PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Abstract cognitive maps of social network structure aid adaptive inference

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

Edited by Michael Gazzaniga, University of California Santa Barbara College of Letters and Science, Santa Barbara, CA; received June 28, 2023; accepted October 12, 2023

Social navigation—such as anticipating where gossip may spread, or identifying which acquaintances can help land a job—relies on knowing how people are connected within their larger social communities. Problematically, for most social networks, the space of possible relationships is too vast to observe and memorize. Indeed, people's knowledge of these social relations is well known to be biased and error-prone. Here, we reveal that these biased representations reflect a fundamental computation that abstracts over individual relationships to enable principled inferences about unseen relationships. We propose a theory of network representation that explains how people learn inferential cognitive maps of social relations from direct observation, what kinds of knowledge structures emerge as a consequence, and why it can be beneficial to encode systematic biases into social cognitive maps. Leveraging simulations, laboratory experiments, and "field data" from a real-world network, we find that people abstract observations of direct relations (e.g., friends) into inferences of multistep relations (e.g., friends-of-friends). This multistep abstraction mechanism enables people to discover and represent complex social network structure, affording adaptive inferences across a variety of contexts, including friendship, trust, and advice-giving. Moreover, this multistep abstraction mechanism unifies a variety of otherwise puzzling empirical observations about social behavior. Our proposal generalizes the theory of cognitive maps to the fundamental computational problem of social inference, presenting a powerful framework for understanding the workings of a predictive mind operating within a complex social world.

social networks | cognitive maps | abstraction | successor representation

Significance

Successful navigation through our social communities—from identifying which "weak ties" can help land a job to deciding whether a friend-of-a-friend can be trusted—necessitates that people build cognitive maps of how individuals are connected within a social network. Because our observations of others' relationships are noisy and incomplete, people need to fill gaps in relational knowledge by inferring the existence of unknown and unobserved relationships. Here, we demonstrate that predictive inferences emerge from a simple cognitive mechanism that abstracts observations of known friendships to multistep relations (e.g., friends-of-friends). This mechanism can parsimoniously explain how people infer a variety of social behaviors, like trust and advice-seeking.

In our daily lives, we are constantly navigating social relationships in pursuit of the rich resources embedded within social networks (1, 2). Someone applying for a job might leverage their acquaintances to get a favorable referral (3, 4); a politician might wage a whisper campaign against a rival, seeding salacious rumors likely to reach key constituencies; a crafty rank-and-file employee might cozy up to managers in hopes of being promoted (5). Researchers commonly assume that adaptive social navigation is aided by having accurate knowledge of network structure (6, 7). Indeed, people accurately identify socially important network members (8, 9), invest time and energy tracking each other's relationships (10), and report believing that social power comes from having accurate knowledge of network structure (11). It is therefore a puzzling empirical fact that people's perceptions of social network structure are riddled with errors, deviating substantially from the true network structure (12–16). As just one example, people believe their social networks to be far more interconnected than they actually are and exaggerate how strongly relations are clustered into groups (17–19).

To explain people's lack of accurate network knowledge, past research posits that people maintain a variety of cognitive schemas about network structure (6, 10, 20). For example, people hold the schematic belief that networks are structured into triads, where three individuals are all connected to each other. Upon learning that two individuals share a mutual friend, people can heuristically apply this schema to infer that these two individuals must also be friends (14, 21). Similarly, groups of densely interconnected individuals can be represented using a community schema, which leads people to heuristically invent nonexistent relationships between members of the same group or community (15, 21). Therefore, under a schema-based account, inaccurate network representation reflects compression of relational knowledge into heuristics (14, 22), which reshape memories of previously observed relationships and bias learning of new relationships (6, 10, 20).

Although schemas offer useful descriptions of how people perceive social networks, they are unsatisfactory as a theoretical account for two fundamental reasons. First, a schema-based account lacks parsimony. To explain an individual's overall network representation, it would be necessary to draw upon a collection of different schemas, each specifying a unique network structure to be represented in memory. Without a formal

Author affiliations: [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

Author contributions: J.Y.S., A.B., and O.F.H. designed research; J.Y.S. performed research; J.Y.S. 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.2310801120/-/DCSupplemental.

Published November 14, 2023.

theory specifying which schemas should be used, or when they should be combined, it is difficult to predict how an individual might encode their network. Second, a schema-based account is unfalsifiable because it sidesteps the crucial question of where schemas come from. In the absence of a theory of how schemas develop, which would place constraints on what schemas could even exist, any novel empirical observation can be explained simply by adding an additional schema to the collection. But what is the alternative? Is there a single falsifiable mechanism that might parsimoniously explain how and why these various psychological schemas emerge?

We start from a well-established principle in cognitive science: Abstraction strips away concrete details about specific observations or experiences, leaving behind deep structural knowledge about the world around us. This abstracted knowledge can then be used to draw sophisticated inferences from limited direct experience (23), including about social relationships (24). From this vantage point, social navigation is not just about deciding how to traverse known relationships. Instead, it begins with inference about what pathways exist at the moment, as well as what pathways could be created in the future. For example, executives create value for the company (and for themselves) by noticing and then bridging otherwise isolated teams (7); matchmakers gain prestige by correctly guessing clients' compatibility; political goons and propagandists help their party consolidate power by inciting intergroup violence along dormant us-them boundaries, reshaping how network members relate to each other (25–27). The common cognitive problem underlying both "path-finding" and "path-forging" forms of social navigation is the need to make predictive inferences of how people are—or will be—connected to each other within social networks. We refer to this relational inference problem as social link prediction, adopting a term from the computer science literature where analogous problems include friend recommendation for online social media systems (28, 29).

If social navigation is about solving the link prediction problem, then abstraction offers a solution in the discovery of deep structure; knowledge of such structure subsequently enables inference despite noisy and incomplete relational observation (30). Leveraging the theoretical framework of abstraction, we investigate a mechanism for performing link prediction, which can account for a diverse set of empirical observations and unifies previous schema-based accounts into a single, parsimonious mechanism. This mechanism is called multistep relational abstraction, where network members' relationships are represented as a combination of direct and indirect connections (i.e., one-step and multistep relations) (31, 32). For example, when observing a friendly interaction between two people, we not only obtain evidence about the observed friendship but also infer that each individual in the interaction is more likely to be friends with the other person's friends (Fig. 1). With sufficient observation, we can build a representation that incorporates observed knowledge about one-step relations, as well as inferences about multistep relations (such as friends-of-friends, friends-of-friends-of-friends, and so on). Multistep relational abstraction provides a principled solution for solving the link prediction problem and specifies a mechanism for building cognitive maps that reflect how people are connected in any given network. It is therefore a flexible mechanism, allowing relationships to be probabilistically inferred between any network members who share many intermediary connections, even when evidence of those relationships has never been directly observed. Across four studies, we test the hypothesis that multistep relational abstraction provides a general, parsimonious, and flexible cognitive mechanism that explains how people build mental representations of social networks.

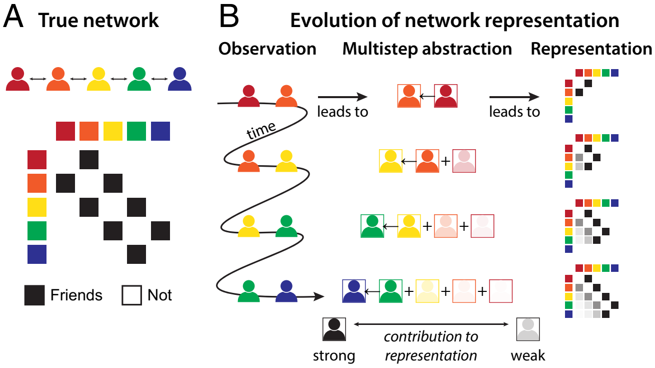

Fig. 1. Conceptual model. (*A*) Network representation can be visualized as a matrix of (friend) relations. In this example, the answer to "Is Red friends with Orange?" is answered by referencing row Red, column Orange. (*B*) Multistep relational abstraction represents network members' relations as a combination of direct (i.e., one-step: observed friendship) and indirect (i.e., multistep: friends-of-friends) connections. Therefore, when observing Blue and Green together, the model represents Blue as being fractionally more likely to be directly connected to Green's friend (Yellow), friends-of-friends (Orange), and so on. Each additional step is down-weighted such that more immediate connections are prioritized in the overall representation.

Results

Study 1: Schema-Like Representation Emerges from Multistep Relational Abstraction. As a first proof of principle, we conducted a simulation study to interrogate whether multistep abstraction can generate schema-like network representations like those observed in empirical studies and tested whether these abstractions are capable of successful link prediction. We implemented multistep abstraction using the reinforcement learning algorithm for learning the Successor Representation (33, 34), which specifies how learning experiences can be used to build cognitive maps in memory (32, 35). We note, however, that other implementations make functionally equivalent predictions (31, 36; see *Discussion* for a comparison between the Successor Representation and related mechanisms).

For any two given network members (e.g., Isa and Asher), the Successor Representation encodes how likely it is that they will be observed together. In our model, this co-occurrence probability informs link prediction. Critically, the Successor Representation computes co-occurrence probabilities through a temporal-difference learning mechanism that combines knowledge of one-step and multistep relations (*Methods*). Generally, the key benefit of using a temporal-difference learning mechanism is that an agent can immediately use new observations to make better predictions; in the context of social networks, the Successor Representation is therefore able to infer unobserved social relations immediately after a novel observation of social interaction. For example, upon observing an interaction between Isa and Asher, the Successor Representation learns that Isa is more likely to be observed together with Asher in the future. The Successor Representation does not stop with this one-step update but also incorporates multistep relations into the update (i.e., Isa will be observed with Asher's friends, friends-of-friends, and so on; Fig. 1). Each successive step is weighted by the successor discount parameter $\gamma \in [0, 1)$ so that the actual observation contributes most strongly to the learning update (i.e., by a factor of γ^0), the inferred relationships between Isa and Asher's friends contribute less strongly (γ^1), and so on. The value of γ therefore dictates the "abstractness" of overall network representation, such that $\gamma = 0$ is pure memorization (i.e., one-step relations), and greater values of γ lead to increasingly abstract multistep inferences.

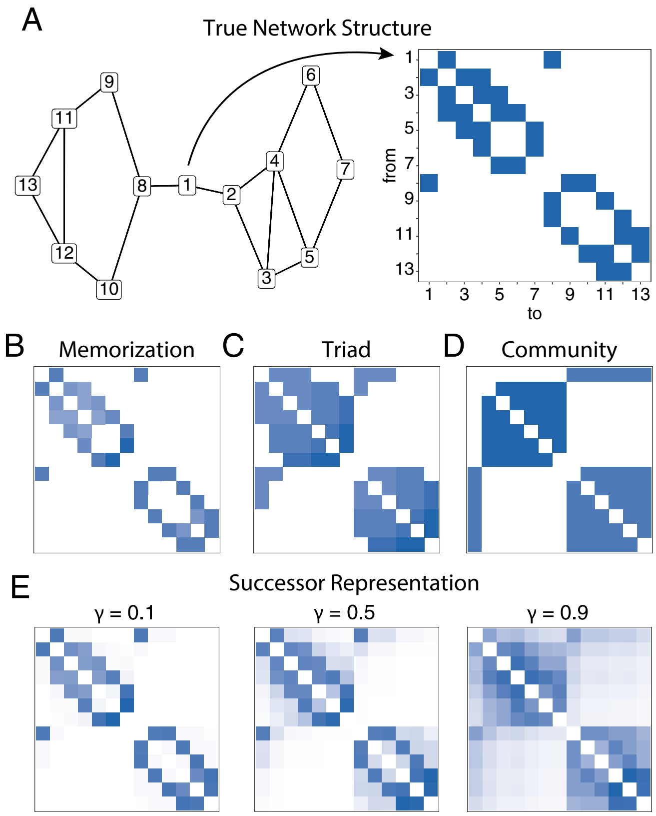

Fig. 2. Study 1: Simulation results. (*A*) The true network topology, visualized equivalently as a graph and adjacency matrix. (*B*) Network representation based on pure memorization. (*C*) Representation predicted by the triad completion schema, i.e., inferring friendship between friends-of-friends. (*D*) Representation based on community detection. (*E*) By adjusting how many steps are integrated over in a multistep relational abstraction, network representation can resemble memorization, triad completion, or community detection. Zero indicates a prediction of no friendship and is color-coded white. Colormaps are rescaled for each panel to emphasize the pattern geometry.

Using only a single parameter γ, multistep abstraction potentially allows us to unite multiple schemas under the same model. To test this, we created an artificial social network (Fig. 2*A*), structured such that memorization (Fig. 2*B*), triad completion (Fig. 2*C*), and community detection (Fig. 2*D*) strategies make divergent predictions about unobserved links in the network representation. We then simulated variants of the Successor Representation with γ ranging from 0.1 to 0.9 (Fig. 2*E*). Simulation results demonstrate multistep abstraction is sufficient for producing representations resembling triad completion, community detection, or a blend of the two, depending on the number of steps being integrated over. In other words, we observed evidence that both triad completion and community detection schemas reflect the same basic cognitive process but at different levels of abstraction. Put simply, multistep abstraction can in principle provide a parsimonious mechanism supporting different schema-like network representations.

Study 2a: Multistep Relational Abstraction Explains People's Network Representations. Following our simulation study showing that multistep abstraction recapitulates various known schemas, we empirically tested whether humans actually deploy multistep abstraction when learning about networks through direct experience. We used two complementary samples to test our hypothesis: one using a "Random Walks" learning task (N = 60) and the other using a "Paired Associates" learning task (N = 28). The Random Walks task consisted of a sequence of network members presented one-by-one, such that observing the $i \rightarrow j$ transition meant that network members i and j were friends. Each "step" of the random walk was chosen randomly from all of a given node's edges, such that upon seeing i, it would be equiprobable to see any of i's friends on the next trial. The Paired Associates task presented a pair of friends on each trial, randomly drawn from all friendships in the network.

Random walks provide firsthand observations of each network member's successors (e.g., friends-of-friends) and most closely match how Successor Representations have previously been studied in nonsocial contexts (31, 36–39). This format also emulates the kinds of relational observations provided by gossip (8), e.g., "Madeleine heard from Asher, who heard from Isa that…" Since participants experience long chains of observations—and, in fact, are directly provided with information about multistep relations—this learning format may facilitate more community-like representation, described in our model by relatively high values of γ. The paired associates format, in contrast, withholds information about multistep relations and breaks the temporal structure provided by random walks. This emulates learning from disjointed pairwise observations (e.g., seeing Madeleine getting lunch with Asher, then later seeing Asher having coffee with Isa) and conveys no additional information about multistep relations. Therefore, building an abstracted cognitive map from paired associates would require stitching together disjointed experiences in memory, which may facilitate the more triad-like representation described by lower values of γ in our model.

Participants in both samples learned about the same network from Study 1. We measured participants' mental representations by showing them every possible pairwise combination of network members and asking them to report a "yes/no" decision about whether each pair was friends. To first verify that participants' mental representations showed evidence of being schema-like, we used mixed-effects regression to predict how people responded to the memory task. In both samples, results reveal that participants' mental representations were significantly predicted by a triad completion schema (Random Walks β = 8.46, Z = 7.67, 95% CI = [6.30, 10.62], P < 0.001; Paired Associates β = 7.15, Z = 6.95, 95% CI = [5.13, 9.16], P < 0.001), and also by a community schema (Random Walks β = 9.29, Z = 6.77, 95% CI = [6.60, 11.98], P < 0.001; Paired Associates β = 7.41, Z = 6.86, 95% CI = [5.29, 9.53], P < 0.001). Both models included random intercepts and slopes.

Given evidence that participants' mental representations were schema-like, we then tested whether multistep abstraction is a viable alternative account of the data. To directly compare schemas against multistep abstraction, we used mixed-effects regression to test how well mental representations are predicted by fixed-γ Successor Representations (i.e., nine variants with $\gamma \in [0.1,\ 0.9]$ in increments of 0.1). Using the Bayesian Information Criterion (BIC) as a goodness-of-fit metric, we found that the Random Walks sample was best explained by a Successor Representation with γ = 0.7 (i.e., more community-like), that the Paired Associates sample was best explained by a Successor Representation with γ = 0.4 (i.e., more triad-like), and that Successor Representations at any value of γ outperformed schemas in both samples (*SI Appendix*). These results 1) support our hypothesis that multistep abstraction subsumes schematic representation, 2) explain how schema-like representation emerges even from complete and veridical observation of a social network, and 3) provide a basis for understanding how people solve the link prediction problem within social networks. However, this analysis is limited by the need to use fixed γ values for the Successor Representation, as well as the assumption that there are no meaningful individual differences (i.e., that everyone learns using the same γ).

Since multistep abstraction can simultaneously explain memorization, triad completion, and community detection, we fit a computational model to each participant's data, formally testing whether the Successor Representation can flexibly account for individual differences in people's use of schema-like strategies. Unlike before, this model estimates γ as a free parameter, as well as weighting terms controlling how strongly the Successor Representation is incorporated into network representation (*Methods*). Even when accounting for individual differences in γ strength, results from the computational model reveal significant group-level use of the Successor Representation in both our samples (mean Random Walks $\beta = 1.74$, $t(59) = 6.49$, 95% CI = [1.20, 2.27], $P < 0.001$; mean Paired Associates $\beta = 1.55$, $t(27) = 6.44$, 95% CI = [1.05, 2.04], $P < 0.001$). Reflecting the difference in learning formats, we found that participants integrated over significantly more steps in the Random Walks sample (mean lookahead = 5.89 steps, equivalent to $\gamma = 0.83$), compared to the Paired Associates sample (mean lookahead = 2.25 steps, equivalent to $\gamma = 0.56$; mean lookahead difference $t(73) = -3.65$ steps, 95% CI = [−5.63, −1.66], $P < 0.001$; Fig. 3*A*). The posterior predictive check also reflects that participants' representations are well characterized as a blend of both triad- and community-like structure (Fig. 3*B*) and confirms that multistep abstraction is capable of flexibly producing schema-like representation.

Study 2b: Multistep Relational Abstraction Explains People's Social Inferences. In addition to structuring mental representations, multistep abstraction also affords link prediction in the service of adaptive social decision-making in situations where no direct experience is available. For example, if Madeleine needed to use the bathroom at a café, whom should she ask to look after her laptop? This is a difficult problem if she is not friends with any of the people around her and has not had a chance to directly assess their trustworthiness. However, Madeleine might leverage her knowledge of multistep relations to quickly infer that a friend-of-a-friend is likely more reliable than a friend-of-a-friend-of-a-friend and ask a cafegoer whom she recognizes as a mutual friend. To test whether multistep abstraction supports link prediction in the service of social decisions, participants in the Paired Associates sample completed a trust prediction task, which required inferences about how much network members trusted each other in an economic decision-making task, based on nothing but knowing who is friends with whom.

Participants were told that one network member (A) was given \$10 and could invest any amount in another network member (B). Any money invested by A would be quadrupled and sent to B, who could then choose how much of the money (if any) to send back to A. Participants then guessed how much money each network member trusted with every other network member (\$0 to 10, in \$1 increments). This task therefore requires participants to make inferences about how much network members, including those with no known relationship, would trust each other.

Before testing whether individual differences in abstraction are associated with differences in inferring how much friends-of-friends trust one another, we checked our intuition that trust behaviors are well-described by the same computational model used to predict mental representations at the group level. That is, friends-of-friends should be trusted with more money than friends-of-friends-of-friends, and so on. As hypothesized, the weights for the trust prediction task reveal significant group-level use of the Successor Representation with an average $\gamma = 0.89$ (mean $\beta = 1.86$, $t(27) = 3.60$, 95% CI = [0.80, 2.93], $P = 0.001$; Fig. 3*A*), suggesting that the same cognitive mechanism may underlie both mental representation of the network and link prediction in the service of social decisions (Fig. 3*B*).

We then used participant-specific memory Successor Representations predicted by the computational model (Fig. 3*B*) to predict inferences in the trust prediction task. For example, a subset of participants seemed to prioritize memorization over abstraction in memory (i.e., $\gamma < 0.1$; 33% in Random Walks, 43% in Paired Associates). Although this would indeed be the ideal strategy if optimizing for accurate representation, insufficient abstraction diminishes an individual's ability to perform link prediction, and as a consequence, the ability to make adaptive social inferences. If this is true, we should observe that participants with greater

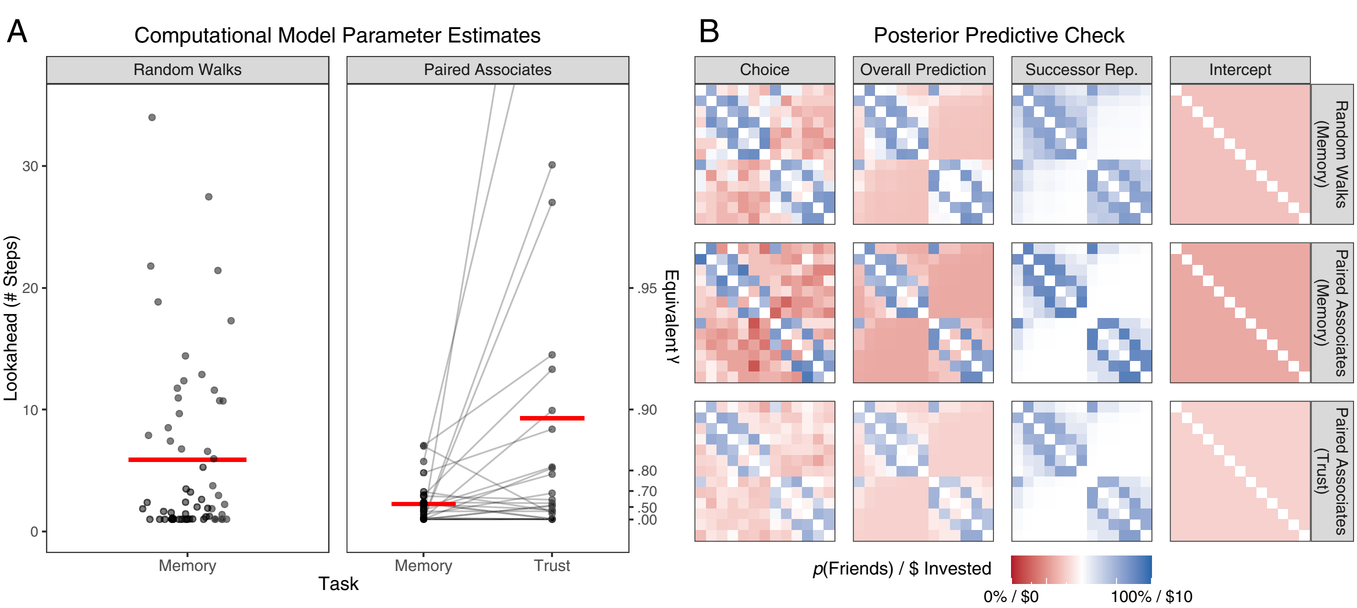

Fig. 3. Study 2: Computational modeling results. (*A*) Estimated values of γ. Because γ is on an exponential scale, it is useful to convert γ into a "lookahead" horizon, indicating the number of steps being integrated over in the model. We find a wide spread of individual differences, with a subset of participants prioritizing memorization (i.e., small values of γ), others representing triad-like structure (i.e., middling values of γ), and still others representing community-like structure (i.e., larger values of γ). Very large values of γ are omitted from the plot to aid visualization of more typical values of γ. (*B*) Posterior predictive check comparing the average of participants' choices in the memory task (Choice) with the computational model's averaged predictions (Overall Prediction, Successor Representation, and Intercept).

memory abstraction (i.e., higher γ) infer greater trust between more distant connections (e.g., friend-of-a-friend-of-a-friend), compared to participants who use less abstraction. This is precisely what we found: The greater a participant's ability to rely on relational abstraction, the more likely they were to infer that money was entrusted across longer-range distances within the network (β = 15.20, t = 5.50, 95% CI = [9.52, 20.87], P < 0.001; full mixed-effects regression model specification and results in *SI Appendix*). Together, these results provide evidence that multistep abstraction not only shapes people's memories of relationships but also enables link prediction for never-before-seen friendships in the service of socially consequential behaviors.

Study 3: Multistep Relational Abstraction Explains Link Prediction from Minimal Network Information. We have so far demonstrated that multistep abstraction explains how schema-like representation of social networks emerges from direct experience and that the same mechanism enables link prediction for unobserved relationships, shaping both mental representations and socially consequential inferences. What remains unknown is whether multistep abstraction is capable of explaining empirical phenomena that are less readily characterized by known schemas and just how little direct experience is necessary for multistep abstraction to support link prediction. To answer these questions, we reanalyzed a dataset by Sehl et al. (40), which tested how participants performed link prediction in artificial social networks after being provided summary information about relationships within these networks (40). Participants were presented with two "target" network members and told how many friends these targets had in total, as well as how many mutual friends they shared. The task was to infer how likely it was that the targets were themselves friends. Participants in this experiment (N = 180) were informed (in various conditions) that in a network of 26 individuals, the two targets had 6 or 12 friends each (Fig. 4*A*). Of these, the targets shared mutual friends in ratios of 1:6, 1:2, or 5:6.

The authors observed two specific effects. First, participants were more likely to infer friendship between the targets when they had a greater proportion of mutual friends. Second, participants were more likely to endorse friendships when the targets each had 12 friends compared to 6 (Fig. 4*B*). Although these inferences seem to suggest heuristic application of schema-like knowledge, a single schema cannot fully account for the results. Although a triad completion schema predicts a link between two individuals who share a mutual friend and might even (with some elaboration) produce the belief that two individuals are increasingly likely to be linked if they share many mutual friends, it cannot explain why people are more likely to infer a link between individuals with a greater number of friends overall (holding the proportion of mutual friends constant). One could explain this effect with a different schema that more popular individuals are more likely to

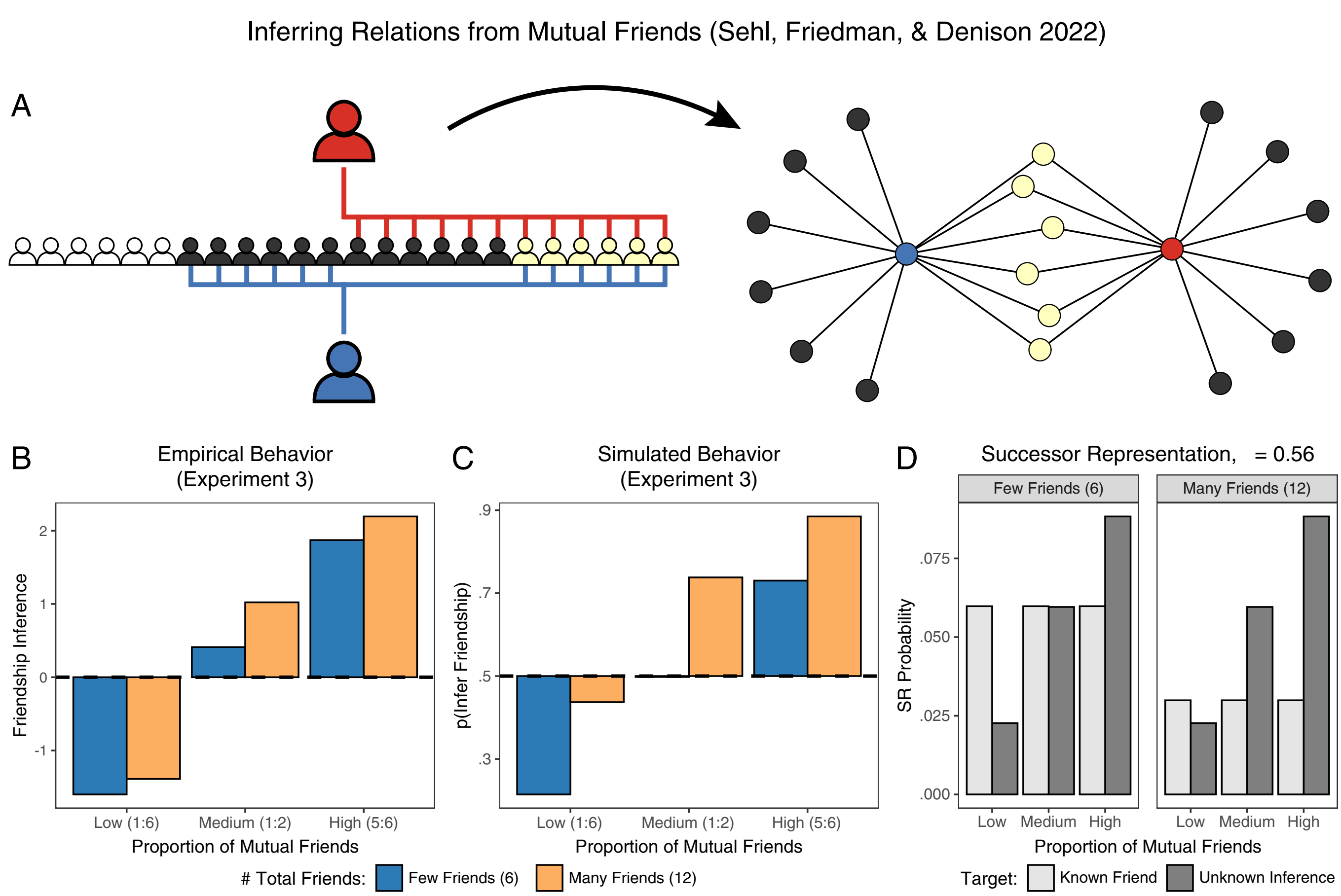

Fig. 4. Study 3: Reanalysis of Sehl et al. (40). (*A*) An example of the kind of schematic participants were shown on each trial (*Left*), and the network graph corresponding to that configuration (*Right*). Participants judged how likely they found it that the target individuals (red and blue) were friends, given information about mutual friends (yellow), exclusive friends (black), and isolates (white). Lines connecting the Red/Blue individuals to other network members indicate known friendships. Isolates are omitted from the graph because they are not connected to anyone else. (*B*) Group-level empirical behavior from a laboratory study where participants inferred friendship between two individuals, given how many mutual and total friends they had. (*C*) Simulated choices based on the Successor Representation. The model qualitatively reproduces participants' behaviors. (*D*) Simulated co-occurrence probabilities for the experimental conditions, which inform the softmax choice function.

be connected. However, it is not parsimonious to use multiple heuristics to account for these effects in piecemeal, if a single alternative mechanism can account for both.

Rather than relying on an account that combines an arbitrary set of schemas, can multistep abstraction parsimoniously explain how people make inferences about unseen friendships, even in situations where people have very limited knowledge about the network? Specifically, can it reproduce the two effects observed by Sehl et al.? To test this, we modeled their experimental setup by first computing the Successor Representation from the limited information provided on every trial (Fig. 4*A*; see *Methods*) and then using the Successor Representation to compute two quantities: 1) the likelihood of observing a friendship between a target and one of their true friends and 2) the likelihood of observing the unknown friendship between the targets. In our model, the probability of inferring friendship is derived from a softmax choice rule computing the probability that the targets are more likely to be observed with each other than one of their known friends.

We fit the model to the empirical data (*Methods*) and found that multistep abstraction qualitatively reproduces the same pattern of results empirically observed in participants' behavior (Fig. 4*C*). Mechanistically, the model reveals that the likelihood of observing a target with one of their known friends depends exclusively on their total number of friends and is lower when the target has 12 rather than 6 friends (Fig. 4*D*). In contrast, the likelihood of observing the unknown friendship between targets steadily increases as the targets share a greater proportion of mutual friends (Fig. 4*D*). Therefore, the ratio of Successor Representation likelihoods for known and unknown friendships determines how people make inferences. As the proportion of mutual friends increases, the model increasingly infers that the targets are indeed friends (Fig. 4*C*). This effect is amplified when the targets have a greater number of friends, as this makes it relatively more likely that the targets will be observed together, compared to observing the targets with one of their numerous known friends. Given that inferences were based on minimal information about the underlying network, the use of abstraction to solve the link prediction problem is particularly notable in this study, and our results demonstrate just how flexibly the multistep abstraction mechanism can adapt to various formats of relational observation.

Study 4: Multistep Relational Abstraction Explains Representation of Real-World Networks. Our results have so far provided consistent and convergent evidence for multistep abstraction as a parsimonious, unified, and flexible mechanism underlying how people solve a variety of link prediction problems. They are limited, however, in that they have only used artificial networks to test mental representation in carefully controlled laboratory settings. These artificial networks may not reflect the complexities often found in real-world social networks, such as more nuanced grouping structure (19), or rich information about latent social properties like popularity and connectedness (8, 10, 41). To address this gap, we used a dataset previously published by Krackhardt (42) to test how well multistep abstraction explains representations of a real-world network composed of 21 managers at a technology manufacturing company (42). The managers were asked to report which other managers they gave advice to and/or got advice from, enabling us to establish a "ground truth" estimate of the network's true structure. Managers were also asked to report their perception of the other managers' advice-giving and -receiving.

Relative to the artificial networks we have considered so far, this network shows considerably less clustering into groups (Fig. 5*A*), which is mirrored in managers' perceptions. (Fig. 5*B*). A model that is only capable of generating groups or clusters would therefore struggle to explain how people are actually connected to one another within this particular network. To test our multistep abstraction hypothesis, we simulated Successor Representations at various levels of γ, then used mixed-effects logistic regression to predict participants' mental representations (i.e., using a nearly identical procedure as in Study 2; see *Methods*). At the group level, we found that $\gamma = 0.7$ was the best-fitting model (BIC = 8,780.27; Fig. 5*C*), with significant use of the Successor Representation at the group level ($\beta = 14.75$, $Z = 6.30$, 95% CI = [10.16, 19.33], $P < 0.001$), even after controlling for memorization ($\beta = 3.53$, $Z = 8.77$, 95% CI = [2.74, 4.32], $P < 0.001$). In contrast, the triad-based model was one of the worst-performing models in the set (BIC = 8,898.79), as was the community-detection model (BIC = 8,976.85), and both were outperformed by every Successor Representation model. Therefore, even when schema-like strategies fall short, our results show that multistep abstraction predicts people's network representations in real-world settings.

As noted by the original author (42), participants' mental representations additionally reveal a strange empirical phenomenon defying simple explanation. Among the most well-connected managers, there are discrepancies between a manager's true centrality (i.e., how often a manager is asked for advice) and his centrality in the minds of others. There are two informative test cases where a pair of managers are equally central in reality, but one is believed by his peers to be more central than the other (Fig. 5*B*). To our knowledge, there are no existing schemas that can explain how such network representations emerge. A stringent test of multistep abstraction, therefore, is to reproduce this phenomenon and explain why the most well-connected managers vary in their perceived centrality, despite having the same number of connections in reality. Using the best-fitting Successor Representation model from our previous analysis, we find that the Successor Representation does indeed reproduce these discrepancies (mean Spearman's $\rho = 0.43$, $P < 0.001$; Fig. 5*D*). In fact, the model is able to recapitulate individual managers' idiosyncratic beliefs, even when that manager's beliefs go against the majority (Fig. 5*E*). Using an exact binomial test as a prevalence test, we found that the model makes correct predictions for the majority of managers (18/21 for the 18 to 21 test pair, $P < 0.001$; 15/21 for the 1 to 7 pair, $P = 0.039$).

Mechanistically, this is because the Successor Representation represents individuals in terms of their extended relations (i.e., the weighted sum of their multistep connections), rather than their true degree centrality (i.e., number of one-step connections). To illustrate, imagine that Isa and Asher have the same number of friends. If Isa's friends have many friends themselves, and Asher's friends are all loners, then Isa will likely be perceived as being more popular. Or, if the network were configured such that Isa was the sole "bridge" connecting her friends, but Asher's friends would all stay connected without him, then Isa might again be perceived as being more popular. Accordingly, we found a strong association between "successor centrality" (i.e., the column sums of the Successor Representation matrix) and betweenness centrality (i.e., the number of shortest paths that pass through a network member, a proxy for brokerage; Spearman's $\rho = 0.89$, $P < 0.001$). This result is consistent with past work finding that network members often exaggerate the brokerage of the network's most popular individuals (19). Remarkably, it also suggests that people are able to accurately identify socially important individuals in their network, even when they lack full knowledge of people's relationships (8, 43), illustrating yet another application of multistep abstraction for solving the link prediction problem.

The model's success is notable given that the model was not privy to information about the company's leadership hierarchy. Despite this, the model correctly identified 4/5 of the senior leadership: the Chief Executive Officer (node 7 in Fig. 5*A*) and most of the Vice

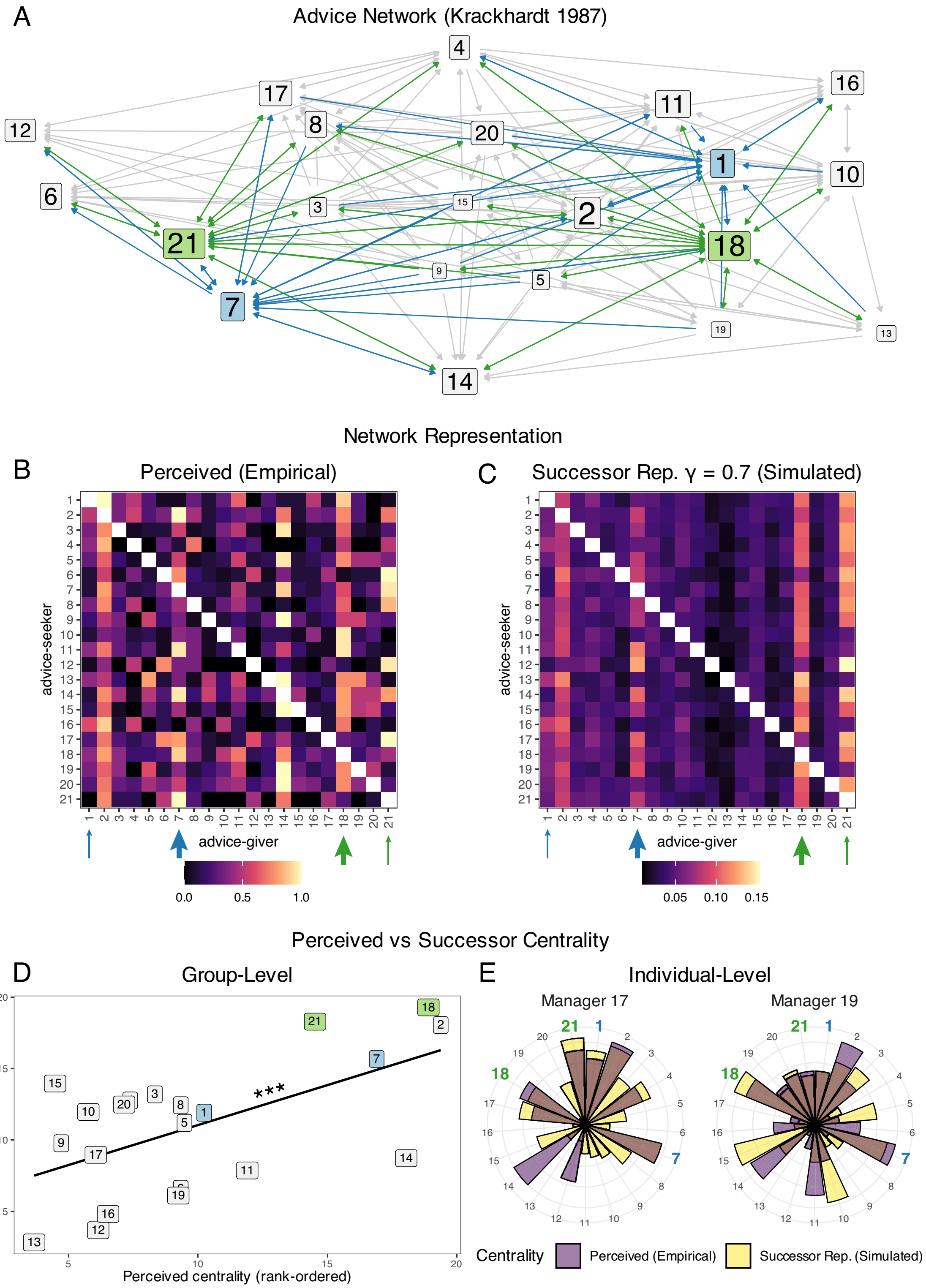

Fig. 5. Study 4: Reanalysis of Krackhardt (42). (*A*) The true advice-giving network among 21 managers in a company. The size of each node corresponds to in-degree centrality (i.e., the number of network members who seek advice from this manager). Managers 1 and 7 (blue) are both approached for advice by 13 other managers, and managers 18 and 21 (green) are both approached for advice by 15 other managers. (*B*) Subjective representation of the advice-giving network, averaged across all managers. A heatmap value of zero in a given cell means that no managers believed those two individuals share advice, and a value of one means that all managers believed those individuals share advice. Managers 1 to 7 and 18 to 21 have the same in-degree centrality. Yet, in the minds of their peers, one manager from each node-pair is represented as being better connected than the other. The color-coded arrows correspond to the two node-pairs of interest, and the individual perceived to be more central is indicated by a thicker arrow. (*C*) Multistep relational abstraction can reproduce managers' representation of the advice-giving network. (*D*) The Successor Representation rank-correlates with perceptions of peers' centrality ($P < 0.001$). It is also able to accurately recapitulate the fact that 18 is perceived as being more central than 21 and that 7 is perceived to be more central than 1. (*E*) Each number on the "rose plot" corresponds to a manager, and the length of each "petal" corresponds to rank-ordered centrality. If a petal is mostly purple, the model has underestimated how central a manager is in the mind of his peer; if a petal is mostly yellow, the model has overestimated a manager's perceived centrality. Greater overlap, in brown, indicates that the model makes accurate predictions. The Successor Representation predicts individual managers' perceptions; we visualize data from two representative participants here. Notably, manager 17 diverges from the group-level trend, perceiving manager 21 as being more central than 18; the model correctly recapitulates this belief.

Presidents (nodes 2, 18, and 21). However, in the model's predictions, one of the Vice Presidents (node 14) is conspicuously—and incorrectly—predicted to have low centrality in the minds of his peers. The company's hierarchical structure likely influences the number of times a manager observes advice-giving, with senior leadership likely to be overrepresented. Therefore, it should be possible to improve the model's predictions by providing it with additional observations of senior leadership providing advice. To test this, we simply trained the Successor Representation on the same set of observations as before, but with twice as many observations of senior leadership. Results indicate that this simple change was sufficient to improve the model's predictions: We observed a greatly improved model fit in the mixed-effects regression (BIC = 8,657.39), increased rank correlation between successor and perceived centrality (Spearman's $\rho = 0.51$, $P < 0.001$; *SI Appendix*, Fig. S10), and more accurate predictions of how individual participants perceived manager 14's centrality (*SI Appendix*, Fig. S11). This analysis therefore suggests that an individual's mental representation of their social network is powerfully shaped by their observations of social interaction, which in turn are constrained by where that individual is positioned in the network.

Discussion

To make adaptive social decisions like whom to trust or gossip with, people must build useful cognitive maps of their social networks. But even for a relatively small network of 30 individuals, the number of possible network configurations is many orders of magnitude larger than the number of atoms in the universe (44). Despite the vastness of this space, we must infer who is connected to whom based on noisy and incomplete observations. This is akin to guessing a 65-character password from occasional glances at the keyboard as it is typed in. How do we solve this fundamental and difficult computational problem of social link prediction? Past theory has suggested that people learn a collection of schemas for common network structures such as triads and communities and then apply them in a heuristic manner (6, 20, 22). We offer an alternate proposal explaining how social link prediction naturally emerges from cognitive maps that are built using a single, parsimonious mechanism of multistep relational abstraction. Our proposal accounts for a number of puzzling empirical findings about social network representations and social link prediction and places them within a single principled and mechanistic framework.

If people leverage their observations to drive both learning of one-step relations and abstracted multistep relations, humans can build the kinds of predictive representations required to successfully navigate complex social networks. We demonstrate how multistep relational abstraction explains link prediction across a variety of social contexts and learning scenarios. Specifically, we find that schema-like mental representation of social networks naturally emerges from applying different degrees of multistep abstraction to learning (Study 1); multistep abstraction explains the structure of people's mental representations and their social behaviors (Study 2); multistep abstraction can inform predictive inferences even from exceedingly sparse information about a novel network (Study 3); and even in a real-world workplace network, multistep abstraction is able to explain how and why people's network representations are systematically biased (Study 4). The same model is able to account for participants' memories, judgments, and inferences across all of these contexts, simply by integrating over a different number of steps in each context.

Our work is proximally inspired by research on Cognitive Social Structures (6, 20, 42), which has carefully documented systematic biases in how people represent social networks. Schemas have played an especially large role in Cognitive Social Structures theory, as they marry descriptions of empirical phenomena with cognitive-level explanations of why these phenomena exist. By compressing relational knowledge into compact structures, schemas are thought to aid the efficient learning and storage of social relationships, despite the combinatorial explosion of possible relations within any nontrivial network (10, 21, 22). As a theory of human cognition, schema-based accounts are valuable for their qualitative descriptions of mental representation. We note that multistep abstraction can also be interpreted and used as a descriptive model; we do so, for example, for the real-world network of managers, where we model the outcome of social network learning rather than the process. Even in its capacity of a descriptive model, the present work provides theoretical value by providing a parsimonious yet flexible mechanism for characterizing the content of mental representation and specifying how existing schemas can be quantitatively described along a spectrum of abstraction. In doing so, multistep abstraction not only unifies previously proposed schematic structures like triads and communities but also provides a framework for generating cognitive structures that have not yet been described by existing schemas.

Multistep abstraction provides even greater theoretical value when used as a computationally grounded theory about the nature of social cognitive maps, their function, and how they are mechanistically learned. The Successor Representation can be built with a biologically plausible temporal-difference learning mechanism detailing how disparate observations can be systematically abstracted into coherent mental representations based on experience. This provides a theory explaining how schema-like representations can be built from direct observation of social relationships. Rather than being endowed with a set of discrete and inflexible heuristics, we propose that humans are furnished with a learning mechanism that discovers the generalizable structural features of the social networks they encounter. This allows them to adapt to a variety of social environments with different relational structures, such as friendship groups, kinship groups, or hierarchical organizations like workplaces. We also provide a theoretical perspective explaining why the brain might systematically encode biases in network representation: At the cost of accurate knowledge, we gain the ability to solve the link prediction problem across a variety of contexts where social inference is required for adaptive decision-making. Given evidence in computer science that multistep abstraction is an effective solution for solving a variety of link prediction problems, including in social networks (28), our theory argues that multistep abstraction is a rational strategy for predicting unknown and unobserved social relationships.

Our theory proposes that multistep relational abstraction enables people to build abstract cognitive maps of their social networks. While we used the Successor Representation to implement multistep abstraction in this work (32–34, 45, 46), we note that we are not committed to the view that reinforcement learning is the only (or even dominant) method through which multistep abstraction is realized in the mind or brain (31). The normalized Successor Representation is mathematically identical to a maximal entropy model of memory abstraction (36), and neural network models suggest that the Successor Representation may share computational similarities with other association-chaining phenomena such as transitive inference and acquired equivalence (47). Interestingly, the Successor Representation's mathematical form has deep structural similarity with a host of network centrality measures that quantify psychological qualities such as influence, power, brokerage, and the ability to spread information widely (8, 43, 48–50). Most notably, the computations required to compute Katz centrality are nearly identical to the computations required to compute the Successor Representation (48, 49), and variants of Katz centrality have proven useful for link prediction in computer science (28). However, these computations require an individual to have complete knowledge of all connections within the network, an unrealistic assumption in most nontrivial social networks. In contrast, the Successor Representation can instead be computed through temporal-difference learning, allowing an individual to approximate multistep relations asymptotically to Katz centrality, despite lacking complete knowledge of the social network. This may explain, for example, the curious finding that people are highly accurate at identifying structurally important individuals in the network, despite seeming to lack detailed knowledge about how people are connected to each other in the network at large (8, 17).

The present work establishes a foundation for a general theory of social network representation and link prediction, but several important limitations and open questions remain for future work to address. For example, the Successor Representation performs less well for the real-world social network, compared to the laboratory experiments. Although real-world network studies provide ecologically valid settings, a major limitation of any such study is that there are many phenomena influencing participants' cognitive maps and only a few that researchers might measure or know about. For example, Study 4 examines managers' perceptions of their real-world advice-giving network, which almost certainly incorporates other sources of knowledge, such as informal friendships. Those additional sources of information are not provided to the model, and so it is perhaps unsurprising that the Successor Representation would be less effective in capturing managers' cognitive maps. If the model were given more observations of managers' day-to-day social interactions (e.g., whether two managers played tennis together after work), or any other significant contributors, it is possible that the model's predictions would improve. We found evidence of this, for example, when providing extra observations corresponding to the company's leadership hierarchy. However, even despite lacking all of this rich knowledge about managers' social relationships, we still find that the Successor Representation is a better fit than other candidate models that are commonly relied upon and can even explain why managers' cognitive maps reflect multistep rather than true centrality.

Real-world social networks also pose a variety of unique computational challenges for learning about social networks, and it is not yet clear how the mind solves these problems. One such challenge is that observations of real-world relationships can be separated far apart in time. Our simulation study suggests that the temporal-difference learning used to compute the Successor Representation is equivalent to slowly learning structure from piecemeal observation of relationships, such that acquiring information over longer time intervals does not fundamentally alter the computational demands of network learning. However, additional factors like memory decay or forgetting may have a stronger impact on the fidelity of the learned Successor Representation in real-world settings. Another unique challenge is that the composition of a social network is often dynamic. We assume in this work that while relationships may change in a network, the set of individuals comprising the network (i.e., the state space) do not. However, real-world networks are constantly evolving, and new individuals joining a social network alter the state space itself. This suggests that an additional mechanism is needed for updating the state space in real-world settings that are continually in flux. Whether these additional mechanisms interact with learning Successor Representations, or else operate independently, remains an open question.

It also remains unknown how consciously aware people are of the knowledge structures they build through network learning, which has implications for social psychological theories of bias. In nonsocial contexts, past research suggests that the SR is used to build knowledge structures that are more implicit than explicit (45). It is also commonly argued that much of what people "know" about the social world lies outside the boundaries of consciousness (51). People may, however, still have conscious awareness of a particularly simple schematic or stereotyped knowledge, such as those involved in intergroup cognition. A fruitful future direction may be to formally connect multistep abstraction with other kinds of learning abstractions. For example, instead of learning about multistep individual-to-individual relationships, people could learn that certain social features are indicative of group membership and could leverage this to learn intergroup relations (24).

It is likely that the usefulness and simplicity of the Successor Representation computation makes it an attractive solution to many social problems the brain must solve. How exactly this computation might be instantiated in the brain is presently unclear, but some neural network models of hippocampal learning have converged upon comparable solutions for generalizing beyond limited direct experience (47, 52–55). Past work has demonstrated that representation of both spatial and nonspatial cognitive maps in the hippocampal formation follow similar principles, hinting at the possibility that abstracted social network representations may also be computed or encoded there (39, 52, 56–58). More generally, knowledge of network members' centrality appears sufficiently useful that the brain seems to spontaneously track this information (10, 41), and it is possible that the underlying neural computations depend on some form of multistep abstraction.

Our work helps extend the general theory of cognitive maps and how they are used for inference. While recent theoretical and empirical research has broadened the framework of cognitive maps beyond the traditional domain of physical space (56), this work has largely focused on the "path-finding" problem in fixed and immutable environments. In contrast, a link prediction perspective invites investigation of how cognitive maps are used to solve "path-forging" problems. By analogy, while we understand a great deal about how rats navigate through complex laboratory mazes, we know much less about how rats create their own complex underground labyrinths from a near-infinite set of prospective pathways branching from existing paths. The link prediction capacities of cognitive maps provide a promising framework for understanding such behavior.

Methods

Participants. In Study 2, the Random Walks sample consists of $N = 60$ participants. These data were originally collected to pilot unrelated experiments and were reanalyzed for this work. Although the sample size was unplanned, we note that it is comparable to previous empirical work on the Successor Representation (45). Participants were either recruited from Brown University and surrounding community ($N = 41$) or from the Prolific online labor market ($N = 19$). No demographic information about race, gender, or age was collected. In the Paired Associates sample, we recruited $N = 30$ participants but lost two data points due to experimenter error. The reported dataset therefore consists of $N = 28$ participants (15 female, 16 white, 9 Asian, 3 Hispanic/Latinx, 3 mixed/other; mean age = 22.29 y old, age range 18 to 31 y old; self-identified racial groups do not sum to the total number of participants because participants could indicate multiple identities). All participants received monetary compensation ($15 in-lab, $13 online) and could earn additional monetary bonuses based on learning accuracy. All study procedures were conducted in a manner approved by the Brown University Institutional Review Board, and all participants provided informed consent. Details about the participants from Studies 3 to 4 can be found in the original papers (40, 42).

Procedure. In Study 2, all stimuli were drawn from the Chicago Face Database (59) and were randomly assigned to nodes in the network. Therefore, while the underlying network structure was the same for all participants, "node 1" referred to different stimuli for different participants.

In the Random Walks sample, participants learned about friendships by viewing sequential steps of a random walk through the network, such that seeing two faces in a row meant that they were friends with each other. Each step of the walk involved the presentation of the face associated with each node, which was displayed for 1.5 s with no intertrial interval. In each "run" of the learning task, participants viewed 280 steps of a random walk, took a short break, and then started over with a new random walk. Fifty-three participants (recruited both in-lab and online) observed 837 steps of a random walk through the network during the learning task (i.e., three runs), and seven participants (all in-lab) observed 1,674 steps during the learning task (i.e., six runs). To ensure that participants paid attention to the learning task, participants were required to press a button whenever they saw an upside-down face (12.5% of all trials). The memory task consisted of 78 trials (i.e., getting a single measure for the $i \rightarrow j$ relation).

In the Paired Associates sample, participants learned about the network by observing two network members appear side-by-side on the same trial if they were friends. No information about multistep relations was ever presented. In each run of the task, each friend pair was presented for 2.5 s each, and every participant completed 10 runs. Since there are 17 friendships in the artificial network, the learning task consisted of a total of 170 trials. The memory task procedure was identical, except that we doubled the number of trials (i.e., measured both $i \rightarrow j$ and $j \rightarrow i$ relations). Participants then completed a trust inference task, also consisting of 156 trials.

Detailed descriptions of Studies 3 and 4 can be found in the original papers (40, 42); here, we summarize particularly relevant parts of the procedures. In Study 3, participants were shown a schematic of a social network on every trial (Fig. 4*A*). This schematic provided a visual summary of the following pieces of information: 1) the total number of people in the social network, which was always 26; 2) how many friends each target individual had, constrained such that both targets always had the same number of friends; and 3) how many of those friends were mutually shared between the targets. The schematic made no indication of other friendships that may or may not have existed within the network. From this minimal summary of the network, participants were required to judge how likely it was that the targets were themselves friends.

In Study 4, each participant was given a questionnaire asking which managers asked each other for advice. An example prompt, provided in the original paper, reads: "Who would Steve Boise go to for help or advice at work?" Under this prompt appeared a list of all other managers' names, and participants indicated who they thought Steve Boise sought advice from. They then repeated this procedure for all other managers, including themselves.

Successor Representation. For a network of N network members, the Successor Representation is an $N \times N$ matrix M such that M_{ij} encodes the likelihood of observing network member j, given that we have observed i. Importantly, M_{ij} reflects not only the likelihood of observing the one-step $i \rightarrow j$ transition but also the multistep likelihood of starting from i and eventually observing j through intermediary network members (e.g., $i \rightarrow k \rightarrow j$). In the context of social networks, a useful heuristic is that i will generally be represented as having a greater likelihood of being connected to j to the extent that they have many mutual friends. The Successor Representation is updated using standard reinforcement learning methods (33, 34, 46), with a few minor modifications (Eq. **1**). Observations of friendship are encoded as a one-hot vector $\mathbf{1}(j)$, i.e., a vector of length N filled with zeroes except at the index j. This one-hot vector is used to drive a row-wise update of M_i, tempered by the learning rate α. The successor discount parameter $\gamma \in [0, 1)$ then causes i to become represented more similarly to its successor j by adding a fractional amount of M_j to M_i. It is this successor term $\gamma > 0$ that enables the Successor Representation to integrate over many one-step transitions, thereby encoding multistep relational abstractions.

$$M_i \leftarrow M_i + \alpha\delta,$$
$$\delta = \mathbf{1}(j) + \gamma M_j - M_i. \quad [1]$$

The reinforcement learning implementation was used to model all studies except Study 3. In that study, we used an analytic method to compute the asymptotic Successor Representation (46), which only relies on the identity matrix I and the network's true transition matrix T (Eq. **2**). Since Study 3 tested inference after a single learning opportunity, rather than through trial-and-error learning, the analytic method provides a more natural fit for how people might infer relations in that context. We note that participants in the original study were provided minimal summary information about the network through a schematic of the structure of the network (Fig. 4*A*), which includes detailed information about the network's true adjacency matrix. For our analysis, we assume that this knowledge was available to the model just as it was to the participants, from which it could compute the Successor Representation (Eq. **2**). We note that while the summary information provided in this task does not specify all the individual links identified in the network, it provides sufficient information to constrain the set of possible graphs to a small set of isomorphic graphs that correspond to a single abstract network structure.

$$M = (I - \gamma T)^{-1}. \quad [2]$$

In both implementations, the raw values in the Successor Representation matrix reflect the expected number of times a given network member will be observed in a random walk of length $L = \frac{1}{1-\gamma}$. To normalize these values into transition probabilities, we therefore scalar multiply the matrix by $1 - \gamma$. We also note that L can equivalently be interpreted as a "lookahead" horizon (i.e., the number of steps that a person integrates over), such that the lookahead increases exponentially as γ approaches its upper limit. Due to its exponential scale, seemingly small changes in γ can therefore produce large changes in the lookahead horizon. For this reason, when testing whether the Successor Representation is integrating over a different number of steps across samples, it is misleading to simply compare the mean values of γ; it is more appropriate to convert the model's estimated γ parameters into their corresponding lookahead horizons.

Parameter Estimation. To translate the Successor Representation matrix into probabilistic judgments of friendship, the model estimates weights that amplify or attenuate the predictions made by the Successor Representation. Conceptually, this is exactly equal to estimating a softmax temperature parameter or a beta weight in logistic regression. We assume that an agent could in principle be using an arbitrary number of representational strategies and therefore allow the agent to employ a mixture-of-experts scheme to determine its overall network representation. We model this using a logistic choice rule (Eq. **3**) that takes a linear combination of strategies and weights as input (Eq. **4**).

$$f(x) = \frac{1}{1 + \exp(-x)}, \quad [3]$$

$$p(\text{relation}) = f\left(\beta_0 + \beta_1 SR + \cdots\right). \quad [4]$$

In Study 2, the model allowed γ to take on any of its theoretically defined values, $\gamma \in [0, 1)$. Free parameters for the weights and γ were fit using the Nelder–Mead algorithm implemented in R's default optimizer, using logistic loss to calculate model likelihoods. To allow the softmax to fully saturate, we multiplied all weights by an arbitrary scaling constant of 10 (i.e., the raw values in the Successor Representation matrix are transition probabilities, and therefore very small on an absolute scale). Mixture-of-experts weights were regularized using the L_2 norm and $\sigma = 5$ such that the likelihood additionally added the term $\sum_i \lambda w_i^2$, where $\lambda = \frac{1}{\sigma^2}$. Each participant's data were fit independently, and each model was reestimated 25 times, keeping only the estimates that best maximized the likelihood.

In Study 3, the large number of participants made it prohibitive to fit a model to every participant's data. Instead, we fit a single model to the group-averaged data. Since participants originally responded to the task using Likert scale ratings, we created a synthetic dataset converting these ratings into binary choices. We rescaled the ratings from the original [−3, 3] scale so that they fell in the range [0, 1]. This gave us an approximate percentage of trials on which participants would have responded "yes" in a binary choice task when inferring whether the two network members were friends with each other. We simulated 100 trials per condition accordingly and then used the logistic choice rule (Eq. **3**) to fit the value of γ best explaining the data. Here, instead of modeling choices as a linear combination (i.e., including an intercept), we fit a single temperature parameter τ to temper how strongly the Successor Representation would influence choice. In the best-fitting set of parameters, we found that $\tau \approx 0.03$.

In Study 4, we used mixed-effects logistic regression models to test how well various fixed values of γ predicted participants' mental representations, rather than fitting a model to each participant's data. Since participants were responding about a real-world network, in which they were deeply embedded, we lacked sufficient knowledge of what observations may have been available to each participant. Since mixed-effects models are specifically designed to extract commonalities between participants, we reasoned that this might be a more robust method for approximating what value of γ might best explain group-level representation. We therefore simulated Successor Representation with $\gamma \in [0.1, 1)$ in increments of 0.1 and used these as predictors (i.e., using the same procedure as the statistical models in Study 2). All regression models additionally included a predictor for pure memorization, and all models included random intercepts and random slopes for the Successor Representation.

Since participants in Study 4 were embedded in the social fabric of the company, we needed to make some assumptions about what social interactions each manager was generally able to observe. We made the trivial assumption that people observe interactions between themselves and their immediate advisors (i.e., one-step relations). We additionally assumed that people are generally able to observe interactions between advisors-of-advisors (i.e., two-step relations), either through firsthand observation or secondhand chatter/gossip (8). Although it is of course plausible that some people may be able to observe more distant interactions, we assumed that this happens infrequently enough that excluding these interactions would not unduly bias the model's predictions; we note that this assumption is consistent with past research on large-scale real-world networks (17). Finally, we assumed that it was equally likely that participants would observe one-step and two-step relations. This final assumption is unlikely to be true of real-world interactions, but our goal was to avoid biasing the model predictions in favor of our hypothesis by allowing the model to learn more about one-step relations than two-step relations. We then used these "observations" to generate predicted Successor Representations, just as we did in the laboratory studies where participants' learning observations were truly known. The rest of the Study 4 mixed-effects modeling procedure is near-identical to the mixed-effects modeling procedure used in Study 2. We note that other plausible assumptions could be made about what social interactions managers are able to observe. We tested alternative models using those assumptions and reported them in *SI Appendix*.

Data, Materials, and Software Availability. All data and code needed to reproduce the analyses are available in a publicly accessible GitHub repository: https://github.com/feldmanhalllab/multistep-relational-abstraction. Previously published data were used for this work (40, 42).

ACKNOWLEDGMENTS. We thank Marc-Lluis Vives for providing the Random Walks dataset and thank Isabella Aslarus, Elizabeth Duchan, Mehak Malhotra, and Maya Mazumder for helping collect the Paired Associates dataset. Part of this research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. Advanced access to these computing resources was supported by NIH award 1S10OD025181. This material is based upon work and personnel supported by the NSF Graduate Research Fellowship award 2040433 (J.Y.S.), the NIH Health award R01 MH125497 (A.B.), the Carney Innovation Grant from the Robert J. and Nancy D. Carney Institute for Brain Science (O.F.H.), and the NSF award 2123469 (O.F.H. and A.B.).

1. P. Bourdieu, "The forms of capital" in *Handbook of Theory and Research for the Sociology of Education*, J. Richardson, Ed. Greenwood, Westport, CT, 1985), chap. 1, pp. 241–258.
2. J. S. Coleman, Social capital in the creation of human capital. *Am. J. Sociol.* **94**, S95–S120 (1988).
3. M. Granovetter, Economic action and social structure: The problem of embeddedness. *Am. J. Sociol.* **91**, 481–510 (1985).
4. K. Rajkumar, G. Saint-Jacques, I. Bojinov, E. Brynjolfsson, S. Aral, A causal test of the strength of weak ties. *Science* **377**, 1304–1310 (2022).
5. Z. B. Cullen, R. Perez-Truglia, The old boys' club: Schmoozing and the gender gap. *National Bureau of Economic Research Working Paper Series* No. 26530 (2019).
6. R. A. Brands, Cognitive social structures in social network research: A review. *J. Org. Behav.* **34**, S82–S103 (2013).
7. R. S. Burt, D. Ronchi, Teaching executives to see social capital: Results from a field experiment. *Soc. Sci. Res.* **36**, 1156–1183 (2007).
8. A. Banerjee, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, Using gossips to spread information: Theory and evidence from two randomized controlled trials. *Rev. Econ. Studies* **86**, 2453–2490 (2019).
9. 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**, 8149–8154 (2018).
10. R. Basyouni, C. Parkinson, Mapping the social landscape: Tracking patterns of interpersonal relationships. *Trends Cogn. Sci.* **26**, 204–221 (2022), 10.1016/j.tics.2021.12.006.
11. D. Krackhardt, Assessing the political landscape: Structure, cognition, and power in organizations. *Admin. Sci. Q.* **35**, 342–369 (1990).
12. H. R. Bernard, P. D. Killworth, L. Sailer, Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Soc. Sci. Res.* **11**, 30–66 (1982).
13. L. C. Freeman, A. K. Romney, S. C. Freeman, Cognitive structure and informant accuracy. *Am. Anthropol.* **89**, 310–325 (1987).
14. 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).
15. L. C. Freeman, C. M. Webster, Interpersonal proximity in social and cognitive space. *Soc. Cogn.* **12**, 223–247 (1994).
16. T. Casciaro, Seeing things clearly: Social structure, personality, and accuracy in social network perception. *Social Netw.* **20**, 331–351 (1998).
17. E. Breza, A. G. Chandrasekhar, A. Tahbaz-Salehi, Seeing the forest for the trees? An investigation of network knowledge. *National Bureau of Economic Research Working Paper Series* No. 24359 (2018).
18. G. A. Janicik, R. P. Larrick, Social network schemas and the learning of incomplete networks. *J. Pers. Soc. Psychol.* **88**, 348–364 (2005), 10.1037/0022-3514.88.2.348.
19. M. Kilduff, C. Crossland, W. Tsai, D. Krackhardt, Organizational network perceptions versus reality: A small world after all? *Org. Behav. Hum. Decis. Process.* **107**, 15–28 (2008).
20. E. B. Smith, R. A. Brands, M. E. Brashears, A. M. Kleinbaum, Social networks and cognition. *Annu. Rev. Sociol.* **46**, 159–174 (2020).
21. M. E. Brashears, E. Quintane, The microstructures of network recall: How social networks are encoded and represented in human memory. *Social Netw.* **41**, 113–126 (2015).
22. M. E. Brashears, Humans use compression heuristics to improve the recall of social networks. *Sci. Rep.* **3**, 1513 (2013).
23. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction. *Science* **331**, 1279 (2011).
24. J.-Y. Son, A. Bhandari, O. FeldmanHall, Cognitive maps of social features enable flexible inference in social networks. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021699118 (2021).
25. C. Tilly, *The Politics of Collective Violence* (Cambridge University Press, 2003).
26. W. Berenschot, Rioting as maintaining relations: Hindu-muslim violence and political mediation in Gujarat, India. *Civil Wars* **11**, 414–433 (2009).
27. D. Yanagizawa-Drott, Propaganda and conflict: Evidence from the Rwandan genocide*. *Q J. Econ.* **129**, 1947–1994 (2014).
28. D. Liben-Nowell, J. Kleinberg, "The link prediction problem for social networks" in *Proceedings of the Twelfth International Conference on Information and Knowledge Management* (Association for Computing Machinery, New Orleans, LA, USA, 2003), pp. 556–559.
29. A. Ghasemian, H. Hosseinmardi, A. Galstyan, E. M. Airoldi, A. Clauset, Stacking models for nearly optimal link prediction in complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23393–23400 (2020).
30. M. K. Ho, D. Abel, T. L. Griffiths, M. L. Littman, The value of abstraction. *Curr. Opin. Behav. Sci.* **29**, 111–116 (2019).
31. C. W. Lynn, D. S. Bassett, How humans learn and represent networks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29407 (2020).
32. I. Momennejad, Learning structures: Predictive representations, replay, and generalization. *Curr. Opin. Behav. Sci.* **32**, 155–166 (2020).
33. P. Dayan, Improving generalization for temporal difference learning: The successor representation. *Neural Comput.* **5**, 613–624 (1993).
34. 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).
35. E. D. Boorman, S. C. Sweigart, S. A. Park, Cognitive maps and novel inferences: A flexibility hierarchy. *Curr. Opin. Behav. Sci.* **38**, 141–149 (2021).
36. C. W. Lynn, A. E. Kahn, N. Nyema, D. S. Bassett, Abstract representations of events arise from mental errors in learning and memory. *Nat. Commun.* **11**, 2313 (2020).
37. A. C. Schapiro, T. T. Rogers, N. I. Cordova, N. B. Turk-Browne, M. M. Botvinick, Neural representations of events arise from temporal community structure. *Nat. Neurosci.* **16**, 486–492 (2013).
38. A. Pudhiyidath *et al.*, Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *J. Cogn. Neurosci.* **34**, 1736–1760 (2022), 10.1162/jocn_a_01864.
39. M. M. Garvert, R. J. Dolan, T. E. J. Behrens, A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* **6**, e17086 (2017).
40. C. G. Sehl, O. Friedman, S. Denison, The social network: How people infer relationships from mutual connections. *J. Exp. Psychol. Gen.* **152**, 925–934 (2022), 10.1037/xge0001330.
41. C. Parkinson, A. M. Kleinbaum, T. Wheatley, Spontaneous neural encoding of social network position. *Nat. Hum. Behav.* **1**, 0072 (2017).
42. D. Krackhardt, Cognitive social structures. *Social Netw.* **9**, 109–134 (1987).
43. A. Banerjee, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, The diffusion of microfinance. *Science* **341**, 6144 (2013).
44. M. O. Jackson, *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors* (Pantheon, 2019).
45. I. Momennejad *et al.*, The successor representation in human reinforcement learning. *Nat. Hum. Behav.* **1**, 680–692 (2017).
46. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
47. D. Kumaran, J. L. McClelland, Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychol. Rev.* **119**, 573–616 (2012).
48. L. Katz, A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953).
49. P. Bonacich, Power and centrality: A family of measures. *Am. J. Sociol.* **92**, 1170–1182 (1987).
50. L. C. Freeman, A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
51. A. G. Greenwald, M. R. Banaji, The implicit revolution: Reconceiving the relation between conscious and unconscious. *Am. Psychol.* **72**, 861–871 (2017).

52. 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.e1223 (2020).
53. C. Fang, D. Aronov, L. F. Abbott, E. L. Mackevicius, Neural learning rules for generating flexible predictions and computing the successor representation. *eLife* **12**, e80680 (2023).
54. T. M. George, W. de Cothi, K. L. Stachenfeld, C. Barry, Rapid learning of predictive maps with STDP and theta phase precession. *eLife* **12**, e80663 (2023).
55. J. Bono, S. Zannone, V. Pedrosa, C. Clopath, Learning predictive cognitive maps with spiking neurons during behavior and replays. *eLife* **12**, e80671 (2023).
56. T. E. J. Behrens *et al.*, What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
57. 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).
58. 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.e1228 (2020).
59. 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).