A Bayesian Model of Confirmatory Exploration in Text-based Web Media

Yosuke Fukuchi (fukuchi@tmu.ac.jp)
Faculty of Systems Design, Tokyo Metropolitan University, 6-6, Asahigaoka, Hino-shi, Tokyo 191-0065, Japan

Abstract

As web media, such as social networking services (SNS), become more prevalent, the formation of false beliefs through fake news and propaganda has become a significant problem. This study focuses on the cognitive process of users as actively information-seeking agents in web media exploration and proposes WEB-FEP, a computational model of users forming specific beliefs through interactions with web media. WEB-FEP specifically attempts to computationally reproduce confirmation bias in web media exploration by formalizing the trade-off between belief-confirmatory and exploratory actions inspired by active inference. WEB-FEP is validated by comparing the results of simulations with user experiments conducted on a virtual SNS. The results indicate that the initial belief distributions and learning rates modeled in WEB-FEP can successfully reproduce the diverse behaviors of users including confirmatory exploration.

Keywords: web-media exploration; belief polarization; confirmation bias; Bayesian modeling; active inference; free-energy principle; social network service; large language model

Introduction

The ultimate goal of this study is to construct a cognitive model that simulates the process of belief formation by exploring web media such as social networking services (SNS) on the basis of actual text content. The web is an immense repository of information, and when used effectively, it allows individuals to gain a wide range of insights. Particularly, with the rise of SNS, users are increasingly provided with opportunities to access differing perspectives and opinions. However, recent attention has been drawn to their negative aspects, where users may form incorrect beliefs due to biased information found on the web. Here, a computational model of the user cognitive processes behind the formation of these false beliefs has the potential to enable human-web media interactions to be simulated. This could contribute to predicting how media design influences users' belief formation and behavior and to creating healthier web media that reduce the possibility of users becoming entrenched in false beliefs.

Confirmation bias is a factor that distorts users' web exploration. It refers to the tendency to preferentially seek information that aligns with one's beliefs while avoiding or ignoring information that contradicts them (Nickerson, 1998). Tanaka et al. conducted an experiment using a fact-checking website, reporting that, while approximately half of the participants actively sought information that contradicted their beliefs, the remaining participants avoided such information, resulting in a tendency to maintain incorrect beliefs (Tanaka et al., 2023). That is, confirmation bias can cause users to overly acquire information that strengthens their existing beliefs, resulting in missing opportunities to obtain information that contradicts those false beliefs. Therefore, this paper focuses on confirmation bias while forming belief through web exploration.

Various attempts have been made to model confirmation bias within the framework of Bayesian inference. For example, Pilgrim et al. demonstrated through simulations that confirmation bias could be explained as an approximate Bayesian inference under limited cognitive resources (Pilgrim, Sanborn, Malthouse, & Hills, 2024). Chattoraj et al. demonstrated that active reasoning based on existing beliefs induces confirmation bias during visual exploration (Chattoraj, Shivkumar, Ra, & Häfner, 2021).

Previous research has demonstrated the effectiveness of using Bayesian inference to model the cognitive basis of confirmation bias, particularly in visual exploration tasks. However, a limited number of studies have focused on its application to web media exploration which primarily involves linguistic information.

This paper proposes WEB-FEP, a cognitive model of confirmatory exploration behavior in SNS environments. WEB-FEP is based on Bayesian inference and especially inspired by the Free Energy Principle (FEP) (Friston, 2010), which is a theoretical framework that explains how agents reduce uncertainty and select actions by minimizing the gap between observations and their beliefs. WEB-FEP formalizes the choices of actions that align with their beliefs or actions that seek new information to expand their beliefs as a trade-off between pragmatic and epistemic values in FEP. By implementing the model in the language embedding space of a large language model (LLM), WEB-FEP aims to simulate web exploration behavior based on linguistic content.

To validate WEB-FEP, we first conducted a user study using our virtual news portal platform and compared user behaviors with the results of a simulation with WEB-FEP. The user study involved participants exploring the platform and selecting links (hashtags) to view the posts. Here, confirmatory behavior was defined as more selection of hashtags that aligned with the participants' prior beliefs. The results showed that the human participants exhibited diverse behav-

iors, including confirmatory exploration on the website, and the simulation results demonstrated that the model could successfully reproduce the diversity of human behaviors by controlling the initial belief distribution and learning rate.

Background

User Behavior Modeling on SNS

Social networking services (SNS) have become a major platform for information dissemination and opinion formation. However, the information environment on SNS is often a mix of both correct and incorrect information, so users may be exposed to fake news and propaganda, leading to the formation of false beliefs. It is an urgent issue to understand how users form beliefs through interactions with information on SNS.

One approach to addressing this problem is through the design of SNS platforms. For instance, recommendation algorithms can bias the information to which users are exposed (Mansoury, Mobasher, & van Hoof, 2024). Connections between users can create echo chambers, where users are predominantly exposed to information that aligns with their pre-existing beliefs, further strengthening false beliefs (Cinelli, Morales, Galeazzi, Quattrociocchi, & Starnini, 2021).

This paper, however, focuses on the user side, modeling the cognitive processes of users' active information-seeking in web media. Previous studies have shown that a certain part of users exhibit confirmation bias in web exploration, where they tend to seek information that aligns with their beliefs while avoiding contradictory information (Tanaka et al., 2023). This bias can cause users to overly acquire information that strengthens their existing beliefs, resulting in missing opportunities to obtain information that contradicts those false beliefs.

Free Energy Principle

FEP is a theoretical framework that explains how an agent copes with the world's uncertainty, influencing perception, learning, and decision-making (Friston, 2010). In FEP, expected free energy is interpreted to capture the trade-off between exploration, the process of seeking new information to reduce uncertainty, and exploitation, the process of using existing knowledge to make decisions. Here, epistemic value $V_{\text{epist}}(a)$ quantifies the value of exploration, which represents the gain in information that the agent can obtain by selecting action a:

$$V_{\text{epist}}(a) = \mathbb{E}_{Q(s',o'|a)}\left[D_{\text{KL}}[Q(s'|o')\|Q(s')]\right], \tag{1}$$

where $Q(s',o'|a)$ is the agent's belief about the world after selecting action a, $Q(s'|o')$ is a posterior belief about the world after observing o', and $Q(s')$ is a prior belief before observing o'.

On the other hand, pragmatic value $V_{\text{prag}}(a)$ quantifies the value of exploitation. This represents the expected alignment between future observations and the agent's objective C:

$$V_{\text{prag}}(a) = \mathbb{E}_{Q(o'|a)}[\ln P(o'|C)], \tag{2}$$

where $Q(o'|s')$ is the agent's posterior belief about future observations given the state s', and $P(o'|C)$ represents the likelihood of observing o' assuming the agent's objective C is achieved. An agent aims to maximize the total value $V(a) = V_{\text{prag}}(a) + V_{\text{epist}}(a)$ by selecting an action.

$$a = \arg\max_a V_{\text{epist}}(a) + V_{\text{prag}}(a). \tag{3}$$

This process refers to balancing the trade-off between the two values and characterizes *active inference*, where the agent proactively interacts with the environment through its actions, controlling observations to obtain the most informative data in an efficient manner. Chattoraj et al. demonstrated that the active inference framework effectively explains confirmation bias in a visual task (Chattoraj et al., 2021), and this paper aims to extend this framework to web media exploration.

WEB-FEP model

Formulation

WEB-FEP is a cognitive model that aims to describe the behavior and belief formation process of users when exploring web media in terms of active inference in FEP. To formulate the model, we simplify the web media exploration process as repeatedly clicking links and observing text content. Here, user action a is the decision of which link to click, and observation o is the text content associated with the link. More specifically, in our simulation with the virtual SNS, links refer to hashtags, and text content refers to posts associated with the hashtags.

Belief In WEB-FEP, a user's belief is represented as a probability distribution $q(x)$ in the sentence embedding space. Here, $x \in \mathbb{R}^d$ represents a proposition such as a supporting or opposing opinion in web media. In this paper, o and x are equally represented in the embedding space. $q(x)$ represents the probability distribution of how likely the agent considers x:

$$q(x) \propto P(x \text{ is true}|o \in O), \tag{4}$$

where O is the set of historical observations. In this paper, $q(x)$ is modeled as a multidimensional normal distribution centered around a vector g_q and a covariance matrix V.

Belief Update During web exploration, WEB-FEP updates $q(x)$ so that it can explain the content observed in web media. Given the observation o, a belief is updated by minimizing a cross entropy loss function:

$$-\mathbb{E}_{p(o|a)}[\ln q(o)]. \tag{5}$$

With stochastic gradient descent, q is updated so that its density at the observed x increases. Let α be the learning rate, which conceptually mean how much an agent's belief is influenced by o.

Value Functions and Action Selection In WEB-FEP, the user's exploration behavior is formalized as a trade-off between two values based on the FEP:

$$V(a) = V_{\text{epist}}(a) + V_{\text{prag}}(a). \tag{6}$$

Because a is the selection of a hashtag, it can correspond to the representation of the hashtag itself. Let o_a be the embedding vector of the hashtag a.

$V_{\text{epist}}(a)$ represents the informativeness of the observation x obtained by a to the belief distribution q:

$$V_{\text{epist}}(a) = \mathrm{KL}(q_{\text{new}}(o_a) \| q(o_a)), \tag{7}$$

where $q_{\text{new}}(x)$ is a belief distribution updated with equation 5 after observing x. $V_{\text{epist}}(a)$ quantifies the impact of x on the belief distribution, leading the agent to exploratory actions. Although FEP considers the future observation o' after a, meaning that the agent predicts the posts associated with the hashtag in our setting, this paper assumes that o_a is sufficiently equivalent to o'. That is, the representation of hashtag o_a is a sufficient proxy for the posts o' to be observed with a. It is important to note that $V_{\text{epist}}(a)$ is affected by α, which governs the amount of belief update.

$V_{\text{prag}}(a)$ represents the likelihood of o_a given the belief q:

$$V_{\text{prag}}(a) = \ln q(o_a). \tag{8}$$

This is equivalent to the objective function of the belief update (Eqn. 5), except that the sign is reversed because Eqn. 5 is a minimization objective. Because this paper equates o_a with o', $V_{\text{prag}}(a)$ quantifies the fitness of o' obtained by a for q, and maximizing this term leads the agent to confirmatory exploration. An agent selects a to maximize $V(a)$ (Eqn. 3).

User Study on Virtual SNS Media

Aim

The aim of this study was two-fold: (i) to investigate human participants' exploration behavior in our virtual SNS environment and (ii) to acquire data for evaluating WEB-FEP. For aim (i), we specifically investigated (i-a) whether humans actually show confirmatory exploration and (i-b) how their beliefs change through exploration.

Design of Virtual SNS

For the experiment, we prepared a virtual SNS that mimics a news portal site. This virtual SNS consisted of three elements. (a) News Article: At the start of the experiment, a single news article was presented to the user. (b) Hashtags: Each hashtag represented a summary of opinions posted about the news article. Hashtags also functioned as selectable links for users (Figure 1a). (c) Posts: Opinions either supporting or opposing the topic of the news article were displayed as posts associated with the hashtags (Figure 1b).

The news article was selected from Yahoo! News[1], where users can freely exchange their comments on an article. The selected gossip article was about a debate over a meeting between Mrs. Akie Abe, the wife of former Prime Minister Shinzo Abe, and newly elected President Trump. The actual user comments posted on the article covered a wide range of

(a) Hashtag selection

(b) Posts

Figure 1: User interface of virtual SNS

topics beyond the appropriateness of the meeting. The original user comments were too diverse, introducing noise into the experimental results. To control for this, we fed them to GPT-4o[2] and re-generated the comments for the experiment with prompts that restricted the content to the debate over the meeting itself. GPT-4o was also used to generate hashtags for the re-generated comments.

Procedure

The procedure of this study was as follows. (1) A participant viewed the news article and reported their initial opinion score of whether they agreed or disagreed with the topic on a scale of 0 (disagree) to 100 (agree) using a scale bar implemented in HTML (Gift, 1989). (2) Six hashtags related to the post were presented. The participant selected one of them and viewed eight posts associated with the selected hashtag. For each time, three hashtags were selected respectively from a set of comments that were positive and negative about the topic. (3) After viewing the posts, the participant proceeded to the next round and selected another hashtag. This process was repeated for ten rounds. (4) After the final round, they reported their final opinion score on the news article again.

To control the experiment, the same set of hashtags was used among participants in each round, but the order of the hashtags was randomized to avoid bias in the order of the choices. To avoid showing semantically similar hashtags in the same round, we conducted clustering of both positive and negative hashtags. Specifically, we used llm-jp-3-13b-instruct, a large language model (LLM) published by the National Institute of Informatics[3] to acquire sentence embeddings of the hashtags and posts. The embedding vectors from the final layer of the LLM were aggregated into a single vector by applying mean pooling to each token, and 5120-dimensional vectors were acquired for each post and hashtag. Then, we reduced the number of dimensions to two using the UMAP method (McInnes, Healy, Saul, & Großberger, 2018)

[1] https://news.yahoo.co.jp/

[2] https://openai.com/index/hello-gpt-4o/

[3] https://huggingface.co/llm-jp/llm-jp-3-13b-instruct

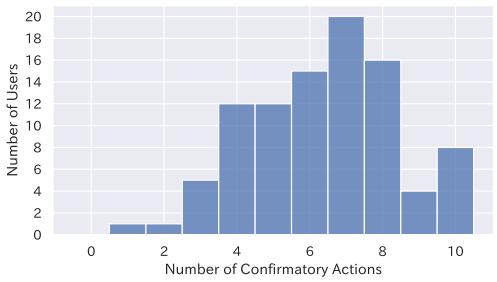

Figure 2: Distribution of number of confirmatory actions

Table 1: Number of participants for each opinion change type

Strengthened	Maintained	Neutral	Reversed
16	47	22	9

and performed k-means clustering. As a result, we obtained three clusters for positive and negative hashtags, respectively. The hashtags were selected from each cluster in a round-robin manner on the basis of the distance between a hashtag and the center of the cluster, and similarly, the posts were selected for each hashtag on the basis of the distance between the two.

The experiment was conducted using Yahoo! Crowdsourcing, and 100 participants were recruited with compensation of 28 JPY (14 females, 85 males, one unspecified; aged 19-65 ($M = 47.7, SD = 10.0$)). Six participants whose initial opinion was exactly neutral (50 points) were excluded from the analysis because we could not determine the direction of the opinion change (strengthened or weakened). As a result, the analyses of the results were conducted on the remaining 94 participants. We did not explicitly instruct the participants to seek facts or determine their opinion through exploration but only to check the posts associated with the news to observe the natural use of an SNS.

Metrics

To analyze participant behavior, we used two metrics. (a) Number of confirmatory actions is a metric that quantifies the extent of confirmation bias in users. In this study, we defined the selection of eight or more hashtags that are on the same side as the initial opinion out of ten rounds as "confirmatory exploration" and two or fewer as "disconfirmatory exploration." (b) Opinion change type is a metric that classifies the user's opinion change pattern. We classified this change pattern into the following four types: (i) strengthened the initial opinion, (ii) maintained the initial opinion, (iii) moved toward neutral, (iv) reversed the opinion. (ii) Maintained the initial opinion was defined as the final opinion being within five points of the initial opinion.

Results

Figure 2 shows the distribution of the number of confirmatory actions. The average number of confirmatory actions was 6.38 ($SD = 2.02$), a little closer to the confirmatory side.

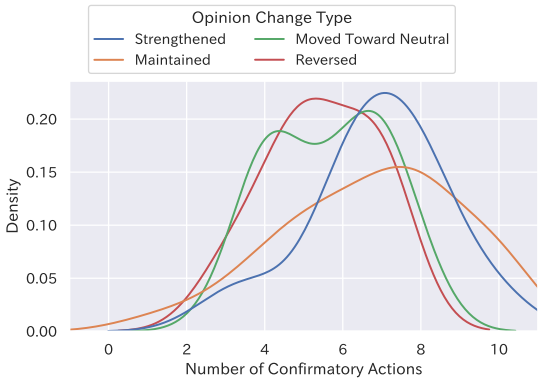

Figure 3: Number of confirmatory actions for each opinion change type

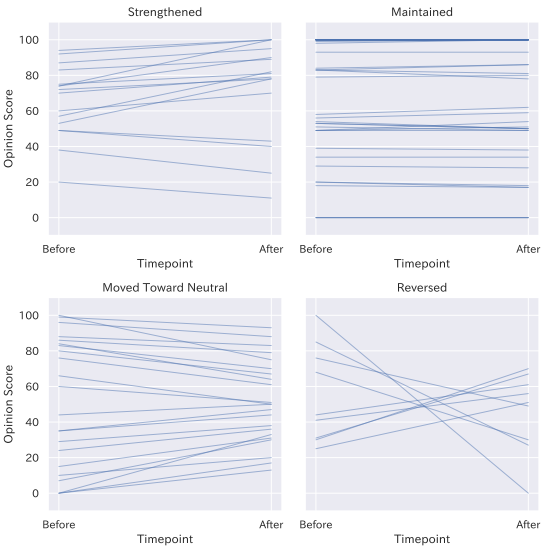

Figure 4: Changes of opinion

28 participants performed confirmatory exploration, 2 performed disconfirmatory exploration, and 64 performed neutral exploration. The results indicate that some actually performed confirmatory exploration on the virtual SNS, though a larger number of users performed more neutral exploration. This aligns with the results of Tanaka et al. (Tanaka et al., 2023), where participants were divided into fact-exposure and fact-avoidance groups. Interestingly, confirmatory and disconfirmatory explorations were asymmetric. Few users actively sought information that contradicted their initial opinion.

Table 1 shows the distribution of opinion change types. The majority maintained their initial opinion. About one-fourth moved toward neutral. 16 participants strengthened their initial opinion. 9 participants reversed their opinion through the exploration.

Figure 3 shows the number of confirmatory actions for each opinion change type. Notably, the participants who

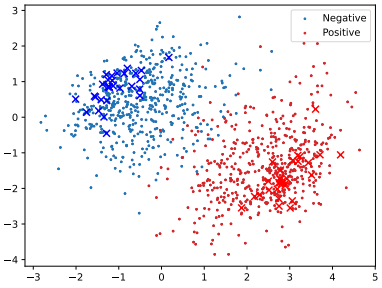

Figure 5: Sentence embedding space. Dots represent posts and crosses represent hashtags.

strengthened their initial opinion performed more confirmatory actions than the others. Those who moved toward neutral and reversed their opinion tended to select hashtags equally on both sides. The participants who maintained their initial opinion also appear to have performed confirmatory actions more than the others. However, we need to carefully discuss the results because they may include participants whose initial opinions were strong and who experienced the ceiling effect, which Figure 4 also indicates. The figure shows the diversity in the change in opinion score.

In conclusion, the results of the user study on a virtual SNS showed that some users performed confirmatory exploration and strengthened their initial belief, while user behavior was diverse, including maintaining the initial opinion and moving toward neutral.

Simulation Experiment

Aim

This experiment aimed to evaluate WEB-FEP's capability to reproduce the behaviors of users in web media exploration. Specifically, we focused on the effects of the initial belief distribution and the learning rate on the model's behavior.

Expected Consequences with Model Parameters

We expected that the model behavior would be predominantly influenced by α and g_q. α controls the extent to which beliefs adapt on the basis of new observations. A higher α increases the influence of $V_{\text{epist}}(a)$ and is expected to lead agents to exploratory behavior, resulting in fewer confirmatory actions. On the other hand, a lower α decreases the influence of $V_{\text{epist}}(a)$, leading agents to retain prior beliefs and thus resulting in confirmatory behavior.

g_q determines the user's prior belief state in the embedding space. Confirmatory behavior is expected to be more likely when g_q is biased toward one side. However, when combined with a larger α, WEB-FEP is expected to exhibit more exploratory behavior because, with the increased influence of $V_{\text{epist}}(a)$, the information of the opposite side appears more informative when g_q is farther from it.

Implementation

For the simulation, we transformed the embedding vectors of the posts and hashtags acquired in the user study with a linear transformation because the original dimensionality of 5,120 was too high for this simple topic. To train this linear transformation, we conducted a crowdsourcing-based study to investigate the pairwise distances of posts. 100 crowdworkers reported the similarity of posts, and we acquired data for 3,600 pairs. With the data, we trained the transformation to maximize the correlation between the distances in the embedding space and the labels reported by the crowdworkers, defined as $1-$similarity. As a result, we obtained a 2D embedding space (Figure 5). Although the number of dimensions was drastically reduced, we considered it to be sufficient for the simulation because the distances between the transformed vectors showed a high correlation with the labels from the crowdworkers (Pearson's $r = .81$).

Procedure

In the simulation, we sampled the initial coordinates of $g_q \in [min(g_x^+, g_x^-,), max(g_x^+, g_x^-)] \times [min(g_y^+, g_y^-), max(g_y^+, g_y^-)]$ from a uniform distribution, where g^+ and g^- are the center coordinates of the positive and negative posts, respectively. α was set to 0.1, 0.2, ..., 0.9. We conducted 100 simulations for each α.

Metrics

We counted the number of confirmatory actions in the same way as the user study. In addition, we defined the bias of g_q, or the opinion score of an agent, as follows:

$$\frac{|g_{neg} - g|}{|g_{pos} - g| + |g_{neg} - g|}, \tag{9}$$

where $|x|$ is the Euclidean norm of x.

Results

Figure 6 shows the relationship between the initial bias of g_q, α, and the number of confirmatory actions. Here, because of the symmetric nature of the opinion scores that were less than and greater than 0.5, we combined the results of the two sides by flipping scores less than 0.5. With a smaller α, WEB-FEP exhibited more confirmatory behavior, and the tendency was more pronounced when g_q was biased. On the other hand, with a larger α, the number of confirmatory actions decreased. This result was consistent with our expectations. Interestingly, as we can see from the histogram, while a smaller $\alpha = 0.1$ formed the peak of the number of confirmatory actions, it was not symmetric. The effect of increasing α diminished for $\alpha \geq 0.6$ and did not increase an agent's disconfirmatory actions, which aligns with the user study results. A larger α leads to neutrality rather than disconfirmatory behavior because as the belief shifts closer to the opposite opinion, the original opinion becomes more informative, prompting the agent to explore it again.

Figure 7 shows the changes of opinion scores. WEB-FEP exhibited a variety of opinion change types and aligned with

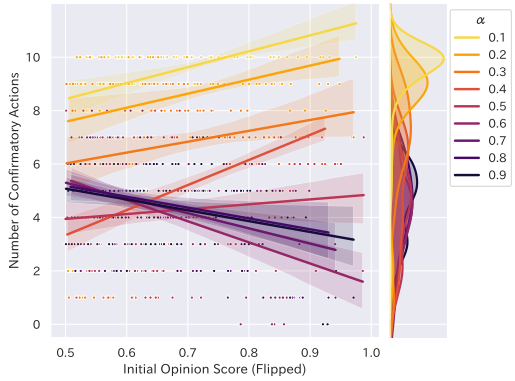

Figure 6: Number of confirmatory actions, initial position of g_q and α

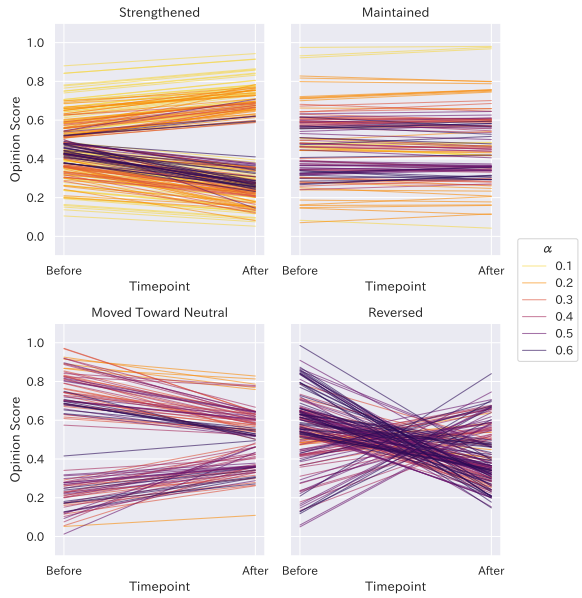

Figure 7: Opinion changes for each opinion type and α

the results of the user study (Figure 4). We also see that the results were largely influenced by α. When α was low, the opinion score tended to be strengthened toward the initial opinion, whereas when α was high, the opinion score was more likely to be neutral or reversed. Table 2 summarizes the distribution of opinion change types for each α. A lower $\alpha(\leq 0.2)$ clearly led to confirmatory results, whereas in a moderate case $(\alpha = 0.3, 0.4)$, the results were more neutral. We can see a slight rebound in the number of Strengthened with an $\alpha > 0.4$. With too high α, belief updates become larger, making beliefs more unstable with each observation. As a result, these agents were more drastically influenced by recent observations, increasing the likelihood of forming a stronger opinion rather than maintaining a neutral stance.

In conclusion, the results aligned with our expectations. The number of confirmatory actions decreased as α increased, and the opinion score was biased toward the initial

Table 2: Distribution of opinion change types across α

α	Strengthened	Maintained	Neutral	Reversed
0.1	86	14	0	0
0.2	64	23	6	7
0.3	31	23	30	16
0.4	3	18	31	48
0.5	10	22	22	46
0.6	25	16	15	44

opinion when α was low. When α was higher, the opinion score was more likely to be neutral. The results were also consistent with the user study results, suggesting that the model could reproduce the diverse user behavior seen in web media exploration.

Future Work

This paper implemented WEB-FEP on a controlled virtual SNS to focus on the fundamental characteristics of the model, but we would like to apply it to open environments, such as real SNS data, and evaluate its generalizability. An important direction for WEB-FEP application is investigating the interaction effects between the diverse user behaviors demonstrated by WEB-FEP and platform design (e.g., bias in information presentation, mixing opposing views). This would contribute to the development of neutral platforms that mitigate belief polarization. Although WEB-FEP is based on the sentence embedding space, there is room for further investigating how far WEB-FEP can capture the nuances of text content and replicate human behaviors. If it has sufficient capability, we might be able to target deceptive web designs such as clickbait. In this paper, WEB-FEP eliminated the influence of visual factors (e.g., font size, order, images), which should affect user exploration behaviors.

Although our results suggest WEB-FEP's capability to reproduce users' diverse behaviors by controlling α and g_q, further research is required to ground the parameters in psychological factors. For example, we would like to investigate the relationship between model parameters and psychological factors such as Cognitive Reflectivity (Pennycook & Rand, 2019), Need for Cognition (Cacioppo & Petty, 1982), or Bullshit Receptivity (Pennycook & Rand, 2020), which are claimed to affect users' cognition in information-seeking and belief formation.

Conclusion

This study proposed WEB-FEP, a cognitive model of users forming specific beliefs through interactions with web media. WEB-FEP attempts to computationally reproduce confirmation bias in web media exploration by formalizing the trade-off between belief-confirmatory and exploratory actions inspired by active inference. The model was validated by comparing the results of simulations with user experiments conducted on a virtual SNS. The results indicate that by adjusting α and g_q, WEB-FEP can successfully reproduce the diverse behaviors of users, including confirmatory exploration.

Acknowledgment

This work was supported by the Telecommunications Advancement Foundation.

References

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116.

Chattoraj, A., Shivkumar, S., Ra, Y. S., & Häfner, R. M. (2021). A confirmation bias due to approximate active inference. In *Annual meeting of the cognitive science society.*

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118. doi: 10.1073/pnas.2023301118

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, *11*(2), 127–138.

Gift, A. G. (1989). Visual analogue scales: measurement of subjective phenomena. *Nursing research*, *38*(5), 286–287.

Mansoury, M., Mobasher, B., & van Hoof, H. (2024). Mitigating exposure bias in online learning to rank recommendation: A novel reward model for cascading bandits. In *Proceedings of the 33rd acm international conference on information and knowledge management* (p. 1638–1648). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3627673.3679763

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, *3*(29), 861. Retrieved from `https://doi.org/10.21105/joss.00861` doi: 10.21105/joss.00861

Nickerson, R. (1998, 06). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220. doi: 10.1037/1089-2680.2.2.175

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39-50. (The Cognitive Science of Political Thought) doi: https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185-200. doi: https://doi.org/10.1111/jopy.12476

Pilgrim, C., Sanborn, A., Malthouse, E., & Hills, T. T. (2024). Confirmation bias emerges from an approximation to bayesian reasoning. *Cognition*, *245*, 105693. doi: 10.1016/j.cognition.2023.105693

Tanaka, Y., Inuzuka, M., Arai, H., Takahashi, Y., Kukita, M., & Inui, K. (2023). Who does not benefit from fact-checking websites? a psychological characteristic predicts the selective avoidance of clicking uncongenial facts. In *Proceedings of the 2023 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3544548.3580826