

Article

# Agent Mental Models and Bayesian Rules as a Tool to Create Opinion Dynamics Models

André C. R. Martins 

Interdisciplinary Research Group in Complex Systems Modelling (GRIFE), Escola de Artes, Ciências e Humanidades (EACH), Universidade de São Paulo, Rua Arlindo Bético, 1000, São Paulo 03828-000, Brazil; amartins@usp.br

**Abstract:** Traditional models of opinion dynamics provide a simplified approach to understanding human behavior in basic social scenarios. However, when it comes to issues such as polarization and extremism, a more nuanced understanding of human biases and cognitive tendencies are required. This paper proposes an approach to modeling opinion dynamics by integrating mental models and assumptions of individuals agents using Bayesian-inspired methods. By exploring the relationship between human rationality and Bayesian theory, this paper demonstrates the usefulness of these methods in describing how opinions evolve. The analysis here builds upon the basic idea in the Continuous Opinions and Discrete Actions (CODA) model, by applying Bayesian-inspired rules to account for key human behaviors such as confirmation bias, motivated reasoning, and human reluctance to change opinions. Through this, This paper updates rules that are compatible with known human biases. The current work sheds light on the role of human biases in shaping opinion dynamics. I hope that by making the model more realistic this might lead to more accurate predictions of real-world scenarios.

**Keywords:** opinion dynamics; Bayesian methods; cognition; CODA; agent-based models; sociophysics



**Citation:** Martins, A.C.R. Agent Mental Models and Bayesian Rules as a Tool to Create Opinion Dynamics Models. *Physics* **2024**, *6*, 1013–1031. <https://doi.org/10.3390/physics6030062>

Received: 18 March 2024

Revised: 14 June 2024

Accepted: 17 June 2024

Published: 31 July 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction: The Need for General Methods

Opinion dynamics modeling [1–9] is a fascinating area of research that seeks to understand how opinions spread through society. A plethora of models have been developed to describe this process, ranging from simple to complex, and covering topics such as the formation of consensus [10], the emergence of polarization [11–14], the different ways one can define it [15], and the spread of extreme opinions [16–24]. Extremism can be defined as the end of a range over a continuous variable [7,25], or as inflexibles who do not change their minds [26], or using mixed models [8,9,27]. To explore the problem of extremism in the real world, not only opinions matter [28], and one must also consider actions as part of what defines an extremist [29,30].

However, despite the wealth of knowledge already gathered, there are still many aspects of opinion dynamics that require greater attention. Community efforts are necessary to fill gaps in research and promote progress in the field [31]. Currently, most models are only comparable to similar implementations, with a lack of translation between different types. While attempts to propose general frameworks and universal formulas exist [9,32,33], they are, so far, isolated efforts. To achieve greater understanding, one needs to explore how different models relate to each other [34] and develop methods to incorporate new effects and assumptions.

For a deeper understanding of the spread of polarization and extremism, we must also consider actions, not just opinions. Incorporating decision making and behavioral aspects is crucial in modeling opinions [35,36], as it allows for a more accurate depiction of how individuals perceive and react to complex information [37]. One promising avenue of exploration is the use of Bayesian-inspired models [9].

Bayesian rules to model opinions have been introduced both in the opinion dynamics community, as extensions of the Continuous Opinions and Discrete Actions (CODA) model [8,9] and similar opinion models [9,21,27,28,34,38–67], by the use of Bayesian belief networks [68], as well as, independently, in models associated with economical reasoning [69–74]. Despite their popularity, there are two aspects of Bayesian-inspired models for opinion dynamics that have not been properly debated so far. They are how to turn assumptions on how the agents reason into dynamical model equations and the problem of the relationship between Bayesianism and rationality.

It is also worth mentioning that Bayesian update rules can even offer deeper insights into the meaning of extreme opinions, by making it straightforward the need to distinguish between being quite sure about something (high probability one is right) from the inability or strong difficulty to change one's position [28]. More than that, it is possible to demonstrate that if each agent includes the possibility that when the neighbor agree with its choice that might be because the agent has influenced the neighbor, one can obtain the dynamics of other discrete models as a limit case [34]. Extensions of concepts such as contrarians [40] and inflexibles [27] have been studied and Bayesian updates have been used even to obtain bounded confidence-like models [38,66].

In this paper, I explore both aspects of the problem. First, I provide a brief explanation on why the use of Bayesian-inspired rules is both supported by experimental evidence [73,75–78] and not the same as assuming rationality [79,80]. And, to illustrate how Bayesian rules can be used in a general problem, I explain how one can create mental models for the agents. And, one can see how models may include any kind of bias and bounded rationality effects [81,82], and turn those assumptions into update rules. More exactly, I show how to introduce agents who distrust opinions opposed either to a certain choice (a direct bias) or against their current preference. Update rules for both kinds of agents are calculated and the effects of those biases on how extreme the positions of the agents become are studied.

## 2. Bayesian Models and Rationality

Bayesian methods are one of the gold standards for rationality and inductive arguments [80]. If one starts with quite simple rules about how induction about plausibilities must be performed, one can show that plausibilities should be updated using Bayes' theorem [83,84]. The same theorem can be obtained from other axioms, such as maximization of entropy [85] or the considerably weaker basis of "dutch books". However, using those ideas to describe how people reason has been considered problematic [86]. On the other hand, Bayesian ideas can be used both in a complicated way, strictly following its rules, or as a soft version, where its basic ideas are used to represent aspects such as updating subjective opinions [87]. That poses the question of whether Bayesian rules can be used to describe human reasoning well. Certainly, one should also ask about the requirements been imposed from human rationality models. Even the definition of bounded rationality can be challenged, as it assumes there is someone to judge if any behavior is entirely rational [88]. Indeed, using Bayesian methods with perfection is impossible as they require infinite abilities [80]. One can only approximate them by considering a limited set of possibilities, and that is compatible with how human brains work.

There is quite a good evidence that this study reasons in ways that are similar to Bayesian methods [73,75–78]. But, if one wants to use them in mathematical and computational models, one needs to go further than just similarity. Indeed, it is well understood that humans are not ideally rational nor good enough at statistics. One sometimes fails at straightforward problems [89,90] and tends to be too confident about human mental skills [91]. At first, experiments about human cognitive abilities seemed to point to a remarkable amount of incompetence.

But that is not the whole story. When looking closer, some of human mistakes are not as serious as they look. While people are not good enough abstract logicians, when the same problems are presented associated with normal day-to-day circumstances, those problems are answered correctly [92]. And, there is evidence that many of human mistakes

can be described as the use of reasonable heuristics [93], short-cuts that allow us to arrive at answers quite fast and with less effort [94,95]. As simplified rules, heuristics fail under some circumstances. If one goes looking for such cases, one for sure finds the answers. But they are not a sign of complete incompetence.

That does not explain human overconfidence problem, surely. But it has been also observed that human reasoning skills might not have evolved to find the best answers, even if we can use the skills for that purpose. Instead, humans show a tendency to defend their identity-defining beliefs [96,97]. More than that, human ancestors had solid reasons to be quite good at fitting inside their groups and, if possible, ascending socially inside them. Human reasoning and argumentative skills were more valuable from an evolutionary perspective if they worked towards agreeing or convincing human peers. Group belonging mattered more than being right about things that affect direct survival [98,99]. Being sure and defending ideas from people social group become more important than looking for better but unpopular answers.

And there is one more issue. In many laboratory problems where it has been observed that humans make mistakes, scientists used questions that never appear in real life [79]. Take the case of the observation of weighting functions, that one seems to alter probability values that are heard [100]. Using changed values might seem to serve no purpose at first. However, the scientists who performed those experiments assumed the values they presented were known with certainty. But there is no such certainty in real life. If someone tells you there is a 30% chance of rain tomorrow, even if based on highly calibrated models, you know that is an estimate. As with any estimates, at best, the actual value is close to 30%. A Bayesian assessment of the problem combines the previous estimate of rain with the information about the forecast to obtain a final opinion. Applying this to the many experiments that showed people use probability values wrong, one can see that human behavior is compatible with Bayesian rules. The observed changes match reasonable assumptions for everyday situations when one hears uncertain estimates [76]. Applying this is actually wrong in artificial cases, such as the laboratory experiments, where there is no (or quite little) uncertainty about the probability values. Even human tendency to change the opinions far less than one should, called conservatism (no relationship to politics implied in the technical term) [101], can straightforwardly be explained. One just needs to include in a Bayesian model a tendency to skepticism about the data heard [76]. Human brains might just have heuristics that mimic Bayesian estimates for an uncertain and unreliable world.

That is, people are not ideal, but human bounded abilities are not those of incompetents. One makes reasonable approximations. People are motivated reasoners, more interested in defending ideas than looking for better answers. Given the right preferences, these can even be described as rational, despite ethical considerations. Even when it seemed one is making mistakes, one might have been behaving closer to Bayesian rules than it was initially assumed. A more complete discussion about these themes, while of high interest, is beyond the scope of the paper and, for further reading, see Refs. [76,77,79,95,102–104].

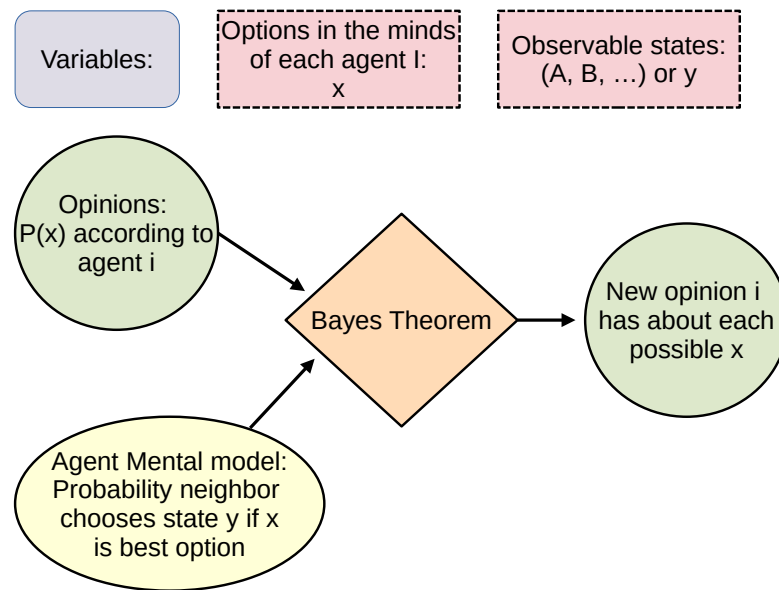
### 3. Update Rules from Agent Mental Models

Humans are not well rational, but can still be described by Bayesian rules. Therefore, it is reasonable trying to apply Bayesian methods as a way to represent human opinions. Thus, the next question one must answer is how one can include human biases and cognitive characteristics in the models. For this, one must consider how the agents think, what the agents show to others, and what the agents expect to see from their neighbors. That is, one needs to describe agent mental models. However, first, let us consider how Bayesian methods work.

#### 3.1. A Brief Introduction to Bayesian Methods

While Bayesian statistics, performed correctly, can become complicated fast enough, it is based on an elementary, almost trivial basis, Bayes' theorem. The theorem works as

follows. There is an issue we want to learn about, and let us represent it by a random variable  $X$ , where each  $x$  represents one possible value. Here,  $x$  can be a quantity, but it can also be nothing more than a label. Let us start with a probability distribution, the initial guess on how likely each possible  $x$  is, represented by a probability distribution  $f(x)$ , called the prior opinion. Once one observes data  $D$ , one must change the opinions on  $x$ . For this, one needs to know, for each possible value  $x$ , how likely it is that one observes  $D$ . That is, one needs the likelihood,  $f(D|x)$ . From that, calculating the posterior estimate  $f(x|D)$  is achieved just by a straightforward multiplication  $f(x|D) \propto f(x)f(D|x)$ . The proportionality constant is calculated by imposing that the final distribution must add (or integrate) to one. Everything in Bayesian methods is a consequence of that update rule and considerations on how to use it. The basic idea, already using an opinion dynamics problem, is represented in Figure 1.



**Figure 1.** Schematics for the general use of Bayes’ theorem as a tool for creating opinion update equations, highlighting the role of agent mental models for what others communicate to them (likelihoods).

To illustrate how it works in practice, let us look at the demonstration of the CODA model rules. In CODA, the agents try to decide between two possible choices,  $A$  or  $B$  (sometimes represented as values of spin,  $+1$  or  $-1$ ). Each agent  $i$  has a time  $t$ , and a probability opinion  $p_i(t)$  that  $A$  is better than  $B$  (and,  $1 - p_i(t)$  that  $B$  is better). But, instead of expressing their probabilistic opinion, they only show their neighbors the option they consider more likely to be better. They also assume their neighbors have a larger than 50% chance,  $\alpha$ , to pick the best option. In general, there can be asymmetric different chances,  $\alpha$  to choose  $A$  when  $A$  is better and  $\beta$  to choose  $B$  when  $B$  is better. As a first example, let us assume the  $\alpha = \beta$  symmetry here. From Bayes’ rule, one obtains the update model for  $p_i(t + 1) \propto p_i(t)\alpha$  and, similarly,  $1 - p_i(t + 1) \propto (1 - p_i(t))(1 - \alpha)$ . As the probabilities must add to one, one divides by their sum and obtains the update rule:

$$p_i(t + 1) = \frac{p_i(t)\alpha}{p_i(t)\alpha + (1 - p_i(t))(1 - \alpha)}. \tag{1}$$

At this step, one has an update rule and can be used it as it is. In this case, however, it is trivial to make a change in variables that provides us a significantly more computationally light rule. The update rule becomes essentially straightforward if let us make the transformation

$$v = \ln\left(\frac{p}{1 - p}\right). \tag{2}$$

The denominators cancel and one obtains

$$v(t + 1) = v(t) \pm C, \tag{3}$$

where  $C = \ln(\frac{\alpha}{1-\alpha})$  and the sign on the sum depends on whether the neighbor prefers  $A (+)$  or  $B (-)$ . One can obtain an even more simplified model by renormalizing  $v$  and making  $C = 1$ .

And that is it. Let us start from the initial opinion, use the Bayes theorem, and is reached an update rule. In this case, the final rule is to add one when a neighbor prefers  $A$  and subtract one when it prefers  $B$ , flipping opinions at  $v = 0$ .

### 3.2. Agent Communication Rules and Their Mental Assumptions

There are two major assumptions in the CODA model. One is how agents communicate. While they have a probabilistic estimate of which option is better, everyone else observes only their best estimate. The second assumption is the mental model of the agents. They think their neighbors are more likely than not to pick the best choice. And, all those neighbors are assumed to have the same chance,  $\alpha > 0.5$ , to be right. Naturally, one can introduce some agents that think their neighbors are more likely to be wrong, that is,  $\alpha < 0.5$ . These is known as contrarians [40], that is, agents who tend to disagreement [105].

Making the model assumption explicit makes it straightforward to investigate what happens if agents behave or act differently. For example, one can have a case when agents look for the best choice between  $A$  and  $B$ . Still, the agents communicate their probability estimates that  $A$  is the better choice,  $p_i(t)$ . In that case, while one can keep the probability  $p_i(t)$  as a measure of the opinion of the agents, those agents no longer state a binary choice but a continuous specific probability value.

#### 3.2.1. Changing What Is Communicated

Mental models become crucial, including which question the agents are trying to answer. Agents may still want to determine which is better,  $A$  or  $B$ , sharing information on their uncertainty. Or they might see the probability values as an ideal mixture. They might accept, for example, that the best position is 60% of  $A$  and 40% of  $B$ . First, let us consider the case where they just want to pick the best choice.

In that case,  $p_i(t)$  is still just a value associated with  $A$ , while, trivially,  $1 - p_i(t)$  gives us the probability  $B$  is better. But one has to evaluate the chance other agents may have any continuous value for  $p_j(t)$  if  $A$  is better (or if  $B$  is). That is, one needs a distribution probability of probabilities to represent the agent not belonging to agents; unified over the paper. mental model. In mathematical terms, one needs a model that says, assuming  $A$  is better, how likely is it that neighbor  $j$  has an opinion  $p_j$  if  $A$  is better,  $f(p_j|A)$ . Certainly, one also needs  $f(p_j|B)$ , but in many situations of interest, that can be obtained by symmetry assumptions. This model was implemented originally by assuming  $f(p_j|A)$  was a beta function  $Be(p_j|\alpha, \beta, A)$ , that is,

$$f(p_j|A) = Be(p_j|\alpha, \beta, A) = \frac{1}{B(\alpha, \beta)} p_j^{\alpha-1} (1 - p_j)^{\beta-1},$$

where  $B(\alpha, \beta)$  is obtained from gamma functions by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Here,  $\alpha$  and  $\beta$  are the traditional parameters of the beta function. Interestingly, the update rule can once more be defined in terms of the log-odds variable  $v = \ln(\frac{p}{1-p})$ , and that leads once more to an additive model. However, the term to be added depends on the probability  $p_j$  communicated by  $j$ , and as the agents become more certain, the size of the



additive term explodes. Consequently, extreme opinions become remarkably stronger than the already extreme opinions in the original CODA model [28,53].

### 3.2.2. Changing the Mental Variables

While in the example in Section 3.2.1 one needs a new likelihood, the beta function that tells us how likely agents think others provide each answer, that need was caused by a change in the communication. But it is not only what is communicated that can be changed. Inner assumptions agents make can also be changed, including the questions they want to answer. That is, one can have quite different assumptions in their mental model.

Assume that instead of having “wisher” agents looking for the best option between  $A$  and  $B$ , each “mixer” agent has an estimate about the best mixture between  $A$  and  $B$ . In this case,  $p_i$  is the percentage of  $A$  in the best blend of two options, and, as such, each value  $0 \leq p_i \leq 1$  must have a probability. One needs probability densities  $f(p)$  as prior and posterior opinion distributions. The more straightforward way to implement such a model is to look for conjugate distributions [106]. Conjugate distributions correspond to those cases where, for a certain likelihood, the prior and the posterior are represented by the same function. In that case, update rules can just update the parameters of that distribution function and not complete distributions. However, this is not the general case for a random application of the Bayes rule. As one builds more detailed models, finding conjugate distributions might not be possible for a given mental model.

Luckily, for the more straightforward cases of interest here, conjugate options exist. Before proceeding to a model, one needs to decide how the agents communicate. Here, once again, one can have a discrete communication, where each agent just tells others which option,  $A$  or  $B$ , should appear in a more significant amount, or communication may include the average estimate for the proportion  $p$  of  $A$ ,  $E[p]$ .

In the first case, with discrete communication, one can find a natural conjugate family by using beta distributions for the opinions and a binomial distribution for the likelihood, that is, the chance that a neighbor chooses  $A$  or  $B$  depending on its average estimate. Surely, other choices of distributions are possible, and they correspond to a similar thought structure but with different dynamics. Under the binomial–beta option, interestingly, the dynamics of the preferred choice, given the more straightforward choice of proper functions for the prior and the likelihood, mimics the original CODA dynamics. However, while the evolution of the preferred option in the mixture is the same, the probability (or proportion) values never reach the same extreme values [53].

### 3.2.3. Other Mental Models Already Explored

Other variations are possible and have been explored. An initial approach to trust was implemented in a fully connected setting by adding one assumption to the agent mental model [47]. Instead of assuming that every other agent had a probability  $\alpha$  to obtain the best answer, each agent assumed there were two types of agents. Agent  $i$  assumed there was a probability  $\tau_{ij}$  that agent  $j$  was a reliable source who picks the best option with chance  $\alpha$ . But other agents may also be untrustworthy (or useless) and pick the best option with probability  $\mu$ , so that  $\mu < \alpha$ , possibly even 0.5 or lower. That is, if  $A$  was the best choice, instead of a chance  $\alpha$  neighbor  $j$  prefers  $A$ , there was a chance given by  $\alpha\tau_{ij} + \mu(1 - \tau_{ij})$ . Applying Bayes’ theorem with this new likelihood led to update rules for  $p_i$  and  $\tau_{ij}$ . Each agent updated both its opinions on whether  $A$  or  $B$  is better. And it also changed its estimates about the trustworthiness of the observed agent. The update rule cannot be simplified by a transformation of variables because no exact way to uncouple the evolution of the opinion and how much agents trusted their neighbors was found.

A similar idea was used for a Bayesian version for continuous communication and the “mixer” type of agents. That model [38] led to an evolution of opinions qualitatively equivalent to what one observes in bounded confidence models [7,25]. That continuous model was later extended to study the problem of several independent issues when agents adjusted their trust based not only on the debated subject but also on their neighbor’s

positions on other matters [66]. Interestingly, that caused opinions to become more clustered and aligned, similar to the irrational consistency one can observe in humans [107].

Even the agent's influence on its neighbors can be used for their mental models. That was introduced as a simplified version by assuming that there were different chances  $a$  and  $c$  that a neighbor prefers  $A$  in the case  $a$  was indeed better, depending on whether the observing agent selected  $A$  or not [34]. That actually weakened the reinforcement effects of agreement, as the other agent may think  $A$  was better not because it was but because the observer also thought that. In the limit of strong influence, the dynamics of the voter model [108,109]—or other types of discrete models, such as majority [110,111] or Sznajd [6] rules, depending on the interaction rules—was recovered. That shows that Bayesian-inspired models are highly more general than the traditional discrete versions.

### 3.3. Introducing Other Behavioral Questions

Bayesian rules can help us explain how humans reason [76,79]. And, one just saw a few examples of introducing extra details in the agent mental models so that new assumptions can be included. In this Section, I discuss how one can move ahead and model some biases observed in human behavior by applying those concepts to create an original model for a specific human tendency.

Let us start by considering a most straightforward one, confirmation bias [112], as that does not even need new mental models. Confirmation bias is just a human tendency to look for information from sources who agree with us. As such, it can be better modeled by introducing rules that reconnect the influence network so that agents are more likely surrounded by those who agree with them. The co-evolution of the CODA model over a network that evolved based on the agreement or disagreement of the agents and their physical location plus thermal noise was studied. Depending on the noise and the strength of the agreement term in the network rewiring, the tendency to polarization and confirmation bias was apparent [63].

Motivated reasoning [96], on the other hand, is not only about who one learns from but how one interprets information depending on whether one agrees with the reasoning or not. This can be implemented in more than one way. Quite a simple version is an approach where trust was introduced in the CODA model [47]. In that model, depending on the initial conditions, as agents became more confident about their estimates, they eventually distrust those who disagreed with them, even when they met for the first time.

#### 3.3.1. Direct Bias

Surely, there are other possibilities, including more heavy-handed approaches. For example, agents might think being untrustworthy is associated with one of the two options. Instead of a trust matrix signaling how much each agent  $i$  trusts each agent  $j$ , one can introduce trust based on the possible choices. For two options,  $A$  and  $B$ , one can assume that each agent has a prior preference. Each agent believes that untrustworthy people defend only the side the agent is biased against. That can be represented by a small addition to the CODA model. Assume agent  $i$  prefers  $A$ , and it thinks that people only go wrong to defend  $B$ . One way to describe that is to assume that there is a proportion  $\lambda$  of reasonable agents who behave like CODA. They pick the better alternative with chance  $\alpha > 0.5$ . However, the remaining  $(1 - \lambda)$  agents choose  $B$  more often, regardless of whether it is true or not, with probability  $\beta > 0.5$ . That is, for agents biased towards  $A$ , the chance a neighbor chooses  $A$ , represented by  $CA$ , if  $A$  (or  $B$ ) is better is given by the equations

$$P(CA|A) = \lambda\alpha + (1 - \lambda)(1 - \beta)$$

and

$$P(CA|B) = \lambda(1 - \alpha) + (1 - \lambda)(1 - \beta).$$

Also,

$$P(CB|A) = \lambda(1 - \alpha) + (1 - \lambda)\beta$$

and

$$P(CB|B) = \lambda\alpha + (1 - \lambda)\beta.$$

Actually, one can introduce an update rule for both the probability  $p_i$  that  $A$  is better and  $\lambda$ . However, for this exercise, let us assume there is an initial fixed value for  $\lambda$ . For example, to illustrate how this bias can change the CODA model, the agents can suppose that a majority of honest people are given by  $\lambda = 0.8$ . Let us also assume that honest people obtain the better answer with a chance of  $\alpha = 0.6$ , while biased people provide their wrong estimate of  $B$  with a probability of  $\beta = 0.9$ . That makes  $P(CA|A) = 0.5$ ,  $P(CA|B) = 0.34$ ,  $P(CB|A) = 0.5$ , and  $P(CB|B) = 0.66$ . So, if agent  $i$  observes someone who prefers  $A$ , it updates its opinion by  $p_i(t + 1) = \frac{p_i(t)^{0.5}}{p_i(t)^{0.5} + (1 - p_i)^{0.34}}$ . That is, for the CODA-transformed variable  $v = \ln(\frac{p}{1-p})$ :  $v_i(t + 1) = v_i(t) + \ln(0.5/0.34) \approx v_i(t) + 0.386$ . On the other hand, if  $B$  is observed, the update rule provides (again for an agent biased towards  $A$ ),  $v_i(t + 1) = v_i(t) + \ln(0.5/0.66) \approx v_i(t) - 0.278$ . That means that steps in favor of  $A$  are larger than those in favor of  $B$  for such an agent. While that agent can be convinced by a majority, the agent moves towards its preference if there is a tie in the neighbors. Depending on the exact values of the parameters, the ratio between step size can become larger.

Let us assume a renormalization of the additive term to implement this bias, as normally performed in CODA applications. Assuming agent  $i$  is biased in favor of  $A$  (the equations for the case where  $i$  is biased in favor of  $B$  are symmetric but are not considered here), one has for the variable  $v_i(t)$ , as defined in Equation (2), when the neighbor also prefers  $A$

$$v_i(t + 1) = v_i(t) + \ln \left[ \frac{\lambda\alpha + (1 - \lambda)(1 - \beta)}{\lambda(1 - \alpha) + (1 - \lambda)(1 - \beta)} \right]. \tag{4}$$

When the neighbor prefers  $B$ , then:

$$v_i(t + 1) = v_i(t) + \ln \left[ \frac{\lambda(1 - \alpha) + (1 - \lambda)\beta}{\lambda\alpha + (1 - \lambda)\beta} \right]. \tag{5}$$

These equations trivially revert to the standard CODA model in Equation (3) when all agents are considered honest, that is, when  $\lambda = 1$ . For ease of further manipulations, one can define the size of the steps in both Equations (4) and (5) as

$$SA = \ln \left[ \frac{\lambda\alpha + (1 - \lambda)(1 - \beta)}{\lambda(1 - \alpha) + (1 - \lambda)(1 - \beta)} \right],$$

$$SD = \ln \left[ \frac{\lambda(1 - \alpha) + (1 - \lambda)\beta}{\lambda\alpha + (1 - \lambda)\beta} \right].$$

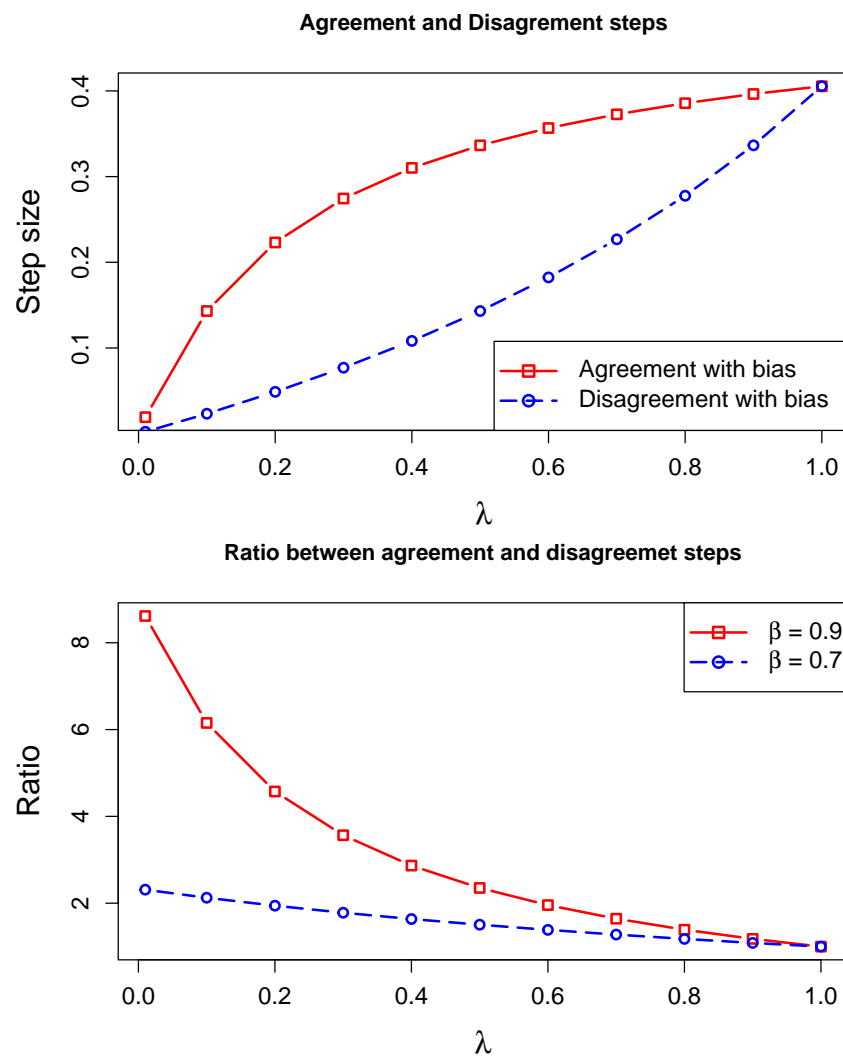
The first question one can ask about those steps is how they relate to each other and the original step size  $C = \ln(\frac{\alpha}{1-\alpha})$ . Straightforward algebraic manipulations show that  $SA < C$  and that  $SD > -C$ , as long as  $\alpha > 0$  in both cases. As  $SD < 0$ , both step sizes are smaller. That was to be expected. Introducing a chance the neighbor might not know what it is talking about should, indeed, decrease the information content of its opinion.

Figure 2 shows how the step sizes change as a function of the estimated proportion  $\lambda$  of honest agents among those each agent is biased against. In Figure 2, upper, one can see both the step sizes when the neighbor agrees with the opinion favored by the bias,  $SA$ , and when the neighbor disagrees with it,  $SD$ . Notice that when  $\lambda$  tends to 1.0, both steps become equal. That corresponds to the scenario where everyone is honest; one recovers the original CODA model with identical steps. The apparent equality when  $\lambda$  tends to zero is not real and is only a product of visualizing minimal steps. Indeed, Figure 2, lower, shows that the ratio between the steps,  $s = SA/SD$ , increases continuously as  $\lambda$  becomes close to zero.



The changes in the step size are indeed not of the same size. If one needs to normalize the steps, as in CODA, one can choose either  $SA$  or  $SD$  as the step made equal to 1.0. Let us initially make, for the implementations, the smaller  $SD = 1.0$ . And, as a simplification, instead of carrying dependencies on  $\lambda$ ,  $\alpha$ , and  $\beta$ , let us just assume there is a ratio  $s$  such that  $SA = sSD$ . That is, it is assumed that disagreement with the bias corresponds to a step size of 1.0 and agreement with the bias to a step size of  $s \geq 1.0$ , where  $s = 1.0$  corresponds to the case when there is no bias. Finally, one has elementary update rules one can implement given by

$$v_i(t + 1) = v_i(t) + \begin{cases} \text{sign}(v_j) & \text{if neighbor } j \text{ disagrees with } i \text{ bias,} \\ s \cdot \text{sign}(v_j) & \text{if neighbor } j \text{ agrees with } i \text{ bias.} \end{cases} \quad (6)$$



**Figure 2. Upper:** size of the steps for agreement,  $SA$ , and disagreement,  $SD$ , as a function of the estimated proportion,  $\lambda$ , of honest agents among those each agent is biased against when it is believed dishonest agents lie with probability  $\beta = 0.9$ . **Lower:** ratio  $s = SA/SD$  for two values of  $\beta$ . See text for details.

### 3.3.2. Conservatism Bias

As a related example, let us introduce the effect called conservatism [101], where people change their opinions less than they should. That can be quickly introduced using a mental model where the agent thinks there is a chance the data are reliable and a chance that the information is only non-informative noise. When that happens, it is only natural

that the update's size is considerably smaller. The larger the chance associated with noisy data, the smaller the update's size.

Here, there is the same mathematical problem as with the direct bias in Section 3.3.1, except that now, instead of believing that defenders of a specific side might be lying, the agents think that defenders of the side who disagree with them might be dishonest. Suppose an agent changes its opinion from  $A$  to  $B$ . In that case, the agent changes its assessment of where there might be dishonesty from  $B$  to  $A$ . That means one has quite similar rules as those in Equation (6). However, the bias is always the same as opinion  $i$ , so the rules depend directly on  $\text{sign}(v_i)$ . That is, then:

$$v_i(t+1) = v_i(t) + \begin{cases} \text{sign}(v_j) & \text{if } \text{sign}(v_j) \neq \text{sign}(v_i), \\ s \cdot \text{sign}(v_j) & \text{if } \text{sign}(v_j) = \text{sign}(v_i). \end{cases} \quad (7)$$

#### 4. Results

To observe how the system might evolve under each rule, I implemented the models using the R software (Version 2.7.0) environment [113]. All cases shown here correspond to an initial neighborhood of agents defined as a square, bi-dimensional lattice with  $40^2$  agents, no periodic conditions, and second-level neighbors. As commented in Section 4.2 below, in some cases, the network is let to evolve into a polarized case before allowing opinions to change using this polarized and quenched network. Once that initial network, lattice or rearranged, is established, agents interact, observing the choice of one neighbor and updating their own opinion based on that observation, according to the update rules of each case. There was an average number of 50 interactions per agent. For all simulations, initial opinions were drawn from a uniform continuous distribution in the range  $-2.0 \leq v_i \leq +2.0$ . All the results for the distribution of opinions correspond to averages over 20 realizations of each case.

##### 4.1. Simulating a Direct Bias

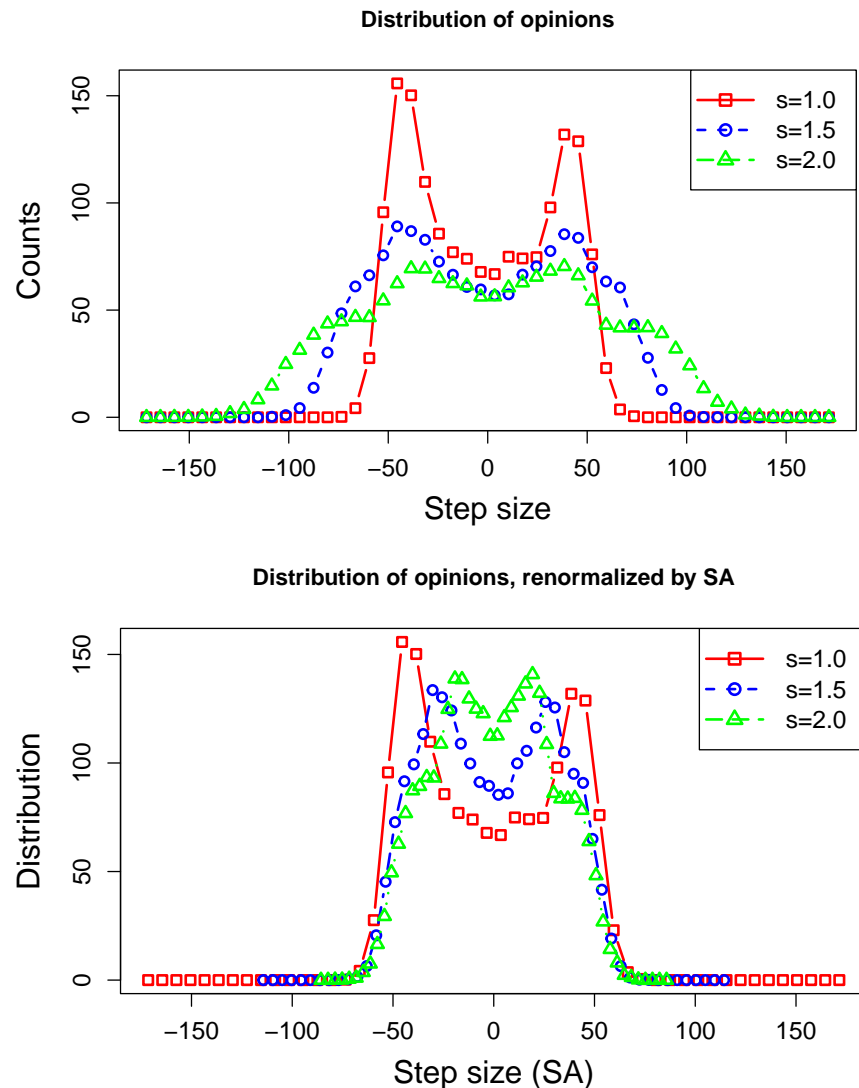
Results for the distribution of opinions in the direct bias case, with no initial rewiring, for three distinct values of the ratio  $s$  can be seen in Figure 3. Each curve corresponds to a different ratio  $s$  between the agreement and the disagreement step. Figure 3, upper, shows how extreme the opinion is measured in the number of disagreement steps (following the algorithm implementation). Figure 3, lower, shows the renormalized distribution if one measures opinions in terms of agreement steps.

One can observe in Figure 3, upper, with the distribution measured in disagreement steps, that as  $s$  increases, the opinions spread further away from the central position. That suggests that the opinions become more extreme. While the peaks associated with the more extreme opinions become softer (and seem to disappear for  $s = 2.0$ ), that happens because the existing extremists become distributed over an extensive range of even stronger opinions.

A different picture emerges if one looks at a renormalized step size, using the agreement step as unit,  $SA = 1$ , instead of the implemented disagreement step. The distribution of opinions for such a case can be seen in Figure 3, lower. What one observes in this case is that when one measures the strength of opinions using  $SA$  as the measuring unit, the distributions become more similar. As  $s$  increases, one observes an increase in the number of agents around the weaker opinions and smaller peaks of extremists.

The apparently contradictory conclusions one can arrive at by looking at only one graphic are another example of the problem with adequately defining what an extremist is [28]. In this model, there are two different ways to define extremism. One of those definitions arises if one was to transform back the number of steps into probability values by inverting the renormalizations and transformations of variables. It is worth noticing that, as can be seen in the upper graphic in Figure 2, both step sizes, for agreement and disagreement, become smaller when bias is introduced when compared to the unbiased case  $s = 1.0$ . That leads to the strange conclusion that when agents think others might be

biased, their final opinion becomes less extreme. That is reasonable as soon as one only cares about the agents' confidence in absolute terms. After all, other agents become less reliable. Their information should convince less.



**Figure 3.** Distribution of opinions after these opinions are (**upper**) measured as disagreement steps and (**lower**) renormalized to agreement steps. See text for details.

But there is another definition of extremism that is also natural and reasonable. That definition comes from asking how straightforward (or complicated) it is to change the choice of a specific agent. This does correspond to the number of steps away from the central opinion. More precisely, since it is necessary to move to the opposite view to change one's choice, using the disagreement step as the unit is the best choice.

That is the case as soon as one was studying results from the conservatism bias, as presented in Section 4.2.1. Here, however, bias was fixed, corresponding to the initial choice of the agent. And that means that there is an essential proportion of agents whose biases do not conform to their opinions. While the proportion of agreement between bias and final opinion increases with  $s$  (observed averages were 57.7% for  $s = 1$ , 66.4% for  $s = 1.5$ , and 74.5% for  $s = 2.0$ ), one a best match between bias and opinion. Even at  $s = 2.0$ , about one in four agents moves towards the opposite choice using agreement steps.

That might seem unrealistic. While there may be a bias toward initial opinions, people usually defend their current position. What size of step humans use when returning to a previously held belief is an interesting question, but that is beyond the scope of this paper.

On the other hand, confirmation bias is described not as an agreement with initial views but as the tendency to look for sources of information that agree with people current beliefs.

That bias was already studied before, with a network that evolved simultaneously with the opinion updates [63]. As agents stopped interacting with the agents they disagreed with, one has a case of traditional confirmation bias. That also means that the results been analyzed in this Section correspond to a direct bias independent of opinions. Real or not, it was introduced here as an exercise and as an example that one can model different modes of thinking using Bayesian tools, regardless of if those tools correspond to reality.

#### 4.2. Rewiring the Network

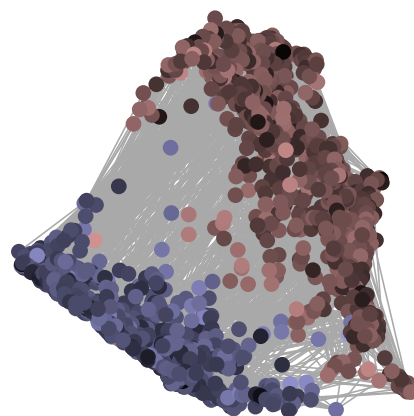
Naturally, one wants to explore more realistic cases. One achieves this in two steps: first, by moving closer to a confirmation bias by introducing rewired networks, and second, by implementing the model of conservatism bias, as defined in Equation (7).

Here, the study follows the rewiring algorithm previously [63] used to study the simultaneous evolution of networks and opinions to generate an initial, quenched network before opinion updates start. At each step, the algorithm tries to destroy an existing link between two agents (1 and 2) and create a new one between two other agents (3 and 4). The decision of whether to accept that change depends on the Euclidean distance between agents 1 and 2,  $d_{12}$ , and agents 3 and 4,  $d_{34}$ , measured by considering a coordinate system over the square lattice where first neighbor distances correspond to 1. That way, there is a tendency to preserve the initial square lattice. And, one also uses a term that makes it more likely to accept the change when the old link was between disagreeing agents and the new one is between agreeing agents. That is, each rewiring is accepted with probability

$$P = \exp(-\Delta H) = \exp[-\beta(d_{34} - d_{12} - J(\sigma_3\sigma_4 - \sigma_1\sigma_2))], \quad (8)$$

where  $J$  is the relative importance between physical proximity and opinion agreement and  $\sigma_s$  denote the choice of each agent. Notice that if one needs a reasonable chance that distant agents connect,  $J$  should be comparable with the initial side of the square network.

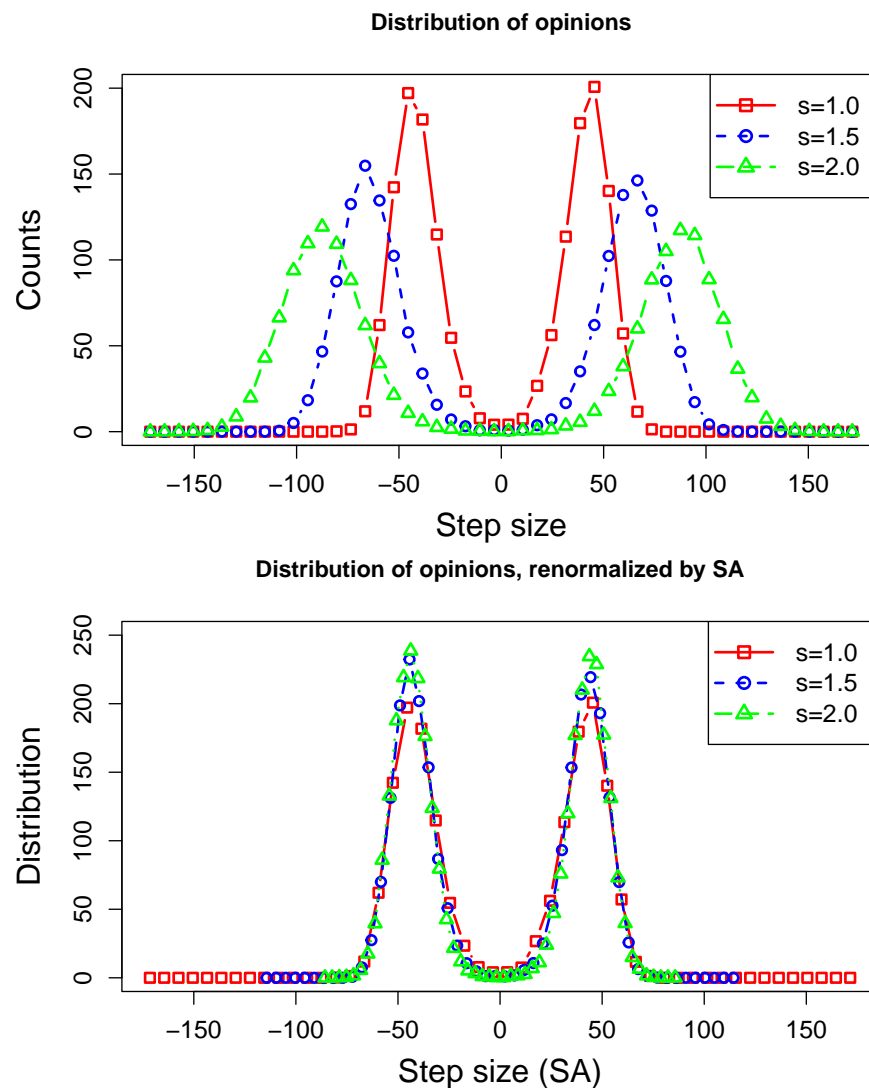
Figure 4 shows a typical case, as an example, of how an initially square lattice is altered after an average of twenty rewirings per agent:  $\beta = 1.0$  and  $J = 20$ . In Figure 4, brown and blue colors correspond to the two choices, and darker hues show a stronger final opinion obtained after the opinion update phase. As the network did not change during the opinion update phase, its final shape after implementing the rule defined in Equation (8) is preserved.



**Figure 4.** Typical network formed after applying the rewiring algorithm with an average of twenty rewirings per agent:  $\beta = 1.0$  and  $J = 20$ . Brown and blue colors correspond to the two choices, and darker hues show a stronger final opinion obtained after the opinion update phase.

Figure 5 shows the distribution of opinions when one uses the direct bias rules after obtaining a quenched network using the parameters to generate networks similar to

those in Figure 4. One can see significant differences between the results with no initial rewiring (Figure 3) and the new ones (Figure 5). With the initial rewiring, one finds almost no moderates in the final opinions, and the extremist peaks, when one uses the disagreement step,  $SD$ , for normalization, move to even stronger values as  $s$  increases. That dislocation also corresponds to smaller peaks distributed over a more extensive range. That, however, is an artifact of using  $SD$ . When renormalized to  $SA$  size steps, One can see that the three curves for different values of  $s$  match almost ideally. That happens because, with the implemented rewiring, most interactions involve agents who already share the same opinions.



**Figure 5.** Distribution of opinions after these opinions are (upper) measured as disagreement steps and (lower) renormalized to agreement steps.

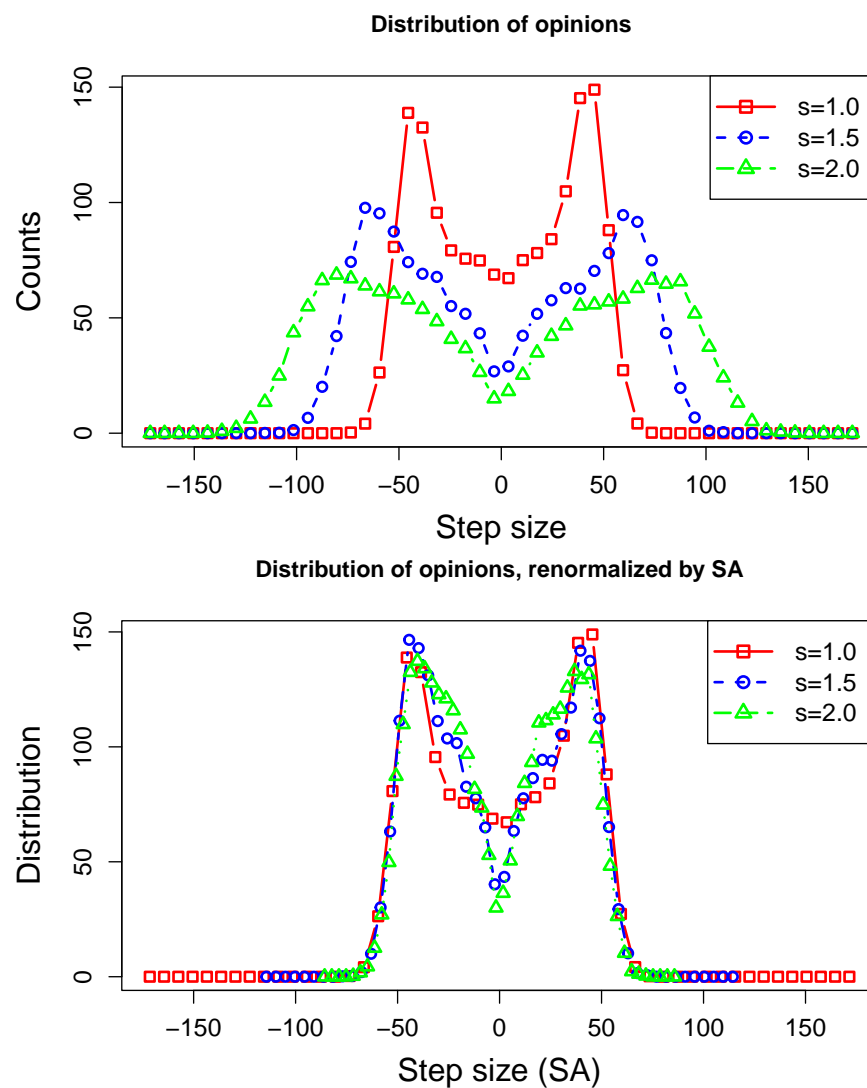
However, the problem of defining extremism under these circumstances becomes even more straightforward. One no longer has a case when the agent bias does not agree with its opinion. Indeed, the observed averages for the proportion of agreement between bias and opinions were 99.8% for  $s = 1$ , 100% for  $s = 1.5$ , and 100% for  $s = 2.0$ . That means that to change position, all agents move one disagreement step at a time. While observing how the curves match when one measures opinions in agreement steps is interesting, the disagreement step case is more informative. And here, as expected, the more bias one introduces, the more complicated it becomes for all agents to change their opinions.



#### 4.2.1. Simulating the Conservatism Bias

The conservatism model defined in Equation (7) was also implemented using the same parameter values used in the previous cases in Section 4.1. As in this scenario there is no distinction between opinion and bias, the simulations presented here do not include an initial rewiring phase to guarantee that most agents are aligned with their own biases.

Figure 6 shows the distribution of opinions as disagreement steps (Figure 6, upper) and agreement steps (Figure 6, lower). As have been discussed in Section 4.2, in this case, the measure that tells us how complicated it is for an agent to change its choice is associated with the disagreement steps. One can see in Figure 6, upper, that, as was observed with the direct bias case with no initial rewiring (Figure 2), the peaks of extreme opinions move to more distant values and become less pronounced. That happens because, contrary to the rewired case when there were few links between disagreeing agents, there is still more extensive borders where one can find agents whose neighbors have distinct choices. Despite that, moderate agents, close to zero, become rarer as the conservatism bias increases. While not as relevant for understanding how extreme opinions are in this model, Figure 6, lower, renormalized for agreement steps of size one, is still of interest. It shows the same tendency of preserving the general shape for different values of  $s$  observed before (in Figure 5), except that one can see that when no conservatism is expected ( $s = 1.0$ ), there are still more agents in the moderate region.



**Figure 6.** Distribution of opinions after these opinions are (upper) measured as disagreement steps and (lower) renormalized to agreement steps.

## 5. Conclusions

Approximating the complete but impossible Bayesian rules provides a reasonable description of human behavior, especially when one accounts for individuals' imperfect trust in others and their tendency to reason in a motivated manner. By modeling these approximations, one can create realistic models of human behavior, as demonstrated in this paper. This paper analysis focuses on introducing variations to the CODA model where agents exhibit biases in favor of a particular opinion. Through the current investigation, it is found that conservatism implementation, where the agents distrust information that goes against their beliefs, results in more extreme opinions and a greater resistance to change.

Furthermore, this research examines how to obtain update rules from assumptions about the agents' mental models, using both previously published cases and the new examples. Another crucial aspect of using Bayesian-inspired rules is that it allows for a better understanding of the relationship between distinct models. By exploring which assumptions lead to the current opinion models, one gets insight into their inter-relatedness and identify cases when each model might be more applicable. Overall, this study sheds light on the potential of Bayesian-inspired modeling to offer a more nuanced description of agent behavior and its impact on opinion dynamics.

**Funding:** This work was supported by the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) under grant 2019/26987-2.

**Data Availability Statement:** The code is available at <https://www.comses.net/codebase-release/d4ab2a25-4233-4e6e-a8c5-a3b919cfd6e2/> (accessed on 13 June 2024).

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **2009**, *81*, 591–646. [\[CrossRef\]](#)
2. Galam, S. *Sociophysics: A Physicist's Modeling of Psycho-Political Phenomena*; Springer Science+Business Media, LLC: New York, NY, USA, 2012. [\[CrossRef\]](#)
3. Latan, B. The psychology of social impact. *Am. Psychol.* **1981**, *36*, 343–365. [\[CrossRef\]](#)
4. Galam, S.; Gefen, Y.; Shapir, Y. Sociophysics: A new approach of sociological collective behavior: Mean-behavior description of a strike. *J. Math. Sociol.* **1982**, *9*, 1–13. [\[CrossRef\]](#)
5. Galam, S.; Moscovici, S. Towards a theory of collective phenomena: Consensus and attitude changes in groups. *Eur. J. Soc. Psychol.* **1991**, *21*, 49–74. [\[CrossRef\]](#)
6. Sznajd-Weron, K.; Sznajd, J. Opinion evolution in a closed community. *Int. J. Mod. Phys. C* **2000**, *11*, 1157. [\[CrossRef\]](#)
7. Deffuant, G.; Neau, D.; Amblard, F.; Weisbuch, G. Mixing beliefs among interacting agents. *Adv. Compl. Sys.* **2000**, *3*, 87–98. [\[CrossRef\]](#)
8. Martins, A.C.R. Continuous opinions and discrete actions in opinion dynamics problems. *Int. J. Mod. Phys. C* **2008**, *19*, 617–624. [\[CrossRef\]](#)
9. Martins, A.C.R. Bayesian updating as basis for opinion dynamics models. *AIP Conf. Proc.* **2012**, *1490*, 212–221.
10. Schawe, H.; Fontaine, S.; Hernández, L. When network bridges foster consensus. Bounded confidence models in networked societies. *Phys. Rev. Res.* **2021**, *3*, 023208. [\[CrossRef\]](#)
11. DiMaggio, P.; Evans, J.; Bryson, B. Have American's social attitudes become more polarized? *Am. J. Sociol.* **1996**, *102*, 690–755. [\[CrossRef\]](#)
12. Baldassarri, D.; Gelman, A. Partisans without constraint: Political polarization and trends in american public opinion. *Am. J. Sociol.* **2008**, *114*, 408–446. [\[CrossRef\]](#)
13. Taber, C.S.; Cann, D.; Kucsova, S. The motivated processing of political arguments. *Polit. Behav.* **2009**, *31*, 137–155. [\[CrossRef\]](#)
14. Dreyer, P.; Bauer, J. Does voter polarisation induce party extremism? the moderating role of abstention. *West Eur. Politics* **2019**, *42*, 824–847. [\[CrossRef\]](#)
15. Bramson, A.; Grim, P.; Singer, D.J.; Berger, W.J.; Sack, G.; Fisher, S.; Flocken, C.; Holman, B. Understanding polarization: Meanings, measures, and model evaluation. *Philos. Sci.* **2017**, *84*, 115–159. [\[CrossRef\]](#)
16. Deffuant, G.; Amblard, F.; Weisbuch, G.; Faure, T. How can extremism prevail? A study based on the relative agreement interaction model. *J. Artif. Soc. Soc. Simul. (JASSS)* **2002**, *5*, 1. Available online: <https://www.jasss.org/5/4/1.html> (accessed on 15 June 2024).
17. Amblard, F.; Deffuant, G. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Phys. A Stat. Mech. Appl.* **2004**, *343*, 725–738. [\[CrossRef\]](#)

18. Galam, S. Heterogeneous beliefs, segregation, and extremism in the making of public opinions. *Phys. Rev. E* **2005**, *71*, 046123. [[CrossRef](#)] [[PubMed](#)]
19. Weisbuch, G.; Deffuant, G.; Amblard, F. Persuasion dynamics. *Phys. A Stat. Mech. Appl.* **2005**, *353*, 555–575. [[CrossRef](#)]
20. Franks, D.W.; Noble, J.; Kaufmann, P.; Stagl, S. Extremism propagation in social networks with hubs. *Adapt. Behav.* **2008**, *16*, 264–274. [[CrossRef](#)]
21. Martins, A.C.R. Mobility and social network effects on extremist opinions. *Phys. Rev. E* **2008**, *78*, 036104. [[CrossRef](#)]
22. Li, L.; Scaglione, A.; Swami, A.; Zhao, Q. Consensus, polarization and clustering of opinions in social networks. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 1072–1083. [[CrossRef](#)]
23. Parsegov, S.E.; Proskurnikov, A.V.; Tempo, R.; Friedkin, N.E. Novel multidimensional models of opinion dynamics in social networks. *IEEE Trans. Autom. Control* **2017**, *62*, 2270–2285. [[CrossRef](#)]
24. Amelkin, V.; Bullo, F.; Singh, A.K. Polar opinion dynamics in social networks. *IEEE Trans. Autom. Control* **2017**, *62*, 5650–5665. [[CrossRef](#)]
25. Hegselmann, R.; Krause, U. Opinion dynamics and bounded confidence models, analysis and simulation. *J. Artif. Soc. Soc. Simul. (JASSS)* **2002**, *5*, 2. Available online: <https://www.jasss.org/5/3/2.html> (accessed on 15 June 2024).
26. Galam, S.; Jacobs, F. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Phys. A Stat. Mech. Appl.* **2007**, *381*, 366–376. [[CrossRef](#)]
27. Martins, A.C.R.; Galam, S. The building up of individual inflexibility in opinion dynamics. *Phys. Rev. E* **2013**, *87*, 042807. [[CrossRef](#)] [[PubMed](#)]
28. Martins, A.C.R. Extremism definitions in opinion dynamics models. *Phys. A Stat. Mech. Appl.* **2022**, *589*, 126623. [[CrossRef](#)]
29. Tileaga, C. Representing the ‘other’: A discursive analysis of prejudice and moral exclusion in talk about romanies. *J. Community Appl. Soc. Psychol.* **2006**, *16*, 19–41. [[CrossRef](#)]
30. Bafumi, J.; Herron, M.C. Leapfrog representation and extremism: A study of american voters and their members in congress. *Am. Polit. Sci. Rev.* **2010**, *104*, 519–542. [[CrossRef](#)]
31. Sobkowicz, P. Whither now, opinion modelers? *Front. Phys.* **2020**, *8*, 461. [[CrossRef](#)]
32. Böttcher, L.; Nagler, J.; Herrmann, H.J. Critical behaviors in contagion dynamics. *Phys. Rev. Lett.* **2017**, *118*, 088301. [[CrossRef](#)] [[PubMed](#)]
33. Galam, S.; Cheon, T. Tipping points in opinion dynamics: A universal formula in five dimensions. *Front. Phys.* **2020**, *8*, 446. [[CrossRef](#)]
34. Martins, A.C.R. Discrete opinion models as a limit case of the coda model. *Phys. A Stat. Mech. Appl.* **2014**, *395*, 352–357. [[CrossRef](#)]
35. Kowalska-Pyzalska, A.; Maciejowska, K.; Suszczyński, K.; Sznajd-Weron, K.; Weron, R. Turning green: Agent-based modeling of the adoption of dynamic electricity tariffs. *Energy Policy* **2014**, *72*, 164–174. [[CrossRef](#)]
36. Müller-Hansen, F.; Schlüter, M.; Mäs, M.; Donges, J.F.; Kolb, J.J.; Thonicke, K.; Heitzig, J. Towards representing human behavior and decision making in earth system models—An overview of techniques and approaches. *Earth Syst. Dyn.* **2017**, *8*, 977–1007. [[CrossRef](#)]
37. Haghtalab, N.; Jackson, M.O.; Procaccia, A.D. Belief polarization in a complex world: A learning theory perspective. *Proc. Natl. Acad. Sci. USA (PNAS)* **2021**, *118*, e2010144118. [[CrossRef](#)]
38. Martins, A.C.R. Bayesian updating rules in continuous opinion dynamics models. *J. Stat. Mech. Theo. Exp.* **2009**, *2009*, P02017. [[CrossRef](#)]
39. Martins, A.C.R.; de Pereira, C.; Vicente, R. An opinion dynamics model for the diffusion of innovations. *Phys. A Stat. Mech. Appl.* **2009**, *388*, 3225–3232. [[CrossRef](#)]
40. Martins, A.C.R.; Kuba, C.D. The importance of disagreeing: Contrarians and extremism in the coda model. *Adv. Compl. Sys.* **2010**, *13*, 621–634. [[CrossRef](#)]
41. Vicente, R.; Martins, A.C.R.; Caticha, N. Opinion dynamics of learning agents: Does seeking consensus lead to disagreement? *J. Stat. Mech. Theo. Exp.* **2009**, *2009*, P03015. [[CrossRef](#)]
42. Si, X.-M.; Liu, Y.; Xiong, F.; Zhang, Y.-C.; Ding, F.; Cheng, H. Effects of selective attention on continuous opinions and discrete decisions. *Phys. A Stat. Mech. Appl.* **2010**, *389*, 3711–3719. [[CrossRef](#)]
43. Si, X.-M.; Yun, Cheng, H.; Zhang, Y.-C. An opinion dynamics model for online mass incident. In *ICASTE 2010. 2010 3rd International Conference on Advanced Computer Theory and Engineering*; Proceedings Volume 5; Desheng, W., Ruofeng, W., Yi, X., Eds.; The Institute of Electrical and Electronics Engineers, Inc.: New York, NY, USA, 2010; pp. V5-96–V5-99. [[CrossRef](#)]
44. Martins, A.C.R. A middle option for choices in the continuous opinions and discrete actions model. *Adv. Appl. Stat. Sci.* **2010**, *2*, 333–346.
45. Martins, A.C.R. Modeling scientific agents for a better science. *Adv. Compl. Sys.* **2010**, *13*, 519–533. [[CrossRef](#)]
46. Deng, L.; Liu, Y.; Xiong, F. An opinion diffusion model with clustered early adopters. *Phys. A Stat. Mech. Appl.* **2013**, *392*, 3546–3554. [[CrossRef](#)]
47. Martins, A.C.R. Trust in the coda model: Opinion dynamics and the reliability of other agents. *Phys. Lett. A* **2013**, *377*, 2333–2339. [[CrossRef](#)]

48. Diao, S.-M.; Liu, Y.; Zeng, Q.-A.; Luo, G.-X.; Xiong, F. A novel opinion dynamics model based on expanded observation ranges and individuals' social influences in social networks. *Phys. A Stat. Mech. Appl.* **2014**, *415*, 220–228. [[CrossRef](#)]
49. Luo, G.-X.; Liu, Y.; Zeng, Q.-A.; Diao, S.-M.; Xiong, F. A dynamic evolution model of human opinion as affected by advertising. *Phys. A Stat. Mech. Appl.* **2014**, *414*, 254–262. [[CrossRef](#)]
50. Caticha, N.; Cesar, J.; Vicente, R. For whom will the bayesian agents vote? *Front. Phys.* **2015**, *3*, 25. [[CrossRef](#)]
51. Martins, A.C.R. Opinion particles: Classical physics and opinion dynamics. *Phys. Lett. A* **2015**, *379*, 89–94. [[CrossRef](#)]
52. Lu, X.; Mo, H.; Deng, Y. An evidential opinion dynamics model based on heterogeneous social influential power. *Chaos Solitons Fractals* **2015**, *73*, 98–107. [[CrossRef](#)]
53. Martins, A.C.R. Thou shalt not take sides: Cognition, logic and the need for changing how we believe. *Front. Phys.* **2016**, *4*, 7. [[CrossRef](#)]
54. Chowdhury, N.R.; Morărescu, I.-C.; Martin, S.; Srikant, S. Continuous opinions and discrete actions in social networks: A multi-agent system approach. In *2016 IEEE 55th Conference on Decision and Control (CDC)*; The Institute of Electrical and Electronics Engineers, Inc.: New York, NY, USA, 2016; pp. 1739–1744. [[CrossRef](#)]
55. Cheng, Z.; Xiong, Y.; Xu, Y. An opinion diffusion model with decision-making groups: The influence of the opinion's acceptability. *Phys. A Stat. Mech. Appl.* **2016**, *461*, 429–438. [[CrossRef](#)]
56. Huang, C.; Hu, B.; Jiang, G.; Yang, R. Modeling of agent-based complex network under cyber-violence. *Phys. A Stat. Mech. Appl.* **2016**, *458*, 399–411. [[CrossRef](#)]
57. Garcia, L.M.T.; Roux, A.V.D.; Martins, A.C.R.; Yang, Y.; Florindo, A.A. Development of a dynamic framework to explain population patterns of leisure-time physical activity through agent-based modeling. *Int. J. Behav. Nutr. Phys. Act.* **2017**, *14*, 111. [[CrossRef](#)]
58. Sun, R.; Mendez, D. An application of the continuous opinions and discrete actions (coda) model to adolescent smoking initiation. *PLoS ONE* **2017**, *12*, e0186163. [[CrossRef](#)] [[PubMed](#)]
59. Sobkowicz, P. Opinion dynamics model based on cognitive biases of complex agents. *J. Artif. Soc. Soc. Simul. (JASSS)* **2018**, *21*, 8. [[CrossRef](#)]
60. Lee, H.K.; Kim, Y.W. Public opinion by a poll process: Model study and bayesian view. *J. Stat. Mech. Theo. Exp.* **2018**, *2018*, 053402. [[CrossRef](#)]
61. Garcia, L.M.T.; Roux, A.V.D.; Martins, A.C.R.; Yang, Y.; Florindo, A.A. Exploring the emergence and evolution of population patterns of leisure-time physical activity through agent-based modelling. *Int. J. Behav. Nutr. Phys. Act.* **2018**, *15*, 112. [[CrossRef](#)] [[PubMed](#)]
62. Tang, T.; Chorus, C.G. Learning opinions by observing actions: Simulation of opinion dynamics using an action-opinion inference model. *J. Artif. Soc. Soc. Simul. (JASSS)* **2019**, *22*, 2. [[CrossRef](#)]
63. Martins, A.C.R. Network generation and evolution based on spatial and opinion dynamics components. *Int. J. Mod. Phys. C* **2019**, *30*, 1950077. [[CrossRef](#)]
64. Martins, A.C.R. Discrete opinion dynamics with  $m$  choices. *Eur. Phys. J. B* **2020**, *93*, 1. [[CrossRef](#)]
65. León-Medina, F.J.; Tena-Sánchez, J.; Miguel, F.J. Fakers becoming believers: How opinion dynamics are shaped by preference falsification, impression management and coherence heuristics. *Qual. Quant.* **2020**, *54*, 385–412. [[CrossRef](#)]
66. Maciel, M.V.; Martins, A.C.R. Ideologically motivated biases in a multiple issues opinion model. *Phys. A Stat. Mech. Appl.* **2020**, *553*, 124293. [[CrossRef](#)]
67. Fang, A.; Yuan, K.; Geng, J.; Wei, X. Opinion dynamics with bayesian learning. *Complexity* **2020**, *2020*, 8261392. [[CrossRef](#)]
68. Sun, Z.; Müller, D. A framework for modeling payments for ecosystem services with agent-based models, bayesian belief networks and opinion dynamics models. *Environ. Model. Softw.* **2013**, *45*, 15–28. [[CrossRef](#)]
69. Orléan, A. Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *J. Econ. Behav. Organ.* **1995**, *28*, 257–274. [[CrossRef](#)]
70. Rabin, M.; Schrag, J.L. First impressions matter: A model of confirmatory bias. *Quart. J. Econ.* **1999**, *114*, 37–82. [[CrossRef](#)]
71. Andreoni, J.; Mylovanov, T. Diverging opinions. *Am. Econ. J. Microecon.* **2012**, *4*, 209–232. [[CrossRef](#)]
72. Nishi, R.; Masuda, N. Collective opinion formation model under bayesian updating and confirmation bias. *Phys. Rev. E* **2013**, *87*, 062123. [[CrossRef](#)]
73. Eguíluz, V.M.; Masuda, N.; Fernández-Gracia, J. Bayesian decision making in human collectives with binary choices. *PLoS ONE* **2015**, *10*, e0121332. [[CrossRef](#)]
74. Wang, Y.; Gan, L.; Djurić, P.M. Opinion dynamics in multi-agent systems with binary decision exchanges. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; Proceedings; The Institute of Electrical and Electronics Engineers, Inc.: New York, NY, USA, 2016; pp. 4588–4592. [[CrossRef](#)]
75. Knill, D.C.; Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* **2004**, *27*, 712–719. [[CrossRef](#)]
76. Martins, A.C.R. Probabilistic biases as Bayesian inference. *Judgm. Decis. Mak.* **2006**, *1*, 108–117. [[CrossRef](#)]



77. Tenenbaum, J.B.; Kemp, C.; Shafto, P. Theory-based Bayesian models of inductive reasoning. In *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*; Feeney, A., Heit, E., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 167–204. [[CrossRef](#)]
78. Tenenbaum, J.B.; Kemp, C.; Griffiths, T.L.; Goodman, N.D. How to grow a mind: Statistics, structure, and abstraction. *Science* **2011**, *331*, 1279–1285. [[CrossRef](#)]
79. Martins, A.C.R. *Arguments, Cognition, and Science: Need and Consequences of Probabilistic Induction in Science*; Rowman & Littlefield Publishers: Lanham, MD, USA, 2020.
80. Martins, A.C.R. Embracing undecidability: Cognitive needs and theory evaluation. *arXiv* **2020**, arXiv:2006.02020. [[CrossRef](#)]
81. Simon, H.A. Rational choice and the structure of environments. *Psychol. Rev.* **1956**, *63*, 129–138. [[CrossRef](#)] [[PubMed](#)]
82. Selten, R. What is bounded rationality? In *Bounded Rationality: The Adaptive Toolbox*; Gigerenzer, G., Selte, R., Eds.; The MIT Press: Cambridge, MA, USA, 2001; pp. 147–171. [[CrossRef](#)]
83. Cox, R.T. *The Algebra of Probable Inference*; The John Hopkins Press: Baltimore, MD, USA, 1961. Available online: <https://bayes.wustl.edu/Manual/cox-algebra.pdf> (accessed on 15 June 2024).
84. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: New York, NY, USA, 2003. [[CrossRef](#)]
85. Caticha, A.; Giffin, A. Updating probabilities. *AIP Conf. Proc.* **2006**, *872*, 31–42. [[CrossRef](#)]
86. Eberhardt, F.; Danks, D. Confirmation in the cognitive sciences: The problematic case of bayesian models. *Minds Mach.* **2011**, *21*, 389–410. [[CrossRef](#)]
87. Elqayam, S.; Evans, J.S.B.T. Rationality in the new paradigm: Strict versus soft bayesian approaches. *Think. Reason.* **2013**, *19*, 453–470. [[CrossRef](#)]
88. Chater, N.; Felin, T.; Funder, D.C.; Gigerenzer, G.; Koenderink, J.J.; Krueger, J.I.; Noble, D.; Nordli, S.A.; Oaksford, M.; Schwartz, B.; Stanovich, K.E.; Todd, P.M. Mind, rationality, and cognition: An interdisciplinary debate. *Psychon. Bull. Rev.* **2018**, *25*, 793–826. [[CrossRef](#)]
89. Watson, P.C.; Johnson-Laird, P. *Psychology of Reasoning: Structure and Content*; Harvard University Press: Cambridge, MA, USA, 1972. Available online: <https://archive.org/details/psychologyofreas0000waso> (accessed on 15 June 2024).
90. Tversky, A.; Kahneman, D. Extension versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychol. Rev.* **1983**, *90*, 293–315. [[CrossRef](#)]
91. Oskamp, S. Overconfidence in case-study judgments. *J. Consult. Psychol.* **1965**, *29*, 261–265. [[CrossRef](#)] [[PubMed](#)]
92. Johnson-Laird, P.N.; Legrenzi, P.; Legrenzi, M.S. Reasoning and a sense of reality. *Brit. J. Psychol.* **1972**, *6*, 395–400. [[CrossRef](#)]
93. Gigerenzer, G.; Todd, P.M.; the ABC Research Group. *Simple Heuristics That Make Us Smart*; Oxford University Press, Inc.: New York, NY, USA, 1999.
94. Tversky, A.; Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cogn. Psychol.* **1973**, *5*, 207–232. [[CrossRef](#)]
95. Gigerenzer, G.; Goldstein, D.G. Reasoning the fast and frugal way: Models of bounded rationality. *Psychol. Rev.* **1996**, *103*, 650–669. [[CrossRef](#)] [[PubMed](#)]
96. Kahan, D.M. Ideology, motivated reasoning, and cognitive reflection. *Judgm. Decis. Mak.* **2013**, *8*, 407–424. [[CrossRef](#)]
97. Kahan, D.M. The expressive rationality of inaccurate perceptions. *Behav. Brain Sci.* **2017**, *40*, e6. [[CrossRef](#)] [[PubMed](#)]
98. Mercier, H.; Sperber, D. Why do humans reason? arguments for an argumentative theory. *Behav. Brain Sci.* **2011**, *34*, 57–111. [[CrossRef](#)] [[PubMed](#)]
99. Mercier, H.; Sperber, D. *The Enigma of Reason*; Harvard University Press: Cambridge, MA, USA, 2017. [[CrossRef](#)]
100. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **1979**, *47*, 263–291. [[CrossRef](#)]
101. Edwards, W. Conservatism in human information processing. In *Formal Representation of Human Judgment*; Kleinmuntz, B., Ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1968; pp. 359–369. Available online: <https://pages.ucsd.edu/~mckenzie/Edwards1968excerpts.pdf> (accessed on 15 June 2024).
102. Plous, S. *The Psychology of Judgment and Decision Making*; McGraw-Hill: New York, NY, USA, 1993.
103. Baron, J. *Thinking and Deciding*; Cambridge University Press: New York, NY, USA, 2023. [[CrossRef](#)]
104. Fitelson, B.; Thomason, N. Bayesians sometimes cannot ignore even very implausible theories (even ones that have not yet been thought of). *Australas. J. Log.* **2008**, *6*, 25–36. [[CrossRef](#)]
105. Galam, S. Contrarian deterministic effect: The hung elections scenario. *Phys. A Stat. Mech. Appl.* **2004**, *333*, 453–460. [[CrossRef](#)]
106. O’Hagan, A. *Kendall’s Advanced Theory of Statistics. Volume 2B: Bayesian Inference*; Edward Arnold: London, UK, 1994.
107. Jervis, R. *Perception and Misperception in International Politics*; Princeton University Press: Princeton, NJ, USA, 1976. [[CrossRef](#)]
108. Clifford, P.; Sudbury, A. A model for spatial conflict. *Biometrika* **1973**, *60*, 581–588. [[CrossRef](#)]
109. Holley, R.; Liggett, T.M. Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* **1975**, *3*, 643–663. [[CrossRef](#)]
110. Galam, S. Modelling rumors: The no plane pentagon french hoax case. *Phys. A Stat. Mech. Appl.* **2003**, *320*, 571–580. [[CrossRef](#)]
111. Galam, S. Opinion dynamics, minority spreading and heterogeneous beliefs. In *Econophysics and Sociophysics: Trends and Perspectives*; Chakrabarti, B.K., Chakraborti, A., Chatterjee, A., Eds.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006; Chapter 13. [[CrossRef](#)]



- 
112. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **1998**, *2*, 175–220. [[CrossRef](#)]
  113. The R Development Core Team. *R: A Language and Environment for Statistical Computing. Reference Index. Version 2.7.0*; R Foundation for Statistical Computing: Vienna, Austria, 2008. Available online: <https://ringo.ams.stonybrook.edu/images/2/2b/Refman.pdf> (accessed on 15 June 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.