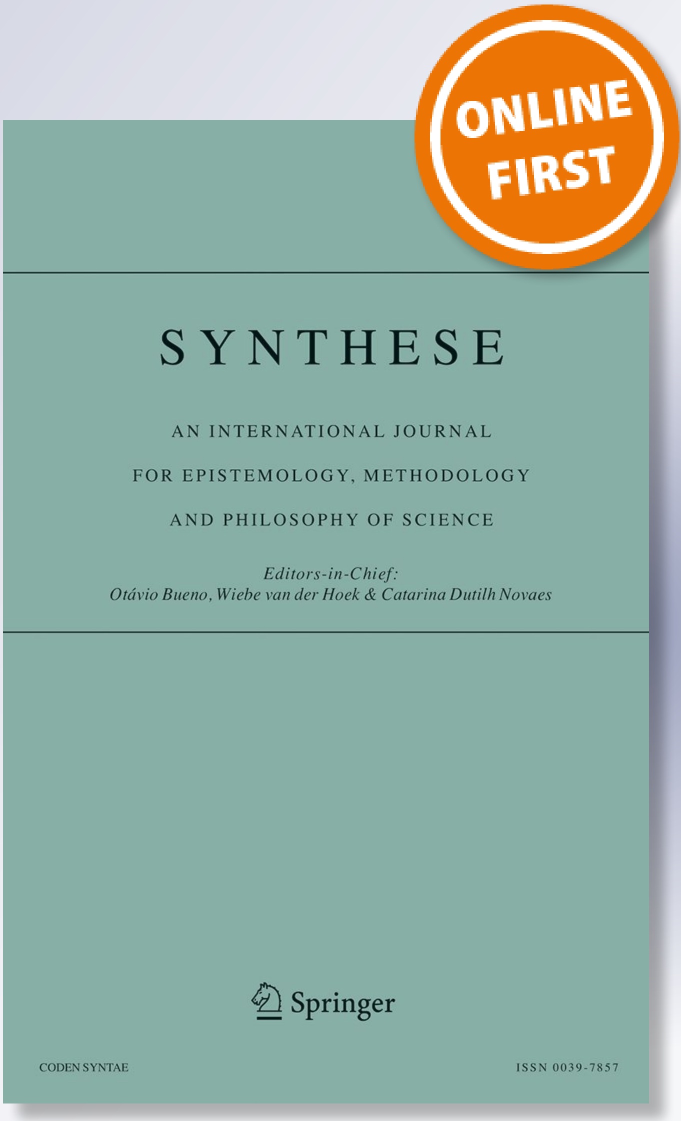
Arational belief convergence

Charles Lassiter

Synthese
An International Journal for Epistemology, Methodology and Philosophy of Science

ISSN 0039-7857

Synthese
DOI 10.1007/s11229-019-02465-6



Synthese
https://doi.org/10.1007/s11229-019-02465-6

COMPUTATIONAL MODELLING IN PHILOSOPHY

Arational belief convergence

Charles Lassiter[1]

Received: 1 March 2019 / Accepted: 4 November 2019

Abstract
This model explores consensus among agents in a population in terms of two properties. The first is a probability of belief change (PBC). This value indicates how likely agents are to change their mind in interactions. The other is the size of the agents audience: the proportion of the population the agent has access to at any given time. In all instances, the agents converge on a single belief, although the agents are arational. I argue that this generates a skeptical hypothesis: any instance of purportedly rational consensus might just as well be a case of arational belief convergence. I also consider what the model tells us about increasing the likelihood that one agent's belief is adopted by the rest. Agents are most likely to have their beliefs adopted by the entire population when their value for PBC is low relative to the rest of the population and their audience sizes are roughly three-quarters of the largest possible audience. I further explore the consequences of dogmatists to the population; individuals who refuse to change their mind end up polarizing the population. I conclude with reflections on the supposedly special character of rationality in belief-spread.

Keywords Agent-based model · Consensus · Arational · Skepticism · Digital humanities

1 Introduction

The defendant's fate is all but sealed at the start of *12 Angry Men*. He's a vaguely "ethnic" boy from the slums on trial for murdering his father. Jurors are skeptical of his moral standing from the start. A thin alibi only serves to fuel their mistrust, and an eye-witness puts the boy at the scene of the crime. Despite the evidence Juror 8 isn't convinced that the boy is guilty. He offers alternative interpretations of the evidence. "How," Juror 8 asks, "how could an old woman witnessing the murder through the windows of a moving L train be certain that *this* kid is the one who did it?" The others

✉ Charles Lassiter
Charles.lassiter@gmail.com

1 Philosophy Department, Gonzaga University, Spokane, USA

Published online: 21 November 2019

begin to change their votes as Juror 8 pokes holes in the prosecutor's case. But there is one hold-out. Juror 3 offers tortured and implausible reasons for why the boy is guilty. His arguments grow desperate as the others increase the pressure for him to see the light. Juror 3 reaches his breaking point and reveals his motivations: he has a strained relationship with his own son and wants the boy to be guilty. He soon after switches his vote, making the "not guilty" verdict unanimous.

The film is intense and dramatic: a case study in prejudice, emotion, and reason. The hook for the play is how the minority opinion eventually becomes the majority. Epistemological questions abound. An important one is: By what model should we conceptualize what happens? One natural way is to describe the events in terms of modifications of credences, where credences are degrees of belief in a proposition. Call these models 'credence-adjustment models.' All jurors except number 8 assign a value less than 0.1 to "not guilty" and greater than 0.9 to "guilty". Juror 8, by contrast, assigns a credence of less than 0.10 to "guilty" and greater than 0.90 to "not guilty". Through some rational process, each agent adjusts his credences with the arguments of Juror 8 to eventually favor a vote of "not guilty."

The Lehrer–Wagner model is a credence-adjustment model (Lehrer and Wagner 1981; cf. DeGroot 1974).[1] In this model, each agent has two properties: a credence concerning some belief and a weight assigned to others in the group.[2] The weight captures how much one agent values the opinions of another agent. Alice might think highly of Betty but quite poorly of Carl; so Alice would assign a higher weight to Betty than to Carl.[3] Agents' credences are updated by taking weighted averages of others' credences. Eventually, consensus is reached.[4]

Credence-adjustment models aren't the only ones on offer. Another natural way is to think of each juror having a unique belief related to whether the boy should be found guilty or not. These beliefs could be something like "the boy is certainly guilty" or "the boy deserves to go to prison." Jurors then compare their evidence, each adopting whichever belief they think is best supported by the evidence. Call these 'unique-belief models.'

Zollman (2015) is a unique-belief model.[5] Here, each agent is tasked with acquiring as many true beliefs and as few false beliefs as possible via testimony. Each agent either believes, disbelieves or withholds judgment concerning a wide range of beliefs. To acquire beliefs, each agent samples the world with a probability of 0.10 (and has around a 0.60 probability of making a correct inference from that observation) and interacting with other agents. When an agent A withholds judgment about a belief and solicits the attitude of agent B towards that belief, A adopts B's attitude. The model isn't

[1] Other credence-adjustment models include, among many others, Hegselmann and Krause (2002), Bala and Goyal (1998), Flache et al. (2017) and Easwaran et al. (2016).

[2] Thanks to an anonymous reviewer for catching poor wording here.

[3] One problem considered by Lehrer and Wagner is the condition under which it is rational to assign someone zero weight, effectively communicating that that agent's beliefs aren't worth consideration. Lehrer's solution is that a person X should assigned a weight of zero to person Y just in case X has no preference between Y's judgments and that of a random device. See Forrest (1985) and Martini et al. (2013) for discussion.

[4] Lehrer (1976) argues that, as long as each member of the group has *some* positive regard for every other member of the group, rational disagreement is impossible.

[5] Baumgaertner (2014) is another example.

interested in consensus per se but rather on how members of doxastic communities come to endorse as many true beliefs as possible. The aim of the community is to converge on the truth.

What both unique-belief models and credence-adjustment models share in common is a commitment to modeling belief- and credence-adjustments as broadly *rational* processes. A process is rational if reasons of some variety can be offered by the agent to explain why she changed her mind. If Alice changes her mind because of what Betty says and cites that Betty is a reliable source of information, then Alice is employing a rational process of belief change. By contrast, if Alice agrees with Betty instead of Carl because she flipped a coin, then Alice does not employ a rational process. Appealing to Betty's believing that *p* as a reason for coming to believe that *p* is a rational process; flipping a coin is neither rational nor irrational. There's a lot of conceptual space between these two cases. Nisbett and Wilson (1977), to take one example, found that students tend to prefer and find more persuasive teachers who are physically attractive, a phenomenon that has come to be called the 'halo effect.' Suppose they're right. Then the reasons students offered for their beliefs were products of the instructor's perceived attractiveness and not the force of the instructor's reasons. Are the students' beliefs rational or not? Since there's nothing lost in the upcoming arguments by adopting a more permissive view with these sorts of cases, I'll do just that. On this permissive view, the students' belief-formation processes count as rational provided they're able to offer some kind of reason in favor of their beliefs, even if those reasons are post hoc and epiphenomenal.

Raz (1999) offers a useful way to characterize the notion of rationality common to unique-belief and credence-adjustment models:

> An account of rationality is an account of the capacity to perceive reasons and conform to them, and of different forms of conforming to reasons, and their appropriateness in different contexts. To explain the capacity to conform to reason the account must explain the possibility of error, failure to perceive reasons correctly, and of failure to respond to them once perceived...The core idea is that rationality is the ability to realize the normative significance of the normative features of the world, and the ability to respond accordingly. In one sense of 'rational', we, or anything else, are rational beings to the extent that we possess that ability, which I will call 'capacity-rationality.'

Unique-belief and credence-adjustment models encode computational agents in a way to facilitate interpreting them as exhibiting capacity-rationality; they're built to be seen as rational. Why think this? Agents in Bala and Goyal's (1998) model update their credences by means of Bayes' rule, which is often presented as a normative rule for adjusting credences (cf. Talbott 2016). Agents in Zollman (2015) attempt to match testimonial beliefs with observations about the world, a conception of belief formation and justification comfortably at home in veritistic social epistemology (Goldman 1999). The Lehrer–Wagner model (which functions much like the DeGroot model) is set out in a book titled "*Rational* Consensus." Even 'naïve' learners who fail to appropriately adjust for repeated information or dependencies in learned information (cf. Golub and Jackson 2010) exhibit capacity-rationality. Such naïve learners may fall short of what the norms of rationality require, but there is nonetheless a reckoning by those norms. Everything speaks for and nothing against interpreting agents in the models as employing rational means for adjusting beliefs and credences.[6]

In this paper, I offer an alternative to capacity-rationality-based views for achieving consensus. The trouble isn't about discerning whether some or other averaging process better captures pre-theoretical intuitions about rationality (cf. Martini et al. 2013). Or if Bayesian adjustments really do capture some ideal of rationality.[7] Rather, capacity-rationality is an unnecessary assumption for explanations of consensus.[8] There is a precedent of sorts in the history of philosophy. Schopenhauer (Schopenhauer 2014/1818) argued that the noumena to worldly phenomena was an arational force called 'Will.' In a similar spirit, I argue that consensus is an instance of an arational construct, Probability of Belief Change (PBC). An agent's PBC is the value indicating the likelihood the agent will change her mind in any given interaction. At the extremes, a PBC of 0 means the agent never changes her mind. A PBC of 1.0 means the agent always changes her mind. Modeling consensus with PBC as the primary mechanism by which agents change their minds is called the model of Arational Belief Convergence (or 'ABC model'). Returning to our jurors, Juror 8 had a PBC of 0: nothing could shake him from his belief that the boy was not guilty. Juror 3 has a very low—but non-zero—PBC.

In what follows, I describe PBC in greater detail and give five arguments for its validity as a psychological construct.[9] After that, I'll describe two simulations employing PBC to reach consensus and results from running the simulation. I'll then argue that this proposes a skeptical hypothesis to belief convergence and then respond to objections.

2 Probability of belief change

An agent's PBC is the likelihood that she will change her mind in a given interaction.To get a feel for the construct, imagine that you loathe—nay, *despise*—salmon. When offered salmon at a wedding reception, you're highly likely to give it a hard pass. When out to dinner, you'll choose just about anything over salmon. But now imagine you're at a meal with people you know and trust and who are well-acquainted with your opposition to salmon. Suppose they tell you that the salmon is delicious; it's a combination of flavors that is absolutely up your alley. You politely decline, but they insist. "You'll love it!" they say. You insist more firmly that you would not, could not,

[6] Thanks to an anonymous reviewer for suggestions to clarify the ideas in these last two paragraphs.

[7] See Silver (2012) for one man's love letter to the right Reverend Bayes.

[8] None of these authors have argued that capacity-rationality is necessary for consensus. Even so, a review of the literature would suggest a broad, de facto consensus that capacity-rationality is at work in convergence. So if any of the above authors wish to deny that capacity-rationality is captured in their models, that's fine! The arguments I develop in this paper would only require a few tweaks to address the descriptive gloss of the models.

[9] This is different from evidence for construct validity as psychologists talk about it (cf. Cronbach and Meehl 1955). There, construct validity is a matter of ensuring that a test actually taps into the hypothetical construct it is intended to tap into. In what follows, I offer conceptual and empirical arguments that the PBC construct has a high degree of empirical plausibility and internal coherence. I happen to mention ways in which an individual's PBC could be discerned, but only in principle.

eat the salmon. But they press on, affirming that they know your tastes fairly well and are confident that you'll enjoy *this* salmon *this* time. After some hemming and hawing, and a few drinks, you decide that your friends might be right and order the salmon. Or consider other cases at the extremes: Someone who is deathly allergic to peanuts has a likelihood of 0.0 for ordering vegetables tossed in peanut sauce. The barfly who always gets a martini has a likelihood of 1.0 for ordering his favorite drink. In our salmon case, the chances of you ordering salmon are slim. It may be a snowball's chance in Hades, but a chance nonetheless.

Considerations of such likelihoods motivate posit of the construct *Probability of Belief Change* (PBC). PBC is the likelihood an agent will change a belief. If your dislike of salmon is accompanied by the belief that salmon tastes disgusting, then your PBC for that belief might be 0.000004. It is highly unlikely that you'll budge on your belief that salmon tastes disgusting but, in principle, you could. In a Quinean vein, observe that the PBC for belief in the excluded middle might not be 0.0, but it is far lower than the PBC for most other beliefs.

A different, but equivalent, way to conceptualize one's PBC is thinking of it as how likely one is to abandon one's current belief. For example, suppose that an agent has a PBC of 0.01 with respect to some belief—say that fluoride in water supplies is overall a good thing for public health. The numerical value provides a probability that the agent will change her mind, that is abandon her belief. For our hypothetical agent, odds are 99 to 1 that she would abandon her belief that public water should be fluoridated. This way of conceptualizing PBC highlights a facet of beliefs that is sometimes overlooked in the modeling literature: that agents will hold onto their beliefs solely because the beliefs are theirs. Instead of thinking about beliefs as endorsed as far as their credences go, PBC asks how likely it is that agents would give up a belief they already have. The insight isn't new: William James (1977/1879) describes the 'sentiment of rationality' as a feeling of "ease, peace, rest" that comes with finding an answer. James approvingly quotes Walt Whitman that in such moments we say "I am sufficient as I am." The belief is ours; we are reluctant to part with it. But even beyond James, the insight that people are sometimes reluctant to abandon some of their beliefs resonates with an effect explored in behavioral economics. The endowment effect notes that agents value more highly that with which they are already endowed. In one study, participants we given a mug and asked how much money it would take to part with it. The median value was \$5.25. When another group of participants were asked how much they would be willing to pay to get the mug, median values were between \$2.25 and \$2.75 (cf. Kahneman et al. 1991). The takeaway? Agents put greater value on the mug because it was theirs. Similarly, one might imagine an endowment effect for beliefs: people are less likely to give up a belief because it is theirs.[10]

The PBC construct embraces black-box epistemology. It abstracts away from many of the usual flora and fauna that populate the epistemological ecology. Let us consider just three: reasons for belief, mechanisms for belief formation and change, and Bayesian probabilities. First, it is clear how PBC abstracts away from reasons for

[10] The metaphor here is obviously far from perfect. Mugs have a clear market value while beliefs do not. Be that as it may, one plausible explanation for why certain deeply entrenched beliefs are difficult to change is that there is something like an endowment effect at work. Wittgenstein (1953) seems to point in this direction with his concept of a 'form of life.'

belief. Those reasons can be utterly outlandish or soberly reasonable; PBC does not discriminate. When calculating a PBC, the only question asked is, did the agent change her mind? Either answer provides more data for discerning an agent's PBC. Similar considerations apply when considering mechanisms involved in belief formation and change. PBC only looks at whether their operations delivered a change in belief. Calculating an agent's PBC with respect to a particular belief takes no account of whether the relevant mechanisms are working in accordance with their proper function or not.

Finally, PBC is not Bayesian reasoning, at least not as Bayes's theorem is typically manifested in models. To be sure PBC is easily modified to be a conditional probability: given the odds of such-and-such conditions, the agent is-and-so likely to change her mind. And an individual's PBC might even best be calculated using Bayes's theorem. But there are important differences. First, PBC makes no claim about ideals of reasoning, while Bayesian reasoning is often offered as a normative or prescriptive ideal.[11] PBC is a report of when and how often agents change their minds. PBC offers no normative guardrails. Second, Bayesian reasoning is an account of how agents do or ought to reason—the difference between the descriptive and normative theories do not matter at this moment.[12] PBC makes no claims about how agents actually reason. It keeps track of when agents do in fact change their minds, or how likely they are to change their minds. In a nutshell, Bayes's theorem is often used in agent-based models as a proxy for how agents reason. PBC is silent on processes of reasoning.

To illustrate the difference between Bayesian reasoning and PBC, consider voting records for members of the House of Representatives in the US Congress. The Affordable Care Act (aka 'Obamacare') was signed into law in October 2009, when the moderately liberal Democratic Party controlled two of the three branches of government. Two years later, the conservative Republican Party took control of the House of Representatives, and between 2011 and 2014 the House voted to repeal the ACA in full or in part 54 times (all of them failing).[13] Consider, then, how Paul Ryan—former Speaker of the House, Republican Party loyalist, and noted Ayn Rand enthusiast—voted on the ACA. Whenever an opportunity came by to repeal or defund areas of Obamacare, he voted in favor of it.[14] A Bayesian gloss on Ryan's epistemic activity would suggest that given his evidence, he believed that the ACA ought to be repealed. For estimating Ryan's PBC, we look instead at his past voting record and voting records of people to whom he's ideologically close. Using these data, we see that Ryan's PBC for voting for Obamacare's defunding or repeal is at or very near 1.0. Why? Because every time he and his ideologically similar colleagues had an opportunity to defund or repeal the ACA, they voted to do so.

Notice that Ryan's PBC takes no account of the mechanisms by which he arrives at his decision. Maybe he engages in some statistical reasoning or considerations of what maximizes utility. Or maybe he has considered the evidence *in toto* and believes that the United States truly is better off without an affordable healthcare system. Or it might

---

[11] E.g. Jeffrey (1983, 1992).

[12] See Griffiths et al. (2008) for review; and see Bowers and Davis (2012) for criticism.

[13] O'Keefe, Ed (March 21, 2014). "The House has voted 54 times in four years on Obamacare. Here's the full list." *Washington Post*

[14] See O'Connor and Weatherall (2019) for fascinating discussion of formal models of propagandists and lobbyists affecting group decision-making processes.

even be that he is exercising a range of intellectual virtues and vices to arrive at his belief that the ACA needs to be eliminated. Looking at the matter using PBCs, it does not matter which account is endorsed: an agent's PBC—as well as how it is figured out—is independent of these details. PBC as an epistemological construct is consistent with a range of other theories about belief-formation and justification. A moment's reflection indicates why. PBC is a diachronic construct; characterizations of it range over multiple points in time. Other theories about belief formation and change are typically atemporal or synchronic. Some form of reliabilism might seem to buck this trend, particularly variants of reliabilism that make capacities' evolutionary histories part of the theory. But even then, what matters is how the capacity is working at a time. The evolutionary story tells us how the capacity got to be that way.

PBC abstracts away from details about mechanisms of belief-formation and justification. Again, exactly what evidence an agent entertains in changing her mind doesn't matter; nor does the nature of the reasoning processes involved concern us. The PBC construct views belief-change from 30,000 feet. Even so, one worry is that such a far-away view doesn't provide anything worth looking at. Arguing from the results of the model to the coherence of the concept gets things the wrong way around: evidence of conceptual coherence is needed prior to testing the concept. Otherwise the quick contra argument is that there *is* some causal mechanism at work but the posited concept isn't the one doing the heavy lifting. To avoid such a fate for the PBC construct, I present five arguments all converging on its coherence.

1. PBC fits with how practicing social psychologists describe their own findings. Some examples: being pressed for time makes one less likely to help others in need (Darley and Batson 1973); saying something makes one more likely to believe the thing uttered (Higgins and Rholes 1978); and unexpected improvements in everyday life, like a sunny day after a rainy streak, makes risky behaviors like buying lottery tickets more likely (Otto and Eichstaedt 2018). In these cases, and scores of others, social psychologists talk about their findings in terms of what makes a particular action or belief more or less likely. In other words, social psychologists describe factors affecting belief change in terms of probabilities. This was stated explicitly in the theoretical work of Egon Brunswik (1955a, b) and is carried on by, among others, Gerd Gigerenzer (2007, 2008). But probabilistically describing changes in behaviors and beliefs is the same thing as the probabilities expressed by PBCs. Social psychologists provide empirical content to the formal concept.
2. PBC is an elegant way of capturing the manifold influences on an agent's behavior. Consider just one area: what the social psychological literature has shown about the many influences on judgments concerning members of minority groups. Begin by noting that racial preferences are evident in babies as young as 6 months (Lee et al. 2017). So while we might not be born with racial preferences, they're developed before a baby's first birthday. If you're an employer reading CVs, you're less likely to request interviews for CVs with stereotypically Black names like 'Jamal' or 'Lakisha' (Bertrand and Mullainathan 2003). But, if you've happened to seen and reflect on an image of a counterstereotypical exemplar, implicit biases have a less powerful effect on your judgments involving members of the relevant minority

group (Govan and Williams 2004). But if you happen to hear a racist dogwhistle like 'inner city' or 'welfare queen', then your discriminatory tendencies are amplified (Horwitz and Peffley 2005). But if you coach yourself saying, "when I see a Black name I will think 'good!' " your judgments are less prone to influence by implicit biases (Mendoza et al. 2010). But if you coach yourself saying, "don't be racist!", you can make your judgments more prone to influence by implicit biases (Macrae et al. 1994). What this brief review suggests is that influences on belief and judgment are legion. Building a model that captures these many influences would be equivalent to building a simulation of the sort that Bostrom (2003) imagines our descendants to be running. But we don't have a detailed enough understanding of these manifold influences nor of their interactions with one another. So if we were to imagine ourselves in a situation in which we had to make a snap judgment about a minority, each of the above forces might affect the likelihood of our forming a particular belief.[15] An agent's PBC abstracts away from these many nudging influences on human behavior to pin a value to how frequently agents change their minds.

3. The concept has been used implicitly in cultural psychology to describe belief change among agents in subcultures. Lassiter et al. (2018) describes and analyzes a model in which agents in different subcultures influence one another. The mechanisms by which members of subcultures are influenced by members of other subcultures is identified as their 'resistance to change.' Higher resistance to change means that agents are less likely to adopt traits from members of other cultures; lower resistance to change means that two agents are more likely to adopt traits from one another. Two agents from the same subculture will always adopt each others' traits; but some subcultures will never adopt traits of the other. (In America, think of the punk and yuppie subcultures of the 1980s and 1990s: neither would be caught dead emulating the other.) Both resistance to change and PBC conceptualize belief change in terms of probabilities. In fact, resistance to change is a species of PBC: the former specifies likelihoods of belief-change given subculture membership while the latter is a more generic concept. The coherence of resistance to change strongly suggests the coherence of the more general PBC-construct.
4. Major online corporations conduct doppelgänger searches as a way of boosting revenue (cf. Stephens-Davidowitz 2017). These searches use huge amounts of data about others' preferences to predict, for some target consumer, what her preferences will be. For example, Netflix will suggest television shows based on how your own watching history matches up to others' histories; Amazon does the same thing with products users have purchased; ads pitched to users by Google function the same way by looking at websites that have been visited. Doppelgänger searches aren't even all that new a concept: when a doctor predicts a patient's likelihood of developing heart disease, the doctor is effectively comparing the patient's health profile against many others. An agent's PBC is a probability that functions like the values generated by doppelgänger searches. The doppelgänger search generates a value indicating how likely a consumer C will purchase a product

[15] Situationists in ethics and epistemology have turned this insight into a minor industry within philosophy. A soon-to-be dying industry one hopes.

P: based on C's history and the purchasing history of people like C, C will purchase P with a probability of 0.70. Provided we had similar data about C's beliefs and the belief histories of people like C, it would be possible to generate a probability predicting the likelihood of C coming to believe B: based on C's history and the belief-changing history of people like C, C change his mind about B with a probability of 0.70. This probability is an agent's PBC.

5. Consider LaPlace's Demon: an omniscient entity that can predict all future states of the universe given knowledge of the starting points as well as the laws governing interactions. Consider a cousin of LaPlace's Demon: Hume's Intelligence.[16] The Intelligence has knowledge of all psychological forces affecting how likely an agent is to change her mind and calculates these forces as probabilities. The Intelligence knows, for example, that I'm more likely to change my mind about household matters when speaking with my spouse rather than my co-workers. The Intelligence, with knowledge of these psychological forces, could calculate for some interaction the likelihood that an agent would change her mind. The Intelligence *just would be* calculating the agent's PBC.[17]

These arguments suggest that PBC is a coherent construct corresponding to something real. The construct's utility lies in the fact that it provides a computationally tractable way to account for changes of opinions that do not overestimate agential abilities for rational decision-making. In a way, the PBC-construct continues a trend begun by the reliabilist approaches initially stated by Ramsey (1931) and honed by Goldman (1986). Painting in incredibly broad strokes, insofar as the reliabilist hangs her hat on how truth-conducive a process is, just how the process works or the contents of the reasons endorsed are not of great concern (remember: broad strokes only). PBC goes one step further and abstracts away from details of those processes. Whether an agent changes her mind because of how attractive the interlocutor is or because of the reasons offered is inconsequential to calculating an agent's PBC. What matters is whether the agent's mind changed. This, as I will argue, has important skeptical implications for social epistemology.

3 Model description

For readers interested in exploring the model for themselves, it can be found at github.com/cslassiter/Arational-Belief-Convergence.

The model bears resemblance to social contagion models of belief-spread.[18] As the name would suggest, some property is conceptualized as moving through a population as a germ does (cf. Abrahamson and Rosenkopf 1997; Kleinberg 2007) . In simple models of contagion, the likelihood of contracting the contagion on some particular exposure to it does not depend on other instances of exposure. That's to say that exposure at time t_{n+1} is not affected by exposure at t_n. The contrary holds in complex

[16] This is, of course, something of a misnomer. Hume was a psychological determinist and hoped that we might one day understand the mind along the lines of classical mechanics.

[17] An assumption of this argument is that opinion dynamics are probabilistic rather than deterministic. This is a reasonable assumption given our best psychological science.

[18] Thanks to Cailin O'Connor for this suggestion.

models of contagion: earlier exposures affect the likelihood of contracting the germ in later exposures (cf. Baronchelli 2008). The ABC models opts for the simple model. Why? Recall that the PBC construct is an instance of black-box epistemology: the only things that matter are the inputs and the outputs with the likelihood of changing states being probabilistically described. If an agent's susceptibility to infection depends on features of the agent that change over time, then by hypothesis these features are already included in the PBC construct. Since the complex model has the distinct potential to duplicate what has already been accounted for, the simple model of contagion has a higher degree of fit.

In the model, there are 331 agents.[19] Every agent is assigned a belief and values for their PBC and audience size.[20] Each agent's belief is represented by a number. In the model, relations among natural numbers have no analogue in relations among beliefs. Beliefs associated with the values '2' and '3' are not necessarily similar to one another; one might be a belief about peach cobbler and the other about the price of tea in China. Each agent, for its PBC, is assigned some value between 0.01 and 0.10 (depending on the trial condition). The values-range for PBC was taken from Hodas and Lerman's (2014) study on the spread of information over social media.[21] There, they consider the 'probability of infection' given how widely liked or shared a news story is, with probabilities ranging between 0.0 and 0.10. PBC is a close analogue to their probability of infection. In both cases, the value expresses the likelihood of an agent adopting a particular belief. The value ranges for the audience size were chosen somewhat arbitrarily. A Speaker addresses a proportion of the population between 0.0 and 0.10. (depending on the run of the model). This seems to capture a wide variety of situations in which one person addresses others. I might run into a few of my friends at the store, where audience-size values are in the 0.01 range. Or I might address a larger group of people at a town hall meeting, where audience-size values are in the 0.10 range.

Figure 1 gives the model schematic.[22] The model setup is creating the agents and assigning values to the parameters PBC and audience size. At each time step, an agent is chosen at random from the agent set to be the Speaker. The Speaker then picks out its Audience. The size of the Audience for a given run of the model is fixed at 0.01–0.10 of the total population. Even though the size is fixed, the agents constituting the Audience for a Speaker changes at each time step. Each member of the Audience picks out a real number $0 \leq n \leq 1.0$. If n falls below the PBC specific to each member of the Audience, then the Audience member adopts the Speaker's belief. If n is equal to

[19] More precisely: There were on average 331 agents. One of the parameters of the model was population density. Every agent is generated from a patch in the model. Each patch draws a random real number and if it is below the value for population density, then the patch generates an agent. For all runs of the model, the density parameter was set to 0.75.

[20] The conception of belief native to the ABC model has more in common with unique-belief models (e.g. Zollman 2015), a small modification to the ABC model can have agents trading credences as opposed to beliefs. Rather than assign each agent a natural number representing a belief, each agent can be assigned a real number $0 \leq n \leq 1.0$ representing a credence held towards a particular belief. The model then runs as usual. Credence adjustment in the revised ABC model is a function of each agent's PBC rather than (e.g.) weighted averaging. Thanks to an anonymous reviewer for identifying the difference.

[21] Thanks to Dunja Šešelja for the suggestion to make this clearer.

[22] Thanks to an anonymous reviewer for suggesting this addition.

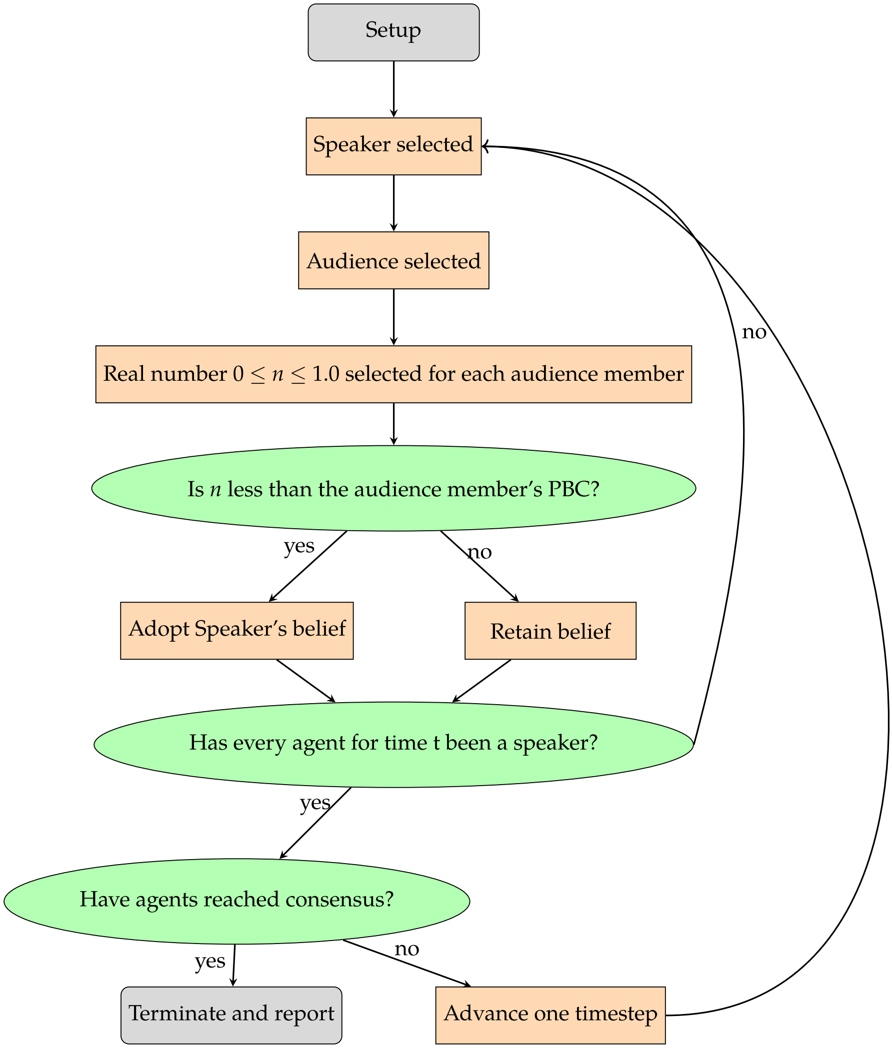


Fig. 1 Model schematic

or greater than the agent's PBC, then no beliefs are changed. This process is repeated for each agent until everyone in the agent set has had a turn as Speaker. This entire process is repeated until consensus is reached. PBC ceilings ranged from 0.01 to 0.10 and Audience size (as a proportion of the total population) ranged from 0.01 to 0.10. The result is a 10×10 experimental condition.

There are two different versions of the model to consider here. In one of them, agents are all the same: each has the same PBC and audience size. Call this the 'Uniform Condition.' In another, everything is exactly the same *except* that the PBC and audience-size values are treated as ceilings, with the values for each agent selected from a uniform random distribution between 0 and the ceiling value for the run of the model. Call this the 'Diversity Condition.' For each condition, there were 50 runs per experimental setup, resulting in N = 5000.

Readers might note an important difference between the above model and others in the literature. First, it is common in the epistemic modeling literature to have a fixed network of agents. Zollman (2007), for example, explores how different networks can speed the spread of truths (and falsehoods) through a network, and this network has a fixed structure. The above model, however, does not resemble such a fixed-structure network. Rather, it's clusters of agents interacting with one another in haphazard ways. In this sense, the above model is more akin to citizens of a small town in rural America—like Winthrop, WA in Okanogan County, with a total population of 394—interacting with one another over long stretches of time than it is to scientists determining what theory best fits the evidence.

Second, this model has no veritistic aspirations. There's no group-independent truth that the agents are after. The model is looking only at the effects of two properties—PBC and audience size—on time to convergence.

Initially the model is run and analyzed with no dogmatists, i.e. agents with a PBC of 0. But we will later consider a few results from model runs with two dogmatists.

4 Results

4.1 Uniform and diversity conditions

For the Uniform Condition, there are both main effects and interaction effects for PBC and audience size on time to convergence, indicated in the upper panel of Fig. 2. The main effects are clear when each of the variables is held to one value. Holding the audience size fixed, time to convergence decreases as PBC increases. Similarly, holding PBC fixed (i.e. looking vertically across lines), time to convergence decreases as audience size increases. Post-hoc pairwise comparison shows significant differences between each of PBC = 0.01–0.03 and audience size = 0.01–0.03 to all other conditions ($p < 0.05$).

An interaction effect is also evident. The audience size for a speaker matters significantly more when PBC values are low than when they are high. But when PBC values are high, the audience size doesn't have a large impact on the time to convergence. A two-way ANOVA reports $F(81, 2900) = 27.36$, $p < 0.001$. There is a large effect size (Cohen 1992), confirmed by $\eta^2 = .214$.

For the Diversity Condition, results were somewhat similar to those of the Uniform Condition. The bottom panel of Fig. 2 shows the effect of audience size and of PBC on time to convergence. Just as in the Uniform Condition, time to convergence decreases significantly as each value increases. Again, there is an important interaction effect. When PBC < 0.04, audience size has a significantly larger impact relative to other conditions. A two-way ANOVA reports $F(81, 4900) = 11.8$, $p < 0.001$. There is a medium effect size confirmed by $\eta^2 = .112$.

Notice that median times to convergence are nearly an order of magnitude higher in the Diversity Condition. The reason is because the Diversity Condition includes

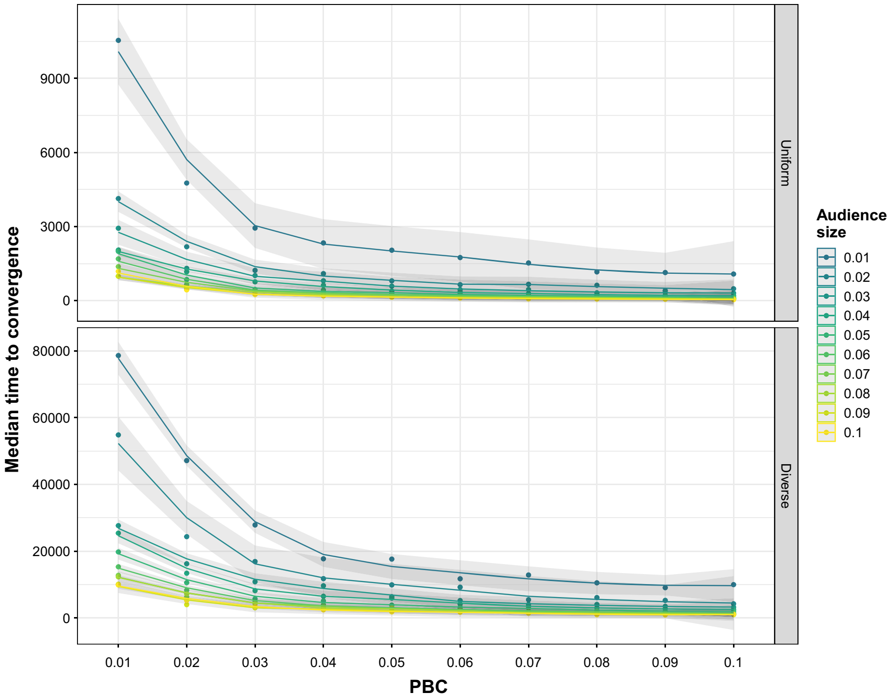


Fig. 2 Time to convergence

agents whose values for PBC and audience size can be closer to 0, thus significantly increasing the time to convergence.[23]

With agents varying by PBC and audience size, it's possible to test an intuitively reasonable prediction: agents with low PBCs and high audience-size values win all (or most of) the time. Define the winner as the agent whose belief is converged upon by the entire population. The conditions for these runs of the model had PBC and audience ceilings of 0.10, 0.33, 0.66, 0.99, creating a 4×4 setup. For each condition, N = 100. The upper panel of Fig. 3 indicates the likelihood of an agent winning relative to the PBC ceiling. The lower panel tells the likelihood of an agent winning relative to the audience size ceiling. While this will be discussed in greater detail in Sect. 5, notice that smaller PBC values increase the likelihood of an agent winning—which is unsurprising–but it's not the case that the highest values for audience size take the cake. That is, while having an exceptionally low PBC greatly increases the chances for getting one's belief to be converged-upon exceptionally high audience size values don't have the same pay-off.

[23] Another difference of note is the number of pairwise significantly different conditions for time to convergence for the Uniform and Diversity conditions: 29% versus 22%. Humility in the power of computational models prohibits from speculating too much about this difference. But if one were to take a guess at an implication of this finding, it might go: (i) the Diversity Condition is much closer to how things stand in the actual world and (ii) it can be difficult to discern differences in PBCs values from cases of convergence in the actual world.

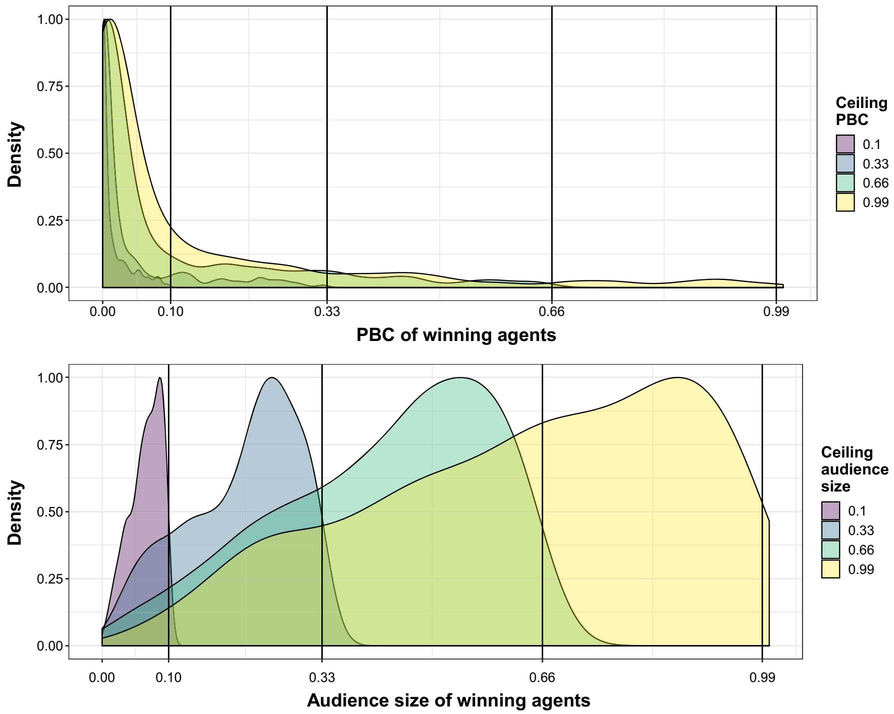


Fig. 3 Distribution of winners

4.2 Dogmatist conditions

In the dogmatist condition, two agents were selected at random to have a PBC of 0. They are the two differently colored strata on the first column of Fig. 4.[24] Everyone else has a PBC $0 < n \leq 0.10$. The strata at each time slice picks out a belief; these are labeled at timesteps 50, 200, and 2500. The size of the stratum reflects how many agents hold that belief. The flows between timeslices track agents. Define three belief-types represented in the model: the beliefs of each of the dogmatists, 'D1' and 'D2', and the beliefs of all the non-dogmatists lumped together, 'ND.' All have the same audience-size value of 0.10. Figure 4 shows the proportion of the population's beliefs at selected time slices. The bands in the plot show agents migrating among three categories of belief. Time was capped at 2500 timesteps. The dogmatist beliefs are in the minority up until timestep 200 and are clearly dominating the market within the next 100 timesteps. The dogmatists' beliefs end up being picked up by the majority of the population by timestep 2500. The one hold-out is an agent with a PBC of 1.29×10^{-5}. But our previous studies show that even this agent will join either D1 or D2 eventually.

[24] The population is limited to one hundred agents; this makes the visualization easier to interpret and doesn't affect any of the underlying philosophical questions.

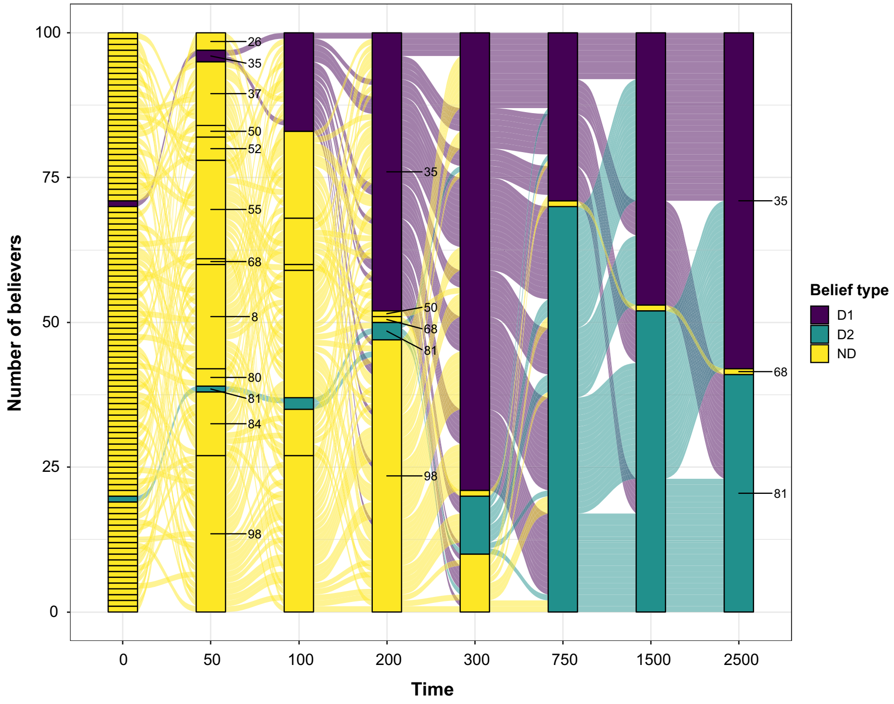


Fig. 4 Population with two dogmatists

4.3 Model limits

The data show that consensus is achieved provided there are no dogmatists and everyone is talking to someone. But what else might prevent convergence on a belief? One very simple mechanism by which convergence is prevented is death. More precisely: if an agent in the model is replaced by another agent with a different belief and parameters fixed randomly (i.e. fixed in the same way as at model initialization), then whether the population converges on a belief depends on the rate at which agents are replaced as well as the values for PBC and audience size. The replacement rate is fixed across all agents and ranges between 0 and 0.01. If an agent A draws a random real number between 0 and 1.0 that is below its replacement rate, then A is replaced by another agent B with different values for belief, PBC, and audience size.

For example, in one run of the model, convergence was reached when the replacement rate was 0.01 and the PBC and audience size ceilings were 0.25. What was funny about this particular run is that the agent with the winning belief had been replaced before the model reached consensus. The agent was gone but its epistemic contribution remained. One is reminded of those artists and scientists whose work is universally acclaimed only after their deaths.

When the PBC and audience size ceilings were set to 0.10, convergence had not been reached by timestep 200,000.[25] The reason is straightforward: when new agents introduce new beliefs to the population, their mere presence reduces how close the population is to consensus. But more importantly, greater diversity in the beliefs available to the population means that agents are less likely to interact with many people all holding the same belief. This means that when an agent eventually changes her mind, there are greater chances that she'll come to endorse a belief that's not in the majority.

5 Discussion

Return to the observation with which this paper began: that unique-belief and credence-adjustment models encode rational criteria for belief change and convergence. Agents are programmed in a way to reflect real-world capacity-rationality. What the ABC model shows is that capacity-rationality is not necessary for convergence. Agents acting arationally are capable of arriving at a consensus. The skeptical challenge, then, is that any case of rational convergence could also plausibly be a case of arational convergence. How? PBC tracks belief-changes, but it is not a normative concept, even if beliefs are themselves normative in some way (cf. Wedgwood 2009). The language used in describing the PBC concept can be done without overt employment of normative language. Because PBC is non-normative and convergence is achieved because agents' have non-zero PBCs, consensus is achieved by non-normative mechanisms. And any instance of consensus that can be modeled by a unique-belief or credence-adjustment model can also be modeled using PBCs. Given how the PBC concept has been defined, there is nothing in individual cases to settle the matter. So then whenever groups come to believe some proposition, they might have done it because they were adjusting their beliefs rationally or arationally. Any case of convergence is possibly an instance of arational belief convergence.

In addition to offering the skeptical challenge, the model identifies several interesting behaviors. In the Diversity Condition, the agents in the best position with respect to PBC values did not always win, though having a lower PBC relative to the group greatly increases one's chances of winning. The data also suggest that the likelihood of winning if one has a higher PBC value can depend on PBC ceilings. An agent with a low PBC when the ceiling is 0.66 is in a better position than an agent with a low ceiling and a PBC ceiling of 0.33. The twin lessons are ones that we try to instill in our students: the value of being open to changing one's beliefs but not *too* open, and the value of being firm in one's beliefs but not *too* firm. The exceptionally long tail on the graphs in Fig. 3 indicate that agents with lower values can still win, but the odds are most definitely not in their favor. The lesson? Something previously noted by Sartre (Sartre 2007/1946): "no signs are vouchsafed in this world." Being very unlikely to change one's mind is a good way to get everyone to eventually agree with you, but it's not guaranteed.

[25] 200,000 timesteps was longer than the longest time to took for models parameterized to PBC ceiling = 0.01 and audience size ceiling = 0.01 to reach consensus.

The results concerning optimal audience-size values are curious. The data suggest that if the audience-size ceiling is C, then an agent's best bet is for her audience-size parameter value be somewhere in the neighborhood of $.75C$ to $.85C$. Why isn't it the case that the highest audience-size value tends to win most often? With a larger value for audience size, agents increase their chances of preaching to the converted. Since further preaching has no effect on agents in this model, it is time wasted by the agent trying to spread the belief: it's telling someone to vote for Candidate C when they've already made up their minds to do so. With a smaller value for audience size, the audience size goes down and so the does time wasted. Obviously, shrinking the value too much will put agents at a severe disadvantage: if I'm talking to three people for your thirty, you can get the message out much faster. But if I'm talking to twenty-five instead of your thirty, the likelihood of attempting to convert believers decreases.

A corollary to these observations is that a belief gains complete adoption by the population once it gains enough traction within the group. If one belief is particularly widespread and if no agents are dogmatists, then repeated exposure to one belief can eventually convince any hold outs. A homey example is when all one's friends and co-workers are talking about a particular movie or television show. If you are the hold-out, then (provided you are not a dogmatist) you will pick up the belief that the show is worth watching from repeated exposure. This is superficially similar to, but distinct from, the 'mere exposure' effect (Zajonc 1968; cf. James 1890). Mere exposure effects rely on repeated exposure to a novel stimulus to increase familiarity with it. The agents in the model are not sophisticated enough to develop anything like a sense of familiarity. Rather, it is because of the ubiquity of the belief in the model that it eventually gets adopted by any hold-outs.

One is reminded of the words from the author of Ecclesiastes: "The race is not to the swift, nor the battle to the strong...nor wealth to the intelligent, nor the favor to the skillful; rather, time and chance happen to all" (9:11). In our own time, these results resonate with happenings like the fame of Schelling (1971) eclipsing that of Sakoda (1971) despite the latter's beating the former to present the same (or superior) results. Or again, Nicholas Maxwell's (1968) anti-physicalist arguments that were echoed by Jackson (1982) and Nagel (1974), but the latter two get the credit and the citations.[26]

Turn now to the dogmatist conditions. At the start of the model, the market is completely saturated: a unique belief is held by each agent. Over time, most non-dogmatists eventually convert to one of D1 or D2. Between timesteps 100 and 300, we find the dogmatist beliefs moving from being marginal to collectively dominating the field. Once converted, they move back and forth among those two camps. There are only a few cases in which we see an agent move from a dogmatist belief to a non-dogmatist one as we do between timesteps 200 and 300. But because the dogmatist beliefs have unyielding defenders, any agents with any room for abandoning their

[26] Maxwell (2003) writes in a footnote, "When I recently drew Thomas Nagel's attention to these publications, he remarked in a letter, with great generosity: 'There is no justice. No, I was unaware of your papers, which made the central point before anyone else.' Frank Jackson acknowledged, however, that he had read my 1968 paper." The paper to which Maxwell alludes has been cited by philosophers other than Maxwell 11 times, according to Google Scholar. Jackson (1982) has been cited 3130 times and Nagel (1974) 8365 times. 'Time and chance' indeed.

beliefs will eventually join a dogmatist camp. Even the one hold-out will succumb eventually.[27]

The dogmatist conditions suggest a way to approach belief-polarization.[28] Polarization into N groups is a product of N dogmatists keeping particular beliefs alive. One sees hints of this in American politics since the early 1990s. When Bill Clinton, a moderate liberal, was voted in as President, Newt Gingrich—noted philanderer and then-member of the House of Representatives from the state of Georgia—began a hyperpartisan campaign that has been a feeder of contemporary American political polarization (Levitsky and Ziblatt 2019). One way to think about this state of affairs is that Gingrich, on some topics, has a PBC of 0. He is unwilling to budge on his beliefs on some matters. On the other side of the political spectrum, Eleanor Norton was ranked in 2018 as the most liberal member of the Democratic Party in the House of Representatives.[29] We might imagine Norton to also have a PBC of 0 on some matters. If we were to imagine Gingrich and Norton arguing about, say, reparations for African-Americans, we could model their interactions as two dogmatists. Gingrich would plausibly have a PBC of 0 for denying reparations and Norton might have a PBC of 0 for enacting reparations. The winning belief is whichever happens to be most widely accepted at a particular snapshot in time.

6 Objection and reply

One concern to have about the model is that there is nothing in it that's particularly about beliefs or anything epistemic. The model might also be used to interpret cases where agents come to consensus about desires or slang or that thick, black glasses frame are cool again.[30] So why interpret the model as being about beliefs as opposed to something else?[31]

In response, this is a feature and not a bug of the model. Issues of interpretation are ubiquitous in formal matters, and the modeling literature is no exception. I look at a scale-free network and see a segment of the Twitter network; you look at it and see a co-citation network. I look at a set of group theoretic relations and see one set of satisfying functions; you look at the same set of relations and see another set of satisfying functions. So multiplicity of interpretations isn't the deep concern. The deep concern is that there is something unique or special about beliefs when considering how beliefs move through large populations. What could this special feature be? Rationality—that agents' grounds for holding or abandoning beliefs are, in some broad sense, rational. This, plausibly, is the driving intuition behind the credence-adjustment and unique-belief models. But it's not at all clear that the spread of beliefs is special in this way. First, there is already a robust literature which does away with the assumption that beliefs are somehow unique or special in how they spread. As mentioned previously,

27 This brings new meaning to the old saw, "science advances one funeral at a time."

28 See Bramson et al. (2017) for an excellent overview of the polarization literature.

29 https://www.govtrack.us/congress/members/eleanor_norton/400295/report-card/2018.

30 To be clear, these frames were never not cool. It was society that got things wrong.

31 Thanks to Carlos Santana for the suggestion.

contagion models of belief spread conceptualize beliefs as behaving like germs. So if there's something special about beliefs, these contagion models would expect it to be shared with germs. Second, the assumption that there is something special about belief has been greeted with skepticism within the history of philosophy as well. Previously mentioned was that curmudgeon Arthur Schopenhauer. In the wake of Kant, German idealists argued about the nature of the noumenal. Schopenhauer (Schopenhauer 2014/1818), bucking broadly rationalist traditions at his time, argues that the noumenal is an arational force called 'Will.' All worldly activities, from human drives to animal instincts to gravitational pulls, are but manifestations of Will. Far from Platonic views about rationality underlying all of nature, Schopenhauer looks around the world to see arational Will at work. Now whether Schopenhauer is right about Will is another matter. At present, it suffices to see that on at least one metaphysical view of the nature of belief, there's nothing special about it. Human beliefs and desires are the manifestations of an indifferent and arational force just like everything else. So from Schopenhauer's vantage point, treating beliefs as spreading like contagion is moving in the right direction: assumptions about the specialness of belief are in need of their own support. The feature that's presumed to be a bug is that belief spread through a population is no more or less rational than the spread of the common cold. So how exactly is that a feature? Just as we have tools for identifying flu outbreaks based on people googling their symptoms, so too might we get a sense of 'belief outbreaks' by people's online search patterns.

7 Conclusion

Let us conclude. We saw that models of belief where convergence happens there is often a commitment to rational adjustment of credences. Lehrer–Wagner, for example, has agents adjust their credences according to the opinions of others in the network as well as how much they value the opinions of others in the network. In both these models, convergence is a product of the means by which credences are adjusted, means that encode rational assumptions about belief change. The ABC model, in contrast, has no such assumptions about the rationality of belief change: belief change is arational. This suggests a skeptical challenge to purportedly rational instances of consensus: any instance of consensus is possibly a case of arational consensus. Even though agents in the model act arationally, there are nonetheless interesting observations to be made about the relative importance of PBC and audience size on both speed of convergence and the likelihood of having one's belief adopted by all. If folks agree that this is broadly right, we can chalk that up to another case of arational consensus.[32]

[32] Model created in Netlogo 6.0.1. Data analysis and graphs produced with *R* 3.5.3 and RStudio. Color palette is from the Viridis package, which renders colors that are both easier for folks with color-vision issues and easier for black and white reproduction. Thanks to Mark Alfino, Nathan Ballantyne, Ted di Maria, Brian Henning, Maria Howard, and Zach Howard for discussion and feedback. Thanks also to Chris Ultican for his willingness to endure thinking through the technical details of the model and the philosophical implications. I'm grateful to the participants of the 2018 Computational Modeling in Philosophy conference at Ludwig Maximilians University for helpful feedback and discussion. Thanks also to the Digital Humanities Initiative at Gonzaga University for financial support. Finally, a special thanks to Michele Lassiter for her unwavering love and support.

References

Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, *8*, 289–309.

Bala, V., & Goyal, S. (1998). Learning from neighbors. *Review of Economic Studies*, *6*, 595–621.

Baronchelli, A. (2008). The emergence of consensus: A primer. *Royal Society Open Science*. https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.172189.

Baumgaertner, B. (2014). Yes, no, maybe so: A veritistic approach to echo chambers using a trichotomous belief model. *Synthese*, *191*, 2549–2569.

Bertrand, M., & Mullainathan, S. (2003). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, No. w9873. National Bureau of Economic Research. http://www.nber.org/papers/w9873.

Bostrom, N. (2003). Are we living in a computer simulation? *The Philosophical Quarterly*, *53*, 243–255.

Bowers, J., & Davis, C. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*, 389–414.

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., et al. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of Science*, *84*, 115–159.

Brunswik, E. (1955a). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Brunswik, E. (1955b). In defense of probabilistic functionalism: A reply. *Psychological Review*, *62*, 236–242.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100–108.

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

Easwaran, K., Fenton-Glynn, L., Hitchcock, C., & Velasco, J. D. (2016). Updating on the Credences of others: Disagreement, agreement, and synergy. *Philosophers' Imprint*. https://quod.lib.umich.edu/p/phimp/3521354.0016.011/--updating-on-the-credences-of-others-disagreement-agreement?view=image.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., et al. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*. http://jasss.soc.surrey.ac.uk/20/4/2.html.

Forrest, P. (1985). The Lehrer/Wagner theory of consensus and the zero weight problem. *Synthese*, *62*, 75–78.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Penguin.

Gigerenzer, G. (2008). *Rationality for mortals*. New York: Oxford University Press.

Goldman, A. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Golub, B., & Jackson, M. (2010). Naï ve learning in socia networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*, 112–149.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*, 357–365.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). New York, NY: Cambridge University Press.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*. http://jasss.soc.surrey.ac.uk/5/3/2.html.

Higgins, E. T., & Rholes, W. S. (1978). Saying is believing: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*, 363–378.

Hodas, N., & Lerman, K. (2014). The simple rules of social contagion. *Scientific Reports*. https://www.nature.com/articles/srep04343.

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, *32*, 127–136.

James, W. (1890). *The principles of psychology* (Vol. 2). New York: Holt.

Jeffrey, R. (1983). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.

Jeffrey, R. (1992). *Probability and the Art of judgment*. New York: Cambridge University Press.

Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: The endowment effect, loss aversion, and the Status Quo Bias. *The Journal of Economic Perspectives*, *5*, 193–206.

Kleinberg, J. (2007). Cascading behavior in networks: Algorithmic and economic issues. In N. Nisan, T. Roughgarden, E. Tardos, & V. Vazirani (Eds.), *Algorithmic game theory* (pp. 612–632). Cambridge: Cambridge University Press.

Lassiter, C., Norasakkunkit, V., Shuman, B., & Toivonen, T. (2018). Diversity and resistance to change: Macro conditions for marginalization in post-industrial societies. *Frontiers in Psychology*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00812/full.

Lehrer, K. (1976). When rational disagreement is impossible. *Noûs*, *10*, 327–332.

Lehrer, K., & Wagner, C. (1981). *Rational consensus in science and society a philosophical and mathematical study*. Dordrecht: Springer.

Lee, K., Quinn, P., & Pascalis, O. (2017). Face race processing and racial bias in early development: A perceptual-social linkage. *Current Directions in Psychological Science*, *26*, 256–262.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808–817.

Martini, C., Sprenger, J., & Colyvan, M. (2013). Resolving disagreement through mutual respect. *Erkenntnis*, *78*, 881–898.

Maxwell, N. (1968). Understanding sensations. *Australasian Journal of Philosophy*, *46*, 127–146.

Maxwell, N. (2003). Three philosophical problems about consciousness and their possible resolution. http://philsci-archive.pitt.edu/2238/.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*, 512–523.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*, 435–450.

Nisbett, R. E., & Wilson, T. D. (1977). The Halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256.

O'Connor, C., & Weatherall, J. (2019). *The misinformation age: How false beliefs spread*. New Haven, CT: Yale University Press.

Otto, A. R., & Eichstaedt, J. C. (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0206923.

Ramsey, F. (1931). In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays*. London: Kegan, Paul, Trench, Trubner & Co.

Raz, J. (1999). *Engaging reason: On the theory of value and action*. New York: Oxford University Press.

Sakoda, J. M. (1971). The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, *1*, 119–132.

Sartre, J.-P. (2007/1946). *Existentialism is a Humanism.* Translated by Carol Macomber. New Haven: Yale University Press.

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

Schopenhauer, A. (2014/1818). *The world as will and representation.* Edited by Christopher Janaway. Translated by Judith Norman and Alistair Welchman. New York: Cambridge University Press.

Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. New York: Penguin.

Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. New York: HarperCollins.

Talbott, W., Bayesian epistemology. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/.

Wedgwood, R. (2009). The normativity of the intentional. In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford handbook of the philosophy of mind* (pp. 421–436). New York: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Translated by Elizabeth Anscombe. Malden, MA: Blackwell Publishing.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1–27.

Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, *74*, 574–587.

Zollman, K. (2015). Modeling the social consequences of testimonial norms. *Philosophical Studies*, *172*, 2371–2383.