Egocentric Bias and Doubt in Cognitive Agents

Nanda Kishore Sreenivas Shrisha Rao

Abstract

Modeling social interactions based on individual behavior has always been an area of interest, but prior literature generally presumes rational behavior. Thus, such models may miss out on capturing the effects of biases humans are susceptible to. This work presents a method to model egocentric bias, the real-life tendency to emphasize one's own opinion heavily when presented with multiple opinions. We use a symmetric distribution centered at an agent's own opinion, as opposed to the Bounded Confidence (BC) model used in prior work. We consider a game of iterated interactions where an agent cooperates based on its opinion about an opponent. Our model also includes the concept of domain-based self-doubt, which varies as the interaction succeeds or not. An increase in doubt makes an agent reduce its egocentricity in subsequent interactions, thus enabling the agent to learn reactively. The agent system is modeled with factions not having a single leader, to overcome some of the issues associated with leader-follower factions. We find that agents belonging to factions perform better than individual agents. We observe that an intermediate level of egocentricity helps the agent perform at its best, which concurs with conventional wisdom that neither overconfidence nor low self-esteem brings benefits.

Keywords: egocentric bias; cognitive psychology; doubt; factions; Continuous Prisoner's Dilemma; opinion aggregation

1 Introduction

Decision making has been a long-studied topic in the domain of social agent-based systems, but most earlier models were rudimentary and assumed rational behavior [56, 19]. Decision making is strongly driven by the opinion that an agent holds; this opinion is shaped over time by its own initial perceptions [44], its view of the world [18], information it receives over various channels, and its own memory [20]. This process of opinion formation is fairly complex even under the assumption that Bayesian reasoning applies to decision making. Years of research in psychology has shown that humans and even animals [28, 10] are susceptible to a wide plethora of cognitive biases [4]. Despite the immense difficulties in the understanding and description of opinion dynamics, it continues to be an area of immense interest because of the profound impact of individual and societal decisions in our everyday lives.

In this work, we model agents with *egocentric bias* and focus on agents' opinion formation based on its perception, memory, and opinions from other agents. Egocentric bias may be described as the tendency to rely too heavily on one's own perspective. The bias has been claimed to be ubiquitous [45] and ineradicable [37]. Egocentric bias is commonly thought of as an umbrella term, and covers various cognitive biases, including the anchoring bias [65, 48, 22] and the "false consensus effect" [54]. Recent research seems to suggest that such bias is a consequence of limited cognitive ability [43] and the neural network structure of the brain [34].

The initial approaches to model opinion dynamics were heavily inspired by statistical physics and the concept of atomic spin states. They were thus detached from real life and allowed only two levels of opinions [12, 63]. There was also the social impact model [47, 46] and its further variations with a strong leader [32].

Later models, such as the ones proposed by Krause [36] and Hegselmann [30] considered continuous values for opinions and introduced the *Bounded Confidence* (BC) model which incorporated the *confirmation bias*. The BC model and the relative agreement model by Deffuant *et al.* [13] inspired by the former, have remained in favor until now [1, 41]. Confirmation bias has also been modeled in other contexts, such as trust [49] and conversational agents [7, 29].

Historically, models concentrating on opinion dynamics have revolved around consensus formation. However, opinions are not formed at an individual or societal level without any consequence. Rather, these opinions lead to decisions and these decisions have a cost, a result, and an outcome. In reality, humans as well as animals learn from outcomes and there are subtle changes introduced in this process of opinion formation during subsequent interactions.

The assignment of weights to all opinions including one's own is a major issue in opinion formation. In the BC model, the weights are taken as a uniform distribution within the interval, and opinions outside of this are rejected. The problem with this model is that it is too rigid. To introduce some level of flexibility into our model, we consider the assignment of weights as per a symmetric distribution centered around the agent's perspective with its flatness/spread varied according to that agent's level of egocentricity.

We consider a game of iterated interactions where an agent, say A, is paired with some random agent, say B, in one such iteration. Each of these interactions is a Continuous Prisoner's Dilemma (CPD) [66], which allows an agent to cooperate at various levels bounded by $[0, 1]$. Here, *opinions* are based upon an agent's knowledge about the opponent's level of cooperation in prior interactions and thus lies between 0 and 1. Thus, A has its own opinion of B and it also takes opinions of B from other sources. A aggregates all the opinions and cooperates at that level, and it then decides the outcome of this interaction based on B's level of cooperation.

Our model also captures an agent's reaction to this outcome. When an agent succeeds, there is a rise in self-esteem and this is reflected in a higher egocentricity in subsequent interactions. We model reaction to failure as a loss of self-esteem i.e., a rise in *self-doubt* on this domain [57]. This domain-based self-doubt is a key aspect of this model as it helps an agent to learn reactively.

In our model, agents can belong to *factions* as well. While most works have modeled factions as a leader-followers structure [59, 2], we model a faction with a central memory, that holds the faction's view on all agents in the system. The faction's view is an unbiased aggregate of individual opinions of its members. To sum up, an agent can have up to three different levels of information—its own opinion, opinions from friends, and the faction's view.

Through simulation, we find results about optimum level of egocentricity and the effect of faction sizes. Varying the levels of egocentricity among agents, it is observed that agents with an intermediate level perform much better than agents with either low or high levels of egocentricity. This is in strong agreement with conventional wisdom that neither overconfidence nor low self-confidence brings optimum results. Agents in larger factions are observed to perform better, and results indicate a linear proportionality between value and faction size as suggested by Sarnoff's Law. Also, to understand the effects of other attributes of the system, we vary the number of interactions, the proportion of different agents, the types of agents, etc.

2 Related Work

In his work on opinion dynamics [61], Sobkowicz writes:

> "Despite the undoubted advances, the sociophysical models of the individual behaviour are still rather crude. Most of the sociophysical agents and descriptions of their individual behaviour

> are too simplistic, too much 'spin-like', and thus unable to capture the intricacies of our behaviours."

Our work thus focuses on three key aspects—egocentricity, self-doubt, and the concept of factions. In this section, we review the existing work in these domains.

Egocentric Bias

Egocentric bias is the tendency to rely too heavily on one's own perspective and/or to have a higher opinion of oneself than others. Ralph Barton Perry [50] coined the term *egocentric predicament* and described it as the problem of not being able to view reality outside of our own perceptions. Greenwald [26] described it as a phenomenon in which people skew their beliefs in agreement with their perceptions or what they recall from their memory. We are susceptible to this bias because information is better encoded when an agent produces information actively by being a participant in the interaction.

Research suggests that this skewed view of reality is a virtually universal trait and that it affects each person's life far more significantly than had been realized [45]. It has also been shown to be pervasive among people and groups in various contexts such as relationships, team sports, etc. [55]. It is closely connected to important traits such as self-esteem and confidence [38]. A high degree of egocentric bias hinders the ability to empathize with others' perspectives, and it has been shown that egocentricity tends to be lower in depressed individuals [24]. Egocentric bias also plays a key factor in a person's perception of fairness: people tend to believe that situations that favor them are fair whereas a similar favor to others is unjust [21, 25]. Perceived fairness is a crucial element in several resource allocation problems. Most importantly, it has been shown to be *ineradicable* even after standard debiasing strategies such as feedback and education [37].

Prior work has been done to model confirmation bias, but the most used model has been the Bounded Confidence (BC) model. The BC model was first introduced by Krause in 2000 [36]. Later, Deffuant *et al.* [13] proposed a relative agreement model (RA) which extended the BC model. In the BC model, an agent considers only those opinions that are sufficiently close to its own, and shuns any opinion outside the confidence threshold. This model has been used to model confirmation bias in many papers [67, 14, 30, 61, 15].

Self-doubt

There can be multiple responses to a perceived failure—lowering of one's aspiration, loss of self-esteem manifested as an increase in doubt, or even leaving the activity altogether [40].

The term *self-esteem* has been used in three ways—global self-esteem, state self-esteem and domain specific self-esteem [9]. We are primarily concerned with an agent's domain-specific self-esteem in this paper, which is a measure of one's perceived confidence pertaining to a single domain. Our work models the self-doubt which is a counterpart of this. Self-doubt is defined as "the state of doubting oneself, a subjective sense of instability in opinions" [6].

Factions

Factions have been broadly considered to be specific sets of agents. However, a faction has been modeled in different ways. Some factions have been modeled as a leader-follower group, where the leader determines the group dynamics [59]. Even if the group does not have an assigned leader to start with, it has been suggested that an agent with high cognitive capacity eventually emerges as a leader [2]. Such a leader eventually impacts the performance of the entire group. Factions can also be modeled as a selfish herd, where each agent is a member for its own gain [27]. However, this structure does not have a single leader and such models have proved useful in modeling certain group behaviors [52, 5].

3 Egocentric Interactions

We consider a system of agents playing a game of iterated interactions. In each iteration, an agent A is paired with some agent B randomly. Since the model is based on the Continuous Prisoner's Dilemma (CPD) [66], an agent can cooperate at any level between 0 and 1, with 0 corresponding to defection and 1 to complete cooperation. $C_B(t)$ denotes B's level of cooperation with A in interaction t, and this value lies between 0 and 1.

The opinion of A about B at the next interaction $t+1$, denoted by $\eta_A(B, t+1)$ is based on A's previous ω previous experiences with B, where ω is the memory size:

$$\eta_A(B, t) = \frac{C_B(t-1) + C_B(t-2) + \ldots + C_B(t-\omega)}{\omega} \tag{1}$$

A has its own opinion of B and also collects opinions of B from its friends (described in Section 3.3) as well. It aggregates these opinions according to its egocentricity, and this aggregate is used as its level of cooperation in the next interaction with B. If B cooperates at a satisfactory level, A decreases its doubt on B and thus is more egocentric in the next interaction with B. These concepts and corresponding formulations are outlined in the following subsections.

3.1 Egocentricity

As discussed in the previous section, current models of egocentricity consider a Bounded Confidence (BC) model and all opinions within this interval get the same weight [36, 13, 61]. This uniform distribution of weights across the confidence interval is not an accurate depiction because such a model would assign the same weight to one's own opinion and an opinion on the fringes of the interval. Also, an opinion that is outside the interval by a mere fraction is to be completely rejected, which is too rigid. This raises the need for some flexibility, and hence we use a Gaussian (normal) distribution to calculate the weights. The use of a symmetric distribution to model the agent's judgments with egocentric bias is a manifestation of the anchoring-and-adjustment heuristic, which has a neurological basis [64] and is well known in studies of the anchoring bias [22, 65]. The same type of distribution is seen in the context of anchoring bias in the work of Lieder *et al.* [43]. The mean of the curve is the agent A's opinion of the other agent B, as the mean gets the highest weight in this distribution.

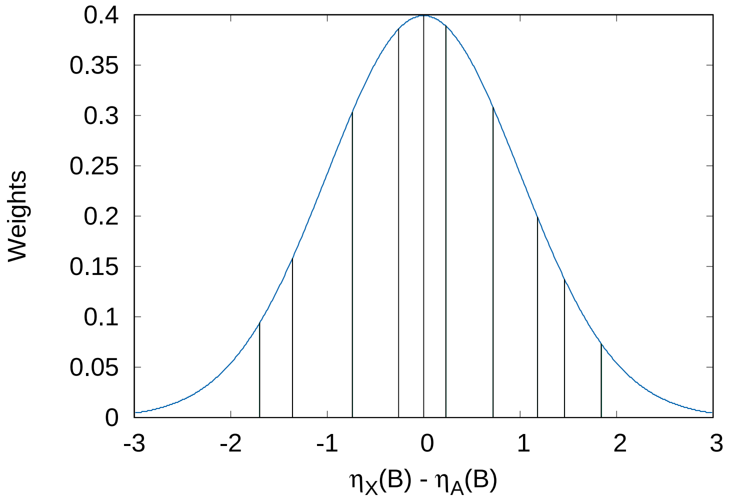


Figure 1: Weights assigned for a range of opinions

Each agent has a base egocentricity E_0, which is a trait of the agent and remains constant. The egocentricity of an agent A with respect to B is obtained by factoring the doubt of A about B on A's base egocentricity. This egocentricity is manifested as the spread of the Gaussian curve, σ. The base spread σ_0 is inversely proportional to the base egocentricity:

$$\sigma_0 = \frac{1}{E_0} \tag{2}$$

The higher the egocentricity of an agent, the lower is the spread of the curve, thus assigning a relatively high weight to its own opinion. The lower the egocentricity of the agent, the flatter is the curve, thus assigning roughly the same weight to all opinions including its own. However, this spread depends on agent B and is an adjusted value of the base spread, σ_0.

3.2 Self-Doubt

Literature in psychology suggests several strategies as responses to failure—quitting, readjusting levels of aspiration, and increasing self-doubt [40]. In this work, we model an increase in domain-specific self-doubt as the response to failure, since self-doubt has been claimed to be useful [68]. In response to a successful interaction, an agent gains in self-confidence, i.e., the self-doubt is decreased. To classify an interaction as success or failure, A has a threshold of *satisfaction* λ; B's level of cooperation has to be higher than λ for A to deem the interaction successful.

Self-doubt is used by agent A as a multiplicative factor on the base spread to obtain the relevant spread for aggregating opinions about agent B. Since doubt is a multiplicative factor, doubt about all agents is initialized as 1, and agent A uses σ_0 as the spread initially. With subsequent interactions, doubt varies as described, and this ensures a constantly changing level of egocentricity, based on outcomes of previous interactions. Since the spread has to be positive at all times, doubt is lower-bounded by 0. Theoretically there is no fixed upper bound for doubt, but beyond a certain value, the curve gets flat enough to a point where all opinions effectively get the same weight. As doubt tends to zero, the agent is completely certain of its opinion and rejects all other opinions. Doubt is subjective, and the doubt of A about B is denoted by $\mathcal{D}_A(B)$. It is updated after each interaction between A and B, being incremented or decremented by a constant c depending on whether A is satisfied or dissatisfied with the interaction.

The spread of the Gaussian curve used for assigning weights to opinions about B is calculated considering the agent-specific doubt and its base spread, σ_0:

$$\sigma = \sigma_0 \mathcal{D}_A(B) \tag{3}$$

This σ defines the spread of the normal distribution used by A to assign weights to different opinions on B.

3.3 Social Structures

Though leader-follower models are seen in certain contexts, such as in the context of modeling countries [58], oligopolies [39], and insurgents [60], there is evidence that other models with no explicit leader-follower structure are appropriate to understand many societal behaviors. The "selfish herd" model [27] was originally suggested for animals seeking to avoid predation and other dangers, but it is seen to explain human social behavior as well [51, 52]. Such models explain economic behavior [53] as well as the evolution of fashions and cultural changes in society [5]. Social media as well as online purchases are also best explained in this way [16]. Group decision making in humans does not follow a strict leader-follower structure even in the presence of bias [62], and the same is also true when it comes to solving complex problems by teams [31]. Gill [23] gives the example of Wikipedia as a well-known example of "collaborative intelligence" at work.

Therefore, we model our agent system as being split into factions, without any single authoritarian faction leader who sets the faction's view. Rather, each faction is modeled to be a herd where all members contribute towards the formation of a *central memory*, which holds an unbiased aggregate of member opinions about each agent in the system.

The contribution of all members is assumed to be authentic and complete. However, the level of adherence of the faction's view is different for each agent. Some agents can be modeled as extreme loyalists, who suspend their own judgment and simply adhere to the faction view, while there are others who are individualists and do not always conform. We introduce the notion of faction alignment κ, which is a measure of an agent's adherence with its faction, with 0 indicating total nonconformance and 1 complete adherence.

The *friends* of an agent A are a small subset of the agents in A's faction. The number of friends may be different for each agent and friendships are defined at random when factions are initialized, but remain intact thereafter. A *friendship* is the two-way connection between two friends. Based on the seminal work of Dunbar [17] on the number of friends in primate societies, a recent paper suggests the concept of Dunbar layers [11]—an individual's network is layered according to strength of emotional ties, with there being four layers in all and the two outermost layers having 30 and 129 members, which suggests that the average number of friends is about 25% of the overall social circle. As the number of friends for an individual is variable, as is the total number of friendships in the faction, we use ${}^{z^2}\!/_8$ as an upper bound for the number of friendships within the faction, where z is the faction size. Friends are the only source of opinions for an agent. Agents fully cooperate when they interact with a friend.

3.4 Game Setting

The standard Prisoner's Dilemma (PD) is discrete, so each agent can choose one of only two possible actions: cooperate or defect. However, not all interactions can be modeled perfectly by such extreme behavior.

In the Continuous Prisoner's dilemma (CPD) [66, 33], a player can choose any level of cooperation between 0 and 1. We borrow the concept and the related payoff equations from Verhoeff's work on the Trader's Dilemma [66]. Here, a cooperation level of 0 and 1 correspond to the cases of complete defection and complete cooperation respectively in the PD.

Consider two agents A and B in a CPD, with their cooperation levels being a and b respectively. The payoff functions are obtained [66] from the discrete payoff matrix by linear interpolation:

$$p_A(a,b) = abC + a\bar{b}S + \bar{a}bT + \bar{a}\bar{b}D \tag{4}$$

where C, T, D, S are the payoffs in the standard PD as shown below

		Player B	
		1	0
Player A	1	(C,C)	(S,T)
	0	(T,S)	(D,D)

The conditions for choosing the values of these variables are $2C > T + S$ and that $T > C > D > S$. Most work on PD, including Axelrod's seminal work on evolution of cooperation [3], uses this set of values: $\langle C = 3, T = 5, D = 1, S = 0\rangle$, and we do the same.

3.5 Opinion Aggregation

There are three phases in each interaction between two agents A and B:

1. *Phase 1*: A adjusts its own opinion $\eta_A(B)$ and all opinions it has received from its friends $\{\eta_{f_1}(B), \eta_{f_2}(B), \ldots\}$, with weights represented by vector W to form an intermediate opinion, O'.

2. *Phase 2*: A incorporates $\mathcal{M}_F(B)$, the faction's view about B, to the intermediate opinion O' using its faction alignment κ as the weight.

3. *Updates*: The interaction takes place, payoff ρ_A is updated, the outcomes classified according to A's satisfaction λ, and doubt $D_A(B)$ is updated.

Consider an agent A, which has m friends $\langle f_1, f_2, \ldots f_m \rangle$, wishing to form an informed opinion about another agent B, given its own and its friends' opinions of B.

Phase 1

As per the definition of opinion in (1), the opinions of B by A and its friends can be structured as a vector E, given by

$$E = \begin{bmatrix} \eta_A(B) \\ \eta_{f_1}(B) \\ \eta_{f_2}(B) \\ \vdots \\ \eta_{f_m}(B) \end{bmatrix} \tag{5}$$

The corresponding weights to each opinion are denoted by the vector W as,

$$W = \begin{bmatrix} w_A & w_{f_1} & w_{f_2} & \ldots & w_{f_m} \end{bmatrix} \tag{6}$$

Our main problem here is to come up with a W that takes A's egocentricity into account. As described in Section 3.1, we consider a normal probability distribution for this purpose.

$$w_x = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\eta_x(B)-\mu)^2/2\sigma^2} \tag{7}$$

$$\text{where } \mu = \eta_A(B), \ \sigma = \sigma_0 \times \mathcal{D}_A(B)$$

So, O', the opinion at the end of Phase 1, is given by

$$O' = W \cdot E \tag{8}$$

This can also be written in an algebraic form as

$$O' = w_A \eta_A(B) + \sum_{i=1}^{m} w_{f_i} \eta_{f_i}(B) \tag{9}$$

Phase 2

Phase 2 of opinion formation focuses on incorporating the faction's view of agent B into the opinion arrived at in phase 1, O'. Let the faction view on B be denoted by $\mathcal{M}_F(B)$ and let κ_A represent A's level of alignment towards its faction. Now, the final opinion about B is a κ-weighted average of O' and $\mathcal{M}_F(B)$

$$O = \kappa_A \mathcal{M}_F(B) + (1 - \kappa_A)O' \tag{10}$$

Updates

The updates phase starts off with updating the payoff ρ_A according to (4).

$$\rho_A = \rho_A + p_A(a, b) \tag{11}$$

Based on the outcome of this interaction with B (the level of cooperation b), A updates $\mathcal{D}_A(B)$, its doubt about B. For A to classify its interaction as successful, b has to be greater than λ. $\mathcal{D}_A(B)$ is decremented by a constant c if the interaction is successful, and it is incremented by c otherwise, as outlined in Section 3.2.

$$\mathcal{D}_A(B) = \begin{cases} \mathcal{D}_A(B) + c, & b < \lambda_A \\ \mathcal{D}_A(B) - c, & b > \lambda_A \end{cases} \tag{12}$$

Thus, A aggregates opinions about B received from its friends, taking into account its level of egocentricity and its doubt about B.

4 Agent Types

An agent pool consisting of three types of agents is considered. The agents are categorized into different types based on their internal working and attributes. The system is initially configured with attributes such as the total number of agents, the proportion of different types, the number of factions in the system, and the number of iterations. In each iteration, an agent is randomly paired with one other agent, then they interact, and finally, each agent updates its experiences and payoff. We formally define the various attributes of an agent before delving into the intricacies of each type.

Basic Attributes

All agent types have four basic attributes as described below:

- α is a unique identifier for each agent, $\alpha \in \{1, \dots, N\}$, where N is the number of agents in the system.
- ρ is the agent's cumulative payoff, a metric to capture the efficiency or performance, $\rho \in \mathbb{Z}^+ \cup \{0\}$, where $\mathbb{Z}^+$ denotes the set of positive integers.
- ω is the memory size of an agent, $\omega \in \mathbb{Z}^+$.

- $\mathcal{E}$, the experiences is a two-dimensional vector with N rows and ω columns. $\mathcal{E}[i][j] = C_i(t-j)$, where $C_X(t)$ denotes X's level of cooperation in interaction t.

$$\mathcal{E} = \begin{bmatrix} C_1(t-1) & C_1(t-2) & \dots & C_1(t-\omega) \\ C_2(t-1) & C_2(t-2) & \dots & C_2(t-\omega) \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ C_N(t-1) & C_N(t-2) & \dots & C_N(t-\omega) \end{bmatrix}$$

Extended attributes

Apart from the basic attributes, an agent type may also have several other attributes that enable their functionality and behavior. We describe them as follows:

- $\mathcal{D}$ represents the self-doubt of an agent. It is a vector indexed by agent id and reflects the level of uncertainty of the agent's own opinion about the corresponding other agent. For agent A, $\mathcal{D}_A = [\mathcal{D}_A(1), \mathcal{D}_A(2), ...\mathcal{D}_A(N)]$, $0 < \mathcal{D}_A(j) < \infty$.

- $\mathcal{F}$ is the set of friends of an agent. Friends of an agent A can be described as a subset of its faction, $\mathcal{F}_A \subset \Upsilon$ (see below).

- σ_0 represents the base spread of an agent, as defined in (2), $\sigma_0 \in \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of positive real numbers.

- λ represents the threshold of satisfaction of an agent, as outlined in Section 3.5; $0.5 \leq \lambda < 1$.

- κ represents the faction alignment of an agent; $0 \leq \kappa \leq 1$.

4.1 Factions and Friends

A faction is formally defined as a 3-tuple, $\Psi = \langle \gamma, \Upsilon, \mathcal{M} \rangle$ where:

- γ is a unique identifier for each faction.

- Υ is the set of member agents.

- $\mathcal{M}$, the central memory is a vector indexed by agent id, and each cell holds the aggregate of members' opinions about the corresponding agent.

$$\mathcal{M} = \begin{bmatrix} \frac{\sum_{k\in\Upsilon} \eta_k(1,t)}{|\Upsilon|} & \frac{\sum_{k\in\Upsilon} \eta_k(2,t)}{|\Upsilon|} & \cdots & \frac{\sum_{k\in\Upsilon} \eta_k(N,t)}{|\Upsilon|} \end{bmatrix}$$

Each faction is uniquely represented by an identifier and holds a set of agents with the condition that an agent can belong to one faction only. Each faction maintains a central memory which indicates past levels of cooperation by all agents in the system (not just the ones in the faction). The faction's memory is updated by all members at the end of each interaction, and is accessible only to its members.

The number of friends are constrained as per the description in Section 3.3. Agents always fully cooperate when they interact with their friends. Any friend can access an agent's experiences.

4.2 Partisan Agents

A partisan agent Π is formally defined as a 9-tuple as given below. A partisan agent uses all the extended attributes in addition to the basic attributes.

$$\Pi = \langle \alpha, \rho, \omega, \mathcal{E}, \mathcal{D}, \mathcal{F}, \sigma_0, \lambda, \kappa \rangle$$

Behavior of a partisan agent

Consider that agent A is paired up with agent B in one iteration. The goal for agent A is to come up with an optimum level of cooperation given its own prior experiences with B and the opinions it receives from its friends. The crux of the problem here is to come up with the necessary Gaussian distribution, defined by μ, σ. Then the opinions are collected, weighed, faction value incorporated, and then the agent cooperates at this level. The interaction takes place and then the agent updates the values and learns. The process is outlined in Algorithm 1, and it can be broken down into four meaningful steps as described below.

1. Initialization

As discussed already, the curve needs to be centered at the agent's own opinion. So, μ is set as A's opinion of B formed on the basis of its prior experiences, $\mathcal{E}_A(B)$. According to (3), σ is set as product of two factors—A's base spread (σ_0) and A's doubt on B, $\mathcal{D}_A[B]$.

Initialize two empty vectors *OpinionSet* and *Weights* to capture the opinions and their respective weights. These vectors correspond to E and W defined by (5) and (6) respectively. This initialization is shown in lines 1–4 of Algorithm 1.

2. Collection of opinions and assignment of weights

First append A's opinion and its weight to *OpinionSet* and *Weights* respectively. This is shown in lines 6–7, where the function $\text{Append}(l, i)$ appends item i to list l. $\text{GaussianPDF}(x, \mu, \sigma)$ returns the value of Gaussian PDF defined by μ and σ at x.

Iterate through the list of A's friends, and for each friend, extract its opinion about B and assign the corresponding weight according to (7). Append the opinion and the weight to *OpinionSet* and *Weights* respectively. This iteration is captured in the for loop at lines 8–13.

3. Deciding on a final level of cooperation

Perform a dot product on *OpinionSet* and *Weights* as per (8) to get the intermediate decision (O') based on local opinions (Line 14). A retrieves its faction's view on B and stores in *FactionView*. $\text{GetFactionRating}(F, X)$ returns the faction F's view about an agent B. The final level of cooperation is taken as the alignment-weighted average of O' and *FactionView* according to (10). This calculation is shown in lines 15–17 of Algorithm 1.

4. Updating payoff and doubt

Calculate payoff according to (4) and update A's payoff (Line 18). Compare B's level of cooperation b with A's threshold of satisfaction and update doubt of A on B according to (12). This condition check is done in lines 19–23 of Algorithm 1. The last line of the algorithm describes the concept of sharing experiences with its faction and that ends this interaction.

4.3 Individual Trust-Based Agents

An individual trust-based agent Ω is defined by a 5-tuple as given below. It uses only one extended attribute, λ.

$$\Omega = \langle \alpha, \rho, \omega, \mathcal{E}, \lambda \rangle$$

The only distinction here is that each cell in experiences ($\mathcal{E}$) stores the outcomes of the corresponding interaction with that agent. The outcomes can either be 0 or 1 signifying failure or success. We model the agents with an attribute called satisfaction (λ) to determine the outcome of an interaction. There is no communication or sharing of experiences among these agents and they strictly operate only based on their experiences.

Behavior of an Individual trust-based agent

Algorithm 1: Behavior of agent A of type Π

```
/* Initialize $\mu$ and $\sigma$ for Gaussian Distribution */
$\mu \leftarrow A$.Experience[$B$];
$\sigma \leftarrow \sigma_0 \times D_A(B)$;
OpinionSet $\leftarrow \emptyset$ ;
Weights $\leftarrow \emptyset$ ;
Friends $\leftarrow A$.Friends ;
Append(OpinionSet,$A$.Experience[B]);
Append(Weights,GaussianPDF($A$.Experience[B],$\mu$,$\sigma$));
/* For each friend, retrieve opinion and weight */
for $i \leftarrow$ Friends do
    $i_{exp} \leftarrow$ i.Experience[$B$];
    Append(OpinionSet, $i_{exp}$);
    w $\leftarrow$ GaussianPDF($i_{exp}$,$\mu$,$\sigma$);
    Append(Weights,w);
end
/* Perform dot product of OpinionSet and Weights */
O' $\leftarrow$ DotProduct(Weights,OpinionSet) ;
FactionView $\leftarrow$ GetFactionRating($A$.factionId,$B$);
FacAlign $\leftarrow A$.falign;
/* LvlCoop represents level of cooperation of $A$ */
LvlCoop $\leftarrow$ O' $\times (1 -$ FacAlign$)$ + FactionView $\times$ FacAlign;
Calculate and update payoff;
if $b$ ¿ $A.satisfaction$ then
    $A$.Doubt[B] = $A$.Doubt[B] - $c$;
else
    $A$.Doubt[B] = $A$.Doubt[B] + $c$;
end
Share experience with faction;
```

Algorithm 2: Behavior of agent A of type Ω

```
  /* $A$.Experiences[$B$] is a vector which represents previous
  outcomes in interactions with $B$                            */
1 LvlCoop $\leftarrow$ Average($A.Experiences[B]$);
2 Calculate and update payoff;
3 if $b$ ¿ $A.satisfaction$ then
4 |   Append($A.Experiences[B],1$);
5 else
6 |   Append($A.Experiences[B],0$);
7 end
```

Individual trust-based agents rely on their history of interactions with other agents as their only source of information to help in decision making. Consider a case where agent A is paired with agent B in an iteration. A retrieves the vector corresponding to B from its Experiences vector $\mathcal{E}_A$ and calculates an average of values and it cooperates at this level (line 1).

A interacts with B and payoffs are calculated (line 2) and updated according to (4). Each agent has an attribute called threshold of satisfaction and this helps to classify an interaction as success or failure. If agent B cooperates at a level greater than the threshold of satisfaction (λ_A), it is classified a success, and a failure otherwise. In case of success, the corresponding vector is appended with 1 and in case of a failure, it is appended with 0. This is captured in lines 3–7 of Algorithm 2.

4.4 Suspicious TFT Agents

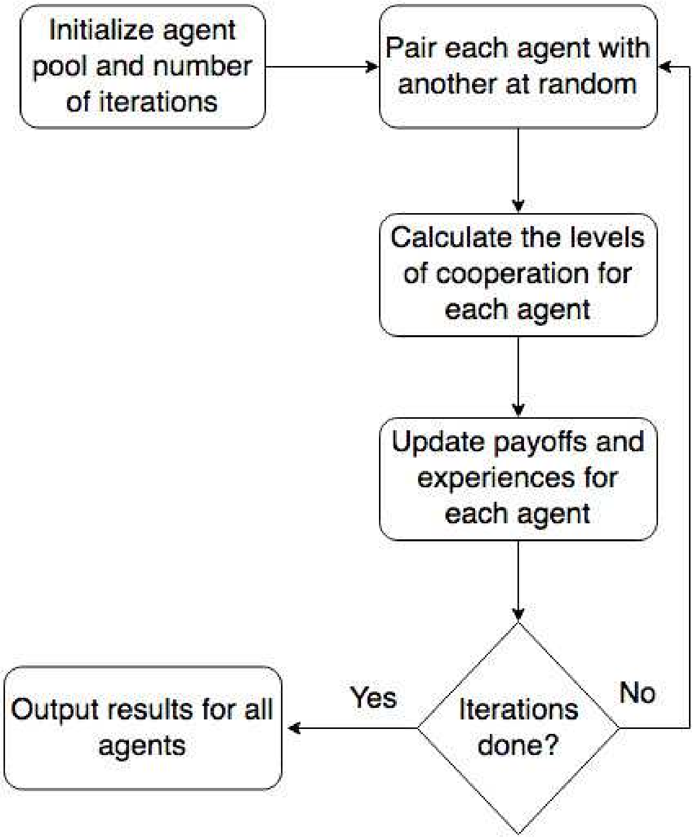
A Suspicious Tit-for-Tat (S-TFT) Agent Δ is defined by a 4-tuple and does not use any extended attribute. The only distinction here is that their experiences vector can only capture the most recent interaction with that agent i.e., $\omega = 1$.

$$\Delta = \langle \alpha, \rho, \omega, \mathcal{E} \rangle$$

S-TFT agents are a standard type of agents which have been well explored in IPD games [3, 8]. As the name suggests, an S-TFT agent A defects completely on its first interaction with B owing to its "suspicious" nature. However, in subsequent iterations, A cooperates at the same level that B has cooperated in the previous interaction.

5 Experiments and Results

The agent pool is configured with all its parameters as described in Section 4 and in each iteration, an agent is paired randomly with one other agent. At the end of an interaction, payoffs and experiences are updated. Agents capable of learning modify their self-doubt based on the outcome. This flow is outlined by Figure 2. We vary several parameters in the configuration of model and individual agents' attributes such as egocentricity to observe their effects on performance. The findings are presented in the following subsections.



Initialize agent pool and number of iterations
Pair each agent with another at random
Calculate the levels of cooperation for each agent
Update payoffs and experiences for each agent
Iterations done?
Yes
No
Output results for all agents

Figure 2: Workflow for system

5.1 The Importance of Egocentricity

To observe the impact of different degrees of egocentricity, we considered a system of 500 agents equally distributed among all 3 types. We consider 5 factions in the system and vary the value of base egocentricity (E_0). We find that payoffs are highest for an intermediate level of egocentricity and is not as good for both extremely high values and extremely low values. Our results concur with the conventional wisdom that egocentricity has to be at a moderate level for better gains (Figure 3).

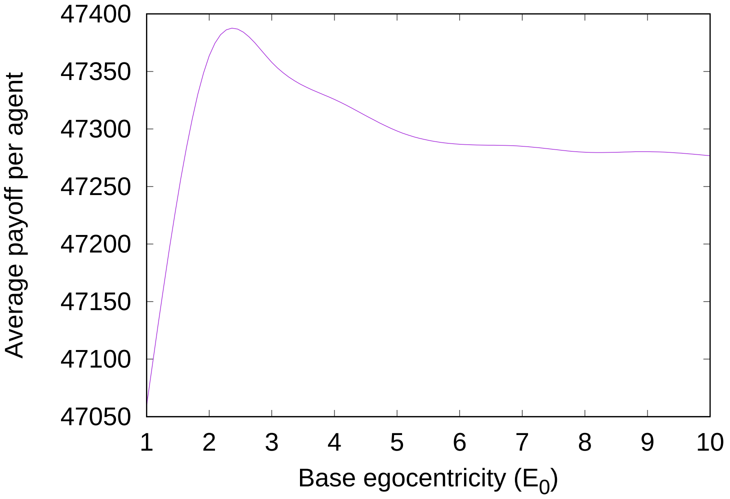


Figure 3: Effect of egocentricity on payoffs

5.2 Comparing Payoffs of All Agent Types

To understand the payoffs for each type and to see how they fare against others, the system's total number of agents is varied along with the number of factions, in such a way that each faction holds about the same number of agents. This is done to avoid any variations resulting from changes in faction size. We increase the number of agents in the system from 50 all the way up to 500. Figure 4 clearly indicates that partisan agents always perform better than the other types.

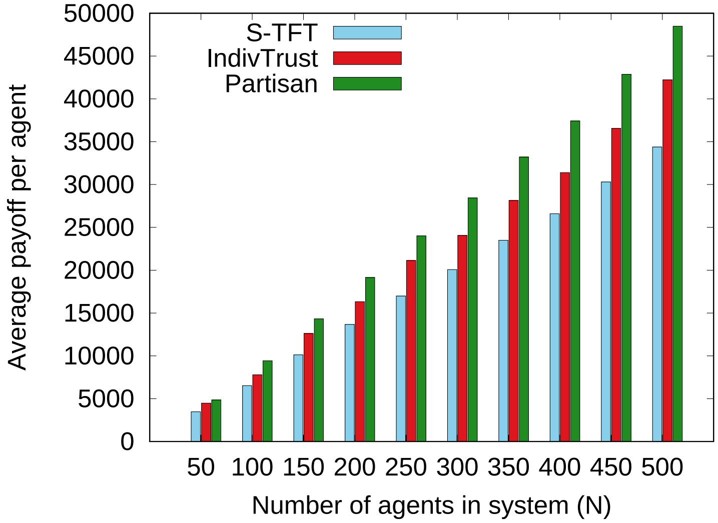


Figure 4: Comparing all types of agents

5.3 Proportion of Partisan Agents

To see if partisan agents perform better at all levels of representation in the system, we vary the proportion of partisan agents in a system of 200 agents with 5 factions. Since we are keeping the total number of agents and the number of factions constant and increasing the representation of partisan agents, factions contain on average a higher number of partisan agents, and hence their payoffs are expected to be higher, as seen in Figure 5.

5.4 Effect of Faction Size

To understand the effect of faction sizes on payoffs, we consider a system of 1300 agents with half of them partisan agents and the rest equally distributed among the other types. We consider 10 factions in the system with sizes ranging from 10 to 225. As the faction size increases, the average payoff per partisan agent also increases and this is seen in Figure 6. (Payoffs for other agent types do not depend on faction sizes, for obvious reasons.)

Network externality can be described as a change in the benefit, or surplus, that an agent derives from a good when the number of other agents consuming the same kind of good changes [42]. Over the years, various network pioneers have attempted to model how the growth of a network increases its value. One such model is Sarnoff's Law which states that value is directly proportional to size [35] (an accurate description of broadcast net-

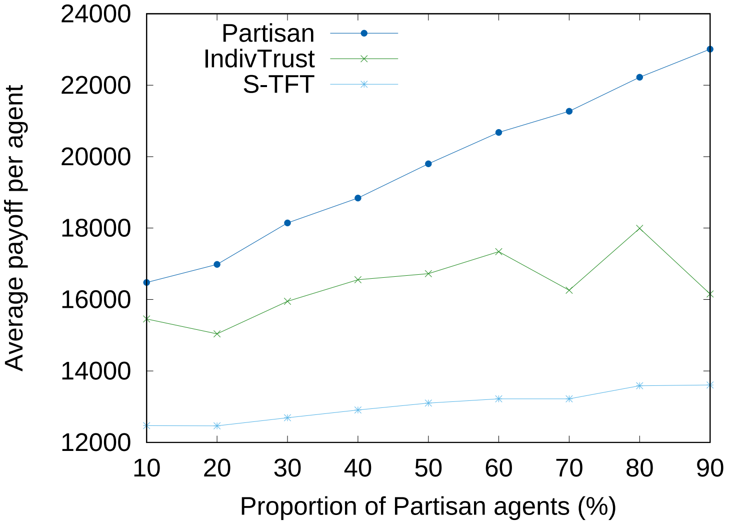


Figure 5: Effect of proportion of partisan agents on payoffs

works with a few central nodes broadcasting to many marginal nodes such as TV and radio).

Since each of our factions has one central memory that caters to all members, it is similar to broadcast networks and Figure 6 exhibits a similar proportionality (with a large offset).

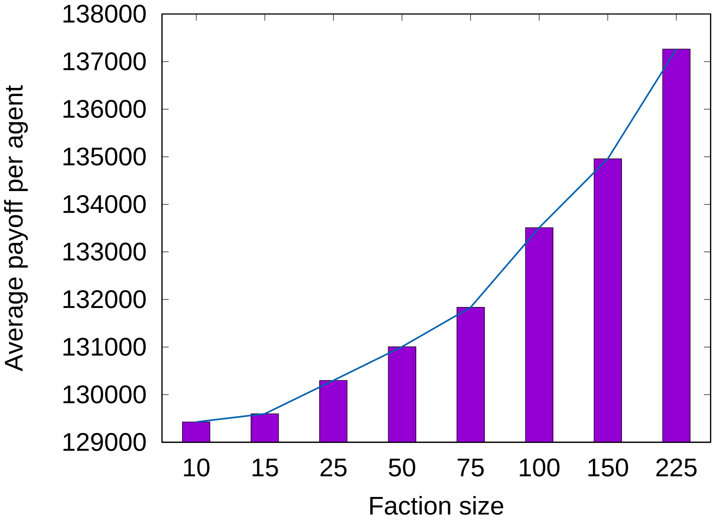


Figure 6: Effect of faction sizes on payoffs of partisan agents

5.5 Number of Interactions and Payoffs

The number of interactions is a crucial aspect when it comes to comparing strategies because S-TFT agents may gain a lot in their first interaction with other agents, and if there are no subsequent interactions with the same agents, it is highly profitable for them. However, partisan agents grow better with each interaction because of the availability of more information. We consider a system of 500 agents equally distributed among all 3 types and vary the number of interactions per agent. As expected, S-TFT agents have their best payoffs for lower numbers of interactions, but their payoffs start to fall rapidly with increasing interactions. Partisan agents steadily receive better payoffs as the number of interactions increases (Figure 7).

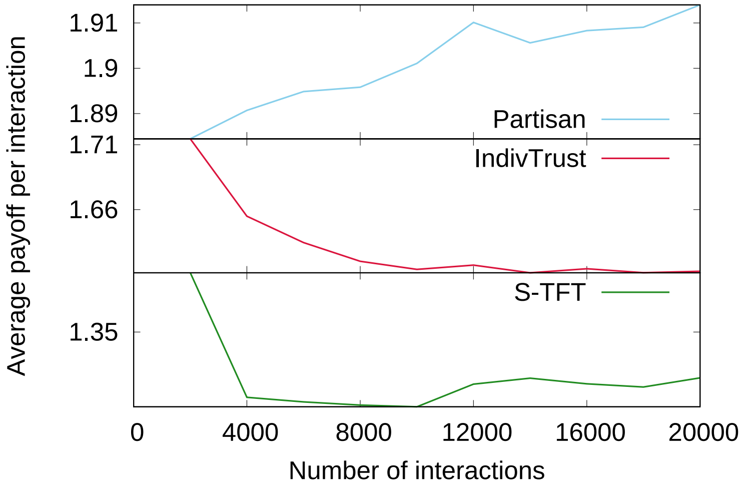


Figure 7: Number of interactions and payoffs

5.6 Number of Factions and Payoffs

For partisan agents, the number of factions in the system plays a vital role. When there are many factions in the system, agents are scattered across factions, thus weakening each faction by reducing the information contained in the faction's central memory. Hence, we expect payoffs to decrease as number of factions are increased. We have considered a system of 200 agents equally distributed among types and vary the number of factions from 10 to 100. It is clear from the Figure 8, that when factions are fewer in number, partisan agents achieve high payoffs, but as the number of factions increase, the advantage of a faction is diluted and the payoff decreases.

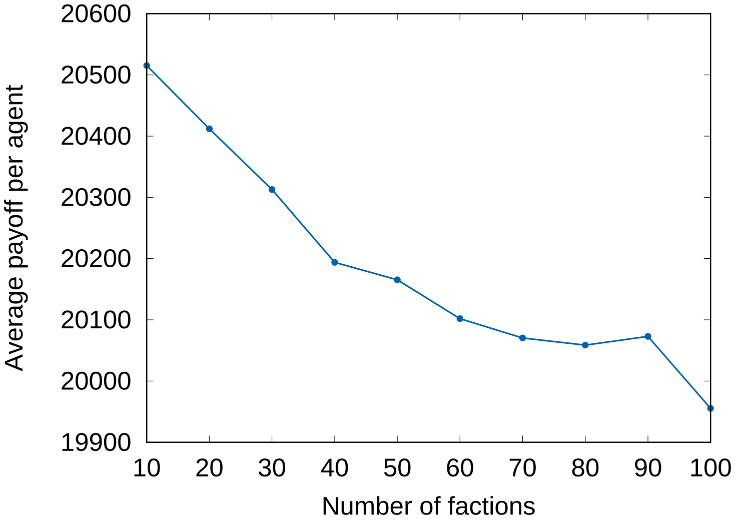


Figure 8: Number of factions and payoffs of partisan agents

6 Conclusions

We live in a deeply fragmented society where differences of opinion are sometimes so high that communication may break down in some instances. A clear model of various biases is important to understand the underlying mechanics of how some hold opinions that may seem irrational to others.

We present a model that closely captures the reality by imbibing the agents with egocentric bias and doubt. We use a symmetric distribution centered at an agent's own opinion to assign weights to various opinions and thus introduce more flexibility than previous models. We also model a response to failure, by altering the self-doubt on that topic. This balance between egocentricity and doubt enables the agent to learn reactively.

Opinion aggregation from multiple sources is now more important than ever owing to the effects of social media and mass communication. Hence, there is a need for appropriate models that realistically capture the way humans form opinions. Group opinion dynamics continue to be an area of immense interest and hence we have also introduced a model of a faction with a central memory. We observe that our model of factions seems to support the theory of network effects, and to be consistent with Sarnoff's

Law.

In people, high egocentricity may be connected with anxiety or overconfidence, and low egocentricity with depression or feelings of low self-worth. Our results also support the notion that egocentricity needs to be moderate and that either extreme is not as beneficial.

It is also observed that partisan agents generally perform much better than the other types that have been considered, which too seems to have parallels in human society.

Acknowledgements

The author S. Rao acknowledges support from an AWS Machine Learning Research Award.

References

[1] Allahverdyan, A. E., and Galstyan, A. Opinion dynamics with confirmation bias. *PLOS One 9* (07 2014), 1–14.

[2] Atwater, L. E., Dionne, S. D., Avolio, B., Camobreco, J. F., and Lau, A. W. A longitudinal study of the leadership development process: Individual differences predicting leader effectiveness. *Human Relations 52*, 12 (1999), 1543–1562.

[3] Axelrod, R., and Hamilton, W. D. The evolution of cooperation. *Science 211*, 4489 (1981), 1390–1396.

[4] Benson, B. Cognitive bias cheat sheet, 2016.

[5] Bikhchandani, S., Hirshleifer, D., and Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*, 5 (Oct. 1992), 992–1026.

[6] Bos, K. V. D., and Lind, E. A. *The Social Psychology of Fairness and the Regulation of Personal Uncertainty.* Routledge, 2009, ch. 7.

[7] Bouchet, F., and Sansonnet, J.-P. Subjectivity and cognitive biases modeling for a realistic and efficient assisting conversational agent. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 02* (2009), WI-IAT '09, IEEE Computer Society, pp. 209–216.

[8] Boyd, R., and Lorberbaum, J. P. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature 327* (1987), 58–59.

[9] Brown, J., and Marshall, M. The three faces of self-esteem. *Self-esteem: Issues and answers* (01 2006), 4–9.

[10] Burman, O. H., Parker, R. M., Paul, E. S., and Mendl, M. T. Anxiety-induced cognitive bias in non-human animals. *Physiology & Behavior 98*, 3 (2009), 345 – 350.

[11] Carron, P. M., Kaski, K., and Dunbar, R. Calling Dunbar's numbers. *Social Networks 47* (2016), 151–155.

[12] Cox, J. T., and Griffeath, D. Diffusive clustering in the two dimensional voter model. *Ann. Probab. 14*, 2 (04 1986), 347–370.

[13] Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. How can extremism prevail? a study based on the relative agreement interaction model. *J. Artificial Societies and Social Simulation 5* (2002).

[14] Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. Mixing beliefs among interacting agents. *Advances in Complex Systems 3* (01 2000), 87–98.

[15] Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H., and Quattrociocchi, W. Modeling confirmation bias and polarization. *Scientific Reports 7* (2016), 40391.

[16] Dhar, J., and Jha, A. K. Analyzing social media engagement and its effect on online product purchase decision behavior. *Journal of Human Behavior in the Social Environment 24*, 7 (2014), 791–798.

[17] Dunbar, R. I. M. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution 22*, 6 (June 1992), 469–493.

[18] Eiser, J. R., and White, M. P. A psychological approach to understanding how trust is built and lost in the context of risk. In *SCARR Conference on Trust* (December 2005).

[19] Epstein, J. M., and Axtell, R. *Growing artificial societies: social science from the bottom up.* MIT Press, 1996.

[20] Fareri, D., Chang, L., and Delgado, M. Effects of direct social experience on trust decisions and neural reward circuitry. *Frontiers in Neuroscience 6* (2012), 148.

[21] Feng, C., Feng, X., Wang, L., Wang, L., Gu, R., Ni, A., Deshpande, G., Li, Z., and Luo, Y.-J. The neural signatures of egocentric bias in normative decision-making. *Brain Imaging and Behavior* (May 2018).

[22] Furnham, A., and Boo, H. C. A literature review of the anchoring effect. *The Journal of Socio-Economics 40*, 1 (Feb. 2011), 35–42.

[23] Gill, Z. Wikipedia: Case study of innovation harnessing collaborative intelligence. In *The Experimental Nature of Venture Creation: Capitalizing on Open Innovation 2.0*, M. Curley and P. Formica, Eds. Springer, Cham, 2013, pp. 127–138.

[24] Goleman, D. A bias puts self at center of everything. *New York Times* (Jun 1984).

[25] Greenberg, J. Overcoming egocentric bias in perceived fairness through self-awareness. *Social Psychology Quarterly 46*, 2 (1983), 152–156.

[26] Greenwald, A. G. The totalitarian ego: Fabrication and revision of personal history. *American Psychologist* (1980), 603–618.

[27] Hamilton, W. D. Geometry for the selfish herd. *Journal of Theoretical Biology 31*, 2 (May 1971), 295–311.

[28] Harding, E. J., Paul, E. S., and Mendl, M. Cognitive bias and affective state. *Nature 427* (2004), 312.

[29] Hayashi, Y., Takii, S., Nakae, R., and Ogawa, H. Exploring egocentric biases in human cognition: An analysis using multiple conversational agents. In *2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing* (Aug 2012), pp. 289–294.

[30] Hegselmann, R., and Krause, U. Opinion dynamics and bounded confidence models, analysis and simulation. *Journal of Artificial Societies and Social Simulation 5*, 3 (2002), 2.

[31] Hung, W. Team-based complex problem solving: A collective cognition perspective. *Educational Technology Research and Development 61*, 3 (June 2013), 365–384.

[32] Kacperski, K., and yst, J. A. H. Opinion formation model with strong leader and external impact: a mean field approach. *Physica A: Statistical Mechanics and its Applications 269*, 2 (1999), 511 – 526.

[33] Killingback, T., and Doebeli, M. The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *The American Naturalist 160*, 4 (2002), 421–438.

[34] Korteling, J. E., Brouwer, A.-M., and Toet, A. A neural network framework for cognitive bias. *Frontiers in Psychology 9* (Sept. 2018), 1561.

[35] Kovarik, B. *Revolutions in Communication: Media History from Gutenberg to the Digital Age.* Bloomsbury Publishing, 2015.

[36] Krause, U. A discrete nonlinear and non-autonomous model of consensus formation. In *Communications in Difference Equations.* Gordon and Breach Pub., Amsterdam, 2000, pp. 227–236.

[37] Krueger, J. I., and Clement, R. W. The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of personality and social psychology 67 4* (1994), 596–610.

[38] Leister, K. D. Relations among perspective taking, egocentrism, and self-esteem in late adolescents. Master's thesis, University of Richmond, 1992.

[39] Leleno, J. M., and Sherali, H. D. A leader-follower model and analysis for a two-stage network of oligopolies. *Annals of Operations Research 34*, 1 (Dec. 1992), 37–72.

[40] Lewin, K., Dembo, T., Festinger, L., and S. Sears, R. Level of aspiration. In *Personality and the behavior disorders.* 1944, pp. 333–378.

[41] Li, J., and Xiao, R. Agent-based modelling approach for multidimensional opinion polarization in collective behaviour. *Journal of Artificial Societies and Social Simulation 20*, 2 (2017), 4.

[42] Liebowitz, S., and Margolis, S. Network externality: An uncommon tragedy. *Journal of Economic Perspectives 8* (02 1994), 133–50.

[43] Lieder, F., Griffiths, T. L., Huys, Q. J. M., and Goodman, N. D. The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review 25*, 1 (Feb. 2018), 322–349.

[44] Magessi, N. T., and Antunes, L. Modelling agents' perception: Issues and challenges in multi-agents based systems. In *Progress in Artificial Intelligence* (Cham, 2015), F. Pereira, P. Machado, E. Costa, and A. Cardoso, Eds., Springer International Publishing, pp. 687–695.

[45] Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology 2*, 2 (1998), 175–220.

[46] Nowak, A., Latane, B., and Szamrej, J. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review 97* (1990), 362–376.

[47] Nowak, A., and Lewenstein, M. Modeling social change with cellular automata. In *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Springer Netherlands, Dordrecht, 1996, pp. 249–285.

[48] Oechssler, J., Roider, A., and Schmitz, P. W. Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization 72*, 1 (Oct. 2009), 147–152.

[49] Pericherla, S., Rachuri, R., and Rao, S. Modeling confirmation bias through egoism and trust in a multi agent system. The 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC2018).

[50] Perry, R. B. The ego-centric predicament. *The Journal of Philosophy, Psychology and Scientific Methods 7*, 1 (1910), 5–14.

[51] Prechter, R. *The Wave Principle of Human Social Behavior*. New Classics Library, 1999.

[52] Raafat, R. M., Chater, N., and Frith, C. Herding in humans. *Trends in Cognitive Sciences 13*, 10 (Oct. 2009), 420–428.

[53] Rook, L. An economic psychological approach to herd behavior. *Journal of Economic Issues 40*, 1 (Mar. 2006), 75–95.

[54] Ross, L., Greene, D., and House, P. The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology 13*, 3 (May 1977), 279–301.

[55] Ross, M., and Sicoly, F. Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology 37* (1979), 322–336.

[56] Schelling, T. C. *Micromotives and Macrobehavior*. Norton, 1978.

[57] Schlenker, B. R., Salvatore Soraci, J., and McCarthy, B. Self-esteem and group performance as determinants of egocentric perceptions in cooperative groups. *Human Relations 29*, 12 (1976), 1163–1176.

[58] Silverman, B. G., Bharathy, G., Nye, B., and Eidelson, R. J. Modeling factions for “effects based operations”: part i—leaders and followers. *Computational and Mathematical Organization Theory 13*, 4 (Dec. 2007), 379–406.

[59] Silverman, B. G., Bharathy, G. K., Nye, B., , and Eidelson, R. J. Modeling factions for 'effects based operations': Part i leader and follower behaviors. *Computational and Mathematical Organization Theory 13* (Sep 2007).

[60] Silverman, B. G., Normoyle, A., Kannan, P., Pater, R., Chandrasekaran, D., and Bharathy, G. An embeddable testbed for insurgent and terrorist agent theories: Insurgisim. *Intelligent Decision Technologies 2* (2008), 193–203.

[61] Sobkowicz, P. Opinion dynamics model based on cognitive biases.

[62] Stasser, G., and Titus, W. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology 48*, 6 (June 1985), 1467–1478.

[63] Sznajd-Weron, K., and Sznajd, J. Opinion evolution in closed community. *International Journal of Modern Physics C 11*, 06 (2000), 1157–1165.

[64] Tamir, D. I., and Mitchell, J. P. Neural correlates of anchoring-and-adjustment during mentalizing. *NAS 107*, 24 (June 2010), 10827–10832.

[65] Tversky, A., and Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science 185*, 4157 (Sept. 1974), 1124–1131.

[66] Verhoeff, T. *A continuous version of the prisoner's dilemma.* Computing science notes. Technische Universiteit Eindhoven, 1993.

[67] Weisbuch, G., Deffuant, G., Amblard, F., and Nadal, J.-P. Interacting agents and continuous opinions dynamics. In *Heterogenous Agents, Interactions and Economic Performance* (Berlin, Heidelberg, 2003), R. Cowan and N. Jonard, Eds., Springer Berlin Heidelberg, pp. 225–242.

[68] Woodman, T., Akehurst, S., Hardy, L., and Beattie, S. Self-confidence and performance: A little self-doubt helps. *Psychology of Sport and Exercise 11*, 6 (2010), 467–470.