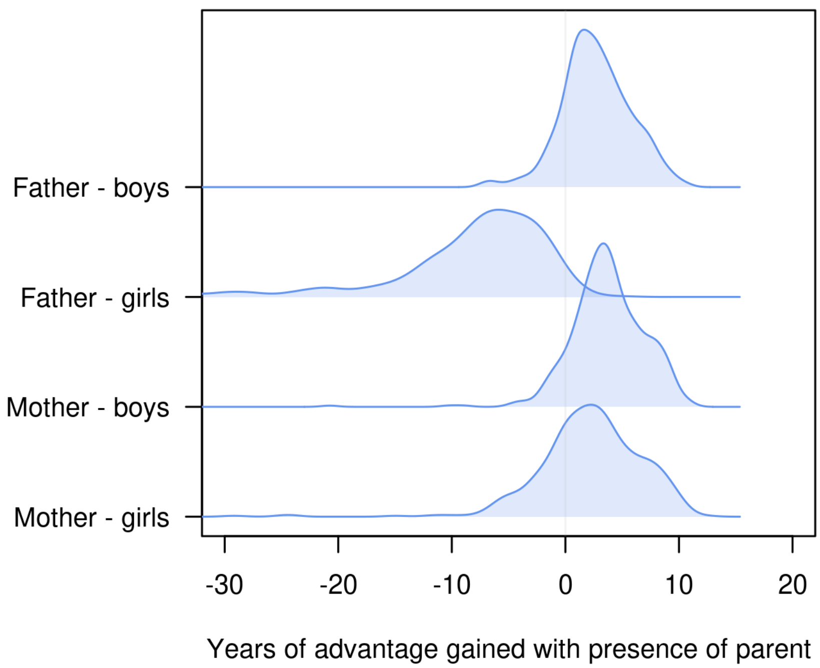

Figure S18: Distribution of years of knowledge gained thanks to the presence of parents, sex specific. Positive values indicate that the presence of one co-residing parent would allow an individual to reach earlier the same knowledge of a 20 years old individual who does not co-reside with a parent.

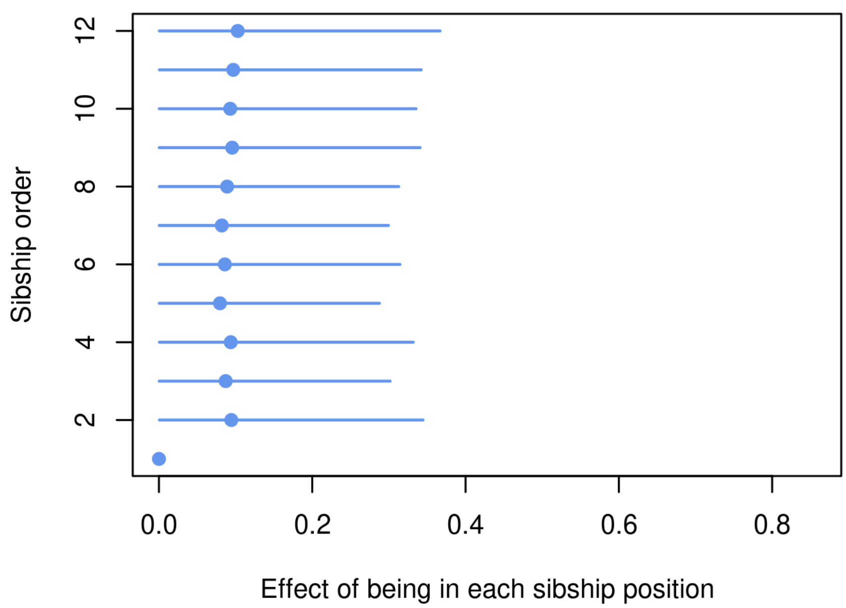

Figure S19: Birth order does not seem to have any effect on knowledge, as none of the positions in the sibship seem to have any different effect.

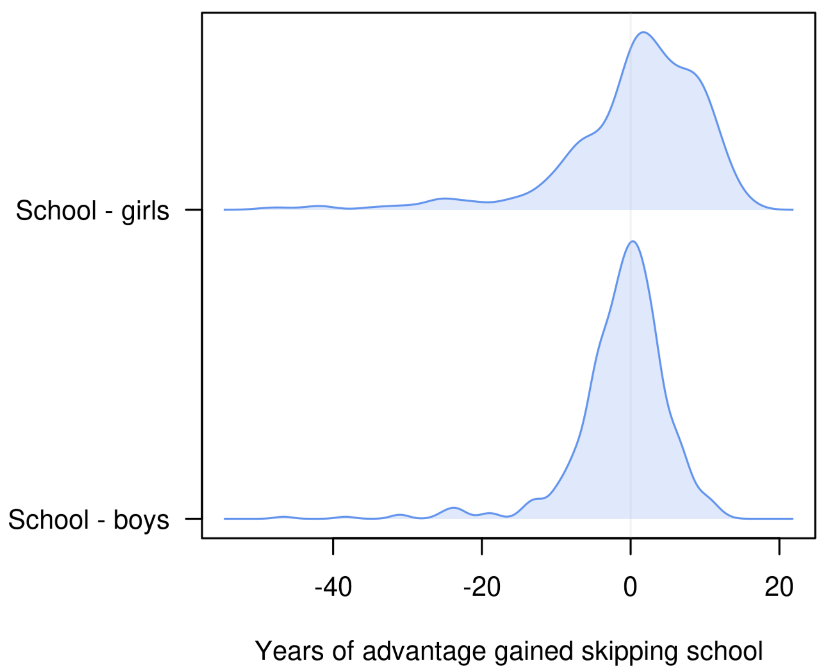

Figure S20: Distribution of years of knowledge gained by not going to school, sex specific. Positive values indicate that an individual who does not go to school would reach earlier the same knowledge of an individual who does.