

Modeling the Social Reinforcement of Misinformation Dissemination on Social Media

Ashleigh T. Aston 

Ink Bowl Publishing, New York, NY, USA

Email: ashleigh0aston@gmail.com

How to cite this paper: Aston, A.T. (2022) Modeling the Social Reinforcement of Misinformation Dissemination on Social Media. *Journal of Behavioral and Brain Science*, 12, 533-547.
<https://doi.org/10.4236/jbbs.2022.1211031>

Received: September 3, 2022

Accepted: November 4, 2022

Published: November 7, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Despite the salience of misinformation and its consequences, there still lies a tremendous gap in research on the broader tendencies in collective cognition that compels individuals to spread misinformation so excessively. This study examined social learning as an antecedent of engaging with misinformation online. Using data released by Twitter for academic research in 2018, Tweets that included URL news links of both known misinformation and reliable domains were analyzed. Lindström's computational reinforcement learning model was adapted as an expression of social learning, where a Twitter user's posting frequency of news links is dependent on the relative engagement they receive in consequence. The research found that those who shared misinformation were highly sensitive to social reward. Inflation of positive social feedback was associated with a decrease in posting latency, indicating that users that posted misinformation were strongly influenced by social learning. However, the posting frequency of authentic news sharers remained fixed, even after receiving an increase in relative and absolute engagement. The results identified social learning is a contributor to the spread of misinformation online. In addition, behavior driven by social validation suggests a positive correlation between posting frequency, gratification received from posting, and a growing mental health dependency on social media. Developing interventions for spreading misinformation online may profit by assessing which online environments amplify social learning, particularly the conditions under which misinformation proliferates.

Keywords

Social Reinforcement Learning, Misinformation Online, Computational Models of Cognition

1. Introduction

Since the rise of social media's popularity in the early 2000s, the spread of misinformation has become an epidemic across all internet platforms [1]. The controversies associated with the 2016 U.S. Presidential Election, in particular, accelerated its propagation explosively. In addition, the expansive latitude of online mass media has expedited information dissemination, providing 4.65 billion social media users [2] with nearly instantaneous access to limitless, unregulated, global information. Hence, internet platforms have become the primary hotbeds and enablers of fake news. This leaves implementing an accurate and successful method of distinguishing between fake and reliable news an increasingly arduous task.

The magnitude of false intel radiated can lead to misguided, potentially detrimental decision-making. Misinformation can be identified as fallacious information accepted as a credible source. In comparison, disinformation and fake news refer to falsities posing as authentic news spread with the intent to deceive [3]. Developing an effective solution will require understanding the underlying psychological processes responsible for motivating people to spread misinformation, and a lapse in critical fact examination, so we can target human cognition's aspects that encourage its recurrence.

On this premise, a computational model of social reinforcement learning online was proposed to illustrate one psychological mechanism behind the spread of misinformation on social media. Social learning theory describes how human behaviors are valued and developed through social evaluative feedback within a network of inter-species interactions [4]. The concept is similar to traditional reinforcement learning and applies to conventional models [5] [6]. However, the two cognitive processes deviate because traditional reinforcement learning stipulates that learning aims to maximize the reward function rather than the acquired value of the action in social learning [6] [7]. The function of a standard computational model simulates, predicts, and explains mental processes through a formal mathematical procedure [8]. For our initiative, this structure was applied to behavioral data apropos to online misinformation movements, interpreting trends in numerical values to identify how they are affected by social reinforcement learning in humans.

1.1. The Consequences of Misinformation

Within the extensive scope of online sharing, misinformation makes up only a small percentage of exposed material [1] [9] [10] [11]. However, it accounts for extensive complications, such as significantly contributing to the growing geo- and socio-political divide. Misinformation circulating across online discussions has been used to incriminate relevant figures, companies, organizations, and political parties [12] through rumors and false accusations, exacerbating polarization and damaging the victim's reputation [13]. The division fosters internet communities of users with shared partisan preferences. These "echo chambers"

are insulated systems that amplify a single belief or opinion and are void of countering viewpoints. The echo chamber affects pre-disposition users to reject alternative news sources and accept those that favor their bias [14]. The evolution of online dynamics has sparked a growing mistrust in news sources. In 2016, a survey reported that merely 14% of Republicans and 51% of Democrats held “a fair” or “a great deal” of trust in the mass media as a news source. Another study revealed that false information is retweeted more rapidly than accurate information on Twitter. This was especially the case when the post concerned politics [3].

In some instances, misinformation transmission can act as a health hazard and generate mass hysteria and anxiety trends. False claims that the Measles, Mumps, and Rubella (MMR) vaccine caused autism elevated fears, contributing to lower vaccination and immunization rates in the late 1990s [15] [16]. The ramifications endured the next 20 years, with measles outbreaks plaguing Washington State and New York in 2017. As recently as 2019, the De Blasio Administration’s Health Department of New York declared the measles crisis a public health emergency in response to regional flares [17]. More contemporaneously, amid the COVID-19 pandemic, the vast output of misinformation circulating across the media has caused people to decline vaccinations, use unproven treatments, and refuse to comply with public health measures, such as social distancing and masking.

Fraudulent information was so salient that a 2021 study revealed that even brief exposure to misinformation regarding COVID-19 vaccinations decreased the likelihood of an individual opting to receive it [18]. Misguided decision-making is a danger to that individual and the surrounding population as collateral damage. In addition, the mental health of adolescents and young adults is especially at risk. A 2022 study found that students who were less aware of misinformation experienced higher rates of anxiety and despair than those who were conscious of its presence [19]. Because of these threats, curtailing the spread of misinformation is imperative.

1.2. Misinformation Prevention

Most attempts to stem misinformation have failed [20]. Websites such as Check-YourFact.com, FactCheck.org, and Snopes.com are sources that aim to debunk false information circulating online. However, research on the efficacy of their fact-checking has received inconsistent findings. Their inconsistency, in part, lies in the fact-checkers themselves, who are also susceptible to deception and biased to their pre-established views. Additionally, the audience of these sites is far more limited than that of other unregulated mass-media information sources, rendering their influence finite [3].

Measures through social media platforms have also been privately undertaken [21]. Social media corporations claim to have administered checks on misinformation but have not disclosed the nature of these systems, which are unreliable

themselves. The elusive nature of the algorithms, and particularly their futility, responsible for regulating content distribution across the platform, raises suspicion that social networks may be contributing to the spread.

1.3. The Role of Social Learning in the Dissemination of Misinformation Online

A recent study utilizing a computational reinforcement learning model indicated that users were more driven to interact with and post content evoking an emotionally outraged response, as revealed by more feedback activity (*i.e.* likes, shares, comments) in response. In addition, receiving more engagement than usual encouraged users to continue to share similar content in pursuit of the same validation. Such trends can be interpreted through reinforcement and normal learning [22].

However, that does not explain the trend of sharing misinformation. Social learning can rationalize how such activity is encouraged. The interaction between a user's activity and the feedback influences the user's future actions. The desirability of behavior is directly linked to the strength of the reinforcement. Positive social signals encourage frequent repetition of a behavior, whereas negative responses generate the opposite effect. Consistent positive feedback solidifies an actor's tendency to disseminate misinformation. Alarming, once a tendency has been entrained, it becomes unremitting even without positive feedback.

As the spread of misinformation gains traction, it reaches a larger audience, who then become susceptible to influence and adversely. This presents another issue. Studies have shown that familiarity and visibility increase a person's faith in information. Hence, trending content is more readily accepted and taken at face value [3]. After exposure, if a user proceeds to post the misinformative content and generates high engagement, the behavior's "value" will increase and motivate them to seek the same reward in the future. The degree of positive feedback dictates the subsequent frequency of the behavior. Numerous studies have proved an absence of online engagement induces undesirable outcomes such as depression, stress, loneliness, and anxiety [23]. These withdrawal effects motivate an increased usage of social media, particularly those (<30), indicating that social reinforcement has a tremendous impact on one's mental state [24]. Ergo, the dynamics of social media platforms are analogous to B.F. Skinner's "Skinner Box," studying the reinforcement of animal behavior when provided with consequential reward through the lens of operant conditioning.

Social media algorithms are programmed to increase the visibility of topics that elicit the greatest activity to maximize user engagement [3]. The sheer size of social media networks and nearly instantaneous access to global information on the platform accelerates information sharing and, therefore, social learning on a massive scale, enabling rapid and abundant positive feedback. Over time, this amplifies unchecked content, compulsive posting [23], and a failure to examine the source's credibility, especially when validation is desperately sought [25].

Therefore, human behavior driven by social learning generates waves of media trends and spreads misinformation.

Human behavioral movements on social media are more effectively predicted by a computational reinforcement learning model [8] than by a standard linear model that relates posting frequency to feedback intensity. Based on the similarities between traditional reinforcement learning (R.L.) and social learning, our objective was to adapt Lindström's reward learning model as an expression of social learning, where the posting frequency of sharing news links depends on the social evaluative feedback of relative engagement. It has already been re-engineered to reflect the role of moral outrage in spreading misinformation [22], implying applicability to our study. By utilizing an R.L. model, we compare the sensitivity of individual users' habits of sharing authentic news and factually-inaccurate news to receiving content engagement.

If the frequency at which misinformation is shared depends on the traffic received, then the posting behavior's value is expected to rise in proportion to the activity rate, implicating social learning as a mechanism whereby misinformation spreads. Formally, we address the question; *can computational models of social reinforcement learning online explain the virality of misinformation across social media platforms?*

2. Methods

To test whether computational models can capture the spread of misinformation in social reinforcement learning, we modified an existing computational model of reinforcement learning on social media. First, Lindström's original 2019 computational model of R.L. online was adapted as an expression of social learning and evaluative social feedback rather than maximizing the reward of a behavior. Then, Lindström's model was applied to datasets of misinformation made available to researchers by Twitter during the 2016 U.S. Presidential Election. This period was an ideal case study for our research since misinformation abounded in that political climate.

The data limits our research, representing human behavior and social learning only within a narrow online ecosystem. Theoretically, the study does not address how or if social learning contributes to the spread of misinformation online. Furthermore, configuring esoteric psychological processes compatible with structured mathematical functions forces us to forfeit an understanding of human cognition more broadly.

2.1. Modeling Reinforcement Learning Online

Our model predicts that the frequency with which a user shares news online—and misinformation in particular—reflects decision-making based on social reinforcement learning ($\tau_{Express}$). Patterns of posting frequency are modified by feedback expectations, a function of an individual's perceived cost of posting misinformation, and that individual's estimate of the feedback rate, R (mc). Predic-

tions are based on the value of behavior determined by prior social interactions. With each social network interaction, $\tau_{Express}$ is adjusted based on the difference between expected (R) and experienced feedback rate.

We have also implemented three parameters:

- 1) The learning rate (α);
- 2) The starting or initial “policy” at the time ($t = 1$);
- 3) The individual’s sensitivity to the subjective cost of expression (C).

Equations (1) - (5) detail the entire model.

$$\tau_{Express^\tau} = e^{Policy^\tau} - \alpha * \bar{R}^\tau \quad (1)$$

$$\delta^\tau = R^\tau - \frac{C}{\tau_{Express^\tau}} - \bar{R}^\tau * \tau_{Express^\tau} \quad (2)$$

$$\Delta \tau_{Express^\tau} = \tau_{Express^\tau} - \tau_{Express^{\tau-1}} \quad (3)$$

$$Policy^{\tau+1} = Policy^\tau + \alpha * \Delta \tau_{Express^\tau} * \delta^\tau \quad (4)$$

$$\bar{R}^{\tau+1} = \bar{R}^\tau + \alpha * \delta^\tau \quad (5)$$

(1) For every decision driven by social reinforcement learning, the model elects ($\tau_{Express}$) from an exponential distribution with a dynamic mean. The initial policy (time ($t = 1$)) is a free parameter. Subtracting the product of the learning rate (α) and average reward rate (\bar{R}^τ) models the latter’s impact on response frequency. (2) The response policy is adjusted congruently to the prediction error (δ^τ), defined by the difference between expected (\bar{R}^τ) and the experienced (R^τ) reward. The former captures effort cost (C), determined by an average subjective reward estimate (\bar{R}^τ). (3), (4) Feedback is maximized by updating the response policy through a reward gradient. The gradient tracks the deviation from maximum feedback at time (t) ($\Delta \tau_{Express^\tau}$). (5) Prediction error updates the average reward rate. An increase in reward rate results in smaller response latencies. Adapted from Lindström et al. [8]

2.2. Misinformation Data Set

Major media and public attention were paid to the role of state-sponsored misinformation dissemination during the 2016 U.S. presidential election. In particular, the Russian Internet Research Agency (IRA) created a network of Twitter accounts to influence the election [26]. A significant component of their online presence consisted of posts linking to political misinformation. We used a comprehensive dataset of the tweets and associated metadata posted by the IRA to test whether Lindström’s model could explain users’ decisions to disseminate misinformation over time.

In 2018, Twitter launched its Information Operations, which made all the tweets they suspected being posted by the IRA or other state-linked entities attempting to manipulate Twitter trends available to academic researchers. The archives can be accessed via a simple online application process [27]. The dataset our research analyzed, labeled the “Elections Integrity Dataset” by Twitter, com-

prised 3613 accounts and was released in October 2018. Twitter specified that although no content had been redacted, to avoid violating ethics and privacy, the personal elements of some accounts had been hashed out of the publicly-available datasets. Namely; screen name, profile photo, and user I.D. [28].

Once we were granted access to the data, we used the statistical computing program R to refine the dataset and derive an automated process from comparing each IRA tweet that contained at least one URL against two databases of web domains: 1) a dataset of known misinformation domains (N = 1699; Existing misinformation domains, 2020) and 2) a dataset of authentic news domains (N = 6378; Authentic news domains, 2020). Both datasets capture national as well as local online news outlets. **Table 1** provides examples of misinformation and authentic news included in the databases. Using this process, we could label the IRA tweets with URLs as either linking to misinformation or authentic news.

The data set was also comprehensively refined to eliminate the activity of another key player in spreading misinformation online: bots. These automated A.I. Twitter accounts are used to manipulate algorithms by impersonating human beings and engaging with or posting information. A recent study, which examined Twitter user behavior, ties, and observable features, estimated that bots made up 9% to 15% of active Twitter accounts, allowing them to precipitate mass media trends. This was especially the case during the 2016 U.S. election, where bots were significant contributors to posting political content [3]. Although the infestation influenced the climate under which the data was collected, the bots remain redacted.

The Twitter accounts operated by the IRA were active on the platform as far back as 2014, and their detection and removal from the platform did not occur until late in 2018. However, they were most active from the summer of 2016 to the same period in 2017. Between 2014 and the period preceding summer 2016, the combined number of tweets posted by all IRA accounts never exceeded 1000 in any given month (**Figure 1**). Across the 12 months from July 2016 to July 2017, the average number of combined posts was greater than 20,000. This date range is unsurprising as it overlaps with the buildup to and aftermath of the 2016 U.S. presidential election, which the IRA presumably sought to disrupt [26]. Therefore, we confined our analyses to this time period.

Table 1. Examples of news domains found in the authentic news and misinformation datasets.

Classification	Internet Domain	Reach
Authentic	washingtonpost.com	National
	nytimes.com	National
	8online.com	Local
Misinformation	100percentfedup.com	National
	winningdemocrats.com	National
	ahtribune.com	Local

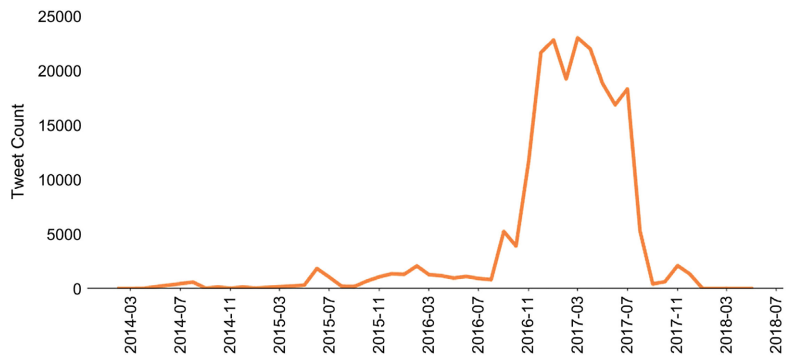


Figure 1. Timeline of the internet research agencies tweet activity. *Note:* The number of tweets posted by IRA accounts during the period they were active, at the month level. The year from 2016-07 to 2017-07 was their most active period.

3. Procedure

To test whether models of online reinforcement learning can explain misinformation dissemination online, we applied the above computational model to the IRA dataset. The procedure we followed mirrors that described in Lindström *et al.* [4].

We first selected Twitter accounts in the data that had posted a link to a misinformation news domain on at least five occasions during the 2016 election period. Most of these accounts (>90%) also tweeted links to factually accurate news domains. This is unsurprising, as most politically-active users post large amounts of political content during periods of heightened political tension, such as during the 2016 election. Therefore, we confined our analysis to these users and applied the Lindström model to the Twitter metadata (*i.e.* time posted and engagement received the sum of replies, retweets, and favorites). All models were processed through the statistical computing program R, package lme4 (version 4.1.1).

4. Results

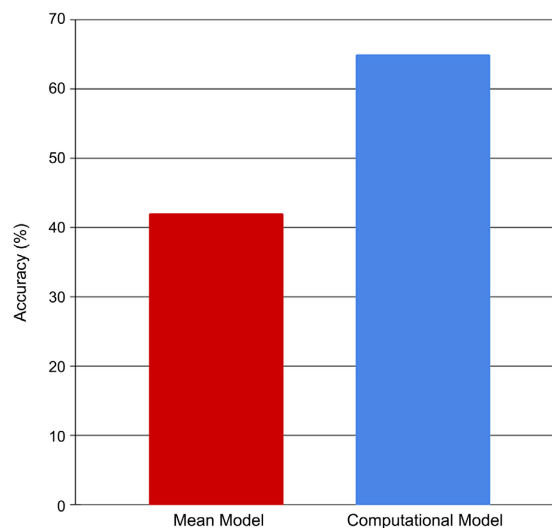


Figure 2 Predictive accuracy of the computational model and mean model. It showcases the computational model to be ~26% more reliable when predicting posting behavior than the mean model.

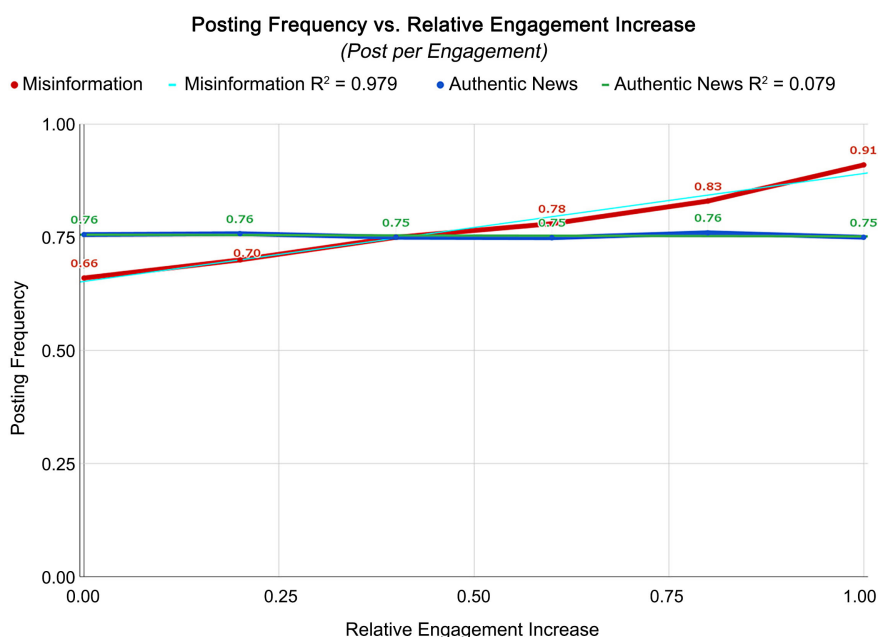


Figure 3. Results of fitting a social reinforcement learning model to misinformation Tweets. It demonstrates that Twitter users who shared misinformation displayed acuity to social reward and a decrease in posting latency compared to those who shared factually accurate news links.

5. Discussion

We replicated Lindström's [4] finding that social reward in the form of engagement on social media—in this case, Twitter—can explain the latency between posts at the user level. Specifically, collapsing across news types (*i.e.* misinformation and factually-accurate news), we found that increases in engagement with a user's content were associated with an increase in posting frequency and that this decrease in posting latency was accounted for by the computational model we applied to the data. Furthermore, the computational was more accurate at predicting the latency between postings than a model based only on mean latency (see **Figure 2**). This suggests that the computational model is a more accurate foreteller of users' posting decisions than a mean model.

We found that this effect of reward on information posting was almost entirely specific to misinformation posts. When we separated tweets that propagated misinformation from tweets that shared authentic news, we found that the reward for misinformation was associated with a decrease in time between misinformation posts. This was not true of factually accurate news posts (**Figure 3**). As we used a within-subjects design, we believe this finding provides strong evidence that reinforcement learning processes sensitive to social reward encourage the spread of misinformation online. Furthermore, we found this effect robust to relative increases in engagement, not just absolute values. When we replaced absolute engagement in our model with a measure of the increase in engagement above a user's baseline, we found that, on average, the effect of the delay between misinformation posts decreased as relative feedback increased.

The adaptation of Lindström's model has unearthed consistencies in behavioral changes when contrasting the online sharing of authentic news and misinformation. **Figure 3**, which illustrates posting frequency vs. relative engagement increase, demonstrates that users expressing veracity generally showed slight variance in routine publications. Posting frequency barely deviated from its baseline level at 0.75. At a relative engagement increase of 0.60, posting frequency dipped slightly below its starting point to ~ 0.75 . Frequency maxed out at ~ 0.76 when relative engagement reached 0.8 but fell back to ~ 0.75 at an engagement increase of 1.0. Relatively, there was no NET change in posting frequency between 0 and 1.0 relative engagement increase. The linear trendline for the authentic news data ($R^2 = 0.079$) affirms that there was no linear relationship between posting frequency and relative engagement increase. In essence, authentic news sharers were unreceptive to reforming their customs online after increases in absolute and relative engagement. Even when social reward increased, posting frequency remained fixed.

In contrast, those posting misinformation exhibited evidence of social reinforcement learning. Inflation of positive feedback from other Twitter users has been associated with an increase in the rate that an individual circulates factually-inaccurate news. Posting frequency elevated exponentially in conjunction with greater relative engagement. The collective average posting frequency of the misinformation data had started at ~ 0.66 , below that of authentic news at ~ 0.75 . However, the posting frequency of misinformation sharers had become comparable to authentic news sharers at a relative engagement increase of 0.4 (misinformation & authentic news = ~ 0.75). It exceeded them at an increase of 0.6 (misinformation = ~ 0.78 , authentic news = ~ 0.75). The NET change of misinformation posting frequency from 0 to 1.0 relative engagement increases was ~ 0.25 . The trendline of the misinformation dataset ($R^2 = 0.979$) corroborates the linear relationship between the two. Posting frequency is dependent on relative engagement increase. Unlike authentic news, the inclination for the rapid spread of misinformation is highly sensitive to social reward.

Our results suggest that we successfully used a computational model of cognition to demonstrate that social reinforcement learning can explain some variances in misinformation proliferation. It has also been proven more accurate in predicting human behavior under the given circumstances than a mean model. The figures present strong relationships or significant independence between social reinforcement learning and authentic or factually-accurate news posting frequency. Such consistent behavior documentation in both instances backs the credibility of our findings. In the case of misinformation spreading, our hypothesis, if a user on a social media platform receives an increase in engagement for sharing misinformation, then they will be encouraged to replicate their action in the future more frequently in proportion because the individual received positive evaluative feedback from the social network, therefore increasing the behavior's value, has been irrefutably supported.

The evidence ruling that social feedback has a linear relationship with misin-

formation circulation rates did not apply to reliable news. Authentic news sharers resisted elevated engagement rates altering their online posting habits. Markedly, there must be explanations for why misinformation is substantially more receptive to social reinforcement than reliable news, even when refined under the same process and circumstances. Therefore, while our computational R.L. model has accurately computed the numerical, superficial values of the dataset, it fails to explain the complete mechanisms by which they are formed.

Projecting the epidemic of misinformation sharing online as a byproduct of social reinforcement learning requires a drastic simplification of human cognition and the establishment of numerous limitations to external and internal factors with which it would typically interact. With this in mind, we consider why social learning is so crucial in prompting behavioral evolution under the conditions of misinformation alone. We speculate that the presence of the climatic factors unaccounted for by our computational model is contributor.

A fundamental barrier to information accuracy begins with Twitter data. We examined the behaviors of individual users 1) on one social media platform and 2) who have posted misinformation or authentic news. Working with numerical datasets (tweets, shares, and likes) requires translating abstract cognitive functions into a quantifiable configuration. This demanded setting a foundation of environmental variable constrictions which had to be incorporated into our adaptation of Lindström's model.

Among the constraints are variances in learning styles and the acquisition of relevant behaviors. For instance, posting imitation is driven by vicarious reinforcement, when a user reshapes misinformation observed to have already received positive feedback for another account. The relationship between social learning and the algorithm also leaves an immense gap in forging accurate computation. Movements online that develop large-scale popularity face increased visibility and influence if meeting an unknown set of programmed factors. Suppose misinformation tends to attract larger quantities of social feedback and aligns with those parameters. In that case, we can theorize that users who post this content receive greater social feedback and, therefore, more positive reinforcement generically. Our analysis of the Twitter data focuses on increases in network engagement relative to the status of a user's prior posts, meaning that we are not comparing the amount of engagement received. If reliable news ordinarily receives less social feedback than misinformation, the behavior's value may not meet the threshold for continued re-sharing.

When greater gratification galvanizes posting more frequently, we can conjecture that a failure to receive social reinforcement may cause a user to experience subsequent mental distress. The consequence, potentially, becomes addictive posting behavior and a growing dependency on social reinforcement for a positive mental state. Failure to receive social rewards can feasibly cause withdrawal-like symptoms such as anxiety and depression, which have been proven to be exacerbated by an obsession with likes, comments, and followers on social media [23]. Such symptoms support the trending decrease in posting latency for mi-

sinformation sharers.

Moral outrage incentivizes engagement with and propagation of misinformation. A study focusing on behavioral trends on Twitter has already demonstrated that misinformation elicits more engagement by provoking outrage expressions [22]. Since misinformation is often harmful or misleading, it can trigger strong emotional responses when it contradicts or validates deeply held personal beliefs. In a charged political climate, outrage can be assumed to be more common than it would be regarding other topics. This may explain, in part, why misinformation is more sensitive to social reinforcement feedback than authentic news, especially within the Twitter dataset.

Nevertheless, social media companies have been loath to release the algorithms that govern peak content circulation and engagement. Without a way to address such an integral component, the Twitter algorithm had to be excluded from our model, and a user's complex stochastic interactions within the online community were discarded. Hence, our computational model attributes all behavioral patterns on online media to social learning alone rather than cooperation with other psychological processes.

That being said, confirmation that computational models of social reinforcement learning online can explain the virality of misinformation on social media platforms is important because it offers a tangible, quantitative comparison of how the abstraction of human cognition is impacted under different online climates and how this affects behavior.

Aspects of misinformation, such as innately provoking outrage, seem to appeal to users. Misinformation seems to captivate the influence of the Twitter algorithm and those of other social media platforms, dictating content exposure on their platforms. When these algorithms promote misinformation visibility, it can trigger imitation of engagement-rewarded behavior across the platform to exponential extents through variant and social reinforcement learning. Once adopted by a user, the action will customarily be continued, increasing posting rates as the projection of their social reward surges. Concurrently, neglect of critical information analysis will also persist.

The implications of these addictive tendencies are a growing misguided population and reliable news sources becoming increasingly difficult to discern. The trends of social learning we have discovered are important because they reflect a need for validation online and concomitant psychological stress. Users with a high dependency on social reinforcement have a strong objective to receive feedback, leading to posting more frequently. When misinformation is shared regularly, it indicates a routine breach in content assimilation. The drive for social reward overrides the consideration of the potential consequences of sharing fake news or questioning the source's reliability. In addition, reckless posting can allude to underlying mental health challenges, such as depression and anxiety, which can regress when a reward is absent and motivate habitual negligence.

A computational model is, by definition, an algorithm. Since our computational cognition model successfully predicted changes in human posting beha-

behavior due to social reinforcement learning, social media algorithms can be reengineered to identify mathematical trends of impulsive posting behavior in response to social reinforcement or online engagement. Future ventures to regulate misinformation spread online should involve adapting computational models with content-recognition capacity. The new algorithms should be designed to monitor information-sharing trends by identifying “keywords” or phrases flagged as fraudulent by an operator or an automated fact-checking complex.

6. Conclusion

Equally fundamental, their function must also assess the corresponding climatic factors which precipitate these activity patterns and conduce to the spread of misinformation. Investing in research that deconstructs how the current algorithms have been structured to expedite the spread of misinformation more than authentic news can help social media companies reassemble them to be capable of eliminating the dissemination of misinformation online in the future.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Watts, D.J., Rothschild, D.M. and Mobius, M. (2021) Measuring the News and Its Impact on Democracy. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e1912443118. <https://doi.org/10.1073/pnas.1912443118>
- [2] Global Social Media Statistic (2022) Data Report. <https://datareportal.com/social-media-users>
- [3] Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J. and Zittrain, J.L. (2018) The Science of Fake News. *Science*, **359**, 1094-1096. <https://doi.org/10.1126/science.aao2998>
- [4] Olsson, A., Knapska, E. and Lindström, B. (2020) The Neural and Computational Systems of Social Learning. *Nature Reviews Neuroscience*, **21**, 197-212. <https://doi.org/10.1038/s41583-020-0276-4>
- [5] Sutton, R.S. and Barto, A.G. (1998) Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, **9**, 1054-1054. <https://doi.org/10.1109/TNN.1998.712192>
- [6] Vélez, N. and Gweon, H. (2021) Learning from Other Minds: An Optimistic Critique of Reinforcement Learning Models of Social Learning. *Current Opinion in Behavioral Sciences*, **38**, 110-115. <https://doi.org/10.1016/j.cobeha.2021.01.006>
<https://www.sciencedirect.com/science/article/pii/S2352154621000073>
- [7] Ho, M.K., MacGlashan, J., Littman, M.L. and Cushman, F. (2017) Social Is Special: A Normative Framework for Teaching with and Learning from Evaluative Feedback. *Cognition*, **167**, 91-106. <https://doi.org/10.1016/j.cognition.2017.03.006>
- [8] Lindström, B., Bellander, M., Schultner, D.T., Chang, A., Tobler, P.N. and Amodio,

- D.M. (2021) A Computational Reward Learning Account of Social Media Engagement. *Nature Communications*, **12**, Article No. 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- [9] Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L.Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J. and Wu, T. (2020) The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 Pandemic. *Nature*, **584**, 262-267. <https://doi.org/10.1038/s41586-020-2404-8>
- [10] Grinberg, N., Joseph, K., Friedland, L., Swire-thompson, B. and Lazer, D. (2019) Fake News on Twitter during the 2016 U.S. Presidential Election. *Science*, **363**, 374-378. <https://doi.org/10.1126/science.aau2706>
- [11] Guess, A., Nagler, J. and Tucker, J. (2019) Less than you Think: Prevalence and Predictors of Fake News Dissemination on Facebook. *Science Advances*, **5**, eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- [12] Wardle, C. and Derakhshan, H. (2017) Information Disorder: Toward an Interdisciplinary Framework for Research and Policy-Making. Council of Europe Report. <http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf>
- [13] Osmundsen, M., Bor, A., Vahlstrup, P.B., Bechmann, A. and Petersen, M.B. (2021) Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, **115**, 999-1015. <https://doi.org/10.1017/S0003055421000290>
- [14] Dapcevich, M. (2022) Snopestationary: What Is an Echo Chamber? <https://www.snopes.com/articles/428074/what-is-an-echo-chamber/>
- [15] Chen, E., Chang, H., Rao, A., Lerman, K., Cowan, G. and Ferrara, E. (2021) COVID-19 Misinformation and the 2020 U.S. Presidential Election. *The Harvard Kennedy School Misinformation Review*, **1**, 1-17. <https://doi.org/10.37016/mr-2020-57>
- [16] Sathyanarayana, R.T.S. and Andrade, C. (2011) The MMR Vaccine and Autism: Sensation, Refutation, Retraction, and Fraud. *Indian Journal of Psychiatry*, **53**, 95-96. <https://doi.org/10.4103/0019-5545.82529>
- [17] The Official Website of the City of New York (2019) De Blasio Administration's Health Department Declares Public Health Emergency due to Measles Crisis. <https://www1.nyc.gov/office-of-the-mayor/news/186-19/de-blasio-administration-s-health-department-declares-public-health-emergency-due-measles-crisis#0>
- [18] U.S. Public Health Service Surgeon General of the United States (2021) Confronting Health Misinformation. The U.S. Surgeon General's Advisory on Building a Healthy Information Environment, 1-22. <https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf>
- [19] Jabbour, D., Masri, J.E., Nawfal, R., Malaeb, D. and Salameh, P. (2022) Social Media Medical Misinformation: Impact on Mental Health and Vaccination Decision among University Students. *Irish Journal of Medical Science*. <https://doi.org/10.1007/s11845-022-02936-9>
- [20] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D. and Rand, D.G. (2021) Shifting Attention to Accuracy Can Reduce Misinformation Online. *Nature*, **592**, 590-595. <https://doi.org/10.1038/s41586-021-03344-2>
- [21] Graves, L. and Mantzarlis, A. (2020) Amid Political Spin and Online Misinformation, Fact-Checking Adapts. *The Political Quarterly*, **91**, 585-591. <https://doi.org/10.1111/1467-923X.12896>
- [22] Brady, W.J., Mcloughlin, K., Doan, T.N. and Crockett, M.J. (2021) How Social

-
- Learning Amplifies Moral Outrage Expression in Online Social Networks. *Science Advances*, 7, eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- [23] Bashir, H. and Bhat, S.A. (2017) Effects of Social Media on Mental Health: A Review. *The International Journal of Indian Psychology*, 4, 125-131. <https://doi.org/10.25215/0403.134>
- [24] Strickland, A.C. (2014) Exploring the Effects of Social Media Use on the Mental Health of Young Adults. Ph.D. Thesis, University of Central Florida, Orlando.
- [25] Crockett, M.J. (2017) Moral Outrage in the Digital Age. *Nature Human Behaviour*, 1, 769-771. <https://doi.org/10.1038/s41562-017-0213-3>
- [26] Thompson, N. and Lapowsky, I. (2018) How Russian Trolls Used Meme Warfare to Divide America. <https://www.wired.com/story/russia-ira-propaganda-senate-report/>
- [27] Gadde, V. and Roth, Y. (2018) Enabling Further Research of Information Operations on Twitter. Twitter Blog. https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter
- [28] Information Operations (n.d.) Twitter Moderation Research Consortium. <https://transparency.twitter.com/en/reports/information-operations.html>