

Article

Not peer-reviewed version

---

# Modeling the Spreading of Fake News Through the Interactions Between Human Heuristics and Recommender Systems

---

[Franco Bagnoli](#)\*, [Tijan Juraj Cvetković](#), [Andrea Guazzini](#), [Pietro Lió](#), [Riccardo Romei](#)

Posted Date: 15 February 2026

doi: 10.20944/preprints202602.1064.v1

Keywords: recommender systems; diffusion of fake news; cognitive psychology; human heuristics







Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

# Modeling the Spreading of Fake News Through the Interactions Between Human Heuristics and Recommender Systems

Franco Bagnoli <sup>1,4,5,6,\*</sup> , Tijan Juraj Cvetkovic <sup>2,4,7</sup> , Andrea Guazzini <sup>2,4</sup> , Pietro Lió <sup>3</sup>  and Riccardo Romei <sup>2,4</sup>

<sup>1</sup> Dept. Physics and Astronomy, University of Florence, via G. Sansone 1, 50019 Sesto Fiorentino, Italy

<sup>2</sup> Department of Education, Literatures, Intercultural Studies, Languages and Psychology, University of Florence, 50135 Florence, Italy

<sup>3</sup> University of Cambridge

<sup>4</sup> Center for the Study of Complex Systems, Florence, Italy

<sup>5</sup> INFN, Sect. Florence, Florence, Italy

<sup>6</sup> Espace-Dev, UMR 228 University of Perpignan via Domitia UPVD, 52 Avenue Paul Alduy, 6600 Perpignan, France

<sup>7</sup> Laboratory of Agent Based Social Simulation, Institute of Cognitive Science and Technology, CNR, via Romagnosi, 18/A, 00196 Roma, Italy

\* Correspondence: franco.bagnoli@unifi.it

† These authors contributed equally to this work.

## Abstract

In many cases, the pieces of information at our disposal come from a recommender source, that can be either an official news system, a large language model or simply a social network. Often, also, these messages are build so to promote their active spreading, which, on the other hand, has a positive effect on one's own popularity. However, the content of the message can be false, giving origin to a phenomenon analogous to the spreading of a disease. In principle, there is always the possibility of checking the correctness of the message by "investing" some time, so we can say that this checking has a cost. We develop a simple model based on the mechanism of "risk perception" (propensity of checking the falseness of a message) and mutual trustability, based on the average number of fake messages received and checked. On the other side, the probability of emitting a fake message is inversely proportional to risk perception and the affinity (trustability) among agents is also exploited by the recommender system. This model represents an integration of cognitive psychology with computational agent-based modeling.

**Keywords:** recommender systems; diffusion of fake news; cognitive psychology; human heuristics

## 1. Introduction

We first present the cognitive psychology background (i.e., the concept of cognitive heuristics), and then the motivation for developing a model that tries to incorporate these aspects and their interplay with a recommender system.

### 1.1. The cognitive background: from gossips to heuristics

Cognitive heuristics and their indirect effects, namely cognitive biases, evolved under ancestral conditions and include phenomena such as majority bias, shared information bias, and status generalization [1–4]. In contemporary digital environments, these mechanisms interact with technologically-mediated biases, such as filter bubbles, echo chambers, deindividuation and phantom emotions [5–8], whose effects are best understood by jointly considering bounded rationality [9], the social brain hypothesis [10] and social locomotion [11] within networked informational environments. At the group level, the coordinated interaction of these mechanisms gives rise to emergent collective outcomes. From this perspective, collective intelligence [12,13] can be defined as a group's capacity to

maximize the fitness of its members. However, the rapid transformation of the human informational ecosystem has rendered many evolutionarily-tuned cognitive mechanisms systematically imperfect under contemporary conditions. The regulatory systems governing human communities emerged in small-scale social environments, as suggested by archaeological evidence associated with early nomadic Homo sapiens populations [14]. Although these systems have undergone adaptation and refinement, many mechanisms structuring contemporary social life, such as gossip, social dominance, stigma, and misinformation, remain deeply rooted in evolutionary contexts that predate modern technological environments by millennia. Within these communities, social norms, roles, status systems, and group cohesion played a central regulatory role [15], supported by strong pressures toward conformity and coordinated goal pursuit (social locomotion) [11]. Among the mechanisms supporting norm enforcement and group cohesion, gossip plays a central role and can be broadly defined as communication about an absent or unaware third party [16]. From a functional perspective, gossip has been interpreted as an evolutionarily grounded, prosocial mechanism supporting cohesion and cooperation [17,18]. While such mechanisms evolved under conditions of bounded relational density, contemporary digital environments profoundly alter the scale, visibility, and accountability of social interaction. As a consequence, prosocial gossip extends into computer-mediated communication, where reduced social costs coexist with preserved or amplified reputational benefits. Under these conditions, the diffusion of false or unverified information may become an efficient strategy for social positioning, helping to explain the persistence and social salience of gossip and misinformation in large-scale digital environments.

### 1.2. Cognitive Heuristics

This paper presents a computational investigation on the interplay of recommender systems and human cognitive heuristics for what concerns the diffusion of fake news.

In many cases of ordinary life, we humans (and presumably all animals) are forced to take decision in the absence of a sufficient quantity of information and often in short times. In addition, we are generally reluctant in investing cognitive resources in elaborating the available information. We therefore often rely on “quick and dirty” procedures called heuristics, which are sometimes prone to errors (and exploited by scammers) but can also provide effective decision techniques [19].

In the early 1970s, Daniel Kahneman and Amos Tversky (K&T) started investigating how humans take decisions in uncertainty conditions [1,20–23]. Their main result was that “*people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations*”.

Although K&T claimed that, as a general rule, heuristics are quite valuable, in some cases, they can lead “*to severe and systematic errors*”. What is important (and what has been already noted by scammers) is that these errors follow certain statistics and, therefore, they could be described and predicted. It is therefore important to include these aspects in any model aiming at representing the typical human behavior (i.e., the algorithmic understanding of the human logic [23]).

K&T described three general-purpose heuristics: **representativeness**, **availability** and **anchoring**.

People exploit the *availability* heuristic when dealing with a probabilistic problem by relying upon the knowledge that is readily available (i.e., when examples are easily recalled) rather than examine all possible alternatives. For example, one may assess the risk of a flight accident by how easy an instance of an accident is retrieved from memory.

Availability is however an useful proxy for assessing frequency or probability, because instances of common events are usually recalled better and faster than instances belonging to less frequent classes. The problem is that this heuristic can lead to excessive fear of small risks and neglect of large ones.

The *representativeness* heuristic occurs when assessing the degree of correspondence between a sample and a population, an instance and a category, an act and an actor or, more generally, between an outcome and a model. Representativeness is composed by categorization and generalization: in order to forecast the behavior of an unknown subject, we first identify the group to which it belongs

(categorization) and then we associate the “typical” behavior of the group to the item. It is apparent that the representativeness heuristic will produce problems whenever people are ignoring base-rates.

K&T also suggested that estimates are often made from an initial value, or *anchoring*, which is then adjusted to produce a final answer. In one investigation, K&T asked participants to estimate whether the following number that was going to be (randomly) extracted will be higher or lower than the relevant percentage. It turned out that the starting point, though clearly random, greatly affected people’s answers. If the starting point was above the average, so it was the estimate, and vice versa.

We have also to consider that, according with time constraints, stress and importance of the context, human thinking exhibits a “dual-process” character [19,24–27]. According to these investigations, people have two systems for making decisions. System I is rapid, intuitive, but sometimes error-prone, while system II involves more reasoning (and more cognitive load) and is therefore slower. We can assume that the less time available for taking a decision, the more probable is the use of a fast (and somewhat stereotyped) heuristics with respect to rational reasoning.

We shall try to include some of these concepts to model autonomous agents that have the task of processing messages from sources that are not always trustable [28–32].

Our main goal is to develop a simple model of the diffusion of hoaxes, gossips, and fake news to be compared with data coming out from a real experiment which is undergoing in this period [33].

### 1.3. Modeling humans and recommender systems interplay

We model here a society of agents who exchange messages through a recommender system. Agents do not corresponds to individuals, but rather to homogeneous communities, grouped together following the representativeness heuristics.

A recommender system can be seen as a device to extract hidden information from a database (also denoted knowledge network or knowledge graph) [34–36].

Let us assume that we can represent a good, e.g., a book or a movie, by a vector of characteristics, expressing for instance the genre, the author/director, actors/characters, and so on. Customers are also represented as complementary (dual) vectors of preferences, corresponding to goods’ characteristics.

The opinion of a customer about a good is assumed to be given by the scalar product of the corresponding vectors (with some nonlinear threshold), i.e., essentially a perceptron model.

The similarity between two users is given by the overlap of their preferences, and that of two goods is given by the overlap between their characteristics.

The goal of a recommender system is that of recommending goods that are expected to be appreciated by a customer (but more on it in the following).

It can be shown that if a matchmaker knows an large enough fraction of the overlaps between customers, above a “rigidity percolation” threshold, then it can anticipate, from this database, the unknown overlaps [34].

However, the direct overlap between customers is not available, but a recommender system may have access to the (partial) database of opinions expressed by customers on goods, either by their direct evaluation, or indirectly by their behavior. For instance, in a social network, the appreciation of customers with respect to received messages can be evaluated by their propensity in forwarding them.

It can be shown that the opinion database can be used as a proxy for the overlap one [37], also in the case of slight nonlinearities in the matching functions [38,39].

The effects of a recommender system is also that of promoting the formation of filter bubbles and artificial communities [40], which here corresponds to agents.

Let us apply a similar model to the case of a social network confronted with the diffusion of fake news. We assume that agents can emit messages, that can be fake or not. A message is successful (and the emitter gains in reputation) if it is forwarded, according with its perceived importance.

We also assume that the message importance is proportional to the probability of being fake. The extreme cases are that of surely true messages, of limited importance, and breakthrough news that are perceived as extremely important and popular, which however have a high probability of being fake.

An agent is confronted with two opportunities: either trust the sender, accept the message and the risk of passing false information, and process it in a short time, or contact a central database (for instance looking for confirmation on the Internet), be sure of the correctness of the message but also waste more time in elaborating it.

Since in social networks news quickly lose their importance, we assume that the fact of checking impedes or makes useless the possibility of forwarding the message, even in the case of a true one.

The social rewarding is also given by the affinity, i.e., one of the components of the propensity of accepting and forwarding a message from a sender, the other ingredient being the risk perception, which models some aspects of human heuristics.

The affinity is increased whenever a message is checked and reveals to be true, and lowered whenever the message is checked and reveals to be fake.

We finally assume that the propensity of emitting fake messages is inversely proportional to one's risk perception, and that the affinity network is available to the recommender system (by the analysis of forwarded and checked messages), which constitutes also a simplification of the simulation structure. The problem is similar to that of detecting forged information in automatic trading [41].

The paper is structured as follows. In Section 2, we introduce the model. Numerical results are presented in Section 3. Suggestions for incorporating human cognitive mechanisms into modern recommender systems are reported in Section 4. A discussion of these results and conclusions are drawn in Section 5.

## 2. The Computer Model

We consider a scenario with  $N$  agents, identified by the index  $i = 1, \dots, N$ . Each agent is characterized by a risk perception factor  $R_i \in [0, 1]$ , between zero and one.

There is also an affinity network  $A_{ij} \in [-1, 1]$ , which characterizes also the "trustability" of agent  $j$  by agent  $i$  and finally determines the reputation of agents, as illustrated below.

At each time step an agent  $i$  is selected at random. He/she emits a message with importance  $I$ , equal to  $1 - R_i$ , but this message can be fake with probability  $IF$ ,  $F$  (fake propensity) being one of the parameters of the system.

The rationale is that agents with larger levels of risk perception (and, as we shall see, with more propensity in checking suspicious messages) are assumed to be more reluctant in emitting "fake" messages (they can however do it, since messages can be fake for not having carefully checked the news). However, we also assume that the larger the importance of messages, the higher their probability of being fake.

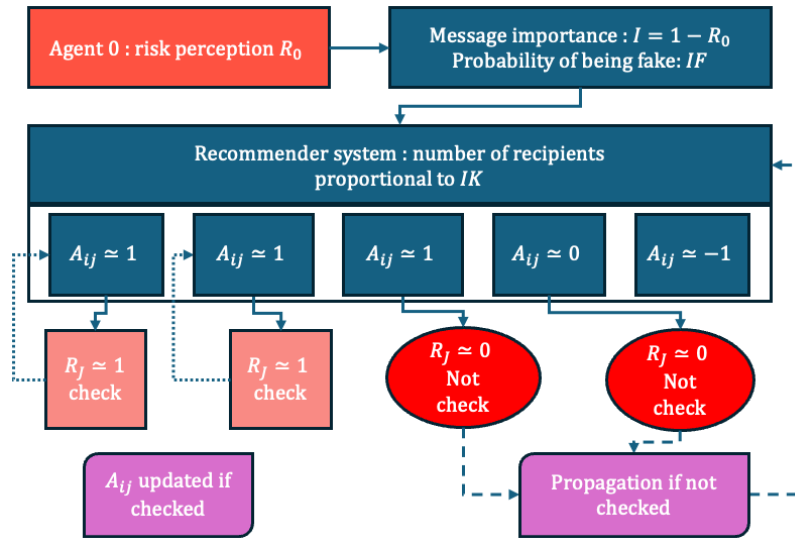
We assume that the recommender system is able to estimate the possible importance of a message by its content (not modeled here).

The message is propagated by the recommender system to up to  $IK$  of other agents,  $K$  being another parameter of the system. Therefore, the more important a message, the larger the number of recipients. We choose a limited amount of possible recipients since otherwise the recommender system would act like a spammer.

The choice of recipients is based on the affinity: an agent  $j$  is selected for receiving the message if he/she has not already received it, and with a probability  $p_{ji} = (1 + A_{ji})/2$ . Therefore, an agent is more likely to receive messages from those with which he/she has a greater affinity.

Receiver  $j$  has the option of checking the message coming from  $i$  or not. If he/she chooses to check and the message reveals fake, the affinity  $A_{ji}$  is decreased by an amount  $r$ , up to a minimum of  $-1$ , if the message reveals true  $A_{ji}$  is increased by an amount  $r$ , up to a maximum of  $1$ . In case of checking, the message is not forwarded.

On the contrary, if the message is not checked, it is forwarded with the same modality. See Figure 1 for a general scheme of the model.



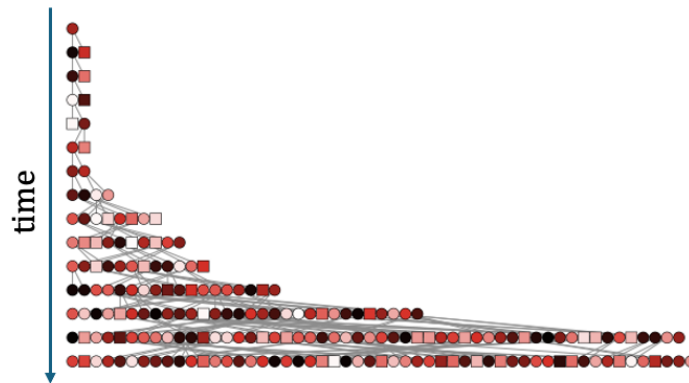
**Figure 1.** The schematic working diagram of the model.

The reputation  $W_i$  of the agent  $i$  at time  $t$  is determined by the sum of its affinity with others,

$$W_i = \sum_j A_{ji}.$$

We included the human heuristics in this simple model by means of risk perception (representativeness) and affinity (availability).

A typical evolution of the resulting communication network among agents is shown in Figure 2.



**Figure 2.** The time plot of the recruitment of agents for  $N = 200$ ,  $K = 3$ ,  $F = 1$ . Circles represent not-checking agents and squares checking ones. Darker shades of red represent smaller values of risk perception  $R$  and lighter ones larger values of  $R$ .

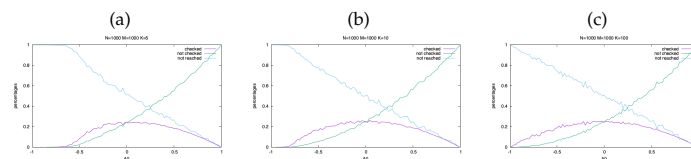
One can see an initial bottleneck, which can bring to the stopping of propagation or is followed by a rapid increase in the number of reached people.

### 3. Simulation Results

Let us first examine the effect of checking the message on a variant of the model, not considering the importance of message nor the affinity, considered constant and equal to  $A_0$ . One agent is randomly chosen to emit a message, which reaches  $K$  other agents. Each of them chooses to forward (without checking) the message to other  $K$  agents (not previously exposed to the message and not chosen by others) with probability  $(1 + A_0)/2$ , or to check it (without forwarding).

As can be seen from Figure 3, the behavior of the fraction of checked messages (whose sum equal to one) is not monotone with  $A_0$ . If  $A_0$  is small, almost all initial messages are checked and not

forwarded, and most of agents are not reached. If  $A_0$  is large, most of messaged are not checked, so many agent are reached, but fake news are not detected. Therefore, the detection of fake messages works better for intermediate values of the affinity.



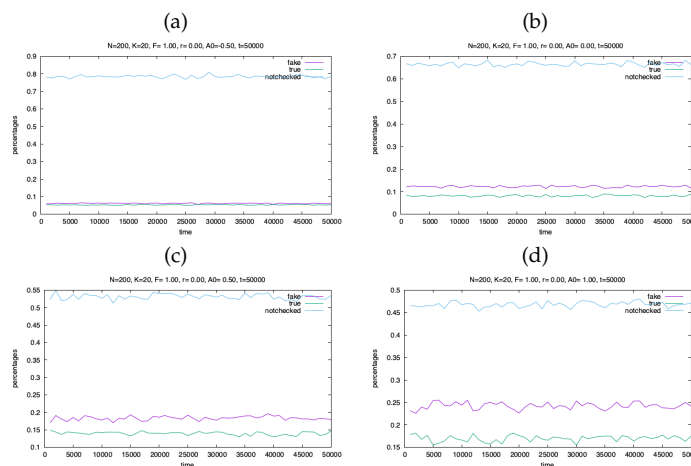
**Figure 3.** The fraction of checked, not checked and not reached agents for the simplified model as a function of the affinity  $A_0$ , for  $N = 1000$ , 100 repetitions and (a)  $K = 5$ , (b)  $K = 10$ , (c)  $K = 100$ . The influence of smaller values of  $K$  is that of preventing the propagation for small values of  $a$ .

A consequent nontrivial effect of the model is that there is a tendency of revealing more fake messages than true ones, even when the probability of generating them is the same. Let us consider again the case in which the affinity is constant ( $A_{ij} = A_0$ ) and not evolving ( $r = 0$ ), but inserting the checking mechanism in the model.

Agents with high values of the risk perception  $R_i$  emits messages that are probably true, but with less importance ( $I = 1 - R_i$ ), while people with low values of  $R_i$  emit more “important” messages, which are probably fake. The recommender system forwards more important messages to a larger number of people (the number of recipients being  $IK$ ).

When all recipients check the message, the propagation chain stops, and this happens with higher probability if the number of recipients is small.

As shown in Figure 4 (with  $F = 1$  the probability of a fake message is 50%), this discrepancy is large for large values of  $A_0$ , and smaller for intermediate values of  $A_0$ . For  $A_0 \simeq -1$  all messages are checked and do not diffuse.

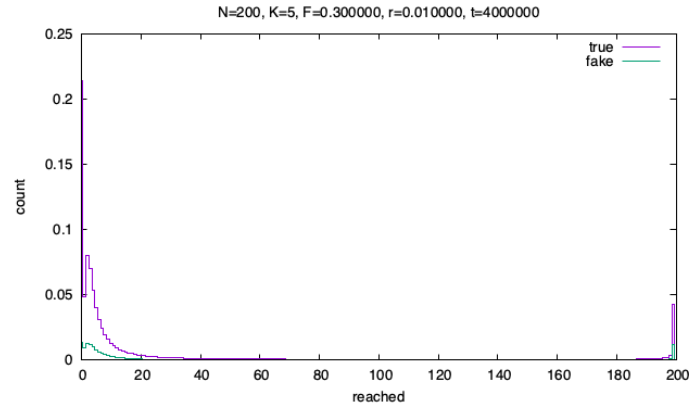


**Figure 4.** The fraction of checked, not checked and not reached agents as a function of time, for  $N = 200$ ,  $r = 0$ ,  $F = 1$  and (a)  $A_0 = -0.5$ , (b)  $A_0 = 0.0$ , (c)  $A_0 = 0.5$ , (d)  $A_0 = 1.0$

We started all subsequent simulation with zero affinity ( $A_0 = 0$ ) and risk perception uniformly distributed between 0 and 1.

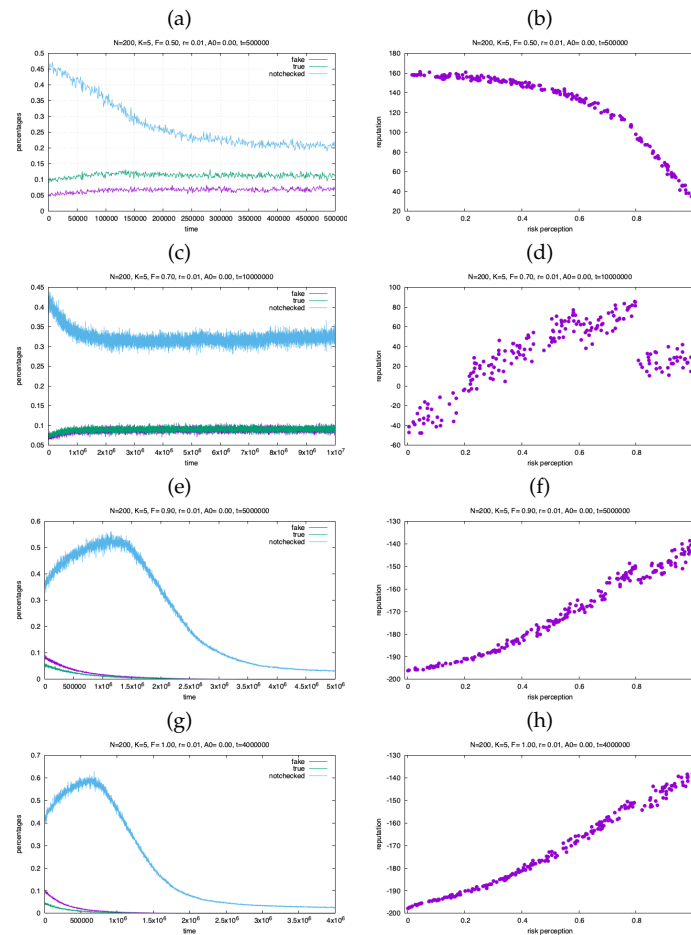
As we shall show, the system can exhibit extremely slow dynamics, and therefore we limited our investigation to a population of 200 agents, and an affinity change rate  $r = 0.01$  (so that it takes roughly 100 coherent steps for affinity to saturate).

In general (see Figure 2), messages either reach the whole population, or stop after a few steps, because they are checked. A typical distribution of reached people is shown in Figure 5.

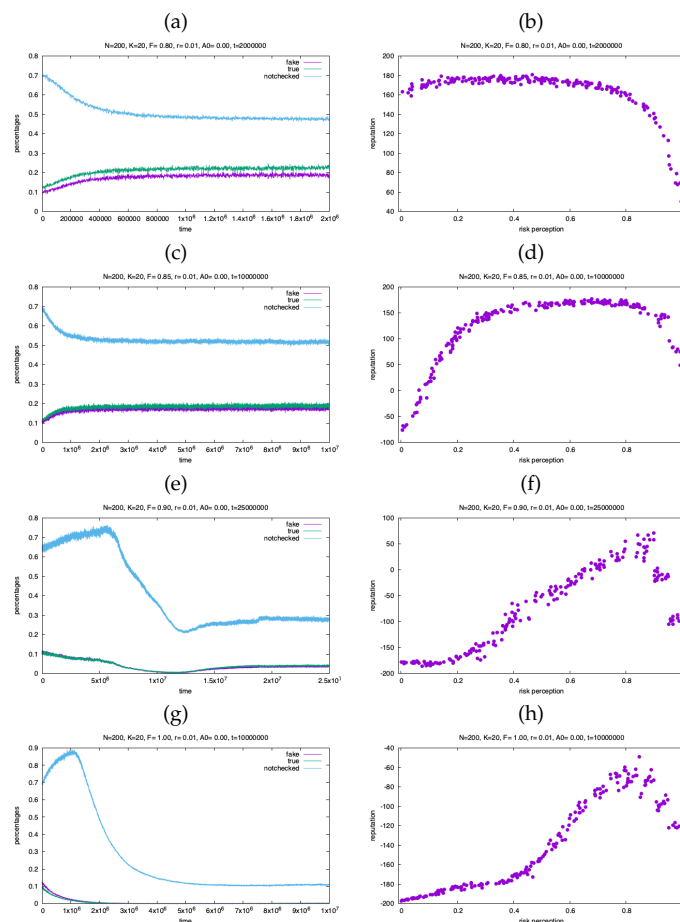


**Figure 5.** Distribution of reached people for  $N = 200$ ,  $K = 5$ ,  $F = 0.3$ ,  $r = 0.01$  and  $T = 4 \cdot 10^6$ . The plot does not change substantially by varying parameters.

The main results are reported in Figure 6 for  $K = 5$  and Figure 7 for  $K = 20$



**Figure 6.** Fraction of checked true, checked false and not checked (left column) and reputation vs risk perception (right column) for  $N = 200$ ,  $K = 5$ ,  $r = 0.01$  and (a-b)  $F = 0.50$ ,  $T = 5 \cdot 10^5$ ; (c-d)  $F = 0.70$ ,  $T = 5 \cdot 10^7$ ; (e-f)  $F = 0.90$ ,  $T = 5 \cdot 10^6$ ; (g-h)  $F = 1.00$ ,  $T = 4 \cdot 10^6$ .



**Figure 7.** Fraction of checked true, checked false and not checked (left column) and reputation vs risk perception (right column) for  $N = 200, K = 20, r = 0.01$  and (a-b)  $F = 0.80, T = 5 \cdot 10^5$ ; (c-d)  $F = 0.85, T = 1 \cdot 10^7$ ; (e-f)  $F = 0.90, T = 2.5 \cdot 10^7$ ; (g-h)  $F = 1.0, T = 1.00 \cdot 10^7$ .

One can see two interesting effects: for large values of the fake factor  $F$  the reputation is larger for agents with higher values of the risk perception, the opposite happens when  $F$  is low. There is also a substantial decrease of the number of reached people for very large times and high  $F$  values. This is consistent with the results of the simplified model of Figure 3: in this case the large number of fake messages lowers the average affinity (which takes negative values) and this implies that most of messages are checked and thus stopped after a few steps.

We see also a kind of transition in the reputation-risk perception dependence, with varying values of the fake factor  $F$ . At transition, which occurs at about  $F = 0.70$  for  $N = 200$  and  $K = 5$ , and at about  $F = 0.85$  for  $N = 200$  and  $K = 20$ , the reputation first grows with risk perception and then decreases. Above this transition the percentage of reached people first increase and then drops to small values, with extremely long transients.

#### 4. Discussion: Bridging Modern Recommender Systems and Cognitive Modeling

Among modern sequential recommender systems, Self-Attentive Sequential Recommendation (SASRec) represents a state-of-the-art architecture based on self-attention mechanisms, capable of capturing both short- and long-range dependencies in user interaction sequences [42]. SASRec consistently outperforms Markovian, convolutional (CNN), and recurrent (RNN) approaches, primarily by adaptively weighting past interactions. Despite its strong empirical performance across datasets of varying sparsity, SASRec optimizes prediction accuracy under the implicit assumption that past user interactions constitute reliable signals of relevance.

This assumption is shared by most contemporary recommender systems; however, such formulations do not explicitly account for cognitive costs, risk perception, or epistemic uncertainty, nor for the downstream societal effects of recommendation-driven information propagation. Our simulation framework complements this line of work by explicitly modeling how cognitive heuristics, trust, and risk perception shape user behavior and information diffusion.

From this perspective, systems such as SASRec can be interpreted as operating in a regime where behavioral regularities are leveraged without direct consideration of epistemic validity. Before proceeding, we emphasize that the following are not part of the simulation described in the previous sections. Rather, they represent a conceptual implementation of the simulation results into modern recommender system architectures.

Based on our results, at least three fundamental architectural extensions can be envisioned. The first concerns the attention formalization used in SASRec (Eq. (2) of Ref [42]):

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V, \quad (1)$$

where  $A$  models the attention. Adding two parameters,  $R$ , that represents specific user or item risk perception and  $\lambda$ , the cognitive cost weight, results in a risk-aware attention algorithm:

$$A_r(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}} - \lambda R\right)V,$$

where  $A_r$  is the attention modulated by risk.

Furthermore, trust-modulated attention to interactions associated with low trust, tackles the conceptual issue that past actions are not equally reliable. This can be achieved by additive logit bias, a very common weighting method in attention variants:

$$\alpha_{t,j} = \frac{\exp\left(\frac{q_t^\top k_j}{\sqrt{d}} + \gamma \tau_{t,j}\right)}{\sum_{j' \leq t} \exp\left(\frac{q_t^\top k_{j'}}{\sqrt{d}} + \gamma \tau_{t,j'}\right)}, \quad s_t = \sum_{j \leq t} \alpha_{t,j} v_j.$$

Such formulation introduces  $\tau_{t,j'}$  as a trust score and  $\gamma$  which controls the strength of the signal. Lastly, in order to penalize the propagation of high-risk or unverified content, the optimization function can be modified to a dual-objective training which considers accuracy and epistemic safety simultaneously:

$$\mathcal{L} = \mathcal{L}_{\text{rank}} + \alpha \mathcal{L}_{\text{risk}}.$$

Taken together, these extensions illustrate how state-of-the-art sequential recommender systems can be conceptually augmented to account for cognitive and epistemic dimensions of user behavior. Importantly, the proposed modifications do not aim to replace existing architectures such as SASRec, but rather to highlight the principal directions through which attention-based recommenders may be aligned with models of human decision-making. By integrating risk perception, trust, and epistemic cost into the recommendation process, such systems may better capture the trade-offs between engagement, accuracy, and information reliability observed in real-world social and informational ecosystems.

Note that Eq. (1) is the scaled dot-product attention; this is not specific to recommender systems. It is exactly the same attention operator introduced in Ref. [43]. Eq. (1) constructs a data-dependent graph and performs one step of message passing on it. Large Language Models (GPT, LLaMA, etc.) use exactly the same operator with a difference scale (billions of parameters and many stacked layers). In other words, LLMs are nothing more than very deep stacks of Eq. (1) with nonlinearities and residuals.

As SASRec can be viewed as a generalization of Markov chains, LLMs generalize n-gram language models in exactly the same way. The addition of the two parameters changes the semantics of attention.  $R$  is a cost / risk matrix, aligned with attention scores while  $\lambda$  controls the trade-off between cognitive cost and relevance. Therefore, adding  $R$  and  $\lambda$ , turns attention from a pure similarity operator into a rational decision mechanism that trades relevance against risk or cost. By setting

$$\text{Attention} = \text{relevance} + \text{trust} - \text{risk},$$

the novelty of our work is that attention becomes a decision mechanism that balances relevance, trust, and risk – not just similarity.

## 5. Conclusions

We have presented a simple model that includes some aspects of human heuristics (risk perception and affinity) coupled to a recommender systems (which is able to “sense” the level of affinity to forward messages to specific agents). The model aims at reproducing some aspects of the diffusion of fake news, and agents face the dilemma of checking the received message (according with their risk perception and affinity with the sender), losing time and thus not forwarding the message within an “importance” timing window, or resubmit it to the recommender system.

We assume that the probability that a message is fake depends inversely on the risk perception of the sender.

The system shows unexpected complexity: people with more propensity of checking determines the diffusion of messages, so that when the affinity is low, all messages (including fake ones) do not propagate, but when affinity is large fake messages are also not easily discovered. However, the affinity evolves in time according with the discovered “truthiness” of messages.

The final reputation of agents is based on their total affinity with others. The main result of the model is that the relation between reputation and the “fake factor”  $F$  that determines (together with the risk perception of the emitter) the truthiness of messages and reputation is not always the same.

When  $F$  is small, less messages are fake, and agents with low risk perception can increase their reputation since the messages that they forward without checking are probably true.

On the contrary, when  $F$  is large, agents with high level of risk perception, while forwarding few messages (due to their propensity towards checking), they get relative higher values of reputation. In this regime however most of messages are checked, so when this patterns establishes there is a decrease in the number of reached people.

This simple model can illustrate two tendencies seen in real recommender system: their reluctance in stopping or discouraging fake news and the fact that in many cases there is a strong correlation between popularity (or reputation) and propensity towards the spreading of fake news.

**Author Contributions:** Conceptualization, F.B., T.J.C, A.G., P.L. and R.R.; methodology, F.B., T.J.C, A.G. and P.L.; software, F.B.; validation, T.J.C, A.G., P.L. and R.R.; formal analysis, F.B.; investigation, F.B.; writing—original draft preparation, F.B.; writing—review and editing, F.B., T.J.C, A.G., P.L. and R.R.; visualization, F.B.; supervision, F.B.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** “Not applicable”

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** No data used. The software code is available upon request.

**Acknowledgments:** The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* **1974**, *185*, 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>.
2. Arima, Y. *Psychology of Group and Collective Intelligence*; Springer International Publishing, 2021. <https://doi.org/10.1007/978-3-030-84698-5>.
3. Petrofsky, J. Facilitating Knowledge Sharing. M.f.a., Indiana University, 2017.
4. Feld, S.L.; Webster, M.; Foschi, M. Status Generalization: New Theory and Research. *Social Forces* **1990**, *68*, 948. <https://doi.org/10.2307/2579371>.
5. Areeb, Q.M.; Nadeem, M.; Sohail, S.S.; Imam, R.; Doctor, F.; Himeur, Y.; Hussain, A.; Amira, A. Filter bubbles in recommender systems: Fact or fallacy—A systematic review. *WIREs Data Mining and Knowledge Discovery* **2023**, *13*. <https://doi.org/10.1002/widm.1512>.
6. Alatawi, F.; Cheng, L.; Tahir, A.; Karami, M.; Jiang, B.; Black, T.; Liu, H. A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation, 2021. <https://doi.org/10.48550/ARXIV.2112.05084>.
7. Wang, Z.; Zhu, H.; Sun, L. Social Engineering in Cybersecurity: Effect Mechanisms, Human Vulnerabilities and Attack Methods. *IEEE Access* **2021**, *9*, 11895–11910. <https://doi.org/10.1109/access.2021.3051633>.
8. Barak, A., Phantom emotions. In *Oxford handbook of Internet psychology*; Oxford University Press, USA, 2007; p. 303.
9. Simon, H.A., Bounded Rationality. In *Utility and Probability*; Palgrave Macmillan UK, 1990; p. 15–18. [https://doi.org/10.1007/978-1-349-20568-4\\_5](https://doi.org/10.1007/978-1-349-20568-4_5).
10. Dunbar, R.I.M. The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews* **1998**, *6*, 178–190. [https://doi.org/10.1002/\(sici\)1520-6505\(1998\)6:5<178::aid-evan5>3.0.co;2-8](https://doi.org/10.1002/(sici)1520-6505(1998)6:5<178::aid-evan5>3.0.co;2-8).
11. Festinger, L. Informal social communication. *Psychological Review* **1950**, *57*, 271–282. <https://doi.org/10.1037/h0056932>.
12. Wolpert, D.H.; Tumer, K. An Introduction to Collective Intelligence, 1999. <https://doi.org/10.48550/ARXIV.CS/9908014>.
13. Leimeister, J.M. Collective Intelligence. *Business & Information Systems Engineering* **2010**, *2*, 245–248. <https://doi.org/10.1007/s12599-010-0114-8>.
14. Nitecki, M.H.; Nitecki, D.V.e. *Origins of Anatomically Modern Humans*; Springer US, 1994. <https://doi.org/10.1007/978-1-4899-1507-8>.
15. Bicchieri, C.; Muldoon, R.; Sontuoso, A., Social norms. In *The Stanford Encyclopedia of Philosophy (Winter 2023 Edition)*; Zalta, E.N.; Nodelman, U., Eds.; 2011.
16. Dores Cruz, T.D.; Thielmann, I.; Columbus, S.; Molho, C.; Wu, J.; Righetti, F.; de Vries, R.E.; Koutsoumpis, A.; van Lange, P.A.M.; Beersma, B.; et al. Gossip and reputation in everyday life. *Philosophical Transactions of the Royal Society B* **2021**, *376*. <https://doi.org/10.1098/rstb.2020.0301>.
17. Dunbar, R.I.M. Gossip in Evolutionary Perspective. *Review of General Psychology* **2004**, *8*, 100–110. <https://doi.org/10.1037/1089-2680.8.2.100>.
18. Dunbar, R. *Grooming, gossip, and the evolution of language*; Harvard University Press: London, England, 2020.
19. Gigerenzer, G.; Todd, P.M.; ABC Research Group. *Simple heuristics that make us smart*; Evolution and Cognition, Oxford University Press: New York, NY, 2000.
20. Tversky, A.; Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* **1973**, *5*, 207–232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).
21. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **1979**, *47*, 363–391.
22. Tversky, A.; Kahneman, D. The Framing of Decisions and the Psychology of Choice. *Science* **1981**, *211*, 453–458. <https://doi.org/10.1126/science.7455683>.
23. Tversky, A.; Kahneman, D.; Slovic, P. *Judgment under uncertainty: Heuristics and biases*; Cambridge, 1982.
24. McNeil, B.J.; Pauker, S.G.; Sox, H.C.; Tversky, A. On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine* **1982**, *306*, 1259–1262. <https://doi.org/10.1056/nejm198205273062103>.
25. Taylor, S.E. The availability bias in social perception and interaction. In *Judgment under Uncertainty*; Tversky, A.; Kahneman, D.; Slovic, P., Eds.; Cambridge University Press, 1982; p. 190–200. <https://doi.org/10.1017/cbo9780511809477.014>.
26. Slovic, P.; Fischhoff, B.; Lichtenstein, S., Facts versus fears: Understanding perceived risk. In *Judgment under Uncertainty*; Cambridge University Press, 1982; p. 463–490. <https://doi.org/10.1017/cbo9780511809477.034>.

27. Christensen, C.; Abbott, A.S. Team Medical Decision Making. In *Decision making in health care: Theory, psychology, and applications*; Chapman, G.; Sonnenberg, F., Eds.; Cambridge University Press: New York, NY, 2003; Vol. 267, chapter 10, pp. 267–284.
28. Weiss, G. *Multiagent systems*; Intelligent Robotics & Autonomous Agents Series, MIT Press: London, England, 2000.
29. Wooldridge, M. *An Introduction to MultiAgent Systems*, 2 ed.; John Wiley & Sons: Chichester, England, 2009.
30. d’Inverno, M.; Luck, M. *Understanding Agent Systems*, 2 ed.; Springer Series on Agent Technology, Springer: Berlin, Germany, 2003.
31. Sun, R., Ed. *Cognition and multi-agent interaction*; Cambridge University Press: Cambridge, England, 2008.
32. Merelli, E.; Armano, G.; Cannata, N.; Corradini, F.; d’Inverno, M.; Doms, A.; Lord, P.; Martin, A.; Milanese, L.; Moller, S.; et al. Agents in bioinformatics, computational and systems biology. *Briefings in Bioinformatics* **2006**, *8*, 45–59. <https://doi.org/10.1093/bib/bbl014>.
33. Dini, S.; Guazzini, A.; Cvetkovic, T.J.; Bagnoli, F. Crowdsourced Fact Checking. In preparation, 2026.
34. Maslov, S.; Zhang, Y.C. Extracting Hidden Information from Knowledge Networks. *Physical Review Letters* **2001**, *87*. <https://doi.org/10.1103/physrevlett.87.248701>.
35. Lü, L.; Medo, M.; Yeung, C.H.; Zhang, Y.C.; Zhang, Z.K.; Zhou, T. Recommender systems. *Physics Reports* **2012**, *519*, 1–49. <https://doi.org/10.1016/j.physrep.2012.02.006>.
36. Zhang, J.C.; Zain, A.M.; Zhou, K.Q.; Chen, X.; Zhang, R.M. A review of recommender systems based on knowledge graph embedding. *Expert Systems with Applications* **2024**, *250*, 123876. <https://doi.org/10.1016/j.eswa.2024.123876>.
37. Bagnoli, F.; Berrones, A.; Franci, F. De gustibus disputandum (forecasting opinions by knowledge networks). *Physica A: Statistical Mechanics and its Applications* **2004**, *332*, 509–518. <https://doi.org/10.1016/j.physa.2003.09.065>.
38. Di Patti, F.; Bagnoli, F., Biologically Inspired Classifier. In *Bio-Inspired Computing and Communication*; Springer Berlin Heidelberg, 2008; p. 332–339. [https://doi.org/10.1007/978-3-540-92191-2\\_29](https://doi.org/10.1007/978-3-540-92191-2_29).
39. Nguyen, V.A.; Koukoliková-Nicola, Z.; Bagnoli, F.; Lió, P. Noise and non-linearities in high-throughput data. *Journal of Statistical Mechanics: Theory and Experiment* **2009**, *2009*, P01014. <https://doi.org/10.1088/1742-5468/2009/01/p01014>.
40. Bagnoli, F.; de Bonfioli Cavalcabo’, G.; Casu, B.; Guazzini, A. Community Formation as a Byproduct of a Recommendation System: A Simulation Model for Bubble Formation in Social Media. *Future Internet* **2021**, *13*, 296. <https://doi.org/10.3390/fi13110296>.
41. Bagnoli, F.; Guazzini, A.; Lió, P., Human Heuristics for Autonomous Agents. In *Bio-Inspired Computing and Communication*; Springer Berlin Heidelberg, 2008; p. 340–351. [https://doi.org/10.1007/978-3-540-92191-2\\_30](https://doi.org/10.1007/978-3-540-92191-2_30).
42. Kang, W.C.; McAuley, J. Self-Attentive Sequential Recommendation, 2018, [[arXiv:cs.IR/1808.09781](https://arxiv.org/abs/1808.09781)].
43. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the Advances in neural information processing systems*, 2017, pp. 5998–6008.

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