PNAS

Sharing of misinformation is habitual, not just lazy or biased

Gizem Ceylan[a,1] (ID), Ian A. Anderson[b], and Wendy Wood[b,c] (ID)

Why do people share misinformation on social media? In this research ($N = 2,476$), we show that the structure of online sharing built into social platforms is more important than individual deficits in critical reasoning and partisan bias—commonly cited drivers of misinformation. Due to the reward-based learning systems on social media, users form habits of sharing information that attracts others' attention. Once habits form, information sharing is automatically activated by cues on the platform without users considering response outcomes such as spreading misinformation. As a result of user habits, 30 to 40% of the false news shared in our research was due to the 15% most habitual news sharers. Suggesting that sharing of false news is part of a broader response pattern established by social media platforms, habitual users also shared information that challenged their own political beliefs. Finally, we show that sharing of false news is not an inevitable consequence of user habits: Social media sites could be restructured to build habits to share accurate information.

misinformation | habits | *Facebook* | social media | outcome insensitivity

The online sharing of misinformation has become a worldwide concern with serious economic, political, and social consequences. Most recently, misinformation has hindered acceptance of COVID-19 vaccines and mitigation measures (1). Misinformation can be defined in various ways (2), and in the present research, we focus on information content that has no factual basis (i.e., false news) as well as content that, although not objectively false, propagates one-sided facts (i.e., partisan-biased news). Such misinformation changes perceptions of and creates confusion about reality (3, 4). What drives the online spread of misinformation?

One answer is that people often lack the *ability* to consider the veracity of such information (i.e., limited reflection, inattention). For example, false claims may seem novel and surprising and thereby activate emotional, noncritical processing (5). Furthermore, older people and those with weaker or less critical reflection tendencies may fail to detect the veracity of information and thus be less discerning in their sharing (6, 7). A second account is that people are *motivated* to evaluate news headlines in biased ways that support their identities (i.e., motivated reasoning) (8). In illustration, rumor cascades in online social platforms are most marked when shared within homogenous and polarized communities of users (9). *Partisanship* may especially influence evaluations of information veracity (8, 10, 11), and false information is more likely to be shared by more conservative users (12). Although these analyses identify distinct predictors of false news acceptance, they all imply that people would spread less false information if only they were sufficiently able or motivated to consider the accuracy of such information and discern its veracity (13, 14).

These personal limitations and motivations, although widely studied, may not be the only mechanisms behind the sharing of false news. Misinformation sharing appears to be part of a larger pattern of frequent online sharing of information: People who share greater numbers of false news items also tend to share more true news [in Study 2: $r(283) = 0.73$, and in Study 3: $r(667) = 0.68$] (7). Furthermore, motivation may not fully account for politically-oriented misinformation, given that people who share more politically liberal news also share more conservative news, $r(802) = 0.47$ (7). Such indiscriminate sharing suggests causes beyond a lack of critical reasoning or a my-side bias.

One psychological mechanism that could account for these broader sharing tendencies is the *habits* that people develop as they repeatedly use a social media site. A habit account shifts the focus away from deficits in individual users onto the patterns of behavior that are learned within the current structures of social media sites. It locates the control of misinformation not in users' recognition of accuracy but instead directly in the structure of online sharing built into most social platforms (12). By tracing misinformation to the broader systems that support its spread, we propose a *system-frame* approach that contrasts with an *individual-frame* focused on fixing user deficiencies (15).

Habits form when people repeat a rewarding response in a particular context and thereby form associations in memory between the response and recurring context cues (16, 17).

Significance

Misinformation is a worldwide concern carrying socioeconomic and political consequences. What drives its spread?. The answer lies in the reward structure on social media that encourages users to form habits of sharing news that engages others and attracts social recognition. Once users form these sharing habits, they respond automatically to recurring cues within the site and are relatively insensitive to the informational consequences of the news shared, whether the news is false or conflicts with their own political beliefs. However, habitual sharing of misinformation is not inevitable: We show that users can be incentivized to build sharing habits that are sensitive to truth value. Thus, reducing misinformation requires changing the online environments that promote and support its sharing.

The simple repetition of social media use is especially likely to form habits. For example, habits strengthen through repeated posting on *Facebook* (18) and repeated sharing of moral outrage on *Twitter* (19). Once habits form, perception of context cues automatically brings the practiced response to mind, and people respond with limited sensitivity to outcomes such as misleading others or acting contrary to personal beliefs (20). Indeed, habits have been found to persist despite conflicting attitudes and social norms (21, 22). Consistent with this analysis, interventions to correct the spread of misinformation by priming accuracy mindsets appear to be short-lived, as old sharing habits apparently continue to be activated (23).

The present research tests whether the habits that people form through repeated use of social media extend to sharing of information regardless of its content. If sharing is a habitual response to platform cues, it will be triggered with minimal deliberation about response outcomes. In four studies ($N = 2,476$), we test whether the strength of habits to share information online increases people's sharing of all kinds of information (reflected in a main effect of habit strength), and whether strongly habitual users are less discerning (a smaller gap between true and false headlines) and less sensitive to partisan bias (a smaller gap between politically concordant and discordant headlines) in sharing information compared with those with weaker habits (reflected in an interaction between habit strength and the type of information shared). In each study, we also assessed how much sharing of false news reflects partisan bias and lack of critical reasoning.

In brief, participants in all studies had a *Facebook* account and chose to share or not share a series of 16 news headlines by pressing a simulated *Facebook* share button displayed below each headline. In Studies 1 and 2, participants saw 8 true and 8 false headlines (see Fig. 1 for examples and *SI Appendix*, section 1 for all headlines); headlines were adapted from Pennycook et. al (24) with new headlines added from mainstream news sources (true headlines) and third-party fact checkers (false headlines). In each study, participants completed a number of additional measures including the strength of their social media sharing habits, which was assessed in two ways: 1) the frequency of past sharing on *Facebook* (25), an antecedent to habit formation, and 2) the automaticity of their sharing [Self-Report Behavioral Automaticity Index (SRBAI), 26; validated for social media use: 15, 27], a consequence of habit formation. As expected, these two measures of habit strength proved to be substantially correlated, $r(839) = 0.58$, $P < 0.001$ (Study 2). All hypotheses (except Study 1) were preregistered at aspredicted.org (See

Materials and Methods for all measures and validation of SRBAI scale for social media).

Initial Test of Habit Effects on Misinformation Sharing: Study 1

In our initial survey of 200 online participants (*SI Appendix*), more true headlines were shared (32%) than false ones (5%), $b = 0.65$, 95% CI [0.29, 1.02], $z = 3.51$, $P < 0.001$. Furthermore, consistent with our primary hypothesis, participants with stronger habits shared more headlines, $b = 0.80$, 95% CI [0.63, 0.96], $z = 9.44$, $P < 0.001$. Even more important, participants with stronger habits were less discerning about headline veracity (Headline Veracity x Habit interaction, $b = -0.27$, 95% CI [-0.44, -0.10], $z = -3.20$, $P = 0.001$). As shown in Fig. 2, those with stronger habits (+1 *SD*) shared a similar percentage of true ($M = 43\%$) and false headlines ($M = 38\%$), odds ratio (OR) = 1.30, 95% critical ratios (CR) = [0.88, 1.91], $z = 1.34$, $P = 0.185$, $d = 0.15$, whereas those with weaker habits (–1 *SD*) shared a greater percentage of true ($M = 15\%$) than false headlines ($M = 6\%$), OR = 2.85, 95% CR = [1.76, 4.62], $z = 4.25$, $P < 0.001$, $d = 0.58$. Thus, weak habit participants were 3.9 times more discerning in their sharing than strong habit ones. Furthermore, the habit effects maintained in models that included individual difference measures of critical reflection and partisanship (*SI Appendix*, section 3).

We also calculated the contribution of highly habitual sharers to spread of misinformation. In fact, the 15% most habitual sharers were responsible for 37% of the false headlines that were shared in this study. Thus, habitual users were responsible for sharing a disproportionate amount of false information. Fig. 3 displays this effect along with the impact of individual differences in cognitive reflection (as measured by need for cognition) and political partisanship (conservatism); all effect sizes based on all studies can be found in *SI Appendix*, section 4).

Considering Accuracy Does Not Deter Habitual Sharing: Study 2

One potential explanation for habitual sharing is that people share indiscriminately when they are not able or motivated to assess the accuracy of information. In this account, habitual sharers spread misinformation just because strong habits limit attention to accuracy. To test this possibility, we examined whether highlighting accuracy prior to sharing would reduce the habitual spread of misinformation and increase sharing discernment (4). We randomly assigned 839 online survey participants to a) first choose

Example True Headline	Example False Headline

BUSINESSINSIDER.COM
Trump spent the past 2 years slashing the government agencies responsible for handling the coronavirus outbreak

HEALTHIMPACTNEWS.COM
Coconut oil's history in destroying viruses, including Coronaviruses.

Fig. 1. Example of true and false headlines, which were presented in *Facebook* format in Studies 1 and 2.
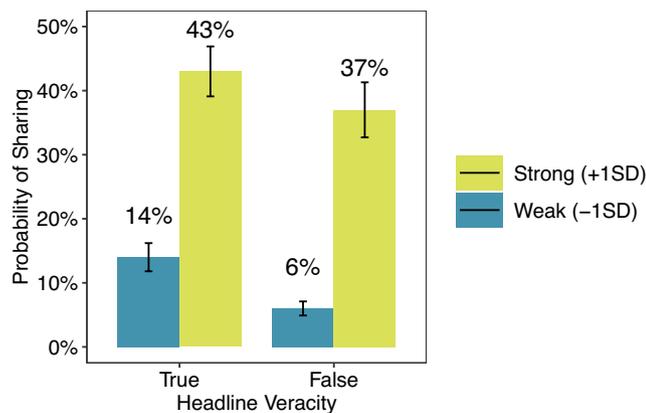
Fig. 2. Study 1: Probability of sharing headlines as a function of headline veracity and habit strength. Note that habit strength was represented as a continuous variable in the analysis. Error bars reflect 95% CI.

to share or not all of the 16 headlines and then determine the accuracy of each or b) first determine all headlines' accuracy and then choose to share each or not. Thus, participants first completed all of the accuracy judgments or all of the sharing choices.

In general, rating accuracy first did not increase the discernment of strongly habitual users any more than less habitual ones. That is, as illustrated in Fig. 4 *A* and *B*, the interaction among headline veracity, question order, and habit strength did not approach significance (*P* > 0.5). Furthermore, the results remained the same when models included covariates (*SI Appendix*, section 9). As predicted, and replicating Study 1, strongly habitual participants continued to share with limited sensitivity to the veracity of headlines, as reflected in the Headline Veracity x Habit Strength interaction (*b* = –0.23, 95% CI = [–0.32, –0.15], *z* = –5.29, *P* < 0.001). Habitual participants shared 42% of the true headlines and 26% of the false headlines, OR = 2.11, 95% CR = [1.88, 2.37], *z* = 12.70, *P* < 0.001, *d* = 0.41. Less habitual participants shared 13% of the true headlines and 3% of the false headlines, displaying a more pronounced tendency to be discerning in their sharing, OR = 3.98, 95% CR = [3.43, 4.62], *z* = 18.27, *P* < 0.001, *d* = 0.76. Similar to the results of Study 1, weak habit participants were 1.9 times more discerning than strong habit ones.

Nonetheless, the accuracy manipulation was broadly effective. Replicating earlier findings of Roozenbeek et al. (23) and Pennycook et al. (4), rating accuracy first reduced participants' sharing of false headlines ($M_{accuracy\ first}$ = 9%; $M_{sharing\ first}$ = 13%),

OR = 1.63, 95% CR = [1.16, 2.31], *z* = 3.64, *P* = 0.002, but not their sharing of true ones ($M_{accuracy\ first}$ = 25%; $M_{sharing\ first}$ = 27%), *P* > 0.8, as reflected in the Headline Veracity x Question Order interaction (*b* = 0.40, 95% CI = [0.15, 0.64], *z* = 3.15, *P* < 0.001). That is, rating accuracy first increased sharing discernment 1.49 times (*d* = 0.70) over that obtained when sharing first (*d* = 0.47). Although our accuracy rating intervention manipulation was stronger than prior tests (which involved only a single accuracy rating), our effect was comparable to that reported by Roozenbeek et al. (23) of 1.4 times increased discernment (accuracy prime *d* = 0.14 vs. no prime *d* = 0.10) and smaller than the original report by Pennycook et al. (4) of 2.8 times (accuracy prime *d* = 0.14 vs. no prime *d* = 0.05). Yet another effect of rating accuracy first was to significantly reduce subsequent sharing rates, *b* = –0.49, 95% CI = [–0.75, –0.23], *z* = –3.64, *P* < 0.001, presumably by making participants generally more cautious.

Thus, highlighting accuracy proved useful in reducing the spread of misinformation but not among the most habitual users. Echoing the first study, 15% of the strongest habit participants were responsible for sharing a disproportionate amount of misinformation—39% across all experimental conditions (habit estimated from SRBAI, 30% with habit estimated from past frequency).

Habitual Sharing Extends to Politically Discordant Information: Study 3

If much of the information shared on social media is done habitually, then habitual social media users may share headlines even when the content conflicts with their own political views. Such an effect would be consistent with evidence that habits in other domains are cued and often performed with little regard for response outcomes (a.k.a., habit insensitivity to reward; 17, 20). To test this idea, we examined whether people with strong sharing habits would be less sensitive to partisan bias and share information that did not align with their political views.

The survey of 836 online participants used a design similar to the prior study, except that: a) participants were exposed to headlines that differed in politics (i.e., liberal or conservative instead of accuracy), and b) participants rated the headlines' politics (rather than accuracy) in one of the two blocks. In this study, all headlines were factually correct but reflected a partisan bias in line with our definition of misinformation. We coded headlines as concordant or discordant based on the headline's political orientation and individual's stated political orientation (e.g., a conservative headline for a conservative participant was coded as
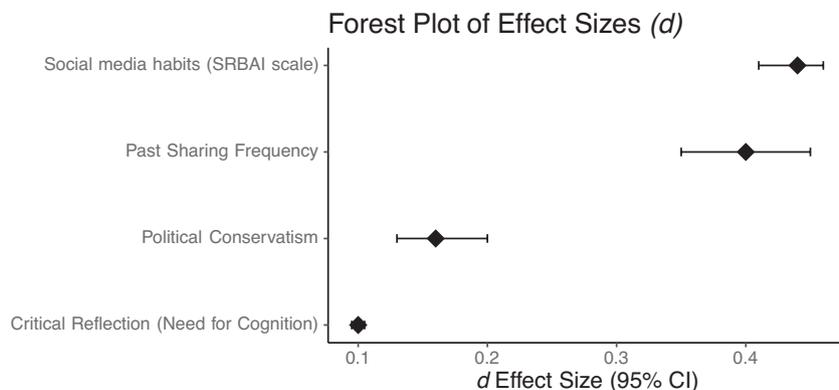

Fig. 3. Effect size (*d*) comparison among each of the predictors (Social Media Habits, Past Sharing Frequency, Political Conservatism, Critical Reflection) of sharing false news.
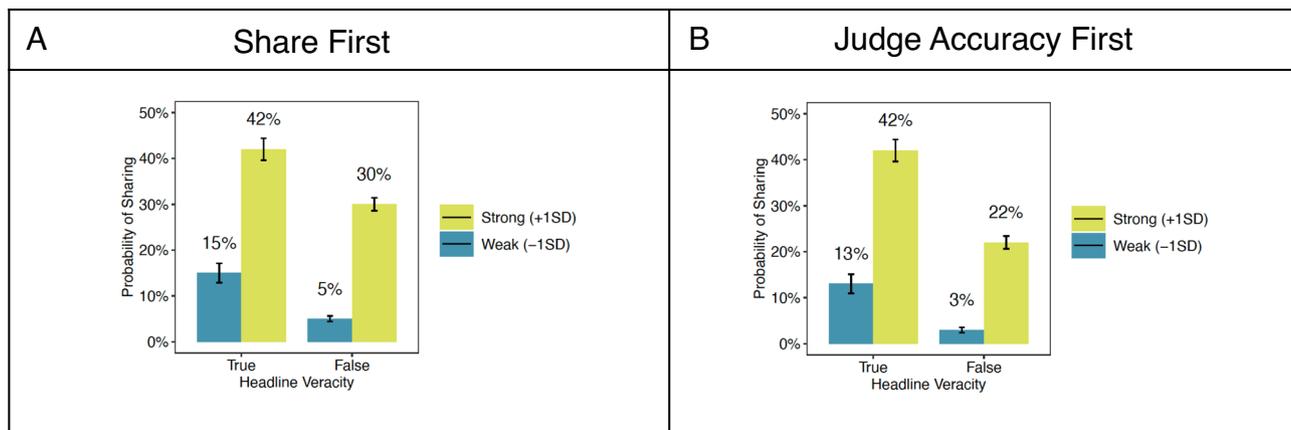
Fig. 4. Study 2: Probability of sharing headlines as a function of headline veracity and habit strength. In the sharing first condition (4*A*), weak habit participants were 2.2 times more discerning than strong habit ones. In the judge accuracy first condition (4*B*), this difference reduced slightly to 1.7 times; however, the three-way interaction did not approach significance. Note that habit strength was represented as a continuous variable in the analysis. Error bars reflect 95% CI.

concordant). We did not include those who rated their political orientation in the middle of the scale (N = 162).

Suggesting a partisan, or myside, bias participants shared more headlines concordant than discordant with their own political views, b = 1.75, 95% CR = [1.49, 2.01], z = 13.38, P < 0.001. Again, participants with stronger habits shared more headlines, b = 0.87, 95% CR = [0.68, 1.05], z = 9.06, P < 0.001. More importantly as shown in Fig. 5*A*, weak habit participants (−1 SD) shared more concordant (M = 21%) than discordant headlines (M = 3%), OR = 9.22, 95% CR = [5.38, 16.96], z = 14.59, P < 0.001, d = 1.22. Strongly habitual participants (+1 SD) showed a less pronounced tendency to share more concordant (M = 45%) than discordant headlines (M = 16%), OR = 4.27, 95% CR = [2.36, 6.68], z = 13.26, P < 0.001, d = 0.80 (interaction between headline concordance and habit strength was significant: b = −0.42, 95% CR = [−1.34, −2.16], z = −4.74, P < 0.001). The three-way interaction among headline concordance, habit strength, and question order was not significant (Fig. 5 *A* and *B*). Results held when we included covariates in the model (*SI Appendix*, section 16). This finding is in line with the idea that habitual sharing is distinct from goal-based sharing and leads people to share headlines that conflict with personal goals and norms.

In sum, strongly habitual sharers showed less partisan bias in their sharing choices than less habitual sharers. Even when rating the political orientation of headlines before sharing,

habitual sharers were less discriminating in the politics of what they shared. Thus, habitual *Facebook* sharers do not just share more false information (see Studies 1 and 2), but also share more information inconsistent with their own political beliefs. Of course, strong partisans may form habits to share information only from highly politicized sources (e.g., Truth Social). The relatively moderate political headlines we used in this study do not capture these extreme patterns. Nonetheless, our findings reveal that sharing misinformation is part of a broader response pattern of insensitivity to informational outcomes that results from the habits formed through repeated social media use.

Social Media Sharing Habits Depend on the Site's Reward Structures: Study 4

The final study identified the reward mechanisms that establish online sharing habits. If habits develop through instrumental learning as people repeatedly respond to rewards (such as likes and comments), then online contexts can be constructed with different reward contingencies to build different sharing habits. In this account, users are forming habits as they share and, following the standard paradigm from animal learning research, these sharing patterns should persist when contingencies change and rewards are no longer available (28).
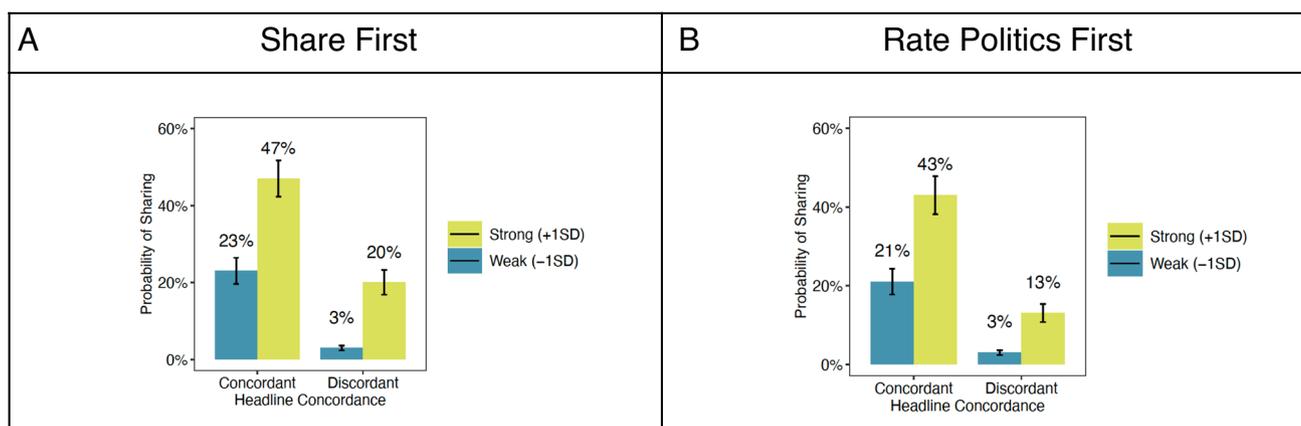

Fig. 5. Study 3: Probability of sharing headlines as a function of political concordance and habit strength. In the sharing first condition (5*A*), weak habit participants showed 1.8 times more partisan bias than strong habit ones. In the rate politics first condition (5*B*), this difference reduced to 1.3 times; however, the three-way interaction did not approach significance. Note that habit strength was represented as a continuous variable in the analysis. Error bars reflect 95% CI.
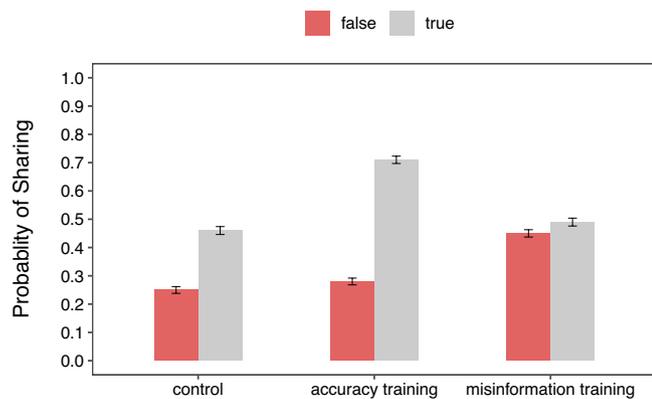
Fig. 6. Study 4: Probability of sharing headlines in training phase as a function of headline veracity and reward training condition. Error bars reflect 95% CI.

Specifically, participants were randomly assigned to one of three *training* conditions. In the *accuracy training* condition, participants received a reward for each of 80 trials in which they shared true information or did not share false information. In the *misinformation training* condition, participants received a reward for sharing information that was false or for not sharing information that was true. Finally, in the *control* condition, participants received no reward. Then, during *test trials*, participants chose to share or not share 16 new headlines (8 true and 8 false) "as they normally would," without any rewards.

During training, the rewards were effective in changing participants' sharing patterns (Fig. 6), with accuracy training increasing sharing discernment (72% true vs. 26% false: OR = 7.12, 95% CR = [6.57, 7.70], z = 48.47, P < 0.001, d = 1.08) and misinformation training reducing sharing discernment (48% true vs. 43% false: OR = 1.20, 95% CR = [1.12, 1.29], z = 4.82, P < 0.001, d = 0.10) vs. control training (45% true vs. 19% false: OR = 3.55, 95% CR = [3.24, 3.87], z = 28.01, P < 0.001, d = 0.70).

As would be expected if the training formed habits, these patterns persisted during the test trials despite the change in reward contingencies. Note that all participants in these analyses had successfully responded to a manipulation check in which they explicitly recognized the absence of rewards during test. As anticipated, prior accuracy training subsequently improved sharing discernment (i.e., a larger gap between true and false news sharing) in the test trials compared with the other conditions. Thus, participants in the accuracy training condition continued to share more true (66%) than false headlines (24%), OR = 12.64, 95% CR = [6.90, 23.17], z = 8.21, P < 0.001, d = 1.40). Also as expected, discernment was lower in the control and misinformation training conditions, although participants still shared more true headlines (control: 40%; misinformation: 54%) than false ones (control: 24%, OR = 4.00, 95% CR = [2.16, 7.40], z = 4.42, P < 0.001, d = 0.76; misinformation: 33%, OR = 3.48, 95% CR = [1.90, 6.37], z = 4.04, P < 0.001, d = 0.69). Thus, sharing discernment was more pronounced in accuracy training (d = 1.40) than control (d = 0.76) or misinformation conditions (d = 0.69).

It is noteworthy that, during test when participants were sharing as they normally would, discernment levels were similar between misinformation training and the control condition, with misinformation training increasing the overall frequency of sharing (b = 0.75, 95% CI = [0.35, 1.15], z = 3.63, P < 0.001; see Fig. 7). Finally, training had comparable influence on the sharing of weak and strong habit participants as measured by our two indices of

habit strength (*SI Appendix*, section 20). Collectively, these results indicate that the reward contingencies in the training trials built habits that carried over to the unrewarded test trials when participants were instructed to share as normal.

Furthermore, our intervention had two critical outcomes: First, given how central user engagement is to the business model of platforms, an important note for application is that our intervention did not reduce the frequency with which users shared. Instead, it increased sharing. Thus, we show that accuracy-based rewards can motivate sharing true information without sacrificing user engagement. Second, our proposed intervention impacted both weakly and strongly habitual users—the ones who are disproportionately responsible for spreading misinformation on social platforms. Thus, this intervention had broad effects.

To further test a hallmark of habit formation—repeated responding that does not depend on the value of response outcomes (20)—we evaluated whether the reward manipulation influenced participants' reported *goals* in the form of desired outcomes of sharing. Consistent with the idea that habits are activated without supportive goals, habit training influenced responding without changing goals. Specifically, after training, all participants across the conditions indicated that it was moderately important to share information that supported their political views ($M_{accuracy}$ = 4.26, $M_{control}$ = 4.31, $M_{misinformation}$ = 4.63; P > 0.12), highly important to share information that was truthful and accurate ($M_{accuracy}$ = 6.64, $M_{control}$ = 6.46, $M_{misinformation}$ = 6.58; P > 0.20), and less important to share information that attracted others' attention so as to be widely read ($M_{accuracy}$ = 3.01, $M_{control}$ = 3.27, $M_{misinformation}$ = 2.98; P > 0.25). Thus, the training directly influenced sharing without altering respondents' goals.

It is interesting to note that stronger habits to share on social media *prior to the training* were associated with greater goals to attract others' attention and get widely read (b = 0.57, 95% CI = [0.47, 0.69], t(599) = 10.12, P < 0.001) and greater goals to support one's political views (b = 0.18, 95% CI = [0.05, 0.33], t(599) = 2.66, P = 0.008). Also, stronger habits were *negatively* correlated with the goal of sharing true information (b = –0.17, 95% CI = [–0.24, –0.09], t(599) = 4.38, P < 0.001). Although we did not predict these relations, they suggest that a significant driver of habit formation on social media is the goal to share sensationalist, surprising content that attracts attention with little regard for its accuracy (2, 5).
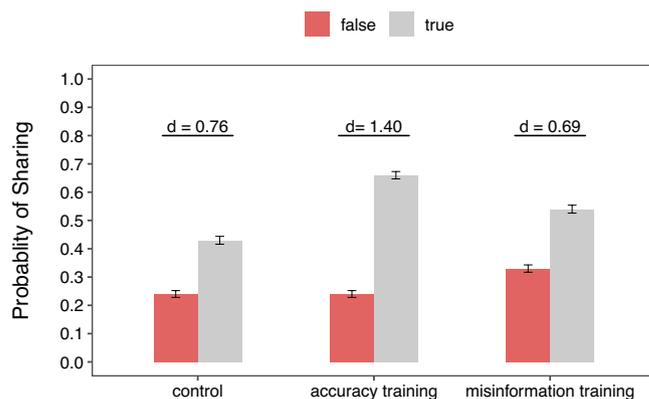

Fig. 7. Study 4: Probability of sharing headlines after rewards ended (test trials) as a function of headline veracity and reward training condition. d indicates the effect size of discernment between true and false information sharing. Error bars reflect 95% CI.

Discussion

Our findings are consistent with a habit-based account of misinformation transmission on social platforms. In the present research, *Facebook* users often shared misinformation out of habit, reacting automatically to the familiar platform cues in which headlines were presented in the standard manner (i.e., *Facebook* format of a photograph, source, headline) with the sharing response arrow underneath. In responding to these cues, habitual sharers gave little consideration to the informational consequences of what they shared.

Our first study revealed the limited discernment of habitual sharers, in that they shared both false and true news. Priming accuracy concerns prior to sharing had only a modest impact on the discernment of everyone and did not ameliorate high habitual sharing of misinformation (Study 2). Next, we showed that habitual sharing of misinformation is part of a broader pattern of insensitivity to informational outcomes. Specifically, habitual sharers showed less partisan bias by sharing politically discordant news that did not align with their own personal views as well as concordant news (Study 3). Finally, our fourth experiment simulated online habit formation with a task that rewarded sharing either accurate news or misinformation. It demonstrated the important point that habitual sharing does not inevitably amplify false news, and forming habits to share true information is also possible.

In addition to habit, information sharing proved to be influenced by political bias and critical reflection. However, when we estimated the influence of each of these other factors in our studies, habits consistently emerged as the major influence. One implication is that many of the current proposals to reduce misinformation online are unlikely to have the desired lasting impact. Once sharing habits have formed, they are relatively insensitive to changing goals through accuracy primes (4) and the display of metrics such as how many people scrolled over a post (29). Thus, existing individual-frame interventions remain relatively ineffective for the habitual sharers who are most responsible for misinformation spread on these platforms.

We recognize that habitual users are integral to social media sites' ad-based profit models (27), and thus these sites are unlikely to create reward structures that encourage thoughtful decisions that impede habits. However, social media reward systems built to maximize user engagement are misaligned with the goal of promoting accurate content sharing, especially among regular, habitual users. Our results show the impact of this now outdated reward system on news sharing, along with the potential of changing the reward structure of the platform to match its current role in the world-wide dissemination of information. Furthermore, our results suggest ways to maintain, and even increase, user engagement.

Thus, our results highlight the restructuring of reward systems on social media to promote sharing of accurate information instead of popular, attention-getting material. Current algorithms rely on engagement (i.e., likes, comments, shares, followers) as a quality signal and rank the most "liked" content at the top of users' feeds. However, given that algorithms that prioritize the popularity of information lower the overall quality of content on a site (30), algorithmic de-prioritization of unverified news content is needed. De-prioritizing this content would mean effectively preventing potential misinformation from being viewed and shared until it is approved by a moderation system. Furthermore, verification would only be needed for the subset of content with high rates of engagement from a small group of users (i.e., more likely to be biased) and could be performed by existing, neutral fact-checking organizations.

Another useful intervention could build on the current structure and design of social media platforms to disrupt habitual news sharing or increase *friction* on it. For example, prior work has shown that updates in *Facebook* platform cues disrupt habitual posting (18, 27). For sharing, adding additional buttons (i.e., disengagement buttons) such as fact-check or skip may disrupt the automatic responses that get activated when users are exposed to only liking and sharing choices. A less habit-forming system would force users to constantly adjust to changes in the process of sharing posts (or better yet, remove single-click sharing), at least for the subset of information most likely to be false. However, as we noted above, social platforms may resist such disruptive solutions that could reduce overall sharing. Potentially assuaging this concern, the present research shows that overall sharing does not decrease by lighter-touch alterations such as rewarding the spread of truthful information. A comprehensive solution might first disrupt the core group of habitual sharers and then subsequently reward all users for sharing accurate information.

Materials and Methods

We preregistered all hypotheses, primary analyses, and sample sizes (except Study 1). Participants provided informed consent, and our studies were approved by the University of Southern California, Institutional Review Board Protocol no. UP-17-00337 and Yale Human Subject Committee, Institutional Review Board Protocol no. 2000033417.

Participants. Study 1 involved 200 participants (38% female, age: range = 21 to 72 y, $M = 40.13$, $SD = 12.15$) from Amazon's MTURK who owned a *Facebook* account, typically accessed *Facebook* through their computers, and gave study responses on their computer. Study 2 was preregistered (#89446) and involved 839 participants with these selection criteria from Amazon MTURK (49% female, age: range = 18 to 78 y, $M = 40.34$, $SD = 12.57$). Study 3, preregistered (#89009), recruited 836 participants with these features from Amazon's MTURK (49% female, age: range = 18 to 89 y, $M = 40.49$, $SD = 12.80$). Study 4, preregistered (#104694), recruited 601 Prolific participants (49% female, age: range = 19 to 84, $M = 40.62$, $SD = 12.90$).

Analysis Strategy. In each experiment, participants chose to share or not share 16 not-before-presented news headlines. We analyzed the data using mixed-effects models with both participants and stimuli as random effects (31). Our power analysis was computed with random targets and participants (32). We incorporated our mixed design (as headline veracity was within-subjects) into our power calculations. With 16 stimuli that varied within-subject, a medium effect size (d) of 0.50, and common assumptions about variance-partitioning components, power of 0.80 could be achieved with 200 participants, which was our minimum sample size. For Studies 2 and 3, with multiple experimental conditions, we increased the sample size to 400 per condition. With 16 stimuli and a sample size of 400 per condition, these studies have a power of at least 0.75 to detect an effect (d) of approximately 0.45. All data were collected in each experiment before conducting any analyses.

All statistical tests were performed using R software (33). Specifically, we fit logistic mixed-effects models with functions from the *lme4* package (34). Participants and headlines were included as random-intercept effects and the predictor variable of headline veracity as a fixed effect.

$$\text{logit Share}_{ij} = b_0 + b_1 \text{HV}_{ij} + b_2 \text{habit\_strength}_i + b_3 \text{HV}_{ij} :$$

$$\text{habit\_strength}_i + (\mu_{0,j} + \mu_{0,i} + \varepsilon_{ij}),$$

in which i refers to participants, j refers to headlines within veracity condition, share is the outcome variable, headline veracity (HV) is the headline veracity, and habit strength is the strength of each participant's habit. This model contains the random participant intercept effect $\mu_{0,i}$, the random headline intercept effect $\mu_{0,j}$, and the residual error term ε_{ij}.

Design. The headlines in Studies 1, 2, and 4 were partially adapted from Pennycook et al. (24) and listed in *SI Appendix*, section 1. We replaced any outdated false headlines using Snopes.com, a third-party fact-checking website, and true headlines with mainstream news sources (e.g., National Public Radio (NPR), The Washington Post). Each headline was presented in the format of a *Facebook* post (i.e., picture accompanied by a headline, byline, and source). Study 3 used headlines published by news sources (e.g., NPR, The Washington Post) that, through our pretesting, were identified as liberal or conservative (see *SI Appendix*, section 11 for headlines and *SI Appendix*, section 12 for the pretest).

The design in Study 1 included a within-subjects factor, *Headline Veracity* (true vs. false) to reflect the 8 true and 8 false headlines read by each participant, and between subjects *Sharing Habit Strength* (weaker to stronger), assessed on a continuous scale. In all studies, the hypothesized effects were maintained when the analyses were recomputed including additional measures as control variables (i.e., politics, need for cognition, impulsiveness, goals for using social media, need to conform, age, gender). The 200 participants, each of whom responded to 16 trials, yielded 3,200 observations.

Study 2 involved a mixed design, with a within-subjects factor of *Headline Veracity* (true vs. false) and 2 between-subjects factors of *Question Order* (rate headline accuracy first vs. make sharing choice first) and *Sharing Habit Strength* (weaker to stronger). The 839 participants each provided 16 accuracy ratings and 16 sharing responses, yielding 13,376 observations for each measure. Initial analysis on participants' accuracy judgments indicated that they correctly perceived headline veracity, with a greater percentage of participants categorizing headlines as accurate when they were true ($M = 77\%$) than false ($M = 33\%$), $b = 2.22$, 95% CI = [1.38, 2.84], $z = 5.75$, $P < 0.001$.

In Study 3, the mixed design included the within-subjects factor of *Headline Politics* (liberal vs. conservative), and 2 between-subjects factors of *Question Order* (rate headline politics first vs. make sharing choice first) and *Sharing Habit Strength* (weaker to stronger). The 674 participants provided 16 politics ratings and 16 sharing responses, yielding 10,784 observations for each measure (excluding 162 participants who indicated a political orientation of the scale midpoint and were excluded from the analyses). Initial analyses on participants' politics ratings revealed that they correctly perceived the headlines' political alignment as liberal ($M = 2.98$) or as conservative ($M = 4.81$), $b = 1.90$, 95% CI = [1.55, 2.26], $t(16.85) = 10.55$, $P < 0.001$, $\eta_P^2 = 0.87$.

Study 4's between-subjects design involved three *Reward* conditions (accuracy training, misinformation training, and no-reward control). Participants first completed training trials in which they pressed the letter "e" or "i" to *share* or *not share* 80 headlines that individually appeared on the screen. After each choice, participants saw a feedback screen. In the accuracy training [misinformation training] condition, if they shared an accurate [false] headline or did not share a false [accurate] one, they saw "won +5 points." If they shared a false [true] headline or did not share a true [false] headline, they saw "won 0 points." No feedback was given in the no-reward condition. Following training, all participants were informed that they won a total of 345 points, which qualified them for an actual lottery to win an additional $20. Then, participants made 16 sharing decisions as they would typically do when on *Facebook* and did not receive rewards. One hundred and twenty respondents failed the no-rewards manipulation check and were removed from the analyses, in line with

our preregistration. Including these participants did not alter the results (reported in the *SI Appendix*, section 19).

Measures.

Sharing choice. In all studies, participants chose to share or not the headline, as if they saw it on *Facebook*.

Accuracy judgment. In Study 2, participants indicated whether each headline was true or false (*true* = 1, *false* = 0).

Judgment of political orientation. In Study 3, participants rated how liberal or conservative each headline was (1 = *extremely liberal*, 4 = *neither*, 7 = *extremely conservative*).

News sharing habit (SRBAI) (26). In all studies, habit strength was measured with four 7-point scales (1 = *never* to 7 = *always,* adapted from the self-report habit index, (36): "*Sometimes I start sharing news on Facebook before I realize I'm doing it,*" "*Sharing news on Facebook is something I do without thinking,*" "*Sharing news on Facebook is something I do automatically,*" and "*Sharing news on Facebook is something I do without having to consciously remember.*" The items were averaged into a composite measure of habit strength ($\alpha = 0.89$). In a set of validation studies (27), it was found that frequent use was largely habitual use. That is, the frequency with which participants posted on a site was significantly related to the automaticity with which they did so (SRBAI), $r(124) = 0.46$, $P < 0.001$, for *Twitter*, and $r(60) = 0.50$, $P < 0.001$, for *Facebook*.

Past frequency of sharing. In Studies 2 and 3, as an alternative measure of habit strength, participants reported their past frequency of sharing news headlines on *Facebook* (7 = *several times a day*, 6 = *daily*, 5 = *several times a week*, 4 = *weekly,* 3 = *monthly*, 2 = *rarely*, 1 = *never*).

Participants' political orientation. In all studies, after responding to all the headlines, participants indicated their political orientation on a 9-point scale (1 = *extremely liberal* to 9 = *extremely conservative),* adapted from ref. 36.

Need for cognition (37). In Studies 1, 2, and 3, on 5-point scales (1 = *not at all characteristic of me* to 5 = *extremely characteristic of me)*, participants completed the short need for cognition scale, which included 6 items such as: "*I would prefer simple to complex problems*" and "*I really enjoy a task that involves coming up with new solutions to problems.*" The items were averaged into an index ($\alpha = 0.91$).

Social media goals. In Study 4, participants responded to three goal importance questions (1 = not at all, 7 = extremely): How important is it to you that the information you provide on *Facebook* 1) supports your political views, 2) is truthful and accurate, and 3) attracts others' attention and gets widely read.

Attention check. In Study 4, participants responded to the attention check question at the end of the study: Were you able to win points in the second part of the study (1 = yes, 2 = no, 3 = maybe).

1. S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, H. J. Larson, Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* **5**, 337–348 (2021).
2. E. C. Tandoc Jr., Z. W. Lim, R. Ling, Defining "fake news": A typology of scholarly definitions. *Digital Journalism* **6**, 137–153 (2018).
3. M. D. Molina, S. S. Sundar, T. Le, D. Lee, "Fake news" is not simply false information: A concept explication and taxonomy of online content. *Am. Behav. Sci.* **65**, 180–212 (2021).
4. G. Pennycook et al., Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).
5. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **1151**, 1146–1151 (2018).
6. G. Pennycook, D. G. Rand, Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
7. G. Pennycook, D. G. Rand, Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *J. Pers.* **88**, 185–200 (2020).
8. M. Osmundsen, A. Bor, P. B. Vahlstrup, A. Bechmann, M. B. Petersen, Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *Am. Political Sci. Rev.* **115**, 1–17 (2021).
9. M. del Vicario et al., The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559 (2016).
10. J. Roozenbeek et al., Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment Decis. Making* **17**, 547–573 (2022).
11. J. J. Van Bavel et al., Political psychology in the digital (mis)information age: A model of news belief and sharing. *Soc. Issues Policy Rev.* **15**, 84–113 (2021).
12. A. Guess, J. Nagler, J. Tucker, Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* **5**, 45–86 (2019).
13. D. Borukhson, P. Lorenz-Spreen, M. Ragni, When does an individual accept misinformation? An extended investigation through cognitive modeling. *Comput. Brain Behav.* **5**, 244–260 (2022).
14. A. Schnauber-Stockmann, T. K. Naab, The process of forming a mobile media habit: Results of a longitudinal study in a real-world setting. *Media Psychol.* **22**, 714–742 (2019).
15. N. Chater, G. Loewenstein, The i-frame and the s-frame: How focusing on the individual-level solutions has led behavioral public policy astray. *Behav. Brain Sci.*, in press.
16. W. Wood, D. Rünger, Psychology of habit. *Annu. Rev. Psychol.* **67**, 289–314 (2016).
17. B. Verplanken, S. Orbell, Attitudes, habits, and behavior change. *Annu. Rev. Psychol.* **73**, 327–352 (2022).
18. I. A. Anderson, W. Wood, Social motivations' limited influence as habits form: Tests of social media engagement". Psyarxiv. https://psyarxiv.com/q9268/ (2022).

19. W. J. Brady, K. McLoughlin, T. N. Doan, M. J. Crockett, How social learning amplifies moral outrage expression in online social networks. *Sci. Adv.* **7**, eabe5641 (2021).
20. W. Wood, A. Mazar, D. T. Neal, Habits and goals in human behavior: Separate but interacting systems. *Perspect. Psychol. Sci.* **17**, 590–605 (2022).
21. G. Itzchakov, L. Uziel, W. Wood, When attitudes and habits don't correspond: Self-control depletion increases persuasion but not behavior. *J. Exp. Soc. Psychol.* **75**, 1–10 (2018).
22. A. Mazar, G. Itzchakov, A. Lieberman, W. Wood, The unintentional nonconformist: Habits promote resistance to social influence. *Pers. Soc. Psychol. Bull.*, 1–13 (2020).
23. J. Roozenbeek, A. L. J. Freeman, S. van der Linden, How accurate are accuracy-nudge interventions? A preregistered direct replication of Pennycook et al. (2020). *Psychol. Sci.* **32**, 1169–1178 (2021).
24. G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, D. G. Rand, Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychol. Sci.* **31**, 770–780 (2020).
25. D. T. Neal, W. Wood, M. Wu, D. Kurlander, The pull of the past: When do habits persist despite conflict with motives? *Pers. Soc. Psychol. Bull.* **37**, 1428–1437 (2011).
26. B. Gardner, C. Abraham, P. Lally, G. J. de Bruijn, Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *Int. J. Behav. Nutr. Phys. Activity* **9**, 1–12 (2012).
27. I. A. Anderson, W. Wood, Habits and the electronic herd: The psychology behind social media's successes and failures. *Consumer Psychol. Rev.* **4**, 83–99 (2021).
28. O. Perez, D. Dickinson, A theory of actions and habits: The interaction of rate correlation and contiguity systems in free-operant behavior. *Psychol. Rev.* **127**, 945–971 (2020).
29. P. Lorenz-Spreen, S. Lewandowsky, C. R. Sunstein, R. Hertwig, How behavioral sciences can promote truth, autonomy and democratic discourse online. *Nat. Hum. Behav.* **4**, 1102–1109 (2020).
30. F. Menczer, Facebook's algorithms fueled massive foreign propaganda campaigns during the 2020 election – here's how algorithms can manipulate you. *The Conversation.* https://theconversation.com/facebooks-algorithms-fueled-massive-foreign-propaganda-campaigns-during-the-2020-election-heres-how-algorithms-can-manipulate-you-168229 (September 20, 2021).
31. C. M. Judd, J. Westfall, D. A. Kenny, Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *J. Pers. Soc. Psychol.* **103**, 54–69 (2012).
32. J. Westfall, D. A. Kenny, C. M. Judd, Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. General* **143**, 2020–2045 (2014).
33. R Core Team, R: A language and environment for statistical computing. *R Found. Stat Comput.* https://www.R-project.org/ (2022).
34. D. Bates, M. Mächler, B. M. Bolker, S. C. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Software* **67**, 1–48 (2015).
35. B. Verplanken, S. Orbell, Reflections on past behavior: A self-report index of habit strength. *J. Appl. Soc. Psychol.* **33**, 1313–1330 (2003).
36. J. T. Jost, J. Glaser, A. W. Kruglanski, F. J. Sulloway, Political conservatism as motivated social cognition. *Psychol. Bull.* **129**, 339–375 (2003).
37. G. L. de H. Coelho, P. HP. Hanel, J. Wolf, The very efficient assessment of need for cognition: Developing a six-item version. *Assessment* **27**, 1870–1885 (2020).