Synthese (2023) 201:211
https://doi.org/10.1007/s11229-023-04163-w

ORIGINAL RESEARCH

Influence theory

Patrick Grim[1,2] · Nicholas Rescher[3]

Received: 20 September 2022 / Accepted: 14 April 2023 / Published online: 8 June 2023

Abstract
Influence theory is a systematic study of formal models of the communicative influence of one person or group of people on another person or group. In that sense influence theory is an overarching philosophical discipline that includes aspects of decision theory and game theory as sub-disciplines as well as established models of de facto segregation, cultural change, opinion polarization, and epistemic networks. What we offer here is a structured outline of formal results that have been scattered across a range of disciplinary contexts from mathematics, physics and computer science to economics and political science, supplemented with a number of new models, emphasizing their place within the philosophical framework of a general theory of influence. What such an outline offers, we propose, is the prospect of new and important cross-fertilizations and expansions in formal attempts to model the diverse patterns of communicative influence.

Keywords Influence · Opinion dynamics · Communication · Modeling · Undecidability

1 Introduction

Fundamentally related at a deep philosophical level are bits and pieces of results and explorations that have been scattered across a range of diverse disciplines: economics, sociology, social psychology, epidemiology, physics, political science, computer science, mathematics, and complex systems. When systematically considered, we will argue, it becomes clear that these can be profitably considered together and pursued as parts of a unified philosophical sub-discipline: influence theory.

✉ Patrick Grim
patrick.grim@stonybrook.edu

1 Stony Brook University, Stony Brook, NY, USA

2 University of Michigan, Ann Arbor, MI, USA

3 University of Pittsburgh, Pittsburgh, PA, USA

Influence theory can profitably be contrasted with *inference* theory. Inference theory, commonly thought of as including logic, examines how one fact or body of facts (the premises) require or obstructs accepting another (the conclusion). Influence theory relates to a different sector of information management. It addresses the broader, statistical issue of how one exponent's acceptance or rejection of certain factual claims impinges on the situation of others. Standard information management addresses the substantive relations of the factual claims themselves. Influence theory addresses the implication for the facts that are exerted via the relationships among their exponents. Both deal with the credentials among bodies of information, in the one case on the basis of substantive considerations regarding assertions, and on the other on the basis of relationship considerations regarding the assertors. Both fact-relations and exponent-relations bear upon acceptability but form profoundly different angles of approach.

Influence theory in the sense we are most immediately concerned with it is the study of formal models of the communicative influence of one person or group of people on another person or group. Aspects of influence theory are clearly relevant to contemporary social and political questions regarding the spread of information and misinformation, some of which are documented below (O'Connor & Weatherall, 2018, 2019). But as a philosophical sub-discipline it also supplements work in philosophy of mind, philosophy of language and philosophy of information with an emphasis on the social dynamics of belief and action change. As an over-arching discipline, influence theory includes aspects of decision theory and game theory as well as established models of de facto segregation, cultural change, opinion polarization, and epistemic networks (Schelling, 1969, 1971, 1978; Axelrod, 1997; Hegselmann & Krause, 2002, 2005, 2006; Zollman, 2007, forthcoming; Grim, 2009; Grim et al., 2013, 2015).

At the core of what we provide here is a systematic outline of results drawn from a range of disciplines, together with new models and extensions, in a structured overview of influence theory as a cohesive area of investigation. Beyond intrinsic interest, a systematic outline of this sort can be of value in at least three ways:

- It can emphasize the similarities and complementarities of results that are closely related in formal structure, though they have been developed and applied in radically different disciplinary contexts. This is an aspect we emphasize in Sect. 3 with regard to influence interpretations of models from physics, biology, and computer science, and in Sect. 4 with application of Markov and epidemiological models to opinion dynamics.

- It can make obvious new or under-noticed applications of formal results across fields, such as the implications of formal undecidability in cellular automata for unpredictability of human patterns of communicative influence, a set of results emphasized in Sects. 10 and 11.

- A systematic outline can also make it clear that there are parts of the over-all picture that have not yet been explored or not yet filled in. We fill in some of those gaps with simple models in Sects. 3 through 6, presented for the first time here, as well as new extensions of population models, information cascades, and the Schelling model in Sects. 5, 7, and 8.

As a whole, influence theory constitutes an overarching discipline of models, hypotheses, general results and limitative theorems. What our intial outline of that discipline offers is not a full catalog but a philosophical framework in which various results, both new and familiar, find their place. What such a framework can make clear is the relation of different patterns of influence, the relation of different patterns to one another, and promises and pitfalls of patterns of influence with regard to particular desiderata of communication, of social decision, of information and information management.

2 Modelling the fundamentals of influence

In relating to our fellows in social context, we humans must in the interest of the general welfare coordinate our own posture with theirs in every significant area of endeavor: thought, action, and evaluation. A great deal of such coordination is overt, explicit, and institutionalized. But much of it is also covert, below the radar screen of instrumentalized explicitness. It is often here, below the level of overt control, that influence comes into operation.

Influence can be exerted in all three of the main spheres of human endeavor: belief, action, and valuation. Our informants and mentors exert influence over our beliefs, our coaches and trainers over our actions, our world models over our values and goals. While explicit control is typically exercised deliberately, influence can be implicit, unintentional and even unwitting. Influence is quite generally a weaker course than control. The controller gets to determine the outcomes at will: the light is on or off in line with what he does. Mere influence—ability to alter outcome probabilities—is generally as far as our efforts go. Parents may exert influence on the career choice of their offspring, for example, but certainly do not control them. In chess one does not—or only rarely—control the opponent's response in forcing a particular move. But if my moves do not influence those of my opponent—i.e. affect the responsive probabilities of various counter-moves—then at least one of us is a very poor player indeed. Again, only rarely do medications actually control our symptomatic responses. But if they do not influence them in altering their likelihood then something is wrong. Ideally and at the level of aspiration, science may aim at giving us control over phenomena. But in general the world we live in is complex and uncooperative, forcing us to settle at best for mere influence.

In the attempt even to boil things down to the simplest aspects of influence there are a number of clear parameters:

- the agents involved
- the influential traits at issue
- the spatial and network patterns of influence contact
- the temporal patterns of influence

- Agents

We take people, or their formal stand-ins, as our basic agents. In the simplest case we can deal with two. But how many agents are involved—and whether singly or in groups or 'schools'—can make a difference in how patterns of influence play out.

- Traits

Is there a single trait at issue in the model—a 'yes' or 'no' regarding some proposition on the ballot, for example—or a set of traits as with the different positions constitutive of a political party platform? Do those traits take a binary value, or real values? And are those traits linked, either in a deterministic pattern or probabilistically?

- Networks

Who are the actors who exert influence on an agent? With even two agents, we can envisage either symmetrical or one-sided influence. With more, we can envisage a line of agents influenced by those to their left or right, a two-dimensional lattice in which agents are influenced by four neighbors (the von Neumann neighborhood) or eight (the Moore neighborhood). Network patterns of any given dimension are open possibilities, as are network patterns in which contact is dictated by similarity of traits or proximity of opinion. Network patterns can change with individual opinion change.

- Timing

Temporal patterns of influence are as important as spatial patterns. Is the model one in which all agents update opinions simultaneously, or in some pattern of sequence? And is that temporal pattern deterministic, probabilistic, or random?

- Fundamental mechanisms

The character of the relevant dynamics can also vary. Even given a set of agents and traits at issue, a network of contacts and timing of interaction, there are basic questions that remain regarding the fundamental mechanisms of influence—the form that individual instances of influence take. Is an agent changed by simple contact imitation, by a majority of its network contacts at a time period, or a threshold proportion? Is that influence deterministic or probabilistic? Does an agent replace a trait or traits, change traits along a spectrum, 'learn' new traits in the manner of neural nets, or hybridize traits in the manner of a genetic algorithm?[1]

Any specific formal model of influence will demand an intertwined specification of all of these factors, and perhaps more. The 'rules of the game' will have to include who is being influenced by whom, in terms of what traits, both spatially and temporally, and in what way. The specific focus of influence theory is the dynamics of influence: how spatial and temporal patterns of influence, of what kind, play out in the formal models at issue. Questions of influence theory include those with a specific target: the influence dynamics within a particular model, or even a particular model with a particular initial configuration. But influence theory also includes more general questions regarding classes of models: similar dynamic patterns exhibited by all models with certain characteristics, for example.

The character of dynamics that is of interest can also vary. Often it will be the end-state of a process of influence that is of interest. Will one trait—an opinion, for example—predominate or become a majority view, for example? Will opinions

[1] Varieties of these different mechanisms within a model for which network, timing, agents and traits are otherwise held constant are explored for example in Grim, Kokalis, Alai-Tafti, Kilb and St. Denis 2004.

crystallize into unchanging polarization? But also of interest may be the path that leads to that end-state. Will that end-state be asymptotically approached or appear as a sudden fixation? We may also be interested in the rates at which an end-state is reached, or the comparative rates of dynamics given different fundamental mechanisms or different networks of interaction. And of course there will be processes of influence that have no tidy end-state: that show oscillation, chaotic behavior, or punctuated equilibrium, for example.

The purposes to which influence theory can be put will include:

- A formal study of influence dynamics in the abstract
- Explanatory models of observed patterns of influence
- Predictive models of expected patterns of influence
- Normative models aimed at optimizing aspects of influence

- Formal theory

The formal development of influence theory will emphasize the quasi-mathematical study of pattern dynamics under iteration. As a formal study, an understanding of the dynamics of influence in formal models is enough. But here as in decision and game theory the possibility of applications can often be expected to be part of the focus.

Influence theory as we conceive it is a formal discipline, much as are decision theory and game theory, aspects of which constitute sub-disciplines. But much of the interest of that formal discipline, as much of the interest in decision and game theory, will be in proposed applications. The formal discipline can tell us what to expect from particular configurations of influence following particular formal rules. But whether those configurations are in fact instantiated in a social reality, and whether those rules are approximated in the interactions of real people, are empirical questions that must be answered in order to apply the formal results, either predictively or normatively. Such is the case with all formal disciplines. It is not the internal mathematics that makes something *applied* mathematics.

- Explanatory models

In terms of application, the goal may be a descriptive understanding of an observed social event or process, either specific or general—the conversion of a majority to a belief in a specific conspiracy theory, for example, or a general increase in opinion polarization with new individually chosen channels of communication.

- Predictive models

It is inevitable that attempts will be made to apply a formal study of influence in the attempt to predict real patterns of influence: the effect of televised hearings on opinion, for example, or a word-of-mouth campaign in advertising a particular product. Predictions regarding the effect of various interventions will inevitably be part of attempted applications as well. But it should be noted that such attempts are fraught with difficulty (Rescher, 1998). An understanding of general dynamics of influence, and even of basic mechanisms, is one thing. Point prediction amidst the complexities of incomplete and inaccurate estimates of parameters in an actual case is another. We may fully understand the physics of balls rolling down frictionless inclined planes, to use a common analogy, and yet be poor predictors of the course of an elephant rolling down a hill—let alone, we might add, predicting the effects of the intervention of putting a large boulder just here…

A particularly important focus for influence theory is the extent to which certain cases of influence are predictable within certain bounds, or predictable at all. In a later section we will emphasize formal limitative results regarding the general (un)decidability of influence systems.

- Normative models

Among the prospective targets of influence models are questions of optimization with particular goals in mind. Given an influence desideratum—approximation to a majority view or an independent truth—in what cases do particular networks of agents using a particular set of updating rules achieve that desideratum? In what cases is a pattern of influence optimal in those terms, and in what cases suboptimal? How should we structure patterns of influence so as to increase effective communication between levels of an organization, for example? What steps should we take to block the influence of misinformation across an existing network? What structure of influence in representation makes a democracy most stable? What structure of influence between theoreticians and experimentalists will give us the most accurate scientific results with the most effective division of labor?

Given this range of parameters—model variations regarding agents, traits, networks, and timing incorporated in the rules of the game; focus on end states or path dynamics; purely formal goals and applicational questions of explanation, prediction, or optimization—it is clear that the overarching discipline of influence theory is an enormous one.

3 Relevant models across disciplines

Influence as we understand it here can be seen as a concept that functions within the broader realm of process theory. A process is a formula or transformation (ϕ) which changes an initial state of something (I) into a resulting state (R), as per:

$$I[\phi] \Rightarrow R$$

A process thus has three components: an input, a mode of operation (modus operandi), and a result.

The objects that will presently concern us as inputs in a process will be states of information. The transformative processes will be modes of information processing, and the result will be a duly determined state of information. The entire process can thus be summarized as follows:

> {Take a body of information} → {Perform a process of information management} → {Inventory a manifold of results}

We have outlined influence theory with a primary eye to psychological, social psychological, and sociological influence in which are agents are taken as people. Influence arises in an interaction between agents, when what is done by the one evokes a reaction in the other. It has two modes: the *cognitive,* which functions when what is done by X affects what is *believed* by Y, and the *practical* which functions when what is done by X affects the *actions* of Y. Our examples will focus on cognitive influence, though many of our formal points apply to both forms.

The principal vehicle for exerting cognitive influence in this sense is *communication*, which is in operation when what someone says is understood by their interlocutors and thereupon evokes a response in their beliefs. And here the result can be either positive (with the recipient's inclination to what is declared being increased) or negative (via a decrease.) When X exercises a positive influence on Y's belief we commonly call this persuasion, which will of course be a matter of degree or extent—total or partial. This is the most common and familiar mode of cognitive influence, and the mode on which we will concentrate. But here again many of our formal points will apply to influence that functions in terms of actions rather than beliefs as well as influence that is not intentional.

Many of the formal points to be made concerning communicative influence among people will also apply, and can profitably draw from, related studies regarding patterns of change that are not cognitive and that need not involve people. There are two-dimensional 'voting models,' interpretable as models of persuasion by neighbors, that take precisely the form of Ising models of spin alignment and magnetism in physics (Galam, Gefen & Shapir, 1982; Galam & Moscovici, 1991; Galam, 1997; Castellano, Fortunato & Loreto, 2009). 'Sociodynamics' and 'sociophysics' represent classes of explicitly physics-analogous models of social change along such lines (Helbing, 1991, Wiedlich, 1971, 1991, 2002, Galam, 2012).

Influence theory can be seen as incorporating aspects of biology as well. Influence can be seen as coordinating a manifold of traits with one's own, realizing a partial identity. The most radical form of influence would consist in total replication of one individual's descriptive content in another. This links the issue of influence with that of self-replication—realizing the same overall descriptive constitution in the influenced 'offspring' that prevails in the influencing 'parent.' In its most formal instantiation, this most radical form of influence is realized in the self-reproducing automata inaugurated by John von Neumann in the 1950s (Bhattacharya, 2021; Burks, 1966). Influence more generally can often be seen as a weakened self-replication, with total replication of descriptive traits as its most extreme form. Reciprocal ties to genetics and evolutionary theory are clear.

The neural network models prominent in artificial intelligence, with both historical inspiration from and application to neurophysiology, are models of influence in which our agents are nodes envisaged as threshold neurons (Haykin, 1994). More generally, causality can quite generally be read in terms of influence. In Mill's methods, causality requires invariable coordination (Mill, 1843), but probabilistic networks can be seen as models of causal influence as well (Glymour & Cooper, 1999; Pearl, 2000; Pearl et al., 2016).

Our aim in what follows is to offer examples of influence modelling at various levels, with an emphasis on the simplest patterns, some of the surprising results and complexities even there, and on ties to established bodies of knowledge and results that take an illustrative place within a broader consideration of influence in general. Those examples are numbered in brackets (example: [4.1]) with a compendium of examples in terms of the parameters above collected as an appendix.

4 Examples from the two-person case

In an attempt to distill things to the simplest case, we might simplify agents, traits, and network parameters by building models with just two agents—A and B, envisaged on the left and right—each with one of two 'traits,' here thought as opinions X and Y:

X Y

If the rules of the game regarding each exchange are that A influences B but not vice versa, we instantly converge to unanimous agreement, which cannot change from that point on. Our end state is simply

X X (3.1)

If the rules of the game are that each agent simultaneously influences the other, on the other hand, we will have an oscillation of opinion at each iteration—a path that clearly has no end state:

X Y
Y X
X Y (3.2)
...

Here we have envisaged the timing of influence in our simplest case as simultaneous. At a given interaction the agents have a simultaneous influence on the other. An alternative is sequential timing, in which one agent first influences the other, the other then influences the first, and so forth ([4.3]). With that change in timing the possibility of oscillation in this simple model disappears. Whichever agent acted first would determine the opinion of both, just as the first play in tic tac toe can guarantee against a loss. Both a fixed first moving agent and a random first-moving agent would leave us with a unanimous opinion of either X or Y.

The ‘rules of the game’ need not dictate deterministic influence, however. A mere probability of influence is a clear alternative as well. We might envisage our agents A and B, again located left and right, as each having a probability of 0.5 of changing the other agent’s mind ([4.4]). Here, if the influence exchange is simultaneous and symmetric, we can expect the following outcomes with the following probabilities:

X	X	A exerts influence, B does not .25
Y	Y	B exerts influence, A does not .25
Y	X	Both exert influence, resulting in a switch .25
X	Y	neither exerts influence .25

In 50% of cases, in this example, we end up with a population of uniform opinion after one exchange. But of particular interest may be iterated influence. What if we have repeated rounds of influence exchange? In this example, after two rounds, we will have as an outcome:

X	X	probability 0.375. This is 0.25 probability of an unchanging X X from the first round, in which no further change influence is possible, plus (0.25 *0.5), which is the probability of going to X X from the proportion of cases that are still open to such a transition: X Y or Y X
Y	Y	probability 0.375 similarly
Y	X	probability 0.25 times 0.50 in the changeable cases = 0.125
X	Y	0.125 similarly

On a third round, we would have.

X	X	0.375 + (0.25 * 0.25) = 0.4375
Y	Y	0.375 + (0.25 * 0.25) = 0.4375
Y	X	0.25 * 0.25 = 0.0625
X	Y	0.0625 similarly

With each round of influence, probabilities that one or the other opinion will dominate increase.

This can of course be generalized to different influence probabilities. Suppose agent A has an 80% probability of influencing agent B’s opinion, whereas agent B has only a 20% probability of influencing A’s. In that case, starting with X Y and with simultaneous influence on a single exchange. We can expect first step outcome as follow:

X	X	probability 0.64 (0.8 probability that A exerts influence times 0.8 probability that B doesn’t exert influence)

Y	Y	probability 0.04 (0.2 probability that B influences times 0.2 probability that A doesn't exert influence)
Y	X	probability 0.16 (0.8 probability that A influences B * 0.2 probability that B influences A)
X	Y	probability 0.16 (0.2 probability that A doesn't influence B * 0.8 probability that B doesn't influence A)

After a first exchange we have a 0.68 probability that both share the same opinion. On a second iteration this becomes.

X	X	0.64 + (0.64 * 0.32) = 0.8448
Y	Y	0.04 + (0.04 * 0.32) = 0.0528
Y	X	0.16 * 0.32 = 0.0512
X	Y	0.16 * 0.32 = 0.0512

and so forth.

All calculations of this form, as long as probabilities of influence are fixed, will asymptotically approach an equilibrium. Here that equilibrium is simple and obvious: the opinion of the agent with the higher probability of influence will dominate in the long run, though the exact pattern of that 'long run' will depend on the rolls of the dice.

Here as above we may also consider a variation in timing from simultaneous possibilities of influence to sequential, which give us importantly different results ([4.5]). Let us return to our two agents A and B, on the left and right, each with a 50% probability of converting the other:

X X

We envisage A going first, followed by B if both don't yet agree, followed by X, and so forth.

With A going first from X Y, the probability of an X X arrangement after the first round is 50%.

With A going first and B going second (in cases that haven't already been 'decided'), the probabilities after the second set of exchanges are:

X	X	0.5
Y	Y	0.5 * 0.5 = 0.25
X	Y	0.5 * 0.5 = 0.25

Note that a final configuration of Y X is not possible if A playing X takes the first move.

The probabilities with a further play by A are.

X	X	0.5 + (0.5 * 0.25) = 0.625
Y	Y	0.25

X	Y	0.5 *. 25 = 0.125

and so forth.

One point of interest is that convergence to a single opinion is 'faster' in the case of sequential influence, if we count by number of plays by A + plays by B. After a single round, the probability of opinion agreement in the case of simultaneous exchange is 0.5; in the case of sequential exchange the probability of agreement is 0.75.

As in the simultaneous case we can also calculate results of sequential exchange for different influence probabilities. If A's influence is 0.8 compared to B's influence of 0.2, probabilities in which A takes the first move will be:

X	X	0.8
X	Y	0.2

With a second influence from B these become:

X	X	0.8
Y	Y	0.2 * 0.2 = 0.04
X	Y	0.8 * 0.2 = 0.16

And so forth. Here too sequential updating gives us a faster convergence toward uniform opinion. After a single play by each of A and B the probability of the same opinion is 0.68 in the case of simultaneous influence, but 0.8 in the case of sequential.

A very general measure of relative influence between agents suggests itself. Let it be that over a succession of time periods (say days) results for A and B show the following pattern, where + or – indicate the presence or absence of a trait at the end of that period and (+) and (-) indicate whether that trait has changed in a way that reflects that of the other agent on the previous period ([3.6]):

	period 1	period 2	period 3	period 4	period 5	period 6	period 7	period 8
Agent A	+	+	(-)	-	+	+	(-)	-
Agent B	-	-	(+)	(-)	-	-	(+)	(-)

In exploring the prospect of influence we can ask:

> For how many periods—i,e., how many times—does A's condition in point of + or - exhibit a change that reflects *B*'s condition on the previous day?

For the two agents the days on which such change in conformity occurs have been parenthesized. For *A* there are two such other-conforming days; for *B* there are four. With influence interpreted along these lines, it becomes a statistically straightforward process to determine its direction and its extent. In this example there are 7 opportunities of influence of which *A* seized four and *B* two. By comparative standards *A* exerted influence to the extent 4/7 and *B* to the extent 2/7/ by relative standards *A* was twice as influential as *B*.

One obvious aspect of simplification in our presentation of these simple models is that we have focused on a single binary trait. A larger area of investigation opens with the consideration of influence across multiple traits, rather than just one. It's clear that similar calculations and a similar measure could be expanded to the case of multiple traits as well.

In the examples above we have treated traits as binary options. But many opinions are best envisaged as on a scale: one's response to questionnaire questions using a Likert scale, for example, the percentage of tax revenues one thinks should be spent on public education, or one's degree of confidence in an improving economy. Here different rules of the game will be appropriate, with their own characteristic dynamics. As a simple example, consider agents A and B with opinions 0.1 and 0.9 respectively and a pattern in which each agent simultaneously influences the other to move halfway toward its opinion ([4.7]). Agent 1 will move to 0.5, as will agent 2, with a clear convergence. If the dynamic is that of the sequential pattern above, however, agent 2 will first move to 0.5 but agent 1 will then move to 0.7, agent 2 will then move to 0.6 and agent 1 to 0.55, resulting in an importantly different outcome. Even in simultaneous updating, our two agents need not follow the same updating rule. If agent 2 stubbornly refuses to budge, but agent 1 follows the 'half-way' pattern of influence, agent 1 will slowly move to meet an opinion of 0.9.

One further note of interest is that our 'rules of the game' in these initial models have here been equally balanced in terms of the number of influence events allotted each agent. In sequential influence it certainly does matter who goes first, but even there we have envisaged alternating influence events. All probabilities will change if agent B somehow manages to take two turns for one of A, for example ([4.8]). Decision theory and especially game theory are known for calculation of optimal strategies. In influence theory one aspect of strategy will be a calculation of how many instances of influence are required in order to compensate for lower probabilities of effective influence on each interaction.

Standard game theory, though it maps gains from interactions, does not standardly include influence. But a simple variation—imitative game theory—clearly would. In such a variant, agents gain or lose points from an interaction (or series of interactions), but then adopt the strategy of a neighbor if that neighbor is doing better. One might envisage this as a version of game-theoretic gains in which an aspect of influence theory is added. But one might also envisage it as a form of influence theory in which motivations for change are more complex and are tied to payoff matrices from previous interactions. Imitative game theory of this sort is played out on a two-dimensional lattice in Nowak & May, 1992 and 1993 and a string of later related work (Grim, 1995). Differences between simultaneous and non-simultaneous influence in that instantiation are noted in Huberman & Glance, 1993. Results in this tradition for stochastic imitative game theory appear in Nowak & Sigmund, 1992 and Grim 1996.

5 The interaction of multiple agents

In the two-agent case, both the rules of the game and the temporal element of, for example, simultaneous versus sequential play make important differences in both outcome and path dynamics of influence. Extension to consideration of multiple agents adds many more complexities, starting with how agent interaction if envisaged: at random across a population, for example, or within a structured network.

Here we might also entertain the possibility that those holding view Y as more persuadable than those holding X: on an occasion of interaction, those holding Y may have a 30% probability of converting to X, while those holding X may have merely a 10% probability of converting to view Y. We can suppose each sub-population to occupy 50% of the whole ([5.1]).

At a first glance, one might think that such a population would quickly become one in which all agents hold opinion X. But here the case differs significantly from that of just two agents. With an exchange in which all agents interact with all others, the expectation is that 10% of X's will become Y's and 30% of Y's will become X's, giving us 60% X's and 40% Y's. But at that point the disparity in the populations becomes crucial. Since the population is 60% X, Y's will interact with a larger population of 'others' than X's will, giving their X-to-Y conversion probabilities a wider field. That pattern continues. On successive interaction instances of all agents with all, the expected proportions of X and Y in the population exhibit the following pattern:

X	Y
50%	50%
60%	40%
66%	34%
69.6%	30.4%
72.03%	27.97%
…	

Although X's proportion slowly increases and Y's slowly decreases, each progresses Zeno-style not toward a population entirely of X's but toward a fixed point equilibrium of 75% X and 25% Y. At those proportions, our probabilities balance out. At 75%, The percentage of X's on the next generation will be (0.75 * 0.9) + (0.3 * 0.25) = 75% again.

Somewhat more surprising is the fact that it doesn't matter what proportion of X's and Y's we start with, as long as our conversion probabilities remain the same. If we begin with 10% X's and 90% Y's, or 1% X's and 99% Y's, progression is to the same equilibrium of 75% X's and 25% Y's. The role of initial proportions is swamped by the iterated application of fixed transition probabilities.

This simple two-party case is a Markov process, for which the Perron-Frobenius theorem applies (Page, 2018). Given a finite set of states (our X and Y), the possibility however unlikely of eventually moving from one state to any other, and the absence of a deterministic cycle through a sequence of states, any such system will asymptotically approach a fixed equilibrium of proportions. The same will hold for any number of states. Were 4 exclusive opinion possibilities at issue, with fixed probabilities of transition between them—20% probability that an X will become a W, 10% that it will become a Y, 0% that it will become a Z (but 1% probability that a Y will become a Z), for example—our system would again approach a fixed probability.

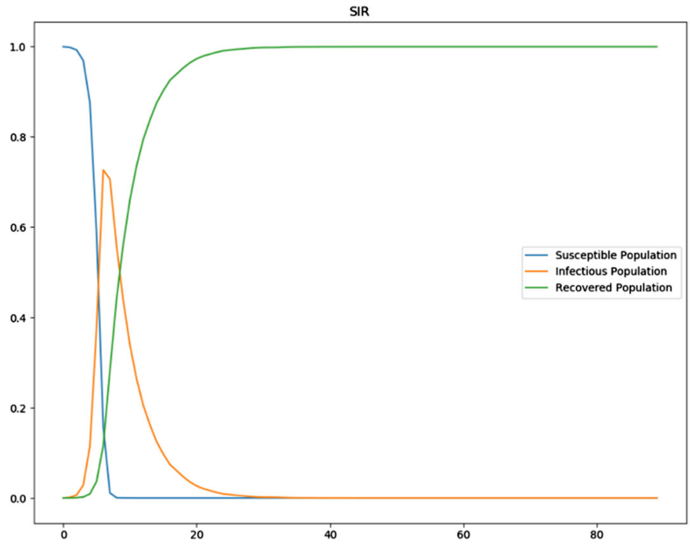

Fig. 1 Standard course of an epidemic in a SIR model

Classical forms of epidemiological calculation follow precisely this pattern (Kermack and McKendrick 1927, 1932) ([5.2]). In a SIR model we have proportions of a population which are uninfected but susceptible (S), currently infected (I), and recovered and permanently immune (R). Given the assumption of random mixing in a population and a specific probability $R_0 > 1$ of a susceptible becoming infected on exposure, the course of a disease standardly follows the type of S-curve also evident in the case of X and Y above (Fig. 1).

A variation is the SIRS model (susceptible-infected-recovered-susceptible) in which recovery is not assumed to guarantee immunity. The recovered thus have some probability of re-entering the susceptible population. Unlike the pattern shown in Fig. 1, in which the susceptible proportion of a population declines to 0 as the recovered proportion climbs to 1, an SIRS model offers the prospect of a continuing equilibrium of endemic disease, such as that expected for Covid and its variants.

The assumptions of the Perron-Frobenius theorem include the lack of a deterministic cycle through states. If X's had a 100% probability of converting to Y's on exposure, and Y's had a 100% probability of converting to X's, our dynamics would be oscillatory, without a fixed point equilibrium ([5.3]). Starting with 25% X's and 75% Y's, expectations would return to that proportion every other iteration, with its mirror image—75% X's and 25% Y's—in between. The Perron-Frobenius theorem also demands fixed probabilities of transition between states. Changing these can easily change the equilibrium, and irregular or chaotic enough change can defeat the possibility of equilibrium entirely. A study of patterns of epistemic chaos when self-reference is at issue appears in Grim, Mar & St. Denis (1998) ([5.4]).

The Granovetter model of propensity to riot is a population model in the same sense as these, in that only percentages of the population are at issue (Granovetter,

1978) ([5.5]). But it differs from the examples above in that different agents have a propensity to convert from an 'unwilling to riot' to a 'willing to riot' position depending on the number of others in the population who are already rioting. Precisely the same dynamics are at issue if we simply consider agents willing to convert from position X to position Y given heterogeneous percentages of the population that already hold position Y.

The point of the Granovetter model is a simple one: that the distribution of 'trigger percentages' is crucial to the outcome of the whole in ways in which a mere average of 'trigger percentages' is not. Consider a population in which 1% already holds view Y, with the rest holding view X. But suppose also that a second 1% is willing to convert to Y if 1% of the population holds Y, a further 1% is willing to convert if 2% of the population holds Y, a third 1% if 3% of the population holds Y, and so forth. The result, trickling over time, will be a population of agents all of whom hold position Y. But a population that starts with 1% Xs but in which the next 2% with the lowest threshold will convert only if 2% of the population holds Y will forever remain with 99% of the population holding view X.

An epistemic model that relies on the Markov process noted above but goes beyond mere population percentages is Lehrer & Wagner's, 1981 model of the formation of rational consensus ([5.6]). Each agent begins with their own probability assignments regarding alternative hypotheses (that the sun is spherical, oblate, or neither, for example) as well as weights representing the credibility or reliability that agent assigns to his own views and those of each of the other members of the group. Using initial probabilities and (positive) other member weights as a representation of background knowledge, each agent revises his or her hypothesis probabilities as the weighted average of the probability assignments of all members of the group. If each agent assigns positive reliability weights to all other members and keeps those weights constant, it can be shown that under iteration such a process converges in the limit toward a consensus in which all agents have the same probability estimates for the hypotheses at issue. As Lehrer and Wagner note, the consensus results of such a model are again those of a Markov process. Earlier consensus models in this tradition include French, 1956, Hararay, 1959, and DeGroot, 1974. Later models that show lack of consensus and polarization when basic trust assumptions are weakened are outlined below (Deffuant et al., 2002; Hegselmann & Krause, 2002, 2005, 2006).

6 Linear influence

Although the models considered in the previous section involve multiple agents, they are largely population models. Like the SIR model, they carry an assumption of random or complete mixing or contact. But with multiple agents also comes the structure of specific networks of interaction, in which it's not true that every agent is in contact with every other.

A simple first case is a one-dimensional model in which our agents form a line.

$$\cdots \quad X \quad Y \quad Y \quad X \quad Y \quad X \quad X \quad Y \quad Y \quad Y \quad X \quad Y \quad \cdots$$

Any of various rules may be in play. For example, let it be that an agent holding position X changes to position Y whenever one of its neighbors is a Y ([6.1)]. If there are any Ys at all, all agents in such an arrangement will eventually hold position Y. We can also consider the case in which a similar rule holds but with a mere p-percent probability: there is a positive p-percent chance than an X will turn an adjacent Y into an X within the next interval of time. If there are any Xs at all, then the entire array will become homogeneously X subject to this mode of linear diffusion, but at a rate of speed contingent on probability p.

Or suppose that for a Y to be turned (for sure) into an X it needs to be *surrounded* by X on both sides, so that X Y X – > X X X ([6.2]). Then some Ys in a random series will predictably be changed, but some may never be, for example the Y's in a formation X X Y Y X.

Given a network spatialization, moreover, rules need not be symmetrical. Consider the alternating sequence.

X Y X Y X Y X Y

subject to the rule that every Y preceded by two Xs will be changed to an X ([6.3]). Given this initial arrangement, nothing will change.

But consider the same change-rule with a random sequence.

X Y X Y Y Y X X Y X Y Y Y ···

Here, once the rule begins to take effect, *everything* further to the right will uniformly become X. So here the diffusion issue depends not (just) on the 'infection' change rule but one the nature of the initial distribution.

We start with a simple line of agents with alternating opinions X and Y:

X Y X Y X Y X Y X Y

If agents convert to an alternative opinion when surrounded by neighbors with that opinion, and if all agents update simultaneously, our array will evolve as follows, 'rearranging' all X's to the left and all Y's to the right, as it were ([6.4]):

X X Y X Y X Y X Y Y
X X X Y X Y X Y Y Y
X X X X Y X Y Y Y Y
X X X X X Y Y Y Y Y

Alternative initial patterns will give us alternative final patterns. The following, for example, will leave us with an island of Y's in the center of X's:

X Y X X Y Y X X Y X
X X X X Y Y X X X X

Of particular interest is a loop in which opinions circle at the point marked z ([6.5]):

z X Y X Y X Y X Y X Y X Y X Y X Y z

Each X here is surrounded by two Y's, each Y by two X's. Thus on each exchange each X will convert to a Y and each Y will convert to an X, resulting in an infinite oscillation across the entire array.

But consider a simple alteration in which we have two X's in a row:

z X X Y X Y X Y X Y X Y X Y X Y X Y z

Those 'leftmost' X's will not convert, though Y's on each side will, and will not revert to Y's on the next iteration. In the end the entire array will become occupied by X's; oscillation in that case will give way to spreading unanimity.

Given simple rules, then, a change in initial configurations can produce radical changes in the evolution and ultimate configuration of an array. It's also obvious that different rules will give us a different influence dynamics. One body of work that fits into linear influence theory is the study of information cascades, perhaps better thought of more generally as influence cascades.

7 Influence cascades

In the simplest case, a linear influence cascade can be modelled in terms of a line of individuals who make decisions not simultaneously, as in the examples above, but sequentially. A first agent considers personally available evidence and makes a decision. A second agent considers his or her personally available evidence, but also the decision of the first agent. A third agent considers personally available evidence together with the information afforded by decisions made by previous agents in the line.

In one ideal case, each agent may share both his or her decision and the evidence on which it is made. In such a case individuals further down the line would have an increasing body of evidence on which to act. If the evidence that each individual personally receives has even a small probability of being correct—51%, say—individuals far enough down the line will have accumulated evidence sufficient to establish the correct decision with a probability approaching 1 ([7.1]).

But in many cases personal information or evidence is not shared, with only the decision or corresponding action of previous agents being observable. Here the observed performative becomes a crucial element in the calculated cognitive response. In a variety of financial transactions, for example, it may be to agents' interests not to reveal their basis of decision, although that decision itself may be publicly observable. Even in communication, social or media-mediated, we may take shortcuts by following actions of other agents rather than ferreting out full evidence and argument for ourselves. In the extreme case, it is only other agents' ultimate decisions that are observable. In such a case, rational observers could infer the evidence received on the basis of that decision or corresponding action, but only on that basis ([7.2]).

It is in this latter case—and the extent to which real cases approach it—that information cascades can be expected to occur. To the extent that we are influenced by others' actions alone as a guide to their individual information, we are open to cascades of both information and misinformation. Here we will also emphasize that sub-groups within a population will increase the probability of a misinformed subgroup and opinion polarization between groups.

We start with a line of agents each making a decision sequentially on the basis of available information. We suppose each agent has private information as to *p* or ~ *p*, for simplicity with a set probability of being correct that is the same for all agents. Our first agent samples the available evidence and, let us say, decides on that ground that *p*. Our second agent samples his private information, which indicates either *p* or ~ *p*.

Suppose that the second agent's evidence indicates that *p*. In that case his decision is clear: he knows that both his own evidence indicates that *p* and that the evidence of the first agent does as well. But now note what happens with the third agent. Even if his private evidence indicates that ~ *p*, the decisions of both predecessors indicates that two other agents, presumed to be peers, have evidence that *p*. On rational grounds, he will take his personal evidence to be an outlier, and will decide that *p* despite his private evidence to the contrary. What holds for our third agent holds even more strongly for all the rest down the line: All subsequent agents will override individual evidence in favor of a decision that ***p***—an information cascade in which everyone decides that *p* simply because of the evidence of two agents at the front of the queue. Had those two agents both decided that ~ *p*, we would have a symmetrical information cascade of uniform decision for ~ *p*.

What if our first agent had decided that *p*, but our second agent's personal information indicated ~ *p*? In that case, if our second agent treats his predecessor as a full peer, available information is simply split. The second agent might then flip a coin, in 50% of cases resulting in precisely the cascade we've tracked.

Should the first agent choose *p* and the second agent choose ~ *p*, their individual influence effectively cancels out and we start the line again with the third player. That third player can be presumed to go with their private information, just as we assumed for the first agent before. The fourth agent will either have information on the same side as the third—starting a cascade as before—or will have contrary information, calling for a flip of a coin with a 50% probability of a cascade or a procedure that effectively 'starts again' with the fifth player.

One thing to note is that it is not merely the balance of evidence between *p* and ~ *p* that is at issue, but the path-dependent order in which that evidence appears. If we suppose that agents' evidence arrives as *p* for agent 1 and *p* for agent 2, followed by ~ *p* for agent 3 and ~ *p* for agent 4, an information cascade for *p* will already have been established after the first two agents, swamping the individual evidence of subsequent agents. But if we suppose a pattern of evidence ~ *p*, ~ *p*, *p*, *p* for agents 1 through 4, the cascading information through the population as a whole will be a cascade of ~ *p* instead.

To this point we have tracked the formation of cascades, but have not tracked their accuracy. Let us suppose that the truth is in fact *p*, but that evidence available is ambiguous or noisy, giving agents only a 51% probability of accurate information.

Private information in that case will have a probability of 51% of giving the correct answer *p*, a probability of 49% of giving the incorrect answer ~ *p*.

In such a case, after just the first two agents' decisions, there is an approximately 75% probability of getting a cascade one way or the other. The probability of a *p* cascade on the basis of identical information between the first two agents that *p* is 0.51 * 0.51 = 0.2601. The probability of a p cascade on the basis of a coin flip given divergent information between the first two agents is 0.51 * 0.49 * 0.5 = 0.12495. Taken together, the probability of a *p* cascade on the basis of *p* information to each of the first two agents is 0.38505. The probability of a ~ *p* cascade is 0.49 * 0.49 = 0.2401, + 0.12495 = 0.36505. Summing all of these, the probability of one form of cascade or another in which everyone goes for *p* or ~ *p* after just the first two agents' decisions is 75.01%.

A cascade can be avoided, we've noted, only if the first two agents differ in decision, followed by the next two agents differing in decision, followed by the third two again differing… With an evidence probability of 0.51, the chance of such a pattern occurring four times in a row is only 0.004, meaning that after 8 agents in line the probability of a cascade one way or the other is greater than 99% – despite the fact that individual evidence leans only 51% on one side (Bikhchandani, Hirshleifer & Welch, 1998).

Of course the information that cascades through such a pattern may in fact be *correct* information. If information has a 51% probability of being correct, very slightly more of the cascades will be cascades of correct information *p* rather than incorrect information that ~ *p*. In the case at issue, the probability that a cascade is a correct *p* cascade rather than an incorrect ~ *p* cascade can be abstracted from the first-two-agent case above: 0.38505 / (0.38505 + 0.36505), or approximately 51.334%. This is slightly better than the 51% probability of agents acting on their private information alone, but of course far less than the probability approaching 1 of a correct decision were the evidence seen by each agent available to all.

The higher the probability that evidence is correct, the higher the probability that a cascade will be a cascade of correct information. Bikhchandani, Hirshleifer & Welch (1998) graph the proportions of correct and incorrect cascades given information probabilities as in Fig. 2. At even a probability of correct information of 0.7, the probability that an essentially inevitable cascade is a cascade of correct information is 0.753. The probability of a cascade of incorrect information infecting the population as a whole is still 0.247.

Although we have envisaged influence theory as a formal discipline, we have also noted that will inevitably be mined in the attempt to understand social and other phenomena at every side. Here it is thus perhaps not out of hand to note that the phenomenon of informational cascades has implications for contemporary concerns with echo chambers and opinion polarization.

The model and calculations above are based on the assumption of a single population of sequentially ordered agents. The same basic pattern will hold, however, if we have a branching tree of sequentially deciding agents. One need only think of the first two agents as the trunk, with lower agents acting on what they observe of agents on a direct line above them.

Consider now the possibility of a population divided into two enclaves ([7.3]). Even if the information received by all agents is the same, we can calculate the probability

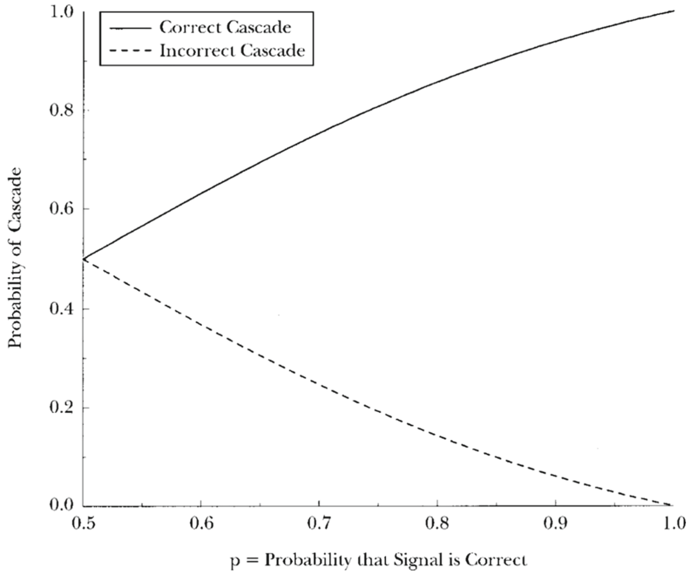

Fig. 2 The probability that a cascade is of correct vs. incorrect information (y axis) given a probability that private agent information has a given probability of correctness (x axis). From Bikhchandani, Hirshleifer and Welch, (1998)

that the population will be polarized in terms of opinion. At information of a probability of 0.7, we've noted, the probability of an essentially inevitable cascade being a correct one is 0.753. The probability of two independent cascades in two independent enclaves both being correct is thus a mere 0.567. If the evidence is noisy or ambiguous the situation is far worse. Given an evidential probability of 51, the probability of an essentially inevitable cascade being correct is a mere 51.334 * 51.334 = 0.2635. The probability of an opinion-polarized population when evidence is that ambiguous or noisy, in other words, is close to 75%.

The more independent enclaves there are in a population, the higher the probability of polarization. With four informationally isolated enclaves, the probability of full unanimity on the correct answer given an evidence probability of 0.7 is merely 0.753 * 0.753 * 0.753 * 0.753 = 0.3215, meaning that the probability of cascade of misinformation in at least one portion of the population is greater than 2/3.

Cas Sunstein and others warn us against the dangers of epistemically isolated sub-groups (Sunstein, 2001, 2007). C. Thi Nguyen distinguishes between 'epistemic bubbles,' which are defined merely in terms of epistemic isolation, from 'echo chambers' defined as discounting information from other areas (Nguyen, 2020). To the extent that the epistemic dynamic of subgroups can be captured in terms of the conditions of epistemic cascades, and given any noise or ambiguity in information or its processing, epistemic isolation of sub-groups would be enough to predict and explain unanimity within 'echo chambers' and radical polarization of sub-groups across a population.

With an eye to applications it is certainly appropriate to ask to what extent the model assumptions of information cascades hold, and to what extent the modelled behavior can be seen in real cases of human influence. It has been proposed that real situations from fads and fashions to the spread of rumors, adoption of new technologies, and even involvement in crime can have at least some of the marks of information cascades (Bikhchandani, Hirshleifer & Welch 1998, Walden & Browne, 2002, Kahan 1997). In the general phenomenon of 'social proof,' well established as a phenomenon in social psychology, the action of others is taken as grounds for imitation, though the strict sequence of the cascade model is not generally in play (Cialdini, 2009). It is a standard advertising and marketing strategy to stimulate and publicize 'early adopters,' and it has been claimed that Donald Trump's political campaign was kicked off using a cheering group of paid actors playing such a role. In a laboratory experiment designed to closely match the assumptions of the model, Anderson and Holt (1997) demonstrate an epistemic cascade. In a line of 94 agents guessing whether an urn contained 2/3 black or 2/3 white balls and privately given one drawn ball, 79 agents acted against their own private evidence and followed a cascade.

But the model assumptions for influence cascades are strong ones. As presented, these include lines of agents making their decisions sequentially and agents that treat others as epistemic peers, such that a decision that *p* by a preceding agent is taken as evidence for *p* as strong as an individual's own private information.

Both conditions can be weakened. As noted, decision structures can be tree-like rather than linear. All that is really required is that an agent take earlier decisions of other agents as evidence on a par with his or her own. It need not even be that an agent treat others on a paras epistemic peers. If an agent discounts the evidence drawn from others' actions—at only 80% the comparative value of his own evidence, for example—it is still the case that a pattern of previous choices will overwhelm individual evidence, leading to a cascade ([7.4]). In the long run that pattern too will be inevitable, with virtually the same conclusions holding with regard to the inevitability of cascades and the percentages of correctness. Although we do need to guard ourselves with the proviso 'to the extent that real situations model the conditions of the model,' therefore, the model is not limited to either strictly linear patterns of decision or full deference to previous agents as epistemic peers, nor does it demand a homogeneous probability of correctness for agents' evidence. Wherever we see actions of earlier actors taken as evidence, rather than their evidence itself, we should expect the possibility of something like an epistemic cascade.

There are a family of influence models from political science which further enrich considerations of multiple agent influence under different decision rules, specifically with implications for the impact of decision rules on the selective sharing of information (Fedderson & Pesendorfer, 1998; Coughlan 2001; Austin-Smith and Feddersen 2005, 2006). Consider for example a jury voting to acquit or convict a defendant, and in which each juror has a single piece of information *i* or *g* (a 'private signal') indicating guilt or innocence, with a given probability of being correct. Jurors are also assumed to each have a 'bias' or standard of reasonable doubt with regard to how much evidence should be required for conviction. Some jurors may require all evidence to be *g* in order to convict, whereas others are willing to convict given a smaller percentage of *g* evidence ([7.5]).

What this family of models emphasizes is the impact of a unanimous decision rule, requiring for conviction or acquittal that all jurors vote the same way, as opposed to mere majority rules. The simplest and perhaps most optimistic case is that in which jurors vote 'informatively,' accurately representing their private signals. In that case the unanimity rule leads to a lower probability of convicting the innocent than does any mere majority rule (Fedderson & Pesendorfer, 1998, p. 24). The possibility of strategic voting and strategic representation of private information changes that result dramatically, however. Under different decision rules and given different juror 'biases,' there can be an incentive for jurors who see themselves as pivotal not to vote in accord with their private information or not to reveal that private information in discussion, deliberation, or preliminary straw polls. Where verdict error is taken as variance from what the full set of information would indicate, the result is that the use of a unanimity rule can be expected to lead to a higher probability of error than a range of majority rules. Given uncertainty regarding other jurors' biases, Austin-Smith & Feddersen (2006) argue that "the unanimity rule is uniquely bad with respect to providing committee members with incentives to share relevant information prior to voting (p. 211)."

Application of influence cascade models to opinion polarization has already been mentioned. Polarization is also the target of a related model in which enclaves or networks of influence are formed on the basis of proximity of real-valued rather than binary opinion, but in which something like an influence cascade can also occur. Hegselmann and Krause offer a 'bounded confidence' model in which mutual influence is modeled between those within only a specific threshold ε of opinion similarity (Hegselmann & Krause, 2002, 2005, 2006) ([7.6]).. Opinions in the Hegselmann-Krause model are mapped onto the [0, 1] interval, with initial opinions spread uniformly at random. Belief updating is done by taking a weighted average of the opinions that are 'close enough' to an agent's own. As agents' beliefs change, a different set of agents or a different set of values can be expected to influence further updating.

The primary results of the model are the formation of consensus given certain thresholds for who counts as 'close enough' and the formation of polarized groups with narrower thresholds. Figure 3 shows the changes in agent opinions over time in single runs with thresholds ε set at 0.01, 0.15, and 0.25 respectively. With a threshold of 0.01, individuals remain isolated in a large number of small local groups. With a threshold of 0.15, the agents form two permanent groups. With a threshold of 0.25, the groups fuse into a single consensus opinion. Very similar results appear in a model by Deffuant et al (2002) in which the sharp cutoff thresholds of the Hegselmann-Krause model are replaced with continuous influence values.

An illustration of average outcomes for different threshold values in the Hegselmann-Krause model appears as Fig. 4. What is represented here is not change over time but rather the final opinion positions given different threshold values. As the threshold value climbs from 0 to roughly 0.20, there is an increasing number of results with concentrations of agents at the outer edges of the distribution, which themselves are moving inward. Between 0.22 and 0.26 there is a quick transition from results with two final groups to results with a single final group. For values still higher, the two sides are sufficiently within reach that they coalesce on a central consensus.

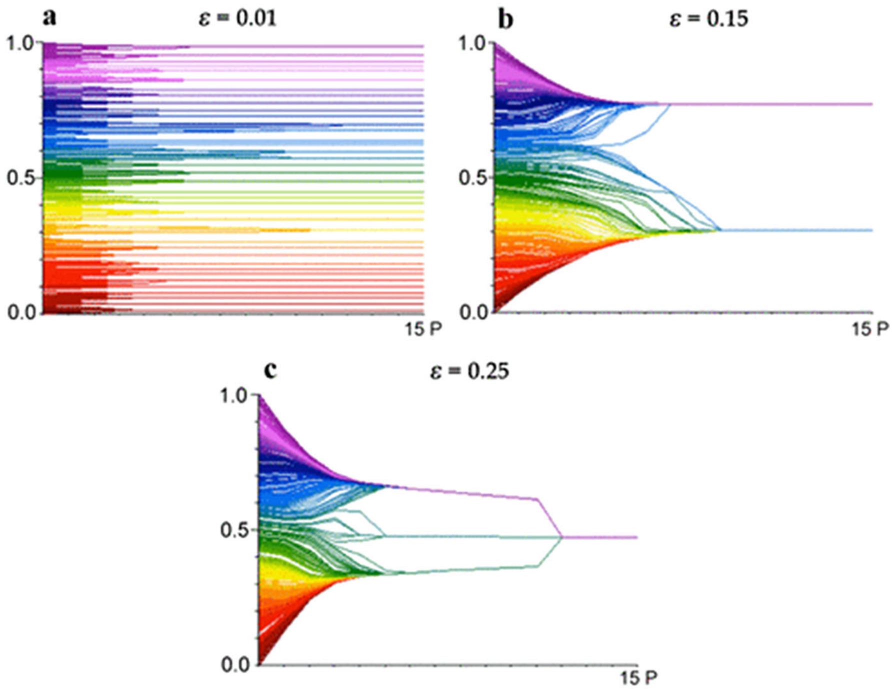

Fig. 3 Example changes in opinion across time from single runs with different threshold values ε ∈ {0.01, 0.15, 0.25} in the Hegselmann and Krause (2002) model

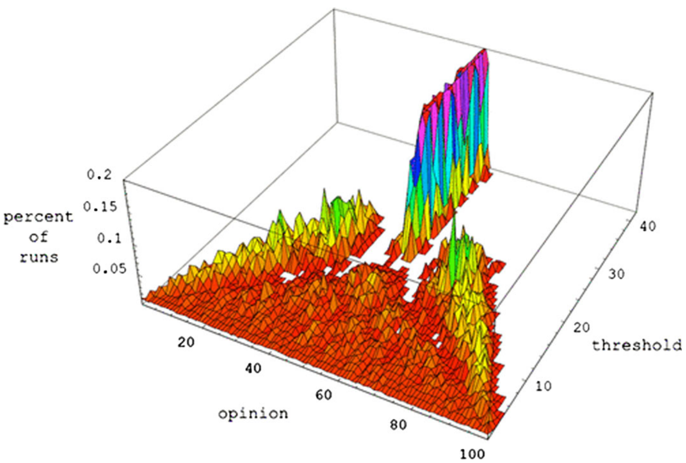

Fig. 4 Frequency of equilibrium opinion positions for different threshold values in the Hegselmann and Krause (2002)model scaled to [0, 100]

8 Influence across two dimensions

Spatial patterns of influence can of course instantiate higher dimensions than the linear. One well-known study represents both (1) the effect of patterns of differential contact and (2) the formation of patterns of contact.

In the standard presentation of the Schelling model (Schelling, 1969, 1971, 1978), one envisages two types of agents (plus spaces) in a two-dimensional checkerboard. Agents have specific thresholds of 'tolerance' for other agents in their immediate neighborhood: the Moore neighborhood of the 8 cells immediately adjacent to them ([8.1]). If the agents surrounding a red agent meet a certain percentage of similar red agents (30% of 'its kind', perhaps), that agent stays put. If it is surrounded by a percentage below its threshold, it moves to another spot until it finds one that meets that threshold.

Schelling's was intended as a model of de facto segregation, demonstrating that recognizable patterns of racial segregation need not require high levels of individual racism: that desire for merely 30% of 'one's kind' in one's immediate neighborhood could nonetheless result in qualitatively distinct segregated enclaves (Fig. 5).

But Schelling's model represents an abstract dynamics. Its space need not be interpreted as geographical residence, and its types need not be interpreted in terms of race. An alternative interpretation is one in which colors represent pro- and anti- attitudes toward a principle or proposition, or conflicting political views. Position on the grid can be read as communication or association with others, and the threshold read as that percentage of associates an agent wishes to be 'like-minded.'

On that alternative interpretation, what Schelling's model shows quite vividly is the formation of patterns of influence and their reinforcement. What one sees in the evolution of an array such as that in Fig. 5 is the formation of 'epistemic enclaves' or 'echo chambers' in which a large percentage of associations are like-minded (74%, in

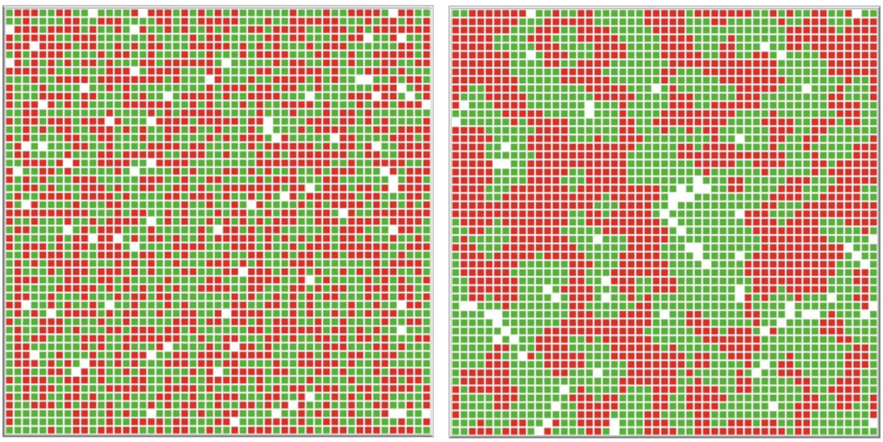

Fig. 5 Typical evolution in the Schelling model with a 97% density and desired neighbor percentage ratio of 30%, resulting in neighbor similarity of 74%

the example of Fig. 5), even when the desire for like-minded association is relatively low (30%, say). There is also a mixed interpretation, in which space remains residential but in which agent colors are taken to indicate 'like-mindedness' or political persuasion. On such a reading, the model would predict residential segregation along political or opinion-based lines, a trend that forms the core of work by Sunstein (2001, 2007).

Two aspects of Schelling dynamics are of particular note with regard to this interpretation. The first is that the configuration of opinion areas clearly depends on the (symmetrical) levels of like-mindedness demanded by the two types of agents. Where the percentage of like-minded neighbors demanded is 50% rather than 30%, the areas occupied by each type are larger and far more contiguous, with a neighbor-similarity of 97% rather than 74% (Fig. 6; note that this is a wrap-around or toroidal display). On an interpretation in terms of patterns of influence, the result of higher 'like-minded' demands is a tendency toward two monolithic areas rather than a scattering of separated 'like-minded' islands.

The second aspect worthy of note is the potential role of different thresholds of 'like-mindedness' between the two agent types. With a fairly low density, allowing many unoccupied spots in the array, a low 'don't care' threshold on the part of the greens together with a high 'like-minded' demand on the part of the reds results predictably in enclaves of reds in a field of scattered greens (Fig. 7a) ([8.2]). But with a fairly high density, allowing few unoccupied spots, and despite a 'don't care' attitude on the part of greens, it proves much more difficult for enclaves of reds to form, however high their 'like-mindedness' demands. This is simply because the greens will have no desire to move, and thus will not open up desired spatial options for the reds (Fig. 7b).

In a two-dimensional model of influence which incorporates multiple traits, Axelrod (1997) proposes that a form of polarization can arise from an intuitive mechanism that would at first sight seem only to promote conformity and cultural convergence

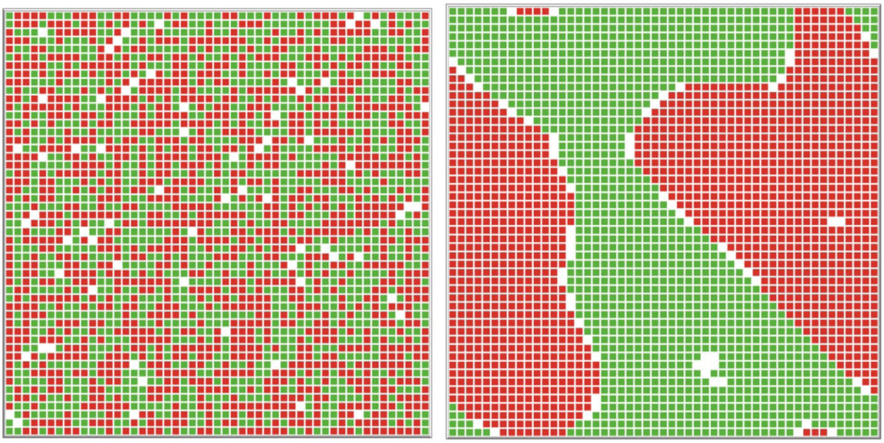

Fig. 6 Typical evolution in the Schelling model with a 97% density and desired neighbor percentage ratio of 50%, resulting in neighbor similarity of 97%. Note that this is a 'wrap-around' or toroidal array.

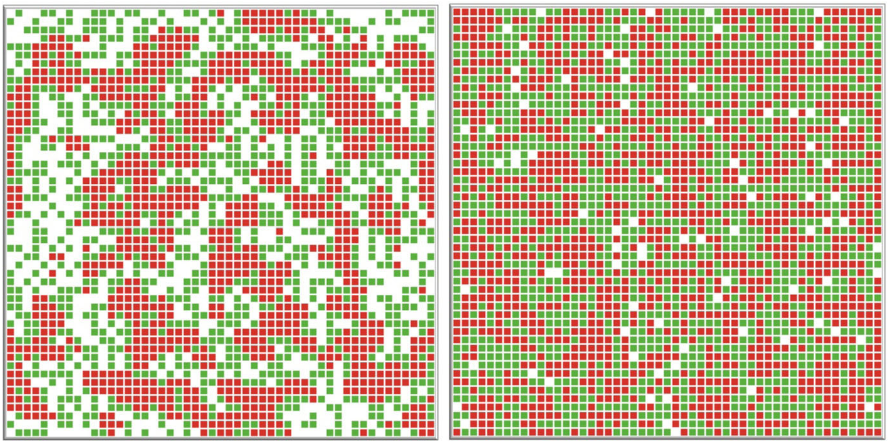

Fig. 7 Contrasting ability of red enclaves to form with low density (left) as opposed to high density (right), despite a 50% like-mindedness demand on the part of reds and a mere 5% like-mindedness demand on the part of greens in both vases

([8.3]). The basic premise is this: that people tend to interact more with those like themselves and tend to become more like those with whom they interact. But if people come to share one another's beliefs (or other cultural features) why do we not observe complete cultural convergence? At the model's core is a spatially instantiated imitative mechanism that produces cultural convergence within local groups but also produces progressive differentiation and cultural isolation from other groups. He refers to that differentiation as 'polarization'.

Axelrod's base model consists of 100 agents arranged on a 10 × 10 lattice such as that illustrated in Fig. 8. Each agent is connected to four others: top, bottom, left, and right. Agents in the model have multiple cultural 'features', each of which carries one of multiple possible 'traits'. One can think of the features as categorical variables and the traits as options or values within each category. For example, the first feature might represent culinary tradition, the second one the style of dress, the third music,

41846	09617	06227	73975	78196	98865	67856	39579	46292	39070
95667	34557	85463	49129	83446	31042	78640	70518	61745	96211
47298	86948	54261	75923	02665	97330	67790	69719	45520	37354
09575	72785	94991	70805	04952	52299	99741	12929	18932	81593
02029	94602	14852	94392	83121	84309	33260	44121	19166	73581
84484	93579	09052	12567	72371	08352	25212	39743	45785	55341
69263	94414	25246	68061	12208	44813	02717	90699	94938	05728
98129	44971	86427	26499	05885	45788	40317	08520	35527	73303
18261	18215	70977	15211	92822	74561	60786	34255	07420	42317
30487	23057	24656	03204	60418	56359	57759	01783	21967	84773

Fig. 8 Typical initial set of 'cultures' for a basic Axelrod-style model consisting of 100 agents on a 10 × 10 lattice with five features and 10 possible traits per agent. The marked site shares two of five traits with the site above it, giving it a cultural similarity score of 40% (Axelrod, 1997)

and so on. In the base configuration an agent's 'culture' is defined by five features, each having one of 10 traits ∈ {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. For example, agent A might have a cultural signature specified by traits {8, 7, 2, 5, 4} while agent B has a cultural signature specified by traits {1, 4, 4, 8, 4}.

Unlike in the Schelling model, Axelrod's agents are fixed in their lattice location and hence their interaction partners. Agent interaction and imitation rates are determined by neighbor similarity, where similarity is measured as the percentage of feature positions that carry identical traits. With five features, if a pair of agents share exactly one such element they are 20% similar; if two elements match then they are 40% similar, and so on. For each iteration, the model picks a random 'active' agent and one of its neighbors. With probability equal to their cultural similarity, the two sites interact and the active agent changes one of its dissimilar elements to that of its neighbor. If agent A {8, 7, 2, 5, 4} is chosen to be active and it is paired with its neighbor agent B {8, 4, 9, 5, 1}, for example, the two will interact with a 40% probability because they have two elements in common. If the interaction does happen, agent A changes one of its mismatched elements to match that of B, becoming perhaps {8, 7, 2, 5, 1}. This change creates a similarity score of 60%, yielding an increased probability of future interaction between the two.

In the course of approximately 80,000 iterations, the model process produces large areas in which cultures of traits on features are identical: Axelrod's 'local convergence.' But arrays such as that illustrated do not typically move to full convergence. They instead tend to produce a small number of stable and culturally isolated regions—groups of identical agents none of whom share features in common with adjacent groups and so cannot further interact. As an array develops, agents interact with increasing frequency with those with whom they become increasingly similar, interacting less frequently with the dissimilar agents. With only a mechanism of local convergence, small pockets of similar agents emerge that become increasingly homogeneous and increasingly isolated from other groups.

In even the simplest two-party models we've noted the importance of time. In two-dimensional models it can again make a great an important difference whether updating is simultaneous, as in the models considered here, or random, for example. Nowak and May instantiate fully cooperative and fully defecting strategies on a two-dimensional array, with agents imitating the strategy of highest-scoring neighbors (Nowak & May, 1992) ([8.4]). The result is both startling images of array evolution and the persistence of a robust percentage of both cooperative and defecting strategies. Those results, however, rely on simultaneous updating of the array. In critique, Huberman and Glance point out that both the startling images and results regarding persistence of cooperation disappear when updating is random rather than simultaneous (Huberman & Glance, 1993) ([8.5]).

Although both the Schelling and Axelrod models embody forms of spatial influence, neither approaches the full range such patterns can play, nor does either incorporate elements of strategy.

Consider the game of Go, in which white and black tokens are progressively put down onto a checkerboard by alternate players ([8.6]). In a simple form, players gain points by either occupying points or by surrounding and thus 'capturing' connected groups of the opponent's stones. With rules such as these, neither proportions of a population nor the simple neighborhood association of a Schelling or Axelrod model will dictate control of the board.

In an informational analogy, neither would be sufficient to map crucial details of patterns of influence. With strategies in play, influence becomes more complex still, here as in the case of strategic voting (Sect. 6).

9 Networks of influence

Two-dimensional lattices represent only one kind of interactive network: the ‘space’ of interactions need not be that of a lattice. A range of network studies thus also fall under the rubric of influence theory.

The ‘small worlds’ of Watts and Strogatz are random networks in which local ties of influence, picturable on a ring, are supplemented randomly with a few longer ties across the ring. Here one relevant network measure is the average path length between randomly chosen nodes. Another relevant measure is average clustering coefficient for nodes of the network: the extent to which the nodes to which a node are linked to each other, or the extent to which ‘one’s friends are friends of one another.’ More formally, the clustering coefficient of a node is the proportion of the number of links between its link-neighbors divided by the number of possible links between them.

A ring network has both high average clustering coefficient and high average path length. A fully random network has both lower average path length and lower average clustering coefficient as well. But path length and clustering coefficient do not change at the same rate in a progression from ring to random network. Between the extremes lie Watts and Strogatz’s ‘small worlds,’ notable for both high clustering coefficient and relatively low average path lengths (Fig. 9) ([9.1]).

Watts and Strogatz’s ‘small worlds’ are quite immediately models of social influence. It has often been proposed that such a model characterizes the way people are

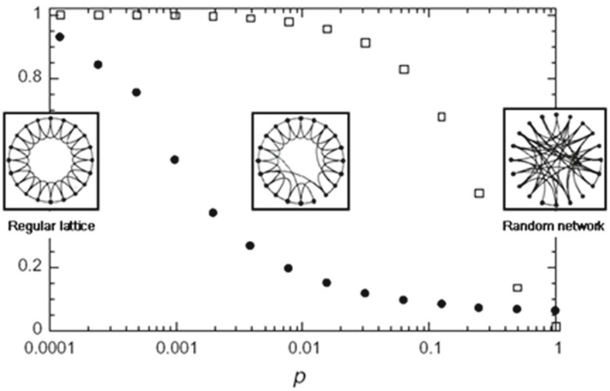

Fig. 9 The transition from a ring or regular lattice to a random network is marked by an increase in p of the probability of link rewiring, shown on a log scale. The upper curve is that of clustering coefficient for the network; the lower is a measure of characteristic path length

connected to each other in actuality, with a great deal of clustering—close groups of friends, communities, or villages—but also a few long-range links that connect those distant groups of friends, communities, or villages. Indications of such a pattern of influence taken from the sociological side, rather than from formal modeling, actually appeared earlier in Granovetter 1983.

A network model that captures network formation as well as structure—a feature noted in discussion of the Schelling model above—is the 'preferential attachment' model of Barabási & Albert, 1999 ([9.2]). Starting with a single node, additional nodes are added to a network with a higher probability of forming links to nodes that already have a higher number of links. The result is a network in which the numbers of links held by nodes show a scale-free or power law distribution. Here again such a model seems immediately applicable to data regarding social influence: power law distributions seem to characterize links on the internet, book sales, academic citations and much else. The preferential attachment model seems to capture not only common network patterns of influence but a plausible mechanism of their formation.

In introduction we mentioned a range of purposes for influence models, including purely formal models, applicational explanation models, and normative models regarding optimization of certain patterns of influence. Our examples have concentrated on formal and potentially explanatory models, but normative models come to the fore in recent network studies in social epistemology. What communication or influence network between scientists, for example, each with their own research data, will maximize scientific accuracy in the community at large? What network structure will maximize convergence on an agreed theory?

One might think that access to more data by more investigators would inevitably optimize the truth-seeking goals of communities of investigators. On that intuition, faster and more complete communication—the contemporary science of the internet—would allow faster, more complete, and more accurate exploration of nature. Surprisingly, however, many of the models at issue offer a robust argument for the potential benefits of *limited* communication. In the spirit of rational choice theory, much of this work was inspired by analytical work in economics on infinite populations by Bala and Goyal (1998), computationally implemented for small populations in a finite context and with an eye to philosophical implications by Kevin Zollman (2007, 2010a, 2010b) ([9.3]). In Zollman's model, Bayesian agents choose between a current method $\phi 1$ and what is set as a better method $\phi 2$, starting with random beliefs and allowing agents to pursue the investigatory action with the highest subjective utility. Agents update their beliefs based on the results of their own testing results—-drawn from a distribution for that action—together with results from the other agents to which they are communicatively connected. A community is taken to have successfully learned when all agents converge on the better $\phi 2$. Zollman's results are shown in Fig. 11 for the three simple networks in Fig. 10. The communication network which performs the best is not the fully connected network in which all investigators have access to all results from all others, but the maximally distributed network represented by the ring. But as Zollman shows, this is also that configuration which takes the longest time to achieve convergence. The normative study of epistemically optimal networks under different assumptions has become something of a growth industry, with important contributions by Hong & Page, 2004, Grim et al., 2013, Weisberg & Muldoon, 2009, O'Connor & Weatherall, 2018, 2019, and others.

The Zollman and related models employ Bayesian techniques as their updating rules, but Bayesian nets in general can also be seen as models of influence. In dynamic Bayesian nets, the value of nodes at a particular iteration are shaped by Bayesian inference on values of nodes at a previous iteration (Korb & Nicholson, 2004) ([9.4]). With node values read as agent opinions, andt with rules written in terms of Bayesian updating, the standard graphics for dynamic Bayesian networks show precisely the kinds of patterns of influence we have tracked throughout (Fig. 12).

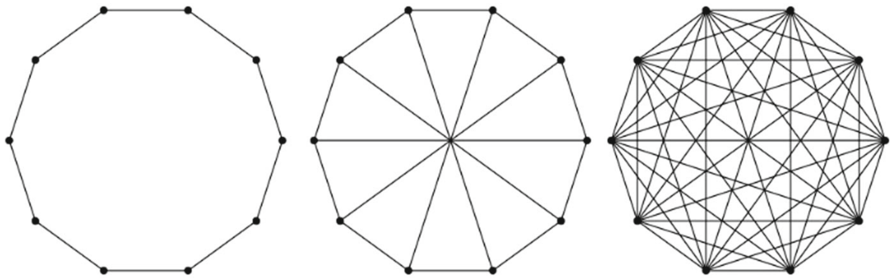

Fig. 10 A 10 person ring, wheel, and complete network

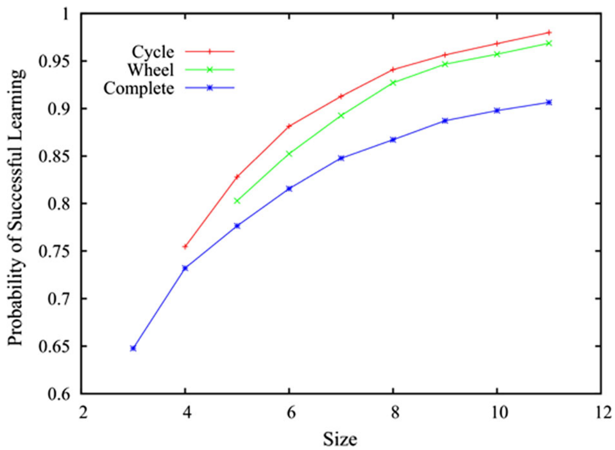

Fig. 11 Successful learning results with different network structures: ring, wheel, and complete networks of Bayesian agents. Adapted from Zollman (2010a)

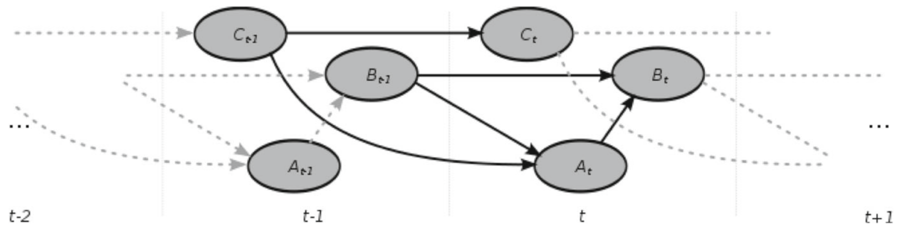

Fig. 12 Patterns of influence as dynamic Bayesian nets. From Korb & Nicholson (2004)

10 Complexity and undecidability in one-dimensional patterns of influence

One of the targets of influence theory is theorems regarding patterns of influence in general. In this regard, for example, it's clear that the complexity of even one-dimensional patterns of influence extends to computational universality and formal undecidability.

With just two opinions—*p* and ~ *p* or X and Y—our linear model constitutes a one-dimensional array of cellular automata (Wolfram, 2002), standardly thought of as extending infinitely to the left and right. We can then read formal results from the literature of such cellular automata as results regarding the potential complexity of patterns of influence.

Consider for example the 'neighborhood' of an agent together with its two immediate neighbors, and the rule that if one or two agents in an agent's neighborhood hold position *p* (perhaps including itself), it will then hold position *p*. If no cells in its neighborhood hold *p*, or if all three agents hold *p*, its opinion will be ~ *p*. Starting from a single agent with opinion *p* in a field that is otherwise occupied by 'doubters' holding ~ *p*, we get the pattern of the Sierpinski triangle on successive generations (Fig. 13) ([10.1]).

The outlined rule is that for linear cellular automata 126 in Wolfram's binary encoding: 01,111,110 for a cell's behavior given the neighborhood configurations of the previous generation shown in Fig. 14.

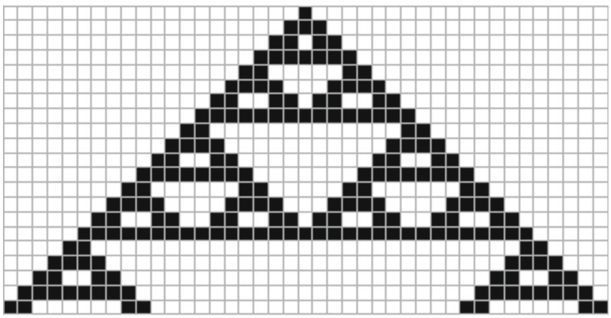

Fig. 13 Against a background of ~ *p* (white), a single *p* generates the well-known Sierpinski gasket using a rule that a cell takes a value of *p* if all three cells in the neighborhood are *p* or ~ *p*, a value of ~ *p* otherwise

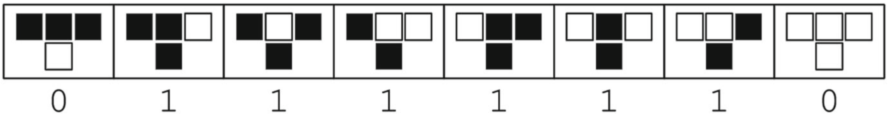

Fig. 14 Wolfram coding for linear cellular automata rules, rule 126 shown

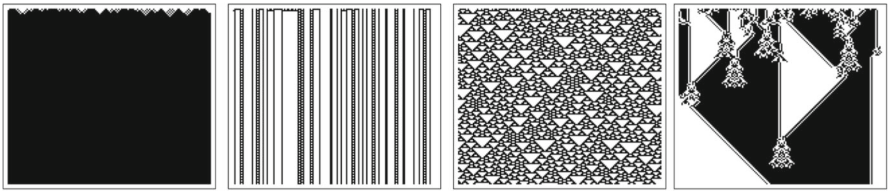

Fig. 15 Examples of Wolfram classes as alternative dynamics of influence given different rules: convergence on a uniform *p* or ~ *p* (class 1), stable or simple periodic states (class 2), apparent randomness (class 3), and the more complex interacting areas of class 4

'The fact that one-dimensional cellular automata can be interpreted as patterns of influence means that their formal properties will hold for influence as well. With the right rules, local influence can show any of the dynamics documented in Wolfram's four classes, for example: convergence on uniform *p* or ~ *p*, the establishment of stable of simple periodic states, apparently random behavior, or complex interacting areas (Fig. 15).

Of particular interest in this regard is a 'right-handed' rule 110, which differs from 126 only in that a single value of *p* to the left of a cell is insufficient to convert it to *p* (Fig. 16).

In a result conjectured by Wolfram but proven by Matthew Cook (2004), rule 110 is computationally universal: any Turing machine can be simulated with an appropriate initial configuration using rule 110. It follows is that the behavior of influence of even this simple linear form is formally undecidable: algorithmically undecidable in principle ([10.2]).

Suppose that there were a calculation which could tell us, with full accuracy and in a finite number of steps, what the eventual dynamics would be of an initial configuration of this sort using a rule such as 110. Since any Turing machine can be instantiated as such a configuration (Cook, 2004), the algorithm we have supposed would be able to tell us whether any arbitrary Turing machine, started on an arbitrary input, would halt or not. By the Halting problem, we know that no Turing machine can tell us, for arbitrary Turing machines and inputs, whether that machine will halt on that input or not (Turing 1937). By Church's Thesis, anything that can be computed at all can be

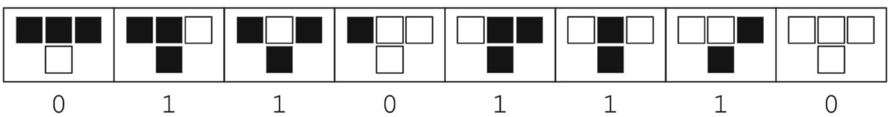

Fig. 16 Wolfram coding for cellular automata rule 110

computed by some Turing machine (Wood, 1987). There can then be no algorithm and no finite decision procedure that can predict the behavior of arbitrary linear cellular automata arrays. Since these can map initial configurations of influence, there is no algorithm and no finite decision procedure that can predict the outcome of arbitrary patterns of influence.

11 Complexity and undecidability in two-dimensional patterns of influence

Computational universality and formal undecidability were instantiated in two-dimensional cellular automata before Wolfram's one-dimensional case: first in the work of John von Neumann, using a spatial neighborhood of four adjacent cells and rules involving, in effect, 29 traits, and later in the work of John Horton Conway's Game of Life (Berlekamp et al., 2001).

In the Game of Life, cells with an immediate 2-dimensional neighborhood of 8, including cells on the diagonal, take binary values of 'alive' or 'dead' according to a simple set of rules:

> If dead, come alive just in case three neighbors are alive.
>
> If alive, remain alive just in case two or three neighbors are alive.

With 'alive' and 'dead' as holding belief *p* and ~ *p*, an information or persuasion interpretation of Conway's rules could take the form:

> If you believe ~ *p*, convert to *p* if exactly three neighbors are convinced of *p*.
>
> If you believe *p*, convert to ~ *p* unless exactly two or three neighbors are convinced of *p*.

Here agents are sensitive to their local population of opinion, but in two different ways. A full three neighbors of the opinion *p* are required for an agent to convert from ~ *p* to *p*, but more than that and the view becomes *too* popular and cliché, and the agent reverts to ~ *p*. It is also The rules of conversion are not symmetrical: changing from *p* to ~ *p* requires a different configuration than changing from ~ *p* to *p*.

As Conway and his followers have demonstrated in detail, the mechanisms of both any Turing machine and any equivalent Minsky register machine can be instantiated in two-dimensional configurations in the Game of Life. By the Halting problem, we know that no Turing machine can tell us, for arbitrary Turing machines and inputs, whether that machine will halt on that input or not. By Church's Thesis, anything that can be computed at all can be computed by some Turing machine. Thus there can be no algorithm and no finite decision procedure that can predict the behavior of two-dimensional cellular automata arrays. Since these can map initial configurations of influence, there is no algorithm and no finite decision procedure that can predict the outcome of arbitrary patterns of influence ([11.1]).

Here again there is a connection with game theory. Iterative game theoretic strategies can be instantiated as cells in the array, and with proper strategies can form the ‘wires’, logic gates and memory units required to model any Minsky register machine. The behavior of such machines can be constructed so as to result in an infinite ‘explosion’ of a given type of strategy or its containment in ways that parallel the undecidability proof for the Halting Problem. Thus spatialized game theory is again an instance of universal computability and formal undecidability in patterns of influence (Grim, 1997; Grim et al., 1998) ([11.2]).

12 Some initial lessons from influence theory

Influence theory, like decision theory and game theory, is an abstract and synoptic venture in bringing a vast variety of different albeit related phenomena into the unified purview on a single discipline.

Obviously the great bulk of relevant models and phenomena will come from the diversified multitude of special situations determined by the relevant range of case studies. What does a generalized integrating theory have to add to this? The answer is a set of assets that no scattered collection of class studies can provide:

- a unifying perspective that reveals the commonalities of cases and lays the basis for their classification and interrelationships.
- a view of how the specificities of cases fit them into the cognitive articulation of a larger sphere of investigation.
- an exposition of the cross-issue commonalities of explanatory processes and procedures.
- A broader theoretical contextualism of a wide range of phenomena and a large-scale mapping of their place in the cognitive scheme of things.

The scientific project, after all, has a highly complex mission, not only to account for phenomena by way of description and explanation, but to give a second order explanation of why our first-order accounts of phenomena take the form they do. Thus science not only addresses individual cases by way of description and explanation but attempts to find their place in a larger cognitive architecture. The organization of our cognitive enterprise into disciplines such as influence theory is an integral part of this larger project of scientific understanding.

Declarations

Conflict of interest The authors declare there are no competing interests or relevant funding sources.

Ethical approval No ethics approval or informed consent was required.

Appendix

A compendium of examples above in terms of parameters outlined in Sect. 2:

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
4.1		2 binary X,Y	1-Way influence	N/A	Influence on neighbor	formal	Immediate unanimity
4.2		2 binary X,Y	Mutual influence	Simultaneous	Influence on neighbor	formal	Oscillation from XY, YX
4.3		2 binary X,Y	Mutual influence	Sequential	Influence on neighbor	formal	First actor determines unanimity
4.4		2 binary X,Y	Mutual influence, Proba-bilistic	Simultaneous	Probability p of influence on neighbor	formal	Asymptotic approach to equilibrium
4.5		2 binary X,Y	Mutual influence, proba-bilistic	Sequential	Probability p of influence on neighbor	formal	Asymptotic approach to equilibrium, faster convergence than simultaneous

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
4.6		2 traits, adaptable to multiple	Mutual influence	Simultaneous & sequential variations	Influence on neighbor at particular trait, deterministic or probabilistic	formal	Allows a measure of degree of influence in terms of number of trait changes
4.7		Continuous scale on single trait	Mutual influence	Simultaneous & sequential variations	Each agent influences the other to move halfway to its position, or recalcitrant agents	formal	Various equilibria given differences in timing, rule and agent recalcitrance
4.8		2 binary, continuous, multiple	Mutual influence	Strategically handled influence	Probabilistic influence	formal	Altered probabilities with unbalanced influence events

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
5.1 Population model	Multiple, proportions of population	2 binary X,Y	Fully connected or random mixing	Simultaneous all with all	Probability of conversion of neighbor	formal	Markov model, unique equilibrium regardless of initial proportions, by Perron-Frobenius theorem
5.2 SIR, SIRS infection models	Multiple, proportions of population	Susceptible infected recovered	Random mixing	Simultaneous all with all	Infection of neighbor with R_0 of pathogen, recovered immunity (SIR) or possible reinfection (SIRS)	explanatory predic-tive	Classical S-curve of infection for SIR model, endemic equilibrium possible in SIRS model
5.3 Deterministic cycle through states	Multiple, proportions of population	2 binary X,Y	Fully connected	Simultaneous all with all	100% probability of conversion of neighbor	formal	Violates conditionsof Markov model: oscillation

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
5.4 Variable transition probabilities, Grim Mar St. Denis (1998)	Multiple, proportions of population	2 binary X, Y and various	Fully connected	Simultaneous or sequential	Probabilistic conversion	formal	Equilibrium, oscillatory, periodic, and chaotic all possible dyanmics
5.5 Granovetter (1978)	Multiple, with hetero-geneous thresholds	2 binary: riot, not	Fully connected	Simultaneous	Agents with heteroge-neous thresholds convert to 'riot' given specific proportions of the population rioting	explanatory	Rioting population depends not merely on average thresholds to riot but on distribution of thresholds

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
5.6 Lehrer and Wagner (1981)	Multiple	Multiple agents have opinion probability and reliability weights for other agents	Fully connected	Iterated simultaneous	Agents change opinion probabilities to weighted average of population	Formal, normative	Consensus in which all agents share same opinion probabilities develops as Markov process
6.1	Multiple	2 binary X, Y	Linear, 1-dimensional	Simultaneous	X changes to Y given a Y neighbor, or does so with probability p	formal	Total conversion to Y given any Y, speed dependent on probability
6.2	Multiple	2 binary X, Y	Linear, 1-dimensional	Simultaneous	Y changes to X if surrounded by X on both sides	formal	Results dependent on initial configuration

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
6.3	Multiple	2 binary X, Y	Linear, 1-dimensional	Simultaneous	Non-symmetrical: Y preceded by 2 X's will convert to X	formal	Results dependent on initial configuration
6.4	Multiple	2 binary X, Y	Linear, 2-dimensional, finite alternation of X and Y	Simultaneous	Conversion to other view when surrounded on each side	formal	Results dependent on initial configuration
6.5	Multiple	2 binary X, Y	Linear loop	Simultaneous	Conversion to other view when surrounded on each side	formal	Oscillation with loop at one point; unanimity with loop at another

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
7.1 shared information	Multiple, with individual evidence as well as cumulative	Binary	Linear, 1-dimensional	Sequential along line	Agents share decisions and evidence down the line	formal	Given a small probability of correctness, individuals down the line will accumulate evidence toward the correct decision with probability 1
7.2 Information cascades Bikhchandani et al. (1998)	Multiple, with individual evidence as well as cumulative	Binary	Linear, 1-dimensional	Sequential along line	Agents share decisions but not individual evidence, make decisions on cumulative data	formal, explanatory	Cascades of decision, with calculable probabilities of misinformation

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
7.3 Information cascades in enclaves	Multiple in isolated enclaves, individual and cumulative evidence	Binary	Linear, 1-dimensional	Sequential along lines of each enclave	Agents share decisions but not individual evidence, make decisions on cumulative data	formal, explanatory	Increased probability of polarization, increased probability of misinformation in at least one enclave
7.4 Information cascades on trees	Multiple in isolated enclaves, individual and cumulative evidence	Binary	Tree network	Sequential from earlier on tree	Agents share decisions but not individual evidence, make decisions on cumulative data	formal, explanatory	Similar cascade phenomena

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
7.5 Fedderson Pesendorfer 1998; Austin-Smith Feddersen (2005, 2006)	Multiple agents	Private signals with 'bias' or standard of reasonable doubt	Fully connected	Simultaneous voting, or voting after straw poll, with either unanimous or majority decision rule	Agents may vote or share information on private signals strategically	Formal, explanatory normative	Incentives for strategic voting are different under unanimity and majority rule; error rates are higher under unanimity
7.6 Hegselmann Krause (2002, 2005, 2006)	Multiple agents	Continuous scale of opinion	Association network with similarity threshold on trait	Simultaneous	Agents' opinions move toward average of those in their threshold range	formal, explanatory	With high threshold, convergence on opinion. With low threshold, polarization of opinion in population

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
8.1 Schelling, (1969, 1971, 1978)	Multiple agents	2 unchanging traits or colors X,Y	2-dimensional lattice with low density (empty cells)	Random	Agents do not change traits, but move on array when number of 'like' neighbors is below threshold t	formal, explanatory	With relatively low threshold (30%), clear clustered areas of residential segregation appear
8.2 Schelling with high, low density and heterogeneous thresholds	Multiple agents	2 unchanging traits or colors X,Y	2-dimensional lattice with low or high density	Random	Agents move when number of like neighbors is below threshold t, but one set of agents have a 'don't care' low threshold	formal, explanatory	At low density, agents do not cluster. At high density, they are forced to

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
8.3 Cultural divfusion Axelrod (1997)	Multiple agents	5 traits, each with 1 of 10 values	Agents fixed in a 2-dimensional lattice	Neighbor pairs chosen at random, interact with a probability correlate to number of traits in common	Given interaction, the 'active' agent changes one of its non-matching traits to that of the other agent	formal, explanatory	Wide areas of convergence a rossall traits with islands of isolated 'cultures'
8.4 Imitative game theory Nowak & May (1992, 1993)	Multiple agents	2 strategies: Cooperate, Defect	2-dimensional lattice	Simultaneous	Play rounds of Prisoner's Dilemma, then imitate most successful neighbor	formal	Maintained proportion of cooperative strategies; symmetrical patterns

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
8.5 Imitative game theory Huberman & Glance (1993)	Multiple agents	2 strategies: Cooperate, Defect	2-dimensional lattice	Random	Play rounds of Prisoner's Dilemma, then imitate most successful neighbor	formal	Symmetrical patterns destroyed; results for cooperation altered from simultaneous case
8.6 Game of Go	Multiple agents	2 traits, black or white	2-dimensional lattice	Sequential	Strategic moves attempting to occupy points or surround opponent	formal	Strategically complex

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
9.1 Small worlds Watts & Strogatz (1998)	Multiple agents	Various	Range of networks from ring with probability of ring rewiring to random	Simultaneous	Influence between nodes on links	formal, explanatory	With ring networks, clustering coefficient high and characteristic path length long. At random, path length short but clustering coefficient low. In 'small worlds,' high clustering with short path lengths

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
9.2 Preferential attachment networks Barabási & Albert (1999)	Multiple agents	Various	Networks formed with new nodes attaching with higher probability to nodes with many links	Sequential addition	Influence between nodes on links	formal, explanatory	Scale free networks form
9.3 Epistemic networks Zollman (2007, 2010a, 2010b)	multiple	Choice of theories as to the higher-paying of 2 'bandits'	Sample networks: ring, wheel, and complete	Simultaneous	Update chosen theory on the basis of individual testing of a chosen bandit and input from linked nodes	formal, normative	Ring formations resuilt in slower community convergence with higher group accuracy probability

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
9.4 Bayesian networks	multiple	Various	Directed acyclic graphs	Sequential	Bayesian updating of nodes from current links and previous values	formal	Various dynamics given differences in timing, rule, and agent recalcitrance
10.1 Wolfram rule 126	multiple	Black or white	Linear, 1-dimensional	Simultaneous	Cell is black just in case self and two neighbors not all blck or all white on previous generation	formal	Forms Sierpinski triangle
10.2 Wolfram rule 110	multiple	Black or white	Linear, 1-dimensional	Simultaneous	'right-handed': like 126 except a single black to the left is insufficient to make a cell black on the next generation	formal	Computationally universal, formally undecidasble (Cook, 2004)

Model	Agents	Traits	Network	Timing	Mechanism	Purpose	Dynamics
11.1 Game of Life	multiple	Alive or dead	2-dimensional lattice	Simultaneous	If alive, stays alive on next iteration iff 2 or 3 neighbors are currently alive. If dead, stays dead unless 3 neighbors are currently alive	formal	Computationally universal, formally undecidasble (Berlekamp, Conway & Guy 1982)
11.2 Spatialized Prisoner's Dilemma	Multiple	Many states to form electrons on wires, logic gates, memory units	2-dimensional lattice	Simultaneous	Cells play Prisoner's Dilemma with neighbors, convert to most successful local strategy	formal	Computationally universal, formally undecidasble (Grim, Mar & St. Denis 1998)

References

Anderson, L., & Holt, C. (1997). Information Cascades in the Laboratory. *American Economic Review, 87*, 847–862.

Austin-Smith, D., & Feddersen, T. J. (2005). Deliberation and voting rules. In D. Austin-Smith & J. Duggan (Eds.), *Social choice and strategic decisions: Essays in Honor of Jeffrey S. Banks.* Springer.

Austin-Smith, D., & Feddersen, T. J. (2006). Deliberation, preference uncertainty, and voting rules. *American Political Science Review, 100*(2), 209–217.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution, 41*, 203–226.

Bala, V., & Goyal, S. (1998). Learning from neighbours. *Review of Economic Studies, 65*(3), 595–621. https://doi.org/10.1111/1467-937X.00059

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5), 09–312.

Berlekamp, E., Conway, J. H., & Guy, R. K. (2001). *Winning ways for your mathematical plays* (2nd ed., Vol. 1). A K Peters/CRC.

Bhattacharya, A. (2021). *The man from the future: The visionary life of John von Neumann*. W. W. Norton.

Bikhchandani, S., Hershleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives, 12*(3), 151–170.

Burks, A. (1966). *Theory of self-reproducing automata*. University of Illinois Press.

Castllano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social phenomena. *Review of Modern Physics, 81*, 591–646.

Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Pearson.

Cook, M. (2004). Universality in elementary cellular automata. *Complex Systems, 15*, 1–40.

Coughlan, P. J. (2000). In defense of unanimous Jury verdicts: Mistrials, communication and strategic voting. *American Political Science Review, 94*, 375–393.

Deffuant, G., Amblard, F., Weisbuch, G., & Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation, 5*(4).

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association, 69*(345), 118–121.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

Fedderson, T., & Pesendorfer, W. (1998). Convicting the Innocent: The Inferioriutyof Unanimous Jury Verdicts under Strategic Voting. *American Political Science Review, 92*(1), 23–35.

French, J. R. P. (1956). A Formal Theory of Social Power. *Psychological Review, 63*(3), 181–194.

Galam, S., Gefen, Y., & Shapir, Y. (1982). Sociophysics: A new approach of sociological collective behavior. I. Mean-behavior description of a strike. *Journal of Mathematical Sociology, 9*, 1–13.

Galam, S., & Moscovici, S. (1991). Towards a theory of collective phenomena: Consensus and attitude changes in groups. *European Journal of Social Psychology, 21*, 49–74.

Galam, S. (1997). Rational group decision making: A random field Ising model at T=0. *Physica a: Statistical Mechanics and Its Applications, 238*, 66–80.

Galam, S. (2002). Minority opinion spreading in random geometry. *European. Physical Journal B, 25*, 403.

Galam, S. (2012). *Sociophysics: A physicist's model of psycho-political phenomena*. Springer.

Glymour, C., & Cooper, G. F. (1999). *Computation, causation and discovery*. AAAI Press.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*(6), 1360–1380.

Grim, P. (1995). The greater generosity of the spatialized prisoner's Dilemma. *Journal of Theoretical Biology, 173*, 353–359.

Grim, P. (1997). The Undecidability of the spatialized prisoner's Dilemma. *Theory and Decision, 42*, 53–80.

Grim, P. (2009). "Threshold Phenomena in Epistemic Networks." *Proceedings, AAAI Fall Symposium on Complex Adaptive Systems and the Threshold Effect*, AAAI Press.

Grim, P., Mar, G., & St. Denis, P. (1998). *The philosophical computer*. MIT Press.

Grim, P., Kokalis, T., Alai-Tafti, A., Kilb, N., & St. Denis, P. (2004). Making meaning happen. *Journal for Theoretical and Artificial Intelligence, 16*, 209–244.

Grim, P., Singer, D. J., Fisher, S., Bramson, A., Berger, W. J., Reade, C., & Sales, A. (2013). Scientific networks on data landscapes. *Episteme, 10*(4), 441–464.

Grim, P., Singer, D. J., Reade, C., & Fisher, S. (2015). Germs, Genes, and memes: Function and fitness dynamics on epistemic networks. *Philosophy of Science, 82*(2), 219–243.

Hararay, F. (1959). A criterion for unanimity in French's theory of social power. In D. Cartwright (Ed.), *Studies in Social Power* (pp. 168–182). Institute for Social Research.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. Macmillan.

Hegselmann, R., & Krause, U. (2002). Opinion Dynamics and Bounded Confirdence Models, Analysis and Simulation, *Journal of Artificial Societies and Social Simulation*, *5*(3).

Hegselmann, R., & Krause, U. (2005). Opinion dynamics driven by various ways of averaging. *Computational Economics, 25*(4), 381–405.

Hegselmann, R., & Krause, U. (2006). Truth and cognitive division of labour: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation, 9*(3).

Helbing, D. (1991). A mathematical model for the behavior of pedestrians. *Behavioral Science, 36*, 298–310.

Hong, Lu., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*(46), 16385–16389. https://doi.org/10.1073/pnas.0403723101

Huberman, B. A., & Glance, N. S. (1993). Evolutionary games and computer simulations. *Proceedings of the National Academy of Science USA, 90*, 7716–7718.

Kahan, M. (1997). Social influence, social meaning, and deterrence. *Virginia Law Review, 83*, 276–304.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A, 115*(772), 700–721.

Kermack, W. O., & McKendrick, A. G. (1932). Contributions to the mathematical theory of epidemics-II. The problem of endemicity. *Bulletin of Mathematical Biol*ogy 1991; 53(1–2): 57–87. https://doi.org/10.1007/BF02464424

Korb, K. B., & Nicholson, A. E. (2004). *Bayesian artificial intelligence*. Chapman and Hall/CRC.

Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society*. D. Reidel.

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive*. 8th revised edition. Harper and Brothers, 1881.

Nowak, M., & May, R. (1992). Evolutionary games and spatial chaos. *Nature, 359*, 826–829.

Nowak, M., & May, R. (1993). The spatial dimensions of evolution. *International Journal of Bifurcation and Chaos, 3*, 35–78.

Nowak, M., & Sigmund, K. (1992). Tit for tat in heterogeneous populations. *Nature, 355*, 250–252.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme, 17*(2), 141–161.

O'Connor, C., & Weatherall, J. (2018). Scientific polarization. *European Journal for Philosophy of Science, 8*(3), 855–875.

O'Connor, C., & Weatherall, J. (2019). *The misinformation age*. Yale University Press.

Page, S. (2018). *The model thinker*. Basic Books.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.

Rescher, N. (1998). *Predicting the future: An introduction to the theory of forecasting*. State University of New York Press.

Schelling, T. C. (1969). Models of segregation. *American Economic Review, 59*(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(2), 143–186.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. W. W. Norton.

Sunstein, C. (2001). *Republic.com*. Princeton University Press.

Sunstein, C. (2007). *Republic.com 2.0*. Princeton Universiy Press.

Turing, A. M. (1937). On Computable Numbers, with an Application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society, Series, 2*(42), 230–265.

Walden, E., & Browne, G. (2002). Information cascades in the adoption of new technology. In *International Conference on Information Systems 2002 Proceedings* (Vol. 40).

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442. https://doi.org/10.1038/30918

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science, 76*(2), 225–252. https://doi.org/10.1086/644786

Wiedlich, W. (1971). The statistical description of polarization phenomena in society. *British Journal of Mathematical and Statistical Psychology, 24*, 251–266.

Wiedlich, W. (1991). Physics and social science—The approach of synergetics. *Physics Reports, 204*, 1–163.

Wiedlich, W. (2002). *Sociodynamics—A systematic approach to mathematical modeling in the social sciences*. Taylor & Francis.
Wolfram, S. (2002). *A new kind of science*. Wolfram Media.
Wood, D. (1987). *Theory of computation*. Harper and Row.
Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science, 74*(5), 574–587. https://doi.org/10.1086/525605
Zollman, K. (2010a). The epistemic benefit of transient diversity. *Erkenntnis, 72*(1), 17–35. https://doi.org/10.1086/525605
Zollman, K. (2010b). Social structure and the effects of conformity. *Synthese, 172*(3), 317–340. https://doi.org/10.1007/s11229-008-9393-8
Zollman, K. (forthcoming). *Network epistemology: how our social connections shape knowledge*. Oxford University Press.