

SNS における情報共有に正確性プロンプトが与える効果

How Accuracy Prompts Shape News Sharing on Social Media

眞嶋 良全, 永澤 昂希
Yoshimasa Majima, Kouki Nagasawa

北星学園大学・社会福祉学部
Hokusei Gakuen University
majima.y@hokusei.ac.jp

概要

本研究は, Capraro & Celadin (2023) の研究 2 について日本人参加者を対象とした追試を行った。その結果, 真のニュースはフェイクニュースより共有されやすい傾向が見られたものの, 正確性プロンプトによる真情報の共有促進と誤情報の抑制について明確な効果は確認されなかった。一方, 偽警告プロンプトは全体的なエンゲージメントを低下させること, 日本人参加者のエンゲージメントは先行研究よりも低い傾向にあることが示された。

キーワード: SNS, プロンプト (prompt), 共有 (share), 誤情報 (misinformation)

1. はじめに

ソーシャルメディアにおける誤情報 (misinformation) やフェイクニュースの拡散は, 現代社会を脅かすリスクとなりうる。例えば 2016 年に相次いで生じた, 米国大統領選における分断とその後の Q アノン陰謀論, 英国における EU 離脱 (Brexit) にまつわる騒動, さらに 2020 年以降に本格化した新型コロナウイルス感染症では, 社会的分断が加速されたと指摘されることも多い。

そのことを反映し, 陰謀論やフェイクニュースに関する研究が増加している。近年行われたレビューによると, 陰謀論やフェイクニュースの信奉について, 直感的思考やヒューリスティック使用などの認知スタイル (Pennycook & Rand, 2021) や, 認識論的・実存的・社会的動機づけの関与が指摘されている (Douglas & Sutton, 2023)。特に, 自己あるいは所属集団に対するアイデンティティ防衛的な認知が影響し, 動機づけられた推論 (motivated reasoning, Kunda, 1990) が見られることが多い。

一方, 誤情報やフェイクニュースの伝達や拡散では, 一般的な予測(自己の信じているものを伝達する)とは異なり, 真実性の判断や信奉と伝達は必ずしも一致せず (e.g., Pennycook et al., 2021), 偽とわかっているにもかかわらず拡散されてしまうことが示されている。この説明としては, アイデンティティ防衛, または美德シグナリング (virtue signaling) の結果として, 意図的に誤情報を拡散するという動機づけ説や, 本来的には正確な情報を共有することを動機づけられていながらも, ソーシャルメディアのコンテキストがその

ような正確な共有を妨害してしまうという不注意起因説などが提唱されている (Pennycook & Rand, 2021)。

特に後者の不注意起因説に対する支持的証拠は, 誤情報の拡散抑制に対する介入研究からも得られている。例えば, Pennycook et al. (2021) は, ニュースの共有を行わせる前にその正確さを判断させるという介入 (正確さナッジ, accuracy nudge) を行うことで, フェイクニュースの共有意図が低下することを示している。ただし, この正確さナッジについては, その後の追試研究の結果, 原著で主張するほどの強い効果は持たないことも指摘されている (Roozenbeek et al., 2021)。

一方で, Capraro & Celadin (2023) は間接的ではなく, より直接的に正確さを保証 (endorse) させる介入の効果について検討している。彼らは 4 つの実験を通じて, 正確さの保証を促すプロンプト (「私はこのニュースが正確だと思っている」, 以後, 正確性プロンプト), 誤情報への注意を喚起するプロンプト (「誤った情報の可能性があることを忘れないように」, 以後, 偽警告プロンプト), およびプロンプト無しの統制条件を比較し, 正確性プロンプトがフェイクニュースの共有を抑制するだけではなく, 真のニュースの共有意図を高めること, 偽警告プロンプトは一部でフェイクニュースの共有を弱めることはあるものの, 真のニュースの拡散が増加することはないこと, 正確さに注意を向けさせるだけでは, 正確さを保証させるプロンプトほどの効果はないことなどを示している。

Capraro & Celadin (2023) は, 先行する Pennycook et al. (2021) とは異なり, 共有行動を起こすかどうかを判断させる時点で, ニュースの正確性を積極的に評価し, 表明させることで, 真の情報の共有と誤情報の拡散抑制をもたらすことを示した。加えて, SNS における共有と「いいね!」を合計した, 投稿に対する反応の総数 (エンゲージメント) はプロンプト間で差がないことも明らかとなっている。このように, 誤情報の拡散を抑制するだけではなく, 真の情報の共有を促進する効果は, 従来型の介入には見られないものであり, 正確さを個人に保証させる介入が, 有効な誤情報対策となりうる可能性を秘めている。

しかしながら、正確さナッジがその後の追試研究によって、当初の想定より効果が弱いことが示されていることを踏まえると、本介入手法についても、その再現性の検討は慎重になされるべきであろう。また、Capraro & Celadin (2023) では、参加者は全て Amazon Mechanical Turk から募集されており、この介入手法の有効性の一般化可能性にも限界があると言わざるを得ない。ソーシャルメディアプラットフォームにおける情報共有のあり方を改善するためには、より幅広い層での検討が必要である。

そのため、本研究では Capraro & Celadin (2023) が行った 4 つの研究のうち、研究 2 を対象とした再現研究を行う。研究 2 が対象として選定された理由は、先行研究で最も主要な検討対象となっている、正確保証および偽警告プロンプトの効果を、より生態学的に妥当な実験課題を用いて検討していることにある。先行研究の知見を踏まえ、本研究では、具体的に以下の 2 つの仮説を検討した。

H1: 正確性プロンプトを与えられた場合、それ以外の 2 条件と比べ、真の情報は共有されやすく、誤情報は共有されにくい。

H2: 偽警告プロンプトとプロンプト無しの統制条件との間では、真情報の共有に明確な差はない。

2. 方法

2.1. 参加者および手続き

クラウドワークスを介して 180 名の参加者 (男性 104, 女性 75, その他 1 名, $M_{age} = 43.0$, $SD = 9.8$) が参加した。ただし、SNS の使用経験がない 4 名を分析の対象から除外したため、最終的に 176 名が分析の対象となった。

参加者はクラウドワークスから、Qualtrics 上の実験サイトに誘導され、研究参加に同意した者のみが、後述のニュース共有課題、人口統計質問、注意チェックの順に回答した。また、ニュース共有課題では、プロンプトの種類に応じた 3 つの条件にランダムに割り当てられた。

2.2. 課題

本実験は、Capraro & Celadin (2023) の研究 2 に準じて行われた。参加者はランダムに、正確プロンプト (以後正確条件)、フェイク警告プロンプト (以後、偽警告)、およびプロンプトの無い統制条件にランダムに割り当てられた (それぞれ, $ns = 59, 59, 58$)。

参加者は、正情報および誤情報それぞれ 12 ずつからなる 24 試行を行い、そのそれぞれで、呈示されたニュースの見出しに対して、SNS 上でシェア、またはいいね！を

するか何もしないかを選択した。このとき、共有といいね！は同時に選択することができたが、何もしないを選択した場合は、残りの選択肢は選べないようになっていた。正情報と誤情報は、可能な限りオンライン上のニュースサイトに近い見目で呈示され、この刺激の下に、[♡いいね！][共有][どちらもしない]と書かれたボタンとともに、条件に応じて[共有]ボタンに、プロンプトが呈示された。参加者は呈示された情報に対して、共有またはいいね！をするか何もしないかを選択した。24 個の見出し (真実・フェイク各 12) はランダムな順で呈示され、参加者は 1 試行あたり 1 つのヘッドラインの評価を行った。

ニュース共有課題の後で参加者は、年齢、性別、学歴、エスニシティ、第一言語、SNS 利用経験の有無を尋ねる人口統計質問に回答し、最後に IMC (増田他, 2019) に回答した。エスニシティと第一言語は、対象者を日本人成人に限定するための、IMC 課題は不注意回答者を除外するためのスクリーニング項目として配置したが、これにより除外された参加者はいなかった。一方、SNS の利用経験のない 4 名は分析の対象から除外された。

2.3. 材料

参加者に呈示された 24 個のニュース記事のうち、正誤各 4 項目計 8 項目は Capraro & Celadin (2023) で用いられていたものであった。それ以外の項目が、日本人参加者には馴染みが薄いことを考慮し、本研究用に新たに 16 項目の記事を作成した。その際、誤情報、すなわちフェイクニュースについては、実際にインターネット上で言及されているものから選抜した。

呈示するニュース見出しは、オンライン上の実際のニュースメディアの記載に近づけるべく、左上に配信社の名称とロゴ、その下に画像付きの見出し (ヘッドライン) を示すような形で作成した。なお、フェイクニュースの場合、実際に偽ニュースメディアとして取り沙汰されている配信社の名称を使用した (例 Figure 1)。

福岡新聞

ルーブル美術館が火災に



Figure 1 Sample fake news headline

Table 1 Distribution of sharing behavior split by news veracity

Veracity	Action					
	Share		Like		None	
True	340	(16%)	380	(18%)	1467	(69%)
Fake	217	(10%)	177	(8%)	1747	(83%)

Note: $N = 2,112$ for each row (12 topics $\times$ 176 participants)

2.4. 研究参加者に対する倫理的配慮

参加者は研究参加に先立ち、研究の目的および方法などの説明を呈示され、その内容を理解した上で研究協力に同意するかどうかを、オンラインフォーム上のボタンのクリックによって回答した。ここで協力に同意した参加者のみが本実験へと進んだ。研究は、ヘルシンキ宣言や日本心理学会倫理規程の定める原則に則り適切に計画されており、著者らの所属機関における研究倫理委員会の承認を受けて実施された(承認番号:24-研倫-15号)。また、本研究は、開始前にその手法等を事前登録してから実施した (<https://doi.org/10.17605/OSF.IO/6875N>)。

3. 結果

ニュースの真偽 (veracity: true / fake) 別に 176 名の各行動 (action: share, like, none) の選択数および比率を求めた結果を Table 1 に示す。表に示される通り、本研究の参加者は、ニュースの共有や、いいね！をつける頻度は

全体的には決して多いとは言えない結果であった。
次に、ニュースの共有が、参加者に与えられたプロンプト (正確 = accuracy, 偽警告 = fake alert, 統制 = control) および、ニュースそれ自体の真実性に影響されるかどうかを検討するため、シェアが選択されたかどうかを従属変数、プロンプトおよび真偽を独立変数としたロジスティック回帰分析を行った。なお、トピックをランダム効果とした一般化線形モデルを適用した (Table 2)。

表に示される通り、ニュースの真偽、およびプロンプトの効果が有意であり、前者について、フェイクニュースは共有されにくく、後者については、偽警告条件が統制条件に比して共有行動が生じにくいことが示された (それぞれ, $ORs = 0.67, 0.64$ $ps = .007, .003$, Figure 2 (a) も参照)。一方、条件と真偽の交互作用として、正確条件における真偽ニュースの共有率は、他条件より差が大きいようにも見えるが、当該交互作用は有意ではなかった ($OR = 0.64, p = .051$)

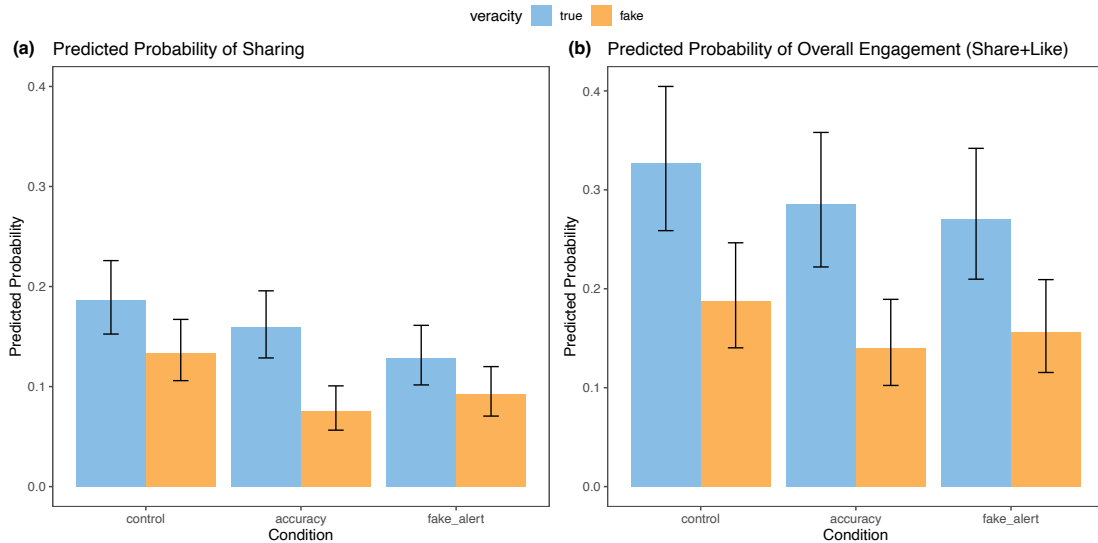
また、共有といいね！を合わせた全体的なエンゲージメント (少なくともどちらか一方を行う) の選択について、同様のロジスティック回帰分析を行った結果、共有のみの場合と同様にフェイクニュースは共有されにくく ($OR = 0.47, p < .001$)、偽警告条件は全般的にエンゲージメントが低いことが示された ($OR = 0.76, p = .022$, Table 2 および Figure 2 (b) 参照)。

Table 2 Logistic Regression Analyses Predicting Share and overall engagement

Predictors	Share (1 = Yes)			Engagement (1 = Yes)		
	OR	SE	p	OR	SE	p
(Intercept)	0.23	0.03	<.001	0.49	0.08	<.001
veracity [fake]	0.67	0.10	.007	0.47	0.06	<.001
condition [accuracy]	0.83	0.12	.178	0.82	0.10	.090
condition [fake_alert]	0.64	0.10	.003	0.76	0.09	.022
[fake] $\times$ [accuracy]	0.64	0.15	.051	0.86	0.16	.422
[fake] $\times$ [fake_alert]	1.02	0.23	.914	1.05	0.19	.772
Random Effects						
σ^2	3.29			3.29		
τ_{00} topic	0.07			0.26		
ICC	0.02			0.07		
N_{topic}	12			12		
Observations	4224			4224		
Marginal R^2 / Conditional R^2	0.035 / 0.055			0.045 / 0.116		

Notes: Engagement = choosing either share or like. OR = Odds ratio.

Figure 2 Predicted probabilities of engagement (share / overall) split by news veracity (true vs. fake) and the type of prompt. (a) = share, (b) = overall engagement (share + like)



4. 考察

本研究は、情報共有行動に対する参加者自身のコミットメント、より具体的には情報の正確さを保証させる介入、すなわち正確性プロンプトによって、誤情報の拡散抑制と真の情報の拡散の促進が生じるかどうかについて、Capraro & Celadin (2023) の研究 2 を対象とした追試研究を行った。その結果、全般的に真情報の方が誤情報より共有されやすいという点は共通していたものの、正確性プロンプトが、偽警告プロンプトやプロンプト無し場合と比べて、真の情報の拡散を促進したり、誤情報の拡散を抑制するといった明確な介入効果があるとは言えない結果であった。一方で、偽警告プロンプトを与えることが、全般的なエンゲージメントを低下させることが示されており、SNS での情報共有行動について、文化差要因の関与が示唆されている。また、Capraro & Celadin との比較において、先行研究では全体的なエンゲージメントがおおよそ 5 割程度に到達していたのに対して、本研究の参加者は概ね 3 割程度以下と、そもそものエンゲージメントが低いという結果も得られている。このことから、日本人参加者においては、情報を拡散させることによる他者からの評価への懸念などが影響している可能性が推察される。一方で、正確さに対するコミットメントの捉え方や、プロンプトの解釈の違いなどが影響した可能性も考えられるため、単に文化的な要因だけではなく、今後は、誤情報拡散防止の介入としてのプロンプトの効果の頑健性について、より詳細な検討が必要であろう。

5. 参考文献

- Capraro, V., & Celadin, T. (2023). “I Think This News Is Accurate”: Endorsing Accuracy Decreases the Sharing of Fake News and Increases the Sharing of Real News. *Personality and Social Psychology Bulletin*, 49(12), 1635–1645. <https://doi.org/10.1177/01461672221117691>
- Douglas, K. M., & Sutton, R. M. (2023). What Are Conspiracy Theories? A Definitional Approach to Their Correlates, Consequences, and Communication. *Annual Review of Psychology*, 74(1), 271–298. <https://doi.org/10.1146/annurev-psych-032420-031329>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- 増田真也・坂上貴之・森井真広. (2019). 調査回答の質の向上のための方法の比較. *心理学研究*, 90(5), 463–472. <https://doi.org/10.4992/jipsy.90.18042>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Roozenbeek, J., Freeman, A. L. J., & van der Linden, S. (2021). How Accurate Are Accuracy-Nudge Interventions? A Preregistered Direct Replication of Pennycook et al. (2020). *Psychological Science*, 32(7), 1169–1178. <https://doi.org/10.1177/09567976211024535>