Cognitive maps of social features enable flexible inference in social networks

Jae-Young Son[a], Apoorva Bhandari[a,1], and Oriel FeldmanHall[a,b,1,2]

[a]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912; and [b]Carney Institute for Brain Sciences, Brown University, Providence, RI 02912

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved August 1, 2021 (received for review October 16, 2020)

In order to navigate a complex web of relationships, an individual must learn and represent the connections between people in a social network. However, the sheer size and complexity of the social world makes it impossible to acquire firsthand knowledge of all relations within a network, suggesting that people must make inferences about unobserved relationships to fill in the gaps. Across three studies (n = 328), we show that people can encode information about social features (e.g., hobbies, clubs) and subsequently deploy this knowledge to infer the existence of unobserved friendships in the network. Using computational models, we test various feature-based mechanisms that could support such inferences. We find that people's ability to successfully generalize depends on two representational strategies: a simple but inflexible similarity heuristic that leverages homophily, and a complex but flexible cognitive map that encodes the statistical relationships between social features and friendships. Together, our studies reveal that people can build cognitive maps encoding arbitrary patterns of latent relations in many abstract feature spaces, allowing social networks to be represented in a flexible format. Moreover, these findings shed light on open questions across disciplines about how people learn and represent social networks and may have implications for generating more human-like link prediction in machine learning algorithms.

social networks | cognitive maps | learning | representation | generalization

Human social life unfolds within the landscape of large, complex social networks. Having a reliable representation of this social environment—knowing how people are linked to each other—aids successful navigation through dense relational webs (1, 2). Simple associative learning mechanisms can help acquire knowledge about relations within a network from observing social interactions (3, 4). However, given the sheer number of relationships in a typical social network (5) and the fact that humans routinely exhibit imperfect memory (6, 7), it is virtually impossible to acquire firsthand knowledge of all relations from observation alone. Therefore, to fill knowledge gaps, people must make informed inferences about unobserved relationships based on indirect clues (1, 8–10). What strategies might humans use to make flexible inferences about the structure of social networks?

One simple solution is to use a similarity heuristic based on homophily, colloquially expressed as "birds of a feather flock together" and formally defined as the tendency of people to disproportionately affiliate with those who share similar traits (11, 12). Homophily has long been established in the social sciences as a fundamental organizing principle for affiliation and group formation (11, 13). Leveraging this principle, friendships can be inferred by identifying individuals who share relevant features, including sociological characteristics such as race, ethnicity, age, religious belief, education level, occupation, and gender (11, 14, 15). In combination with associative learning, a simple similarity heuristic could provide a computationally inexpensive mechanism for representing who is friends with whom in social networks, especially in a world dominated by homophily (5). However, a similarity heuristic lacks flexibility. In situations in which homophily is not the dominant organizing principle (e.g., when "opposites attract"), using a similarity heuristic may lead to inaccurate representation and generalization.

A more sophisticated and flexible solution is to explicitly encode relational knowledge within a cognitive map (Fig. 1*A*), a representational format that binds knowledge about entities and their relations (16). Cognitive maps first arose as an explanation for how rodents learn, represent, and navigate spatial environments (17, 18) and have since been invoked to explain how humans perform a variety of tasks (spatial and otherwise) that require knowledge of abstract relations (16, 19–23), including social knowledge about community structure and social hierarchies (24–26). Indeed, humans seem to spontaneously track community members' social positions in large, real-world networks (27–29), further hinting that cognitive maps underlie the representation of social networks. Critically, cognitive maps provide the key affordance of being able to generalize knowledge beyond direct experience (17, 30–35), which makes them especially well suited for representing social networks (Fig. 1*B*).

What form might a social cognitive map take? One possibility is that individuals are represented as nodes in a cognitive graph (36), where the edges represent relations between individuals (37). This type of individual-based cognitive map can support relational inferences about known network members (38–40), but it cannot support inferences about a stranger's relations (40). We know, however, that even young children are able to make inferences about unknown others (41, 42), which suggests that cognitive maps might take a different form. An alternative possibility is that people build additional feature-based cognitive maps, where each node is a social feature associated with an individual (i.e., a hobby rather than a person), and edges are

Significance

How do people learn the large, complex web of social relations around them? We test how people use information about social features (such as being part of the same club or sharing hobbies) to fill in gaps in their knowledge of friendships and to make inferences about unobserved friendships in the social network. We find the ability to infer friendships depends on a simple but inflexible heuristic that infers friendship when two people share the same features, and a more complex but flexible cognitive map that encodes relationships between features rather than between people. Our results reveal that cognitive maps play a powerful role in shaping how people represent and reason about relationships in a social network.

Author contributions: J.-Y.S., A.B., and O.F.H. designed research; J.-Y.S. performed research; J.-Y.S., A.B., and O.F.H. analyzed data; and J.-Y.S., A.B., and O.F.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]A.B. and O.F.H. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: oriel.feldmanhall@brown.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021699118/-/DCSupplemental.

Published September 13, 2021.

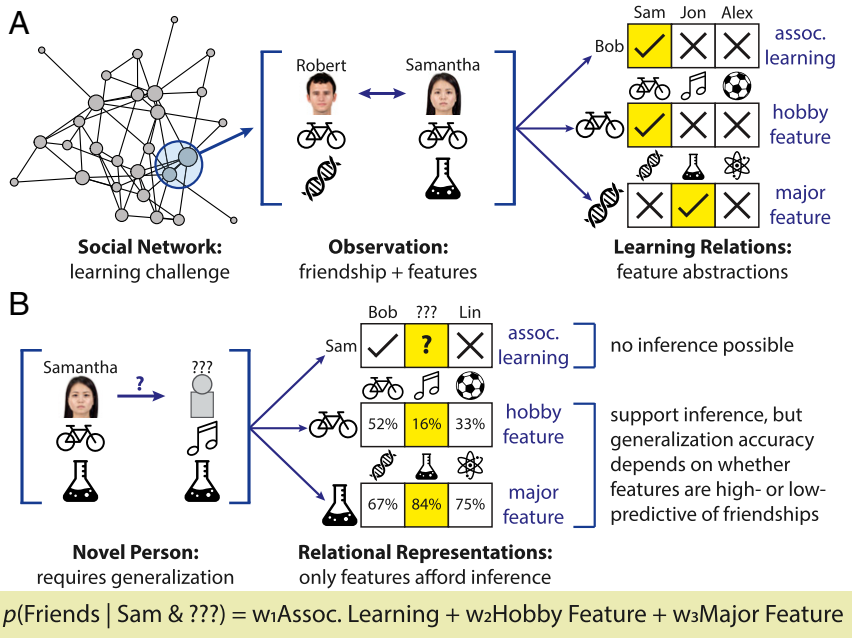

Fig. 1. Conceptual model of feature-based cognitive maps. (*A*) Social networks contain a large number of potential relations, making it difficult to learn about all relationships. One strategy is to use associative learning to encode friendships. Individuals could also be abstracted into social features (e.g., race, gender, personal interests) such that latent feature-to-feature relations are learned. In this example, Robert and Samantha's friendship can be abstracted into the latent relations "biker-to-biker" and "biologist-to-chemist." (*B*) Each learning strategy comes with different affordances for generalization. Associative learning does not support inference about unknown others or unobserved relations. Feature-based cognitive maps support the ability to make flexible generalizations as long as the stranger's features are known. In this example, knowledge of the "hobby" and "college major" feature maps can be used to infer the probability that Samantha and the stranger are friends. However, flexible use of feature-based cognitive maps does not guarantee inferences are accurate. If a particular feature map is reliably more predictive of friendship, placing a higher weight on that map leads to more accurate generalization.

relationships between features (Fig. 1*A*). Critically, this kind of cognitive map would allow knowledge about the statistical dependencies between social features and friendships to be generalized (43, 44), enabling inferences about a stranger's friendships without the need for direct observation of social interactions (Fig. 1*B*). Feature-based cognitive maps can be built using biologically plausible and computationally cheap learning rules (40, 45–47) and naturally capture latent statistical structure between social features and friendships (48, 49). In short, a feature-based cognitive map can predict friendship between any two individuals as long as their social features are known.

Feature-based cognitive maps can support flexible generalization in two ways. First, they can encode the specific statistical dependencies between social features and friendships present in the social environment. For instance, they can encode all of the same relational knowledge as a similarity heuristic when homophily is the dominant organizing principle in the social network. At the same time, they can encode other forms of structure; consider a feature map where individuals are abstracted into their college majors, and edges reflect major-to-major relations (Fig. 1*A*). Such a map can be used to glean that "biology majors tend to be friends" and that "physics majors find each other intolerable and prefer to be friends with chemists" (Fig. 1*B*). Unlike a similarity heuristic, feature-based cognitive maps can encode arbitrary relationships between social features and friendships, which enables flexible prediction based on the specific statistics of the social environment. Second, if a person builds multiple feature maps (Fig. 1*A*), they can then flexibly draw upon different feature maps to guide generalization (Fig. 1*B*). For example, a person could learn that majors are highly predictive of friendships in a college network, while hobbies are far more predictive of friendship in a company network. Although feature maps do not guarantee accurate inference (e.g., this person may mistakenly believe that hobbies are better predictors of friendship than majors in the college network), flexibility is a prerequisite for fine-tuning relational predictions depending on the social environment or contextual goals.

Across three studies, we examine how humans represent the myriad of social relationships that comprise a larger social network, interrogate the learning mechanisms used for social inference, and test whether those mechanisms afford flexible generalization. We test a number of strategies—associative learning, similarity heuristics, and feature-based cognitive maps—that might be leveraged to encode and infer relationships in a social network, either alone or in combination. In all studies, subjects were tasked with learning novel social networks in which different social features (e.g., affiliation with an extracurricular club or sharing similar interests) predicted friendship with varying accuracy. We then estimated the degree to which knowledge of social features influenced the representation of the network's configuration (i.e., the existence of actual friendships). In Study 1, we demonstrate that people readily use features to represent social networks and selectively rely on features that are highly predictive of friendship. In Studies 2 and 3, we test which learning mechanisms and representations enable subjects to make flexible inferences about never-before-seen friendships in the social network.

Results

Study 1: Social Network Representations Are Shaped by Predictive Social Features. Subjects (n = 50) learned about three novel networks in separate phases, each containing 11 network members and 15 friendships. In each phase, subjects associatively learned pairwise friendships and then reported their representation of that network's configuration. In phase 1, we measured network representations when subjects were only able to learn friendships associatively. In phases 2 and 3, subjects were additionally provided with network members' social features, which probabilistically predicted friendships. We operationalized social

features as club affiliation, such that belonging to the same club was highly predictive of friendship in one phase and less predictive in the other. Subjects were explicitly instructed that this was the case, and the order was counterbalanced across subjects. This design allowed us to test three questions: 1) whether subjects' social network representations were shaped by knowledge of social features, 2) whether such information was selectively used when subjects believed it was useful for predicting friendships, and 3) whether subjects' use of associatively learned friendship information decreased once information about social features became available.

We measured subjects' representation of the social network using two measures. First, subjects completed a spatial arrangement task (50), arranging network members within a circle such that greater spatial distance corresponded to greater social distance (Fig. 2*A*). We computed response matrices by calculating pairwise Euclidean distances between all network members so that 1 = maximum similarity (i.e., friendship). Second, subjects completed a more traditional memory task, recalling which network members were friends with each other (Fig. 2*B*), including their confidence on a four-point scale (Very Unsure to Very Sure).

The measures from these two tasks were analyzed using a behavioral variant of representational similarity analysis to examine the mapping between the network's actual configuration and subjects' representations of it. Predictor matrices encoded each piece of information that could be incorporated into a social network representation: associatively learned friendships as well as the club, race, and gender features (Fig. 2*C*). Therefore, if a subject had exclusively encoded the friendships through associative learning, their response matrix would be identical to the friendship predictor matrix, which simply contained the network's true configuration. The influence of social features could then be captured when subjects' representations deviated from the network's true configuration in a manner encoded by the predictor matrices. To estimate how much each feature explained subjects' representations, we used linear regression to predict each response matrix as a weighted sum of all predictor matrices (Fig. 2*C*). We estimated each subject's regression estimates for each task separately and then averaged them into a single composite estimate for robustness (see *Methods* for details and *SI Appendix* for nonaggregated results).

The results reveal that subjects' representations of the social network were significantly shaped by network members sharing social features (i.e., club affiliation) when instructed that the feature was highly predictive of friendships but not when told that the feature was low predictive (Fig. 3*A*). In other words, subjects flexibly incorporated social features into their representation of the network only when they were predictive of friendship. These results are unlikely to be caused by selective learning about specific features, as subjects accurately remembered network members' features in all phases (93% when the feature was predictive and 91% when not). The degree to which associative learning predicted responses did not significantly depend on whether the feature was predictive, nor did it differ from phase 1, in which no information about features was available (Fig. 3*A*). Other observable features (i.e., race and gender) did not influence subjects' representations of the network (Fig. 3*A*), possibly because subjects recognized that they were not predictive in our task.

Study 2: Social Features Enable Flexible Inference of Unobserved Friendships in a Social Network. The results from Study 1 show that subjects' representations of the network deviated from the true, observed configuration (consistent with imperfect associative learning or retrieval) but in a manner that is consistent with using predictive social features. This suggests that people supplement knowledge of directly observed friendships with an inferential process—such as a similarity heuristic or a cognitive map—that relies on predictive social features to make inferences about unobserved friendships. Therefore, we tested in a second study whether predictive features are spontaneously (i.e., in the absence of explicit instruction) detected and flexibly generalized (i.e., inferring unobserved friendships). The experimental task was similar to the one used in Study 1, with a few differences (see *Methods*). Subjects (n = 84) learned about a social network consisting of 12 people and 14 friendships and simultaneously learned about friendship and network members' social features (operationalized as hobbies and college majors, counterbalanced to be high or low predictive of friendship). Unlike Study 1, subjects were not instructed that either social feature was probabilistically predictive of friendships (51). Given past studies showing that there are large individual differences in people's ability to represent (spatial) cognitive maps (52) and that there is a cognitive cost of building structured representations (53–55), we examined how representation and generalization are affected by an individual's ability to accurately learn which features are associated with which network members.

Representation. Replicating the results of Study 1, social network representations in Study 2 reliably reflected subjects' use of both associative learning and social features (Fig. 3*B*). At the group level, and diverging from Study 1, the results indicate that subjects made significant use of all available features (i.e., high- and low-predictive features as well as demographic features like race and gender). We tested whether the use of high- versus low-predictive features depends on the ability to accurately remember social features, using mixed-effects linear regression with random intercepts. Subjects demonstrated divergent use of high- and low-predictive features depending on their feature memory accuracy (Fig. 3*C*; interaction $\beta = 0.21$, $SE = 0.07$, $t = 2.96$, $P = 0.004$). Those with greater feature memory accuracy placed a greater weight on high-predictive features compared with subjects with less accurate memory (Fig. 3*C*, purple line; $\beta = 0.30$, $SE = 0.05$, $t = 5.61$, $P < 0.001$). In contrast, the same weight was placed on low-predictive features regardless of an individual's feature memory accuracy (Fig. 3*C*, orange line; $\beta = 0.09$, $SE = 0.05$, $t = 1.68$, $P = 0.095$). Put simply, the more accurately subjects were able to remember features, the more their representation resembled those from Study 1, in which high-predictive features were explicitly instructed.

Generalization. These results demonstrate that social features can bias representations away from the true configuration of the social network. Given this cost, why do subjects use social features to represent the network? One possibility is that social features unlock the ability to infer friendships that are not remembered and to infer friendships that were never observed. To test this, we administered a generalization task requiring subjects to infer which network members were most likely to become friends with new transfer students joining the network (see *Methods*). Generalization could not rely on direct experience, as the transfer students had never been encountered. Instead, this task allowed us to test whether subjects generalized knowledge about the predictive relationship between friendships and social features, and actively inferred unobserved friendships based on shared social features. In our analyses, we controlled for potential confounding effects of social network position (i.e., network centrality), as subjects could simply have used a heuristic that, for example, popular network members are more likely to become friends with new students.

The results revealed that subjects were significantly more likely to infer friendship when the transfer student and known network member shared a feature compared to when they shared no features ($\beta = 0.14$, $SE = 0.04$, $z = 3.08$, $P = 0.002$). This was especially pronounced in those with better feature memory accuracy (Fig. 4*A*; interaction $\beta = -0.99$, $SE = 0.05$, $z = -19.06$, $P < 0.001$). Although these results demonstrate that people use feature knowledge to guide inference about unknown friendships in the network, subjects did not preferentially infer friendships when high- versus low-predictive features were shared, nor did this inference depend on feature memory accuracy (all Ps > 0.500; *SI Appendix*, Table S1).

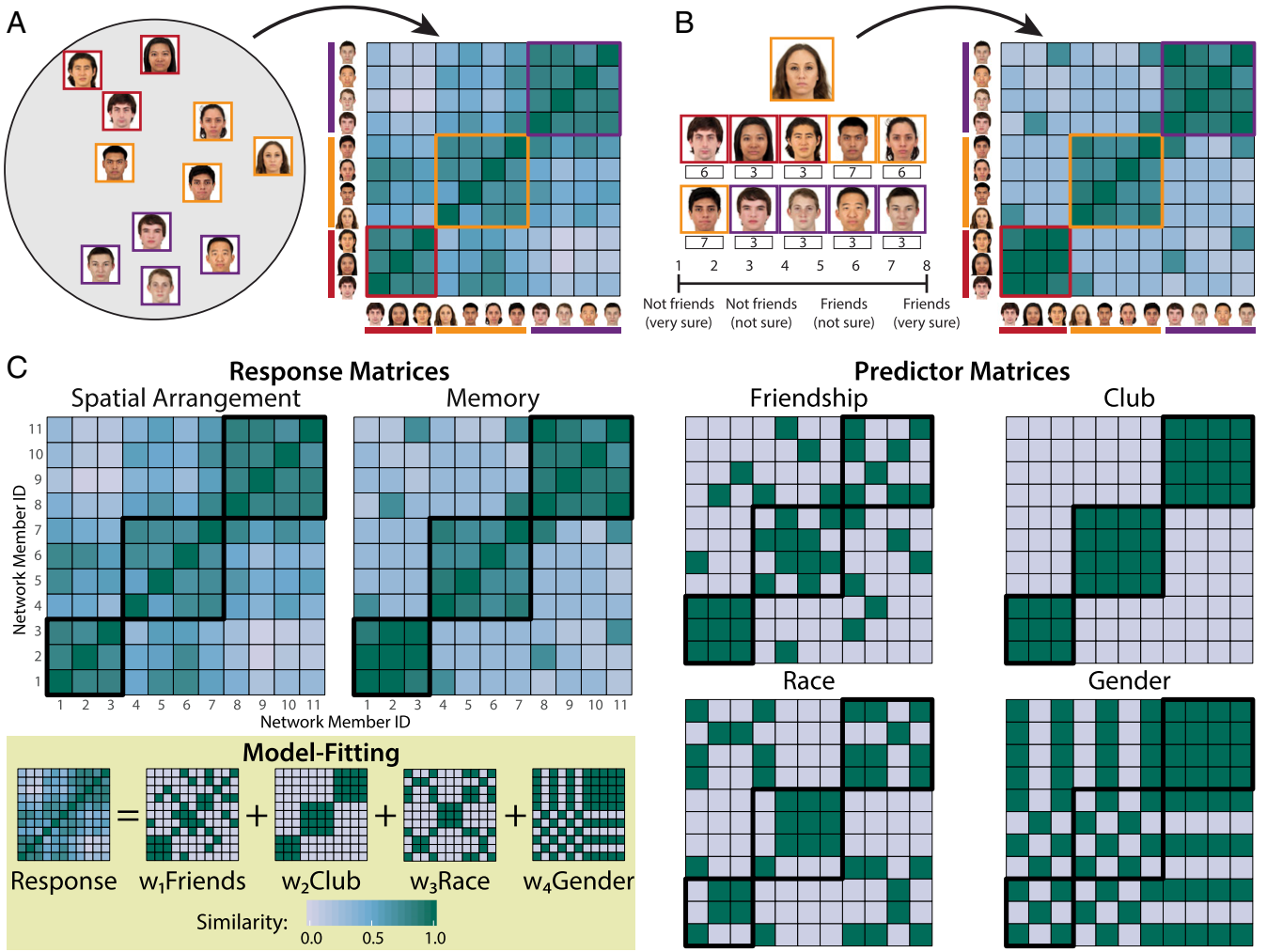

Fig. 2. Representational Similarity Analysis methods. (*A*) Subjects were asked to spatially arrange network members according to how socially close they were perceived to be, such that shorter distances reflected closer friendship. A representative subject's arrangement is displayed here. Using subjects' arrangements, we created a representational similarity matrix of responses using pairwise Euclidean distances. Network members are outlined in colors corresponding to what clubs they belonged to in this study session. (*B*) We also created a memory response matrix by weighting recall with confidence ratings. (*C*) We created predictor matrices reflecting how a subject would have responded if only a single type of information shaped their representation of the network. We then estimated how much individual subjects used each type of information. A representative subject's predictor matrices are shown for illustration; subjects were shown different networks containing session-specific feature mappings.

At first blush, the lack of a preference for the high-predictive feature suggests that subjects' ability to predict friendship in unknown parts of the network reflects the use of an inflexible similarity heuristic that does not rely on learning the predictive statistical relationship between features and friendships. If this were the case, the feature-based bias that we observe during representation should have no bearing on generalization. On the other hand, it is possible that these results reflect the combination of a similarity heuristic and the use of feature-based cognitive maps. In that case, representations shaped by high-predictive features should lead to greater inference of friendship between individuals who share a high-predictive feature. In other words, evidence that the content of an individual's representation guides their generalization decisions would point to the use of a cognitive map in addition to a similarity heuristic.

Representation informs generalization. To test whether representation guides generalization, we examined whether subjects' use of features in the representation tasks predicts their use of features during generalization. We observed that when high-predictive social features played a stronger role in shaping the network representation, subjects were more likely to infer the existence of a friendship between a transfer student and network member who shared a high-predictive feature (Fig. 4*B*; $\beta = 4.02$, $SE = 0.99$, $z = 4.05$, $P < 0.001$; reference *SI Appendix*, Table S2 for full results). This result reveals that subjects' representations of the social network are shaped by social features, which subsequently guide inferences during generalization. To the extent that subjects are able to accurately remember what features are associated with which network members, they are then able to flexibly draw upon feature knowledge to place greater weight on high- over low-predictive features. Taken together, this suggests subjects' behavior in the generalization task cannot be explained by a similarity heuristic alone. Rather, people additionally draw upon feature-based cognitive maps to infer friendships.

Mechanisms supporting social inference. While these results hint at the joint influence of a simple but inflexible similarity heuristic (i.e., a prior belief in homophily as an inductive bias) as well as a flexible cognitive map (i.e., representing the statistical structure of latent relations in an abstract feature space), they do not formally disentangle these accounts. These two uses of social features are not mutually exclusive, and it is also possible that subjects rely on a combination of a similarity heuristic as an inductive bias and

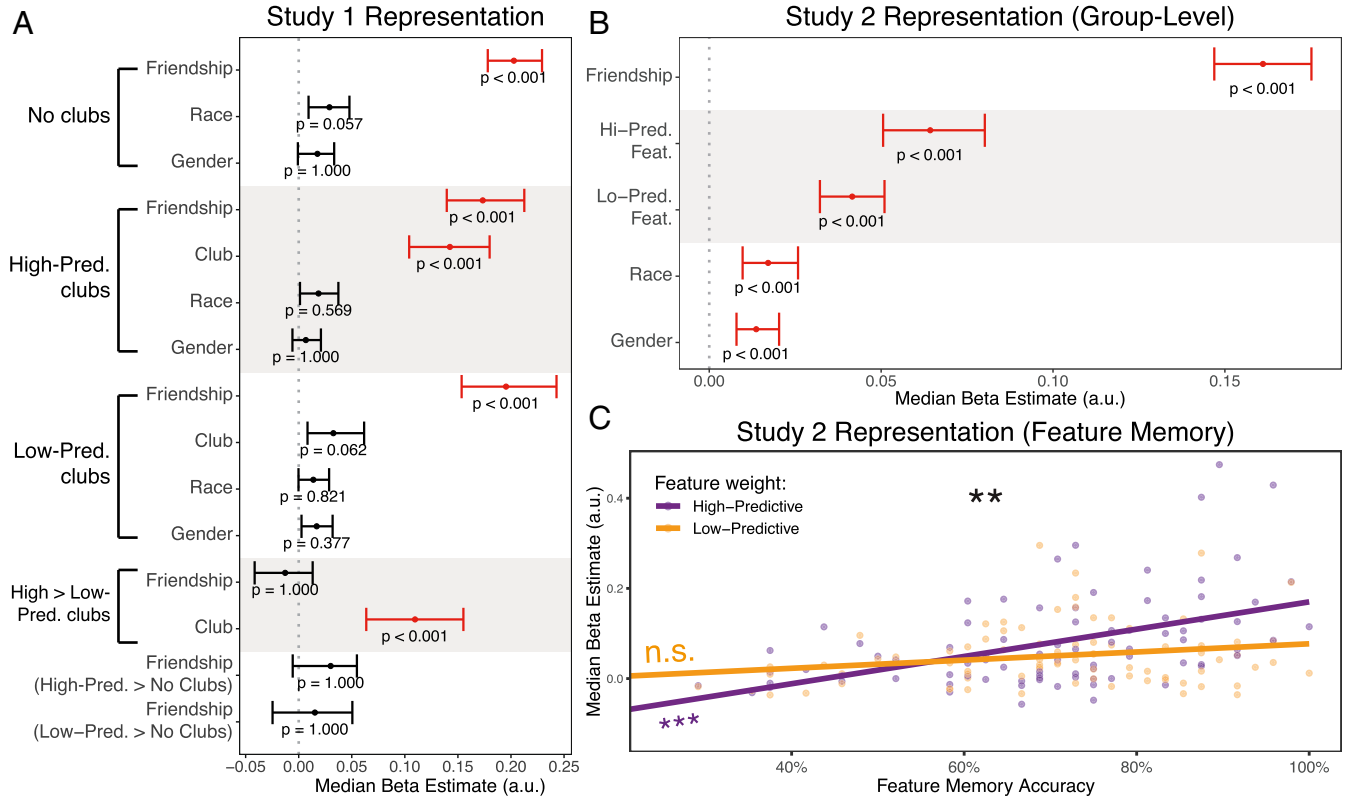

Fig. 3. Representation results from Studies 1 and 2. In all plots, datapoints reflect median β estimates, and error bars reflect uncorrected nonparametric 95% CIs. (*A*) Study 1 forest plot of all contrasts tested. All *P* values are Bonferroni corrected for 15 comparisons. (*B*) Study 2 forest plot of all tests against 0. All *P* values are Bonferroni corrected for five comparisons. (*C*) In Study 2, subjects' use of high- versus low-predictive features depended on how accurately they could remember which features were associated with which network members. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, n.s., non-significant.

feature-based maps, according to how well they predict friendships. To test these possibilities, we built a number of psychologically plausible computational models pitting these mechanisms against one another. In total, we tested five models varying in their level of sophistication and flexibility. Our aim was to identify which mechanism (or combination of mechanisms) best explains subjects' behavior on the memory and generalization tasks.

Feature-based cognitive maps represent network members in abstract feature spaces such that a particular individual is represented as a feature rather than a distinct entity. For example, the network member Robert might be represented as the feature "biking" in the hobby map and as "biology" in the college major map (Fig. 1*A*). Therefore, Robert's friendship with Samantha the biking chemist would be represented as a biker-to-biker relation in the hobby map and as a biologist-to-chemist relation in the majors map (Fig. 1*A*). Using a simple updating rule (38, 40, 44, 45), our computational model learned an approximation of $p(\text{friendship} \mid \text{features})$ for each feature map (Fig. 5*A* and *SI Appendix*, Fig. S7 and see *Methods*).

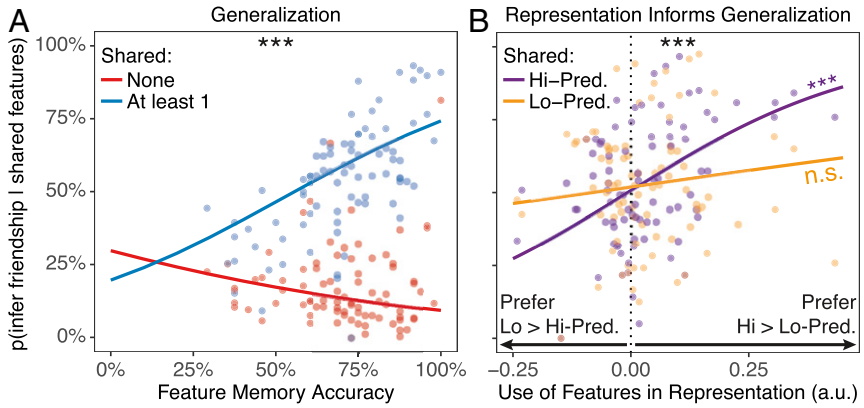

Fig. 4. Generalization results from Study 2. (*A*) Subjects with greater memory accuracy for network members' features were less likely to infer a friendship when no features were shared between the transfer student and known network members, and more likely to infer friendship when at least one feature was shared. (*B*) The more subjects relied on high-predictive features to represent the social network, the more they inferred a greater likelihood of friendship when the transfer student and network member shared a predictive social feature. The significance asterisk reflects the plotted interaction effect. Each datapoint represents one subject's average likelihood of inferring friendship. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, n.s., non-significant.

Each feature map therefore predicts the likelihood of friendship between two network members given their latent relation in abstract feature space.

We found that models lacking a mechanism for building feature-based cognitive maps are a poor fit for subjects' data, suggesting that cognitive maps play a critical role in inference (Fig. 5 *B*, *Left*). The best-fitting model relied on a combination of associative learning ($M = 0.13$, t (81) = 8.44, 95% CI = [0.10, 0.17], $P < 0.001$), a similarity heuristic (high-predictive $M = 0.26$, t (81) = 12.97, 95% CI = [0.22, 0.30], $P < 0.001$; low-predictive $M = 0.25$, t (81) = 18.19, 95% CI = [0.22, 0.28], $P < 0.001$), and feature-based cognitive maps (high-predictive $M = 0.11$, t (81) = 2.22, 95% CI = [0.01, 0.22], $P = 0.029$; low-predictive $M = 0.24$, t (81) = 5.46, 95% CI = [0.15, 0.33], $P < 0.001$). Therefore, neither a similarity heuristic nor a feature-based cognitive map is sufficient on its own for explaining network representations or inferences during generalization. Instead, we find there was variability in subjects' use of each strategy. Some individuals relied more heavily on a similarity heuristic, while others leveraged a cognitive map (Fig. 6*A*). Given the difficulty of building and representing cognitive structures (53–55), we tested whether these individual differences can be explained by subjects' ability to accurately remember what features were associated with which network members. While feature memory accuracy did not influence associative learning (Fig. 6 *B*, *Left*; $\beta = -0.06 \pm 0.07$, $t = -0.84$, $P = 0.400$) or how much subjects preferred a similarity heuristic for high- over low-predictive features (Fig. 6 *B*, *Middle*; $\beta = 0.00 \pm 0.11$, $t = 0.02$, $P = 0.988$), it did have a significant effect on how features were used for cognitive maps. Subjects with better feature memory accuracy had a significantly stronger preference for using high- over low-predictive features in their cognitive maps (Fig. 6 *B*, *Right*; $\beta = 1.50 \pm 0.37$, $t = 4.02$, $P < 0.001$).

Study 3: Feature-Based Cognitive Maps Act as the Dominant Mechanism for Learning Social Networks. In Study 2, the predictions made by a similarity heuristic are highly correlated with the predictions made by feature-based cognitive maps, which may be obfuscating a reliance on feature-based cognitive maps. We tested the possibility that behavior appearing consistent with a similarity heuristic may actually be the result of using feature-based cognitive maps in a third study ($n = 194$). In this network (8 nodes and 14 edges), a similarity heuristic and cognitive map make diverging predictions about friendship (see *Methods*), such that high-predictive features perfectly determine whether two network members are friends (e.g., biologists and chemists were always friends) and does not reliably depend on homophily (e.g., physicists were never friends with each other). While homophily is mildly predictive of friendship for low-predictive features, this set of features is not generally informative of the friendships in the network. Therefore, to the extent that subjects continue to rely on an inflexible similarity heuristic despite its lack of predictive power, it would provide evidence for a similarity heuristic as a dominant inductive prior. If, however, subjects demonstrate flexible use of features to infer friendship, this would provide strong evidence that people build and use cognitive maps of social features. We tested three computational models which implemented learning mechanisms for a similarity heuristic and/or cognitive maps and found that the majority of subjects were best described by the model that included mechanisms for both the similarity heuristic and cognitive maps (Fig. 5 *B*, *Right*; see *Methods* for details of model selection).

Results from the model indicate that subjects were sensitive to the fact that homophily had ceased to be the organizing principle underlying friendships in this study (Fig. 6*A*): subjects no longer placed significant positive weight on a similarity heuristic (high-predictive $M = -0.04$, $t(191) = -2.05$, CI = [−0.08, −0.00], $P = 0.042$; low-predictive $M = 0.01$, $t(191) = 0.58$, CI = [−0.03, 0.06], $P = 0.570$) but did continue to make significant use of cognitive maps (high-predictive $M = 0.52$, $t(191) = 13.00$, 95% CI = [0.44, 0.60], $P < 0.001$; low-predictive $M = 0.51$, $t(191) = 11.71$, 95% CI = [0.42, 0.59], $P < 0.001$). Moreover, friendship memory accuracy did not affect subjects' preference for high- over low-predictive features when using a similarity heuristic (Fig. 6 *C*, *Left*; $\beta = -0.16 \pm 0.15$, $t = -1.11$, $P = 0.270$) but did when relying on a cognitive map (Fig. 6 *C*, *Right*; $\beta = 1.15 \pm 0.45$, $t = 2.55$, $P = 0.012$). In other words, the more accurately subjects remember friendships/features in the observed network, the more weight they place on cognitive maps built from high- rather than low-predictive features. Together, these results clearly demonstrate that people can build and use feature maps to

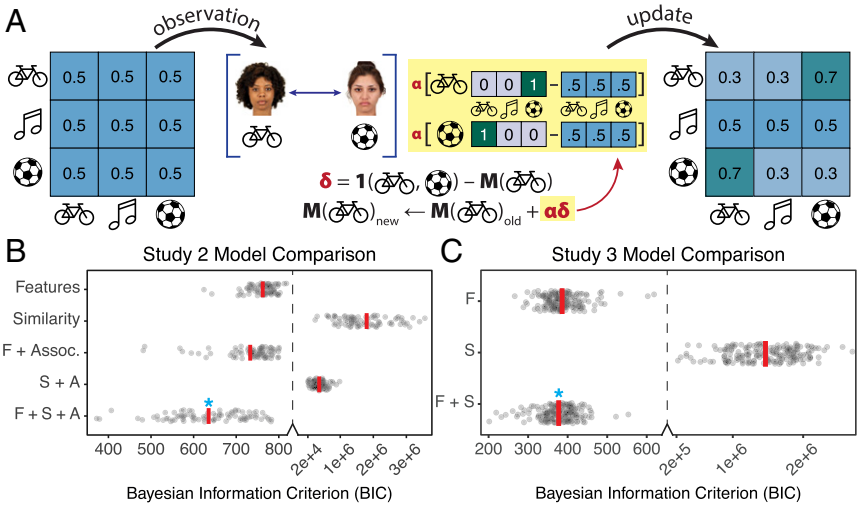

Fig. 5. Computational framework for Studies 2 and 3. (*A*) The feature matrix **M** represents the probability of two individuals being friends given their features and is learned using a simple update rule. Upon observing a friendship, a one-hot vector (denoted by **1**) encodes the relation between the features. The resulting prediction error δ is used to update **M**, tempered by the learning rate α. Since friendships are mutual in our study, updates occur symmetrically. The indexing and updating of **M** is row wise but is depicted here as a single-input function for simplicity. (*B*) We estimated computational models that draw upon feature-based cognitive maps, a similarity heuristic, and/or associative learning. The best-fitting model in Study 2 uses all three and is indicated by the blue asterisk. (*C*) The best-fitting model in Study 3 uses feature-based cognitive maps and a similarity heuristic.

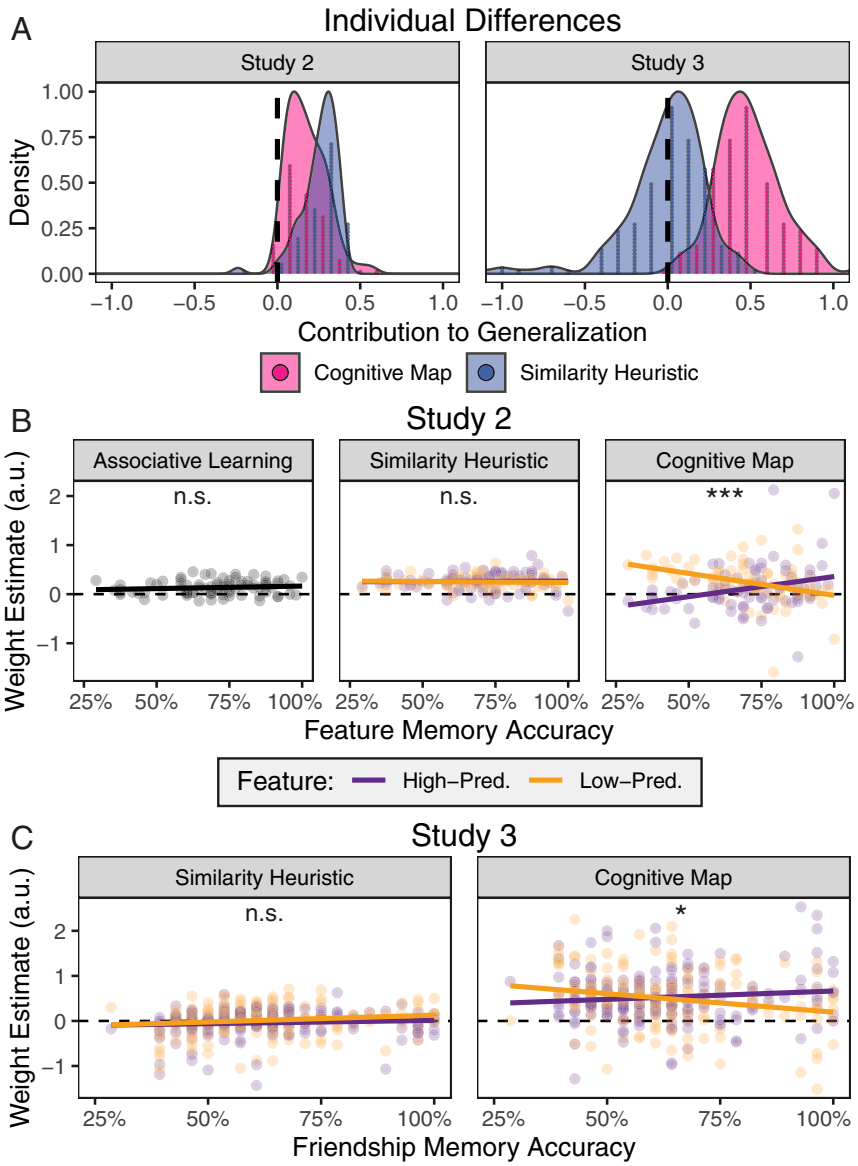

Fig. 6. Computational modeling results from Studies 2 and 3. (*A*) Subjects demonstrated variability in their use of a similarity heuristic and cognitive maps. Each dot represents one subject's parameter weight, which reflects how much a particular strategy contributed to generalization. (*B*) Although the use of associative learning and similarity heuristics does not vary according to feature memory accuracy in Study 2, those with better feature memory accuracy relied on high- (compared to low) predictive features. Significance tests reflect changing preference for the high- over low-predictive feature as a function of feature memory accuracy. (*C*) In Study 3, unlike a similarity heuristic, subjects with better friendship memory place more weight on high- than low-predictive features.

flexibly infer unobserved friendships in social networks, above and beyond the use of a similarity heuristic based on homophily.

Discussion

Due to the vastness of social networks, it is all but impossible to gain firsthand knowledge of all the relationships in a network, raising the question of how people build reliable representations of their complex social world. Here, we show that people supplement direct experience with relational inference based on social features (e.g., clubs or hobbies). We find that social features shape how people represent friendships within a social network, which in turn affects how people make inferences and generalize about unobserved relations. When informed that features are highly predictive of friendship, people adjust how much they incorporate a particular feature into their representation of the social network (Study 1). Even when this information is not explicitly provided, some people can spontaneously discover which features are predictive of friendship and use these features to shape their representation of the network and make predictions about other, unobserved friendships in the network (Study 2). What cognitive mechanisms enable such flexible generalization? We find that people leverage a similarity heuristic based on homophily and feature-based cognitive maps (Study 2), and cognitive maps become especially useful when homophily does not predict friendships (Study 3).

Given that homophily is a fundamental organizing principle for friendship and social networks writ large (5, 11), it is reasonable to expect that people will exploit a simple similarity heuristic to predict friendship, as it provides an easy solution to a challenging problem. Our data suggests that many people do at times deploy such a heuristic. However, once homophily is no longer an underlying generative principle (in a network which

bears little resemblance to the real world), most people prefer to build flexible cognitive maps in lieu of a similarity heuristic. In short, flexible generalization relies on a combination of a computationally inexpensive heuristic and a more complex strategy of building cognitive maps that encode the latent statistical structure underlying friendships in an abstract feature space.

While we operationalized similarity-based inference as an inflexible heuristic and contrasted it with more flexible cognitive maps, our computational models' implementation of feature-based cognitive maps can encode a similarity heuristic when homophily is the dominant generative principle of friendship. That is, if the observed statistics of friendships are consistent with homophily, as they were in Study 2, feature-based cognitive maps will in fact learn to represent a similarity principle. It is therefore possible that what appeared to be a reliance on a similarity heuristic in Study 2 may actually reflect the use of cognitive maps. Indeed, in Study 3, the widespread preference for feature-based maps over a similarity heuristic suggests this might be the case and hints at the possibility that feature-based cognitive maps could act as a unifying representational mechanism.

In our data, individual differences in cognitive map use were associated with the ability to remember the features linked to network members, suggesting a privileged role for declarative memory in map building. This would be consistent with theories in which declarative memories modify the state space over which relational learning operates (56, 57) and therefore how people make generalization decisions (58, 59). It is likely that the hippocampal formation plays a key role in the interplay between these learning mechanisms, as it contributes to both declarative and relational memory encoding (60–62), underlies predictive relational representations such as the successor representation and successor features (40, 47, 63), and provides episodic context for guiding generalization decisions (64–66).

To date, little work has explicitly studied how humans represent cognitive maps of friendships within a social network, despite the prevalence of graph-centric representations in social network analysis (27–29, 67) and recent interest in characterizing the "cognitive graph" in spatial navigation (36). Inspired by research on how spatial maps are encoded in the human hippocampal–entorhinal system (19), much of the work on abstract cognitive maps has studied how people infer relations in a two-dimensional social space (21). The hippocampal–entorhinal system seems to encode people's positions along two-dimensional social hierarchies within a Euclidian space (25, 26), which is especially important when making inferences integrating over both dimensions (26). In contrast to this more traditional view of the cognitive map operating in two-dimensional Euclidean space, our results demonstrate that people can also build cognitive maps encoding arbitrary patterns of latent relations in many abstract feature spaces, which allows social networks to be represented in a more flexible format. This is consistent with some recent models of hippocampal–entorhinal cognitive maps in which abstract maps encode structural features of the transition structure of experience (37).

Prior research in network science suggests people can make inferences about unknown social relations using schemas or heuristics that bias perception (1, 10). For example, network scientists have noted that people routinely make systematic errors when trying to remember social network ties, including categorizing people by group membership (13, 68–70), perceiving networks as being much more densely interconnected than they actually are (71), and inferring triadic closure [e.g., the belief that if Abby and Beth are friends, and Beth and Cathy are friends, then it is likely that Abby and Cathy are also friends (72–74)]. These errors have yet to be construed as an inferential mechanism that jointly contributes to representation and generalization. Thus, we extend this work by demonstrating that people can use social features to build cognitive maps, which can then be drawn upon to support inference of social relations. In this case, inferences about unknown parts of the social network are made not by tracking observable relations between specific people but by tracking latent relations between abstract features.

Our work also has implications for how link prediction is achieved in the field of computer science, which uses machine learning to predict unknown relationships within social networks (75, 76). The goal of link prediction is not to build explanatory models of how the human mind might infer relations but to generate descriptive models that maximize prediction accuracy. Despite this difference in objective, many of the approaches used in this literature share commonalities with the cognitive strategies examined here. For example, some classes of machine learning models implement a similarity heuristic, identifying social features in which homophily is predictive of relations (75, 76). Others resemble the cognitive maps used in our work, probabilistically inferring the existence of a relationship between individuals by learning relations between features (75, 76). Empirical work demonstrates that social features can be a powerful tool for link prediction (77), raising the possibility that detailing the mechanisms humans use to infer unknown social relations can offer new algorithms to improve link prediction in machine learning.

Our work therefore connects the problem of link prediction in a social network to a rich, multidisciplinary literature on relational inference with cognitive maps in humans and animals (40). Cognitive maps are well suited for representing a variety of relationships that comprise a social network, and recent theoretical work has led to biologically plausible computational models that can be deployed to study such representations (16, 37–39, 44–46, 63, 78, 79). We focused on a computational model that was inspired by the successor features framework in reinforcement learning. Related ideas, such as successor representations, could similarly be applied to the study of topology-based social link prediction, raising the tantalizing possibility that similarity and other previously identified heuristics (e.g., triadic closure, social groups, etc.) are predictive rules that emerge from a fundamental set of flexible learning mechanisms used for building cognitive maps. Taken together, these results lay groundwork for understanding how people mentally represent and navigate social networks, while also paving the way for examining the neural basis of feature-based social network representation and social link prediction.

Methods

Subjects. In Study 1, 50 subjects (36 female; 26 non-White or mixed race; mean age = 22.10, SD ± 6.98) were recruited from Brown University and the surrounding community. Subjects were paid $15 or received partial course credit for participating in 1.5-h study sessions. In addition, subjects could earn up to a $3 bonus depending on how accurately they were able to remember information from the tasks. The sample size was based on a preregistered power analysis comparing subjects' use of associative versus feature-based learning using a paired Student's t test, which estimated that we would need data from 41 subjects to achieve 80% power with an assumed effect size of Cohen's $d = 0.4$. In Study 2, 84 subjects (58 female or gender nonbinary; 51 non-White or mixed race; mean age = 20.1, SD ± 2.81) were recruited in an identical manner. The doubled sample size was based on past research on generalizing newly learned structures, which has found that about half of subjects seem unable to make such generalizations (53, 55), possibly because of its cognitive costs (54). In Study 3, 200 subjects were recruited online using the Prolific study pool. Due to technical errors, six subjects' data were lost, and the final sample size was therefore $n = 194$ (94 female, 3 gender unknown; mean age = 33.8, SD ± 10.8). Subjects were paid $6.50 for participating in a 1-h study session and could earn up to a $3 bonus depending on their performance in the study. All studies were approved by Brown University's Institutional Review Board, and informed consent was obtained from each subject.

Study 1: Social Network Representations Are Shaped by Predictive Social Features. Networks were generated using tools from the igraph package in R (80). Subjects were told the cover story that networks were derived from undergraduates' real friendships. Network members were randomly assigned a photograph of a face from the Chicago Face Database (81) with

two constraints: that the distribution of stimuli match Brown University's racial demographics (Asian, 20%; Black, 10%; Latinx, 16%; and White, 54%) and that they be evenly split between male and female faces. In phases 2 and 3 of the study, clubs were randomly assigned a label from a list of 10 real extracurricular organizations at Brown University. To avoid suspicion, subjects were told that all of the photographs and club labels had been deidentified using stock photos and a shuffled list of clubs to protect network members' privacy. Except in phase 1, network members were associated with one of three extracurricular clubs, and different clubs were used in each phase to avoid ordering or familiarity confounds.

We took the following steps to rule out potential alternative explanations. First, since subjects might use social features (i.e., information about clubs) simply because it is easier than memorizing individual friendships, we provided subjects with a monetary bonus for accurately remembering friendships but not club affiliation. Accordingly, by providing extrinsic motivation to learn and use observable friendship information, we conducted a strong test of whether subjects learn and use social features even when it is not monetarily rewarding to do so. Second, to avoid biasing responses, we did not explicitly instruct subjects to incorporate features during learning. Indeed, they were told that club affiliation was only one of many factors that potentially contribute to friendship. Third, in our analyses, we accounted for other social features that subjects may have built from network members' physical appearances (i.e., race and gender), which are often predictive of real-world friendships (11), even though they were not predictive of friendships in our study.

In the learning task, two network members were presented side by side on the same trial to indicate friendship. These pairs were displayed five times (in five separate runs) for 3 s each. In the memory representation task, subjects made a single rating for each distinct pair (i.e., responses for 55 unique pairs) to avoid response duplication. We computed response matrices by weighting subjects' binary judgment (yes/no) by their confidence and then scaling distances such that 0 = high confidence that two people are not friends, 1 = high confidence that they are friends, and intermediary values correspond to lesser confidence.

As face stimuli and clubs were randomly assigned for each subject, the predictor matrices (reflecting the contents of each feature space) were sometimes correlated with each other in a given study session and made overlapping predictions. Therefore, we used linear regression to estimate semipartial beta weights for subjects' use of each feature space, which reflect each predictor's unique influence upon the network representation. For the same reason, our analysis controls for subjects learning the true pattern of friendships through associative learning, and beta weights therefore reflect features' unique contribution to representation. We flattened all matrices' upper triangles into vectors before individually regressing each subject's response vectors onto the predictor vectors. Although the beta coefficients from the linear regression are unbiased estimates, this procedure is necessary for group-level inference, as the variance estimates would otherwise conflate the number of pairwise judgments (55 per subject) with the number of observations (one per subject). Group-level inference was performed using nonparametric Wilcoxon sign-rank tests (82).

Study 2: Social Features Enable Flexible Inference of Unobserved Friendships in a Social Network. In Study 2, we used the same network configuration in both phases. To experimentally control for the influence of the demographic characteristics that were observable from the photographs, the gender distribution was evenly split between male and female faces in both networks, as was the distribution of race. Moreover, since clubs (the features used in Study 1) explicitly bind people together into discrete social groups, we alternatively operationalized social features as personal hobbies and college majors in Study 2. Counterbalanced across subjects, one feature was manipulated to be highly predictive of friendships across all networks, and the other was nonpredictive. Because subjects were less accurate at learning which network members were associated with specific social features compared to what was observed in Study 1 (71% for high-predictive and 75% for low-predictive features), we reasoned that subjects' feature maps would likely deviate from the ones we presented them. Therefore, we defined the predictor matrices to reflect subjects' (potentially mistaken) beliefs about network members' social features. Importantly, none of our conclusions change when using network members' true hobbies and majors (*SI Appendix*).

The representation tasks were identical to those used in Study 1, with the following exceptions. First, in the memory task, we accounted for possible memory asymmetries (i.e., representing a directed graph despite learning an undirected graph) by showing subjects every network member as a target. Second, we provided a more psychologically intuitive cover story for the spatial arrangement task by telling subjects that we had previously taken network members to a "happy hour" at a local bar (*SI Appendix*, Fig. S2*E*), where we recorded their physical positions as they walked around the room and interacted with different people. Subjects had to indicate where people were spatially located during the event. Third, we introduced a new representation task requiring subjects to group network members together (*SI Appendix*, Fig. S2*F*). We provided the cover story that we took the network members to dinner after the happy hour and made a note of what tables people seated themselves at. On each trial, subjects were shown a target and guessed which table the target had sat at. As in Study 1, we first estimated subject-specific weights for each measure separately, and then averaged estimates into a composite metric before performing group-level tests (reference *SI Appendix* for nonaggregated measures).

To test how subjects used social maps to predict unobserved friendships in the generalization task, transfer students were sometimes presented with no features, a single feature (i.e., their hobby or college major), or both features. In total, there were nine repetitions of four trial types: only hobbies, only majors, both hobbies and majors, or neither. To avoid gender and race confounds, transfer students were presented as silhouettes. We tested for generalization using mixed-effects logistic regressions examining the effect of shared hobbies and college majors on friendship. Given that features could only be used for generalization if they were accurately remembered, we included subjects' memory accuracy for network members' social features as a predictor. As network members' social status is potentially confounding (e.g., popular people are more likely to get nominated), we also included predictors in our regression analyses that controlled for popularity (degree centrality), brokerage (betweenness centrality), and influence (eigenvector centrality). All mixed-effects regressions were performed using the lme4 library in R (83, 84).

In our computational modeling analysis, we tested five models that used the learned feature maps to make friendship inferences by weighting 1) only features, 2) only a similarity heuristic, 3) associative learning and features, 4) associative learning and a similarity heuristic, and 5) associative learning, features, and a similarity heuristic. The fifth model outperformed the others, as assessed using the median Bayesian Information Criterion, and was formally tested using the likelihood ratio test (*SI Appendix*). We used the Nelder–Mead optimizer implemented in SciPy (85) to estimate free parameters. To ensure that we had thoroughly sampled the parameter space, we estimated free parameters for each subject 50 times, keeping only the parameter values from the greatest maximum likelihood estimate.

Our feature-based learning model is inspired by reinforcement learning models that build predictive relational representations (39, 40, 44, 45) and uses a similar delta rule to learn an approximation of p(friendship | features). Network members are first mapped to an abstract feature space, which is encoded by a N × N feature matrix $\mathbf{M}$, where N is the number of unique features in a particular cognitive map (e.g., biking, music, and sports in the hobby map). Each row of $\mathbf{M}$ represents a particular feature's latent relations with all features, and every row–column pair $\mathbf{M}(i, j)$ encodes the probability of there being a friendship given the latent relation between the features i and j. When a friendship is observed, a one-hot vector $\mathbf{1}(i, j)$ encodes what features have a relation with each other. This one-hot vector is used to compute a prediction error $\delta = \mathbf{1}(i, j) - \mathbf{M}(i, j)$. This prediction error is then used to update the feature matrix using the equation $\mathbf{M}(i, j)_{\text{new}} \leftarrow \mathbf{M}(i, j)_{\text{old}} + \alpha\delta$, where α is the learning rate tempering the update. Since friendships are mutual in our study, we assume symmetric updating for both features. We initialized the feature matrices with subject-supplied priors [i.e., the subject's reported p(friendship | features)], measured prior to learning (*SI Appendix*, Fig. S2*A*).

To assess how much weight subjects placed on each strategy, we estimated subject-specific weights on high- and low-predictive features, a similarity heuristic for the high- and low-predictive features, and use of pure associative learning. Weights were estimated using data from both the friendship memory and generalization task. For the purpose of estimating the associative learning weights, we provided the model with the true network configuration (i.e., the network's adjacency matrix) in the memory task as well as the subject-supplied prior in the generalization task. We operationalized the similarity heuristic mechanism as always predicting a friendship when any feature was shared and never predicting a friendship in the absence of shared features. The weights were permitted to take any real value, and we report standardized weights here. Two subjects were excluded from the computational modeling analysis due to their estimated weights having extremely large values.

Study 3: Feature-Based Cognitive Maps Act as the Dominant Mechanism for Learning Social Networks. Due to the tendency for homophily to be correlated with social features in real-world social networks (11), we changed the cover story of Study 3 so that subjects would not be able to rely on homophily as an inductive prior. Subjects were required to learn friendships between cartoon aliens, and features were operationalized as home planet (non-Earth planets

in our solar system) and occupation (nonsensical jobs such as growing and scheduling). Subjects were randomly assigned to view planets or occupations as the high-predictive feature. To sidestep potential problems related to the ability to accurately remember features, all tasks displayed all aliens' features. There were four high-predictive features that were perfectly (100%) predictive of alien friendship and in which using a similarity heuristic would be an unreliable predictor of friendship. There were also three low-predictive features in which a similarity heuristic was slightly more predictive of friendship but in which using the remaining features would result in an inaccurate network representation. The exact networks and feature matrices used in this study can be found in *SI Appendix*.

The learning task in this study was active rather than passive, requiring subjects to initially guess when observing a friendship for the first time. All 28 possible friendships were shown four times each. The memory and generalization tasks were identical to the ones used in Study 2, except that aliens' features were always displayed.

Our computational modeling approach was similar to Study 2, with two key differences. First, since high-predictive features determined all friendships in this study's network, pure associative learning would make the same predictions as high-predictive features in the memory task. For this reason, we did not use subjects' responses in the memory task to estimate weights. Second, since high-predictive features perfectly dictated what friendships existed in the network, we used friendship memory accuracy as a covariate when examining how much subjects weighted features and a similarity heuristic during learning and generalization, mirroring our analysis in Study 2.

We estimated free parameters for each subject 100 times, keeping only the parameter values from the greatest maximum likelihood estimate. For the majority of subjects (67%), the best-fitting model included mechanisms for both the similarity heuristic and cognitive maps. For the remaining subjects (33%), this model fit was not significantly better than a simpler model that only included a cognitive map. We report results from the model that tests both the similarity heuristic and cognitive maps, as it is the best-fitting model for most subjects and additionally allows us to draw comparisons between Studies 2 and 3. We note that this simpler model produces a qualitatively similar pattern of results. Two subjects were excluded from the computational modeling analysis due to the estimated weights having extremely large values.

Data Availability. The data and code supporting the findings of this manuscript are available online at the Open Science Framework at the following URL: https://osf.io/v8ucz/.

ACKNOWLEDGMENTS. We thank Danai Benopoulou, Logan Bickel, Tina Tan, and Hanzhang Nina Zhao for helping collect data; David Badre, Michael J. Frank, Seongmin A. Park, and Linda Yu for helpful comments on previous versions of this manuscript; Joseph Heffner and Amrita Lamba for computational modeling advice; and Erin E. Burman for creating stimuli for the third study. Part of this research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. This material is based upon work supported by the NSF Graduate Research Fellowship award number 2040433 (J.-Y.S.), the National Institute of Neurological Disorders and Stroke of the NIH under award number R21NS108380 (A.B.), and a Carney Innovation Grant from the Robert J. and Nancy D. Carney Institute for Brain Science (O.F.H.).

1. R. A. Brands, Cognitive social structures in social network research: A review. *J. Organ. Behav.* **34**, S82–S103 (2013).
2. D. Krackhardt, Assessing the political landscape: Structure, cognition, and power in organizations. *Adm. Sci. Q.* **35**, 342–369 (1990).
3. O. FeldmanHall, J. E. Dunsmoor, Viewing adaptive social choice through the lens of associative learning. *Perspect. Psychol. Sci.* **14**, 175–196 (2018).
4. O. FeldmanHall, J. E. Dunsmoor, M. C. W. Kroes, S. Lackovic, E. A. Phelps,Associative learning of social value in dynamic groups. *Psychol. Sci.* **28**, 1160–1170 (2017).
5. M. O. Jackson, *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors* (Pantheon, 2019).
6. P. D. Killworth, H. R. Bernard, Informant accuracy in social network data. *Hum. Organ.* **35**, 269–286 (1976).
7. H. R. Bernard, P. D. Killworth, Informant accuracy in social network data II. *Hum. Commun. Res.* **4**, 3–18 (1977).
8. S. J. Gershman, H. T. Pouncy, H. Gweon, Learning the structure of social influence. *Cogn. Sci. (Hauppauge)* **41** (suppl. 3), 545–575 (2017).
9. T. Lau, H. T. Pouncy, S. J. Gershman, M. Cikara, Discovering social groups via latent structure learning. *J. Exp. Psychol. Gen.* **147**, 1881–1891 (2018).
10. E. B. Smith, R. A. Brands, M. E. Brashears, A. M. Kleinbaum, Social networks and cognition. *Annu. Rev. Sociol.* **46**, 159–174 (2020).
11. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
12. P. F. Lazarsfeld, R. K. Merton, "Friendship as a social process: A substantive and methodological analysis" in *Freedom and Control in Modern Society*, M. Berger, T. Abel, C. H. Page, Eds. (Van Nostrand, New York, NY, 1954), 18, pp. 18–66.
13. H. Tajfel, J. C. Turner, "The social identity theory of intergroup behavior" in *Political Psychology: Key Readings*, J. T. Jost, J. Sidanius, Eds. (Psychology Press, 2004) pp. 276–293.
14. J. A. Smith, M. McPherson, L. Smith-Lovin, Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *Am. Sociol. Rev.* **79**, 432–456 (2014).
15. A. Mayer, S. L. Puller, The old boy (and girl) network: Social network formation on university campuses. *J. Public Econ.* **92**, 329–347 (2008).
16. T. E. J. Behrens *et al.*, What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
17. E. C. Tolman, Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
18. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon Press, Oxford, 1978).
19. C. F. Doeller, C. Barry, N. Burgess, Evidence for grid cells in a human memory network. *Nature* **463**, 657–661 (2010).
20. J. L. S. Bellmund, P. Gärdenfors, E. I. Moser, C. F. Doeller, Navigating cognition: Spatial codes for human thinking. *Science* **362**, eaat6766 (2018).
21. A. O. Constantinescu, J. X. O'Reilly, T. E. J. Behrens, Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016).
22. N. W. Schuck, M. B. Cai, R. C. Wilson, Y. Niv, Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91**, 1402–1412 (2016).
23. S. Mark, R. Moran, T. Parr, S. W. Kennerley, T. E. J. Behrens, Transferring structural knowledge across cognitive maps in humans and models. *Nat. Commun.* **11**, 4783 (2020).
24. S. H. Tompson, A. E. Kahn, E. B. Falk, J. M. Vettel, D. S. Bassett, Functional brain network architecture supporting the learning of social networks in humans. *Neuroimage* **210**, 116498 (2020).
25. R. M. Tavares *et al.*, A map for social navigation in the human brain. *Neuron* **87**, 231–243 (2015).
26. S. A. Park, D. S. Miller, H. Nili, C. Ranganath, E. D. Boorman, Map making: Constructing, combining, and inferring on abstract cognitive maps. *Neuron* **107**, 1226–1238.e8 (2020).
27. C. Parkinson, A. M. Kleinbaum, T. Wheatley, Spontaneous neural encoding of social network position. *Nature Human Behaviour* **1**, 0072 (2017).
28. S. A. Morelli, Y. C. Leong, R. W. Carlson, M. Kullar, J. Zaki, Neural detection of socially valued community members. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 201712811 (2018).
29. M. Peer, M. Hayman, B. Tamir, S. Arzy, Brain coding of social network structure. *J. Neurosci.* **41**, 4897–4909 (2021).
30. H. F. Harlow, The formation of learning sets. *Psychol. Rev.* **56**, 51–65 (1949).
31. R. N. Shepard, Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
32. J. B. Tenenbaum, T. L. Griffiths, Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* **24**, 629–640, discussion 652–791 (2001).
33. A. G. E. Collins, M. J. Frank, Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
34. A. G. E. Collins, M. J. Frank, Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition* **152**, 160–169 (2016).
35. M. Rmus, H. Ritz, L. E. Hunter, A. M. Bornstein, A. Shenhav, Humans can navigate complex graph structures acquired during latent learning. *bioRxiv* [Preprint] (2019). https://doi.org/0.1101/723072. Accessed 25 August 2021.
36. M. Peer, I. K. Brunec, N. S. Newcombe, R. A. Epstein, Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn. Sci.* **25**, 37–54 (2021).
37. J. C. R. Whittington *et al.*, The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249–1263.e23 (2020).
38. E. M. Russek, I. Momennejad, M. M. Botvinick, S. J. Gershman, N. D. Daw, Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput. Biol.* **13**, e1005768 (2017).
39. I. Momennejad *et al.*, The successor representation in human reinforcement learning. *Nat. Hum. Behav.* **1**, 680–692 (2017).
40. I. Momennejad, Learning structures: Predictive representations, replay, and generalization. *Curr. Opin. Behav. Sci.* **32**, 155–166 (2020).
41. E. S. Spelke, K. D. Kinzler, Core knowledge. *Dev. Sci.* **10**, 89–96 (2007).
42. A. Pun, S. A. J. Birch, A. S. Baron, The power of allies: Infants' expectations of social obligations during intergroup conflict. *Cognition* **211**, 104630 (2021).
43. M. K. Ho, D. Abel, T. L. Griffiths, M. L. Littman, The value of abstraction. *Curr. Opin. Behav. Sci.* **29**, 111–116 (2019).
44. A. Barreto *et al.*, Successor features for transfer in reinforcement learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017), pp. 4058–4068. https://arxiv.org/abs/1606.05312.
45. L. Lehnert, M. L. Littman, Successor features combine elements of model-free and model-based reinforcement learning. *J. Mach. Learn. Res.* **21**, 1–53 (2020).
46. L. Lehnert, M. L. Littman, M. J. Frank, Reward-predictive representations generalize across tasks in reinforcement learning. *PLoS Comput. Biol.* **16**, e1008317 (2020).
47. W. de Cothi, C. Barry, Neurobiological successor features for spatial navigation. *Hippocampus* **30**, 1347–1355 (2020).
48. C. Kemp, J. B. Tenenbaum, Structured statistical models of inductive reasoning. *Psychol. Rev.* **116**, 20–58 (2009).

49. C. Kemp, J. B. Tenenbaum, S. Niyogi, T. L. Griffiths, A probabilistic model of theory formation. *Cognition* **114**, 165–196 (2010).
50. N. Kriegeskorte, M. Mur, M. D. S. Inverse, Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Front. Psychol.* **3**, 245 (2012).
51. R. C. Wilson, Y. Niv, Inferring relevance in a changing world. *Front. Hum. Neurosci.* **5**, 189 (2012).
52. S. M. Weisberg, N. S. Newcombe, Cognitive maps: Some people make them, some people struggle. *Curr. Dir. Psychol. Sci.* **27**, 220–226 (2018).
53. A. G. E. Collins, J. F. Cavanagh, M. J. Frank, Human EEG uncovers latent generalizable rule structure during learning. *J. Neurosci.* **34**, 4677–4685 (2014).
54. A. G. E. Collins, The cost of structure learning. *J. Cogn. Neurosci.* **29**, 1646–1655 (2017).
55. A. R. Vaidya, H. M. Jones, J. Castillo, D. Badre, Neural representation of abstract task structure during generalization. *eLife* **10**, e63226 (2021).
56. Y. Niv, Learning task-state representations. *Nat. Neurosci.* **22**, 1544–1553 (2019).
57. S. J. Gershman, N. D. Daw, Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annu. Rev. Psychol.* **68**, 101–128 (2017).
58. V. P. Murty, O. FeldmanHall, L. E. Hunter, E. A. Phelps, L. Davachi, Episodic memories predict adaptive value-based decision-making. *J. Exp. Psychol. Gen.* **145**, 548–558 (2016).
59. A. M. Bornstein, M. W. Khaw, D. Shohamy, N. D. Daw, Reminders of past choices bias decisions for reward in humans. *Nat. Commun.* **8**, 15958 (2017).
60. H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, H. Tanila, The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron* **23**, 209–226 (1999).
61. H. Eichenbaum, N. J. Cohen, Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* **83**, 764–770 (2014).
62. D. Schiller *et al*., Memory and space: Towards an understanding of the cognitive map. *J. Neurosci.* **35**, 13904–13911 (2015).
63. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
64. D. Shohamy, A. D. Wagner, Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron* **60**, 378–389 (2008).
65. A. M. Bornstein, K. A. Norman, Reinstated episodic context guides sampling-based decisions for reward. *Nat. Neurosci.* **20**, 997–1003 (2017).
66. O. FeldmanHall, D. F. Montez, E. A. Phelps, L. Davachi, V. P. Murty, Hippocampus guides adaptive learning during dynamic social interactions. *J. Neurosci.* **41**, 1340–1348 (2021).
67. E. C. Baek, M. A. Porter, C. Parkinson, Social network analysis for social neuroscientists. *Soc. Cogn. Affect. Neurosci.* **16**, 883–901 (2020).
68. G. W. Allport, *The Nature of Prejudice* (Addison-Wesley, Oxford, United Kingdom, 1954).
69. C. Tilly, *The Politics of Collective Violence* (Cambridge University Press, 2003).
70. J. Bakonyi, B. B. de Guevara, *A Micro-Sociology of Violence: Deciphering Patterns and Dynamics of Collective Violence* (Routledge, 2014).
71. M. Kilduff, C. Crossland, W. Tsai, D. Krackhardt, Organizational network perceptions versus reality: A small world after all? *Organ. Behav. Hum. Decis. Process.* **107**, 15–28 (2008).
72. L. C. Freeman, Filling in the blanks: A theory of cognitive categories and the structure of social affiliation. *Soc. Psychol. Q.* **55**, 118–127 (1992).
73. M. E. Brashears, Humans use compression heuristics to improve the recall of social networks. *Sci. Rep.* **3**, 1513 (2013).
74. M. E. Brashears, E. Quintane, The microstructures of network recall: How social networks are encoded and represented in human memory. *Soc. Networks* **41**, 113–126 (2015).
75. P. Wang, B. Xu, Y. Wu, X. Zhou, Link prediction in social networks: The state-of-the-art. *Sci. China Inf. Sci.* **58**, 1–38 (2015).
76. S. Haghani, M. R. Keyvanpour, A systemic analysis of link prediction in social network. *Artif. Intell. Rev.* **52**, 1961–1995 (2019).
77. R. Schifanella, A. Barrat, C. Cattuto, B. Markines, F. Menczer, "Folks in Folksonomies: Social link prediction from shared metadata" in *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (Association for Computing Machinery, New York, 2010), pp. 271–280.
78. R. M. Mok, B. C. Love, A non-spatial account of place and grid cells based on clustering models of concept learning. *Nat. Commun.* **10**, 5685 (2019).
79. M. S. Tomov, E. Schulz, S. J. Gershman, Multi-task reinforcement learning in humans. *Nat. Hum. Behav.* **5**, 764–773 (2021).
80. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal. Complex Syst.* **1695**, 1–9 (2006).
81. D. S. Ma, J. Correll, B. Wittenbrink, The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **47**, 1122–1135 (2015).
82. H. Nili *et al*., A toolbox for representational similarity analysis. *PLOS Comput. Biol.* **10**, e1003553 (2014).
83. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **1** (2015).
84. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest package: Tests in linear mixed effects models. *J. Stat. Soft.* **1** (2017).
85. P. Virtanen *et al*, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).