Why meaning matters for belief diffusion in social networks

By

James P. Houghton
B.S. Aerospace Engineering
Massachusetts Institute of Technology, 2004

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2019

Signature of Author: Signature redacted
Department of Management
July 16, 2019

Certified by: Signature redacted
Hazhir Rahmandad
Associate Professor of System Dynamics
Thesis Supervisor

Accepted by: Signature redacted
Catherine Tucker
Sloan Distinguished Professor of Management Science
Faculty Chair, MIT Sloan PhD Program

Why meaning matters for belief diffusion in social networks

By James P. Houghton

Submitted to the MIT Sloan School of Management on July 16, 2019 in partial fulfillment of the requirements for the degree of Master of Science in Management Science.

ABSTRACT

It is well known that human beings preferentially adopt beliefs that are consistent with what they already know *(1)*. At its best, this process helps knowledge cumulate, and at its worst facilitates motivated reasoning and pseudoscience. Recent research in social contagion shows that the tendency to treat new information in light of what is already known creates *interdependence* in the diffusion patterns of simultaneously diffusing beliefs, and that this is sufficient to generate societal polarization and competing worldviews *(2-8)*. This paper explains the mechanisms by which interdependence between beliefs can lead to fundamentally different patterns of adoption than would have occurred under traditional assumptions of independent diffusion. First, when beliefs facilitate one another's adoption, they spread to more individuals than any could have reached spreading on its own. Secondly, as individuals become more alike, they increase their likelihood of exchanging beliefs in the future, and of forming around themselves a faction of likeminded peers. These mechanisms explain why the most popular beliefs tend to be related to one another, and how polarization may spontaneously emerge in homogeneous and well-connected populations. Simulations in this paper make a direct comparison between interdependent and independent diffusion, explaining why the mechanisms of interdependent diffusion reverse many predictions of standard (independent) diffusion models. For example, while independently diffusing beliefs can make a population more homogenous, interdependent diffusion leads the same population to polarize. While the most successful independent beliefs are those with central network positions, interdependent beliefs become popular by facilitating the diffusion of related beliefs.

Thesis Supervisor:

Hazhir Rahmandad
Associate Professor of System Dynamics

Beliefs about the world (i.e. ideas and information) spread from person to person through a process of social contagion. A large body of research has documented how personal relationships and social network structures lead some beliefs to diffuse widely while others receive limited attention *(9-14)*. More recent research explores how relationships between beliefs can affect their diffusion *(2-8)*. This literature is motivated by the idea that individuals are more likely to adopt beliefs that are consistent with those they have acquired in the past. By adopting one belief, individuals can become susceptible to adopting another related belief (and vice versa) such that the processes of each belief's diffusion through the population become interdependent. Simulation models indicate that interdependent diffusion can generate previously unrecognized consequences of social contagion, for example the spontaneous polarization of opinions in political discourse *(2)*, or the emergence of political issue alignment *(3,8)*.

While the emerging focus on belief interaction has expanded our understanding of the potential outcomes of social contagion, room for theoretical clarity remains. In particular, the literature on interdependent diffusion has yet to articulate the mechanisms by which individuals' preferences for consistency aggregate to cause outcomes at the scale of society. Instead, many of these works test the consequences of an assumed belief structure on the shape of diffusion *(2, 5, 6)*. These models do not address how such belief structures come to exist, or to be shared by members of the population (itself a diffusion process). In these models, polarization and heterogeneity of belief adoption are strongly dependent upon the assumed belief structure, and so the effect of belief interactions in the diffusion process is difficult to assess *(15)*. Other models demonstrate how certain collections of beliefs can become popular when individuals imitate similar neighbors, or intentionally distinguish themselves from an out-group *(3, 8)*. However, in these models, beliefs interact primarily through their role as attributes the population uses to sort itself into groups, rather than through the adopter's assessment of their internal consistency. As a result, none of these works describe the behavioral mechanisms by which belief interaction shapes social contagion, and do so in a way that is clear enough to test empirically.

In what follows, I illustrate two mechanisms by which a preference for internal consistency leads to polarization and the emergence of competing worldviews. First, "reciprocal facilitation" allows sets of closely related beliefs to snowball through a population, while other equally-plausible sets languish. Secondly, "agreement cascades" make individuals who share beliefs more likely to exchange beliefs in the future, such that diffusion tends to amplify within-faction similarity more rapidly than across-faction similarity.

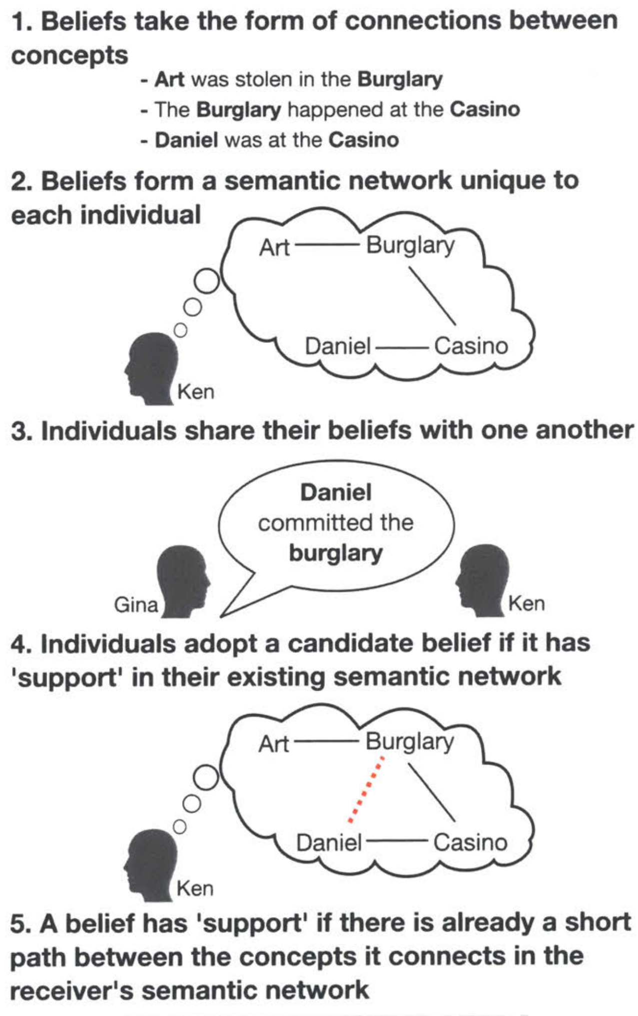
I use a simplified model of belief interaction, with no exogenous assumptions about belief compatibility, to reveal every link in the ladder of aggregation between belief interaction within an individual to outcomes at the scale of society. Step-by-step examples reveal how an individual's decision-making process shapes the likelihood of exchange with particular social network neighbors, and the likelihood that particular beliefs will be exchanged. Simulations then explain how these patterns of interpersonal exchange create heterogeneity in the rate at which particular beliefs spread, and the rate at which particular pairs of individuals become similar to one another. I then show how heterogeneous rates of diffusion and increasing similarity shape polarization and the emergence of shared worldviews. I directly compare the case of interdependent diffusion with what would have been expected had the same beliefs diffused independently of one another, in order to highlight how the mechanisms of interdependent diffusion modify what we expect from traditional models of independent diffusion. I explore consequences arising first from the interaction between beliefs as they spread, and secondly from the interaction between individuals as they share beliefs.

Results:

To focus this paper, I look specifically at beliefs *about* the world, rather than beliefs *in* a value or ideal, as illustrated in Fig. 1. I represent an individual's beliefs using a "semantic network" model borrowed from the cognitive science literature (*16-18*), and consistent with other formal models of belief formation *(19)*. Nodes in a semantic network represent concepts such as people, locations or activities. A semantic network edge represents the belief that two concepts are connected in some way. This formalization allows beliefs to interact (e.g. beliefs about a place interact via their connection to the node representing that place) without having to pre-specify which beliefs are compatible with one another (e.g. that belief i is compatible with j, but not k). Because we don't need to assume exogenous compatibility relationships, belief structures emerge naturally through the diffusion process, and any systematic variation in the popularity of beliefs can be attributed to the diffusion process rather than to the choice of belief structure (*15)*.

When an individual is exposed to a new belief by her neighbor in the social network, she decides whether or not to adopt it by seeing how it relates to beliefs in her existing semantic network. She is likely to adopt a belief connecting two concepts that are already close together in her semantic network, as it seems consistent with the beliefs she already holds (16, *19*). Conversely, she is unlikely to adopt a belief that two distant concepts are connected, as doing so would dramatically reshape her belief structure. The simplest representation of this assumption is that a simulated individual will adopt any belief that her neighbors possess, as long as the existing distance in her semantic network is below a threshold. To make the simulation easy to follow, I use a threshold of 2 links distance, such that a belief will be adopted if it closes a triangle in the adopter's semantic network *(20)*.

In the simulation presented in this paper, each individual is seeded with an equal number of beliefs drawn randomly from a complete semantic network *(20, 21)*. Individuals are given enough beliefs that each individual's semantic network is initially sparse, while maintaining good coverage of the set of all beliefs in the population as a whole. Random seeding eliminates any correlation between sets of beliefs and particular locations in the social network, or any correlation between the popularity of a belief and the popularity of the beliefs it is connected to. Random seeding also ensures that the simulation starts without polarization or systematic variation in belief popularity. I simulate adoption as binary, permanent, and deterministic, as these simplifications make belief interactions clearly observable and repeatable. I place individuals in a random, connected social network to remove any possibility that the network structure may contribute to polarization *(11, 13)*. See the 'Methods' section for more details.



Path	Length
Daniel -> Casino	1
Daniel -> Burglary	2
Daniel -> Art	3

Fig. 1. Beliefs as edges in an individual's semantic network.

I begin by following the diffusion of a focal belief through the social network. The focal belief is able to spread when the prior beliefs of an exposed individual make him susceptible to the focal belief (i.e. in this simplified model the focal belief "closes a triangle" in his semantic network). Continuing to spread to other neighbors in the social network, the focal belief may then reciprocate that facilitation by creating conditions of susceptibility to the beliefs which had previously supported its diffusion, and repeat the cycle. Fig. 2A illustrates this mechanism of "reciprocal facilitation", in which simultaneously-diffusing beliefs alternately create susceptibility to one another, and together are adopted by more individuals than any one belief could have been by diffusing on its own.

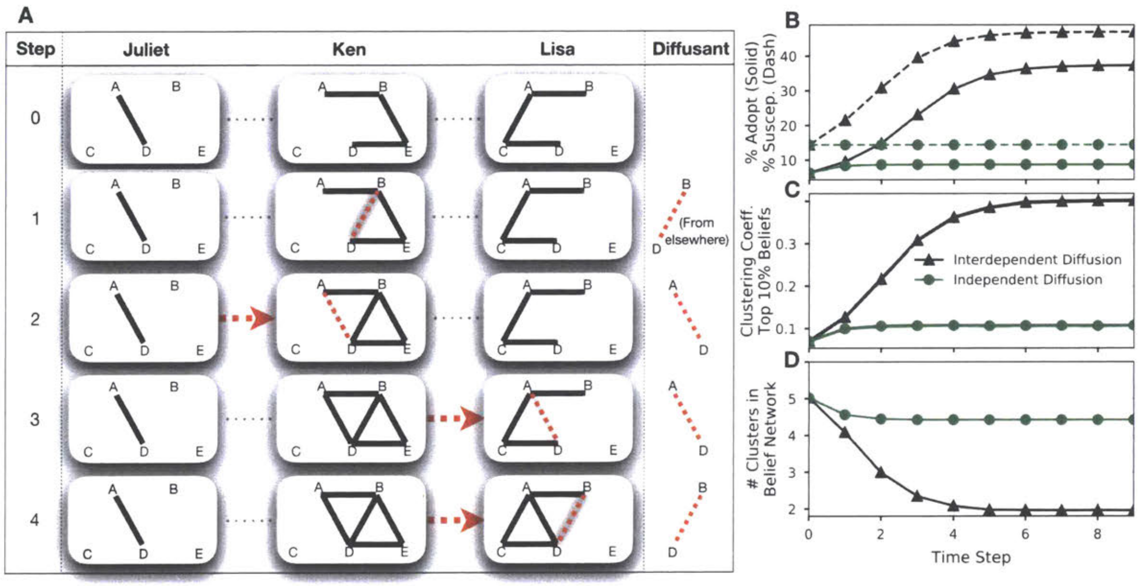


Fig. 2. The effect of reciprocal facilitation on diffusion and belief clustering. (A) *Juliet and Lisa are strangers, but talk regularly with Ken. No diffusion occurs in step 0, as no individual could close a triangle in their semantic network by adopting a neighbor's belief. In step 1, Ken acquires belief BD from an external source. BD facilitates Ken's adoption of belief AD in step 2, making him a bridge for AD to diffuse to Lisa in step 3. In step 4, AD reciprocates by facilitating Lisa's adoption of BD.* ***(B)*** *The average percentages of adopters (solid lines) and susceptible individuals (dashed lines) is shown for a comparison between interdependent diffusion (black triangles) and independent diffusion (green circles) in a simulation of forty individuals.* ***(C)*** *The density of connections between the 10% most popular beliefs is measured as the clustering coefficient of all beliefs that make up the sample.* ***(D)*** *Clusters are counted as the maximum number of separable components in the aggregate semantic network when the least popular beliefs are progressively removed (21).*

The first effect of reciprocal facilitation is to allow the average numbers of susceptible and adopting individuals to grow simultaneously, as shown by the logistic growth curves in Fig. 2B (black triangles). In contrast, independent beliefs that diffuse from the same starting conditions may only be adopted by the fraction that are originally susceptible (green circles). Put another way, to explain the same level of final adoption, models of independent diffusion must assume significantly more initial susceptibility to each belief.

It would be reasonable to make such an assumption about initial susceptibility were it not for the second effect of reciprocal facilitation. Reciprocal facilitation does not support the diffusion of all beliefs to the same degree, but preferentially amplifies those that are connected to other widely adopted beliefs. A belief that is supported by many popular beliefs will have many opportunities to diffuse, and then to facilitate the adoption of other closely related beliefs. Conversely, a belief that is connected only to unpopular supporting beliefs will have trouble reaching even the few individuals who are susceptible to adopting it. As a consequence, the most popular beliefs consolidate into fewer and fewer clusters, as shown in Fig. 2C and D, respectively. These clusters are densely connected islands of popular beliefs that are only linked to one another by less widely adopted beliefs, each residing in a particular (but *a priori* unpredictable) area in the semantic space.

The clusters of popular beliefs generated by this simulation are consistent with previous models of issue alignment. The simulation presented here suggests that alignment occurs not because the most popular beliefs develop relationships with one another, but because the most popular beliefs bring popularity to their neighbors. In contrast, for an independent diffusion model to explain this correlated variation in belief popularity, it must pre-specify which sets of related beliefs are likely to become popular, rather than letting these sets emerge endogenously from the diffusion process itself.

To explain the second mechanism of interdependent diffusion, I focus on the developing semantic networks of individuals. When two individuals exchange beliefs, they become more similar to one another. Because their existing belief sets influence the way they respond to new beliefs, shared beliefs make the two individuals more likely to adopt (or reject) the same new beliefs in the future, independent of any preference to align or distinguish themselves from one another. Similar individuals expand one another's access to beliefs that the two may adopt in common, and filter each other's exposure to beliefs that would set them apart from one another. In contrast, individuals who adopt differing beliefs are likely to diverge further as they become susceptible to beliefs in different parts of the semantic space. Fig. 3A illustrates this "agreement cascade" mechanism, in which the similarity between individuals is reinforced by common adoption and exchange.

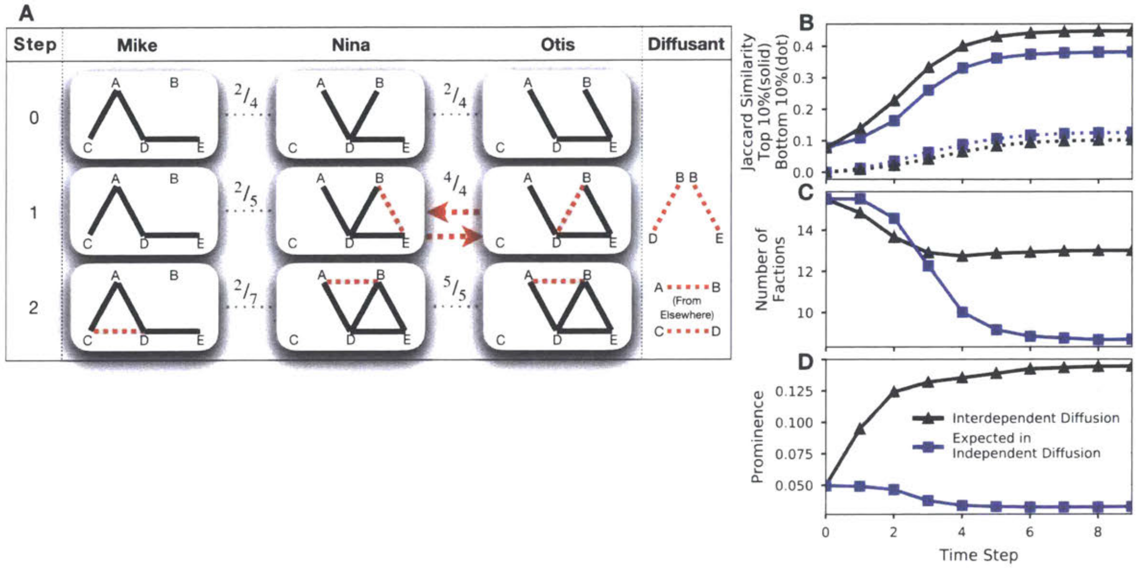


Fig. 3. The effect of agreement cascades on the formation of factions. ***(A)*** *Mike and Otis are strangers, but know Nina. Initially, all possess beliefs AD and DE, with four beliefs among each pair of individuals, giving a Jaccard similarity of ½. In step 1, Nina and Otis adopt BE and BD from one another, becoming perfectly similar and decreasing their similarity with Mike to ⅖. In step 2, all are exposed to AB and CD by an external source. Nina and Otis adopt AB alike, but Mike differentiates himself further by adopting CD.* ***(B)*** *The Jaccard similarity of the 10% most and least similar pairs of individuals (solid and dotted lines, respectively) is compared between interdependent diffusion (black triangles) and what would be expected by chance had independent beliefs diffused to the same extent (blue squares).* ***(C)*** *Factions are identified by hierarchical clustering, interpreted as mapping individuals to a landscape whose height represents the similarity between each pair of individuals. Each "peak" indicates a faction.* ***(D)*** *Prominence is the difference between the maximum similarity within a faction and the average similarity to the next closest faction, or the height of the peak compared to the surrounding landscape, using the same units as Jaccard Similarity (21).*

As simulated individuals do not forget their beliefs, diffusion will raise the average similarity of the population purely by chance. However, Fig. 3B shows that agreement cascades drive the most similar pairs of individuals (solid lines) to become yet more similar to one another, and prevent the least similar pairs (dotted lines) from acquiring even the similarity they might gain by chance adoption of the same beliefs. This leads to the spontaneous emergence of social factions: groups whose members are more similar to one another on average than they are to individuals outside of the faction.

Polarization is characterized by a clear distinction between factions, and high similarity within them. The tendency of agreement cascades to generate polarization is illustrated by Fig. 3C and D. The simulation's random starting conditions exhibit a large number of low-prominence factions. If beliefs diffused independently of one another (blue squares), we would expect that increases in the average similarity between individuals would lead factions to blend together and lose distinctiveness, as individuals in separate factions came to share more beliefs in common. Instead, while the agreement cascade mechanism (black triangles) increases similarity more quickly than expected by chance, it does so primarily by amplifying the similarity within factions, increasing their prominence and distinctiveness. While traditional models of social contagion would predict that the diffusion of beliefs would help to solve the problem of polarization, when beliefs interdepend, the opposite is true.

Discussion:

When dense clusters of popular beliefs are observed empirically *(3, 22)*, it is easy to suppose that they represent internally-consistent worldviews that spread because they make sense as a whole, or represent some underlying truth about the world. Implicit in such a supposition is the idea that other combinations of beliefs would not make sense together, and as a result, alternate worldviews could not become popular. This simulation demonstrates that with the most basic assumptions about belief interdependence, clusters of popular beliefs may emerge even in settings where all beliefs are equally compatible with one another and all beliefs are equally "true". In contrast to the above supposition, large groups of people may come to share a particular set of interconnected beliefs, and reject others to which they had equal initial exposure, without any external reference for which beliefs are adopted and which are discarded.

It is not surprising to see factions develop when a social network is formed from relatively isolated communities *(13)*, or to see polarization grow when individuals choose to associate with people like themselves *(10)*. What is surprising is that when beliefs interact, factions may emerge even in a random, fixed social network *(4)*, without any conscious imitation or desire for differentiation *(3, 8)*, and without exogenous structures of beliefs to which agents may be attracted or repelled *(2,5,6)*. This simulation shows that polarization is an expected and understandable consequence of agreement cascades occurring in the diffusion process itself.

The mechanisms identified here generalize to a broader class of problems involving interdependent diffusion, and have consequences for social policy development. For example, reciprocal facilitation may amplify the dynamics of coinfection between diseases such as TB and HIV, which reinforce susceptibility to one another. If so, targeting treatment to break the feedback between the two diseases may be a high-leverage intervention strategy. As a second example, the agreement cascade mechanism suggests that organizations working to de-escalate a conflict need to do more than just make social connections between rival groups, as any social network can support polarization. They must also establish a shared understanding to serve as a basis for further reconciliation.

I have explored two mechanisms which distinguish interdependent diffusion from independent diffusion: reciprocal facilitation, and agreement cascades. These mechanisms shape diffusion dynamics in sociologically relevant ways: by driving polarization and leading to the emergence of shared worldviews. I show that these dynamics are a fundamental consequence of interdependent diffusion itself, and are not dependent upon external belief structures, heterogeneous preferences, or social network communities. The mechanisms elaborated here predict observable outcomes at the individual, dyad, faction and population levels, allowing the theory of interdependent diffusion to be empirically tested. The mechanisms help generalize the theory to new contexts, make new sociological predictions, and support social policy development.

Methods:

Simulation Details: In the simulation presented in Figs. 2B-D, and 3B-D, the social network is a connected Erdős–Rényi (G_{nm}) random graph with 40 agents total, and in which agents have an average of 4 neighbors. There are 30 concepts (nodes) available in the semantic network, yielding 435 possible beliefs (edges). Each agent is initialized with 25 beliefs (edges) selected randomly from the 435 available. These values ensure good coverage of beliefs in the population, while individual semantic networks are initially sparse. The measures for mean diffusion and number of susceptible individuals are averaged over all beliefs in the simulation, and the presented results are an average of 4500 simulations. In each step, individuals are selected in random order, and update their beliefs by incorporating into their semantic networks any belief (edge) their neighbors possess that closes a triangle in their existing semantic networks.

Clustering of the most popular beliefs: Reciprocal facilitation leads the most popular beliefs to be more clustered than independent diffusion would predict. Fig. 4 shows that any threshold for the most popular beliefs between ~5% and ~40% will yield a qualitatively similar difference between the independent and interdependent diffusion cases. The simulations presented in the paper use a threshold value of 10%, to show that the effect on clustering persists to very high popularity thresholds.

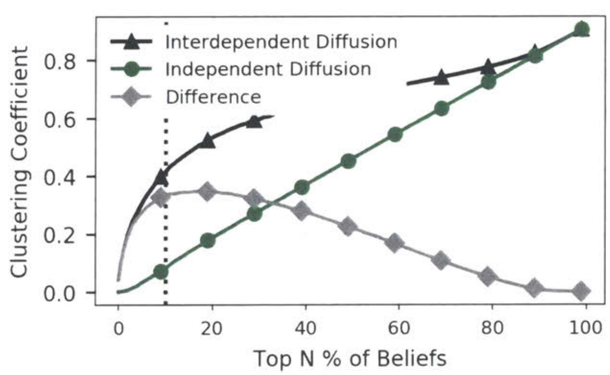


Fig. 4*. The average clustering coefficient of a sample of the N% most popular beliefs grows with N for interacting beliefs (black triangles), and*

when beliefs diffuse independently (green circles). Grey diamonds show the difference between the two cases.

Number of belief clusters: Fig. 2D plots the maximum number of separable components that the aggregate semantic network can be broken into when the least popular beliefs are progressively removed. In their study of belief clustering revealed in social media data, Houghton *et al.* (*22*) identify k-clique communities for varying levels of k and filter threshold. In order to ensure that the clustering identified in this paper is conceptually equivalent to that identified in the empirical study, I use a conservative version of the same technique. Identifying the number of separable components in the semantic network is equivalent to identifying the number of 2-clique communities. As every n-clique community is a subset of an (n-1)-clique community, if more than one 2-clique community is present, we have confidence that these are meaningfully separable, and not a function of the choice of k.

Similarity between individuals: The Jaccard similarity between two individuals is the number of beliefs they have in common divided by the total number of beliefs held by at least one individual in the pair. Agreement cascades drive the most similar individuals to be more alike than we should expect given the same number of beliefs adopted by chance. Fig. 5 shows that this effect only gets stronger as we look at smaller percentiles of the population. Figure 3B uses the top and bottom 10% of comparisons, to include a meaningful sample of comparisons. With a 40-agent simulation, this gives 78 pairs of individuals in each decile.

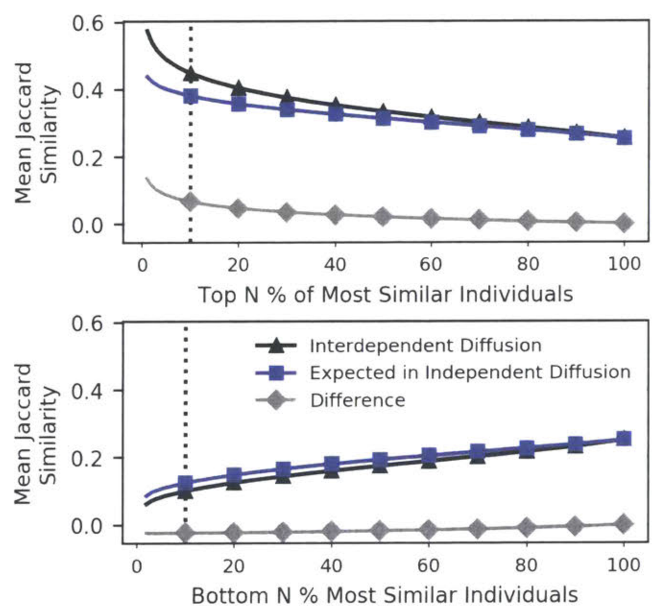


Fig. 5. This figure shows the Jaccard similarity between each pair of individuals for all pairs above (or below) a threshold Nth percentile. Curves represent the difference (grey diamonds) between interdependent

diffusion (black triangles) and what would be expected given the same belief distribution under independent diffusion (blue squares).

Comparison case for the agreement cascade mechanism: As diffusion occurs, both the numerator and denominator of the Jaccard similarity measure grow, and both must saturate when each member of the compared pair has adopted all beliefs. This means that regardless of *which* beliefs the individuals adopt, as they adopt more beliefs, individuals become more similar to one another purely by chance. Fig. 3B-D aims to demonstrate the effect of agreement cascades in organizing factions, without the component that is due to the gross number of adoptions. The best comparison case is therefore what we would expect to see had independently diffusing beliefs reached the same level of adoption. This comparison is created by shuffling beliefs between individuals, so that the distribution of adoptions is equal between the interdependent diffusion case and the comparison case.

Number and Prominence of Factions: A faction is a group of individuals who are more similar to one another on average than they are to individuals outside of the faction. Fig. 3C-D uses measures of polarization that aggregate from local features of the relationships between individuals, rather than partitioning the entire population into groups. Doing so allows for the potential absence of factions, and for the possibility that some individuals in the population may not be a member of any faction at all. The number of factions is measured by counting the pairs of individuals who are more similar to one another than they are to any other individual. To illustrate, if individuals were mapped to an n-dimensional landscape, where the height of the landscape was the similarity between individuals, these pairs would represent local maxima. The prominence of each faction is the difference between the similarity of most similar individuals in the cluster and the average similarity to the next closest cluster, or how far down the landscape you have to go from one peak before you can start climbing the next closest peak.

The landscape is constructed using a hierarchical clustering algorithm, which progressively groups individuals or sub-clusters that are most similar to one another *on average*. (Alternate algorithms are shown in the supplement to give qualitatively similar results.) Fig. 6 shows a dendrogram illustrating hierarchical clustering on a 20-agent simulation (dendrograms typically plot height as increasing difference, or decreasing similarity). The number of factions is the number of pairs of individuals who are more similar to one another than to any other individual (highlighted in bold black). The prominence of each faction is the difference between the height of the peak and the closest link in the dendrogram that would connect it to another peak (red T-shapes). The remaining population may have its own underlying form, with some individuals more connected than others, but not organized into recognizable groups.

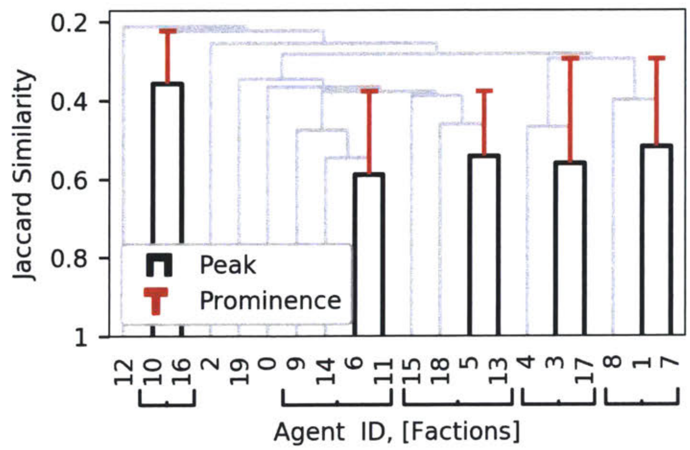


Fig. 6. The number and prominence of factions. In this illustrative example, there are five peaks (dark-shaded pairs), indicating five factions (marked with brackets); two of which have four members ({9, 14, 6, 11} and {15, 18, 5, 13}), two have three members ({4, 3, 17} and {8, 1, 7}) and one has two members ({10, 16}). There are four individuals (12, 2, 19 and 0) who do not participate in any faction.

References and Notes:

1. Nickerson, Raymond S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* **2** 175–220 (1998).
2. N. E. Friedkin *et al.*, Network science on belief system dynamics under logic constraints. *Science.* **345**, 321-326 (2016).
3. D. DellaPosta *et al.*, Why do liberals drink lattes? *AM. J. Sociol.* **120** 1473-1511 (2015).
4. A. Goldberg, S. Stein, Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation. *Am. Sociol. Rev.* **83** 897-932 (2018).
5. S. E. Parsegov *et al.*, Novel Multidimensional Models of Opinion Dynamics in Social Networks. IEEE Trans. Automat. Contr. **62** 2270-2285 (2017).
6. F. Xiong *et al.* Analysis and application of opinion model with multiple topic interactions. *Chaos.* **27** 083113 (2017).
7. C. T. Butts, Why I know but don't believe. *Science.* **354** 286-287 (2016).
8. D. Baldassarri, P. Bearman, Dynamics of political polarization. *Am. Sociol. Rev.* **72** 784-811 (2007).
9. M. Granovetter, The Strength of Weak Ties. *Am. J Sociol.* **78**, 1360-1380 (1973).
10. T. Schelling, "Sorting and Mixing" in *Micromotives and Macrobehavior.* (Norton, Toronto, 1978) chap. 4.
11. D. J. Watts, A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. U.S.A.* **99** 5766-5771 (2002).

12. R. Reagans, B. McEvily, Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. *Adm. Sci. Q.* **48** 240-267 (2003).
13. D. Centola, M. Macy, Complex Contagions and the Weakness of Long Ties. *Am. J Sociol.* **113** 702-734 (2007).
14. S. Aral, M. Van Alstyne, The Diversity-Bandwidth Trade-off. *Am. J Sociol.* **117** 90-171 (2011).
15. SI Part A demonstrates how exogenously assumed belief structures may bias simulation outcomes.
16. A. M. Collins, E. F. Loftus, A Spreading-Activation Theory of Semantic Processing. *Psychol. Rev.* **82** 407-438 (1975).
17. M. Steyvers, J. B. Tenenbaum, The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cogn. Sci.* **29** 41-78 (2005).
18. R. J. Brachman, "On the epistemological status of semantic networks" in *Associative Networks.* 3-50 (Acad. Press, 1979).
19. M. Schilling, A 'small-world' network model of cognitive insight. *Creat. Res. J.* **17** 131-154 (2005).
20. SI Part C explores the consequences of relaxing these assumptions.
21. The 'Methods' section and SI Part B explain additional details of the simulation and measurements.
22. J. P. Houghton *et al.*, Beyond Keywords: Tracking the Evolution of Conversational Clusters in Social Media. *Sociol. Methods Res.* (2017).

Acknowledgments: Special thanks to Ray Reagans, Hazhir Rahmandad, Carolyn Fu, Joshua Becker, and the System Dynamics and Economic Sociology groups at MIT Sloan for their guidance.

Data and materials availability: All code and materials are available at https://github.com/JamesPHoughton/interdependent-diffusion.

Appendices to

Why meaning matters for belief diffusion in social networks

James P. Houghton

Appendices include:

A. Non-biasing Formalization of Belief Interaction
B. Additional Simulation Details and Explanation of Measures
C. Sensitivity and Robustness Tests
D. Simplified Model Code

Figs. S1 to S13

A: Non-biasing Formalization of Belief Interaction

When we use models to study the effect of diffusion on the emergence of factions or polarization, it's important that the assumed decision logic relating current beliefs to future adoptions not be informative of the shape of factions or the beliefs around which the population polarizes. This isn't to say that in the real world there is no shared decision logic, or that the shared decision logic has no influence over what beliefs are most likely or what clusters form in the population. The problem is that if the observed pattern of belief clustering is predictable from the decision logic, then the simulation is unable to identify whether clustering is shaped by social contagion, or is merely a result of the decision logic. A simulation in which clustering is a product of either the decision logic or the diffusion process is indeterminate for questions about which creates clustering.

An interdependence matrix which maps the presence of belief A to the likelihood of adopting belief B will always be informative of the final configuration (or trivial). We can demonstrate this with an extremely simple deterministic model, with no diffusion at all. This model is in the style of Friedkin *et al.(2)*, but omits all social influence, stochasticity, and continuous levels of belief. Simplifying the model this way lets us see intuitively why the interdependence matrix is problematic for our purpose in a way that is less obvious when the various complicating factors are in place. When stochasticity, social influence, etc. are present, the problem doesn't go away, it is just masked by the complexity of the model.

In this simplified interdependence-matrix belief adoption model:

1. Each agent is exposed to all beliefs in every timestep (making this an individual learning model, not a social learning model)
2. Having a belief 'A' either contributes (+) to an agent's likelihood of adopting belief 'B', or takes away from it (–)
3. Not having belief 'A' has no direct influence on an agent's likelihood of adopting belief 'B'
4. Belief adoption is binary and permanent
5. Agents adopt a belief if the majority of their current beliefs suggest they should. (Other thresholds are of course possible.)

To show that the matrix of influence is informative, I'll use a population of 64 individuals, each with a unique combination of six beliefs, denoted A-F. The top center chart in Fig. S1 illustrates this starting condition. Each column represents an individual (0-63), each row represents a belief (A-F). I darken the corresponding square to show that an individual has adopted a particular belief. To the right-hand side of the adoption plot, I show a histogram of the total number of individuals adopting each belief. In the initial condition, all beliefs have been adopted by 32 individuals.

I randomly draw two matrices of influence in the style of Friedkin *et al.*, Logic 1 and Logic 2, plotted on the left-hand side of Fig. S1, with equal weights for + (red) and – (blue). In each matrix, if an individual already has belief A, then the influence of that belief on the adoption of belief B is found in Column A, Row B. In Logic 1, this influence is positive, and in Logic 2, this influence is negative.

Now, for each individual, I use the starting belief set to calculate the net influence on each as-yet-unadopted belief. If the candidate belief gets a majority of + votes from the existing belief set, then the candidate belief is adopted. (This assumes that each individual has equal access to all candidate beliefs at all times, removing the influence of social contagion altogether.) I repeat the process until the individual has converged on a stable combination of beliefs. That is, no new beliefs are added. I repeat the process, again from the starting condition, for each logic, displaying the results in the second and third row of the figure.

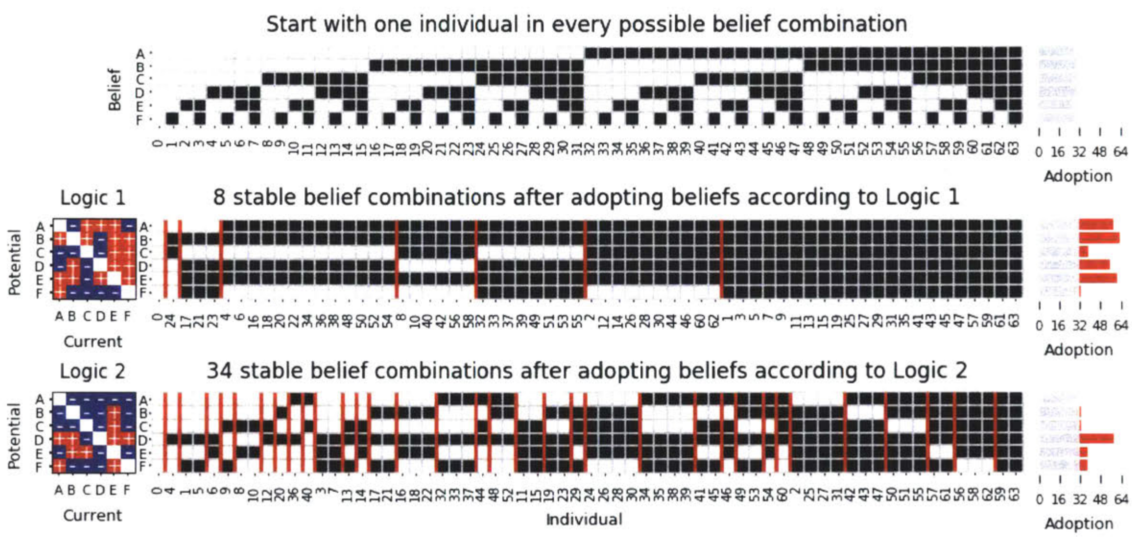


Fig. S1. *With an interdependence-matrix style belief structure, are the extent of factionalism or differences in the level of belief adoption due to social contagion, or can they be explained just with the belief structure?*

Any model of social learning with binary and permanent belief adoption and a finite number of beliefs will show some form of belief consolidation, and an increase in mean similarity between individuals. We expect to see that from our maximally-differentiated initial conditions, some groups of stable belief combinations should emerge. In the second and third adoption plots, I group individuals according to their stable sets of beliefs, indicating the groups with a red divider.

The two Friedkin-style logics, despite being drawn randomly, yield very different stable combinations of beliefs. In Logic 1, there are 8 different stable combinations, and in Logic 2, there are 34. In a diffusion model, we would expect to see a lot more consolidation of beliefs, and clustering of individuals, using Logic 1 rather than Logic 2, regardless of the network structure. Moreover, the two different logics strongly influence which beliefs we should expect to be widely diffused in the population, and which we expect to see only because of their presence in the initial conditions. Logic 1 suggests strong new adoption (shown in red) of beliefs A, B, D, and E. Logic 2 shows that only belief D is widely adopted, with belief A having no new diffusion at all.

In the real world, it may well be that the natural logic which ties some beliefs together will strongly influence the shape of social clustering and the extent of diffusion of particular beliefs. In our simulation, however, this influence makes it extremely difficult to identify clustering and amplification of beliefs that is genuinely due to a social process. In this paper, I have suggested an alternative formulation, in which beliefs are formalized as edges in a semantic network, and the adoption rule is that beliefs are adopted if they close a triangle in an individual's semantic network.

In Fig. S2, I repeat the above analysis with this new formulation. Again there are six beliefs, representing all of the possible edges in a semantic network with concepts P, Q, R and S. Again, each of 64 individuals is initialized with a unique combination of beliefs, and has access to all six beliefs. Each individual adopts new beliefs that are consistent with the triangle-closing decision rule, and I plot the stable belief combinations after each individual has individually converged. As there is only one decision rule, there can be only one outcome for grouping the sets of stable beliefs.

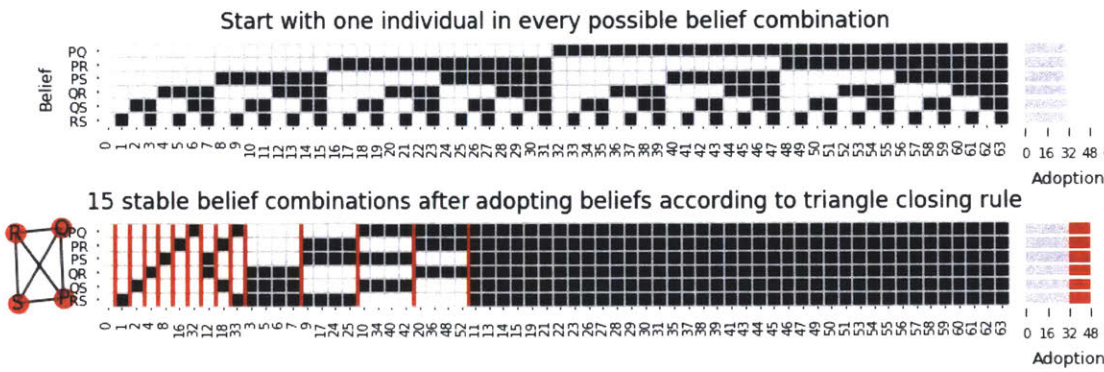


Fig. S2. *With the semantic network representation of belief structure, we have confidence that any variance in adoption patterns, or differences in social clustering, are due to the interaction of the belief structure with the diffusion process.*

With this formalization, we know exactly how the decision rule for adoption influences the number of possible stable states; and most importantly, these states are symmetric with respect to the individual beliefs. The histogram to the right of the second row of S2 shows that each belief has an equal number of new adoptions, and that we should not expect any beliefs to preferentially diffuse as a result of the decision rule itself.

In this paper I have shown that some beliefs do indeed diffuse much more widely than others, and that the population clusters into a subset of the possible stable states.

Because the decision rule does not influence which of the beliefs diffuse most widely, or suggest variation in the number of social clusters we can expect, we have confidence that the results are genuinely due to the interaction of the belief structure with the diffusion process.

B. Additional Simulation Details and Explanation of Measures

<u>Comparison cases for the reciprocal facilitation mechanism</u>

In Fig. 2 I compared interdependent diffusion (with reciprocal facilitation between beliefs) with a case in which beliefs diffused independently from one another, in which the susceptibility to each belief is fixed from the start of the simulation (no facilitation between beliefs). A third case is also plausible, in which the diffusion of one belief precedes the diffusion of the second, and may facilitate its adoption (in a "forwards" direction, chronologically) but not vice versa. Fig. S3 adds sequential diffusion with forward facilitation to the charts from Fig. 2.

Averaged over all beliefs in the simulation, each allowed to diffuse one after another, we see that forward facilitation allows average susceptibility to rise somewhat above the independent case. With sequential diffusion and forward facilitation, the number and density of clusters lean towards the result we saw with simultaneous diffusion and reciprocal facilitation. However, the results of forward facilitation are more qualitatively similar to the independent case than the dramatic outcomes under simultaneous diffusion. This suggests that it is truly the interactive nature of reciprocal facilitation that generates the results we see.

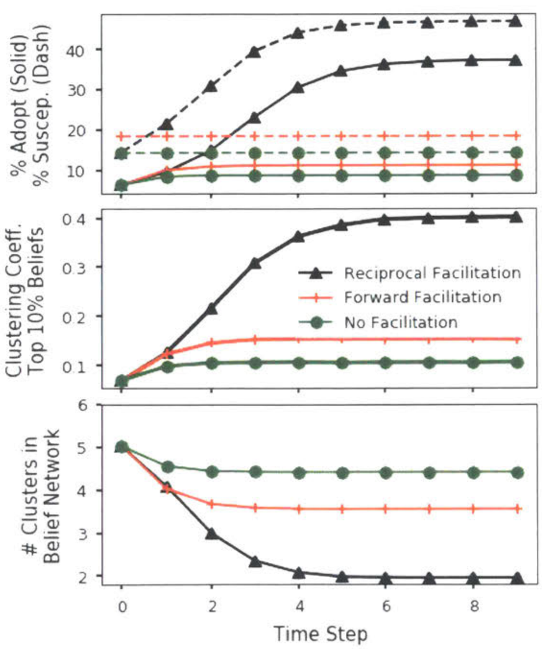


Fig. S3. *Sequential diffusion with forward facilitation suggests the same direction of result as simultaneous diffusion with reciprocal facilitation, but represents a qualitatively different phenomenon.*

C. Sensitivity and Robustness Tests

The model in this paper distills its representation of interdependent diffusion to the absolute minimum structure needed to explain reciprocal facilitation and agreement cascades. While this is appropriate for its purpose, we need to have confidence that the results are not an artifact of the simplifications. When we relax the simplifications that are not part of the sociological theory being advanced, we should still arrive at the same conclusion. When we move away from the structural components of the theory, we should see that the behavioral components attenuate appropriately.

For example, to create the behavior-over-time graphs in Figs. 2 and 3, I seeded each agent with a small number of beliefs (25 seeds each) randomly selected from a finite set of all possible beliefs (435 possible beliefs). In the real world, we don't know how many beliefs there are, or what fraction are acquired independently of social diffusion. To test the robustness of the model to changes in the number of seeds, we should vary the number of beliefs individuals start with and see if the reciprocal facilitation and agreement cascade mechanisms are still active. If so, do they lead to the same conclusions about clustering of beliefs and social polarization?

<u>Sensitivity to the number of starting beliefs</u>

The reciprocal facilitation mechanism is active when susceptibility and adoption evolve together. For adoption to happen, an individual must be simultaneously susceptible to a belief and exposed to it. If the simulation starts with too few beliefs, these factors will cooccur too infrequently for meaningful diffusion, whether beliefs diffuse simultaneously or individually.

Fig. S4A shows how the ultimate number of individuals susceptible to a belief (dashed lines) and final adoption of beliefs (solid lines) varies with the number of starting beliefs per agent. (All other parameters are as they were in the simulations presented in the main body of the paper.) At the start of the simulation (light blue stars) adoption is by definition proportional to the number of beliefs we seed. The initial number of susceptible individuals is dependent upon the chance that two facilitating beliefs create an open triangle that the candidate belief may close. Increasing from low seeding densities, the average frequency of susceptibility to a belief grows faster than linearly with the number of seeds, and saturates as it approaches 100%.

After the simultaneous diffusion of beliefs (black triangles), susceptibility and adoption have both grown, and with more than 75 ~ 100 seed beliefs (*ceteris paribus*) the simulation saturates with full adoption of all beliefs. This saturation is an artifact of the finite belief universe we stipulate in the model, and influences some of the behavior we will see at higher seed densities as we perform our sensitivity tests.

Fig. S4B shows the same outcome as S4A, except that beliefs diffuse independently from the initial condition. This is the case in which reciprocal facilitation is "switched off", the comparison case I use in the main body of the paper. In this case, the frequency of susceptibility is fixed at what is provided by the initial belief seeds. With enough starting beliefs, however, this susceptibility also saturates at 100%, and the initial adoption is such that diffusion (green circles) reaches nearly all susceptible individuals.

Fig. S4C shows the difference between the simulation condition and the comparison condition. We can interpret this as the component of diffusion which is due to the reciprocal facilitation mechanism. When agents start with very few beliefs, the two cases are similar in their poor diffusion. After moderate increases to the number of seed beliefs, reciprocal facilitation is able to increase the level of diffusion in excess of what is possible without belief interaction. When agents have many starting beliefs, adoption in either case saturates at 100%, and so reciprocal facilitation makes little difference.

Varying the number of starting beliefs here shows that we don't expect reciprocal facilitation to have a large impact when contagion is weak, or when susceptibility is so strong that beliefs will diffuse even without facilitation. However, the nontrivial diffusion processes most likely to be of interest to sociologists will occur in the middle range, in which the mechanism is active.

How do the outcome measures change with the number of starting beliefs? Fig. S4D shows how the clustering of the top 10% of beliefs varies with the initial number of beliefs. As we expect, the clustering of the most popular beliefs increases as the reciprocal facilitation mechanism becomes active. Why does it decrease when starting beliefs are densely seeded? As adoption saturates near 100%, there is less difference between the most popular beliefs and the least popular beliefs, until at full saturation all beliefs are equally popular. At this point, the most popular 10% is just a random sample of the full set of beliefs, and the clustering of that sample returns to what we expect due to chance.

In Fig. S4E, we see the same process occurring in the independent diffusion case. Though not shown in the diagram, when the initial number of beliefs is high enough that diffusion saturates at 100%, the clustering coefficient must be that due to randomness. Because the diffusion saturates at much higher numbers of starting beliefs, the curve peaks further to the right. In Fig. S4F, we see the difference between the two curves. Interestingly, with sufficiently many starting beliefs, the reciprocal facilitation case exhibits less clustering than the no facilitation case. This reversal is purely an artifact of saturation, and occurs because our simulation has a finite number of beliefs. In discussion of Fig. S5, we'll explore what happens as we increase the number of concepts in the semantic network, and the number of beliefs an agent may choose between.

In Fig. S4G, H, and I, we see that the reciprocal facilitation mechanism encourages the formation of a small number of clusters when seeded with many fewer beliefs than we would predict if beliefs did not interact. When diffusion is complete, we see by definition a single large cluster in both cases. The fact that the curve for reciprocal facilitation approaches a single cluster smoothly, and with relatively fewer seed beliefs, is significant. When diffusion is independent, the number of clusters is driven by chance even for very densely seeded semantic networks (the curve in H does not bend downward until the simulation shows over 80% average adoption). In contrast, the reciprocal facilitation mechanism begins pruning and merging smaller clusters at relatively low levels of average adoption. As there is no natural expectation for the number of clusters we should expect, Fig. S4I shows the ratio of the two cases, rather than the difference.

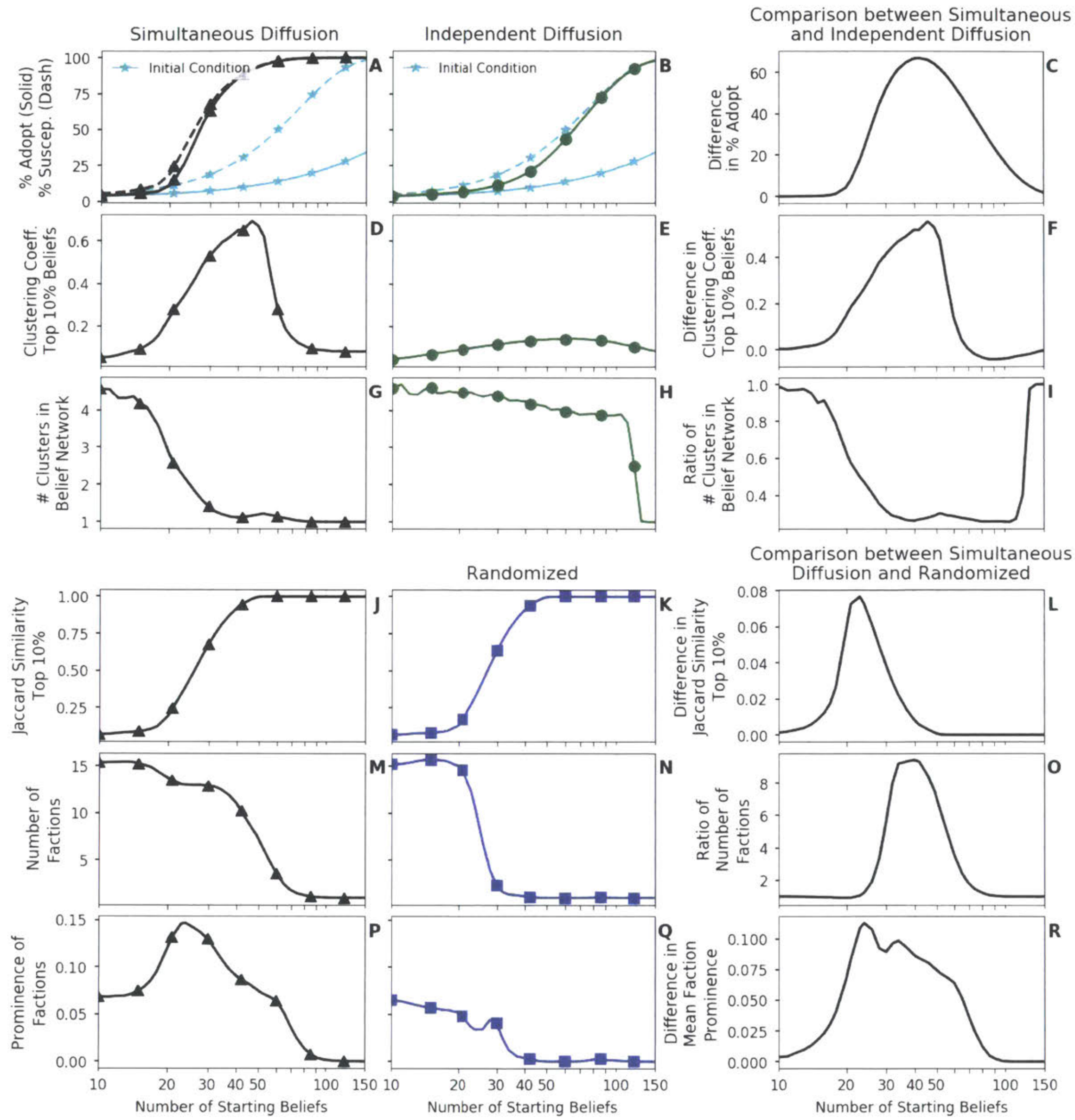


Fig. S4. *A detailed exploration of how the model results change as we vary the number of beliefs each individual is seeded with at the beginning of the simulation. The leftmost column shows each metric under interdependent diffusion, when beliefs diffuse simultaneously, support one another's diffusion, and participate in agreement cascades. The middle column shows the same measures in the appropriate comparison case: independent* *diffusion without any facilitation for B, E, and H; and the case in which beliefs are shuffled between agents for K, N and Q. The rightmost column gives a comparison of the first two columns, representing the impact of reciprocal facilitation and agreement cascades. When a measure has clearly defined upper and lower bounds (Percent, Clustering Coefficient, Jaccard Similarity/ Prominence) I give the difference between the cases. When the measure does not have a natural scale, I show the ratio, to give a better sense of whether the comparison shows a large distinction or a small one.*

As in Fig. 3 in the main body of the paper, I randomize the beliefs of individuals to serve as a comparison case for the remaining measures. Shuffling the beliefs that individuals possess gives a more conservative comparison for these measures, because so much of the similarity between individuals is driven by the average number of beliefs they have adopted. As the diffusion of beliefs under the independent diffusion case used in plots B, E, and H is so much lower than the reciprocal facilitation case, if we were to use this for comparison, we would not be able to separate the true effect of belief interaction on interpersonal similarity from the generic effect of diffusion.

At very low densities of seed beliefs, individuals have very little in common. With very high densities of seed beliefs, they have very little to differentiate themselves. In between, what they share is a function of how they exchange beliefs – the mechanism I describe as agreement cascades. In Figs. S4J, K, and L, we see that the primary effect of this mechanism is to shift the S-shaped similarity curve leftwards. That is, the most similar individuals are able to achieve their similarity not because of the *number* of beliefs they possess, but because of coordination among *which* beliefs they possess.

Figs. S4M, N, and O count the number of factions in the population, with the last plot again showing the ratio rather than the difference between cases, as there is no natural expectation for the number of factions we should expect. At very low levels of starting beliefs, the randomized case can generate slightly more factions than the simulated case, as small amounts of diffusion in the simulated case occur between already similar individuals, increasing their similarity further. This tends to amplify the existing clusters, but not to create more of them. When beliefs are randomized, the number of clusters is still determined by chance, but there is more heterogeneity in the number of beliefs each individual has, and so more opportunity for peaks to emerge by chance.

Plots S4P, Q, and R highlight the average prominence of the revealed factions. With low numbers of starting beliefs, the number and prominence of factions is due to chance. When there are high numbers of starting beliefs, and the diffusion begins to saturate, factions fade and the heterogeneity between individuals shrinks as all individuals adopt all beliefs. In between, chance acts to smooth the landscape and decrease the number and prominence of factions, while the agreement cascade amplifies them.

Sensitivity to the size of the semantic network

The curves in the rightmost column of Figs. S4 (C, F, I, L, O and R) provide a useful way to compare the impact of changes to various parameters in the simulation. Fig. S5A shows that when we increase the number of concepts in the semantic network, the maximal effect of reciprocal facilitation also grows. However, the maximum occurs with a larger number of seed beliefs; as with more concepts, any two beliefs are less likely to create an open triangle (and consequent opportunity for diffusion) in the larger semantic network. If we were to make comparisons across semantic network sizes with a fixed number of starting beliefs, we might observe either an increase or decrease in the effect of reciprocal facilitation, depending upon the point we chose. For the rest of the sensitivity tests, I will show the full comparison curves between the simultaneous and independent diffusion cases.

The effect of a larger semantic network size is much the same on the other measures. The fact that clustering of the top 10% of beliefs is comparable between a semantic network of 20 concepts and 190 beliefs and a semantic network with 80 concepts and 3160 beliefs gives confidence that the effect will persist at even larger scales. With a larger semantic network, a larger number of clusters is possible in the semantic space under random conditions, and so the ratio between the two cases can be greater.

When the semantic network is larger, the effect of agreement cascades on faction prominence declines somewhat, as there is a larger stock of beliefs to differentiate individuals from one another, increasing the denominator in the calculation of Jaccard similarity. A simulation in which individuals forgot about beliefs that were not socially or semantically reinforced, we should expect this trend to reverse.

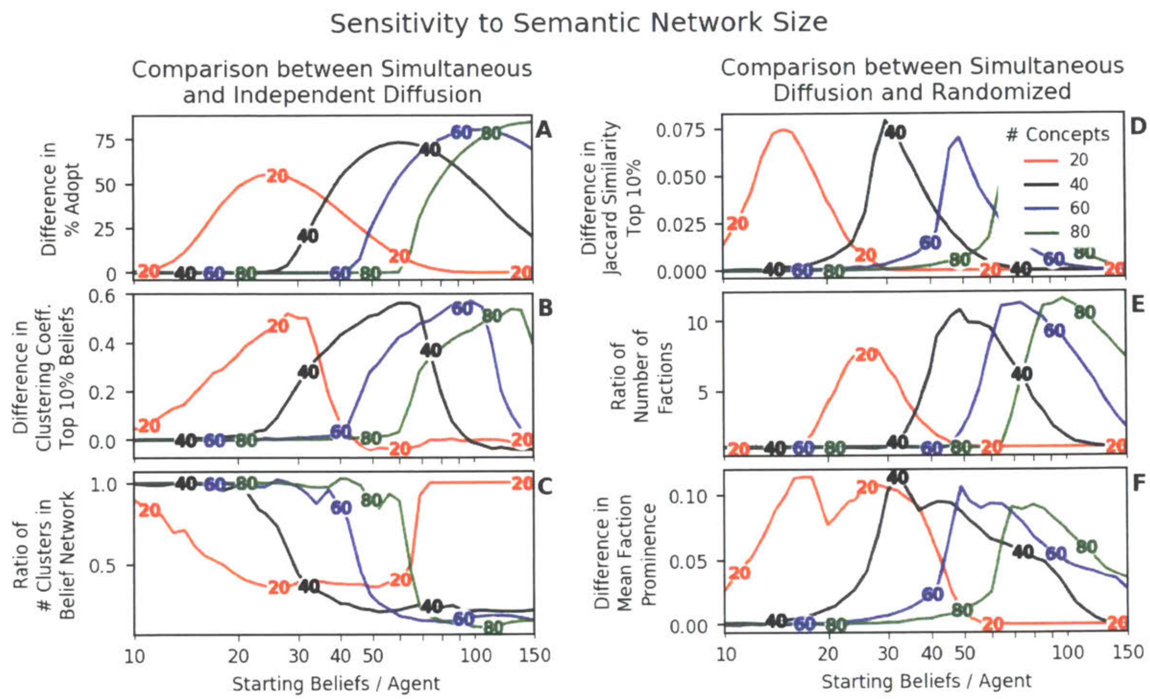


Fig. S5. The magnitude of the effect of reciprocal facilitation and agreement cascades is moderately influenced by the number of concepts in the semantic network.

Sensitivity to the size of the social network

The size of the social network has little effect on reciprocal facilitation's contribution to belief diffusion. However, with more agents in the social network, we are better able to identify the 10% most popular beliefs, and so the randomizing effect of belief saturation is minimized. This generates the rightward shift we observe for large social networks for the curves in Figs. S6B and C; although the sociological relevance of this shift is debatable.

A larger social network has more opportunities for the formation of local factions. In the individual diffusion case without reciprocal facilitation, the population is still formed into something of an amorphous mass, and so the ratio of the number of factions in the simultaneous diffusion case grows with the size of the population.

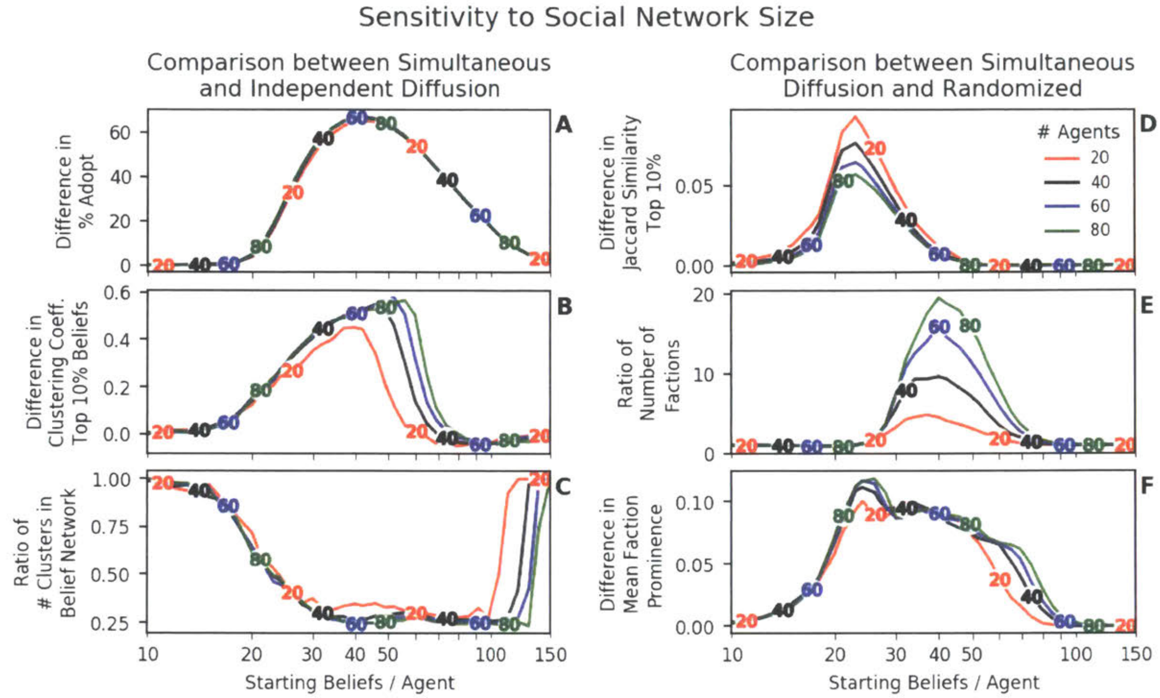


Fig. S6. The effect of reciprocal facilitation is generally insensitive to the size of the population. Larger populations provide opportunities for the development of more local factions.

<u>Sensitivity to the shape of the social network</u>

In the main body of the paper, I conducted simulations on a random social network to minimize the effect that social network structure would have on the formation of factions. Fig. S7 reintroduces social network structure with the Barabasi-Albert, Watts-Strogatz (p=.02), and connected caveman graphs. Social network degree is held fixed (or as close as possible for B-A) along with network size.

Adding network structure serves to limit the extent of diffusion, with a larger retarding effect on independent diffusion than simultaneous diffusion. This increases the observed impact of the reciprocal facilitation mechanism on diffusion for cases with more seed beliefs.

Just as reciprocal facilitation has little impact on the extent of diffusion when initial susceptibility is high enough to saturate diffusion on its own, we expect that when the social network already has built in communities there will be less room for the agreement cascade mechanism to make a difference in the observed number of factions. Fig. S7E shows this to be the case. The number of factions present in the connected caveman graph is greater than in any of the other network structures, under both simultaneous and independent diffusion. With so many clusters present even under independent diffusion, the relative impact of agreement cascades is minimized.

Figs. S7D and F show that when communities are identifiable in the social network, the impact of agreement cascades on faction distinctiveness is amplified. We see a more skewed similarity distribution, and greater prominence in the resulting factions.

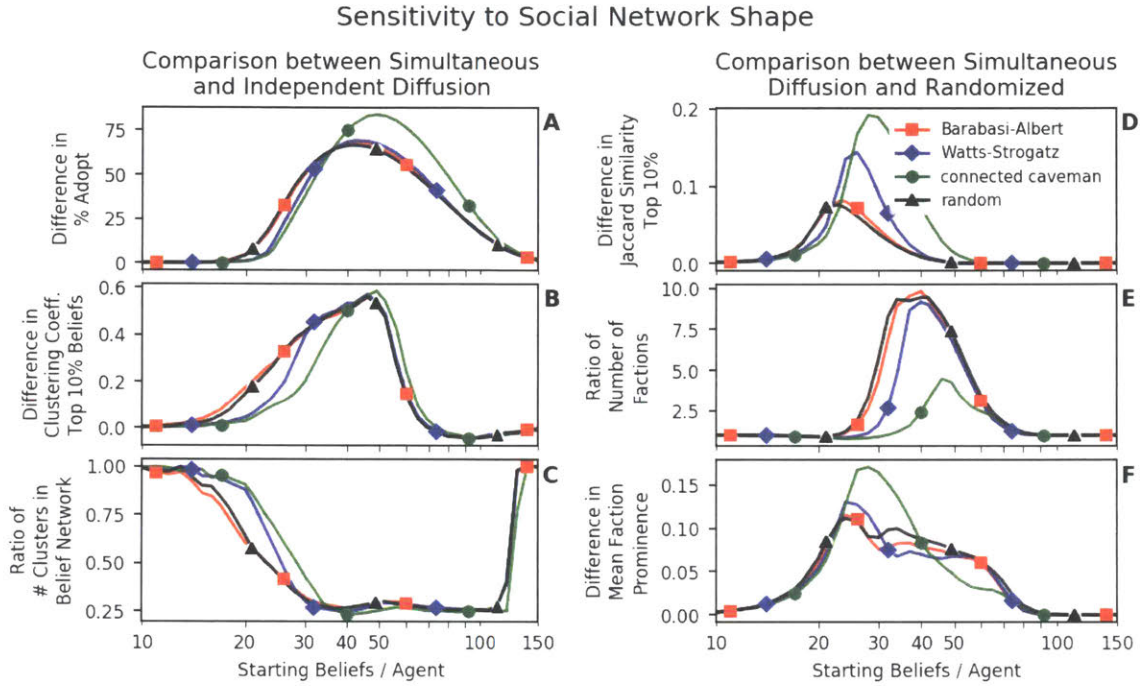


Fig. S7. The presence of communities in the network structure generates its own factions under either simultaneous or independent diffusion. Communities accentuate the impact of agreement cascades on faction prominence.

<u>Sensitivity to the average number of neighbors in the social network</u>

The primary impact of having more social connections is having more exposure to beliefs. When you have more beliefs to choose from at the start of the simulation, the relative importance of reciprocal facilitation for the final extent of diffusion declines. As with communities in the structured social networks in Fig. S8, having fewer social network connections amplifies the impact of agreement cascades on the prominence of clusters and the similarity of the most similar individuals.

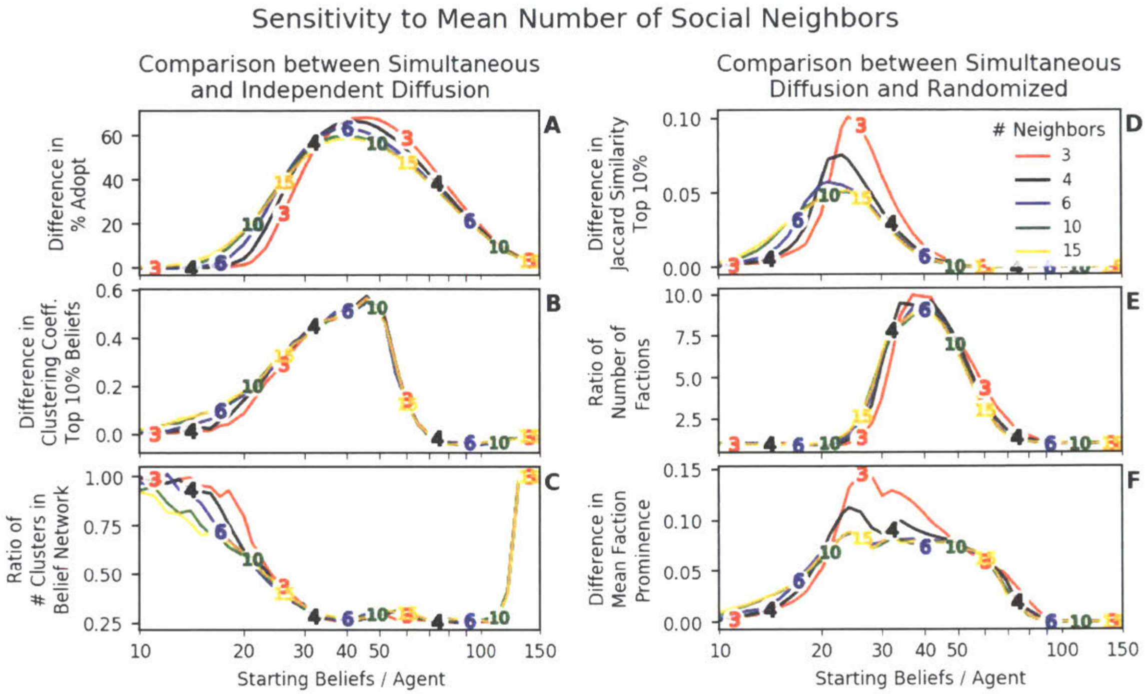


Fig. S8. Understanding how the results of the simulation depend on the average number of neighbors each agent is connected to in the social network.

<u>Sensitivity to the adoption decision rule</u>

In this paper, I simplified individuals' adoption decision rules to say that individuals adopted a belief if it closed a triangle in their semantic networks, as long as at least one of their neighbors has also adopted – that is, they are exposed to the information. In contrast, threshold models of diffusion ignore the semantic dimension altogether, and pay attention instead to the fraction of an individual's neighbors who have adopted the belief. These two dimensions form a space in which we can define more complex decision rules. Fig. S9A shows the triangle closing decision rule as a region occupying the full height of the decision space, for a certain maximum semantic network distance. S9B shows the threshold model as occupying the full width of the space above a given threshold. S9C shows one way we could combine the two rules while maintaining a binary decision rule.

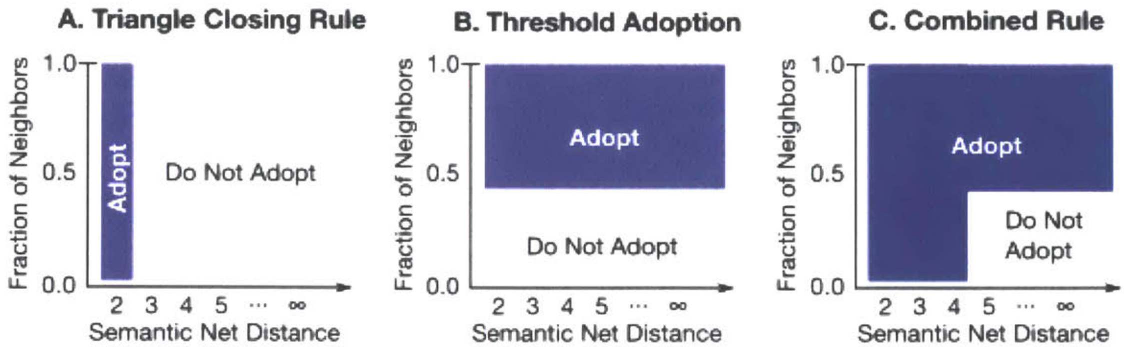


Fig. S9. A polygon-closing rule can be combined with a threshold rule for a rule based upon both social and semantic considerations.

Naturally, if the space was full, this would correspond to the case in which an agent would always adopt a belief they were exposed to, regardless of the semantic or social conditions under which it is presented. Under this condition we would see no difference between simultaneous or independent diffusion by definition, as we have entirely removed the interaction between beliefs. In Fig. S10, we progressively approach this condition from the left, by increasing the distance in the semantic network that a belief may span and be adopted, from a path length of 2 through 30, encompassing all beliefs that have any existing path in the semantic network. When individuals use semantic information in a less discriminating way, the clustering of beliefs naturally approaches the always-adopt condition.

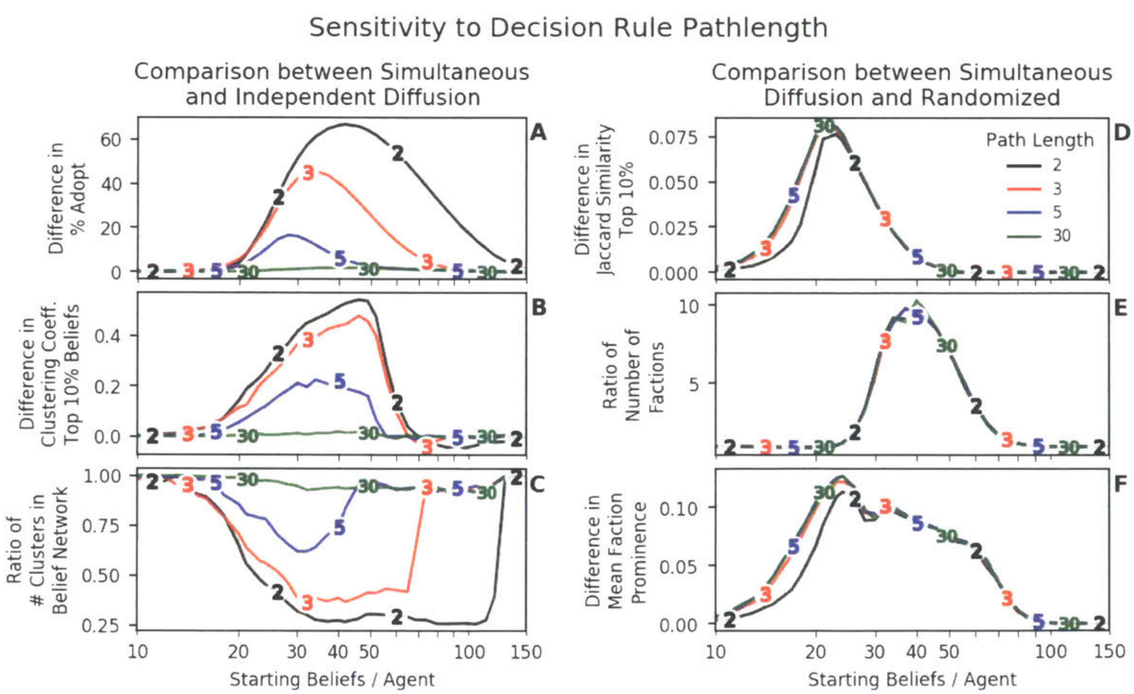


Fig. S10. When we increase the semantic network distance that a new belief may span, the behavior of the simulation approaches that under the always-adopt decision rule.

Contrary to our expectation, Figs. S10D, E and F show very little response to the less discriminating decision rule. Fig. S11 helps to explain why. When individuals become more relaxed about using semantic information to discriminate between beliefs, we see that the extent of diffusion changes dramatically under the individual diffusion case, but hardly at all under the simultaneous diffusion case. This is because reciprocal facilitation makes it easier to fill in the intermediate triangle that it would take to close a square, and so the ability to close a square makes less of a difference, similarly for larger polygons.

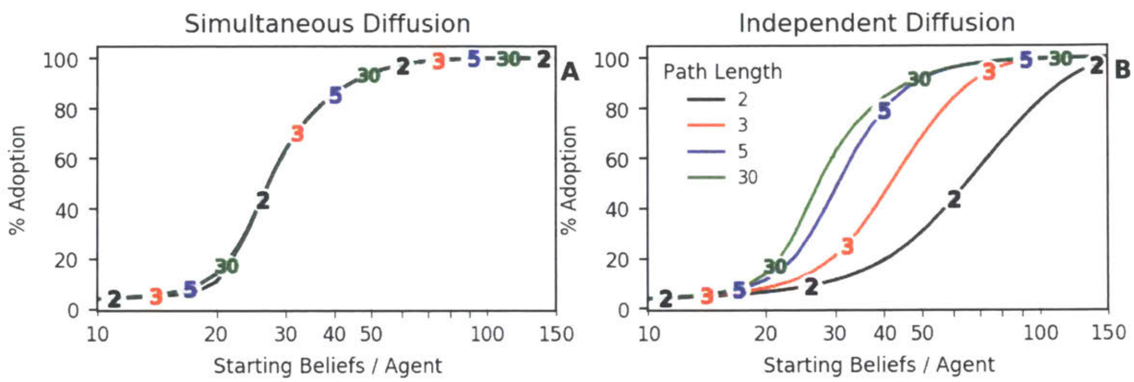


Fig. S11. Less discriminating decision rules increase the mean extent of diffusion under individual diffusion by increasing the fixed frequency of susceptibility to each belief. They have little impact on the simultaneous diffusion case, because reciprocal facilitation allows susceptibility to evolve throughout the simulation, and so the initial level of susceptibility is less important.

As the righthand column of Fig. S10 is based upon the simultaneous diffusion case alone (with beliefs shuffled to make the comparison case) and there is little difference in the adoption patterns for different path lengths in the decision rule, we also see little change in the factionalism measures.

So far, individuals have only been able to adopt beliefs where there are existing paths in their semantic networks. Now I will add to the individual's decision rule the ability to adopt a belief if it closes a triangle in their semantic network, or if the fraction of their neighbors who have adopted the belief is greater than some threshold. When the threshold is 1, we see the original behavior. When the threshold is less than 1, individuals now have the potential to adopt beliefs that are not connected to their existing semantic networks in any way. Fig. S12 shows that as we reduce the threshold, the final extent of diffusion shifts under both simultaneous and sequential adoption, as we expect.

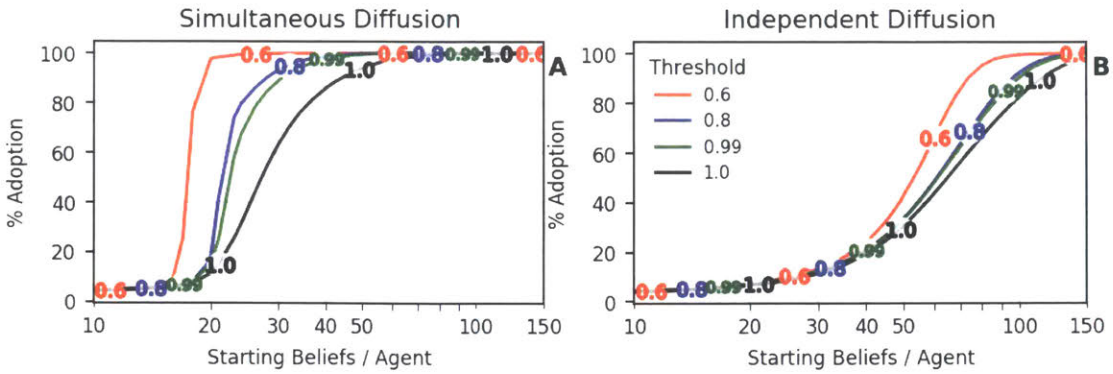


Fig. S12. Threshold adoption increases the mean extent of diffusion for a given set of seeds, under either simultaneous or independent diffusion.

Fig. S13 shows that when the threshold is high (0.99, 0.8), the introduction of threshold adoption allows the adoption of beliefs that are not connected to the existing semantic network; while the dynamics are still primarily driven by reciprocal facilitation and agreement cascades. The ability to adopt unconnected beliefs increases the importance of both reciprocal facilitation and agreement cascades, with the maximum impact of the mechanisms occurring with fewer seed beliefs. The increase in the impact on diffusion is a straightforward conclusion, as anything that supports the diffusion of a belief will lead to greater levels of adoption, which will be amplified by reciprocal facilitation. The impact on agreement cascades is to increase the ability of similar individuals to become yet more similar, as those whose neighbors all agree in their adoption of a belief will be able to adopt that belief as well.

As the threshold is lowered (0.6), diffusion is driven more by social considerations than semantic consistency. As we should expect, when beliefs are more likely to be adopted because they exceed the social threshold for adoption than because they are consistent with the agent's semantic network, adoption patterns start to look more like what we would see under threshold diffusion. The importance of reciprocal facilitation and agreement cascades declines. As we saw with the less discriminating decision rules in Fig. S10, when agents are less driven by a desire for internal consistency in their semantic networks, the dynamics that emerge from the interdependence between beliefs are less observable.

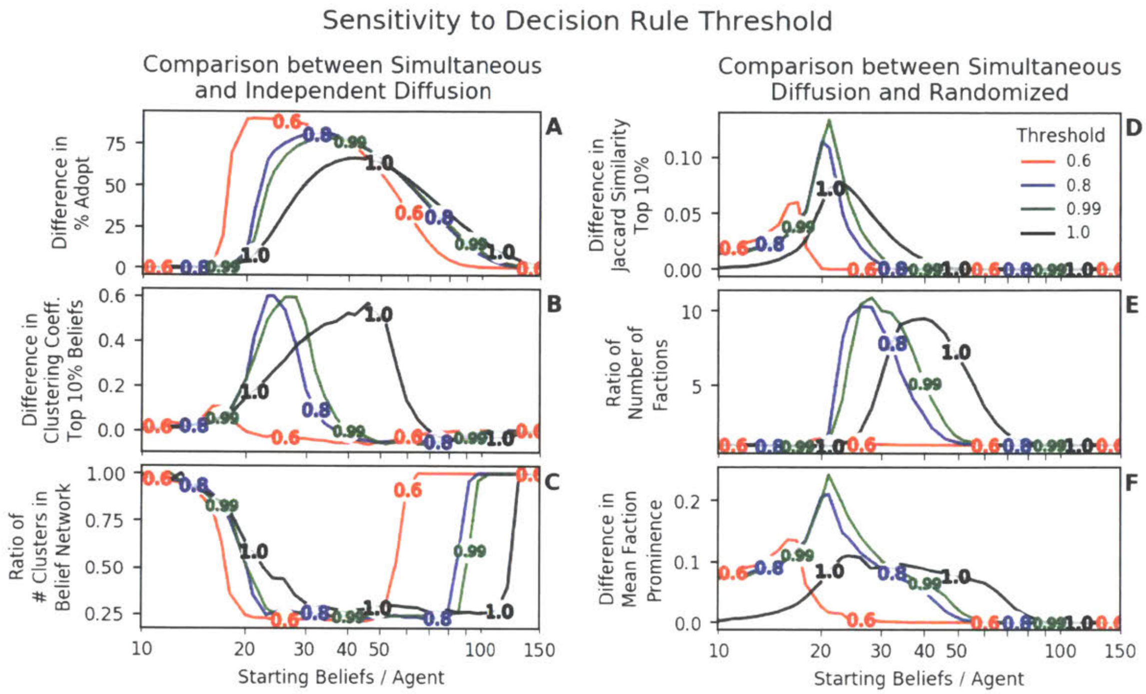


Fig. S13. When threshold adoption acts to allow access to beliefs whose diffusion is still primarily dependent upon consistency with individuals' semantic networks, it accentuates the effect of reciprocal facilitation and agreement cascades on belief clustering and social factionalism. When threshold adoption is the dominant mode of diffusion, reciprocal facilitation and agreement cascades become less relevant.

D: Simplified Model Code

The following (Python 3) code is sufficient to simulate the social contagion of interdependent beliefs, using the semantic network representation of beliefs and the triangle-closing adoption rule. I also include code to measure semantic network clustering, and social factionalism. A full listing of all code necessary to generate the figures in this paper, and faster ways of conducting the simulation and measures, is available at: https://github.com/JamesPHoughton/interdependent-diffusion.

The following code formalizes this paper's theory of the social contagion of interdependent beliefs. Lines 4-7 define the parameters for the simulation. Lines 9-27 initialize the social network, assign beliefs to agents, and cache a list of the beliefs present in the network. Lines 29-38 represent the decision rule that each agent uses to decide whether to adopt a belief. Lines 40-50 perform the simulation itself.

```python
import networkx as nx
import numpy as np

n_agents = 40  # number of agents in the population
deg = 4  # number of neighbors each agent has, on average
n_concepts = 30  # number of nodes in the complete belief space
n_beliefs = 25  # number of starting beliefs given to each individual

# create a randomly connected social network
# if it isn't connected, try again
connected = False
while not connected:
    g = nx.gnm_random_graph(n=n_agents, m=int(n_agents*deg/2))
    connected = nx.is_connected(g)

# give agents their initial beliefs
nx.set_node_attributes(
    g,
    name='M',  # set node attribute 'M' (for 'mind')
    # create a semantic network with a different random set of beliefs
    # for each agent, and assign to nodes in the social network
    values={agent: nx.gnm_random_graph(n_concepts, n_beliefs) for agent in g}
)

# capture a list of all the beliefs in the population
beliefs = np.unique([tuple(sorted(belief)) for agent in g
                     for belief in g.node[agent]['M'].edges()], axis=0)

# define a rule for whether an agent adopts a candidate belief
def adopt(g, agent, edge):
    try:  # there may be no path between the nodes
        length = nx.shortest_path_length(g.node[agent]['M'], *edge)
    except (nx.NetworkXNoPath, nx.NodeNotFound):
        length = 1000  # assume that length is very large

    exposure = np.mean([edge in g.node[neighbor]['M'].edges()
                        for neighbor in g[agent]])
    return length == 2 and exposure > 0 # triangle closing decision rule

# perform the simulation
n_steps = 10
for step in range(n_steps):
    # cycle through agents in random order
    for ego in np.random.permutation(g):
        # cycle through all possible beliefs in random order
        for edge in np.random.permutation(beliefs):
            # check whether the selected agent adopts the selected belief
            if adopt(g, ego, edge):
                # add the belief to the agent's semantic network
                g.node[ego]['M'].add_edges_from([edge])
```

The following code computes the various measures used in Figs. 2 and 3 in this paper.

```
import scipy.cluster.hierarchy as sch
from scipy.spatial.distance import squareform

# Measure the number of adopters of each belief
diffusion = dict()  # keys will be tuples of the belief
for edge in beliefs:
    # count the agents with the belief in their semantic network
    diffusion[tuple(edge)] = np.sum([edge in g.node[agent]['M'].edges()
                                     for agent in g])

# Measure the clustering coeff. of all beliefs above qth percentile
q = 90
# find index of qth percentile belief
thresh = int(np.ceil(len(diffusion)*(q/100)))
# rank beliefs from least to most popular
edges = sorted(diffusion, key=diffusion.get)
# create a subgraph with only the beliefs above threshold
subgraph = nx.from_edgelist(edges[thresh:])
# measure clustering of the subgraph
belief_clustering = nx.average_clustering(subgraph)

# Measure the number of separable clusters in aggregate semantic network
# identify unique values for extent of diffusion
levels = set(diffusion.values())
num_components = []
for level in levels:
    # identify the edges that are above the level
    edges = [belief for (belief, adopters) in diffusion.items()
             if adopters > level]
    # create a subgraph with only the beliefs above
    subgraph = nx.from_edgelist(edges)
    # count the number of components in the subgraph
    num_components.append(nx.number_connected_components(subgraph))
# find the maximum number of components discovered
num_semantic_clusters = np.max(num_components)

# Measure jaccard similarity between each pair of individuals
jaccards = dict()
# for each pair of agents in the simulation
for a, b in itertools.combinations(g.nodes, r=2):
    # identify the edges of each agent
    a_edges = set(g.node[a]['M'].edges())
    b_edges = set(g.node[b]['M'].edges())
    # jaccard similarity is the intersection divided by the union
```

```python
    intersect = len(a_edges.intersection(b_edges))
    union = len(a_edges.union(b_edges))
    jaccards[(a, b)] = intersect/union

# Measure the mean of similarities above qth percentile
# Find out what index represents the qth percentile individual
thresh = int(np.ceil(len(jaccards)*(q/100)))
# Average over all similarities above the qth percentile
mean_interpersonal_similarity = np.mean(sorted(jaccards)[thresh:])

# identify factions
# distance measure is used by hierarchical clustering (HC)
distances = 1-np.array(list(jaccards.values()))
# HC on the average distance btw sub-clusters
link = sch.linkage(distances, method='average')
# a peak is anywhere the HC pairs two individuals
peaks = np.argwhere(link[:,3]==2).flatten()
n_peaks = len(peaks)

# calculate the prominence of all factions
if len(peaks) > 1:
    # calculate matrix of distances between all agents
    sf = squareform(sch.cophenet(link))
    prominences = []
    # look at each peak
    for node, height in link[peaks,1:3]:
        distances = []
        for othernode in link[peaks,1]:
            # compare to each alternate peak
            if node == othernode:
                continue
            # extract cophanetic distance between peaks
            distance = sf[int(node), int(othernode)]
            distances.append(distance)
        #prominence is the smallest cophanetic distance
        #net the distance between the nodes making up
        #the current peak
        prominences.append(min(distances)-height)
    mean_prominence = np.mean(prominences)
else:
    # if there is only one peak, cophanetic distance
    # to nearest peak is equal to height -> net 0
    mean_prominence = 0
```