Interdependent Diffusion: The social contagion of interacting beliefs

James Houghton[a,b]

[a]Massachusetts Institute of Technology; [b]University of Pennsylvania

This manuscript was compiled on May 29, 2021

Social contagion is the process in which people adopt a belief, idea, or practice from a neighbor and pass it along to someone else. For over 100 years, scholars of social contagion have almost exclusively made the same implicit assumption: that only one belief, idea, or practice spreads through the population at a time (1–26). It is a default assumption that we don't bother to state, let alone justify. The assumption is so ingrained that our literature doesn't even have a word for "whatever is to be diffused" (1), because we have never needed to discuss more than one of them. But this assumption is obviously false. Millions of beliefs, ideas, and practices (let's call them "diffusants") spread through social contagion every day. To assume that diffusants spread one at a time – or more generously, that they spread independently of one another – is to assume that interactions between diffusants have no influence on adoption patterns. This could be true, or it could be wildly off the mark. We've never stopped to find out. This paper makes a direct comparison between the spread of independent and interdependent beliefs using simulations, observational data, and a 2400-subject laboratory experiment. I find that in assuming independence between diffusants, scholars have overlooked social processes that fundamentally change the outcomes of social contagion. Interdependence between beliefs generates polarization, irrespective of social network structure, homophily, demographics, politics, or any other commonly cited cause. It also coordinates structures of beliefs that can have both internal justification and social support without any grounding in external truth.

Social Contagion | Social Learning | Simulation | Experiment

On June 17th, 2015, the U.S. Treasury announced that a portrait of a woman would appear on the ten-dollar bill (27). The same day, news emerged of a mass shooting at a historically black church in Charleston, South Carolina (28, 29). Within twenty-four hours both stories had spread throughout U.S. social media. It is reasonable to study the spread of these diffusants independently of one another, because the probability that an individual will share news of the shooting is not likely to be causally influenced by whether they have understood news about the $10 bill, and vice versa.

The day after the Charleston shooting, two other news items emerged. The first was a report that the shooter had been motivated by racial hatred symbolized in the Confederate flag. The second was a call to remove the Confederate flag from the South Carolina state capitol grounds (28, 29). Even though these are distinct ideas, each spreading through social contagion, we cannot ignore one in trying to understand the diffusion of the other. If an individual has previously adopted the belief that the flag should be removed from the capitol grounds, they will be more likely to believe that the shooter's identification with the flag is politically relevant, and vice versa. Rather than being independent diffusants, these beliefs are interdependent.

With good reason, nearly all social contagion research assumes that diffusants spread independently of one another. The independence assumption makes for parsimonious theory (1–7, 10, 12–16, 20–22, 25, 26), and it reduces the complexity and expense of experiments (8, 9, 11, 17–19, 23, 24). The most influential authors on social contagion have used this assumption to study the effect of social network structure (1–11), social reinforcement (12–19, 26), homophily (20–22), and network rewiring (22–24) on contagion outcomes. Unfortunately, scholars assume independence between diffusants so frequently and to such productive ends that we generally forget we are doing so, and fail to question whether the assumption is appropriate.

In contrast with independent diffusion, "interdependent diffusion" describes any social contagion process in which individuals' likelihood of adopting diffusant A is a function of their current state of adoption of B (C, D, ...) and in which their likelihood of adopting B (C, D, ...) is a function of their state of adoption of A. Very few theoretical models include any form of interaction between diffusants (30–35), and none are empirically verified. It remains to be seen whether interdependence between diffusants 1) generates new sociological processes, 2) creates new observable outcomes, and 3) has practical consequences for communication and social policy. In short, does interdependence matter for the theoretical and empirical study of social contagion, or are our models of independent diffusion sufficient?

Theoretical development

Interdependent diffusion can be studied through two theoretical lenses. First, we will observe the spread of a "focal" belief as it participates in a process of "reciprocal facilitation" with other diffusants. Secondly, we will observe a pair of individuals as they exchange multiple beliefs with one another in a process we might call an "agreement cascade".


Significance Statement

When studying the spread of beliefs from person to person, scholars generally look at only one belief at a time. This paper shows that by focusing on a single belief, we miss interactions between beliefs that lead to polarization. Because polarization is such an important part of why we study social contagion in the first place, recognizing this oversight encourages us to reexamine much of what we know, and opens up a rich sub-field for future research.

A. Real World

1. Humans co-create a dynamic, directed, weighted, structured social network.

2. A belief statement asserts a relationship between two concepts.

Belief Statement

The shooting *was a case of* **racial violence.**

concept — relationship — concept
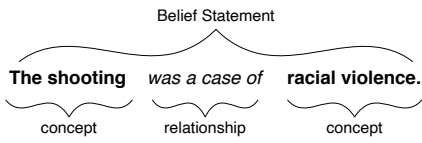
3. Each human has a unique set of beliefs.

- **The state government** *flies* **the Confederate flag.**
- **The Confederate flag** *symbolizes* **racial violence.**

- **The state government** *endorses* **racial violence.**
- …

4. A human is "exposed" to new beliefs by their neighbors, the media, or their own experience.

5. Some humans are susceptible to adopting a particular new belief, while others are not.

- **Independent diffusion** assumes that susceptibility is fixed *a priori.*

- **Interdependent diffusion** assumes that susceptibility is a function of other (possibly changing) beliefs.

Existing beliefs:
- **The state government** *flies* **the Confederate flag.**
- **The Confederate flag** *symbolizes* **racial violence.**

Likely candidate:
- **The state government** *endorses* **racial violence.**

Unlikely candidate:
- **The state government** *endorses* **shoplifting.**

6. Humans update their beliefs continuously and to varying degrees, depending on their exposure and susceptibility.

B. Abstraction

1. Simulating a static, undirected, unweighted, and unstructured social network isolates the effect of interdependent diffusion.

2. Concepts map to nodes in a "knowledge graph", and relationships become edges.

relationship

was a case of

The shooting — **racial violence**
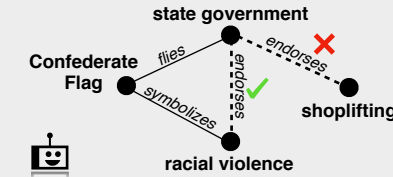
concept — concept
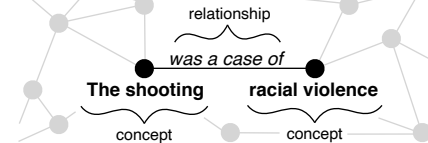
3. Each agent is assigned its own knowledge graph, initialized with random edges.
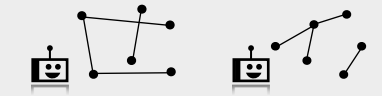
4. An agent is "exposed" to any beliefs held by their social network neighbors.

5a. **Independent diffusion condition**: Each agent is randomly assigned to be susceptible to each belief with probability *p.*

5b. **Interdependent diffusion condition**: An agent adopts beliefs that have short existing paths in its evolving knowledge graph.

state government

Confederate Flag — *flies* — *endorses* ✗

symbolizes — *endorses* ✓ — shoplifting

racial violence

6. Agents' beliefs are binary. Agents update their beliefs once per time period, in random order.

C. Implementation

```
social_network = Erdős-Rényi(
  n_agents = 60, mean_degree = 3,
  enforce_connected = True)
```

```
// Define all possible beliefs
B = complete_graph(n_nodes=25, n_edges=300)
beliefs = all edges in B
```

```
for each agent in social_network
  knowledge_graph = random_choice(
      options = beliefs,
      n_choices = 25,  // out of 300 total
      with_replacement = False
  )
  assign knowledge_graph to agent
```

```
function exposed(belief, agent)
  for each of agent's neighbors
    if belief in neighbor's knowledge_graph
      return True
  return False  // no neighbors hold belief
```

```
// Independent only, run once at start
for each agent in social_network
  for each belief in beliefs
    if random(0,1) <= p then
      assign agent susceptible to belief

// Interdependent only, run each decision
function susceptible(belief, agent)
  a, b = nodes in belief
  l = shortest path length from a->b in
      agent's knowledge_graph
  threshold = 2
  if l <= threshold then
    return True
  else
    return False
```

```
for step in (0-9)
  for shuffle each agent in social_network
    for shuffle each belief in beliefs
      if agent is exposed to belief and
         agent is susceptible to belief then
        add belief to agent's knowledge_graph
```

Fig. 1. Simulating interdependent diffusion as knowledge graph edges spreading over a social network. A. In the real world, the effect of interdependence on social contagion is enmeshed with other influences and complexities. **B.** A useful abstraction simplifies reality to isolate the mechanisms of interest, and **C.** allows us to simulate the behavior we wish to understand.

Reciprocal Facilitation. A focal belief spreads when an exposed individual's existing beliefs lead her to think it is true – that is, they "facilitate" her adoption of the focal belief. From its new position in the social network, the focal belief may then spread further, and subsequently facilitate the adoption of the beliefs that had previously supported its diffusion. This cycle creates a reinforcing feedback, such that when diffusants alternately create susceptibility to one another they can be adopted by more individuals than any single belief could have reached on its own. We see this "reciprocal facilitation" dynamic at work in beliefs about the confederate flag. The existing political conflict facilitated the spread of information about the shooter's identification with the flag, and as it spread, news of the shooter's identification with the flag brought more attention to its contentious display at the state capitol.

We can use a simple agent-based simulation (described in Fig. 1) to explore the macro-scale effects of reciprocal facilitation. In this simulation, agents populate a fixed, random social network, and are initially assigned a random set of beliefs. Each belief asserts a relationship between two concepts (*e.g.*

the shooting *was a case of* racial violence). By analogy to the "knowledge graph" interpretation of semantic networks[36–38], an agent's beliefs can be aggregated to form a graph in which nodes represent concepts (*e.g.* **people**, **places**, **items**, **activities**, etc. . .), and edges represent relationships that connect concepts to one another (*e.g. ownership*, *membership*, *co-location*, *usage*, etc. . .). When an agent adopts a new belief, it becomes a new edge in the agent's knowledge graph.

When an individual is exposed to new beliefs, she prefers to adopt those that connect two concepts that are already close together in her knowledge graph [39, 40], as doing so does not require dramatic changes to her belief structure. The simplest representation of this tendency is that a simulated individual will adopt any edge present in at least one neighbor's knowledge graph, as long as the distance it spans in her existing knowledge graph is below some threshold. This decision rule ignores social reinforcement (*i.e.* no complex contagion) and the exposer's attributes or other beliefs (*i.e.* no homophily or heterogeneous influence), and so for each belief is equivalent to a simple contagion model in all respects other than the influence of the

individual's internal state. Please see the **Methods** section and section S1 of the supplement for simulation details.

Figure 2 shows the average results of 20,000 simulations of 60 agents in the above model. Each agent is assigned random initial beliefs, and adopts beliefs from its social network neighbors that span up to two steps distance in its knowledge graph. First we see that under interdependent diffusion, reciprocal facilitation allows the average number of people susceptible to a belief to grow endogenously with the number of adopters (Fig. 2A, red). By comparison, if the same beliefs diffuse independently they can only be adopted by individuals who are susceptible at the start, and so they spread less widely in the population (Fig. 2A, black). Put another way: to explain the same level of final adoption, models of independent diffusion need to assume that significantly more people start out susceptible to each belief.

It might be reasonable for us to ignore interdependence (and instead assume more widespread initial susceptibility) were it not for the second effect of reciprocal facilitation. When diffusants do not interact, the number of people initially susceptible to a belief is an excellent predictor of how many will eventually adopt it (Fig. 2B). However, when beliefs are interdependent, individuals can acquire susceptibility by adopting other supporting beliefs, such that initial susceptibility is no longer a strong predictor of eventual adoption. For example, imagine that a "focus group" is selected from our artificial population before the simulation starts, and that amongst this group belief A is adopted by 25% more people than belief B. Unsurprisingly, if we simulate the spread of these beliefs independently of one another, belief A will be more popular than B in over 99% of cases. On the other hand, when we simulate interdependent diffusion, A is adopted by more people than B only 57% of the time – just slightly better than chance.
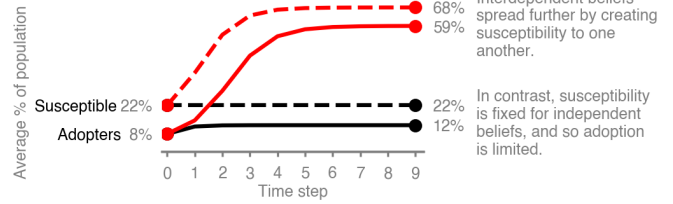
The popularity of interdependent beliefs is hard to predict because the reciprocal facilitation process does not support the diffusion of all beliefs to the same degree. Instead, it preferentially distributes those that are supported by other widely adopted beliefs. Fig. 2C shows how this reinforcing feedback leads a belief's popularity to become correlated with that of the most popular belief adjacent to it in the knowledge graph. This is because 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 only supported by unpopular beliefs will have trouble reaching even the few individuals who are susceptible to adopting it.

As a result, patterns emerge when individuals' knowledge graphs are aggregated to the level of the population. After interdependent diffusion, clusters of mutually-supporting beliefs are held by large fractions of the population (Fig. 2D), and they dictate which new beliefs can be adopted. This is quite remarkable, as *post-hoc*, individuals have both internal support for their beliefs (*i.e.* each belief is supported by many other beliefs) and external support (*i.e.* other individuals share both their beliefs and their justifications for believing so). This happens even though *a priori* we have no way to tell which sets of beliefs will become popular, and no ground truth.
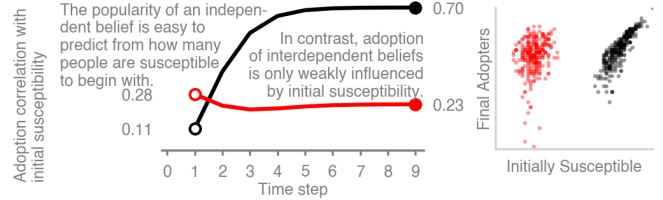
Reciprocal facilitation yields an emergent macro-scale outcome analogous to the micro-scale effect of confirmation bias, in which whole groups of people can collectively deceive them-
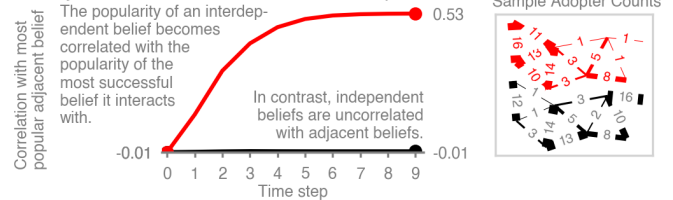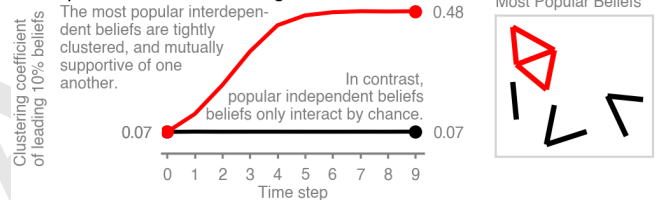

Fig. 2. Simulated effect of interdependent and independent diffusion on shared belief structures. A. Average popularity (number of adopters) of all beliefs over time, given an equal number of initially susceptible individuals in each condition. **B.** Correlation between the initial number of susceptible individuals and the popularity of each belief, given the same *final* distribution of adoption between conditions; detail shows $t = 9$ popularity vs. number of individuals initially susceptible. **C.** Correlation between the popularity of each belief and the popularity of the most popular belief it shares a 'node' with; sample $t = 9$ popularity of each belief overlaid on the knowledge graph. **D.** Clustering coefficient of a hypothetical knowledge graph made up of the most popular 10% of all beliefs; sample $t = 9$ network filtered on popularity.
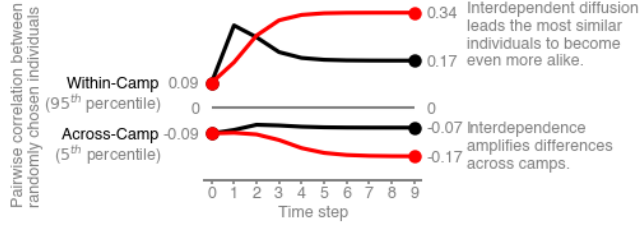
selves through the rational action of comparing new beliefs for consistency with what they already know. This effect doesn't require agents to have any prior preference for believing one thing over another, nor does it depend on any in-group vs. out-group sentiment, affinity for similar individuals, network structure, or social reinforcement. It's merely a phenomenon of the social contagion of interdependent beliefs.

Agreement Cascades. Switching lenses, we now focus on a pair of neighboring individuals, and observe a process of "agreement cascades". When two people exchange beliefs, they become more similar to one another. Because their existing beliefs influence the way people respond to new diffusants, shared beliefs make the two individuals more likely to adopt (or reject) the same new beliefs in the future. As a result, they become more similar still, regardless of any preference to align or distinguish themselves from one another.

It is helpful to contrast agreement cascades with homophily

Interdependent diffusion | Independent diffusion

A: "Camps" become more distinct

Pairwise correlation between randomly chosen individuals

Interdependent diffusion leads the most similar individuals to become even more alike. — 0.34
— 0.17

Within-Camp 0.09
(95th percentile)
0 / 0

Across-Camp -0.09
(5th percentile)

Interdependence amplifies differences across camps. -0.07
-0.17

Time step: 0 1 2 3 4 5 6 7 8 9

B: Beliefs align along an axis

Percent of variation falling along a single axis

When beliefs interact, individuals align along a left-right axis and beliefs become associated with each other. 12%

6%

In contrast, independent beliefs are more evenly mixed in the population, and a left-right axis is less obvious. 3%
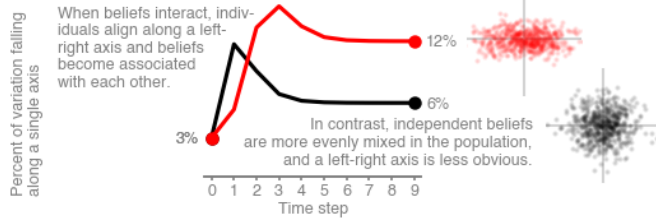
Time step: 0 1 2 3 4 5 6 7 8 9

Fig. 3. Simulated effect of interdependent and independent diffusion on social polarization. A. Correlation between beliefs of randomly sampled pairs of individuals, 95th (5th) percentile represents relationships conservatively within (across) ideological "camps". **B.** An individual's beliefs locate her in a space with one dimension for each possible belief (300). The percent of variation explained by first principal component of this space describes how well individuals map to a "political spectrum". Offset diagrams are exaggerated and show a larger population to illustrate how a component can explain more or less variation.

(*i.e.* an individual's tendency to give extra weight to the opinions of more similar neighbors ([35])). When homophily is active, an individual is (consciously or unconsciously) aware of their similarity to a neighbor, and allows that assessment of similarity to influence their adoption decisions. In contrast, agreement cascades increase an individual's likelihood of adopting a belief from a similar neighbor, even if she is blind to every other belief her neighbor possesses. Homophily increases the attractiveness of all beliefs from a similar neighbor, while the same belief shared by a dissimilar neighbor would be less attractive. By contrast, under interdependent diffusion each belief has a particular likelihood of adoption regardless of its source, but agreement cascades mean that similar neighbors provide a higher concentration of beliefs that an individual is likely to adopt.

Fig. 3A shows the similarity between pairs of highly similar individuals (whom we might describe as being in the same "ideological camp") and highly dissimilar individuals (different camps). Under interdependent diffusion, agreement cascades lead camps to become increasingly self-similar as they expand members' access to beliefs they can hold in common, and filter out beliefs that would set them apart from one another. Moreover, differences between camps are amplified as existing beliefs drive dissimilar individuals to adopt beliefs from different parts of the belief space.

As individuals organize into camps, it becomes easier to predict each person's position on one belief from their position on other beliefs. For example, belief A may co-occur with belief B 70% of the time, but co-occur with belief C only 10% of the time. Interdependent diffusion compresses the population's variation in the space of beliefs into a few principal axes (*e.g.* liberal-conservative, or libertarian-populist). As a result, individuals can be more easily described by their position along a "left-right axis", as seen in Fig. 3B, with individuals

at any point on the axis relatively similar to one another. 210
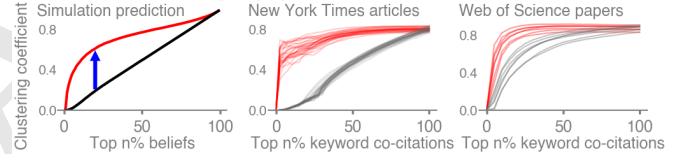
Consistency with observational data 211

If reciprocal facilitation and agreement cascades are active in the real world, we should be able to observe the simulation's predictions in real-world data. For example, when a document asserts a connection between two concepts, then the most popular connections drawn from a corpus of related documents should exhibit more clustering than we expect by chance. Figure 4A shows this effect in article keywords from the New York Times ([41]) and academic paper keywords listed in the Web of Science ([42]), over the full range of possible thresholds for a "popular" keyword co-occurrence. 212–221

The simulation also predicts that the most similar individuals in the population should be more alike than expected by chance, and those who are dissimilar more different. Figure 4B shows that this prediction is borne out in surveys of political opinions by the Pew Research Center ([43]) and in surveys of social values by the World Values Survey ([44]). 222–227
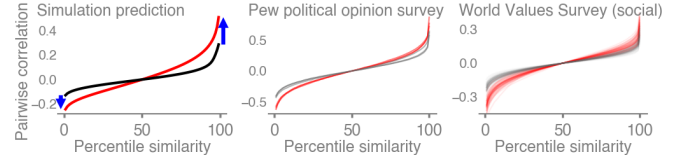
Finally, the simulation predicts that the variation between individuals in these surveys should be more aligned with a few principle axes of polarization than we expect by chance. Figure 4C shows this to be true. Please see the **Methods** section and section S2 of the supplement for analysis details. 228–232

Interdependent/Data | Independent | Expected

A. Increased clustering of popular beliefs

Clustering coefficient

Simulation prediction — Top n% beliefs
New York Times articles — Top n% keyword co-citations
Web of Science papers — Top n% keyword co-citations

B. Amplified similarity and dissimilarity at extremes

Pairwise correlation

Simulation prediction — Percentile similarity
Pew political opinion survey — Percentile similarity
World Values Survey (social) — Percentile similarity

C. Increased alignment along primary axes

Percent variation

Simulation prediction — Principal component
Pew political opinion survey — Principal component
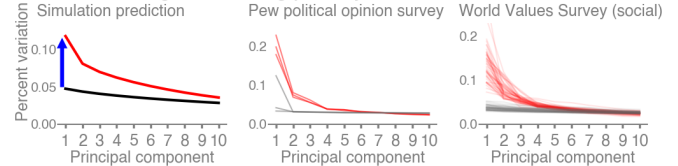World Values Survey (social) — Principal component

Fig. 4. Observing the predicted effects of interdependent diffusion in real-world data. A. The clustering coefficient of a hypothetical knowledge graph constructed from the $n\%$ most popular beliefs, for $n \in (0, 100)$. Interdependence is predicted to increase clustering among the most popular connections between concepts. This increase is observed in keyword co-citation networks from 289k New York Times articles (curves for each of 27 sections)([41]), and 47k academic papers from Web Of Science (7 fields)([42]), compared to a randomized baseline. **B.** Correlation in beliefs between randomly selected pairs of individuals. Interdependence is predicted to amplify extreme values of similarity and dissimilarity. This amplification is observed among 10k responses to the Pew Political Polarization Survey (3 waves)([43]), and 71k responses to questions about social values in the World Values Survey (49 countries)([44]). **C.** The percentage of variation present in the first 10 components of a principal component analysis. Interdependence is predicted to increase the variance explained by the first few components. This increase is observed in the Pew and World Values surveys.

Testing predictions of population-level outcomes in a randomized controlled experiment

The above simulation describes two mechanisms involved in the social contagion of interacting beliefs, and predicts outcomes that are observed in real world data. However, this model is a simplification of reality, and like all such models may mischaracterize human behavior in important ways. Similarly, observational analysis is limited in this case by the lack of an independent counterfactual; it can only show consistency with the predicted outcomes, rather than causal evidence of the mechanisms themselves. To rigorously test the simulation's predictions, I conducted a fully preregistered randomized-controlled experiment with 2400 participants. In an online laboratory context, I systematically varied the level of interaction between otherwise identical diffusants in identically constructed populations, and measured macro-scale polarization in behavior and self-reported beliefs.

The experiment took the form of a "detective game" in which participants were given clues to a burglary and were asked to sort those clues into "Promising Leads" and "Dead Ends", using the interface in Fig. 5A. When a participant categorized a clue as a promising lead, it was immediately shared with her three neighbors in a 20-person social network, and so could diffuse through the network.

In our simulation model, agents could be programmed to either pay attention to interactions between beliefs or to ignore them. Human beings, on the other hand, are wired to see connections between ideas, and so it is impossible to conduct a perfect test in which all clues interact in one experimental condition, and yet *the same clues* are fully independent in another. To approximate the ideal manipulation, I partitioned the clues such that some (22 clues) were common to both interdependent and independent conditions and were used for analysis, while the remainder (55 clues) varied across conditions to manipulate the level of interaction between analysis clues. In both conditions, analysis clues linked a crime scene and stolen object to three suspects, two articles of clothing, two physical descriptions, two tools, and two vehicles (Fig. 5B). In the interdependent condition, "cross-link" clues connected each suspect, vehicle, etc. to one another (Fig. 5C), to create a fully-connected knowledge graph. In the independent condition, "filler" clues gave additional, non-interacting details (Fig. 5D) that allowed analysis clues to remain as independent as possible. Clues were designed to equally implicate any suspect or burglary method, and each was randomly assigned to exactly one player at game start.

The experiment was run in 30 blocks of matched treatment/control games, each block using a unique set of randomly generated clues. Each social network position was randomly assigned four clues, common (apart from the manipulation) across conditions. Each participant was randomly assigned to a condition and social network position. Groups played for eight minutes before estimating the likelihood that each suspect, vehicle, etc. was involved in the crime.

The measures of similarity and left-right axis alignment used in Fig. 3 were computed based on participants' behavior and self-reported opinions. "Behavioral" measures were constructed from each participant's final categorization of the 22 analysis clues. "Self-report" measures were constructed from the 11-item post-game assessments, reflecting how participants internalized the information they encountered to create
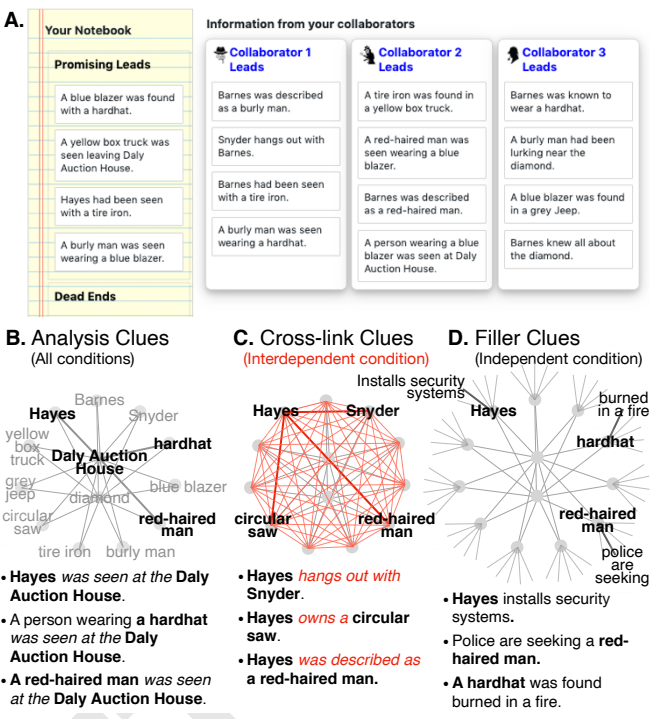

Fig. 5. The Detective Game. A. Player Interface. *"A priceless diamond has been stolen from the Daly Auction House. To solve the mystery, drag clues that you think are true into the "Promising Leads" section of your notebook, and clues you think are false into the "Dead Ends" section. You are rewarded for each correct 'lead' and penalized for each incorrect 'lead'. There is no reward or penalty for 'dead ends'."* **B.** 22 "analysis" clues link the crime scene and stolen object to three suspects, two articles of clothing, two descriptions, two tools and two vehicles. **C.** In the interdependent condition, 55 "cross-link" clues connect each analysis clue to every other analysis clue. **D.** In the independent condition, "filler clues" take the place of cross-link clues, allowing analysis clues to remain independent of one another.

opinions. As our simulation gives a strong theoretical prior for the direction of each effect, I used uncorrected pairwise one-sided t-tests to assess how each measure differed between baseline and treatment conditions. For more details, please see the **Methods** section and supplement section S3.

In a testament to the pervasiveness of belief interaction, even this manipulation (and ideal laboratory conditions) could not create a control condition with perfectly independent beliefs. For example, even though the independent clue set contained no explicit links between suspects, participants could draw implicit connections between the guilt of one suspect and the presumed innocence of another. This imperfect control means that measured differences between independent and interdependent conditions are likely to underestimate the true effect of belief interaction.

Nevertheless, the results of this experiment support the above macro-level predictions. Figure 6A shows that interdependence measurably increased the population's alignment along a left-right axis among both behavioral ($+2.1\%$, $p = .013$, 90% $CI = (0.6, 3.7)$) and self-reported ($+2.8\%$, $p = .022$, $CI = (0.6, 5.0)$) measures of belief. While not all measures were significant, camps were more self-similar (behavioral measures: $+.026$, $p = .015$, $CI = (0.008, 0.045)$, Fig. 6B) and more distinct from one another (self-report measures: $-.034$, $p = .058$, $CI = (-0.07, 0.0015)$, Fig. 6C) in the interdependent condition than in the independent condition.

Houghton

PNAS | **May 29, 2021** | vol. XXX | no. XX | **5**

Comparison effect of social network structure. To gauge whether the effect of interdependence is large enough to be worth attention when compared to other drivers of polarization, I ran a parallel experimental condition that fixed diffusants to be independent and varied the social network structure between non-polarizing and polarizing extremes. The baseline was a "dodecahedral" network (inset diagram in Fig. 6A), in which none of a participant's neighbors were directly connected to any other (network clustering coefficient 0.0), and the average social network distance between individuals was short (2.6 steps). We should expect to find very little polarization in this network, as information can diffuse across the network readily, and coordination among subgroups is impeded by the lack of mutual connections. The second social network was a "regular connected caveman" structure, in which neighbors shared 50% of their remaining contacts in common (clustering coefficient 0.5), and there are large average distances between individuals (4 steps). We should expect to find high levels of polarization in this network regardless of the level of interaction between clues, as strong clustering makes it easy for subgroups to converge on a shared set of clues, and long average path lengths make it harder for information to spread between camps.
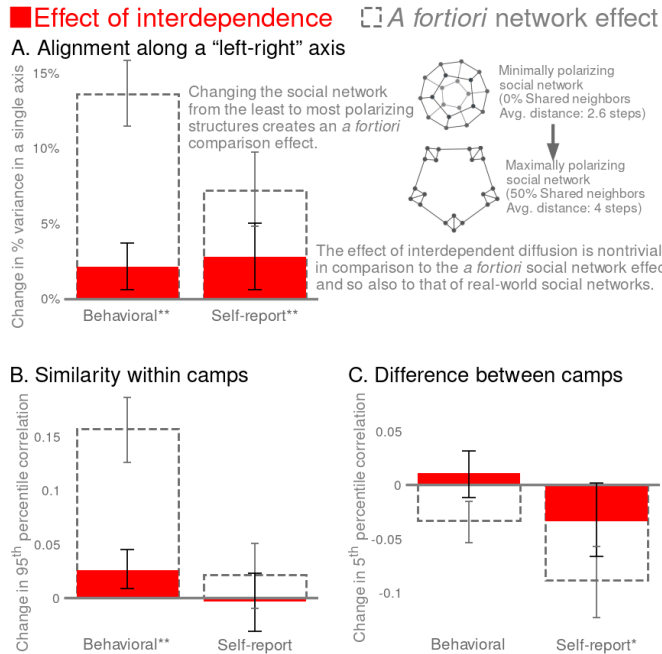

Fig. 6. **Testing the macro-scale effects of interdependent diffusion in an online lab experiment.** The predicted effects of belief interaction are borne out in experiment, with meaningful effects in comparison to an *a fortiori* comparison that varied the social network structure underlying independent diffusion. Paired T-test $^*p < .1$, $^{**}p < .05$. Error bars show 90% bootstrap CI. $n = 30$ pairwise comparisons.

Real-world social networks lie somewhere between these two extremes, and so this social network manipulation creates an *a fortiori* comparison effect. If the effect of interdependence is nontrivial in this comparison, then we have confidence that it is meaningful in the real world. In the statistically significant measures of this experiment, interdependent diffusion creates an effect between 16% and 39% of the *a fortiori* reference, as shown in Fig. 6. We should always use caution when generalizing effect sizes from laboratory groups to large-scale social networks. However, this comparison suggests that interdepen-

dence plays an unignorable role in social contagion, and that scholars' almost exclusive focus on network structure over belief interaction is out of proportion to the relative importance of the two effects. Please see the suplement section 3.6 for comparison details.

Discussion

While it cannot explore all of the implications of interdependent diffusion, this paper demonstrates that belief interaction enables at least two new sociological processes that influence polarization and shared belief structures. The social contagion literature is replete with studies that explain polarization as a consequence of social network structure (45, 46), homophily (35, 47, 48), complex contagion (49, 50), or other personal or network characteristics. While they certainly influence polarization, this paper shows that when beliefs interdepend, none of these explanations are necessary for polarization to emerge - polarization is a natural consequence of diffusion itself. Likewise, we have ample research asking how fringe beliefs and conspiracy theories emerge and persist (49–51). Interdependent diffusion suggests that they are a natural consequence of social contagion.

Given that truly independent diffusion is likely to be rare among real world beliefs, we can no longer assume that consensus and truth should be the default outcome of social contagion, absent other polarizing factors. No matter what changes we make to the social network structure, or how we reduce interpersonal bias, we should still expect polarization to emerge. This motivates several new research questions: What are the conditions under which populations can overcome the polarizing effect of interdependent diffusion to reach consensus? How can populations avoid deceiving themselves about the truth of the world? Answering these questions will require us to drop the assumption of independence and confront the complex reality of belief interaction in social contagion.

This simulation and experiment remind us that the simplifying assumption of independence between diffusants is just that – an assumption. Despite its ubiquity, the assumption should be carefully made and frequently challenged. With luck we will find that most of what we know about diffusion is robust to interaction between diffusants. We may also find that relaxing the assumption of independence helps us explain new sociological phenomenon and better understand social contagion.

Materials and Methods

All materials used to create this paper (including code for running simulations, observational analysis, designing and conducting the experiment, and analyzing experimental results, along with anonymized, timestamped experimental data) is available at: https://github.com/JamesPHoughton/interdependent-diffusion.

Simulation. In the simulation presented in **Figs. 2 and 3**, the social network is a connected Erdős–Rényi (G_{nm}) random graph with 60 agents, each with an average of 3 neighbors. Each agent is initialized with 25 beliefs (edges) selected randomly from the 300 edges available in a complete knowledge graph with 25 concepts (nodes). These values ensure good coverage of beliefs in the population, while individual knowledge graphs are initially sparse. Random seeding ensures that the simulation starts without polarization or systematic variation in belief popularity, and also that the social network structure itself does not contribute to polarization. Because beliefs are drawn from a complete knowledge graph, there is no natural

belief structure around which polarization may nucleate. Results are qualitatively similar with other types and sizes of social networks and different sizes and seeding densities of initial knowledge graphs, so long as there are enough beliefs seeded initially for diffusion to occur and not so many that adoption is complete.

In each step, individuals are selected in random order and update their beliefs by incorporating into their knowledge graphs all beliefs (edges) that their neighbors possess and they are susceptible to adopting. In the interdependent case, individuals are susceptible to any belief with an existing path length of 2 (*i.e.* that closes a triangle) in their knowledge graphs at the current time. In the comparison (independent) case for **Fig. 2A**, a random selection of the population is defined to be susceptible to each belief in the same proportion as are initially susceptible to the belief in the interdependent diffusion condition. In **Figs. 2B, 2C, 2D and 3**, a random selection of susceptible individuals is made in proportion to the final number of susceptible individuals in the interdependent diffusion case. As a result, for all graphs other than **Fig. 2A**, a histogram of the extent of diffusion of each belief is approximately the same under both independent and interdependent treatments. This ensures that the subsequent presentation of results reflects purely the effect of interdependence between diffusants, and not the effect of different levels of adoption in the compared populations. Results are qualitatively similar when calculated based upon the initial susceptibility.

Measures. In **Fig. 2A**, the measure of adopters is the number of individuals with each belief in their knowledge graph, averaged over all beliefs, divided by the total population. Similarly, the measure of the susceptible population (and all discussion of the susceptible population) represents the fraction of individuals who would adopt each belief if exposed to it according to the appropriate decision rule for independent vs. interdependent diffusion, plus the fraction that has already adopted the belief. **Fig. 2B** shows the Pearson correlation between the number of people who have adopted each belief at time t and the number who were initially susceptible to the belief at $t = 0$ but did not start with it. As this has no meaningful value at $t = 0$, the curve is drawn from t1-t9. **Fig. 2C** assesses the correlation between the number of individuals who have adopted each belief (knowledge graph edge) and the number who have adopted the most popular belief it shares a 'node' with, averaged over all beliefs. **Fig. 2D** uses the clustering coefficient of a knowledge graph constructed from the most popular 10% of beliefs as a demonstration that the most popular beliefs are mutually interrelated, and not merely all related to a single leading belief (*e.g.* a star or barbell pattern). Clustering only makes sense when beliefs are conceptualized as some form of semantic network. Other conceptualizations of belief interaction might prefer to plot the number of top decile beliefs that each top decile belief interacts with. This measure gives essentially the same result (*i.e.* large fractional growth over time in the interdependent case, with no change from randomness in the independent case) but fails to capture the mutual inter-relatedness indicated by the clustering coefficient. The measure is generally insensitive to the specific threshold used to define a 'popular' belief for any thresholds between about 5% and 40%. See the supplement for sensitivity analysis.

There are many complex measures of polarization in the literature (30, 31, 33, 45–51, for a sample), which generally attempt to represent three basic intuitions. First, that individuals within the same ideological camp come to be more similar to one another. Secondly, that individuals in different ideological camps become more dissimilar to one another. Lastly, that an individual's position on one dimension of belief becomes informative of their position on other dimensions. As my purpose is not to identify camps and their members, but to suggest that one set of conditions is more generative of polarization than another, these measures add more complexity than value. Instead I report heuristic measures characterizing the above three intuitions.

A simple and reproducible way to assess the similarity of individuals within an ideological camp – absent exogenous labels such as demographic or party – is to measure the similarity between all pairs of individuals and define a certain percentile as belonging to the same ideological camp. The more exclusive we are (*i.e.* the higher the percentile), the more conservative the claim that these represent "within-camp" relationships. To define across-camp similarity, we can choose a percentile that (conservatively) represents relationships between individuals in different ideological camps. In **Figs. 3A, 6B, and 6C**, I use the 95th and 5th percentiles respectively. See the supplement for sensitivity analysis.

Figs. 3B, 4C and 6A measure belief alignment as the percent of variation between individuals that can be explained by the best fitting axis in the space of possible beliefs, using singular-value decomposition. In **Fig. 3B**, the original feature space has one dimension for each belief in the simulation (300), and points representing each individual's position in that feature space (60) according to the beliefs they have adopted. The offset graphs are exaggerated and show a larger population to illustrate how a component can explain more or less variation. In **Fig. 6A**, the feature space has 22 dimensions for the behavioral measures, and 11 dimensions for self-report measures (described below), and points representing 20 individuals.

Analysis of observational data. The first column in **Fig. 4** shows simulated measures of belief clustering, inter-subject similarity, and percent variance in a given principal component over a range of thresholds and principal components, using the same simulation parameters described previously.

In order to assess clustering of popular beliefs in **Fig. 4A**, a dataset needs to include a large diversity of connections between concepts. New York Times article data is taken from the KPTimes dataset(52), including 13 years of articles for all categories with $>=$ 3000 documents. One line is shown for each of 27 news sections, each computed independently. When multiple categories were indicated, a document was assigned to the smallest category with $>= 3000$ documents. A weighted undirected graph was constructed for each category, in which article keywords formed nodes, and edges were weighted by the number of articles tagged with the two keywords in the edge. An alternative method in which edges are constructed by random draws without replacement gives the same qualitative effect, and is detailed in the supplement. Data on Web of Science paper keywords(42) used the same analysis, showing one line for each of 7 major academic fields.

In order to assess similarity and political axis alignment in **Figs. 4B and 4C**, a dataset needs dense information about individual beliefs. Political opinion data comes from 58 political opinion questions from the Pew Political Typology Survey(43), conducted in three waves in early 2014. The social values analysis uses 144 questions regarding non-political social values in the World Values Survey(44) collected from 2017-2020. Missing, unknown, and refused responses ($< 5\%$) are filled with the mean values for each question for each wave or country. As in the simulations above, similarity between individuals is assessed as the Pearson correlation of their responses to survey questions. Similarity curves are median-shifted in order to cleanly show the effect on all waves in the same plot. Principal components are computed from a space initially consisting of one dimension for each survey question. A full list of survey questions used is available in the supplement.

As we have no counterfactual world in which diffusion is assuredly independent, I compare the observed data to a synthetic control generated by shuffling beliefs among documents or individuals.

Experiment. The preregistration for this experiment included all code necessary to conduct the experiment using Empirica(53) and perform all statistical analyses reported in this paper. The preregistration is available at https://osf.io/239ns.

Over the course of eight days in July 2020, I recruited 2768 U.S. and Canadian Mechanical Turk workers under the criteria that they were 18+ years old, and have completed 100+ HITs with a 90%+ approval rating. Of these, 2400 completed training and were randomized into 20-person social networks. Each network was assigned to one of four (matched) experimental conditions, yielding $n = 30$ samples per condition. This sample size was set by budget constraints. The participant population was 45% female; mean 37.1 years old; 27% high-school, 49% bachelors, 16% masters+ graduates. 96.8% of players who completed training went on to complete all steps of the experiment, with less than 0.4% difference in dropout between conditions. The average payout was just over $4, and the experiment took about 20 minutes.

Similarity between self-reported measures is assessed using Pearson's correlation on the vectors of individuals' responses. This

Houghton

PNAS | **May 29, 2021** | vol. XXX | no. XX | **7**

measure has the advantage of being easily interpretable and having a well-defined range that is independent of the number of features in the vector of attributes being compared, and the negative region of which can be interpreted as expressing dissimilarity. To assess the similarity of the binary "behavioral" data I use the Phi coefficient, an analogous measure to Pearson's correlation with the same interpretable range.

The behavioral measures in the experiment are sensitive not only to interdependence and network structure but also to the average level of diffusion of beliefs. To minimize noise due to differences in the level of activity between games, each of the behavioral measures is assessed compared to what would be expected due to chance, keeping the number of adopters of each clue and the number of clues adopted by each participant fixed. This correction was designed in simulation and preregistered.

1. MS Granovetter, The strength of weak ties. *Am. journal sociology* **78**, 1360–1380 (1973).
2. RS Burt, *Structural holes.* (Harvard university press), (1992).
3. DJ Watts, SH Strogatz, Collective dynamics of 'small-world'networks. *nature* **393**, 440–442 (1998).
4. DJ Watts, A simple model of global cascades on random networks in *The Structure and Dynamics of Networks.* (Princeton University Press), pp. 497–502 (2011).
5. R Reagans, B McEvily, Network structure and knowledge transfer: The effects of cohesion and range. *Adm. science quarterly* **48**, 240–267 (2003).
6. D Kempe, J Kleinberg, É Tardos, Maximizing the spread of influence through a social network in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* pp. 137–146 (2003).
7. D Lazer, A Friedman, The network structure of exploration and exploitation. *Adm. science quarterly* **52**, 667–694 (2007).
8. J Travers, S Milgram, An experimental study of the small world problem in *The structure and dynamics of networks.* (Princeton University Press), pp. 130–148 (2011).
9. M Kearns, S Suri, N Montfort, An experimental study of the coloring problem on human subject networks. *science* **313**, 824–827 (2006).
10. D Acemoglu, A Ozdaglar, Opinion dynamics and learning in social networks. *Dyn. Games Appl.* **1**, 3–49 (2011).
11. S Suri, DJ Watts, Cooperation and contagion in web-based, networked public goods experiments. *PloS one* **6**, e16836 (2011).
12. G Le Bon, *The crowd: A study of the popular mind.* (TF Unwin), (1897).
13. MH DeGroot, Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).
14. TC Schelling, Sorting and mixing. *Micromotives macrobehavior* (1978).
15. M Granovetter, Threshold models of collective behavior. *Am. journal sociology* **83**, 1420–1443 (1978).
16. D Centola, M Macy, Complex contagions and the weakness of long ties. *Am. journal Sociol.* **113**, 702–734 (2007).
17. MJ Salganik, PS Dodds, DJ Watts, Experimental study of inequality and unpredictability in an artificial cultural market. *science* **311**, 854–856 (2006).
18. J Lorenz, H Rauhut, F Schweitzer, D Helbing, How social influence can undermine the wisdom of crowd effect. *Proc. national academy sciences* **108**, 9020–9025 (2011).
19. L Muchnik, S Aral, SJ Taylor, Social influence bias: A randomized experiment. *Science* **341**, 647–651 (2013).
20. CR Shalizi, AC Thomas, Homophily and contagion are generically confounded in observational social network studies. *Sociol. methods & research* **40**, 211–239 (2011).
21. B Golub, MO Jackson, How homophily affects the speed of learning and best-response dynamics. *The Q. J. Econ.* **127**, 1287–1338 (2012).
22. NA Christakis, JH Fowler, Social contagion theory: examining dynamic social networks and human behavior. *Stat. medicine* **32**, 556–577 (2013).
23. DG Rand, S Arbesman, NA Christakis, Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl. Acad. Sci.* **108**, 19193–19198 (2011).
24. A Almaatouq, et al., Adaptive social networks promote the wisdom of crowds. *Proc. Natl. Acad. Sci.* **117**, 11379–11386 (2020).
25. S Galam, S Moscovici, Towards a theory of collective phenomena: Consensus and attitude changes in groups. *Eur. J. Soc. Psychol.* **21**, 49–74 (1991).
26. R Hegselmann, U Krause, , et al., Opinion dynamics and bounded confidence models, analysis, and simulation. *J. artificial societies social simulation* **5** (2002).
27. J Calmes, Treasury says a woman's portrait will join hamilton's on the $10 bill. *New York Times* (2015).
28. JP Houghton, et al., Beyond keywords: Tracking the evolution of conversational clusters in social media. *Sociol. Methods & Res.* **48**, 588–607 (2019).
29. JB Hawes, *Grace Will Lead Us Home: The Charleston Church Massacre and the Hard, Inspiring Journey to Forgiveness.* (St. Martin's Press), (2019).
30. D Baldassarri, P Bearman, Dynamics of political polarization. *Am. sociological review* **72**, 784–811 (2007).
31. D DellaPosta, Y Shi, M Macy, Why do liberals drink lattes? *Am. J. Sociol.* **120**, 1473–1511 (2015).
32. NE Friedkin, AV Proskurnikov, R Tempo, SE Parsegov, Network science on belief system dynamics under logic constraints. *Science* **354**, 321–326 (2016).
33. A Goldberg, SK Stein, Beyond social contagion: Associative diffusion and the emergence of cultural variation. *Am. Sociol. Rev.* **83**, 897–932 (2018).
34. F Xiong, Y Liu, L Wang, X Wang, Analysis and application of opinion model with multiple topic interactions. *Chaos: An Interdiscip. J. Nonlinear Sci.* **27**, 083113 (2017).
35. R Axelrod, The dissemination of culture: A model with local convergence and global polarization. *J. Confl. Resolut.* **41**, 203–226 (1997).
36. JF Sowa, Semantic networks. (1987).
37. AM Collins, EF Loftus, A spreading-activation theory of semantic processing. *Psychol. review* **82**, 407 (1975).
38. R Popping, Knowledge graphs and network text analysis. *Soc. Sci. Inf.* **42**, 91–106 (2003).
39. MA Schilling, A" small-world" network model of cognitive insight. *Creat. Res. J.* **17**, 131–154 (2005).
40. RS Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. general psychology* **2**, 175–220 (1998).
41. Y Gallina, F Boudin, B Daille, KPTimes: A large-scale dataset for keyphrase generation on news documents in *Proceedings of the 12th International Conference on Natural Language Generation.* (Association for Computational Linguistics, Tokyo, Japan), pp. 130–135 (2019).
42. K Kowsari, et al., Hdltex: Hierarchical deep learning for text classification in *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on.* (IEEE), (2017).
43. Beyond red vs. blue: The political typology, (Pew Research Center), Technical report (2014).
44. European values study and world values survey: Joint evs/wvs 2017-2021 dataset (2020).
45. A Flache, MW Macy, Small worlds and cultural polarization. *The J. Math. Sociol.* **35**, 146–176 (2011).
46. M Del Vicario, et al., Echo chambers: Emotional contagion and group polarization on facebook. *Sci. reports* **6**, 1–12 (2016).
47. P Dandekar, A Goel, DT Lee, Biased assimilation, homophily, and the dynamics of polarization. *Proc. Natl. Acad. Sci.* **110**, 5791–5796 (2013).
48. VV Vasconcelos, SA Levin, FL Pinheiro, Consensus and polarization in competing complex contagion processes. *J. Royal Soc. Interface* **16**, 20190196 (2019).
49. D Spohr, Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Bus. Inf. Rev.* **34**, 150–160 (2017).
50. P Törnberg, Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PloS one* **13**, e0203958 (2018).
51. G Pennycook, DG Rand, Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
52. Y Gallina, F Boudin, B Daille, Kptimes: A large-scale dataset for keyphrase generation on news documents (2019).
53. A Almaatouq, et al., Empirica: a virtual lab for high-throughput macro-level experiments. *Behav. Res. Methods*, 1–14 (2021).