Learning Dissemination Strategies for External Sources in Opinion Dynamic Models with Cognitive Biases

Abdullah Al Maruf[1] , **Luyao Niu**[1] , **Bhaskar Ramasubramanian**[2] , **Andrew Clark**[3] and **Radha Poovendran**[1]

[1]Network Security Lab, Department of Electrical & Computer Engineering, University of Washington

[2] Electrical and Computer Engineering, Western Washington University

[3]Electrical & Systems Engineering, Washington University in St. Louis

{maruf3e, luyaoniu, rp3}@uw.edu, ramasub@wwu.edu, andrewclark@wustl.edu

Abstract

The opinions of members of a population are influenced by opinions of their peers, their own predispositions, and information from external sources via one or more information channels (e.g., news, social media). Due to individual cognitive biases, the perceptual impact of and importance assigned by agents to information on each channel can be different. In this paper, we propose a model of opinion evolution that uses *prospect theory* to represent perception of information from the external source along each channel. Our prospect-theoretic opinion model reflects traits observed in humans such as loss aversion, assigning inflated (deflated) values to low (high) probability events, and evaluating outcomes relative to an individually known reference point. We consider the problem of determining information dissemination strategies for the external source to adopt in order to drive agent opinions towards a desired value. However, computing such a strategy faces a challenge that agents' initial predispositions and functions characterizing their perceptions of information disseminated might be unknown. We overcome this challenge by using Gaussian process learning to estimate these unknown parameters. When the external source sends information over multiple channels, the problem of jointly selecting optimal dissemination strategies is in general, combinatorial. We prove that this problem is submodular, and design near-optimal dissemination algorithms. We evaluate our model on three widely-used large graphs that represent real-world social interactions. Our results indicate that the external source can effectively drive opinions towards a desired value when using prospect-theory based dissemination strategies.

1 Introduction

The opinions of members of a population are influenced by interactions with family, friends, and organizations around them. They may have an initial predisposition or bias, which determines how their opinions will evolve over time. Information from sources such as advertisements and media can have a profound impact in shaping opinions. This influence has been observed in many social settings, such as in voting patterns during elections [Cohen and Tsfati, 2009], assessing the importance of wearing a mask during the COVID-19 pandemic [González-Padilla and Tortolero-Blanco, 2020], and in public policy [Shanahan *et al.*, 2011]. Understanding the evolution of opinions in shaping individual behavior and reasoning about interactions among groups of agents has led to significant research in multiple domains, including biology and social networks [Leonard and Levin, 2022].

One way to model opinion evolution is through a *weighted average* update of opinions [Altafini, 2012; DeGroot, 1974; Deffuant *et al.*, 2000; Hegselmann *et al.*, 2002], where a weight quantifies the importance an agent assigns to an opinion of another agent. Initial predispositions [Friedkin and Johnsen, 1990], stubbornness [Tsang and Larson, 2014], and influence of external sources [Li and Zhu, 2020; Quattrociocchi *et al.*, 2014] can be incorporated as additive terms in the model. However, these frameworks do not consider the possibility of agents exhibiting cognitive biases.

Agents might show various cognitive biases [Tversky and Kahneman, 1992], including (i) having different perspectives on losses and gains, (ii) unconsciously assigning inflated values to low probability events and and deflated values to high probability events, and (iii) evaluating outcomes relative to individually adopted reference points. These factors can affect individuals' perceptions of information from the external source [Giardini and Vilone, 2021; Sobkowicz, 2018], and hence influence dynamics of opinion formation of the group. A realistic model of opinion evolution must account for cognitive biases of individual agents. Prospect theory, introduced in [Kahneman and Tversky, 1979], has been shown to effectively characterize behaviors (i) - (iii) in empirical evaluations on single individuals. While prospect-theory has been studied from the perspective of a single agent, its effectiveness in modeling and shaping the evolution of the opinions among multiple interacting agents has not been investigated.

In this paper, we propose a model of opinion evolution that uses prospect theory to represent the perception of information provided by an external source. Our model provides a computational framework to reason about opinion formation when agents have cognitive biases, and enables the design of information dissemination strategies for the external source. The opinion of each individual is updated as a linear combination of (i) a weighted average of opinions of their peers, (ii) their own initial biases, and (iii) the perception of information from the external source along each channel, proportionally scaled by the level of trust in this source.

We analyze our prospect-theoretic model from the perspective of the external source. The external source could be an advertiser who wishes to market a product to a population, and is interested in nudging individuals towards showing interest in purchasing the product. One approach to accomplish this nudging is by disseminating information about the product over multiple channels. In our model, information from the external source along each channel is represented as a probability distribution over a random variable, which represents possible outcomes of a phenomenon. For example, the random variable could offer indicators of the efficacy of a new vaccine, such as *effective with no side-effects, effective with some side-effects*, and *not effective*.

We consider a source that attempts to compute optimal information dissemination strategies, which consists of choosing the probability distribution over the random variable to steer agents' opinions towards a desired value. Computing a dissemination strategy for the external source can be challenging when initial predispositions of agents and functions characterizing agents' perceptions of information disseminated are unknown. We overcome this challenge by using Gaussian process learning [Seeger, 2004] to estimate these unknown parameters. We use these estimates to compute an optimal dissemination strategy for the external source and provide a probabilistic bound on the gap between agents' opinions computed using Gaussian process learning and their true values. When the external source can send information along multiple channels, selecting jointly optimal strategies is combinatorial. We prove that this problem is submodular, and leverage this property to develop near-optimal information dissemination algorithms with provable guarantees.

We evaluate our model on three widely used large graph networks that represent real-world social interactions: (i) the Watts-Strogatz small world graph [Watts and Strogatz, 1998], (ii) the Barabási-Albert scale-free graph [Barabási and Albert, 1999], and (iii) a Facebook friendship graph [Rossi and Ahmed, 2015]. Our results indicate that the external source can drive opinions towards a desired value more effectively when using our prospect-theoretic opinion model than a baseline which uses an expectation-based update. We also examine sensitivity of our model to changes in parameter values that characterize prospect-theoretic agent perceptions.

The remainder of this paper is organized as follows. Sec. 2 surveys related work. Sec. 3 presents our prospect-theoretic model of opinion dynamics. We analyze the model in Sec. 4 and present results of our experiments on three graph networks in Sec. 5. Sec. 6 concludes the paper.

2 Related Work

Characterizing the opinion dynamics of a group of networked agents has been a subject of research across multiple disciplines. These models allow reasoning about mechanisms that nudge the population to consensus or persistent disagreement. This section surveys related work on opinion evolution.

The DeGroot model proposed in [DeGroot, 1974] assumed that any single agent's opinion was a weighted average of opinions of its neighbors. The Friedkin-Johnsen (FJ) [Friedkin and Johnsen, 1990; Friedkin and Johnsen, 2011] model extended [DeGroot, 1974] to include a term corresponding to agents' initial opinions or prejudices. Equilibria and convergence of the FJ model was analyzed in [Ghaderi and Srikant, 2014]. When agents had opinions on multiple topics, a multidimensional FJ model was proposed in [Parsegov *et al.*, 2016]. Graph-based opinion models that additionally use a threshold to characterize interactions among agents include the Hegselmann-Krause (HK) [Hegselmann *et al.*, 2002] and Deffuant-Weisbuch [Deffuant *et al.*, 2000] models.

A weighted-median opinion update mechanism in [Mei *et al.*, 2022] allowed reasoning about scenarios that had a well-defined ordering among multiple options. However, the weighted median model does not create new opinions, and does not provide conditions for convergence or persistent disagreement. Opinion evolution has also been studied from a game-theoretic perspective in [Acemoglu and Ozdaglar, 2011; Bauso and Cannon, 2018; Etesami and Başar, 2015].

A signal from the external source was included as an additive term to the FJ-model in [Li and Zhu, 2020]. In comparison, [Quattrociocchi *et al.*, 2014] considered the effect of communication among media outlets on opinion formation. Emergent behaviors in emergencies when there was mistrust in sources of information was studied in [Giardini and Vilone, 2021]. For agents with prospect-theoretic preferences, the relationship between Nash and correlated equilibria was characterized in [Phade and Anantharam, 2019; Phade and Anantharam, 2021]. When agents had to learn behaviors in an unknown environment, prospect-theoretic learning algorithms were designed in [Borkar and Chandak, 2021; Ramasubramanian *et al.*, 2021]. These works do not provide provable optimality guarantees or consider the multiple agent problem setting.

Our Approach. Different from the above methods, our approach considers multi-agent opinion dynamics and we use prospect theory to model perception of information from an external source. We develop a submodular formulation that yields near-optimal dissemination strategies for the external source to adopt in order to drive opinions towards a desired value. We also provide provable probabilistic bounds on the gap between agents' opinions computed using Gaussian process learning and their true values.

3 Model

This section presents our model of opinion evolution that uses prospect theory to represent the perception of signals disseminated by an external source. Agents are represented as a set of nodes $\mathcal{V}$ of a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $|\mathcal{V}| = N$. There is an edge in $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ directed from node i to j in $\mathcal{G}$ if agent i's opinion can be influenced by agent j's opinion. The weight of an edge w_{ij} quantifies the importance of agent j's opinion on i's opinion. We assume that $w_{ij} \geq 0, \sum_j w_{ij} = 1$

for all i, j, and that $\mathcal{G}$ is strongly connected i.e., there is a directed path between all pairs of nodes of $\mathcal{G}$.

The opinion of an agent i at time k is denoted $x_i(k) \in \mathbb{R}$. We assume that $x_i(k) = 0$ represents a *neutral opinion*, while large negative (positive) values indicate *strong disagreement (agreement)*. We consider an external source that can broadcast information on multiple channels- for example, aiming to sell a product by advertising on television, the Internet, and newspapers. We assume that there are m channels. For each channel $l = 1, \cdots, m$, the external source chooses a dissemination strategy from the set $\mathcal{Q}_l = \{q_l^1(\theta), \cdots, q_l^r(\theta)\}$. Each $q_l(\theta) \in \mathcal{Q}_l$ is a probability distribution over a discrete random variable θ. For e.g., if $x_i(k)$ indicates an opinion of whether agent i wants to invest money in a new portfolio, their initial predisposition could be an indicator of their preference for investing an amount larger than \$1000, and each $q_l(\theta) \in \mathcal{Q}_l$ could contain information about predictive trends of the stock market from a different expert. The external source in this case will be an aggregator who provides investment advice by collecting information from multiple experts.

Due to their cognitive biases, each individual might perceive information broadcast by the external source differently. Such a perception will depend on the level of trust of the individual in the source, and on the actual information broadcast. We use insights from *prospect theory* [Kahneman and Tversky, 1979; Tversky and Kahneman, 1992] to characterize perception of information from the external source. For each agent i, we use a nonlinear *value function* $v_i : \mathbb{R} \to \mathbb{R}$, and a nonlinear *probability weighting function* $p_i : [0, 1] \to [0, 1]$, to transform the value of the random variable and the distribution $q(\theta)$. These functions are taken from empirical models of human behavior developed in the social sciences [Kahneman and Tversky, 1979] and address a tendency of humans to (i) be loss-averse, (ii) assign inflated (deflated) values to low (high) probability events, and (iii) evaluate outcomes relative to an internal reference. Choices of $p_i(\cdot)$ and $v_i(\cdot)$ that have been examined in the literature [Kahneman and Tversky, 1979] which exhibit the above properties include:

$$v_i(y) = \begin{cases} |y - Ref|^{c_+}, & y \geq Ref \\ -c_*|y - Ref|^{c_-}, & y < Ref \end{cases} \tag{1}$$

$$p_i(\eta) = \frac{\eta^\alpha}{\left(\eta^\alpha + (1-\eta)^\alpha\right)^{\frac{1}{\alpha}}}, \tag{2}$$

where $0 \leq \alpha \leq 1$, $c_* > 1$, and $c_+, c_- \in (0, 1]$. An individual's perception of information broadcast by the source on a single channel is characterized by a *prospect*, defined below.

Definition 1. *The* prospect *of a distribution $q_l(\theta) \in \mathcal{Q}_l$ for agent i on channel l is defined as $u_i^{q_l} := \sum_\theta p_i(q_l(\theta))v_i(\theta)$.*

When there are N agents, the opinion $x_i(k)$ of agent i is modeled as a linear combination of (i) a weighted average of opinions of its peers $\sum_j w_{ij}x_j(k)$, (ii) their own initial predisposition x_i^b, and (iii) the prospect $u_i^{q_l}$, appropriately scaled by the level of trust T_i that the agent has in information disseminated by the external source. We remove the explicit dependence on θ, and write $q_l \equiv q_l(\theta)$.

Let $\mathcal{S} := \{(l, q_l) : l = 1, \ldots, m; q_l \in Q_l\}$. With $2^\mathcal{S}$ denoting the power set of $\mathcal{S}$, we define a collection of sets $\mathbb{A} \subseteq 2^\mathcal{S}$ by $A \in \mathbb{A}$ if $|A \cap \{(l, q_l)\}| = 1$ for all $l \in \{1, \ldots, m\}$. We observe that $\mathbb{A}$ is the basis of a partition matroid $\mathbb{M} = (2^\mathcal{S}, Y)$ where Y are independent sets defined by $\{Y \in 2^\mathcal{S} : |Y \cap (l, q_l)| \leq 1 \; \forall l = 1, \ldots, m\}$. Each set $A \in \mathbb{A}$ thus maps to a valid dissemination strategy $q \in \mathcal{Q} = \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_m$ for the external source, and the strategy for channel l is given by the unique q_l such that $(l, q_l) \in A$. For a fixed dissemination strategy q (equivalently, for fixed $A \in \mathbb{A}$) the opinion of agent i can be written as:

$$x(k+1) = \Lambda_1 W x(k) + \Lambda_2 x^b + \Lambda_3 T \sum_{l=1}^{m} \sum_{q_l:(l,q_l)\in A} u^{q_l}, \tag{3}$$

where $x(k) := [x_1(k) \cdots x_N(k)]^T$, $x^b := [x_1^b \cdots x_N^b]^T$, $u^{q_l} := [u_1^{q_l} \cdots u_N^{q_l}]^T$, W is a $N \times N$ matrix with entries w_{ij}. We define Λ_1, Λ_2, Λ_3 as diagonal matrices with entries $\lambda_{i1}, \lambda_{i2}, \lambda_{i3}$ respectively with $\lambda_{i1}, \lambda_{i2}, \lambda_{i3} \in [0, 1]$, and T as an $N \times N$ diagonal matrix with entries T_i. In Eqn. (3), W is *row-stochastic*. We also have $\Lambda_1 + \Lambda_2 + \Lambda_3 = I$ where I is an $N \times N$ identity matrix. In the remainder of this paper, we will assume that $\Lambda_2 + \Lambda_3 \neq 0$ which implies $\Lambda_1 \neq I$.

4 Analysis

In this section, we analyze our prospect-theoretic model in Eqn. (3) and identify an information dissemination strategy that the external source can adopt in order to drive opinions of agents towards a desired value. We first define notions of convergence and stability for our model and establish conditions to satisfy these notions. Then, we describe how the external source can estimate agent opinions when some model parameters are unknown, and consequently compute an optimal strategy. Proofs of all results are available in an extended version at https://rb.gy/2fg1c.

4.1 Model Convergence and Stability

We define notions of convergence and stability to analyze the steady-state behavior of the opinion dynamics in Eqn. (3).

Definition 2. *The opinion evolution described in Eqn.* (3) *is* convergent *if there exists $x_{ss} := \lim_{k\to\infty} x(k)$ for initial agent predisposition x^b and strategy q adopted by the external source. The opinion dynamics is* stable *if it converges and x_{ss} is independent of the initial values of opinions of agents, $x(0)$.*

The values of x_i^b and $x(0)$ need not be equal. Let $u^q := \sum_{l=1}^m \sum_{q_l:(l,q_l)\in A} u^{q_l}$ where A is a valid dissemination strategy in $\mathcal{Q}$. When the strategy adopted by the external source is fixed, the value of u^q is constant, making the dynamics in Eqn. (3) linear and time-invariant. We state a result from [Parsegov *et al.*, 2016; Wolfowitz, 1963] when $\Lambda_3 = 0$.

Proposition 1 ([Parsegov *et al.*, 2016; Wolfowitz, 1963])**.** *Assume that the graph $\mathcal{G}$ is strongly connected. Then, the opinion dynamics in Eqn (3) with $\Lambda_3 = 0$ is convergent if the graph is aperiodic. The dynamics is convergent and stable if all the eigenvalues of $\Lambda_1 W$ have magnitude < 1, i.e., are located within the unit circle in the complex plane.*

The above result implies that the opinion dynamics in Eqn. (3) converges to a unique steady-state value when the external source uses a fixed strategy $q \in \mathcal{Q}$ after some time-step $\hat{k}$.

Theorem 1. *Assume $\mathcal{G}$ is strongly connected, and let $\Lambda_1 \neq I$. Suppose the strategy $q \in \mathcal{Q}$ selected by the external source is fixed from a certain time-step. Then, the opinion dynamics in Eqn.* (3) *will be stable. Moreover, the steady-state value x_{ss}^q will be independent of any strategies the external source uses prior to using the strategy q.*

Theorem 1 implies that we can limit our focus to the case where the external source uses a fixed strategy q for the analysis of the steady-state behavior of the opinion dynamics in Eqn. (3). We use x_{ss}^q to denote the steady-state value of agent opinions when the external source adopts a strategy q.

4.2 Dissemination Strategy: One Channel

We now develop techniques to determine a dissemination strategy that will enable the external source to drive agents' opinions towards a desired value when there is only one channel available to disseminate information ($m = 1$). A challenge faced by the external source is that agents' initial predispositions and functions characterizing perceptions of information broadcast by the source in Eqn. (3) may not be known. For a fixed strategy $q \in \mathcal{Q}$, we can write Eqn. (3) as

$$x(k+1) = \underbrace{\Lambda_1 W x(k)}_{\text{known}} + \underbrace{\Lambda_2 x^b + \Lambda_3 T u^q}_{\text{unknown}}. \tag{4}$$

We denote the unknown term in Eqn. (4) as $h^q := \Lambda_2 x^b + \Lambda_3 T u^q$. Though the external source does not know h^q, it can use observations of the dynamics following Eqn. (4) to estimate this term. Specifically, the external source is assumed to have access to a set of noisy observations $\mathcal{D} = \{z_d, y_d\}_{d=1}^D$ corresponding to strategies $q_d \in \mathcal{Q}$, with $z_d := (x_d, q_d) \in \mathbb{R}^N \times \mathbb{R}^m$ and $y_d \in \mathbb{R}^N$ is given by $y_d = \Lambda_1 W x(k) + h^{q_d} + \omega_d$, where $\omega_d \sim \mathcal{N}(0, \xi^2)$ for all d and $\xi \in \mathbb{R}$. The set $\mathcal{D}$ can be obtained by the external source in several ways, including through surveys or from previous history. From Sec. 4.1, we know that the steady-state value of the opinion dynamics only depends on the adopted fixed dissemination strategy after a certain time-step $\hat{k}$. Thus, the source may temporarily employ different dissemination strategies in order to construct elements of $\mathcal{D}$ without affecting the steady-state value of opinions. Once such $\mathcal{D}$ is available, the external source can use insights from Gaussian process learning [Seeger, 2004] to estimate the unknown function h^q. Using Gaussian process learning enables us to derive probabilistic bounds on the quality of estimates of h^q, and use these estimates to compute an optimal dissemination strategy.

To use Gaussian process learning, we assume that h^q can be written as $h^q = \sum_{\gamma=0}^{\infty} \alpha_\gamma \, \kappa^{q,q_\gamma}$, where $\alpha_\gamma \in \mathbb{R}$ and κ^{q,q_γ} is a bounded and continuous *kernel function*. We further assume that h^q has a bounded norm on an appropriately defined Hilbert space [Wahba, 1990]. For each $z_d \in \mathcal{D}$, let $\tilde{y}_d := y_d - \Lambda_1 W x(k)$. The posterior distribution of h^q is characterized by a Gaussian distribution $\mathcal{N}(\mu_D^q, (\sigma_D^q)^2)$ where

$$\mu_D^q = (\kappa_D^q)^T (K_D + \xi^2 I)^{-1} \tilde{y}, \tag{5}$$

$$(\sigma_D^q)^2 = \kappa^{q,q} - (\kappa_D^q)^T (K_D + \xi^2 I)^{-1} \kappa_D^q, \tag{6}$$

where $\tilde{y} = [\tilde{y}_1, \dots, \tilde{y}_D]$, K_D is a $D \times D$ matrix with entries $[K_D]_{dd'} = \kappa^{q_d, q_{d'}}$, and κ_D^q is a vector whose d-th entry is $[\kappa_D^q]_d = \kappa^{q,q_d}$. The external source uses the mean $\mu_{i,D}$ and standard deviation $\sigma_{i,D}$ to estimate the opinion update of each agent i. We state a result from [Berkenkamp *et al.*, 2017].

Proposition 2 ([Berkenkamp *et al.*, 2017])**.** *Let $\|h^q\|_\kappa \leq B$ and ω_d be ξ-sub-Gaussian. Then there exists β_D and δ such that for each $i = 1, \dots, N$,*

$$Pr[|\mu_{i,D}^q - h_i^q)| \leq \beta_D \sigma_{i,D-1}^q] \;\geq\; 1 - \delta, \tag{7}$$

where $\beta_D = B + 4\xi\sqrt{\gamma_D + 1 + \ln(1/\delta)}$.

When the external source adopts strategy q, Eqns. (5) - (7) provide a way to estimate the mean and variance of the unknown part of the opinion dynamics for agent i when using Gaussian process learning. The term h^q in Eqn. (4) can then be represented as a random vector with mean $\mu^q = [\mu_{1,D}^q, \cdots, \mu_{N,D}^q]^T$ and covariance $\Sigma^q = diag\ [(\sigma_{1,D}^q)^2, \cdots, (\sigma_{N,D}^q)^2]$, which can be used to compute the mean and covariance of x_{ss}^q, as described below.

Proposition 3. *The mean and covariance of steady-state value of agents' opinions, x_{ss}^q when the external source adopts strategy q are respectively given by:*

$$\mu(x_{ss}^q) = (I - \Lambda_1 W)^{-1} \mu^q, \text{ and} \tag{8}$$

$$\Sigma(x_{ss}^q) = (I - \Lambda_1 W)^{-1} \Sigma^q (I - W^T \Lambda_1)^{-1}. \tag{9}$$

The objective of the external source is to drive x_{ss}^q towards a desired opinion x^* by selecting a dissemination strategy from the set $\mathcal{Q}$. This can be characterized by a loss metric

$$L(q, x^*) := \frac{1}{N} \mathbb{E}[(x_{ss}^q - x^*)^T (x_{ss}^q - x^*)]. \tag{10}$$

The objective of the external source will then be to compute the optimal strategy q^* such that $L(q, x^*)$ is minimum. With $Tr[\cdot]$ denoting the trace of a square matrix (i.e., the sum of its diagonal entries), we have the following result.

Proposition 4. *The optimal policy q^* of the external source that minimizes the metric $L(q, x^*)$ is given by*

$$q^* = \arg\min_{q \in \mathcal{Q}} \Big(Tr[\Sigma(x_{ss}^q)] + (\mu(x_{ss}^q) - x^*)^T (\mu(x_{ss}^q) - x^*) \Big),$$

where $\mu(x_{ss}^q)$ and $\Sigma(x_{ss}^q)$ is given by Eqns. (8) *and* (9)*.*

Algorithm 1 in the *Appendix* presents a procedure to determine the optimal dissemination strategy for the external source based on Propositions 3 and 4.

Since we use the mean and covariance of samples to construct the unknown term h^q in Eqn. (4), there could be a gap between the true value of $(x_{ss}^{q^*} - x^*)$ and the value obtained using $(\mu(x_{ss}^{q^*}) - x^*)$. Defining $G(q, x^*) := \frac{1}{N} \Big(\big((x_{ss}^q - x^*) - (\mu(x_{ss}^q) - x^*) \big)^T \big(x_{ss}^q - x^*) - (\mu(x_{ss}^q) - x^*) \big) \Big)$, our next result provides a probabilistic bound on the magnitude of this term.

Proposition 5. *Let the constants δ and β_D be as specified in Proposition 2. Then, with $\sigma^{q^*} := [\sigma_{1,D}^{q^*}, \cdots, \sigma_{N,D}^{q^*}]^T$,*

$$Pr[G(q^*, x^*) \leq \frac{\beta_D^2}{N} (\sigma^{q^*})^T (I - W^T \Lambda_1)^{-1} (I - \Lambda_1 W)^{-1} \sigma^{q^*}] \geq (1 - \delta)^N.$$

The above bound holds for any strategy $q \in \mathcal{Q}$.

4.3 Dissemination Strategy: Multiple Channels

When the external source can broadcast information on multiple channels, its objective is to determine $q^* := (q_1^*, \dots, q_m^*) \in \mathcal{Q}$ to minimize $||x_{ss} - x^*||^2$. From Eqn. (3), $x_{ss} = (I - \Lambda_1 W)^{-1}(\Lambda_2 x^b + \Lambda_3 T \sum_{l=1}^{m} \sum_{q_l:(l,q_l)\in A} u^{q_l})$. Define $\Phi := (I - \Lambda_1 W)^{-1}\Lambda_3 T; \Psi := (I - \Lambda_1 W)^{-1}\Lambda_2 x^b$. With Φ_j denoting the columns of Φ, the objective of the external source is to determine q^* to minimize:

$$||\Psi + \sum_{j=1}^{N}\sum_{l=1}^{m}\sum_{q_l:(l,q_l)\in A} \Phi_j u_j^{q_l} - x^*||^2 \tag{11}$$
$$= \sum_{i=1}^{N}(\Psi_i - x_i^* + \sum_{j=1}^{N}\sum_{l=1}^{m}\sum_{q_l:(l,q_l)\in A} \Phi_{ij} u_j^{q_l})^2 =: g(A).$$

Since the problem of computing an optimal strategy across multiple channels is combinatorial in nature, we will leverage insights from submodular optimization to efficiently compute near-optimal strategies [Fujishige, 2005]. The next result establishes submodularity of the objective in Eqn. (11). As a consequence, the problem of selecting an optimal strategy q^* for the external source will be equivalent to identifying a basis $A \in \mathbb{A}$ (Eqn. (3)) that minimizes $g(A)$.

Theorem 2. *There exists a supermodular function $\bar{g}(A)$ such that $\bar{g}(A) = g(A)$ for all $A \in \mathbb{A}$.*

Theorem 2 allows us to recast our goal of learning a dissemination strategy as solving an equivalent constrained submodular maximization problem. Submodularity of the problem implies that an efficient local search algorithm can be used to obtain a dissemination strategy with a provable constant-factor optimality bound [Feige *et al.*, 2011]. Algorithm 2 in the *Appendix* presents a procedure to determine a dissemination strategy. At each iteration, we search over all channels $l = 1, \dots, m$ and all actions $q_l \in Q_l$. For each q_l, we check if replacing the current information dissemination strategy for channel l with q_l will reduce the cost function of Eq. (11). If so, we choose q_l as the information dissemination strategy of channel l. The procedure terminates when it is not possible to further reduce the cost.

Theorem 3 provides a bound on the gap between the solution returned by Algorithm 2 and the actual optimal solution. The proof is presented in the Appendix.

Theorem 3. *There exists a sufficiently large constant Z for which the dissemination strategy A returned by Algorithm 2 satisfies $6(Z - \bar{g}(A)) < Z - \bar{g}(A^*)$, where A^* is the optimal solution to $\min\{\bar{g}(A) : A \in \mathbb{A}\}$.*

When information about initial predispositions and functions characterizing the perceptions of information broadcast by the external source is not available, we can use the Gaussian process learning procedure from Sec. 4.2 to determine probabilistic bounds on the quality of estimates of the unknown function h^{q_l} for each channel l.

5 Experiments

In this section, we describe our experimental setup, parameters of our prospect-theoretic model, and the graph networks we use to evaluate our model. We then present our results.

5.1 Evaluation Setup and Models

We consider three graph networks:
(i) **Watts-Strogatz small world graph** [Watts and Strogatz, 1998]: We assume $N = 1000$ nodes. Edges between nodes are generated with avg. degree 8 and rewire probability 0.3.
(ii) **Barabási-Albert scale-free graph** [Barabási and Albert, 1999]: We assume $N = 1000$ nodes. We start with $N_0 = 100$ nodes and add nodes until $N = 1000$. An edge is added to connect a newly added node with existing nodes with probability proportional to the existing node's degree.
(iii) **Facebook friendship social network graph** [Rossi and Ahmed, 2015]: We use a subset of the Facebook friendship graph with $N = 2235$ nodes and 91000 edges from a dataset called 'Amherst41' collected in 2015.
The Watts-Strogatz and Barabási-Albert graphs are random graphs where the connectivity of the graphs are determined from a probability distribution. In each graph (including the deterministic Facebook friendship social network graph), we assign edge weights $w_{ij} \sim Unif(1, 10)$ and normalize the weights to ensure that the matrix W is row-stochastic.

In order to evaluate the impact of information disseminated by the external source, we compare our prospect-theoretic model of opinion evolution in Eqn. (3) with an expectation-based update model, given by:

$$x^E(k+1) = \Lambda_1 W x^E(k) + \Lambda_2 x^E(0)$$
$$+ \Lambda_3 T\, \mathbf{1}_N \big(\sum_{l=1}^{m}\sum_{q^l:(l,q^l)\in A}\sum_{\theta} q^l(\theta)\theta\big), \tag{12}$$

x^E are agent opinions when following the expectation based update model and $\mathbf{1}_N$ is an $N-$dimensional vector with all entries equal to 1. We assume that $x_i^b = x_i(0)$ for all agents. Eqn. (12) is a generalization of the model proposed in [Li and Zhu, 2020], and serves as a benchmark to evaluate the role of prospect-theory in determining an information dissemination strategy for the external source. Eqn. (12) is a special case of the opinion model in Eqn. (3) when $Ref = 0, c_+ = c_- = c_* = \alpha = 1$ in Eqns. (3) - (2). Unless indicated otherwise, $\lambda_{i1} = 0.1$, $\lambda_{i2} = (0.3 + \epsilon_i)$, and $\lambda_{i3} = (0.6 - \epsilon_i)$ for each agent i in Eqn. (3), where $\epsilon_i \sim Unif(0, 0.05)$.

The external source aims to drive the opinions of agents towards a desired value $x^* = 1$. In Eqns. (3) and (12), the parameter θ is a discrete random variable. The set of dissemination strategies $\mathcal{Q}_l$ is such that $|\mathcal{Q}_l| \in \{5, 20\}$, and each $q_l(\theta) \in \mathcal{Q}_l$ is some valid probability distribution over outcomes of the random variable θ. We set the number of channels $m = 3$ for experiments with multiple channels. The objective of the external source is to select $q \in \mathcal{Q}$ to minimize the average distance $L(q, x^*)$ of agents' final opinions from the desired opinion x^*. Agents' perceptions of information disseminated by the external source is characterized by Eqns. (3) - (2). We choose $Ref \sim Unif(-1, 1)$, c_+, c_-, α from $Unif(0, 1)$, and $c_* \sim Unif(1, 2)$. To carry out Gaussian Process learning, we assume that the external source collects $|\mathcal{D}| = 200$ noisy observations of opinions, where the noise $\omega_d \sim \mathcal{N}(0, 0.1)$. We use the squared exponential kernel κ^{q, q_γ} in Eqns. (5) - (6) to determine the mean and standard deviation of the posterior of the unknown term h^q in Eqn. (4).

Initial Opinion	WS: % Final Opinion > 0.5		BA: % Final Opinion > 0.5		FB: % Final Opinion > 0.5	
	PT (Eq. (3))	Exp. (Eq. (12))	PT (Eq. (3))	Exp. (Eq. (12))	PT (Eq. (3))	Exp. (Eq. (12))
$Unif(-1, 0)$	$\mathbf{18.5}\%$	0	$\mathbf{12.1}\%$	0	$\mathbf{7.91}\%$	0
$Unif(0, 1)$	$\mathbf{21.6}\%$	0.51%	$\mathbf{22.1}\%$	0.65%	$\mathbf{15.26}\%$	0
$Unif(-1, 0.5)$	$\mathbf{14.5}\%$	0	$\mathbf{14.7}\%$	0.1%	$\mathbf{9.49}\%$	0
$Unif(-1, -0.5)$ OR $Unif(0.5, 1)$	$\mathbf{17.7}\%$	6.5%	$\mathbf{17.9}\%$	5.6%	$\mathbf{12.13}\%$	0

Table 1: This Table summarizes the final opinions of agents when the external source aims to drive opinions towards $x^* = 1$ in the Watts-Strogatz (WS), Barabasi-Albert (BA), and Facebook social network (FB) graphs. We compare opinion evolution under our proposed prospect-theoretic model (PT) with the expectation-based update model (Exp). For different distributions of initial opinions, we examine the fraction of agents whose final opinions are in *strong agreement* ($x > 0.5$) with x^*. We observe that our PT model consistently results in a significantly larger fraction of agents whose final opinions are in strong agreement with the opinion desired by the external source.

5.2 Results

Given initial opinion values of the agents, we compare their final values when opinions evolve following (i) our model with prospect-theoretic information dissemination in Eqn. (3), (ii) the FJ model in Eqn. (20), and (iii) the expectation-based update model in Eqn. (12).

When $x^* = 1$, agent i is said to be in *strong agreement* if $x_i > 0.5$. Table 1 shows that our prospect-theoretic opinion model consistently results in a greater fractions of final opinions moving towards the desired opinion x^* for all three graph networks, and across different distributions of agents' initial predispositions when the external source disseminates information along a single channel. We believe this is because our model adequately represents the agents' heterogeneous perceptions of information (characterized by different values of u_i^q for each i in Definition 1) from the external source, which aids in guiding their final opinions towards x^*. In comparison, the expectation-based update is less effective because it does not leverage the heterogeneity of agents' perceptions.

We present additional results demonstrating the effectiveness of our model in terms of final opinions of agents when agents' initial predispositions are in *disagreement* with x^* in the *Appendix*. We observe that the external source is able to drive a larger fraction of agent opinions towards x^* when it chooses a dissemination strategy that takes into account prospect-theoretic perceptions of information (Eqn. (3)).

We compare values of the average distance of agents' final opinions from the desired opinion x^* ($L(q, x^*)$ in Eqn. (10)) when the external source takes into account prospect-theoretic perceptions of agents and when it does not during computation of an optimal dissemination strategy, given that agents' opinions evolve according to Eqn. (3). Tables 2 and 3 respectively show cases when the external source can disseminate information along one and multiple channels. We select initial predispositions of agents according to $Unif(-1, 0)$. When considering prospect-theoretic agents, the external source computes the optimal dissemination strategy q^* using Algorithm 1 when opinion evolution is determined by Eqn. (3). When the source ignores possible cognitive biases of agents, it calculates a dissemination strategy $\hat{q}$ assuming opinions evolve according to Eqn. (12). Explicitly accounting for PT-based behaviors when computing a dissemination strategy results in a smaller average distance of agents' final opinions to x^*.

Network	PT (Eqn. (3))	Exp. (Eqn. (12))
Watts-Strogatz	$\mathbf{1.488}$	2.223
Barabási-Albert	$\mathbf{1.709}$	2.457
Facebook	$\mathbf{2.343}$	2.625

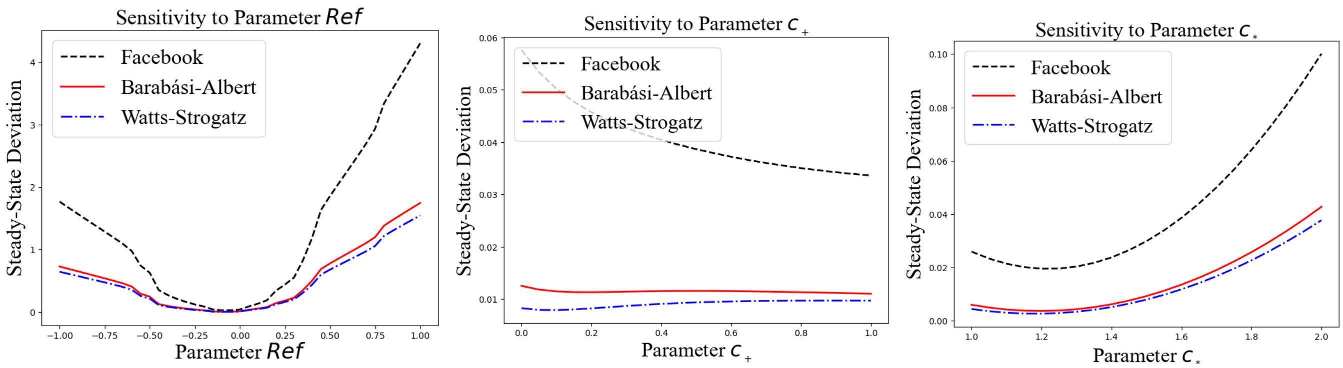
Table 2: This Table compares the average distance, $L(q, x^*)$ (Eqn. (10)), of the final opinions of agents from the desired opinion x^* when the external source sends information along exactly one channel. For all three graph networks, $L(q, x^*)$ is lower when the external source computes an optimal dissemination strategy that considers prospect-theoretic agents (*Column 2*) compared to a strategy that does not consider such behavior (expectation based-update, *Col. 3*).

Network	PT (Eqn. (3))	Exp. (Eqn. (12))
Watts-Strogatz	$\mathbf{5.518}(\mathbf{5.070})$	6.316
Barabási-Albert	$\mathbf{5.527}(\mathbf{5.078})$	6.328
Facebook	$\mathbf{5.408}(\mathbf{4.191})$	6.983

Table 3: This Table compares the average distance, $L(q, x^*)$ (Eqn. (10)), of the final opinions of agents from the desired opinion x^* when the external source sends information along three channels. We observe that for all three networks, $L(q, x^*)$ is lower when the external source computes dissemination strategies that consider prospect-theoretic agents (*Column 2*) compared to a strategy that does not consider such behavior (expectation based-update, *Column 3*). Values in parentheses in *Column 2* indicate the average distance when using the true optimal policy. We also observe that local search (Algorithm 2) is effective in finding a solution that is near-optimal.

When the external source can transmit information along $m = 3$ channels, we again observe in Table 3 that explicitly accounting for prospect-theoretic agent behavior when computing a dissemination strategy results in a smaller average distance of agents' final opinions to x^*. Further, we observe that using local search (Algorithm 2) returns a solution that is within a constant-factor gap of the actual optimal solution, which follows from Theorem 3.

Fig. 1 illustrates the role of the parameters Ref, c_+, and c_* in Eqns. (3) - (2) that characterize prospect-theoretic agents. We achieve this by comparing the average distance between steady-state values of opinions when following our prospect-



(a) (b) (c)

Figure 1: This Figure shows the sensitivity of our prospect-theoretic opinion model to parameters Ref, c_+, c_* in Eqns. (3) - (2) that characterize prospect-theoretic perceptions of agents. We measure the sensitivity by computing the *steady-state deviation,* J (defined as the average distance between steady-state opinion values when following our prospect-theoretic opinion model in Eqn. (3) and the expectation-based update model in Eqn. (12)). Fig. (a) shows that J increases as the value of the reference Ref is farther away from zero, since opinions following dynamics in Eqn. (3) and Eqn. (12) could evolve in different directions when $Ref \neq 0$. Fig. (b) shows that J is minimum when $c_+ = 1$ for the Facebook social network (black curve), since the perceived value of the signal from the external source will be equal to the actual value offset by Ref. Smaller values of c_+ will affect the signal from the external source in a nonlinear manner, which increases the value of J. The Watts-Strogatz (blue) and Barabási-Albert (red) graphs are relatively less sensitive to changes in c_+. The parameter c_* quantifies the increase in the perceived value of the signal from the external source when its true value is lower than Ref. Fig. (c) indicates that the value of J is proportional to c_*. The richness of our prospect-theoretic opinion model is underscored by the insight that the expectation-based update model corresponds to the case when $Ref = 0, c_+ = c_* = 1$.

theoretic opinion model in Eqn. (3) and when following the expectation update model in Eqn. (12). The *steady-state deviation* is $J := \frac{1}{N|Q|} \sum_{q \in Q} (x^q_{ss} - x^{E^q}_{ss})^T (x^q_{ss} - x^{E^q}_{ss})$, where x^q_{ss} and $x^{E^q}_{ss}$ are steady-state values of agent opinions in Eqns. (3) and (12) for the external source strategy q.

In Fig. 1a, we compute the value of J while varying the agent's individual reference point Ref in the range $[-1, 1]$, and keeping c_+, and c_* fixed. We observe that the average steady-state deviation is smaller when Ref is closer to 0. A reason for this is that when Ref is closer to 0, the actual value of the signal provided by the external source and the value of this signal as perceived by the agent will have the same sign. However, for values of Ref farther away from 0, the direction of opinion evolution following Eqn. (3) or Eqn. (12) could be different, which results in an increased value of J.

Fig. 1b shows values of J for $c_+ \in (0, 1)$ in Eqn. (2), while Ref, c_* are fixed. For the Facebook graph (black curve), the average steady-state deviation is minimum at $c_+ = 1$, when the perceived value of the signal from the external source will be equal to the actual value offset by Ref. For smaller values of c_+, the nonlinear value function in Eqn. (2) increases the value of signals from the external source that are lower in magnitude, while decreasing that of signals with larger magnitude. The Watts-Strogatz (blue) and Barabási-Albert (red) graphs are less sensitive to changes in c_+. This could be due to higher average degrees of nodes in these graphs.

Fig. 1c shows values of J when $c_* \in (1, 2]$ for fixed Ref and c_+. A larger c_* indicates that signals from the external source with value $y < Ref$ are increased more than signals with value $y > Ref$, which results in a larger value of J.

Setting the parameters Ref, c_+, c_* to $0, 1, 1$ respectively yields the expectation-based update model (thereby capturing identical agent preferences), showing the generality of our framework. This analysis also provides a mechanism to tune values of these parameters based on observations of agent behavior to adaptively learn optimal dissemination strategies in an adaptive manner. Determining strategies to adaptively tune these parameter values is an area of future research.

6 Conclusion

This paper proposed a new model of opinion evolution in the presence of information provided by an external source over multiple channels (e.g., Internet, television) when individuals exhibit cognitive biases. We used prospect theory to represent agents' perception of information disseminated along each channel. When initial predispositions and functions characterizing agents' perception of information were not known, we used Gaussian process learning to compute optimal strategies for the external source to adopt in order to drive opinions of individuals towards a desired value. In the multi-channel case, we additionally proved that the problem of jointly selecting optimal dissemination strategies across each channel was submodular and developed near-optimal information dissemination algorithms. Experiments on three widely adopted social network models- the Watts-Strogatz, Barabási-Albert, and Facebook friendship graph networks- showed that the external source was more effective in driving opinions of individuals towards a desired value when using our prospect-theoretic model than a baseline model that used an expectation-based update. Analyzing the sensitivity to changes in parameters associated to prospect-theoretic agent perceptions provided fine-grained insight into the role of each parameter in opinion evolution.

Ethics Statement

The opinion model developed in this paper captures cognitive biases of individuals in a population, different perceptual impacts, and importance assigned by individuals to information disseminated by an external source. Models that enable an external entity to shape opinions of members in a population raise ethical concerns, since they can be used for both societal good (e.g., promoting best practices to eliminate disease) and for harm (e.g., spread of propaganda).

Potential concerns on the role of the external source include the ability to spread propaganda, fake news, or hateful discourse by authoritarian regimes or malicious actors to manipulate public opinion in support of harmful policies. Even in democratic societies, there have been multiple reports of disinformation spread, for e.g., textbooks being rewritten in ways that marginalize or malign sections of the population, automated fake news generation [Zellers and others, 2019], and through social media platforms. Each case requires designing and implementing a different mitigating strategy.

Within a democracy, a multi-pronged solution mechanism that includes a combination of fact-checking, verification by trusted third parties, and deploying spam filters [Nakov and Da San Martino, 2020] can be adopted. Educational and outreach initiatives involving synergies among organizations with widespread reach, including governments, businesses, academic institutions, and non-profit groups across cultural and political boundaries will be critical to developing transparent solutions to understand, distinguish, and verify information disseminated by external sources [Kokuryo *et al.*, 2020]. Developers of such tools and mechanisms must periodically update their systems, and maintain continuous engagement with all stakeholders involved to improve efficacy. Research efforts that focus on adapting automated fake news generators to serve as fake news detectors have reported some success [Zellers and others, 2019]. Harmful actions by authoritarian regimes to manipulate public opinion can be thwarted through coordinated implementation (by groups of countries, or by a world-body like the UN) of preventive measures such as international sanctions to ensure adherence to international laws and norms such as well-established human rights conventions [Maclaurin *et al.*, 2019].

On the other hand, in some settings, public communication of information is of paramount importance, and recognizing heterogeneities in perceptions of, and varying levels of trust placed by individuals in information provided by an external source is critical [Rimal and Lapinski, 2009]. One such setting is public health communication, in which the source is an agency that aims to communicate best practices to eliminate diseases, or advertise effects of smoking/ sugary products to reduce cancer, obesity, and diabetes. Different from heuristics-based approaches (e.g., data-driven vaccine dissemination strategies to reduce infant mortality in Nigeria [Nair and others, 2022]), we present a principled approach to design and optimize the allocation of remedial interventions by the source that takes into account cognitive preferences exhibited by humans in the real world.

Acknowledgements

This work was supported by the Air Force Office of Scientific Research (AFOSR) through grants FA9550-23-1-0208, FA9550-22-1-0054, and FA9550-20-1-0074, and by the National Science Foundation (NSF) via grant CNS 2153136.

References

[Acemoglu and Ozdaglar, 2011] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.

[Altafini, 2012] Claudio Altafini. Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, 58(4):935–946, 2012.

[Barabási and Albert, 1999] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[Bauso and Cannon, 2018] Dario Bauso and Mark Cannon. Consensus in opinion dynamics as a repeated game. *Automatica*, 90:204–211, 2018.

[Berkenkamp *et al.*, 2017] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based RL with stability guarantees. *Advances in Neural Information Processing Systems*, 30, 2017.

[Borkar and Chandak, 2021] Vivek S Borkar and Siddharth Chandak. Prospect-theoretic Q-learning. *Systems & Control Letters*, 156:105009, 2021.

[Cohen and Tsfati, 2009] Jonathan Cohen and Yariv Tsfati. The influence of presumed media influence on strategic voting. *Communication Research*, 36(3):359–378, 2009.

[Deffuant *et al.*, 2000] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98, 2000.

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

[Etesami and Başar, 2015] Seyed Rasoul Etesami and Tamer Başar. Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions. *IEEE Trans. on Automatic Control*, 60(7):1886–1897, 2015.

[Feige *et al.*, 2011] Uriel Feige, Vahab S Mirrokni, and Jan Vondrák. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4), 2011.

[Friedkin and Johnsen, 1990] Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[Friedkin and Johnsen, 2011] Noah E Friedkin and Eugene C Johnsen. *Social influence network theory: A sociological examination of small group dynamics*, volume 33. Cambridge University Press, 2011.

[Fujishige, 2005] Satoru Fujishige. *Submodular functions and optimization*. Elsevier, 2005.

[Ghaderi and Srikant, 2014] Javad Ghaderi and Rayadurgam Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014.

[Giardini and Vilone, 2021] Francesca Giardini and Daniele Vilone. Opinion dynamics and collective risk perception: An agent-based model of institutional and media communication about disasters. *Journal of Artificial Societies and Social Simulation*, 24(1), 2021.

[González-Padilla and Tortolero-Blanco, 2020] Daniel A González-Padilla and Leonardo Tortolero-Blanco. Social media influence in the COVID-19 pandemic. *International Braz J Urol*, 46:120–124, 2020.

[Hegselmann *et al.*, 2002] Rainer Hegselmann, Ulrich Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[Kahneman and Tversky, 1979] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

[Kokuryo *et al.*, 2020] Jiro Kokuryo, Toby Walsh, and Catharina Maracke. *AI for Everyone: Benefitting from and Building Trust in the Technology*. AI Access, 2020.

[Lee *et al.*, 2009] Jon Lee, Vahab S Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Non-monotone submodular maximization under matroid and knapsack constraints. In *ACM Symp. on Theory of Computing*, 2009.

[Leonard and Levin, 2022] Naomi Ehrich Leonard and Simon A Levin. Collective intelligence as a public good. *Collective Intelligence*, 1(1):26339137221083293, 2022.

[Li and Zhu, 2020] Tingyu Li and Hengmin Zhu. Effect of the media on the opinion dynamics in online social networks. *Physica A: Statistical Mechanics and its Applications*, 551:124117, 2020.

[Liu *et al.*, 2017] Zhipeng Liu, Andrew Clark, Phillip Lee, Linda Bushnell, Daniel Kirschen, and Radha Poovendran. Submodular optimization for voltage control. *IEEE Transactions on Power Systems*, 33(1):502–513, 2017.

[Maclaurin *et al.*, 2019] James Maclaurin, Toby Walsh, Neil Levy, Genevieve Bell, Fiona Wood, Anthony Elliott, and Iven Mareels. *The effective and ethical development of AI: An opportunity to improve our wellbeing*. Report for the Australian Council of Learned Academies, 2019.

[Mei *et al.*, 2022] Wenjun Mei, Francesco Bullo, Ge Chen, Julien M Hendrickx, and Florian Dörfler. Micro-foundation of opinion dynamics: Rich consequences of the weighted-median mechanism. *Phys. Rev. Research*, 2022.

[Meyer, 2000] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. SIAM, 2000.

[Nair and others, 2022] Vineet Nair et al. ADVISER: AI-Driven vaccination intervention optimiser for increasing vaccine uptake in Nigeria. In *IJCAI*, 2022.

[Nakov and Da San Martino, 2020] Preslav Nakov and Giovanni Da San Martino. Fake news, disinformation, propaganda, and media bias. In *IJCAI Tutorial*, 2020.

[Parsegov *et al.*, 2016] Sergey E Parsegov, Anton V Proskurnikov, Roberto Tempo, and Noah E Friedkin. Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285, 2016.

[Phade and Anantharam, 2019] Soham R Phade and Venkat Anantharam. On the geometry of Nash and correlated equilibria with cumulative prospect theoretic preferences. *Decision Analysis*, 16(2):142–156, 2019.

[Phade and Anantharam, 2021] Soham R Phade and Venkat Anantharam. Learning in games with cumulative prospect theoretic preferences. *Dynamic Games and Applications*, pages 1–42, 2021.

[Quattrociocchi *et al.*, 2014] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. Opinion dynamics on interacting networks: Media competition and social influence. *Scientific Reports*, 4(1):1–7, 2014.

[Ramasubramanian *et al.*, 2021] Bhaskar Ramasubramanian, Luyao Niu, Andrew Clark, and Radha Poovendran. Reinforcement learning beyond expectation. In *IEEE Conf. on Decision and Control*, pages 1528–1535, 2021.

[Rimal and Lapinski, 2009] Rajiv N Rimal and Maria K Lapinski. Why health communication is important in public health. *Bulletin of the WHO*, 87, 2009.

[Rossi and Ahmed, 2015] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.

[Seeger, 2004] Matthias Seeger. Gaussian processes for machine learning. *Int. Jrnl. Neural Systems*, 2004.

[Shanahan *et al.*, 2011] Elizabeth A Shanahan, Mark K McBeth, and Paul L Hathaway. Narrative policy framework: The influence of media policy narratives on public opinion. *Politics & Policy*, 39(3):373–400, 2011.

[Sobkowicz, 2018] Pawel Sobkowicz. Opinion dynamics model based on cognitive biases of complex agents. *Journal of Artificial Societies and Social Simulation*, 2018.

[Tsang and Larson, 2014] Alan Tsang and Kate Larson. Opinion dynamics of skeptical agents. In *Autonomous Agents and Multi-Agent Systems*, pages 277–284, 2014.

[Tversky and Kahneman, 1992] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[Wahba, 1990] Grace Wahba. *Spline models for observational data*. SIAM, 1990.

[Watts and Strogatz, 1998] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[Wolfowitz, 1963] Jacob Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 14(5):733–737, 1963.

[Zellers and others, 2019] Rowan Zellers et al. Defending against neural fake news. *Adv. in Neural Information Processing Systems*, 32, 2019.

Learning Dissemination Strategies for External Sources in Opinion Dynamic Models with Cognitive Biases

7 Appendix

This appendix presents the proofs of our results in Sec. 4 and additional experimental results.

7.1 Proofs of Results

Two strategies q and q' in are said to belong to a *constant sum class* if and only if $u^q = u^{q'}$ (when $m = 1$, we can call it *equivalent class*). With this notion, we present a necessary condition for the convergence of the dynamics in Eqn. (3).

Proposition 6. *Assume $\mathcal{G}$ is strongly connected, and let $\Lambda_1 \neq I$. In Eqn. (3), let $\lambda_{i3} \neq 0$ and $T_i \neq 0$ for all $i = 1, \cdots, N$. Then, if the opinions of agents converge to a value x_{ss}, then there exists a time-step $k^* \geq 0$ such that the dissemination strategies adopted by the external source belong to the same constant sum class for all $k > k^*$.*

Proof. Suppose agent opinions in Eqn. (3) converges to a steady-state value x_{ss} at time k^*, for some sufficiently large value of k^*. Then, $x(k) = x_{ss}$ for all time-steps $k \geq k^*$. If the dissemination strategies adopted by the external source do not belong to the same equivalent class for all time-steps $k > k^*$, then there exists a time $\tilde{k} > k^*$ at which the source will change its strategy from q to q' where $u^q \neq u^{q'}$ according to our assumption. Using the fact that $x(\tilde{k}-1) = x(\tilde{k}) = x(\tilde{k}+1) = x_{ss}$, Eqn. (3) at time-step $k = \tilde{k} - 1$ can be written as $x_{ss} = \Lambda_1 W x_{ss} + \Lambda_2 x^b + \Lambda_3 T u^q$. Equivalently,

$$x_{ss} = (I - \Lambda_1 W)^{-1}(\Lambda_2 x^b + \Lambda_3 T u^q). \tag{13}$$

Similarly, Eqn. (3) at time-step $k = \tilde{k}$ can be written as:

$$x_{ss} = (I - \Lambda_1 W)^{-1}(\Lambda_2 x^b + \Lambda_3 T u^{q'}). \tag{14}$$

Comparing Eqns. (13) and (14) and using the fact that $\lambda_{i3} \neq 0$ and $T_i \neq 0$ for all $i = 1, \cdots N$, we obtain $u^q = u^{q'}$ which is a contradiction. Thus, dissemination strategies adopted by the external source belong to the same equivalent class for all time-steps $k > k^*$. □

Theorem 1:

Proof. Suppose the external source uses a fixed strategy starting from time $\hat{k}$. Consider the dynamics in Eqn. (3) for time-steps $k \geq \hat{k}$ with initial condition $x(\hat{k})$. Since $\Lambda_1 \neq I$ and the graph $\mathcal{G}$ is strongly connected, $\Lambda_1 W$ is sub-stochastic and irreducible. Leveraging an insight that all eigenvalues of an irreducible and sub-stochastic matrix are contained within the unit circle in the complex plane [Meyer, 2000], and Proposition 1, we can conclude that the opinion dynamics in Eqn. (3) is stable. From Definition 2, we can further conclude that the steady-state opinion x_{ss} is independent of $x(\hat{k})$. Consequently, x_{ss} will be independent of all strategies used by the external source prior to using the strategy q. □

Proposition 3:

Proof. We rewrite the dynamics in Eqn. (4) as $x(k+1) = \Lambda_1 W\ x(k) + h^q$ where the mean and covariance of h^q are given by μ^q and Σ^q, respectively. With initial condition $x(0)$,

$$x(k) = (\Lambda_1 W)^k x(0) + \sum_{j=0}^{k-1} (\Lambda_1 W)^{k-(j+1)} h^q. \tag{15}$$

When $k \to \infty$, $x(k) \to x_{ss}^q$. Substituting in Eqn. (15),

$$x_{ss}^q = \lim_{k\to\infty} \big((\Lambda_1 W)^k x(0) + \sum_{j=0}^{k} (\Lambda_1 W)^k h^q\big). \tag{16}$$

Moreover, since all the eigenvalues of $\Lambda_1 W$ have magnitude < 1, we have that $\lim_{k\to\infty}(\Lambda_1 W)^k = 0$ and $\lim_{k\to\infty}\sum_{j=0}^{k}(\Lambda_1 W)^k = (I - \Lambda_1 W)^{-1}$. Thereby we can write Eqn. (16) as:

$$x_{ss}^q = (I - \Lambda_1 W)^{-1} h^q. \tag{17}$$

With $\mathbb{E}[\cdot]$ denoting the expectation, from (17) we have:

$$\begin{aligned}
&\mu(x_{ss}^q) = \mathbb{E}[(I - \Lambda_1 W)^{-1} h^q] = (I - \Lambda_1 W)^{-1}\mu^q, \text{ and}\\
&\Sigma(x_{ss}^q) = \mathbb{E}[(x_{ss}^q - \mu(x_{ss}^q))(x_{ss}^q - \mu(x_{ss}^q))^T]\\
&\quad = (I - \Lambda_1 W)^{-1}\mathbb{E}[(h^q - \mu^q)(h^q - \mu^q)^T]\big((I - \Lambda_1 W)^{-1}\big)^T\\
&\quad = (I - \Lambda_1 W)^{-1}\Sigma^q (I - W^T\Lambda_1)^{-1},
\end{aligned}$$

which completes the proof. □

Proposition 4:

Proof. Using the fact that $\mathbb{E}[x_{ss}^q - \mu(x_{ss}^q)] = 0$, we can write

$$\begin{aligned}
L(q, x^*) &= \frac{1}{N}\ \mathbb{E}[(x_{ss}^q - \mu(x_{ss}^q) + \mu(x_{ss}^q) - x^*)^T\\
&\qquad (x_{ss}^q - \mu(x_{ss}^q) + \mu(x_{ss}^q) - x^*)]\\
&= \frac{1}{N}\ \mathbb{E}[Tr[(x_{ss}^q - \mu(x_{ss}^q))(x_{ss}^q - \mu(x_{ss}^q))^T]]\\
&\qquad + (\mu(x_{ss}^q) - x^*)^T(\mu(x_{ss}^q) - x^*)\\
&= \frac{1}{N}\ \big(Tr[\Sigma(x_{ss}^q)] + (\mu(x_{ss}^q) - x^*)^T(\mu(x_{ss}^q) - x^*)\big),
\end{aligned} \tag{18}$$

which completes the proof. □

Proposition 5:

Proof. First note that we can write

$$\begin{aligned}
\big(x_{ss}^{q^*} - x^*\big) - \big(\mu(x_{ss}^{q^*}) - x^*\big) &= (x_{ss}^{q^*} - \mu(x_{ss}^{q^*}))\\
&= (I - \Lambda_1 W)^{-1}(h^{q^*} - \mu^{q^*}).
\end{aligned} \tag{19}$$

From Proposition 2, $Pr[|h_i^{q^*} - \mu_{i,D}^{q^*}| \leq \beta_D\ \sigma_{i,D-1}^{q^*}] \geq 1 - \delta$ for each agent i. We assume that the stopping criterion for computing statistics of h^{q^*} yields $\sigma^{q^*}_{i,D-1} = \sigma^{q^*}_{i,D}$ and that the unknown portion $h^{q^*}_i$ in Eqn. (4) corresponding to each agent i is independent of every other agent. Then, $Pr[|h^{q^*} - \mu^{q^*}| \leq \beta_D \, \sigma^{q^*}] \geq (1-\delta)^N$. Combining this with Eqn. (19) and the fact that $(I - \Lambda_1 W)^{-1}$ is a non negative matrix, we recover the inequality in the statement of Proposition 5. □

Algorithm 1 presents a procedure to determine the optimal dissemination strategy based on Propositions 3 and 4.

Algorithm 1 Algorithm to compute optimal strategy q^* for the external source

- **Input**: Dataset $\mathcal{D}, \Lambda_1, W, \mathcal{Q}, x^*$.
- **Output**: q^*.
- Find μ^q and σ^q $\forall q \in \mathcal{Q}$ using Eqns. (5) and (6).
- Calculate $\mu(x^q_{ss})$ and $\Sigma(x^q_{ss})$ for all $q \in \mathcal{Q}$ using Proposition 3.
- Compute $L(q, x^*) = \frac{1}{N}\Big(Tr[\Sigma(x^q_{ss})] + (\mu(x^q_{ss}) - x^*)^T (\mu(x^q_{ss}) - x^*)\Big) \; \forall q \in \mathcal{Q}$.
- Choose $q^* = \arg\min_{q \in \mathcal{Q}} L(q, x^*)$.

Theorem 2:
We first state an intermediate result from [Liu *et al.*, 2017].

Proposition 7 ([Liu *et al.*, 2017]). *If $\hat{g}(A) = -h(\sum_{l \in A} \alpha_l - \beta)$ for some convex function $h(\cdot)$, where $\alpha_l \geq 0, \beta \in \mathbb{R}$, then $\hat{g}$ is submodular in A.*

Proof. (Theorem 2) We have that $g(A) = \sum_i g_i(A)$, where $g_i(A) = (\Psi_i - x^*_i + \sum_{j=1}^N \sum_{l=1}^m \sum_{q_l:(l,q_l)\in A} \Phi_{ij} u^{q_l}_j)^2 = (\Psi_i - x^*_i + \sum_{l=1}^m \sum_{q_l:(l,q_l)\in A} \sum_{j=1}^N \Phi_{ij} u^{q_l}_j)^2$. Let $h(\cdot) = (\cdot)^2$ in Proposition 7. Define $\Theta_i = \{(l, q_l) : \sum_{j=1}^N \Phi_{ij} u^{q_l}_j < 0\}$ and $\Theta_{il} = \{q_l : (l, q_l) \in \Theta_i\}$. Finally, let $\bar{\beta}_i = x^*_i - \Psi_i - \sum_{(l,q_l)\in\Theta_i} \sum_{j=1}^N \Phi_{ij} u^{q_l}_j$ and define

$$\bar{\alpha}_{ilq_l} = \sum_{j=1}^N \Phi_{ij} u^{q_l}_j - \sum_{\hat{q}_l \in \Theta_{il}} \sum_{j=1}^N \Phi_{ij} u^{\hat{q}_l}_j$$

We have that $\bar{\alpha}_{ilq_l} \geq 0$. Hence, according to the Proposition 7, the function $\bar{g}_i(A) = \Big(\sum_{(l,q_l)\in A} \bar{\alpha}_{ilq_l} - \bar{\beta}_i\Big)^2$ is supermodular in A. It remains to show that $\bar{g}_i(A) = g_i(A)$ for all $A \in \mathbb{A}$. Note, $\sum_{(l,q_l)\in A} \sum_{\hat{q}_l\in\Theta_{il}} \sum_{j=1}^N \Phi_{ij} u^{\hat{q}_l}_j - \sum_{(l,q_l)\in\Theta_i} \sum_{j=1}^N \Phi_{ij} u^{q_l}_j = 0$. As $A \in \mathbb{A}$, there is exactly one q_l with $(l, q_l) \in A$ for all $l = 1, \ldots, m$. Thereby we can write,

$$\left(\sum_{(l,q_l)\in A} \bar{\alpha}_{ilq_l} - \bar{\beta}_i\right) = \sum_{l=1}^m \sum_{q_l:(l,q_l)\in A} \sum_{j=1}^N \Phi_{ij} u^{q_l}_j + \Psi_i - x^*_i$$

which completes the proof. □

Algorithm 2 gives a procedure to determine a dissemination strategy for the external source when it can send information along multiple channels. Starting from a randomly selected information strategy that spans the basis of $\mathcal{G}$, at each iteration, we construct a new dissemination strategy as $A \cup \{(l, q_l)\} \setminus \{(l, q'_l)\}$, where $\{(l, q_l)\}$ is a new strategy for channel l and (l, q'_l) is the previously chosen strategy for channel l. If the newly constructed strategy in *Line 7* of Algorithm 2 satisfies $g(A \cup \{(l, q_l)\} \setminus \{(l, q'_l)\}) < (1-\epsilon) g(A)$, we then the dissemination strategy is updated to the newly constructed one. Theorem 3 shows that Algorithm 2 yields a feasible solution to the constrained submodular maximization problem with a constant-factor optimality bound.

Algorithm 2 Local search for dissemination strategies

1: Set $\epsilon \in [0, 0.5)$
2: Initialize A to define a valid dissemination strategy
3: $flag \to 1$
4: **while** $flag == 1$ **do**
5:     $flag \to 0$
6:     **for** $(l, q_l) \notin A, \; (l, q'_l) \in A$ **do**
7:         **if** $A \cup \{(l, q_l)\} \setminus \{(l, q'_l)\} \in \mathbb{A}$ **then**
8:             **if** $g(A \cup \{(l, q_l)\} \setminus \{(l, q'_l)\}) < (1-\epsilon) g(A)$ **then**
9:                 $A \leftarrow A \cup \{(l, q_l)\} \setminus \{(l, q'_l))\}$
10:                 $flag \leftarrow 1$
11:                 **Break**
12:             **end if**
13:         **end if**
14:     **end for**
15: **end while**
16: **return** A

Theorem 3:
By selecting a sufficiently large constant Z and using Theorem 2, we have that $Z - \bar{g}(A)$ is positive and submodular in A. The optimality guarantee then follows from [Lee *et al.*, 2009, Theorem 22].

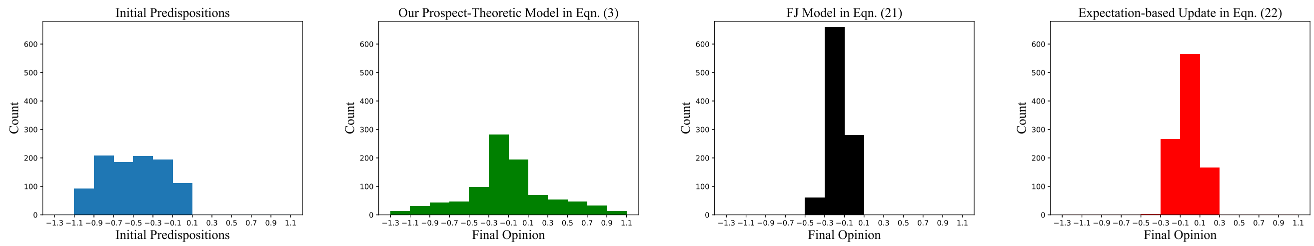
7.2 Additional Experiments

In order to evaluate the impact of information disseminated by the external source, we compare our prospect-theoretic model of opinion evolution in Eqn. (3) with two other opinion models: the FJ model [Friedkin and Johnsen, 2011] and an expectation-based update model, given by:

$$x^{FJ}(k+1) = \Lambda_1 W x^{FJ}(k) + \Lambda_2 x^{FJ}(0), \tag{20}$$

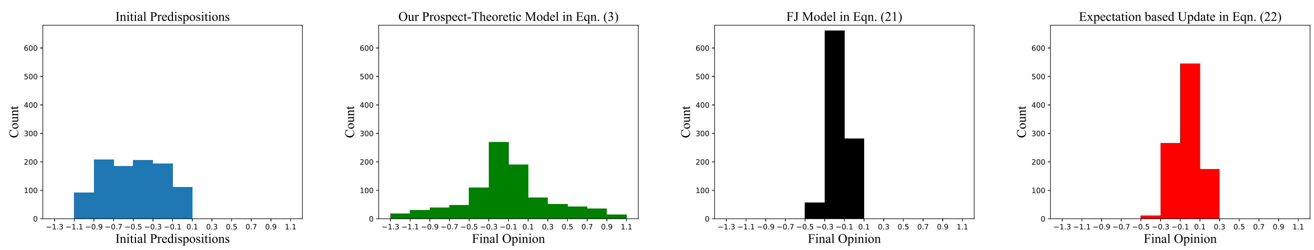
$$x^{E}(k+1) = \Lambda_1 W x^{E}(k) + \Lambda_2 x^{E}(0) + \Lambda_3 T \, \mathbf{1}_N \big(\sum_{l=1}^m \sum_{q^l:(l,q^l)\in A} \sum_{\theta} q^l(\theta)\theta\big), \tag{21}$$

where x^{FJ} and x^E are agent opinions when following the FJ model and expectation based update model, respectively. We assume that $x^b_i = x_i(0)$ for all agents. The FJ model does not have an external source, and serves as a benchmark to assess the impact of the external source. The expectation update model is a generalization of the model proposed in [Li and Zhu, 2020], and serves as a benchmark to evaluate the role of prospect-theory in determining an information dissemination strategy for the external source.



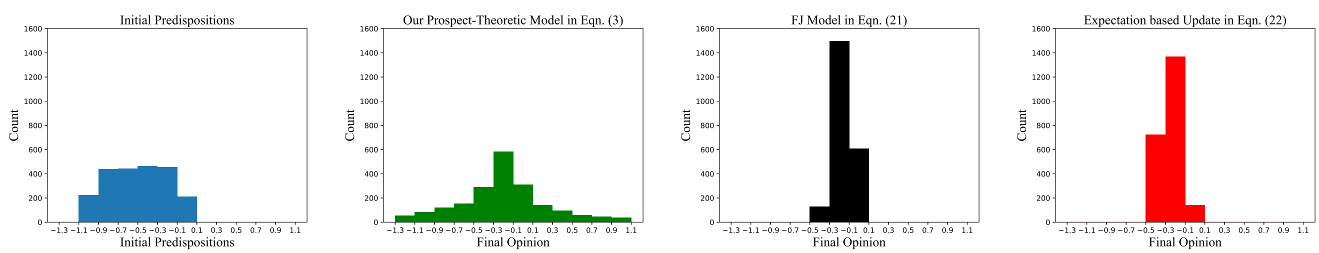
(a) (b) (c) (d)

Figure 2: This Figure shows histograms of agents' final opinions when initial predispositions of agents are in *disagreement* ($x_i(0) < 0$) with the desired value $x^* = 1$ towards which the external source aims to drive agent opinions for the Watts-Strogatz graph with $N = 1000$ agents. We assume that initial predispositions are randomly selected from $Unif(-1, 0)$ (Fig. (a)). Using a prospect-theoretic information dissemination strategy allows the external source to drive a larger fraction of agent's final opinions towards x^* (Fig. (b)) than when there is no external source (FJ-model, Fig. (c)) or an expectation-based update model (Fig. (d)).



(a) (b) (c) (d)

Figure 3: This Figure shows histograms of agents' final opinions when initial predispositions of agents are in *disagreement* ($x_i(0) < 0$) with the desired value $x^* = 1$ towards which the external source aims to drive agent opinions for the Barabási-Albert graph with $N = 1000$ agents. We assume initial predispositions are randomly selected from $Unif(-1, 0)$ (Fig. (a)). Using a prospect-theoretic information dissemination strategy allows the external source to drive a larger fraction of agent's final opinions towards x^* (Fig. (b)) than when there is no external source (FJ-model, Fig. (c)) or an expectation-based update model (Fig. (d)).



(a) (b) (c) (d)

Figure 4: This Figure shows histograms of agents' final opinions when initial predispositions of agents are in *disagreement* ($x_i(0) < 0$) with the desired value $x^* = 1$ towards which the external source aims to drive agent opinions for the Facebook social network with $N = 2235$ agents. We assume initial predispositions are randomly selected from $Unif(-1, 0)$ (Fig. (a)). Using a prospect-theoretic information dissemination strategy allows the external source to drive a larger fraction of agent's final opinions towards x^* (Fig. (b)) than when there is no external source (FJ-model, Fig. (c)) or an expectation-based update model (Fig. (d)).

We show the final opinions of agents for the desired value $x^* = 1$ when agents' initial predispositions are in *disagreement* with x^*, i.e., $x_i(0) \sim Unif(-1, 0)$ for all i for the Watts-Strogatz (Fig. 2), Barabási-Albert (Fig. 3), and the Facebook social network (Fig. 4) graphs. The initial disposition of agents in each case is chosen randomly according to $Unif(-1, 0)$ (Fig. 2a, Fig. 3a, Fig. 4a). Fig. 2b indicates that the external source is able to drive a larger fraction of agent opinions towards x^* when it chooses a dissemination strategy that takes into account prospect-theoretic perceptions of information following Eqn. (3) for the Watts-Strogatz graph. In comparison, Fig. 2c and Fig. 2d indicate that agents' opinions cannot be driven towards x^* when there is no external source (Eqn. (20)) or when the external source chooses a dissemination strategy that does not consider prospect-theoretic perceptions of agents (Eqn. (21)).

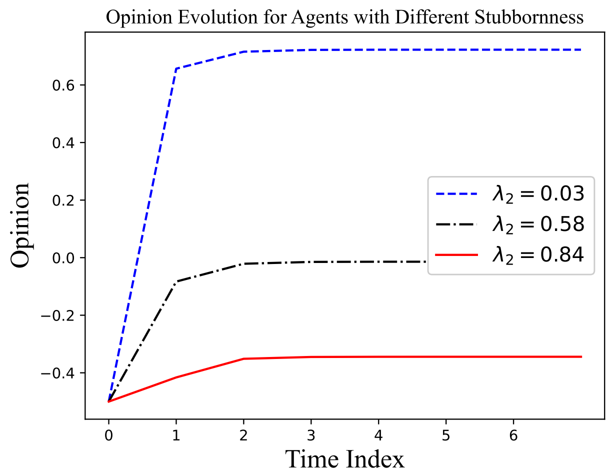


Figure 5: This Figure shows the temporal evolution of opinions of three agents in the Facebook social network following our prospect-theoretic opinion dynamics model in Eqn. (3). Each agent has a different value of the stubbornness parameter λ_2. We observe that the external source is most effective in changing the opinion of the least stubborn agent ($\lambda_2 = 0.03$) in terms of the change from the agent's initial opinion (blue curve). The opinion of the most stubborn agent ($\lambda_2 = 0.84$) is not significantly affected by opinions of its neighbors or by inputs provided by the external source (red curve).

Fig. 3b - 3d and Fig. 4b - 4d indicate similar results for the Barabási-Albert and Facebook social network graphs.

We additionally investigate how opinions of agents with different stubbornness levels ($\lambda_2 \in \Lambda_2$ in Eqn. (3)) evolve over time. Fig. 5 shows the evolution of the opinions of three agents in the Facebook social network when the external source adopts our proposed dissemination strategy. The initial opinions of all the agents is set to $x_i(0) = -0.5$. We make the following observations. First, the opinions of all three agents converge to a steady-state value within 6 time-steps, independent of the value of the stubbornness parameter λ_2. This is consistent with our result in Theorem 1. We further observe that the external source is most effective in driving the opinion of the least stubborn agent (lowest value of $\lambda_2 = 0.03$) in terms of the change from the agent's initial opinion (blue curve in Fig. 5). On the other hand, the opinion of the most stubborn agent (largest value of $\lambda_2 = 0.84$) is not significantly affected by opinions of its neighbors and the external source (red curve in Fig. 5). This behavior is because such an agent ascribes greater importance to its own initial predisposition than to opinions of its neighbors and information provided by the external source. The opinion of a 'moderately stubborn' agent ($\lambda_2 = 0.58$) is shown by the black curve in Fig. 5).