nature human behaviour

Article

https://doi.org/10.1038/s41562-025-02221-6

Cognitive representations of social networks in isolated villages

Received: 18 August 2024

Accepted: 17 April 2025

Published online: 16 June 2025

Check for updates

Eric Feltham [1,2], Laura Forastiere [1,3,4] & Nicholas A. Christakis [1,2,4]

People not only form social networks, they construct mental maps of them. We develop a sampling strategy to evaluate network cognition in 10,072 adults across 82 Honduras villages and systematically map the underlying village networks. In 17 villages, we also discern the genetic relatedness of all 1,333 residents. Observers overestimate the social interactions among kin and are 33.38 percentage points (*J*) more accurate in judgements of ties between non-kin (95% confidence interval: 31.27–35.49). Counterintuitively, observers had more accurate beliefs about non-kin pairs, especially when the observers were popular, middle-aged, or educated. Observers were less able to accurately judge ties across different religions or wealth. Individuals in villages that cultivate coffee, requiring coordinated effort, demonstrated greater bias to view networks as connected. Finally, more accurate respondents had better access to information that we experimentally introduced to their peers. Overall, people inflate the number of connections in their networks and exhibit varying accuracy and bias, with implications for how people affect and are affected by the social world.

Social networks are crucial for social coordination[1,2] and for the interpersonal diffusion of health-related[2–4] and social, economic and political[5–8] phenomena. Decades of research on social networks has focused on the social effects of patterns of social interaction. Separately, research on social cognition has largely focused on how people represent other people and their attributes[9–11]. Much less work has examined how people think about social structure per se.

Yet, the ability to accurately 'see' one's social network may facilitate individual and collective action. Humans possess a coalitional psychology[12,13], and beliefs about the existence of relationships are probably a crucial mechanism by which people form alliances or infer the trustworthiness of others as information sources or as partners in joint action. Knowledge of social connections is also an important means by which individuals assign status to others[14,15]. Humans also exhibit an innate predilection to gossip about other people[16,17], which has been shown to be important for bonding, trust and social cohesion[17–19].

Scholars have tendered strong and contradictory theoretical accounts of individuals' capacity for knowledge of social structure. Some social theorists have supposed that individuals have little to no knowledge of their broader social world[20]. At the other extreme, social learning models assume that people have complete network knowledge[21,22]; however, some have argued that individuals' beliefs about geodesically distant connections play a role in social network dynamics[23].

Existing work has examined social network cognition in the laboratory[24–27] and in very small networks (usually less than 30 individuals) in circumscribed settings (such as classrooms)[28,29]. Other work has demonstrated that non-human primates have knowledge of third-party relationships[30,31], and that young humans can infer the closeness of two people[32]. Work has also suggested that animals build 'cognitive maps'[33] that capture their spatial environment[34] and that serve as a general mechanism to represent abstract cognitive space[35–37], including social domains. In humans, such maps can be used for introductions, information disclosure and access to network resources.

Social networks are typically characterized using surveys, where each member of a defined population is asked about their personal relationships. For example, a villager is asked to list individuals in response to the question, 'With whom do you spend free time?' In the present work, however, we develop a procedure to assess individuals' beliefs

[1]Human Nature Lab, New Haven, CT, USA. [2]Department of Sociology, Yale University, New Haven, CT, USA. [3]Department of Biostatistics, Yale University, New Haven, CT, USA. [4]Department of Statistics and Data Science, Yale University, New Haven, CT, USA. e-mail: nicholas.christakis@yale.edu

about others' relationships in their broader network, asking individuals to evaluate whether 'others' are connected; for example, we ask person k 'do i and j spend free time together?' Previously, the approach has been to collect data on social network beliefs exhaustively, where each person in the network, k, is asked whether every distinct pair of individuals in the network is connected[28,29,38]. This approach has limited the study of network beliefs to very small groups. Here, however, we develop a sampling technique to collect beliefs about networks, and we deployed it in 10,072 adults in 82 villages in Copán, a rural region of Honduras. Conjointly, we collected complete village-wide sociocentric network data and comprehensive sociodemographic data from all the inhabitants.

In the simplest scenario, k has evidence of the relationship between i and j either through gossip or direct observation. However, when evidence is absent, the respondent faces a more challenging epistemic task and must determine whether the absence of evidence should be interpreted as evidence of absence. In general, there need not be any necessary connection between the types of error people make: people may be strictly better able to identify ties that either do or do not exist. However, given a mental model and set of evidence, a perceiver can trade false positives for false negatives by varying their threshold for classification. The ability of k to form such a judgement may depend on the visibility of the tie to k. For example, ties that are geodesically further away in a village network may be ones for which k is less able to acquire any evidence; consequently, higher overall accuracy may simply be associated with a comparatively lower commission of type I (false positive) errors.

Since a person's own direct connections represent a small proportion of the village network in which a person is embedded, activating their broader network to serve a purpose can often involve chains of multiple actors[39,40]. Hence, network knowledge may be important for access to social capital. Furthermore, the relationship between social attributes or arrangements (from education to religion, from geographic isolation to the kind of local farming), on the one hand, and accuracy in network cognition, on the other hand, are important to understanding how individuals differentially reason about and access the social capital that the relationships in their community represent.

For example, previous work has shown that efforts to acquire resources are often buoyed specifically by heterophilous interactions (between dissimilar individuals), which are more likely to provide non-redundant information and resources to an individual[39,41]. Despite their utility, heterophilous ties are harder to maintain[42], generally attributed to structural (for example, shared social contexts make ties easier to maintain) and preferential (individual taste for homophily) factors; however, these explanations do not generally consider the 'visibility' of diverse ties. Namely, individuals' access to social capital might in principle be limited by the visibility of ties that cross social boundaries or connect socially dissimilar individuals.

Finally, here, we focus on kinship. Kinship has long been an object of inquiry[43,44], and kinship classifications have been taken to be fundamental to social order and cohesion[45] and determinative of social position. Some work has argued that knowledge of social relationships may be more important to social outcomes than mentalizing other individuals, especially in societies characterized by tight kinship structures[46]. We discover that while respondents are very accurate in determining the patterns of kinship in their networks, they wrongly use the existence of a kinship tie as a heuristic for other types of social interactions.

Overall, we find that a diverse set of demographic, village and network factors is associated with accuracy. As a result, people generally form a distinctly mistaken impression of the social structure around them. Finally, we investigate whether people who conceive of networks more accurately are better able to acquire novel information that we experimentally introduced at random into the villages, and we find that accuracy in social network cognition is associated with their receipt of novel information.

Surveying social network beliefs

Using a sampling procedure (see Methods and Fig. 1), we presented each villager with up to 40 unique pairs of individuals (mean = 33.6, s.d. = 7.7) that represented either real ties in the reference sociocentric network or ties that did not exist (see Methods and Fig. 1). We collected the reference village-wide sociocentric data directly from the villagers approximately 3 weeks before administration of the network cognition survey. Specifically, we asked respondents to evaluate whether each pair of individuals: (1) spend free time together ('free-time'); (2) discuss personal or private matter together ('personal-private'); and (3) are immediate 'kin' (see Extended Data Table 1 for the full questionnaire). We compared respondents' beliefs to the undirected sociocentric network, where a relationship is said to exist between two individuals if at least one of them reports the existence of the relationship.

Importantly, 'free-time' and 'personal-private' differ in their visibility to external observers. An individual can judge whether a pair spends free time together on the basis of direct observation of public behaviour, without reference to the intentions or dispositions of the occupants of the relationship. By contrast, an inference about a personal-private relationship between two people requires more substantial knowledge.

Villages ranged in size from 57 to 875 persons and in total contained 45,614 existing ties. The pairs (whether real ties or not) assessed in the survey were evaluated by an average of 4.43 respondents (s.d. = 4.90), and individual pairs (comprising the ties queried) appeared 128.9 times on average (s.d. = 72.0) as a pair member (see Supplementary Results for more details).

Statistical models

We use multilevel logistic regression models to assess the determinants of accuracy in network cognition in our data, where survey respondents have repeated measures and are clustered into villages. Our binary outcome is whether a respondent k believes that a tie exists between a pair of individuals, i and j. We separately estimate the two dimensions of accuracy: the true positive rate (TPR) and the false positive rate (FPR). FPR is equivalently represented as the true negative rate (TNR), TNR = 1 − FPR. We also estimate Youden's J statistic[47] to summarize performance: J = TPR − FPR. J ranges from −1 to 1, where a value of 0 denotes chance performance (see Supplementary Fig. 2 and Supplementary Methods).

We examine the individual rates along with the J statistic, since unidimensional measures of accuracy conflate sensitivity and specificity. In Fig. 2, we delineate three distinct changes in accuracy that may be associated with changing levels of an attribute. Changes may be associated with a pure change in performance, where an individual is better able to identify ties overall, or instead be associated merely with a shift in the type of error committed. Most of the changes in accuracy we observe are changes both in performance and in trade-offs between errors. We report accuracy results in hundredths, where, for example, 1.2 'points' corresponds to a J statistic value of 0.012. For the TPR and FPR rate estimates, this reporting corresponds to percentage points.

We include a range of fixed effects in these models, including the relationship of the pair in the reference network as well as demographic and other attributes and network characteristics. We report adjusted predicted probabilities and contrasts, showing how the belief that a tie exists changes with the values of a characteristic while holding other covariates fixed at their population means or typical values in the population. For analyses of the relationship between the two accuracy rates, and between accuracy and knowledge of exogenously introduced information, we estimate second-stage models that regress on the individual accuracy scores. All reported statistical tests correspond to two-tailed tests (see Methods for a full description).

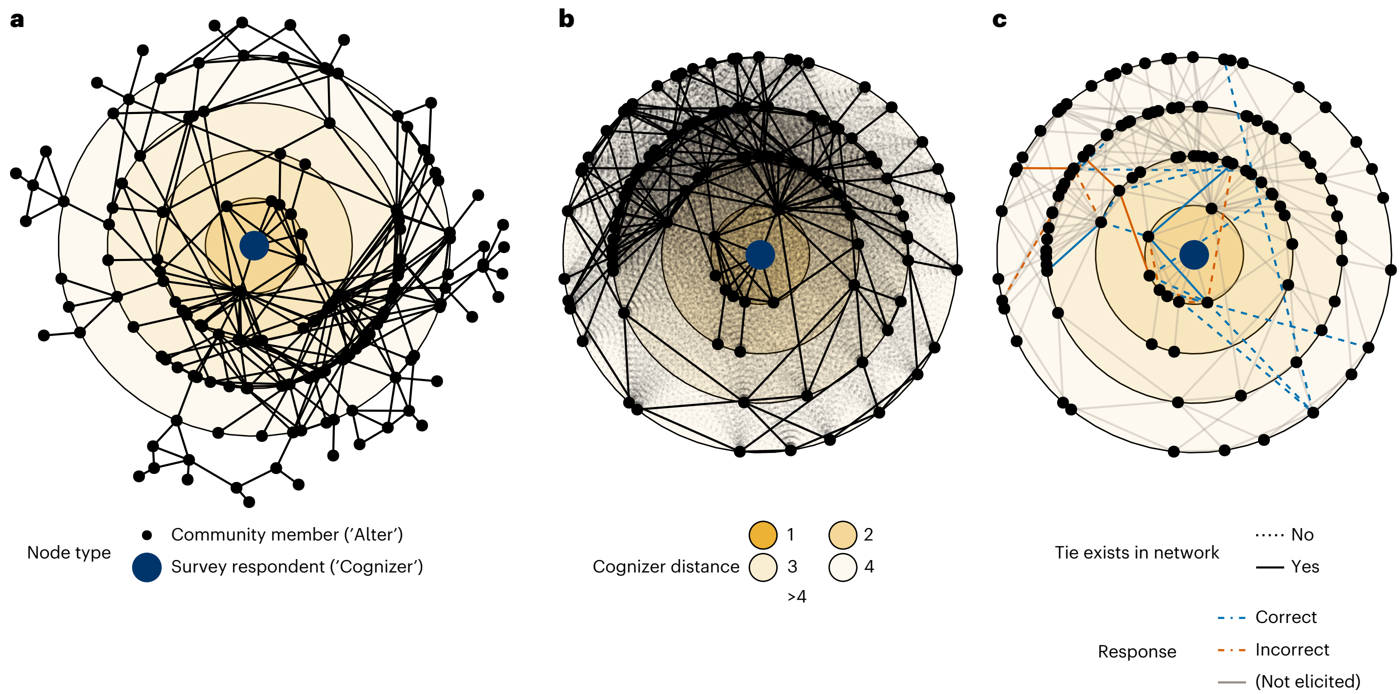


Fig. 1 | Outline of the survey procedure. a, The network for a specific relationship (for example, free-time) in a representative village for an individual. Circles represent geodesic distances from 1 to 4 steps from the respondent in the underlying sociocentric network. **b**, Conceivable ties within 4 geodesic steps from the respondent. Solid lines represent existing ties, dotted lines represent non-existent ties. **c**, Survey queries and responses. We present up to 40 ties to each respondent, drawn from **b**. Of these, 20 are among individuals within 2 degrees of the respondent, 10 are 3 degrees away, and 10 are 4 degrees away, measured in the network defined by the union of kin, personal-private and free-time networks. Individuals judge the existence of pairs that represent real ties (solid lines) or not (dashed lines) in the sociocentric network. Rings correspond to sampling bins. Individuals were queried about 33.6 pairs (median = 37.0, mode = 40.0) on average across free-time and personal-private relationships, and may be queried about fewer than 40 ties if they do not recognize individuals in the pair (see Methods). Since we sample on the union of networks, a wide range of geodesic distances between observers and elicited pairs is present in the data, with an average distance of 5.67 between a respondent and a displayed pair, and over 10% of elicited ties at 8 or more geodesic steps away.

Overall accuracy by tie type and kinship status

Figure 3a presents the bivariate distributions of accuracy, stratified by both the type of relationship judged (free-time or personal-private) and whether the judged tie is between kin or not (as reported in the reference network by *i* and *j*) (see also Supplementary Fig. 3). Each point in this space represents an individual respondent, *k*, who may be thought of as a binary classifier.

When measured by unadjusted individual accuracy scores, individuals perform above the level of chance in their relationship judgements (observe that the black dots are above and to the left of the diagonal indicating chance performance in Fig. 3a). Nonetheless, estimates for both types of relationship are close to the line of chance in practical terms when they occur between kin (observe that the magenta dots are near the diagonal line for the top two graphs in Fig. 3a) after adjustment for individual attributes of the observers (such as gender and age) and their network attributes (for example, the geodesic distance between a respondent and a pair they assess).

The chance-level performance in judgements of ties that are between kin is driven by a very high FPR for beliefs about personal-private (estimate = 91.42 percentage points, 95% CI: 90.8–92.04, $P < 0.001$) and free-time (estimate = 91.61 percentage points, 95% CI: 91.01–92.21, $P < 0.001$). In fact, both adjusted rate estimates are close to 1.0 in this case. Individuals appear to simply assume that individuals who are kin must be connected in these two other relationship types (see also Extended Data Table 2). Moreover, we observe that the J statistic is near zero for kin judgements, which can be appreciated visually since the quantity J is indicated by the vertical distance of the points from the diagonal lines. Conversely, we find that individuals are 33.38 points (J) more accurate in judgements of ties between non-kin compared with those between kin (95% CI: 35.49–31.27, $P < 0.001$) (bottom two graphs in Fig. 1a; see also Extended Data Table 2). Furthermore, we observe a substantial number of kin ties that correspond to neither free-time nor personal-private ties (Supplementary Table 1), indicating that kinship does not imply the existence of relationships in these village network in general.

Whether a tie is between kin is the single most important factor that affects social network cognition (as compared with all other effects considered in Extended Data Tables 2 and 3). We also find that individuals are very accurate in their knowledge of the kinship patterns themselves (Extended Data Table 4). Yet, this stands in sharp contrast to their ability to accurately determine whether kin are also connected in the other two network relationships we study. In other words, while people know who is related to whom in their villages, they are less aware whether others spend free time together or discuss personal matters, and incorrectly assume that kin necessarily have relationships in these other domains.

We collected genetic data in a subset of 17 villages (n = 2,293 respondents) for the members of the cognized pairs (n = 1,333 people) and estimated the kinship coefficient between the pairs (see Methods for details). We observe a strong relationship between accuracy and (genetic) relatedness (Fig. 3b), validating the self-reported kinship result. Furthermore, respondents are most accurate in judgements of individuals who are somewhat unrelated, with a kinship coefficient of −0.063 and overall accuracy at 48 points (95% CI: 43.2–52.7, $P < 0.001$; TPR: 66.9 points, 95% CI: 62.7–71.0, $P < 0.001$; FPR: 18.9 points, 95% CI: 16.0–21.8, $P < 0.001$). However, they approach chance performance both for judgements of very unrelated individuals or very close kin (Supplementary Table 2). These results are consistent with alternative specifications that use the specific kinship category (for example, sibling) or distance in the kinship network (Extended Data Fig. 1).

 https://doi.org/10.1038/s41562-025-02221-6

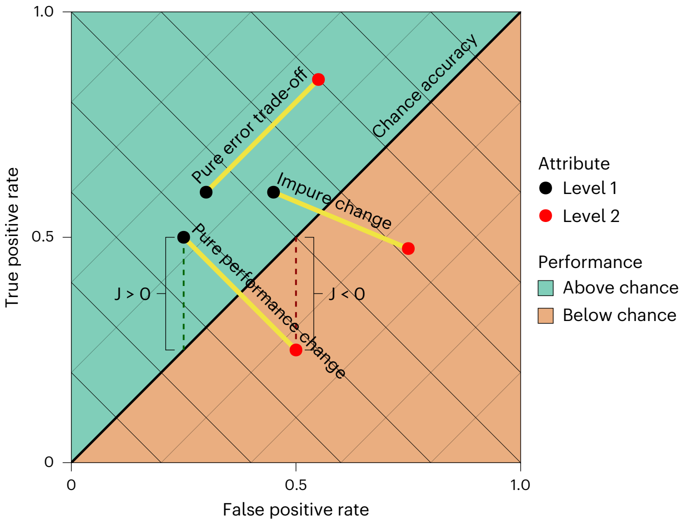


Fig. 2 | Receiver operator characteristic space. Observations represent estimated means for each dimension of accuracy (TPR and FPR). Each point may be thought of as a binary classifier. Classifiers perform better-than-chance (light-green region) or worse-than-chance (light-orange region), and those on the diagonal black line (where TPR = FPR) perform at the level of chance. The vertical distance to the line of chance corresponds to overall performance and is negative below this line, where FPR > TPR. Movement parallel to line TPR = FPR indicates changes in *J* with an equal increase for both rates. Conversely, movement parallel to the line TPR = FPR represents a change in error type only where *J* is constant. We show three illustrative examples where a change in the value of an attribute (the move between a black dot and a red dot) leads to one of the three types of change in accuracy. In the first ('Pure performance change'), the change in an attribute leads to a pure change in performance, where the *J* decreases from level 1 to level 2 with an equal change in the rate of true positives and false positives. In the second ('Pure error trade-off'), we see no change in overall performance, but a sizable shift in types of error committed across the range of the attribute. Note that, along this line, the value of an attribute is associated with a more liberal or more conservative tendency to render judgements of the existence of a tie. In the third ('Impure change'), we see a change in both bias and in overall performance.

As expected, we find that personal-private relationships are conceived of less accurately than free-time relations, in the full sample of 82 villages. We see that *J* for the non-kin ties is 40.48 (95% CI: 38.45–42.42, $P < 0.001$) for free-time, which is 3.63 points lower than for judgements of personal-private ties (95% CI: −6.5 to −0.75, $P = 0.013$; TPR: −9.39, 95% CI: −11.4 to −7.38, $P < 0.001$; FPR: −5.77, 95% CI: −7.96 to −3.57, $P < 0.001$). Furthermore, the two rates move in opposite directions, where FPR and TPR are lower for judgements of personal-private compared with free-time ties (Extended Data Table 2).

Respondent characteristics are associated with accuracy

Attributes of individual respondents are associated with how accurate they are. Here we specifically focus on individual judgements of non-kin ties, since there are no meaningful differences in accuracy when respondents assess ties between kin. Note that the effect estimates for kin are near the diagonal in the upper right of the graphs in Fig. 3c, such that the accuracy estimates are very similar in size regardless of the value of a studied attribute (see also Extended Data Fig. 2 for more details).

First, we do not find evidence of an effect of gender on overall accuracy (*J*). Women are non-significantly more accurate than men, at an increase of 2.42 points (95% CI: −0.72 to 5.56, $P = 0.131$) in terms of overall accuracy. However, we do find that men have a 3.2-point higher FPR than women (95% CI: 0.79–5.6, $P = 0.009$) (TPR: −0.78, 95% CI: −2.97 to 1.42, $P = 0.488$) (Extended Data Fig. 2). Age exhibits a curvilinear relationship with accuracy, at its highest around middle age. From middle age, accuracy declines to chance performance in elderly participants (Fig. 3c). Most of the change is driven by an increase in the false positive rate. Over the age range, estimates of overall accuracy reach a maximum *J* statistic of 39.28 points (95% CI: 37.37–41.18, $P < 0.001$; TPR: 67.07, 95% CI: 65.72–68.42, $P < 0.001$; FPR: 27.80, 95% CI: 26.37–29.23, $P < 0.001$) at age 31. While this is not significantly different from the estimates at the lowest ages, with a difference of 2.51 points (95% CI: −1.02 to 6.04, $P = 0.163$; TPR: 5.26, 95% CI: 2.53–7.98, $P < 0.001$; FPR: 2.74, 95% CI: 0.31–5.18, $P = 0.027$), this is 37.45 points higher than the oldest participants (95% CI: 29.68–42.41, $P < 0.001$), who have scores below zero, indicating worse-than-chance performance overall (TPR: 1.4, 95% CI: −5.23 to 8.03, $P = 0.679$; FPR: −36.04, 95% CI: −42.41 to −29.68, $P < 0.001$).

Both wealthier and more educated individuals are significantly more accurate. The wealthiest individuals have a 6.94 point increase in *J* (95% CI: 2.22 to 11.6, $P = 0.0.004$; TPR: −13.32, 95% CI: −16.77 to −9.87, $P < 0.001$; FPR: −20.26, 95% CI: −23.76 to −16.76, $P < 0.001$) over villagers with the least wealth. In addition, they are also more conservative in their judgements, with a greater bias to assume that their networks are sparse compared with those lower in wealth or education (Fig. 3c). The results for education are presented in Extended Data Fig. 2 and Table 3. Finally, individuals with higher network degree exhibit a clear increase in performance on all three accuracy measures, with an increase in *J* of 8.3 points for those with just one tie up to those with 18 ties (95% CI: 3.47–13.12, $P < 0.001$) (TPR: 5.55, 95% CI: 1.63–9.46, $P = 0.005$; FPR: −2.75, 95% CI: −5.75 to 0.25, $P = 0.073$; Fig. 3c) (see Extended Data Table 3 for detailed results on the contrasts).

Tie characteristics are associated with observer accuracy

Accuracy is related to properties of the ties judged by the respondents (Fig. 4, and Extended Data Figs. 3 and 4). Here we again examine differences in judgements of non-kin ties. As respondents are asked about pairs of individuals who are further away in the underlying sociocentric network, we observe a strong decrease in overall accuracy. *J* declines by 21.6 points (95% CI: 17.61–25.59, $P < 0.001$; TPR: −36.82, 95% CI: −40.4 to −33.25, $P < 0.001$; FPR: −15.22, 95% CI: −17.27 to −13.17, $P < 0.001$) from judgements about directly connected individuals (that is, about whether two alters connected to an ego are in turn connected to each other) to those furthest away (Fig. 4b).

We also find that accuracy in respondents' judgements changes with the number of geodesic steps that separate two target individuals in the network, when they are not connected. The observer judged more accurately as the geodesic distance increases by over an order of magnitude (from 2 to over 15 degrees of separation); pairs are less likely to be seen as connected when they are further apart. This amounts to an FPR change of −34.55 points (95% CI: −36.33 to −32.77, $P < 0.001$) (Fig. 4b). Hence, individuals are responsive to the social distance between individuals.

Furthermore, we see a decrease in overall accuracy regarding ties connecting people with high average degree (of *i* and *j*), with a change of −7.1 points (95% CI: −11.14 to −3.07, $P = 0.001$; TPR: −16.1, 95% CI: −18.38 to −13.82, $P < 0.001$; FPR: −23.2, 95% CI: −26.76 to −19.65, $P < 0.001$) in Fig. 4a. Here we observe a substantial trade-off in the error type that a respondent makes when the average degree of the pair changes. We also observe a significant interaction between the degree of the respondent and the average degree of the pair (TPR: $b = 3.325$, $P < 0.001$, FPR: $b = 3.397$, $P < 0.001$). This means that high-degree individuals are the most accurate in judgements of low-degree pairs, but their performance declines more rapidly than low-degree respondents for judgements of high-degree pairs (Supplementary Fig. 4).

Separately, as the average age of a pair increases, we find a decrease in accuracy. This is driven solely by a drop in TPR. Here, overall accuracy (*J*) changes by −24.34 points (95% CI: −27.67 to −21.01, $P < 0.001$;

https://doi.org/10.1038/s41562-025-02221-6

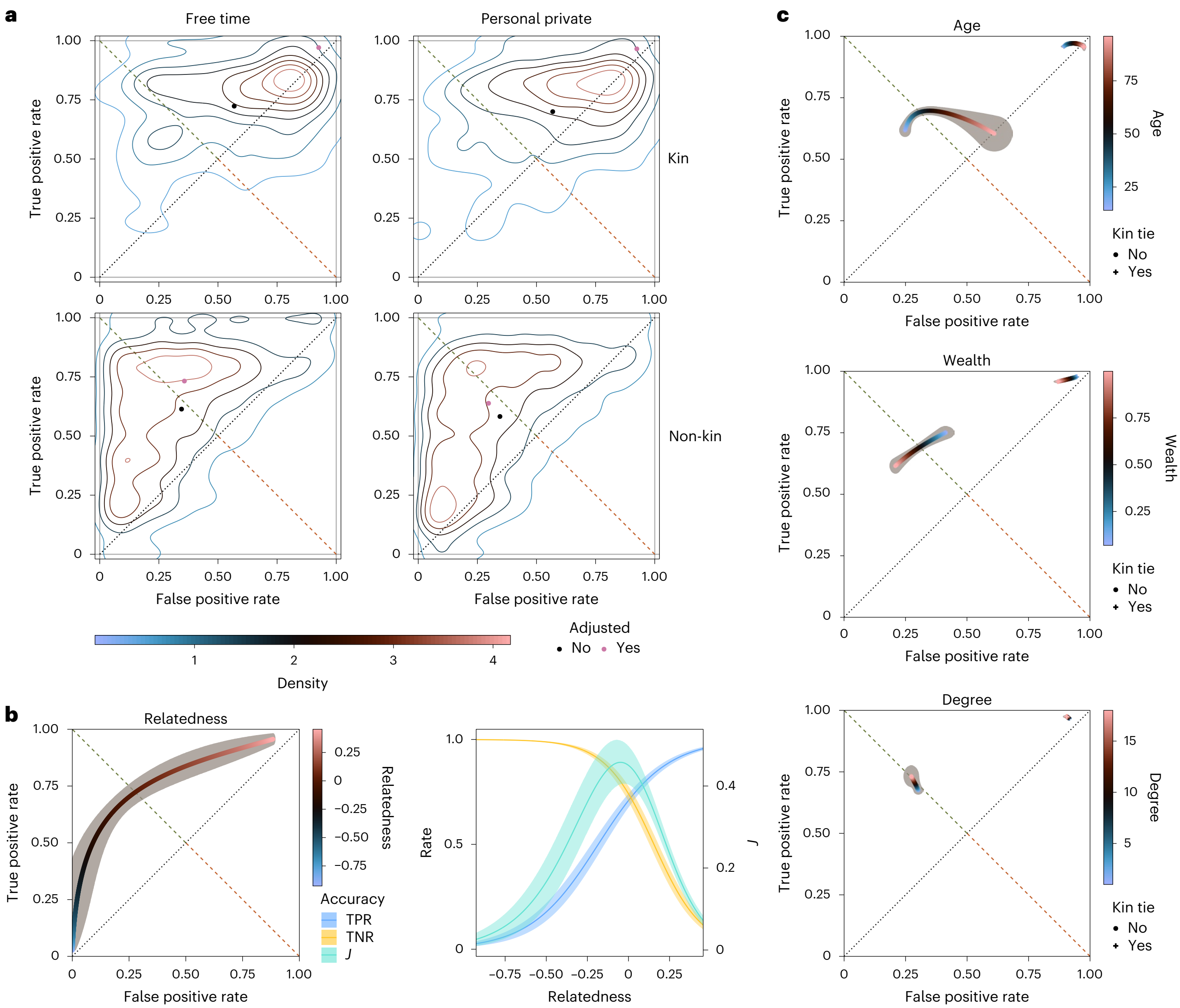


Fig. 3 | Accuracy of social network beliefs. a, Distributions of respondent accuracy. The green dotted-line segment above the black dotted line represents better-than-chance performance, and the orange section below the dotted line represents worse-than-chance performance. Top: distributions of assessments for kin ties. Bottom: assessments for non-kin ties. Participant-level accuracy rates are stratified by actual kinship status and relationship type. Black dots indicate overall unadjusted average participant-level accuracy; magenta dots show model-adjusted means. Unadjusted estimates restrict to respondents evaluating at least 3 true and 3 false ties (n = 9,305). Adjusted estimates account for demographic controls, network degree, kinship status, relationship type, network distance and random effects for village and respondent. Density values reflect interpolated counts of respondent-level true and false positive rates. **b**, Marginal effect of genetic relatedness on network cognition accuracy. In 17 villages (n = 2,248 respondents), we estimated kinship coefficients between cognized pairs (1,333 individuals). Kinship coefficient values <0 indicate unrelated individuals and 1/2 indicates monozygotic twins. Left: grey bands represent bootstrapped 95% confidence ellipses around the mean estimates. Right: bands represent 95% confidence intervals around the mean estimates. **c**, Marginal effects of observer characteristics on accuracy. Effects of age, wealth and network degree (count of first-degree neighbours for personal-private or free-time relationships) are shown. Grey shading represents 95% bootstrapped confidence ellipse around the mean estimates.

TPR: −26.71, 95% CI: −29.26 to −24.16, P < 0.001; FPR: −2.37, 95% CI: −4.69 to −0.06, P = 0.045) over the range of values (Extended Data Fig. 3c). By contrast, when there is a greater difference in age between those in a judged pair, ties are substantially more salient, with an increase in accuracy of 22.01 points (95% CI: 19.0–25.01, P < 0.001; TPR: 24.78, 95% CI: 22.8–26.76, P < 0.001; FPR: 2.77, 95% CI: 0.41–5.13, P = 0.021) over the range of values in the population (Extended Data Fig. 3d).

We observe significant differences for each gender combination of the ties, with mixed ties identified at an overall accuracy (J) of 17.37 points higher than those between men (95% CI: 14.33–20.41, P < 0.001; TPR: 3.27, 95% CI: 1.05–5.49, P = 0.004; FPR: −14.1, 95% CI: −16.33 to −11.86, P < 0.001) and at 7.76 points higher than those between women (95% CI: 4.86–10.65, P < 0.001; TPR: −2.93, 95% CI: −5.01 to −0.86, P = 0.006; FPR: −10.69, 95% CI: −12.85 to −8.53, P < 0.001). Likewise, ties between women are also more accurately conceived than those between men, at a difference of 9.61 points (95% CI: 6.52–12.7, P < 0.001; TPR: 6.21, 95% CI: 4.14–8.27, P < 0.001; FPR: −3.41, 95% CI: −5.86 to −0.96, P = 0.006).

Next, we consider important characteristics related to social identity: religion, indigeneity (whether respondents identify as being

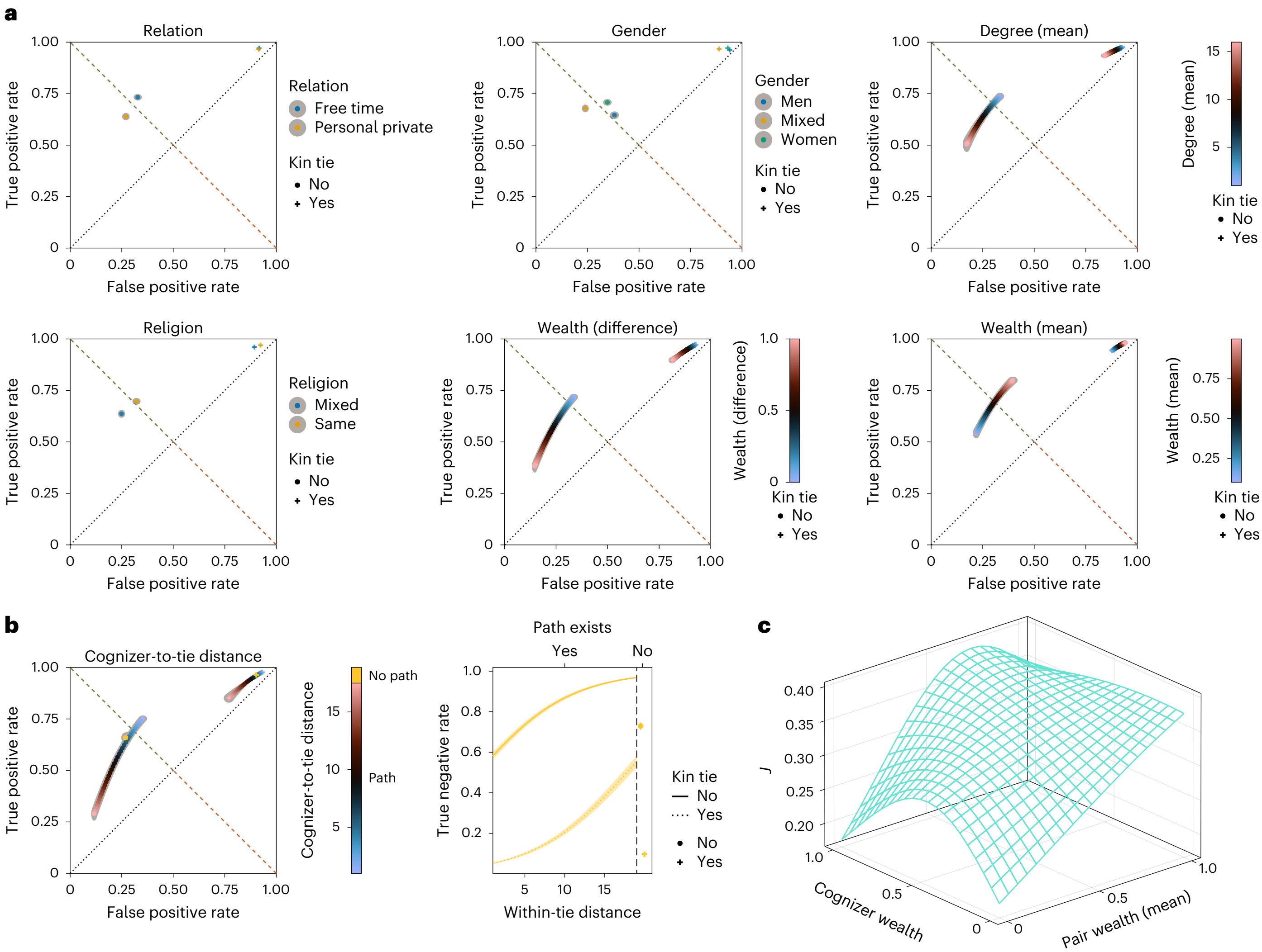


Fig. 4 | Tie determinants of respondent accuracy. **a**, We find that a range of properties of ties have statistically significant associations with their tendency to be accurately identified. In each panel, the marginal effect on accuracy in ROC space is shown. Grey shading represents the 95% bootstrapped confidence ellipse of the predictions from the two models. Estimates are stratified by whether they are of a tie among kin or not. Dotted line TPR = FPR indicates the line of chance, and the dotted green and red line TPR = 1 − FPR indicates changes in performance. **b**, Network distances. Left: respondent-to-tie geodesic distance. Individuals may or may not have a defined path between them in the reference network; when there is a path, individuals exist at a geodesic distance defined as the minimum number of steps between them; note that individuals who do not have a path between them necessarily have a path in at least one of other networks considered in this study, by design. Right: within-tie distance. When a direct tie does not exist between two individuals, they are separated by a specific geodesic distance (or they may have no path between them in the network). The scatter points to the left of the dashed linerefer to estimates where there is no geodesic path in the network from person *i* to person *j*. Since the tie does not exist, only the false positive rate (or true negative rate) is defined. Parameters are fit from separate models of each rate, conditional on tie verity in the reference network (see Methods for details). **c**, Interaction between the average wealth of a pair and the respondent's wealth on the summary measure, *J* (see Methods for details).

of Mayan descent) and wealth. We do not find a significant relationship between the indigenous composition of a tie and how well it is conceived (Extended Data Fig. 4b). We find that while individuals answer correctly at similar overall rates for mixed-religion pairs as for same-religion pairs (Fig. 4a and Extended Data Fig. 4a), there is a substantially different error pattern: same-religion pairs are associated with a TPR rate that is 6.06 percentage points higher than for mixed-religion pairs (95% CI: 3.94–8.17, $P < 0.001$), and the FPR changes by an even larger 7.24 percentage points (95% CI: 5.11–9.38, $P < 0.001$) (J: −1.19, 95% CI: −4.08 to 1.71, $P = 0.423$). In short, individuals are less able to identify connections between individuals of different religions and more likely to falsely assume that ties exist between co-religionists.

Furthermore, we find relatively large and significant effects related to the wealth of the cognized pair. Observers judge ties between individuals of similar wealth more accurately. In Fig. 4a (Extended Data Fig. 4c), when we compare assessments of those closest in wealth to those with the greatest absolute difference in wealth, we find that J declines by 14.49 points (95% CI: 11.02–17.96, $P < 0.001$; TPR: 33.47, 95% CI: 30.46–26.47, $P < 0.001$; FPR: −18.98, 95% CI: 16.99–20.96, $P < 0.001$). In Fig. 4a (Extended Data Fig. 4d), we observe that wealthier pairs (those with a high average wealth level of the individual pair members) are distinguished with significantly higher overall accuracy, a difference of 7.92 points (95% CI: 4.45–11.39, $P < 0.001$; TPR: 25.57, 95% CI: 22.94–28.2, $P < 0.001$; FPR: 17.65, 95% CI: 15.23–20.07, $P < 0.001$). In addition, changes in average wealth also constitute a significant trade-off in error type; villagers tend to blindly assume that wealthier pairs are more likely to be connected, exhibiting both high TPR and FPR (see Extended Data Table 2 for all estimates).

Moreover, we examined the interaction between the wealth of the respondent (k) and the pair (of i and j) with respect to accuracy. We find evidence that the relationship between accuracy and the wealth of a cognized pair changes substantially with the wealth of the respondent. The interaction is statistically significant for both the TPR ($b = 3.515$, $P < 0.001$) and FPR ($b = 3.959$, $P < 0.001$) (Extended Data Fig. 4e and Supplementary Table 3). For both underlying accuracy rates, we see that poorer individuals are more accurate in rendering judgements of ties between other individuals who are poor, compared with wealthier individuals. This pattern diminishes and then flips for pairs with relatively high wealth. When we consider overall accuracy (J), as shown in Fig. 4c, we see that accuracy increases linearly with the wealth of the pair for judgements by individuals who are low in wealth (see Supplementary Fig. 5 for additional results).

In sum, while both poor and wealthy individuals have low accuracy when queried about the ties between poor individuals, the accuracy of the wealthy respondents increases at a faster rate and is more accurate, for other wealthy pairs. Individuals who are less well-off are less able to judge ties among the wealthier members of their village networks than the (relatively) wealthy themselves, even after adjusting for social network distance, demographic characteristics and kinship. Strikingly, poorer individuals are also less able to judge relationships among other poor individuals.

Village-level network accuracy

We find relatively little variation across the 82 villages in the accuracy of their residents (Fig. 5a). We assessed whether village size, geographic isolation, or agricultural focus were associated with the accuracy of social network cognition of their inhabitants after adjusting for the characteristics of the respondents. Neither village size nor isolation were associated with villager accuracy (see Supplementary Results). Furthermore, given that we adjust for the geodesic distance between the observer and the pair, we specifically note that we fail to find an effect of village size independent of this distance (although we do find an effect of village size in a simplified model, in Supplementary Fig. 6). However, villages engaged in coffee cultivation had more accurate villagers (Fig. 5b and Supplementary Fig. 7) with a difference in the TPR of 5.07 points (95% CI: 2.09–8.06, $P = 0.001$). Most generally, villagers exaggerate the connectivity of their network, when aggregating over all respondents, although the general shape appears visually similar to the reference network (as illustrated in Fig. 5c,d).

Relationship between the accuracy rates

Next, we consider the relationship between the two underlying measures of accuracy within respondents. We estimate respondent-level accuracy scores and regress TPR on FPR, with second-stage adjustment for the age, gender and the network degree of the respondent (see Supplementary Methods for details).

There is a clear concave relationship, where more sensitive individuals are less specific in their judgements about network structure (Fig. 6a and Supplementary Table 4). Generally, we see that the distribution of accuracy scores is defined by individuals' tendency to trade errors rather than perform better or worse. However, we do find that more accurate individuals have a greater relative ability to detect true negatives, with a sparser view of their network. Specifically, the most accurate have the lowest false positive rates ($J = 42.73$ points, 95% CI = 41.91–43.55, $P < 0.001$), while the least accurate have the highest false positive rates ($J = 8.82$ points, 95% CI = 6.42–11.22, $P < 0.001$). While we observe a TPR change of 29.09 percentage points (95% CI: 26.56–31.6, $P < 0.001$) over the observed range of respondent-level false positive rates, the simultaneous change in the FPR is over twice as large.

Network accuracy and acquisition of exogenous information

Our investigation was conducted amid a randomized controlled trial to discern optimal methods for the delivery of public health interventions to individuals[48]. In 44 (out of 82) villages, we targeted randomly chosen individuals for information sessions focused on health behaviours (delivered monthly for 22 months) where participants were taught previously unknown riddles related to health outcomes. Knowledge of the riddles was assessed at the conclusion of the intervention, roughly a year before the collection of network beliefs (Supplementary Table 5). The percentage of villagers targeted in each village was randomized, by design, over the range from 0% to 100%. Here we examine riddle knowledge among those ($n = 2,700$ respondents) 'not' targeted for the intervention, who would have learned the riddles through social learning. The three riddles summarized the use of zinc as an effective treatment for diarrhoea, the use of a prenatal vitamin, and proper umbilical cord care for newborns.

Overall, we find that knowledge of the riddles is positively associated with network accuracy (J). The most accurate individuals are 27.70 percentage points more likely to also learn the riddles (95% CI: 17.50–38.00, $P < 0.001$), although those who are less able to identify true positives are less likely to know the riddles. These effects represent a residual association, obtained after adjusting for demographic and network characteristics (Fig. 6b, and Supplementary Tables 6 and 7).

Performance, bias and trade-offs across attributes

Finally, we examine accuracy performance and error type bias across characteristics (Fig. 7). We transform the coordinates, TPR and FPR, so that they represent the performance (J) and bias (illustrated in Fig. 7a). We measure positive predictive bias (PPB) as the extent to which a classifier has a greater ability to detect true positives over true negatives and errs toward stating the existence of ties regardless of their status (see Supplementary Methods). PPB ranges from 0 to 2, where an individual assumes that no ties exist at 0, assumes that every pair is a tie at 2, and is unbiased at PPB = 1. We observe that the most accurate individuals are those with a higher TNR than TPR, such that individuals with highest TNR are 4.61 points more accurate (95% CI = 3.76–5.45, $P < 0.001$) than those without bias, at $J_{PPB} = 1$ (Fig. 6a). In Fig. 7b on the left, we display the maximum absolute amount of change in performance and bias, and on the right, we display the ratio.

Ratios above 1 indicate greater change in performance than a trade-off. Here, the network degree of a survey respondent is the characteristic most clearly associated with a pure performance, although other attributes (for example, age) do represent a larger performance change over their full observed range of values in the population. By contrast, wealthier people are primarily more conservative in their judgements of ties in their network, with a comparatively more modest increase in overall accuracy: they are less likely to falsely see ties that do not exist, but at partial expense to their ability to ascertain real ties. Most attributes have a ratio below one, implying a greater tendency to trade errors.

Robustness checks

We repeated our main analyses using two alternative measures of the reference (underlying) sociocentric network. First, we allowed a tie to be considered true if reported in any one of the three previous waves of data collection (and false otherwise) reaching back roughly 5 years. Separately, we required that a tie be nominated in both of the two most recent waves of data collection (the current round and two years previously). These analyses revealed only modest changes in our findings. In addition, we duplicated our main analysis on a restricted subset of the data, where survey respondents were within 3 geodesic steps of the pairs they judge, in line with smaller network cognition datasets[49]. Here, we found results consistent with the results in the larger network (see Supplementary Results).

Discussion

We find that people have systematic, discernible biases in how they perceive social relationships among others in their communities. This bias is socially patterned. First, people systematically overstate the role

 https://doi.org/10.1038/s41562-025-02221-6

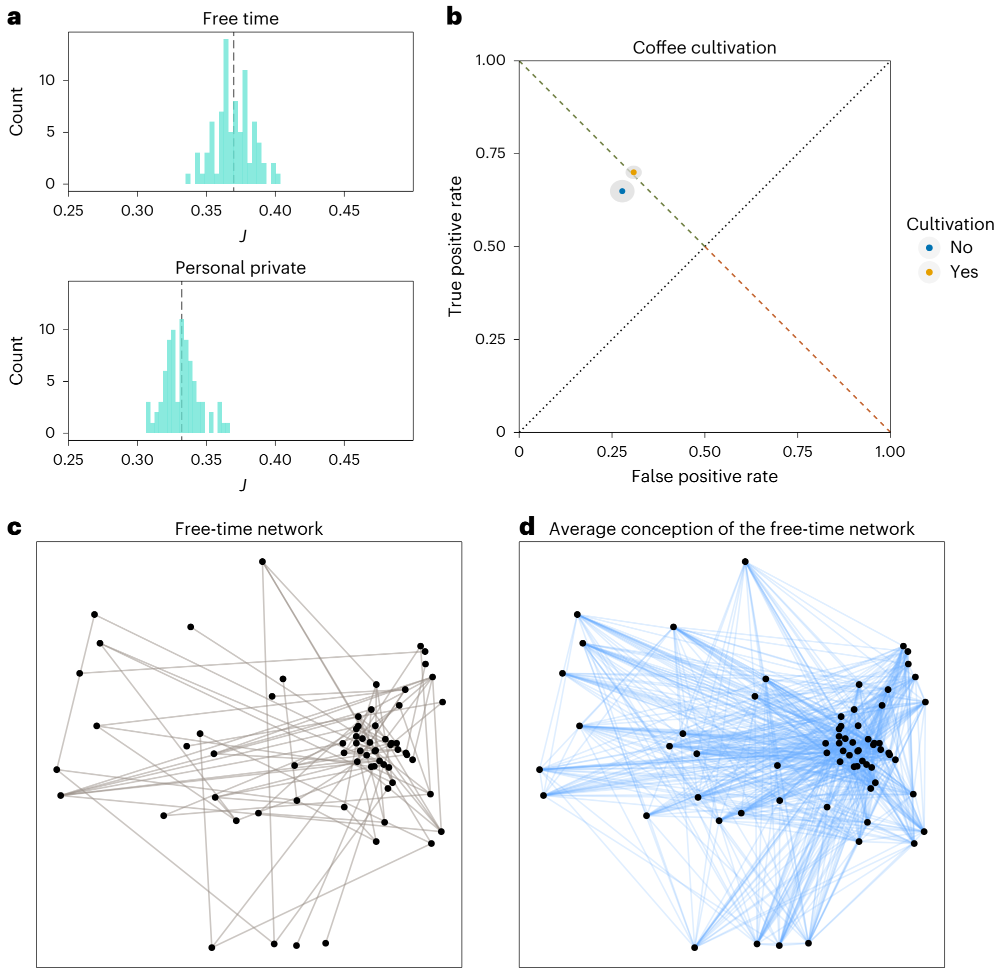


Fig. 5 | Village-level perspectives. a, Village-level distributions of overall accuracy (*J*). We find that the village-average scores are narrowly distributed (in contradistinction to the much wider individual-level distributions in Supplementary Fig. 3), just above 1/3 of the way to perfect accuracy from chance performance for each network. Consistent with the individual-level analyses (Fig. 3), we see greater accuracy for the free-time network. The dashed lines represent the distribution mean, for each relationship. **b**, Effect of whether a village cultivates coffee on social network accuracy. Parameters are fit from separate models of each rate, conditional on tie verity in the reference network. Adjusted predictions at the mean on accuracy in ROC space are shown; grey shading represents the 95% bootstrapped confidence ellipse of the predictions from the two models. **c**,**d**, We show the contrast between the underlying sociocentric network (**c**) and the network as an average of the individual predictions across the whole free-time network (**d**) for a single village. Here predicted ties are either true or false positives. We see that individuals predict a network that is much denser (by a factor of 8), adding a total of 748 ties to the 108 that exist in the sociocentric network. See Supplementary Information for further village-level analyses.

of kinship in social structure by assuming kin interact more than they do. That kinship is systematically misread as a guide to other domains of social life is notable given that so much work emphasizes its role in social classification and collective action[43–45]. Further, overestimating kinship's influence may negatively impact trust and cooperation, particularly with out-groups[46,50–52], and may limit individuals' ability to form links beyond family ties that are important for collective efficacy and social learning[53].

Nonetheless, that individuals have accurate knowledge of the kinship patterns themselves accords with the importance of family structures in settings where they often act as substitutes for markets or other institutions (for example, by securing property rights and guiding exchange)[54,55]. Kin ties are often highlighted at community events (for example, at weddings) and are likely to be topics of conversation (for example, who got married) in ways that informal ties are not. While it may appear surprising that a tie in the kinship network 'does not' necessarily imply that two individuals discuss personal or private matters, people do not always activate kin ties for personal advice[56,57], and even small-scale societies are characterized by large numbers of non-kin ties[58–62]. More broadly, this tendency to conflate distinct types of ties (for example, kin connections and free-time relationships) may represent a more general pattern whereby observers (wrongly) infer that ties overlap. Perhaps, individuals may presuppose multiplexity in relationships in an attempt represent complex patterns more simply[26].

Individuals also misconceive relations among non-kin. We find that a variety of key demographic characteristics predict accuracy in network knowledge. Furthermore, many characteristics indicate a change in error type bias rather than genuine changes in performance. Notably, degree, a social network property of an individual, is most clearly associated with a strict increase in performance.

Consistent with limited work on network accuracy[63], we find that it is crucial to disaggregate the dimensions of accuracy, rather than relying only on a summary measure (*J*), given the different social meanings of the error rates: failing to see a tie that exists may have a very different implication than falsely seeing one that does not. Relatedly, different individuals may have identical representations of a relationship but apply different classification thresholds and thus judge ties very differently, potentially due to considerations of practical interest[64], where tendency to assert a tie may hang on the stakes for an individual. Benefits may even accrue to false beliefs in ties: individuals who act as

 https://doi.org/10.1038/s41562-025-02221-6

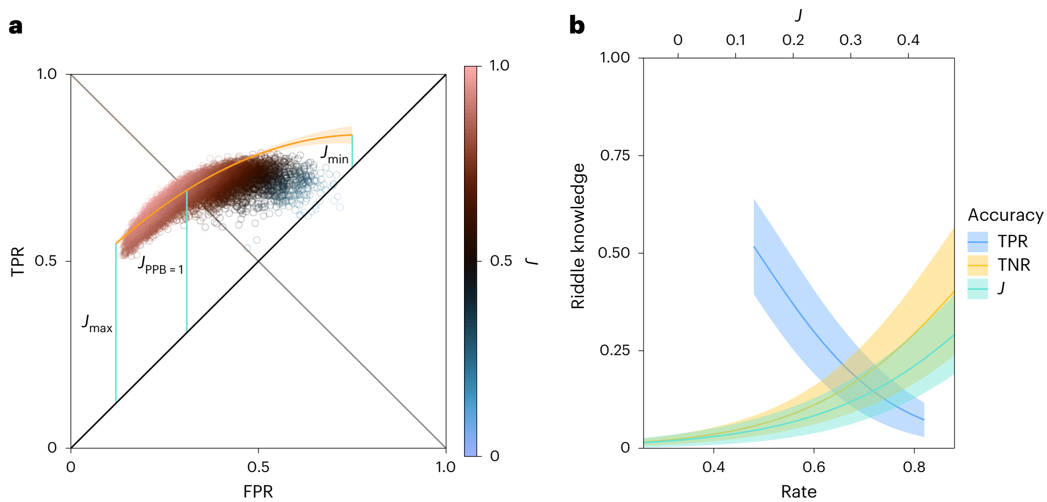


Fig. 6 | Accuracy within individual respondents and association with exogenous information. a, We find that the individual-level accuracy rates (TPR and FPR) are strongly related, such that increases in a tendency to identify true positives is associated with an increase in the tendency to identify false positives. The orange line represents the adjusted predictions at the mean for the individual accuracy rates, estimated from an OLS regression of an individual's TPR on their FPR, adjusting for age, gender, network degree and the relationship (free-time and personal-private). The points represent the predicted accuracy rates for each respondent, marginalized over relationship, and are coloured by performance (J). The estimates concern observations among non-kin. Bootstrapped 95% confidence intervals are shown (orange band). In addition, J is represented graphically at its extrema (J_{max}, J_{min}) and the point at which individuals are unbiased (PPB = 1). **b**, Network accuracy is associated with acquisition of novel non-social information. Estimation is performed with a logistic model of knowledge of three exogenously introduced riddles on predicted respondent-level social network accuracy. Models adjust for the specific riddle, demographic characteristics (age, gender, education) and network degree. Bands represent bootstrapped 95% confidence intervals around the mean (see Supplementary Fig. 8; see Supplementary Tables 7 and 8 for details).

if friendship ties exist may inadvertently forge new ties. Investigating the broader social consequences of network knowledge is an important direction for future work.

Strongly connected communities are better able to solve coordination problems[65,66], while networks fragmented along sectarian or ethnic lines are often less effective and often have lower levels of economic development[67,68]. Intergroup contact is important for social cohesion[69] and has been linked to effective community governance[70]. Evangelical Protestantism has emerged as an important and contentious identity marker in Latin America[71]. We find that individuals tend to assume that more ties exist between those of the same religion and are less able to identify ties across religious lines. Similarly, individuals are less able to identify ties that exist across different socioeconomic levels. While some of these findings accord with the tendency toward homophily in social networks[42,72], individuals' errors may overemphasize this phenomenon. For age and gender, dissimilar ties are judged with greater accuracy, and individuals expect connections between pairs with differing network degrees. Future work may consider the specific conditions and extent to which individuals rely on homophily as a heuristic to judge their networks. These biases may impede the activation of existing resources for social action. Consequently, our findings suggest that, in addition to structural interventions facilitating new connections across identity lines[69,73], it may be important to focus on strategies that make salient the overlooked intergroup ties that already exist in communities.

In the randomized controlled trial of introduced information, we found that individuals with a higher TNR are more likely to absorb novel 'non-social' information transmitted through social channels. This fits with the general pattern that accurate individuals seem to possess a less 'oversocialized' view of their social environment; one that less strongly inflates network density and highlights the actual paths of influence.

While social capital is conventionally held to inhere in the social environment, external to individuals network beliefs are probably crucial to social capital in groups[5,40,74,75]; solutions to collective action problems may be inhibited when individuals fail to properly see and activate the social ties around them. Similarly, coffee cultivation in Honduras is labour intensive, relying on seasonal community labour for the months-long annual harvest. The practice of mobilizing social ties to coordinate production, in combination with sustained interpersonal contact during the harvest, may foster this tendency of villagers to inflate the number of connections in their network, thereby seeing their communities as more cohesive. This further aligns with research showing that labour-intensive crop farmers exhibit less individualistic orientations and greater abilities in relationship building[76,77].

Work on network search has generally assumed that individuals only have knowledge of their own connections in reaching out to indirect connections[78,79]. But partial network knowledge may impact this process. Connections to individuals high in socioeconomic status is a key element of economic mobility[80], and we find that the visibility of ties among poor individuals is lower than for those of the better-off villagers. Similarly, the observed lack of awareness of ties among the less wealthy by individuals who are also poor may constrain the ability for social coordination among the poor[81].

The manifold functions of network cognition may be especially important in settings where formal institutions are absent or weak, such as the developing world, and where individuals' ability to access social capital depends on informal networks[45,82], and their broader network knowledge. Still, while we focus on villages in Honduras, social networks exhibit a common 'small-world' topology across societies[61,83,84] with short path lengths and high clustering coefficients[85,86]. Nonetheless, whether network beliefs manifest context-dependent patterns around the world is a subject for future work. Our analysis has focused on beliefs about social networks from the bottom up: we ask individuals about the existence of dyads rather than their impressions of entire networks, and we expect that individuals learn about networks through their experience of social relationships.

In sum, how individuals think about the relationships around them may substantially shape the way behaviours and ideas flow through networks—often in ways that cannot be captured by the network structure alone.

https://doi.org/10.1038/s41562-025-02221-6

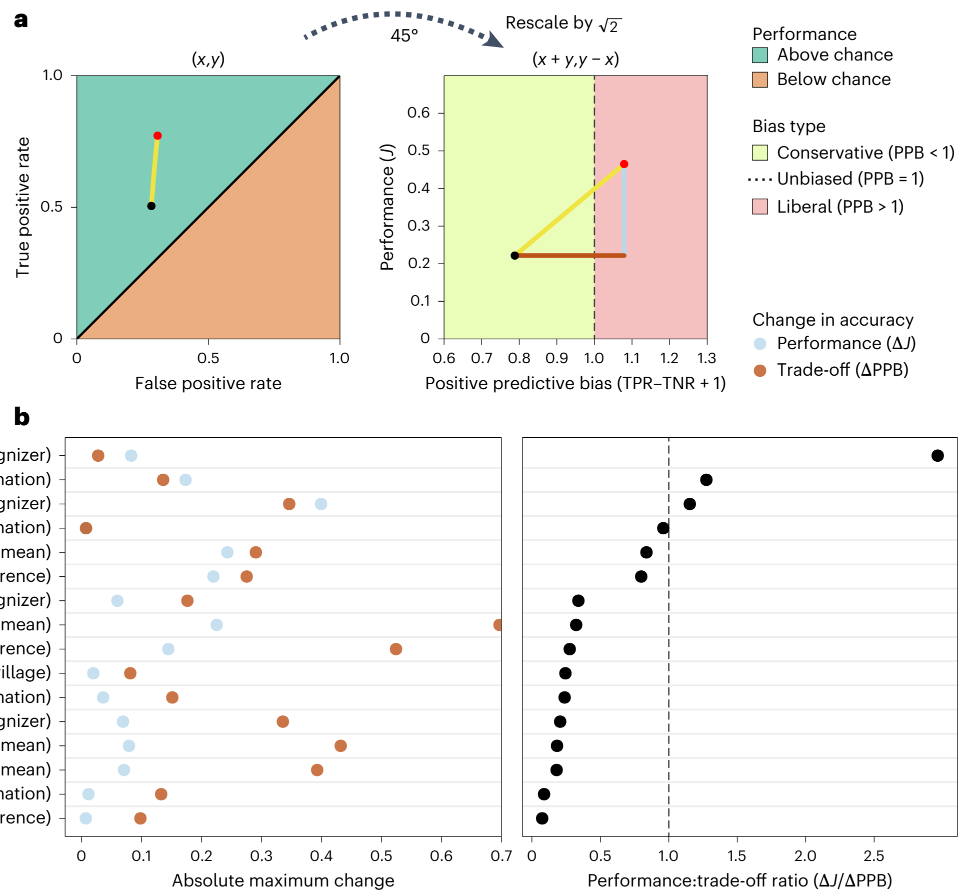


Fig. 7 | Bias in error commission. a, To further summarize the change in accuracy, we change the basis of the ROC space. The transformed axes (right) represent performance (*J*) and bias (positive predictive bias). After this operation, we decompose the vector formed by the maximum change in each dimension (light-blue and orange lines). See Supplementary Methods for details. **b**, Performance:trade-off ratio (right) and maximum change (left). The maximum change over the range of each studied attribute, whether an attribute of a survey respondent, or of a tie, is shown, as either the change in the *J* statistic, or in the positive predictive bias (left) and the ratio of the two (right). In the case of attributes of ties, an attribute may be either the mean values of the pair (mean), the absolute difference between the two (difference), or the unique combinations of qualitative values (combination).

Methods

The Yale IRB (Protocol #2000036654) and the Honduran Ministry of Health approved all data collection procedures, and all participants provided informed consent.

Local involvement

We worked closely with the local population of Copán, Honduras, sought feedback and approval from officials at the Ministry of Health (MOH) of Honduras, and endeavoured to provide practical benefits to the local community. Here we briefly summarize this history and outline some of our principles and actions in this regard.

When we began designing the underlying cohort project in 2013, the Gates Foundation introduced us to the Inter-American Development Bank (IDB), which has been supporting and doing work throughout Latin America, and IDB in turn introduced us to the Honduras MOH. Because of this pathway to getting the project launched, we worked with local and regional public health organizations and with local leaders rather than with local academic institutions.

From the outset when the original cohort was impanelled (for a randomized controlled trial initiated in 2013)[48], we sought extensive local involvement, beginning with a needs assessment where local village residents told us about topics of concern to them in a series of meetings in villages throughout Copán. In addition to extensive community input, we sought input from the MOH.

Copán is a very isolated area. Over the years, as we built our data collection team in Copán, we developed deep ties to the local community, to local village leaders, to the few local health clinics, and to local transportation and infrastructure providers. Because of these ties and our commitment to the local community, we periodically presented our results directly to these constituencies at proper intervals.

We furthermore provided specific material benefits to the local community. For instance, when people were tested for parasites as part of our study, we furnished test results and arranged their treatment. When people had their vision tested, we provided corrective glasses. We also solicited ideas from the local community about what infrastructure improvements we could make, and we repaired many local playgrounds and clinics as a result (a detailed summary is available upon request). In addition, we built capacity for development goals in the region: we hired and trained over 100 local people, and many of our former data collectors have gone on to work for other public health and development entities.

This work is not likely to result in stigmatization, incrimination, or discrimination for the participants, and we have carefully safeguarded all data from threats to the privacy or security of our participants, which has constrained the individual-level data we can release.

Survey design

We obtained individuals' beliefs about the existence of ties between others in their social network for three distinct relationship types. In other words, we asked individuals to judge whether pairs of individuals who live in their villages are connected in each of three ways: whether they are immediate kin, whether they spend free time

https://doi.org/10.1038/s41562-025-02221-6

together, and whether they discuss personal or private matters together. For example, we asked survey respondent, José, 'do María and Eduardo spend free time together?' The survey is presented in Extended Data Table 1.

We collectively represent each villager's beliefs about their social network as a three-dimensional $N \times N \times N$ array Y, where each element Y_{kij} represents person k's belief about the relationship between two individuals, i and j. We may think of Y as a series of N matrices, where each $N \times N$ 'slice' Y_{ij} represents a single individual's estimation of the underlying sociocentric network in the village they inhabit. Note that Y is the collection of beliefs for all individuals about a network defined by a particular relationship. We collected network beliefs separately for each of three relationships in 82 villages in rural Honduras.

Since it is infeasible to collect each person's beliefs about their whole network, we developed a sampling strategy that leverages our knowledge of the underlying sociocentric social networks to make the collection of data on relationship conceptions possible in such large-scale networks. Under this sampling design, we showed up to 40 distinct pairs of individuals to each survey respondent to sample beliefs from Y for each of the three relationships.

In contrast to our approach, existing work on social network cognition has almost exclusively relied on a complete survey in which researchers obtain a response from each individual in a social network about every possible tie in that network (Supplementary Fig. 10). As a result, data collection has only been performed in small networks, since surveying all ties in a network of even moderate size would require asking individuals thousands of questions. While there have been various approaches to this problem[87–89], previous attempts to model cognitive social structures data has generally not focused on the development of strategies to sample it[90–92].

For methodological simplicity and given findings on social cognition, we asked only about symmetrized relations. While social relationships are not always reciprocated[93–96] and reciprocation itself is socially patterned[97], there is some evidence that individuals are not able to accurately track the directionality of ties among others in their group[24,89]. For example, Smith may know that Jones and Brown are friends but is unlikely to know whether Jones considers Brown his friend, Brown considers Jones her friend, or both.

The broader research effort in Honduras has involved the collection of a broad set of ties, including a range of kin and non-kin ties that are either instrumental or affective in nature. All three relationship questions for the new network cognition survey used here correspond to questions used to generate the reference sociocentric networks. Moreover, we queried respondents about relationships (for example, 'do i and j spend free time together?') that were obtained just a few weeks before the data collection focused on social network beliefs.

Specifically, we asked each villager to nominate those with whom they spend free time together, discuss personal or private matters, and consider their immediate kin. We combined these responses together to form an undirected sociocentric network according to the 'union' rule, such that we say that a tie exists between two individuals if either of them reports it; for example, we say that José and Julio spend free time together if either of them reports it (see Supplementary Results for robustness checks based on alternative definitions).

The survey consisted of two parts, represented by questions 1–3 and 4–6 in Extended Data Table 1. We successively displayed the candidate pairs to each survey respondent. In addition, we displayed the faces of individuals rather than their names, to ensure that individuals are correctly identified (for example, more than one individual may share the same name in a community). When a new pair was selected, a cognizer was first asked whether they recognize each individual in the pair: for example, for a selected pair of individuals Rosa and Hector, the respondent was asked 'Do you know Rosa?' and 'Do you know Hector?' (questions 1 and 2). Then, if the respondent recognized both individuals, respondents were presented with both faces in the pair together, and the surveyor asked whether the two individuals know each other at all (question 3). If a respondent answered 'No', we assumed that the respondent believes that no tie exists between the pair, meaning that they would respond 'No' to the remaining questions. If respondents answer 'Yes' to this question, the surveyor moved to ask the main relationship questions (4–6).

The survey consists of these two stages to minimize the total survey time and avoid asking redundant questions to participants. This conditional survey logic follows the reasoning that individuals cannot meaningfully answer a query about a tie between two individuals when one or both individuals are unknown to the respondent. Therefore, participants are not asked about such ties. However, the respondents recognized the displayed individuals in the vast majority of cases (93.5% of the time, see Supplementary Table 9).

These specific relationship questions have been used to collect social network data in previous data collection waves and have been carefully defined for the study population. The two primary questions of interest for this study, 'with whom do you spend free time?' (free-time) and 'with whom do you discuss personal or private matters?' (personal-private) can be independent from each other and can apply to ties between kin or non-kin. We expected that estimation of these relationships involves the respondent's drawing on distinct social information to demonstrate knowledge for each of the two tie types.

The survey was conducted with the TRELLIS network data collection platform[48,98]. This new survey of social network knowledge took an average of 14.23 min to complete.

Sampling procedure

We developed a method to sample 40 dyads to each of the 10,072 survey respondents. To circumvent the limitations of traditional methods, we sampled the array Y for each network and relationship type. Under our sampling procedure, we made use of sociocentric networks we mapped just a few weeks earlier to show each respondent a different random set of dyads. Consequently, the reference network was current, and it was also a relevant basis of comparison for data on individual beliefs collected here.

One of the major concerns in designing a sampling method in this context is to ensure that participants are asked about ties that are socially relevant and to vary important features of ties so that we are statistically powered to answer our key research questions.

More specifically, we sampled up to 40 ties to each respondent, stratifying on geodesic distance between a respondent and a pair, and the existence of the tie in the underlying network. We sampled a fixed number of ties at 1 up to 4 degrees of separation between the respondent and the pair. Furthermore, half of the ties presented represent real connections that exist in the reference network and half represent ties that do not exist in the network. In the latter case, the pair of individuals shown are not truly connected, such that neither individual reports the other as a relation (for example, both individuals do not report the other as someone with whom they spend free time). We stratified on this distance under the reasoning that ties relatively close to an individual in their network are of fundamentally deeper relevance for network cognition, and that faraway ties are simply less likely to be cognized. We define the distance between a cognizer, k, and a pair, (i, j), as the mean of the geodesic distance from k to i and from k to j.

In addition, we stratified sampling on the basis of distance in the network defined by the union of the kin, free-time and personal-private networks rather than the individual networks. We used this combined network to simplify the method for implementation. In this 'union' network, we say that a tie exists if at least one person in the pair nominates the other for at least one of the three relationships (for example, i and j are connected if j reports that they spend free time together, even if neither nominates the other for the other relationships). Consequently, we stratified on distance between a pair and a cognizer in this network. This design choice means that while 4 is the maximum distance in the union network, respondents may be further from the pairs they judge in any of the individual network types; consequently, individuals judge ties over a wider range of geodesic distances in the network (Supplementary Figs. 1 and 12). We graphically present the survey and sampling design in Fig. 1.

Below, we characterize the sampling procedure more formally, and we detail the steps of the procedure in Supplementary Table 10. To select the ties shown to survey respondents, we specified (1) the maximum geodesic distance from which an individual that is shown in the survey may be from the respondent, and the numbers of (2) real and (3) counterfactual ties to be inquired about at each distance.

The neighbourhood of individual i up to degree d is the subset of the social network, N, restricted to nodes that are no more than d degrees away from the individual. For instance, i's neighbourhood up to degree 1 only includes i's friends and also corresponds to i's egocentric network. The neighbourhood up to degree 2 not only contains i's friends, but also the friends of i's friends. Similarly, we can define the 'social orbit' of an individual, up to some degree d, as the ego network up to degree d, with the pertinent addition of counterfactual ties. Specifically, we took the ego network of some cognizer, k, up to degree d. We then took the set of real relationships among the nodes, excluding k, and the set of possible relationships that do not exist among these nodes. Note that as d increases, the number of real ties approaches the number of ties in the whole social network minus the number of direct connections to k, and the number of counterfactual ties approaches the number of possible ties minus the number of ties that exist in the network.

In our setting, we defined a procedure to display up to 40 ties for each cognizer, k, within their social orbit up to degree 4. Consequently, we sampled up to 5 real and 5 counterfactual relationships for each distance 1 to 4 degrees away in the 'union' network, for a desired total of 20 connections that exist in the underlying network and 20 that do not exist in that network. This ensured that we were powered to make comparisons both across distances within this range and over the underlying reality of the ties. For a participant, k, we stratified the set of valid pairs, those within 4 degrees of k, into distinct bins: one bin for the ties that exist and one for those that do not, at each distance from 1 to 4; for example, one bin is composed of real ties among first-degree neighbours, another bin is composed of real ties among second-degree neighbours and between first- and second-degree neighbours of k and so on. Note that these bins would contain many more than the desired number of ties, especially as distance from the respondent increases. Moreover, there are many more ties that could exist but do not in these social networks, where density is much less than unity (the mean density of the union network is 3.42%, s.d. = 0.021). However, it is also possible that a bin would contain fewer than the desired number of ties; for example, a respondent would have fewer than five real ties among their first-degree neighbours if they only have two direct connections. Therefore, a participant might have fewer than 40 ties selected. We sampled ties from each bin uniformly at random.

We executed this procedure separately for each individual k. As a result, each person received a unique set of up to 40 ties which they were then asked about. Figure 1 illustrates the sampling procedure for a particular individual in a village network, and Fig. 1c gives an example of the survey responses for an individual survey participant, giving the underlying data structure that we analysed. No participant with a complete survey was excluded from any of the analyses presented. The mean age of the respondents was 38.901 (s.d. 17.059), and 64.873% were women.

Primary model

As a general description, we used mixed-effects logistic regression models to model the individual binary responses. As described above, survey respondents answered 'Yes' or 'No' when queried about up to 40 ties drawn from their village social networks. Furthermore, they answered three questions for each displayed pair, stating whether they believe that the two individuals are immediate kin, whether they spend free time together, and whether they discuss personal or private matters. Hence, we have repeated measures such that individuals provide three responses for each unique pair that they judge. Individuals were also nested within villages. We included random effects to account for each of these nested levels.

In addition, we split the responses into two separate datasets, containing ties that truly exist in the reference network and those that do not. We modelled each dataset separately, and thereby estimated the conditional probability that a response is 'Yes' when it exists (the true positive rate) and the probability that a response is 'Yes' when it does not exist (the false positive rate) with separate logistic regression models. Together, these separate models represented the two independent dimensions of accuracy. We further combined these models to estimate Youden's J statistic, which is the true positive rate minus the false positive rate.

We modelled the responses with a range of individual, network and village-level characteristics. In addition, we included the geodesic distance from the survey respondent to the pair as a covariate. In the model of the false positive rate, we also included the distance between the individuals in the judged pair (which is necessarily 1 in the true positive rate model). The specifications of the true and false positive rate models are identical except for this difference.

Furthermore, all the results that pertain to cognizer and tie characteristics are given by a single model, unless otherwise noted. We also present effect estimates rather than coefficient estimates since they are more interpretable and less sensitive than odds ratios[99]. Specifically, we included adjusted predictions at the mean and pairwise contrasts that correspond to discrete marginal effects at the mean. We varied a characteristic of interest (for example, age) over its observed range in the data, holding other model coefficients at their typical values, to observe its relationship with accuracy.

We give a more detailed description of our modelling strategy below. We modelled tie conception with multilevel logistic regression models that are appropriate for a binary outcome and which account for the nested structure of our data. This model allowed us to assess the determinants of accuracy in belief, and the effect of a respondent's individual, village and network characteristics on beliefs about a tie.

The data were structured such that a respondent $k \in 1, \ldots, N_g$ in a specific village g has beliefs about the relationship between, at most, 40 pairs of other individuals $i, j \in 1, \ldots, N_g$ in the network of village g, for a given relationship type $R_{[ij]g}$, where $r = 0$ for 'spends free time with' and $r = 1$ for 'discusses personal matters with'. Separately, $H_{[ij]g} \in 0, 1$ indicates whether i and j is a tie between kin or non-kin. Thus, we let $Y_{r[ij]kgc} \in 0, 1$ be our dependent variable, representing respondent k's propositional attitude (hence the superscript 'c' for 'cognized') toward relationship of type r between individuals $i, j \in 1, \ldots, N_g$ in village g. Observe that this contrasts with $Y_{r[ij]g}$, which represents whether the respective tie exists in the underlying sociocentric ('reference') network. Furthermore, we let our independent variables be individual, village and network characteristics of the respondent X_{kg}; for example, age, gender, wealth, coffee cultivation and degree centrality. We also included the average geodesic distance between individual k and the two individuals i and j in the network of type r, $D_{r[ij]kg}$, and the geodesic distance between i and j when a direct tie between them does not exist in the sociocentric network (hence, only when $Y_{r[ij]g} = 0$). In addition, we denoted by $X_{[ij]g}$ characteristics of ties such as means, absolute differences, or unique combinations of the characteristics of the individuals who compose a tie; for example, the average of the network degrees of i and j. Continuous-valued characteristics were standardized to range of 0 to 1. We fit the following mixed-effects logistic regression model separately for ties that exist in the sociocentric network ($b = 1$) and those that do not ($b = 0$):

$$\begin{aligned}\text{logit}\left[P\left(Y^{c}_{r[ij]kg} = 1 | Y_{r[ij]g} = b\right)\right] = {} & \beta^{b}_{0} + \beta^{b}_{1} r_{[ij]g} + \beta^{b}_{2} H_{[ij]g} + \mathbb{1}[D_{r[ij]kg} < \inf] \\ & \left(\beta^{b}_{3} + \beta^{b}_{4} D_{r[ij]kg}\right) + \mathbb{1}[D_{r[ij]g} < \inf, b = 0] \\ & \left(\beta^{b}_{5} + \beta^{b}_{6} D_{r[ij]g}\right) + \left(\Lambda^{b}_{1}\right)^{T} X_{kg} + \left(\Lambda^{b}_{2}\right)^{T} X_{[ij]g} \\ & + \left(\Lambda^{b}_{3}\right)^{T} X^{\upsilon_1}_{kg} X^{\upsilon_2}_{[ij]g} + \left(\Lambda^{b}_{4}\right)^{T} X_{g} + u^{b}_{g} + s^{b}_{kg}\end{aligned} \quad (1)$$

where P is probability. All coefficients were indexed by b to reflect the fact that we estimated separately for the true and false positive rates. We included random intercepts for the village (u_g) and respondent (s_{kg}). β refers to a single estimated coefficient and Λ refers to a vector of coefficients for a vector of characteristics. Specifically, Λ_1 represents the association between response and individual and network features, Λ_2 between response and features of ties, Λ_3 the interaction between selected individual and tie attributes (for example, cognizer wealth and the average wealth of the pair), and Λ_4 between response and village-level characteristics (where υ_1 and υ_2 indicate selected subsets of respondent and tie characteristics, respectively).

Note that a path may or may not exist between two individuals in the sociocentric network: i and j may exist in disconnected subcomponents of the free-time network, even if they are linked in the union of the free-time, personal-private and kin networks. When a path does not exist, we say that the distance between those individuals is effectively infinite. Consequently, we modelled the integer-valued geodesic distance as the interaction between an indicator for whether a path exists between the nodes (between both k and i, and k and j) and the geodesic distance (which is undefined when a path does not exist). We analogously represented the distance within a tie (between i and j when $b = 0$) and the distance in the kinship network (in Extended Data Fig. 1).

With this model, we assessed the extent to which individual and network characteristics of the respondent and the ties affect the likelihood of k believing that a relationship of type h exists between i and j, regardless of its existence in the reference sociocentric network. Hence, we fit equation (1) conditional on the existence ($Y_{r[ij]g} = 0$), or not ($Y_{r[ij]g} = 1$), of the relationship of type r in the reference network to investigate the determinants of accuracy of prediction. Generally, we estimated effects from the model in equation (1) ranging over a characteristic of interest while holding others at their population means or typical values in the population.

Over each condition, $b = 1$ and $b = 0$, respectively, we used these models to predict the true positive rate (TPR) and the false positive rate (FPR) as a function of an individual, tie, or network characteristic, as well as geodesic distance and tie type:

$$\text{TPR}\left(\theta^{p}; h, \theta\right) = P\left(Y^{c}_{r[ij]kg} = 1 | Y_{r[ij]g} = 1, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \theta^{p}\right) \quad (2a)$$

and

$$\text{FPR}\left(\theta^{p}; h, \theta\right) = P\left(Y^{c}_{r[ij]kg} = 1 | Y_{r[ij]g} = 0, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \theta^{p}\right) \quad (2b)$$

where $\mathbf{\Theta}_{r[ij]kg} = (R_{[ij]g}, X_{kg}, X_{[ij]g}, X_g, D_{r[ij]kg}, D_{r[ij]g})$ combines all properties, except the kinship status of the tie, into a single vector. $\mathbf{\Theta}^{\neg p}$ denotes the exclusion of the pth characteristic, and $\mathbf{\Theta}^{p}$ only contains the pth characteristic. Confidence intervals on these model predictions were estimated by the delta method. For a focal characteristic, θ^p, we estimated each rate over its (standardized) range $[0, 1]$, and held characteristics besides θ^p at their population means (that is, we held $\mathbf{\Theta}^{\neg p}$ at $\bar{\theta}^{\neg p}$). This allowed us to investigate how, on average, for instance, the distance between the cognizer and the potential tie affects, in a different manner, the specificity and sensitivity of beliefs. For example, individuals might be more likely to overestimate the number of ties in their immediate social orbit and underestimate ties further away, such that their sensitivity might increase with geodesic distance while specificity might decrease. Note that when Θ is not specified, it means that all the characteristics in Θ are not specified. We additionally estimated the marginal effect of kinship status, in which case we held each attribute in Θ at its mean or typical value in the population, such that, for example, we computed $\text{TPR}(h; \theta) = P(Y^{c}_{r[ij]kg} = 1 | Y_{r[ij]g} = 1, H_{[ij]g} = h)$.

In addition, we combined the effect estimates for each rate into a unidimensional measure of overall accuracy, Youden's J statistic. Consequently, we then derived this overall effect directly from two rates:

$$J\left(\theta^{p}; h, \theta\right) = \text{TPR}\left(\theta^{p}; h, \theta\right) - \text{FPR}\left(\theta^{p}; h, \theta\right) \quad (2c)$$

The J statistic, ranges from −1 to 1, where 0 indicates chance performance and accuracy improves from −1 up to perfect accuracy at 1. We describe the calculation and interpretation of this statistic and its relationship to receiver operator characteristic (ROC) analysis in Supplementary Information. Confidence intervals for J were calculated through a bootstrapping procedure to account for the uncertainty associated with combining the TPR and FPR estimates from the two models. We conducted a parametric bootstrap appropriate for mixed models, for the TPR and FPR models, with 10,000 iterations. For each iteration, we estimated the effects defined in equation (2a,b) to construct the bootstrap distribution of each effect. Then, we used these replicates to repeatedly calculate equation (2c) to compute the bootstrap distribution of the estimate for J and then calculated the standard error for the estimate J. We further constructed the confidence ellipses by approximation to a bivariate normal distribution. Confidence intervals were then constructed using a normal approximation. We further calculated and report contrasts (Extended Data Tables 2 and 3). We calculated discrete marginal effects at the extremes of the observed values in the population for the TPR (where $(\theta^p)^{\min}$ and $(\theta^p)^{\max}$ are the extrema of characteristic p on the standardized range):

$$\begin{aligned} & = \text{DTPR}\left(\left(\theta_p\right)^{\text{extrema}}; h, \theta\right) \\ & = P\left(Y^{c}_{r[ij]kg} = 1 | Y_{[ij]hg} = 1, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{\max}\right) \\ & - P\left(Y^{c}_{r[ij]kg} = 1 | Y_{[ij]hg} = 1, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{\min}\right)\end{aligned} \quad (3a)$$

and

$$\begin{aligned} & = \text{DFPR}\left(\left(\theta_p\right)^{\text{extrema}}; h, \theta\right) \\ & = P\left(Y^{c}_{r[ij]kg} = 1 | Y_{[ij]hg} = 0, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{\max}\right) \\ & - P\left(Y^{c}_{r[ij]kg} = 1 | Y_{[ij]hg} = 0, H_{[ij]g} = h, \mathbf{\Theta}^{\neg p} = \bar{\theta}^{\neg p}, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{\min}\right)\end{aligned} \quad (3b)$$

and analogously for J. Similarly, we calculated marginal effects over the unique combinations of categorical (and, degenerately, binary) characteristics. Furthermore, for covariates that exhibited a curvilinear or parabolic relationship with accuracy, we derived the value of the covariate with the maximum accuracy score, $(\theta^p)^*$, over the observed range of θ^p. We then contrasted $(\theta^p)^*$ with both the minimum and maximum of θ^p. We did this analogously to the contrasts in equation (3a,b). For example, for the true positive rate, we calculated

$$\begin{aligned}\text{DTPR}\left(\left(\theta^{p}\right)^{\text{extremum}}, \left(\theta^{p}\right)^{*}; h, \theta\right) = P\left(\ldots, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{*}\right) \\ - P\left(\ldots, \mathbf{\Theta}^{p} = \left(\theta^{p}\right)^{\text{extremum}}\right)\end{aligned} \quad (4)$$

As above, we did so separately for each accuracy metric. For example, we estimated accuracy to be highest at around age 31, which we compared to the maximum and minimum age in the observed population. In general, we present estimates and effects that stratify on the kin status of the tie, R_{ijg}, due to its large effect on the model. ROC-space plots (explained in Fig. 2) present the effects stratified by kin status, while the effect plots (for example, right of each panel in Extended Data Fig. 2) only present estimates for ties between individuals who are not kin.

Finally, each participant k in each village g received an accuracy score for each rate (including J) and relationship (free-time and personal-private), corresponding to the model-adjusted effect estimate using the average cognizer-to-tie distance, the average i to j distance (for the FPR only), the average values of the tie properties and the cognizer-specific values of cognizer properties (that is, we used the respondent's particular characteristics). We defined the respondent-level accuracy scores, respectively, as

$$\mathrm{TPR}_{kg}(r,h) = \mathrm{TPR}\left(H_{[ij]g} = h, R_{[ij]g} = r, X_{kg}, X_g\right) \quad (5a)$$

$$\mathrm{FPR}_{kg}(r,h) = \mathrm{FPR}\left(H_{[ij]g} = h, R_{[ij]g} = r, X_{kg}, X_g\right) \quad (5b)$$

and

$$J_{kg}(r,h) = \mathrm{TPR}(r,h) - \mathrm{FPR}(r,h) \quad (5c)$$

where the set of attributes at the cognizer and village (X_{kg} and X_g, respectively) were held at their values for each respondent. All characteristics left implicit were held at their population means or typical values. Note also that we stratified separately on the kinship status of the tie and the relationship type.

Two-stage estimation

As described above, we used equation (5a–c) to calculate a score for each individual and rate type, which were further combined to a measure of accuracy for each participant for each rate. These respondent-level estimated accuracy scores held $H_{ijg} = 0$, such that we only considered the case of non-kin ties between individuals (i and j).

We used these scores in a second-stage regression model to estimate the relationship between the true positive and false positive rates (Fig. 6a), and between social network acuity and knowledge of an exogenous health-related intervention.

To model the relationship between the TPR and FPR, we estimated an ordinary least squares (OLS) model:

$$\mathrm{TPR}_{kg}(r,h) = \beta_0 + \beta_1 \mathrm{FPR}_{kg}(r,h) + \beta_1 FPR^2_{kg}(r,h) + \beta_2 r_{r[ij]g} + \mathbf{\Lambda}_1^T \boldsymbol{X}_{kg} + \epsilon \quad (6)$$

where TPR_{kg} and FPR_{kg} correspond to the first-stage model predictions for each participant and rate. We made further adjustments at the second stage for key characteristics (degree, age, age^2, gender, religion, wealth, indigenous status) and relationship type. ϵ represents an error term. We present this analysis in Fig. 6a, as estimated marginal means over the observed range of false positive rates. We included FPR and its square to account for the diminishing marginal returns in the ability to detect true positives associated with additional increases in the FPR.

In addition, we modelled riddle knowledge, R_{lkg}, which is a binary 'Yes'/'No' (coded as 1 or 0) response variable for a specific riddle (indexed by l), and L_l is a categorical variable that refers to the specific riddle. Specifically, we included three distinct riddles, which are described in Supplementary Table 5. Commensurately, we fit a logistic model, adjusting for cognizer-level demographic characteristics:

$$\mathrm{logit}\left[P(R_{lkg} = 1)\right] = \beta_0 + \beta_1 A_{kg}(r,h) + \beta_2 \mathbb{1}[l=2] + \beta_3 \mathbb{1}[l=3] + \mathbf{\Lambda}_1^T \boldsymbol{X}_{kg} \quad (7)$$

Here, R_{lkg} represents a particular binary riddle outcome (whether the cognizer knows riddle l), where A denotes one of TPR, FPR, or J, representing k's accuracy score, and X_{kg} demarcates the cognizer characteristics. Estimated marginal means across the range of predicted responses are presented in Fig. 6b. We estimated separate models for each accuracy metric.

For each estimated two-stage model, we accounted for uncertainty at both the first (equation 1) and second stage (equations 6 and 7) of estimation via bootstrapping. Specifically, for the first-stage model (equation 1), we conducted a parametric bootstrap appropriate for mixed models with 1,000 iterations. We simulated 1,000 response vectors, and refit the model in equation (1) each time. At each first-stage iteration, we estimated the cognizer accuracy scores (according to equation 5a–c) and then estimated the second-stage model (equation 6 or 7, as appropriate) 1,000 times in a second-level parametric bootstrap. The outer iterations captured the uncertainty in estimation of the mixed-effects response model, and the repeated second-stage fits likewise captured the uncertainty in estimation of the outcome of interest. We collected the 1,000,000 replicates of each second-stage model parameter to calculate the adjusted standard errors of each model coefficient.

Genetic relatedness index

In addition to using self-reported kinship, we also collected genetic data for 17 villages ($n = 1{,}333$ individuals). We used the KING framework to measure genetic relatedness, developed for genome-wide association studies (GWAS)[100]. The assumption that the genotypes for all individuals in our population arise from a common set of allele frequencies was not met in our setting, where individuals come from different ethnicities. Consequently, we used the KING-robust method. While this method may be less directly interpretable than KING-homo, which ranges in the 0 to 1 interval, KING-robust gives accurate estimates of relatedness in the presence of population stratification.

This method estimates a kinship coefficient φ_{ij}, for pair of individuals (i, j), which is defined as the probability that alleles sampled at random are identical by descent. This measure is unbounded from below, such that individuals are considered unrelated if they have a relatedness score $\varphi \le 0$, and ranges to $\varphi = \frac{1}{2}$ marking monozygotic twins. Established thresholds were used to infer that, for example, a relatedness score of $\frac{1}{4}$ is taken to imply that individuals are full siblings, and a score of $\frac{1}{16}$ is taken to represent 3rd degree relatives. While individuals were considered effectively unrelated with $\varphi \le 0$, the measure was defined[100] as

$$\hat{\varphi}_{i,j} = \frac{N_{\mathrm{Aa;Aa}} - 2N_{\mathrm{AA;aa}}}{N_{i;\mathrm{Aa}} + N_{j;\mathrm{Aa}}} \quad (8)$$

where $N_{\mathrm{Aa};Aa}$ represents the total count of single nucleotide polymorphisms (SNPs) where both individuals i and j are heterozygous; analogously, $N_{\mathrm{AA;aa}}$ represents the count of SNPs where both individuals are differently homozygous (that is, i is AA and j is aa, or the reverse). In the denominator, $N_{i;Aa}$ represents the count of SNPs where i is heterozygous (and the same for j). Consequently, the measure is negative when the count of heterozygous locations is less than twice the sum of differently homozygous counts in the pair. Moreover, the index has a natural continuous interpretation such that individuals with increasingly negative values may be considered more genetically distant. This may reflect population structure[101–103], which is known to be present in the current setting where the population is an admixture of European and Mayan indigenous populations.

In Supplementary Fig. 12, we observed that in the sample, the average pair of individuals shown to the survey respondents have a relatedness score of around 0.1. The kinship coefficient is provided by the KING-robust algorithm used specifically to execute GWAS studies and identify relatedness without relying on previous kinship information or assuming population homogeneity (which has been found to lead previous methods to be biased towards relatedness).

In Fig. 3b, we estimated the relationship between social network accuracy using each accuracy metric and this bounded continuous measure of kinship. We estimated separate models for the TPR and FPR that replace the binary self-reported kinship measure $H_{[ij]g}$ with the quantity defined here, in equation (1). We made the analogous switch for the estimates for distance in the kinship network and the specific kin category, in Extended Data Fig. 1.

In addition, as a further robustness check on the estimates from the 'KING-robust' method, we have added results from the KING-homo method in Supplementary Fig. 13. Both the KING and KING-robust algorithms are included in the second-generation version of the PLINK software[103] used to calculate the relatedness measures. Furthermore, we present the distribution of kinship coefficients for each self-reported immediate kinship relation in Supplementary Fig. 14, which indicates that the self-reports yielded genetic kinship values within the expected ranges (with a mean close to 0.25)[100].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Compliant with our privacy and confidentiality assurances to our research participants and with other legal obligations, data will be made available on our secure server, subject to data release provisions in force at Yale and the Yale Institute for Network Science (or successor entities) at the time of release. Access to data requires proof of IRB approval and human participants certification. Contact nicholas.christakis@yale.edu for inquiries regarding the data.

Code availability

All analysis was conducted in the Julia programming language[104] (v.1.10.2). The sampling procedure was executed with the 'SamplingPerceivedNetworks.jl' Julia package[105] (which we are pleased to release). See the Supplementary Methods for details on software packages used. Additional paper replication materials are available on GitHub at https://github.com/emfeltham/honduras-css-paper-release.git (ref. 106). See Supplementary Results for further details.

References

1. White, H. C. *Identity and Control: A Structural Theory of Social Action* (Princeton Univ. Press, 1992).
2. Christakis, N. A. *Blueprint: The Evolutionary Origins of a Good Society* (Little, Brown, 2019).
3. Bearman, P. S. & Moody, J. Suicide and friendships among American adolescents. *Am. J. Public Health* **94**, 89–95 (2004).
4. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
5. Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
6. Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
7. Beaman, L., BenYishay, A., Magruder, J. & Mobarak, A. M. Can network theory-based targeting increase technology adoption? *Am. Econ. Rev.* **111**, 1918–1943 (2021).
8. Campbell, D. E. Social networks and political participation. *Annu. Rev. Polit. Sci.* **16**, 33–48 (2013).
9. Carlston, D. in *The Oxford Handbook of Social Cognition* (ed. Carlston, D.) 2–15 (Oxford Univ. Press, 2013).
10. Wegner, D. M. & Vallacher, R. R. *Implicit Psychology: An Introduction to Social Cognition* (Oxford Univ. Press, 1977).
11. Saxe, R. & Kanwisher, N. People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *NeuroImage* **19**, 1835–1842 (2003).
12. Bowles, S. & Gintis, H. *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton Univ. Press, 2013).
13. Boyer, P. *Minds Make Societies: How Cognition Explains the World Humans Create* (Yale Univ. Press, 2018).
14. Mascaro, O. et al. Human and animal dominance hierarchies show a pyramidal structure guiding adult and infant social inferences. *Nat. Hum. Behav.* **7**, 1294–1306 (2023).
15. Bothner, M. S., Smith, E. B. & White, H. C. A model of robust positions in social networks. *Am. J. Sociol.* **116**, 943–992 (2010).
16. Dunbar, R. I. M., Marriott, A. & Duncan, N. D. C. Human conversational behavior. *Hum. Nat.* **8**, 231–246 (1997).
17. Dunbar, R. I. M. *Grooming, Gossip, and the Evolution of Language* (Harvard Univ. Press, 1998).
18. Jackson, M. O., Rodriguez-Barraquer, T. & Tan, X. Social capital and social quilts: network patterns of favor exchange. *Am. Econ. Rev.* **102**, 1857–1897 (2012).
19. Krems, J. A., Williams, K. E. G., Aktipis, A. & Kenrick, D. T. Friendship jealousy: one tool for maintaining friendships in the face of third-party threats? *J. Pers. Soc. Psychol.* **120**, 977–1012 (2021).
20. Bourdieu, P. *Distinction: A Social Critique of the Judgement of Taste* (Harvard Univ. Press, 1979).
21. Dessí, R., Gallo, E. & Goyal, S. Network cognition. *J. Econ. Behav. Organ.* **123**, 78–96 (2016).
22. Banerjee, A. V. A simple model of herd behavior. *Q. J. Econ.* **107**, 797–817 (1992).
23. White, H. C. Notes on the constituents of social structure. Soc. Rel. 10 - Spring '65. *Sociologica* https://www.rivisteweb.it/doi/10.2383/26576 (2008).
24. De Soto, C. B. Learning a social structure. *J. Abnorm. Soc. Psychol.* **60**, 417–421 (1960).
25. Parkinson, C., Kleinbaum, A. M. & Wheatley, T. Spontaneous neural encoding of social network position. *Nat. Hum. Behav.* **1**, 0072 (2017).
26. Brashears, M. E. Humans use compression heuristics to improve the recall of social networks. *Sci. Rep.* **3**, 1513 (2013).
27. Son, J., Vives, M.-L., Bhandari, A. & FeldmanHall, O. Replay shapes abstract cognitive maps for efficient social navigation. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-024-01990-w (2024).
28. Krackhardt, D. Cognitive social structures. *Soc. Netw.* **9**, 109–134 (1987).
29. Casciaro, T. Seeing things clearly: social structure, personality, and accuracy in social network perception. *Soc. Netw.* **20**, 331–351 (1998).
30. Seyfarth, R. M., Cheney, D. L. & Marler, P. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094 (1980).
31. Kubenova, B. et al. Triadic awareness predicts partner choice in male–infant–male interactions in Barbary macaques. *Anim. Cogn.* **20**, 221–232 (2017).
32. Thomas, A. J., Woo, B., Nettle, D., Spelke, E. & Saxe, R. Early concepts of intimacy: young humans use saliva sharing to infer close relationships. *Science* **375**, 311–315 (2022).
33. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
34. O'Keefe, J. & Nadel, L. The hippocampus as a cognitive map. *Behav. Brain Sci.* **2**, 520–533 (1979).
35. Bellmund, J. L. S., Gärdenfors, P., Moser, E. I. & Doeller, C. F. Navigating cognition: spatial codes for human thinking. *Science* **362**, eaat6766 (2018).
36. Tavares, R. M. et al. A map for social navigation in the human brain. *Neuron* **87**, 231–243 (2015).
37. Basyouni, R. & Parkinson, C. Mapping the social landscape: tracking patterns of interpersonal relationships. *Trends Cogn. Sci.* **26**, 204–221 (2022).

38. Brands, R. A. Cognitive social structures in social network research: a review. *J. Organ. Behav.* **34**, S82–S103 (2013).
39. Lin, N. *Social Capital: A Theory of Social Structure and Action* (Cambridge Univ. Press, 2001).
40. Burt, R. S. *Neighbor Networks: Competitive Advantage Local and Personal* (Oxford Univ. Press, 2010).
41. Burt, R. S. Structural holes and good ideas. *Am. J. Sociol.* **110**, 349–399 (2004).
42. Kossinets, G. & Watts, D. J. Origins of homophily in an evolving social network. *Am. J. Sociol.* **115**, 405–450 (2009).
43. Lévi-Strauss, C. *The Elementary Structures of Kinship* (Beacon Press, 1969).
44. White, H. C. *An Anatomy of Kinship: Mathematical Models for Structures of Cumulated Roles* (Prentice-Hall, 1963).
45. Bearman, P. Generalized exchange. *Am. J. Sociol.* **102**, 1383–1415 (1997).
46. Henrich, J. *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous* (Penguin, 2020).
47. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
48. Airoldi, E. M. & Christakis, N. A. Induction of social contagion for diverse outcomes in structured experiments in isolated villages. *Science* **384**, eadi5147 (2024).
49. Kilduff, M. & Krackhardt, D. Bringing the individual back in: a structural analysis of the internal market for reputation in organizations. *Acad. Manage. J.* **37**, 87–108 (1994).
50. Moscona, J., Nunn, N. & Robinson, J. A. Keeping it in the family: lineage organization and the scope of trust in Sub-Saharan Africa. *Am. Econ. Rev.* **107**, 565–571 (2017).
51. Alesina, A. & Giuliano, P. The power of the family. *J. Econ. Growth* **15**, 93–125 (2010).
52. Enke, B. Kinship, cooperation, and the evolution of moral systems. *Q. J. Econ.* **134**, 953–1019 (2019).
53. Migliano, A. B. et al. Characterization of hunter–gatherer networks and implications for cumulative culture. *Nat. Hum. Behav.* **1**, 0043 (2017).
54. Greif, A. Family structure, institutions, and growth: the origins and implications of western corporations. *Am. Econ. Rev.* **96**, 308–312 (2006).
55. La Ferrara, E. in *Culture, Institutions, and Development: New Insights Into an Old Debate* (eds Platteau, J.-P. & Peccoud, R.) Ch. 5 (Routledge, 2010).
56. Small, M. L. *Someone to Talk To* (Oxford Univ. Press, 2017).
57. Small, M. L., Brant, K. & Fekete, M. The avoidance of strong ties. *Am. Sociol. Rev.* **89**, 615–649 (2024).
58. Bird, D. W., Bird, R. B., Codding, B. F. & Zeanah, D. W. Variability in the organization and size of hunter–gatherer groups: foragers do not live in small-scale societies. *J. Hum. Evol.* **131**, 96–108 (2019).
59. Hill, K. R. et al. Co-residence patterns in hunter–gatherer societies show unique human social structure. *Science* **331**, 1286–1289 (2011).
60. Glowacki, L. et al. Formation of raiding parties for intergroup violence is mediated by social network structure. *Proc. Natl Acad. Sci. USA* **113**, 12114–12119 (2016).
61. Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. Social networks and cooperation in hunter–gatherers. *Nature* **481**, 497–501 (2012).
62. Graeber, D. & Wengrow, D. *The Dawn of Everything: A New History of Humanity* (Farrar, Straus and Giroux, 2021).
63. Yenigün, D., Ertan, G. & Siciliano, M. Omission and commission errors in network cognition and network estimation using ROC curve. *Soc. Netw.* **50**, 26–34 (2017).
64. Stanley, J. *Knowledge and Practical Interests* (Oxford Univ. Press, 2005).
65. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proc. Natl Acad. Sci. USA* **111**, 17093–17098 (2014).
66. Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl Acad. Sci. USA* **108**, 19193–19198 (2011).
67. Eubank, N. Social networks and the political salience of ethnicity. *Q. J. Polit. Sci.* **14**, 1–39 (2019).
68. Koopmans, R. & Veit, S. Cooperation in ethnically diverse neighborhoods: a lost-letter experiment. *Polit. Psychol.* **35**, 379–400 (2014).
69. Mousa, S. Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science* **369**, 866–870 (2020).
70. Bowles, S. & Gintis, H. The moral economy of communities: structured populations and the evolution of pro-social norms. *Evol. Hum. Behav.* **19**, 3–25 (1998).
71. Hartch, T. *The Rebirth of Latin American Christianity* (Oxford Univ. Press, 2014).
72. Fu, F., Nowak, M. A., Christakis, N. A. & Fowler, J. H. The evolution of homophily. *Sci. Rep.* **2**, 845 (2012).
73. Paluck, E. L. Interventions Aimed at the Reduction of Prejudice and Conflict. in *The Oxford Handbook of Intergroup Conflict* (ed. Tropp, L. R.) 179–192 (Oxford Univ. Press, 2012).
74. Coleman, J. S. Social capital in the creation of human capital. *Am. J. Sociol.* **94**, S95–S120 (1988).
75. Chetty, R. et al. Social capital I: measurement and associations with economic mobility. *Nature* **608**, 108–121 (2022).
76. Talhelm, T. et al. Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science* **344**, 603–608 (2014).
77. Thomson, R. et al. Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proc. Natl Acad. Sci. USA* **115**, 7521–7526 (2018).
78. Watts, D. J., Dodds, P. S. & Newman, M. E. J. Identity and search in social networks. *Science* **296**, 1302–1305 (2002).
79. Dodds, P. S., Muhamad, R. & Watts, D. J. An experimental study of search in global social networks. *Science* **301**, 827–829 (2003).
80. Chetty, R. et al. Social capital II: determinants of economic connectedness. *Nature* **608**, 122–134 (2022).
81. Gould, R. V. *Insurgent Identities: Class, Community, and Protest in Paris from 1848 to the Commune* (Univ. Chicago Press, 1995).
82. Banerjee, A. et al. Changes in social network structure in response to exposure to formal credit markets. *Rev. Econ. Stud.* **91**, 1331–1372 (2024).
83. Christakis, N. A. *Blueprint: The Evolutionary Origins of a Good Society* (Little, Brown Spark, 2019).
84. Perkins, J. M., Subramanian, S. V. & Christakis, N. A. Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries. *Soc. Sci. Med.* **125**, 60–78 (2015).
85. Watts, D. J. Networks, dynamics, and the small-world phenomenon. *Am. J. Sociol.* **105**, 493–527 (1999).
86. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature* **393**, 440–442 (1998).
87. Iorio, A. Brokers in disguise: the joint effect of actual brokerage and socially perceived brokerage on network advantage. *Adm. Sci. Q.* **67**, 769–820 (2022).
88. Flynn, F. J., Reagans, R. E. & Guillory, L. Do you two know each other? Transitivity, homophily, and the need for (network) closure. *J. Pers. Soc. Psychol.* **99**, 855–869 (2010).
89. Almaatouq, A., Radaelli, L., Pentland, A. & Shmueli, E. Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. *PLoS ONE* **11**, e0151588 (2016).

90. Butts, C. T. Network inference, error, and informant (in)accuracy: a Bayesian approach. *Soc. Netw.* **25**, 103–140 (2003).
91. Batchelder, W. H., Kumbasar, E. & Boyd, J. P. Consensus analysis of three-way social network data. *J. Math. Sociol.* **22**, 29–58 (1997).
92. Sosa, J. & Rodríguez, A. A latent space model for cognitive social structures data. *Soc. Netw.* **65**, 85–97 (2021).
93. Shah, A. K. & LaForest, M. Knowledge about others reduces one's own sense of anonymity. *Nature* **603**, 297–301 (2022).
94. Freeman, L. C. Filling in the blanks: a theory of cognitive categories and the structure of social affiliation. *Soc. Psychol. Q.* **55**, 118–127 (1992).
95. Heider, F. *The Psychology of Interpersonal Relations* (Psychology Press, 2013).
96. Krackhardt, D. & Kilduff, M. Whether close or far: social distance effects on perceived balance in friendship networks. *J. Pers. Soc. Psychol.* **76**, 770–782 (1999).
97. Vaquera, E. & Kao, G. Do you like me as much as I like you? Friendship reciprocity and its effects on school outcomes among adolescents. *Soc. Sci. Res.* **37**, 55–72 (2008).
98. Lungeanu, A. et al. Using Trellis software to enhance high-quality large-scale network data collection in the field. *Soc. Netw.* **66**, 171–184 (2021).
99. Norton, E. C., Dowd, B. E. & Maciejewski, M. L. Marginal effects—quantifying the effect of changes in risk factors in logistic regression models. *JAMA* **321**, 1304–1305 (2019).
100. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
101. Thornton, T. et al. Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91**, 122–138 (2012).
102. Wei Min, C. KING Tutorial: Relationship Inference. *KING: Kinship-based INference for Gwas* https://www.kingrelatedness.com/manual.shtml (2023).
103. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
104. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).
105. Feltham, E. SamplingPerceivedNetworks.jl. *GitHub* https://github.com/human-nature-lab/SamplingPerceivedNetworks.jl (2023).
106. Feltham, E. Honduras-CSS-Paper-NHB. *GitHub* https://github.com/emfeltham/honduras-css-paper-release.git (2025).

Acknowledgements

We thank all the study participants and the local organizations, doctors and community leaders in Honduras with whom we have interacted. We have benefited from many local connections and support, including the following local partner organizations that played a role in the development or surveying of our cohort: Soluciones para Estudios de la Salud (SES), MURE Consultores, and World Vision. We thank the Ministry of Health in Honduras and Inter-American Developmental Bank for their extensive support and cooperation; J. E. Gámez and E. J. Urrea Carbajal for coordinating the field work in Honduras; R. Negron, L. Nicoll, A.L. Rodriguez de la Rosa, and T. Keegan for support with respect to field operations, data collection and data management; W. Israel for work on modifying the TRELLIS interface to collect the data; B. Valentin and E. Liu for research assistance; D. Gilbert, S. Lou and M. Gerstein for helpful comments. This research was supported by NIH grants R01AG081814 (N.A.C.) and R01AG062668 (N.A.C.) from the National Institute on Aging. Development of the underlying cohort was supported by the Bill and Melinda Gates Foundation (N.A.C.) and the Pershing Square Foundation (N.A.C.), and the genotype information was collected with partial support from the Rothberg Catalyzer (N.A.C.) and the NOMIS Foundation (N.A.C.). Additional support was also obtained from The Paul Graham Foundation (N.A.C.). The funders played no role in the design, data collection or analysis of this paper.

Author contributions

E.F. conceptualized the project with L.F. and N.A.C. N.A.C. acquired funding. E.F. developed the methodology with L.F. and N.A.C. E.F. conducted formal analysis and wrote the original draft. All authors reviewed and edited the paper.

Competing interests

The authors declare no competing interests.

Additional information

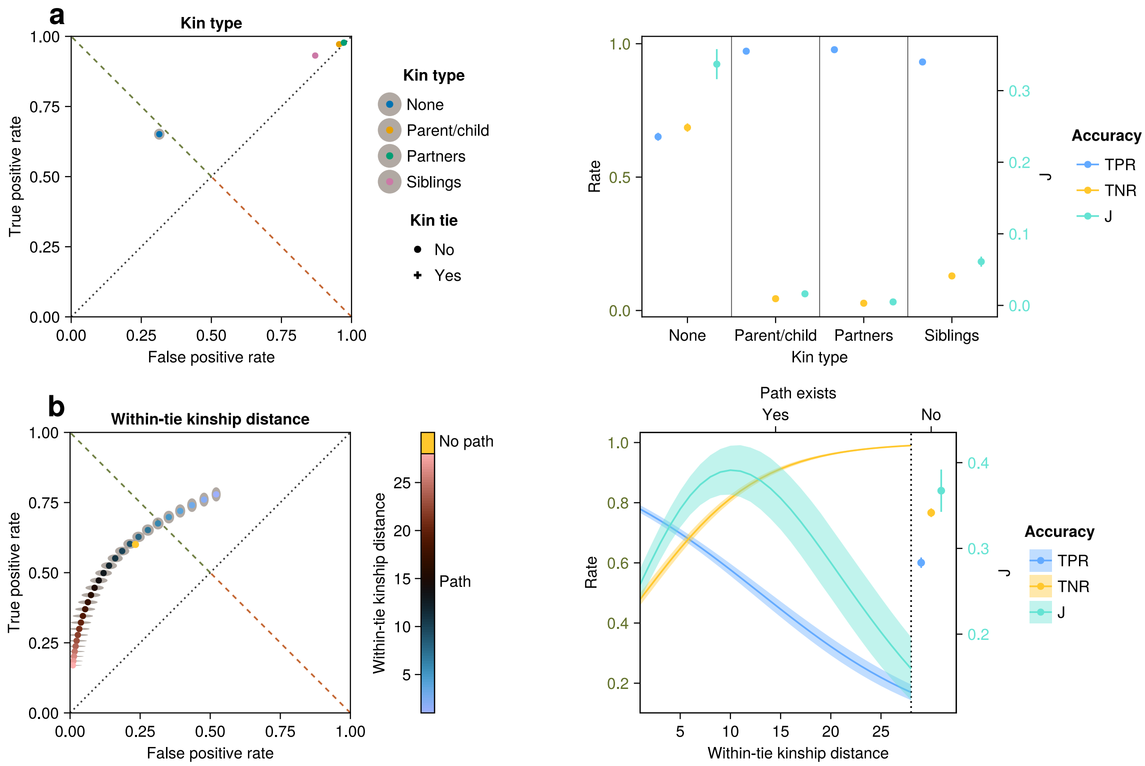
Extended data is available for this paper at https://doi.org/10.1038/s41562-025-02221-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-025-02221-6.

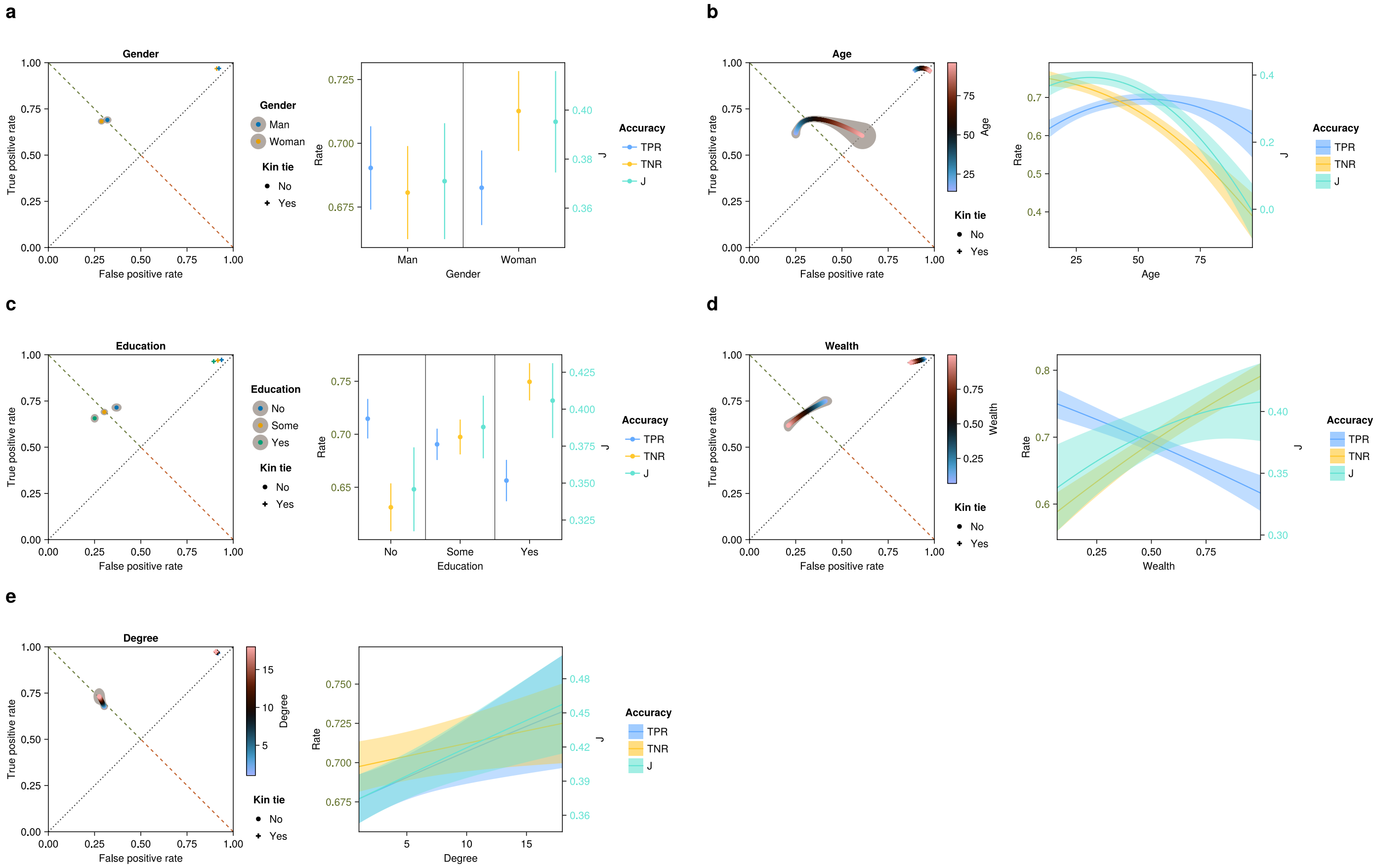
Correspondence and requests for materials should be addressed to Nicholas A. Christakis.

Peer review information *Nature Human Behaviour* thanks Michael Macy, Javier Mejia, Doug Speed and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41562-025-02221-6

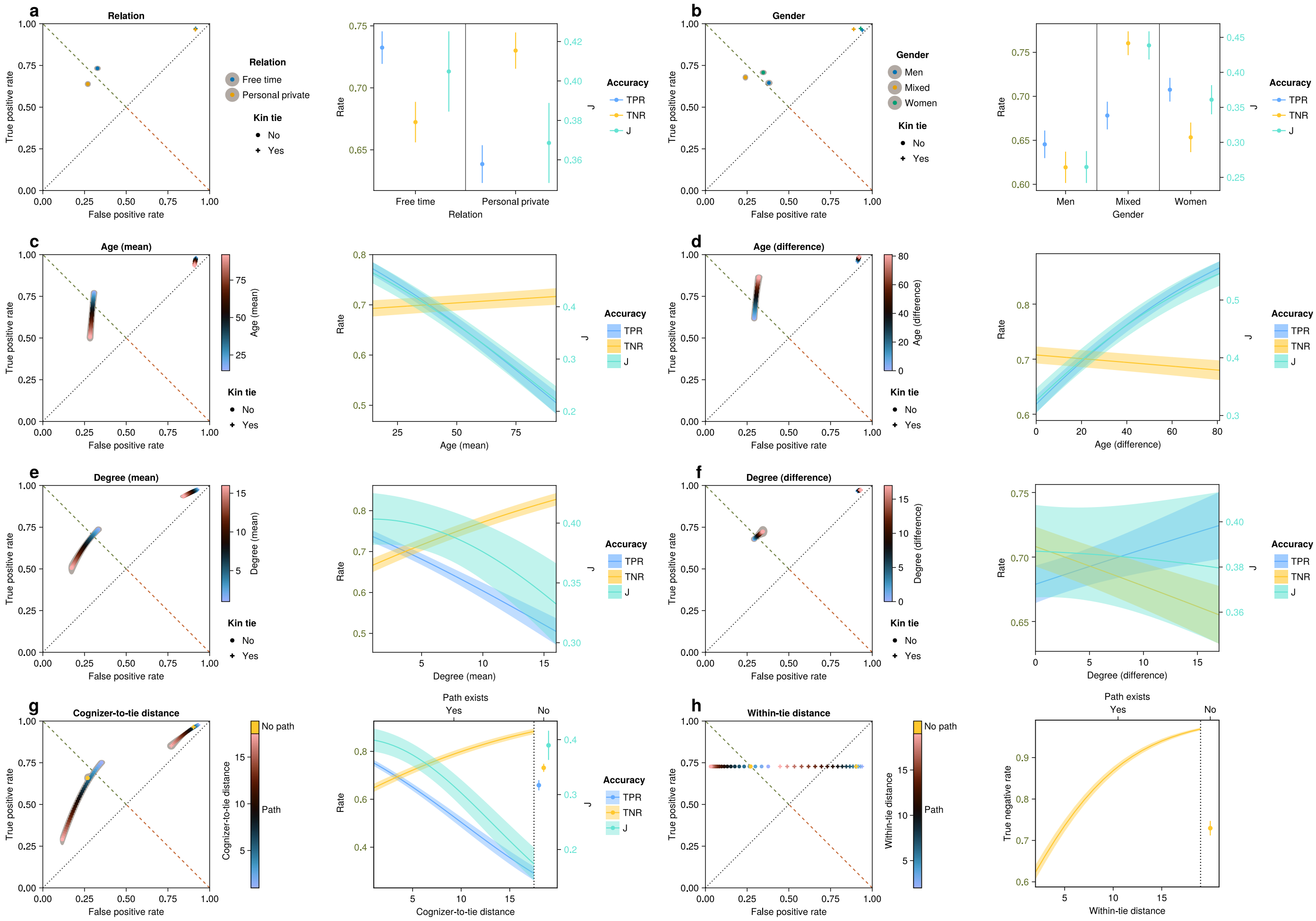


Extended Data Fig. 1 | Alternative definitions of kinship. In addition to the binary definition of kinship (used in the primary analyses) and genetic relatedness (Fig. 3b), we consider the effects of **(a)** specific type of kinship tie as a categorical variable and **(b)** distance in the kinship network. We find that the categorical results are consistent with the binary definition, and distance in the kinship network broadly corresponds to that of genetic relationship. In both panels, gray bands (LHS) displays 95% confidence ellipses around the mean estimates. Error bars (RHS) display 95% confidence interval around the mean estimates. Results are from n = 9,998 survey respondents in both panels, corresponding to 177,928 individual responses for the TPR estimates, and 477,393 responses for the FPR estimates.



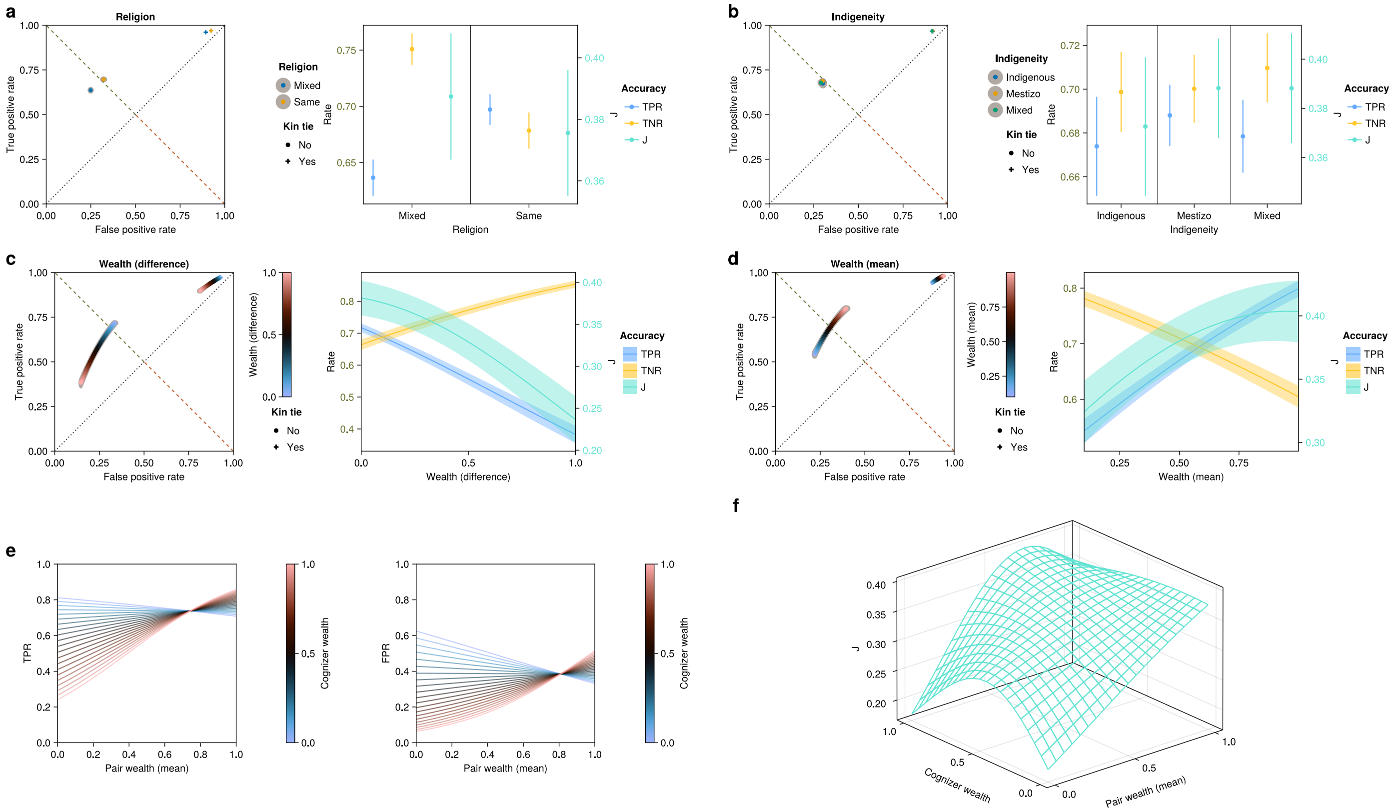
Extended Data Fig. 2 | Individual determinants of respondent accuracy. We observe that several key demographic characteristics are associated with an individual's ability to accurately predict the ties in their village network. In each panel, the left-hand image shows the marginal effect of the cognizer characteristic on accuracy in ROC-space (grey shading represents the 95% bootstrapped confidence ellipse of the predictions from the two models), and the right-hand image shows the marginal effect with respect to each individual accuracy measure: the true positive rate, false positive rate, and the overall summary measure of accuracy (Youden's *J*). Intervals represent 95% confidence levels, calculated via normal approximation for the two rates, and bootstrapped for the *J* statistic. **(a)** Gender, **(b)** Age, **(c)** Education, **(d)** Wealth and **(e)** Network degree (here, effectively an average of the count of first-degree neighbors for the two relationships analysed, personal-private or free-time). Supplementary Fig. 7 presents additional characteristics. Results are from n = 9,998 survey respondents in both panels, corresponding to 177,928 individual responses for the TPR estimates, and 477,393 responses for the FPR estimates.

https://doi.org/10.1038/s41562-025-02221-6



Extended Data Fig. 3 | Tie determinants of respondent accuracy. We find that a range of properties of ties have statistically significant associations with their tendency to be accurately conceived. In each panel, LHS, marginal effect on accuracy in ROC-space. Grey shading represents the 95% bootstrapped confidence ellipse of the predictions from the two models. RHS, marginal effect of each individual accuracy measure: the true positive and false positive rates and the summary measure, Youden's *J*. Intervals represent 95% confidence levels, calculated via normal approximation for the two rates, and bootstrapped for *J*, around the mean estimates. Estimates are stratified by whether they are of a tie among kin or not. **(a)** Relationship type; we include a covariate for the two relationships considered, free-time or personal-private. **(b)** Gender combination of tie members, *for example*, both women or both men. **(c)** Average age of tie members. **(d)** Difference in age between tie members. **(e)** Average degree of tie members. **(f)** Difference in degree between tie members. **(g)** Cognizer-to-tie geodesic distance. Individuals may or may not have a defined path between them in the reference network; when there is a path, individuals exist at a geodesic distance defined as the minimum number of steps between them; note that individuals who do not have a path between them necessarily have a path in at least one of other networks considered in this study, by design. **(h)** Distance between tie members. When a tie does not exist between two individuals, a specific geodesic distance may separate them (or they may have no path between them in the network). The TPR is set to the population average; but it does not have a meaningful interpretation in assessments of ties that do not exist. Parameters are fit from separate models of each rate, conditional on tie verity in the reference network. See Methods for details of model specification. Results are from n = 9,998 survey respondents in both panels, corresponding to 177,928 individual responses for the TPR estimates, and 477,393 responses for the FPR estimates.

https://doi.org/10.1038/s41562-025-02221-6



Extended Data Fig. 4 | Tie social identity determinants of respondent accuracy. We find that characteristics related to the social identity of a pair of individuals (*i* and *j*) affects how well that tie is conceived of by individuals *k*. **(a-d)** LHS, marginal effects on accuracy in ROC-space. Grey shading represents the 95% bootstrapped confidence ellipse of the predictions from the two models. RHS, marginal effect of each individual accuracy measure: the true positive and false positive rates and the summary measure, Youden's *J*. Intervals represent 95% confidence levels, calculated via normal approximation for the two rates, and bootstrapped for *J*. **(a)** Religion combination of tie members. **(b)** Indigenous status of the pair. Parameters are fit from separate models of each rate, conditional on tie verity in the reference network. **(c)** Absolute difference in wealth between the tie members. **(d)** Average wealth of the tie members. **(e)** Interaction between the average wealth of a pair and the cognizer's wealth on the (LHS) TPR and (RHS) FPR. **(f)** Interaction between the average wealth of a pair and the cognizer's wealth on the summary measure, *J*. See Methods for details of model specification. Results are from n = 9,998 survey respondents in both panels, corresponding to 177,928 individual responses for the TPR estimates, and 477,393 responses for the FPR estimates.

https://doi.org/10.1038/s41562-025-02221-6

Extended Data Table 1 | Social network belief questionnaire

Question	Text	Response
1	Do you know [person a]? [photo of a shown with question 1]	Yes No I don't know / I refuse to answer*
2	Do you know [person b]? [photo of b shown with question 2]	Yes No I don't know / I refuse to answer*
3	Do [person a] and [person b] know each other? [photo of pair shown with question 3]	Yes No I don't know / I refuse to answer*
4	Do [person a] and [person b] spend free time together?	Yes No I don't know / I refuse to answer*
5	Do [person a] and [person b] trust each other to talk about something personal or private?	Yes No I don't know / I refuse to answer*
6	Are [person a] and [person b] one of the following?	Parent / child Sibling Partner I don't know / I refuse to answer*

Each survey respondent is asked about their beliefs about the existence of relationships between pairs of individuals drawn from their village social network.

Extended Data Table 2 | Contrasts for tie characteristics

	Variable	Contrast	Difference		
			TPR	FPR	J
1	Age (difference) 95% CI *P*	81.0 > 0.0	24.78 (22.8, 26.76) <0.001	2.77 (0.41, 5.13) 0.021	22.01 (19.0, 25.01) <0.001
2	Age (mean)	91.99 > 14.5	−26.71 (-29.26, −24.16) <0.001	−2.37 (-4.69, −0.06) 0.045	−24.34 (-27.67, −21.01) <0.001
3	Cognizer-to-tie distance	1.0 > No path	9.3 (6.8, 11.8) <0.001	8.35 (5.93, 10.77) <0.001	0.95 (-2.43, 4.34) 0.582
4		17.5 > 1.0	−46.12 (-49.25, −42.99) <0.001	−23.57 (-25.65, −21.5) <0.001	−22.55 (-26.19, −18.92) <0.001
5		17.5 > No path	−36.82 (-40.4, −33.25) <0.001	−15.22 (-17.27, −13.17) <0.001	−21.6 (-25.59, −17.61) <0.001
6	Degree (difference)	17.0 > 0.0	4.55 (1.58, 7.52) 0.003	5.29 (2.57, 8.01) <0.001	−0.74 (-4.67, 3.19) 0.712
7	Degree (mean)	15.99 > 1.0	−23.2 (-26.76, −19.65) <0.001	−16.1 (-18.38, −13.82) <0.001	−7.1 (-11.14, −3.07) 0.001
8	Education	Same > Mixed	0.02 (-2.0, 2.03) 0.988	0.77 (-1.43, 2.97) 0.493	−0.75 (-3.63, 2.13) 0.609
9	Gender	Mixed > Men	3.27 (1.05, 5.49) 0.004	−14.1 (-16.33, −11.86) <0.001	17.37 (14.33, 20.41) <0.001
10		Women > Men	6.21 (4.14, 8.27) <0.001	−3.41 (-5.86, −0.96) 0.006	9.61 (6.52, 12.7) <0.001
11		Women > Mixed	2.93 (0.86, 5.01) 0.006	10.69 (8.53, 12.85) <0.001	−7.76 (-10.65, −4.86) <0.001
12	Indigeneity	Mestizo > Indigenous	1.41 (-1.24, 4.07) 0.297	−0.15 (-2.55, 2.25) 0.905	1.56 (-1.91, 5.03) 0.378
13		Mixed > Indigenous	0.46 (-2.35, 3.26) 0.751	−1.1 (-3.52, 1.32) 0.374	1.55 (-2.04, 5.15) 0.397
14		Mixed > Mestizo	−0.96 (-3.13, 1.21) 0.387	−0.95 (-3.16, 1.26) 0.4	−0.01 (-3.02, 3.0) 0.996
15	Kin	Yes > No	28.29 (26.89, 29.69) <0.001	61.67 (60.02, 63.32) <0.001	−33.38 (-35.49, −31.27) <0.001
16	Relation	Personal Private > Free Time	−9.39 (-11.4, −7.38) <0.001	−5.77 (-7.96, −3.57) <0.001	−3.63 (-6.5, −0.75) 0.013
17	Religion	Same > Mixed	6.06 (3.94, 8.17) <0.001	7.24 (5.11, 9.38) <0.001	−1.19 (-4.08, 1.71) 0.423
18	Wealth (difference)	1.0 > 0.0	−33.47 (-36.47, −30.46) <0.001	−18.98 (-20.96, −16.99) <0.001	−14.49 (-17.96, −11.02) <0.001
19	Wealth (mean)	1.0 > 0.1	25.57 (22.94, 28.2) <0.001	17.65 (15.23, 20.07) <0.001	7.92 (4.45, 11.39) <0.001
20	Within-tie distance	19.0 > 2.0		−34.55 (-36.33, −32.77) <0.001	
21		19.0 > No path		−23.95 (-25.74, −22.15) <0.001	
22		2.0 > No path		10.6 (8.11, 13.08) <0.001	

Contrasts for tie characteristics. Each accuracy measure represents the difference between the predicted value for each level of the contrast. 95% confidence intervals are presented in parentheses.

 https://doi.org/10.1038/s41562-025-02221-6

Extended Data Table 3 | Contrasts for respondent and village characteristics

			Difference		
	Variable	Contrast	TPR	FPR	J
1	Age	30.65 > 14.0	5.26	2.74	2.51
	95% CI		(2.53, 7.98)	(0.31, 5.18)	(-1.02, 6.04)
	P		<0.001	0.027	0.163
2		96.0 > 14.0	−1.4	36.04	−37.45
			(-8.03, 5.23)	(29.68, 42.41)	(-46.19, −28.7)
			0.679	<0.001	<0.001
3		96.0 > 30.65	−6.66	33.3	−39.96
			(-12.99, −0.32)	(27.08, 39.52)	(-48.4, −31.51)
			0.039	<0.001	<0.001
4	Coffee cultivation	Yes > No	5.07	3.08	2.0
			(2.09, 8.06)	(-0.16, 6.32)	(-2.26, 6.25)
			0.001	0.063	0.358
5	Degree	18.0 > 1.0	5.55	−2.75	8.3
			(1.63, 9.46)	(-5.75, 0.25)	(3.47, 13.12)
			0.005	0.073	0.001
6	Education	Some > No	−2.42	−6.63	4.21
			(-4.8, −0.04)	(-9.42, −3.84)	(0.67, 7.76)
			0.047	<0.001	0.02
7		Yes > No	−5.83	−11.84	6.0
			(-8.53, −3.13)	(-14.7, −8.98)	(2.2, 9.81)
			<0.001	<0.001	0.002
8		Yes > Some	−3.42	−5.2	1.79
			(-5.87, −0.97)	(-7.61, −2.8)	(-1.52, 5.09)
			0.006	<0.001	0.289
9	Gender	Woman > Man	−0.78	−3.2	2.42
			(-2.97, 1.42)	(-5.6, −0.79)	(-0.72, 5.56)
			0.488	0.009	0.131
10	Indigeneity	Mestizo > Indigenous	−2.04	−0.42	−1.62
			(-4.76, 0.68)	(-3.23, 2.39)	(-5.41, 2.18)
			0.141	0.767	0.404
11	Religion	No Religion > Catholic	−2.06	−0.64	−1.42
			(-6.0, 1.87)	(-4.45, 3.17)	(-6.66, 3.81)
			0.304	0.741	0.594
12		Protestant > Catholic	−0.04	−0.68	0.64
			(-2.26, 2.17)	(-3.09, 1.72)	(-2.51, 3.79)
			0.971	0.578	0.69
13		Protestant > No Religion	2.02	−0.04	2.06
			(-1.93, 5.97)	(-3.85, 3.77)	(-3.19, 7.31)
			0.316	0.983	0.441
14	Wealth	1.0 > 0.07	−13.32	−20.26	6.94
			(-16.77, −9.87)	(-23.76, −16.76)	(2.22, 11.66)
			<0.001	<0.001	0.004

Each accuracy measure represents the difference between the predicted value for each level of the contrast. 95% confidence intervals are presented in parentheses.

Extended Data Table 4 | Accuracy on kinship ties

Response \ Status	None of the above	Parent/child	Partners	Siblings
None of the above	96.74	1.17	0.33	1.77
Parent/child	2.23	96.99	0.24	0.55
Partners	2.11	0.47	96.95	0.47
Siblings	3.15	0.76	0.13	95.97

Survey respondents are remarkably accurate in their knowledge of the kinship relations in their networks. Rows indicate the response to survey question 6, where respondents indicate the type of kin relationship (if any) that holds between a presented pair. Columns indicate the status in the underlying sociocentric (reference) network. Respondents make correct identifications are made around 96.66% of the time, on average across the categories.