Monetization of social media engagements increases the sharing of false (and other) news but penalization moderates it

Meysam Alizadeh[1*], Emma Hoes[1] and Fabrizio Gilardi[1]

[1]Department of Political Science, University of Zurich.

*Corresponding author(s). E-mail(s): alizadeh@ipz.uzh.ch;
Contributing authors: hoes@ipz.uzh.ch; gilardi@ipz.uzh.ch;

Abstract

Social media companies are introducing new features for users to monetize engagements, derived from blockchain-based decentralized social media. These steps are potentially worrisome. The monetization of engagements might create incentives to post objectionable content. However, it is unclear to what extent such negative outcomes are likely to occur. To address this question, we first administered a survey that showed that many users have a poor understanding of the mechanisms behind monetization. Second, we conducted a survey experiment to examine the effects of hypothetical monetary incentives. We find that a simple nudge about the possibility of earning money for user engagements increases the willingness to share different kinds of news, including misinformation. The presence of penalties for objectionable posts diminishes the positive effect of monetary rewards on misinformation sharing, but it does not eliminate it. These results have policy implications for content moderation practices if platforms embrace decentralization and monetization.

Keywords: Decentralized Social Media, Misinformation, Polarization, Content Moderation, Content Monetization

1 Introduction

Many decentralized social media (DeSo) platforms are offering monetary incentives to users for their activities on the platform. A DeSo is a platform that does not have a central governing body and operates on a public ledger, usually a blockchain [1]. Users are rewarded with cryptocurrencies, commonly called token or coin (see Appendix S1 for background), for their 1) content creation, 2) content curation, 3) content attention, and 4) community management [2]. For example, *Steemit* is a DeSo, analogous to Reddit, which has a price tag for each post that represents 'Steem tokens' that creators would earn for their achieved user engagement (e.g. up-vote or comment), encouraging them to post quality content [3]. The Steem token can then be exchanged for fiat money or other cryptocurrencies in supported online markets (see [1] for a review of existing DeSo and Appendix S2 for a list of major existing DeSo platforms).

Major social media companies are enacting to DeSo and incorporate similar features. Reddit introduced 'Community Points', which are a measure of reputation to reward users for their activity and quality content (see Appendix S3 for references on this and other similar plans of the major platforms). The points are on the blockchain, which means they can be converted to cryptocurrencies, commonly called token or coin, and be spent as money. In the beta version, for which users are invited to join the waitlist, users can only spend the 'Reddit Coin' on the platform. This includes tipping other users for their posts or comments, purchase special memberships, or reward developers for making tools for a subreddit. There are speculations about Reddit's plan to eventually turn the coins to an Ethereum-based token, which would enable the owners to spend it anywhere.[1] Twitter introduced 'Tips', which allows users to support content by sending fiat money, Bitcoin, or Ethereum to creators. More recently, on the 5th of November 2022, Elon Musk announced in a tweet that "creator monetization for all forms of content" will soon be added to Twitter.[2] Medium introduced its 'Partner Program' in which writers can get paid proportional to the user engagement they achieve.

While the idea of DeSo is not new, the possibility of major platforms embracing monetary incentives of DeSo is both promising and worrisome. On the one hand, it might help to alleviate problems linked to content moderation, (e.g., algorithmic ranking and managing objectionable content), data protection, advertisement pricing, and compensating content developers [2, 4]. On the other hand, it may increase the spread of polarizing content [5] and misinformation [6] on social media if users are rewarded financially for their achieved user engagements, but it may also decrease it if monetary penalties are put in place [7].

Even if major platforms do not incorporate these monetization features, DeSo users are increasingly growing [1, 8], which highlights the necessity of studying the consequence of their monetary incentives. For example, after the

[1] https://cointelegraph.com/news/reddit-to-reportedly-tokenize-karma-points-and-onboard-500m-new-users.
[2] https://www.businessinsider.com/elon-musk-twitter-monetization-model-forms-content-youtube-2022-11.

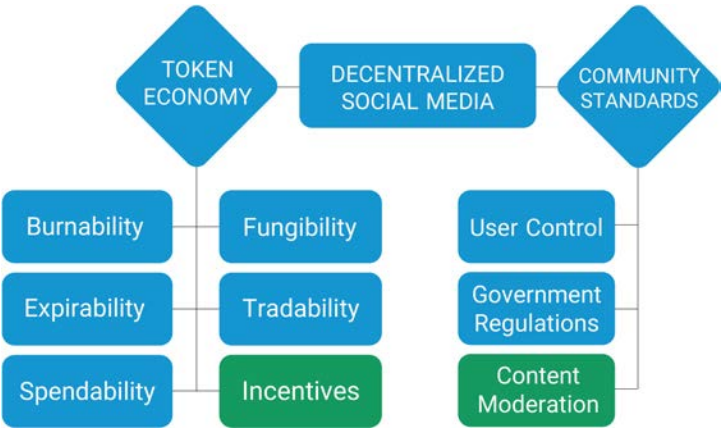

Fig. 1 Main features of a decentralized social media that may affect the posting activity of users. Green boxes denote the parts studied in this paper. Burnability indicates whether a token can burned to terminate a right or revoke access. Expirability refers to whether a token can be expired after some time. Spendability indicates whether a token can be used to gain access to services or pay fees. Fungibility says if a token is interchangeable with other tokens. Tradability illustrates if a token can change ownership within a platform or on secondary markets. Finally, incentives schemes can be divided to two categories: 1) incentive enablers, which refer to what token holder can do with the token, and 2) incentive drivers, that indicate why a token holder engages with incentivized behaviour (e.g. gain reward or reputation).

recent purchase of Twitter by Elon Musk, the users of Mastodon, a decentralized social media platform similar to Twitter, increased by 100,000 in just a few days [9]. In addition, the monetary mechanisms of DeSo are susceptible to misuse by free-riders who just promote others' content [10] or bots [11]. In fact, a recent study showed that more than 16% of transfers of cryptocurrency in Steemit are sent to bots [3].

To better understand how DeSo may affect user behavior, we put forward the framework shown in Figure 1. The first element is "token economy", which is the redistribution of financial benefits to users through a platform-issued token, that conveys a right to its services and participation in its governance, but not necessarily to benefit from its revenue [4]. In general, there are five groups of token characteristics [12]: 1) technical, 2) behavior, 3) inherent value, 4) coordination, and 5) pseudo-archetypes (see [12] for definitions). While many of them are important for investment analysts, we argue that at the current stage of technology and public literacy, only few characteristics from the 'behavior' and 'coordination' domains to be the key features that may affect the posting behavior of users. The second group of DeSo features are linked to community standards (Figure 1). From this perspective, users control over their content and profile, content moderation procedure, and government regulations are critical factors affecting the behavior of users.

While all features of DeSo in Figure 1 can potentially affect user behavior and require research, our focus is on those that are most likely to create

problematic motivations: monetary incentives and content moderation regulations. More particularly, we are interested in studying the effect of rewarding and penalizing social media users with a monetary asset on their willingness to share misinformation as well as neutral and hyperpartisan news headlines. As for content moderation, we consider a specific case that users would loose a portion of their rewards should they post misinformation or hate speech. In short, we thus examine a situation in which users get paid for their achieved user engagements, and further explore under which circumstances this posting behavior changes.

Previous research shows that partisan news is shared more frequently on Twitter [13], political out-group posts are shared about twice as much as as in-group posts on Facebook and Twitter [14], and false news spread faster, farther, deeper, and more broadly than true news on Twitter [15]. However, it it remains unclear whether the implementation of monetary rewards and penalties would bear either positive or negative consequences. While existing research shows that accuracy nudging decreases the probability of sharing fake news [16], and that monetary incentives may serve as such nudges [17], we know little about the effects of monetary incentives on the willingness to share different kinds of content.

As an increasing number of social media platforms are on the brink of incorporating monetary incentives, insights about their potential negative consequences is urgently required for two reasons. First, inaction could exacerbate the spread of misinformation and its costs [18] such as reduction in vaccination intent [19]. Second, there are currently some DeSo applications that have incorporated such incentives [1]. To this end, we propose a framework to understand how DeSo may affect user behavior and conduct 1) an online survey to explore how well these new financial incentives of DeSo are understood by people, and 2) an online survey experiment to examine how the monetary incentives of DeSo might affect the likelihood of posting different kinds of content. We find that 1) most respondents have a mediocre understanding of the mechanisms behind the monetization of engagements, and 2) monetization of engagements increases the sharing of misinformation (as well as other kinds of news), which penalization compensates only partially. These results have policy implications for designing effective intervention strategies to reduce the spread of misinformation [18] and preventing potential misuse of monetary features by state-sponsored coordinated networks on social media [20].

2 Results

2.1 Study 1

We preregistered a target sample of 1,000 representative US participants recruited from Prolific. The survey was conducted between March 8, 2022, and March 9, 2022. In total, 1,067 participants started the survey and 1,022 of them completed it. The final sample (mean age = 45.40) included 517 women,

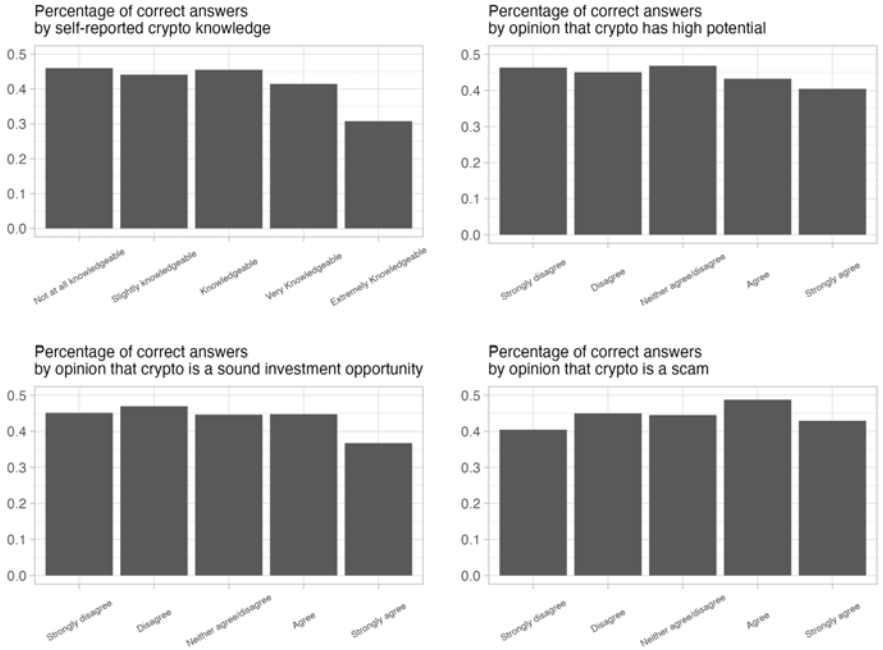

Fig. 2 Fraction of participants who answered 5 questions about cryptocurrency correctly by their self-reported knowledge and opinion about cryptocurrency. In study 1, n = 1,022 American individuals from Prolific were presented with five statements about cryptocurrencies of which two were correct and three incorrect which they had to rate as either as 'right' or 'wrong'. In general, the results show that the more participants think they know about cryptocurrency or have more positive opinion about it, the less likely they answer correctly to questions. (top left) Participants who reported to be 'extremely knowledgeable' about cryptocurrency are less likely to answer correctly to questions. (top right) Participants who reported to be 'strongly agree' with the high potential of cryprocurency are less likely to answer correctly to questions. (bottom left) Participants who reported to be 'strongly agree' that cryptocurrency is a sound investment opportunity are less likely to answer correctly to questions. (bottom right) Participants who reported to be 'strongly disagree' that cryptocurrency is a scam are less likely to answer correctly to questions.

496 men, and 9 other gender identities. In the survey, respondents were presented with five statements about cryptocurrencies of which two were correct and three incorrect which they had to rate as either 'right' or 'wrong' (see Section 4.3 in *Materials and Methods* for the list of questions).

We observe that there is little variation in correct response rates. On average, respondents in the reference groups answered about 46% of the questions correctly, that is about 2.3 questions. The only significant differences (see regression table in Appendix S5) in a negative direction are found for older respondents (65+, relative to 18-24 years old) and —strikingly—the self-indicated most knowledgeable. More specifically, older respondents and respondents who reported to be "extremely knowledgeable" about crypto relative to those reporting to be "not at all knowledgeable" (Fig 2) had a poorer

understanding of crypto. These results suggest that the understanding of crypto is, on average, mediocre, with little variation across most groups. A low degree of knowledge of the mechanisms behind monetization is not a prerequisite for using social media that include those features. However, a lack of knowledge might prevent users for engaging with monetization productively as well as critically.

2.2 Study 2

To provide evidence on the possible effects of the monetary incentives of DeSo on users' willingness to share different kinds of content, we conducted an online survey experiment with $n = 1,500$ participants from a nationally representative U.S. sample recruited from Prolific. The survey was conducted between June 9, 2022, and June 12, 2022. In total, 1,568 participants started the survey and 1,501 of them completed it. The participants were presented with five true neutral, five misinformation, and five true hyper-partisan news headlines about COVID19. Following [21], the headlines were randomly selected from the dataset in [22] (see Section 4.2 in *Materials and Methods*). Participants were then asked about their willingness to share each of fifteen headlines on social media, which is our outcome variable. Participants in the treatment group were shown a set of statements about a hypothetical (but realistic) world in which major platforms reward their users with digital asset for their so-called *reputation points* (such as the Karma point in Reddit), calculated based on their achieved user engagements. While in all treatment conditions participants earn a (hypothetical) fixed reward, the treatments vary across two dimensions: a) whether users could be penalized for posting misinformation or hate speech, and b) whether the calculation of reputation points is explicit (see Section 4.3 in *Materials and Methods*). The experimental design, data collection, and data analysis plan were preregistered (see *Materials and Methods*).

To gain initial insights into whether monetary incentives of DeSo impacts participants willingness to share various types of news headlines, we plot the fraction of headlines that participants across the control and treatments groups said they would consider sharing in Fig 3. Compared to the control group, participants in all four treatment groups show higher sharing intention across all three types of news (12.8 percentage point difference on average). This includes 6.3 percentage point higher sharing intention of misinformation, 13.95 percentage point higher sharing intention of hyperpartisan, and 15.98 percentage point higher sharing intention of neutral news headlines. Participants in the 'penalization' condition show less sharing intention of misinformation compared to those in the 'no penalization' treatment condition (4.81 percentage point difference). However, penalization does not completely eliminate the the positive effect of monetary rewards. Participants in the 'penalization' condition show higher sharing intention of misinformation compared to the control group (3.89 percentage point difference). Whether the calculation of reputation points was made explicit or not does not seem to have any effect on the sharing intention of participants.

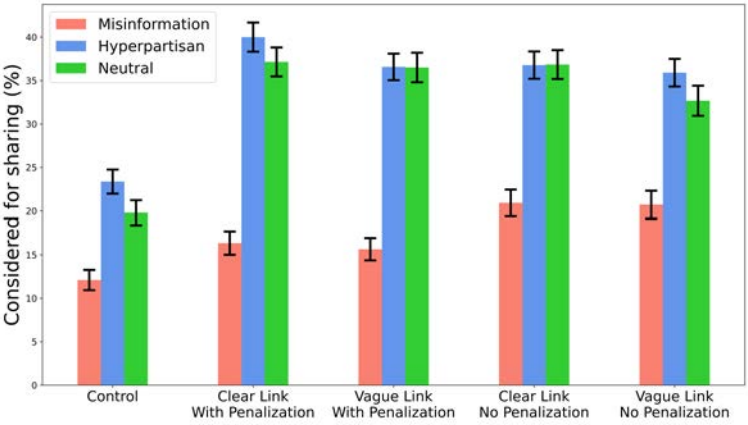

Fig. 3 Fraction of various headlines that participants across the control and four treatment groups said they would consider sharing. In study 2, n = 1,520 American individuals from Prolific were presented with a set of 15 headlines and asked to indicate whether they would consider sharing them on social media. Compared to control group, participants in all treatment conditions show higher sharing intention for all three types of headlines including misinformation and hyperpartisan news. Penalization would decrease the sharing intention of misinformation headlines but not to the level of the control group. Participants in the 'clear link, with penalization' condition show higher sharing intention of hyperpartisan news. Error bars indicate standard errors.

As per our preregistered analysis plan, Figure 4 shows the effects of hypothetical monetary rewards, penalties, and calculation methods of reputation points (clear vs. vague) on the reported willingness to share neutral, hyperpartisan, and misinformation news headlines about COVID-19 on social media. The figure compares different scenarios: (1) rewards but no penalties, relative to the control group (i.e., neither rewards nor penalties); (2) both rewards and penalties, relative to the control group; (3) rewards but no penalties, relative to both rewards and penalties; (4) clear link, relative to vague link, between rewards and user engagements (no penalties in both scenarios); and (5) vague link, relative to clear link, between rewards and user engagements as well as penalties and posting behavior.

Taken together, we find that monetary rewards for user engagements increase the willingness to share all kinds of news including misinformation, and that monetary penalties for such behavior decrease the positive effect of rewards on the sharing of misinformation, but they do not eliminate it. More specifically, we highlight four sets of results. First, in the absence of penalties, monetary rewards for user engagements increase the willingness to share all kinds of news, including misinformation. Respondents are about nine percentage points more likely to share misinformation when they are rewarded monetarily for the user engagements their posts receive, relative to respondents for which no such rewards exist ($coef = 0.089, se = 0.015, p = 0.000$). Second, the same effect holds if respondents can be penalized for sharing misinformation, although effect size decreases to about four percentage points

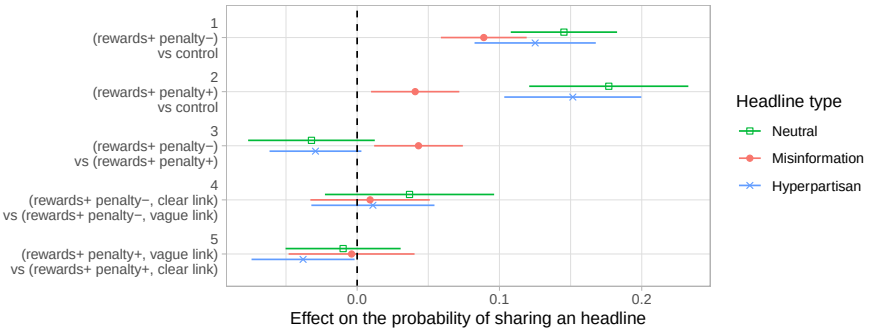

Fig. 4 Point estimates from OLS regressions with 95% CI. Effect of monetary rewards for user engagements and penalties for problematic sharing behavior, as well as of the clarity of the link between rewards/penalties and user engagements/behavior, on the willingness to share neutral and hyperpartisan news, and misinformation. Without penalties, rewards increase the sharing intention of misinformation by about nine percentage points, compared to the control group. With penalties, the effect decreases to about four percentage points. The clarity of the link between rewards/penalties and user engagements/behavior has no effect.

($coef = 0.041, se = 0.016, p = 0.03$). Third, the effect of penalties is confirmed when comparing respondents within the treatment group, where rewards are in place. Given the presence of rewards, respondents that cannot be penalized for sharing misinformation are about four percentage points more likely to share misinformation than those who can be subject to penalties ($coef = 0.043, se = 0.016, p = 0.02$). Fourth, it does not matter whether respondents are told how exactly rewards and penalties are calculated. The clarity of the link between rewards and user engagements, and between penalties and sharing behavior, has no effect on the willingness to share content of any kind. As a robustness test, we re-estimated all models excluding those respondents who failed the manipulation check (see Fig S.2 in Appendix). The results are unchanged, with the exception of the effect of rewards and penalties relative to the control group. The point estimate remains positive but is smaller and not statistically significant ($coef = 0.026, se = 0.016, p = 0.355$).

3 Discussion

The decentralization of social media will introduce a new set of incentives for users, which may affect their online behavior in both positive and negative ways. Given the existence of decentralized social media platforms and their increasing growth, and that most major social media platforms have announced plans in this direction, there has been speculation but little evidence regarding possible outcomes. Our results, based on an online survey experiment, are of course limited (see [23] for a discussion), but put forward evidence on the possible consequences of social media decentralization.

Our main conclusion is that rewarding users monetarily for their achieved user engagements increase the news sharing intention of them including misinformation, hyperpartisan, and neutral news. However, if monetary penalties are set in place, that would disincentivize sharing intention of misinformation. Moreover, we find that rewards increase the willingness to share misinformation even when penalties are in place. Given that we also find that respondents tend to have a mediocre understanding of the mechanisms behind monetization features, future research should consider how knowledge may be linked to the effects of monetization. If monetization is adopted widely by mainstream social media platforms, user knowledge may increase in the future.

These findings have policy implications for content moderation efforts [24] specially that of combating the spread of misinformation [25], coordinated information operations [20], conspiracy theories [26], and bot activities [3]. More particularly, our findings that 1) monetary incentive outweighs monetary penalization, and 2) participants show higher sharing intention of hyperpartisan news when misinformation and hate speech are penalized are in agreement with the idea that people engage disproportionately higher with 'borderline content' including more sensationalist and provocative content [7]. The results further help to identify relevant directions for further research as well as potential issues that social media companies may need to consider as they incorporate elements of DeSo into their platforms. The decentralization of social media may further increase the misinformation problems that social media already face.

4 Materials and Methods

The study was approved by the Ethics Committee of the Faculty of Arts and Social Sciences, University of Zurich. The experimental design, data collection, and data analysis plan were preregistered. Preregistration materials are available at the Open Science Foundation (OSF). In all survey experiments, we do not exclude participants for inattentiveness or straightlining to avoid selection effects that can undermine causal inference.

4.1 Participants

Study 1

We preregistered a target sample of 1,000 representative American participants recruited from Prolific. The actual experiment conducted between March 8, 2022, and March 9. In total, 1,067 participants started the survey and 1,022 of them completed it. The final sample (mean age = 45.40) included 517 women, 496 men , and 9 other gender identities.

Study 2

We preregistered a target sample of 1,500 participants based on a power analysis aiming to obtain 0.80 power to detect a small effect size of 0.02 at the

standard 0.05 alpha error probability. We also used a Bonferroni adjustment by dividing the nominal alpha level, 0.05, by the number of models tested, 15 (5 hypotheses, 3 outcomes), yielding an alpha of 0.003. The adjusted alpha level was used in the power analysis. The analysis revealed that we needed to recruit at least 730 participants (see Fig S.1 in Appendix S6). We obtained a sample of nationally representative US participants from Prolific. In total, 1,569 started the survey and 1,520 finished it. Only 98 participants indicated they have no social media account. However, we did not remove them from the dataset. We made this decision because one might expect that those without any social media account would join one in future if they know they can make money out of it. A further 459 participants (30%) did not pass the manipulation check (see Section 4.4 in Materials and Methods), but we did not remove them from the dataset. However, we re-run the analyses on only those who pass the manipulation check and report the results in Appendix S10 as a robustness check. The focused sample (mean age = 44.89) included 778 females, 727 males, and 15 other gender. This study was run on 10-12 June 2022.

4.2 Materials

Study 1

In this study we examined the knowledge of participants regarding cryptocurrencies. In the survey, respondents were first presented with a general background information about the recent trends in major social media platforms in mimicking the monetary incentives of decentralized social media counterparts and then presented them with five statements of which two were correct and three incorrect which they had to rate as either 'right' or 'wrong'. A detailed description of the questions and workflow is included in the preregistration materials.

Study 2

The participants were presented with 5 false, 5 true, and 5 true hyper-partisan news headlines about COVID19 in random order for each participant. All three types of news were randomly selected from a politically-balanced pool of news headlines. Following [21], we selected all news headlines from [27], which labeled news headlines in terms of accuracy and partisan orientation in the US. Participants in the treatment groups were presented with a set of statements about a hypothetical world in which major social media platforms decided to incorporate the monetary incentives of decentralized social media platforms and reward their users with a digital asset for their so called *reputation points*, that is, being calculated in an explicit or implicit way based on the user engagements they received (i.e. like, share, comment, and video views). The statements further mentioned that users can exchange all or part of their reputation points for digital assets in an online market, keep them to maintain their reputation level, or sell it in the future if they think the value

of the assets would increase (see Appendix S8 for the full text shown to all treated participants).

We then presented participants with news headlines and asked them: "Would you consider posting this news headline on your social media timeline?". The response options were limited to "no" and "yes", and the order in which these two options appeared was fully randomized. Although hypothetical, previous research shows that self-reported and actual sharing decisions of news articles are strongly correlated [28]. A detailed description of the questions and workflow is included in the preregistration materials.

4.3 Procedure

Study 1

First, all participants answer questions about their socio-demographic characteristics, social media use, and various political attitudes and beliefs (see pre-survey questions in Appendix S7). Then, all participants were presented with the following background information about the recent trends in social media platforms:

> "[An image of a news headline titled 'Twitter Fully Incorporates Ethereum Tipping and Wallet Support'.] There have been some recent developments in social media companies to provide an opportunity for their users to gain monetary rewards for the user engagements (number of likes, shares, comments, and video views) they receive. For example, Reddit announced its plan to convert "community points" to digital credits, which users would be able to sell in an online market. When social media platforms such as Facebook, Instagram, TikTok and Twitter will introduce this new feature, users will be able to convert their engagement points to actual money with a one-click button in their profile.
>
> To make these features more concrete, imagine a scenario in which you can earn money as a result of the amount of likes, shares, comments, and video views you receive from your activity on a social media platform (e.g. Facebook or Reddit). Here is what is going to happen in that hypothetical world for you:
>
> 1. First, you have to be patient and build a reputation for yourself by posting content and attracting likes, shares, comments, or video views from other users. This may take a few months.
> 2. After the above warm up period, you start earning "reputation points" from that platform based on the engagements (e.g. likes, shares, comments, or video views) you received. Generally, the more engagement, the more points you receive.
> 3. Your reputation points will be shown to others in your profile. The more points you have the more trustable and attractive you will appear to other users. Also, the more points you have, the higher the probability that your posts will be shown to other users by the platform's ranking algorithms.
> 4. Once you have enough reputation points, there will be an online market where you can sell all or part of them. Like in the stock market, the price for points in that market is variable: it might go up in the future, go down, or stay the same."

Then, participants were asked "Consider the following statements regarding the new features that social media platforms are considering, described above.

Table 1 Treatment Groups

	Link: Clear[1]	Link: Vague[2]
Rewards and Penalties	Group 1	Group 2
Rewards but no penalty	Group 3	Group 4

[1]This condition captures the situation in which the method being used to convert user engagements to reputation points is clearly stated for the participants.

[2]This condition captures the situation in which the method being used to convert user engagements to reputation points is vaguely stated for the participants.

Please check if they are right or wrong, based on the short text you have just read." which they had to rate as either 'right' or 'wrong':

"1. Users can convert engagements (e.g., likes) into actual money.
2. The monetary value of engagements depends on how much they are worth when users sell them in the online market in which they can be exchanged for actual money.
3. Social media platforms set a fixed price at which engagements can be converted into actual money.
4. Reputation points automatically increase over time.
5. The increased visibility users enjoy thanks to their reputation points remains even after selling them."

At the end of the experiment, all participants were presented with a debriefing statement (see Section S9 in Appendix) about the goal of the study and names and contact information of the researchers and the University of Zurich's Ethics Committee.

Study 2

First, all participants answer questions about their socio-demographic characteristics, social media use, and various political attitudes and beliefs (see pre-survey questions in Appendix S7). Then each participant is randomly assigned to either the control group (20% of the sample) or one of four treatment conditions (20% of the sample each). The four treatment conditions combine two factors: (1) whether users can only be rewarded for the content they post (Group 3 and 4), or also penalized (Group 1 and 2); and (2) whether the link between their behavior and rewards/penalties is clear (Group 1 and 3) or not (Group 2 and 4). We use simple randomization, not conditional on any observable characteristics of participants.

After treatment assignment, all participants in the treatment condition were presented with the following background information about the recent trends in social media platforms:

"[An image of a news headline titled 'Twitter Fully Incorporates Ethereum Tipping and Wallet Support'.] There have been some recent developments in social media companies to provide an opportunity for their users to gain monetary rewards for

the user engagements (number of likes, shares, comments, and video views) they receive. For example, Reddit announced its plan to convert "community points" to digital credits, which users would be able to sell in an online market. When social media platforms such as Facebook, Instagram, TikTok and Twitter will introduce this new feature, users will be able to convert their engagement points to actual money with a one-click button in their profile.

To make these features more concrete, imagine a scenario in which you can earn money as a result of the amount of likes, shares, comments, and video views you receive from your activity on a social media platform (e.g. Facebook or Reddit). Here is what is going to happen in that hypothetical world for you:

1. First, you have to be patient and build a reputation for yourself by posting content and attracting likes, shares, comments, or video views from other users. This may take a few months.

2. After the above warm up period, you start earning "reputation points" from that platform based on the engagements (e.g. likes, shares, comments, or video views) you received. Generally, the more engagement, the more points you receive.

3. Your reputation points will be shown to others in your profile. The more points you have the more trustable and attractive you will appear to other users. Also, the more points you have, the higher the probability that your posts will be shown to other users by the platform's ranking algorithms.

4. Once you have enough reputation points, there will be an online market where you can sell all or part of them. Like in the stock market, the price for points in that market is variable: it might go up in the future, go down, or stay the same."

Participants in the Group 1 condition were given the following instructions: "Imagine a scenario in which you have the possibility to post a link to a news article on your social media timeline. In this hypothetical scenario, you also have the opportunity to earn "reputation points" which can be converted into a digital currency (cryptocurrency). More particularly, you will receive one reputation point for every ten likes, shares, comments, or video views you get. This means that the more engagement your post attracts, the more reputation points you will receive, and therefore the more money you can make. However, posting a link to a news article containing misinformation or hate speech, will cost you half a reputation point. On the following pages, you will be shown a number of headlines. Keeping in mind that the popularity and content of your post determine how much money you can make, for each headline, please indicate whether you would consider posting it on your timeline. First, however, answer the question on the next page."

In the Group 2 condition, the 'More particularly, you will receive one reputation point for every ten likes, shares, comments, or video views you get.' sentence was replaced with 'The number of reputation points you receive depends on the amount of likes, shares, comments or video views you get.' In the Groups 3 and 4 conditions, the 'However, posting a link to a news article containing misinformation or hate speech, will cost you half a reputation point.' sentence from Groups 1 and 2 instructions was removed respectively.

In both conditions, the response options were simply 'no' and 'yes'. Moreover, participants saw the response options listed as either yes/no or no/yes

(randomized across participants—that is, an individual participant only ever saw 'yes' first or 'no' first).

4.4 Manipulation Check

After the survey experiment, each participant was presented with the following two questions:

> Please answer the following questions based on the scenarios you have just read:
>
> 1. Could you be penalized for posting misinformation on social media? (Yes/No)
> 2. Was it clear exactly how many reputation points you would gain for the engagements you get on social media? (Yes/No)

Following our pre-analysis plan, we do not exclude those who failed the manipulation check from our dataset, but we rerun the analyses results for those who pass the manipulation check as well as a robustness test.

4.5 Dependent Variables

The study considers three sets of outcomes linked to the reported willingness to share different kinds of news headlines.

4.5.0.1 Fake news:

Participants will be presented with 5 fake news headlines and asked for each: "If you were to see the above article, would you consider posting it on social media?" (Possible responses: yes/no).

4.5.0.2 Hyperpartisan news (polarization):

Participants will be presented with 5 true, hyperpartisan news headlines and asked for each: "If you were to see the above article, would you consider posting it on social media?" (Possible responses: yes/no).

4.5.0.3 Neutral news:

Participants will be presented with 5 true, non-partisan news headlines and asked for each: "If you were to see the above article, would you consider posting it on social media?". (Possible responses: yes/no).

Following [21], the headlines will be randomly selected from the dataset in [22]. Specifically, we will randomly select 5 fake news headlines, 5 hyperpartisan headlines, and 5 neutral headlines from [22]. All pools are balanced in terms of political bias. The news headlines are sequentially presented in random order for each participant.

4.6 Independent Variables

The following variables are measured in the pre-survey and used as covariates in the OLS regression (see Appendix S7 for full item wording of these survey questions): gender, age, race, party identification, political interest, social media accounts, investment experience, and knowledge of cryptocurrencies.

4.7 Statistical Analysis

Our preregistration specified that following [16] all analyses would be performed at the level of the individual response (that is, one data point per response per participant; 0 = no, 1 = yes) using linear regression with robust standard errors clustered on participant and headline. Formally, we estimate the following OLS model separately for each type of headlines (true, false, and hyperpartisan) and for each relevant comparison (see Figure 4 and Table 1):

$$y_{i,j} = \alpha + \beta_1 \mathbf{T} + \beta_2 \mathbf{X}_i + n_{i,j} + h_{i,j} + \epsilon_{i,j} \quad (1)$$

where i indexes the participants, and j indexes the news headlines. The independent variable $y_{i,j}$ is whether participant i show willingness to share news headline j. β_1 is the effect of treatment, $\mathbf{T}$, on the willingness of participants to share news headlines. $\mathbf{X}_i$ is a vector of covariates associated with participant i (see Section 4.6 in *Materials and Methods* for a full list of covariates). Since our analysis estimates effects on three different outcomes (neutral and hyperpartisan news, and misinformation), we adjusted p-values for multiple hypothesis-testing using the Benjamini-Hochberg procedure [29].

Supplementary information. This manuscript contains supplementary material.

Acknowledgments. This project received funding from the European Research Council under the EU's Horizon 2020 research and innovation program (Grant Agreement No. 883121). We thank conference participants at the EPSA 2022 and APSA 2022 conferences and the Digital Democracy Workshop 2022 for their helpful feedback. Andrew Guess, Pablo Barbera, Gordon Pennycook, and Jacob Shapiro all shared specific ideas we have tried to incorporate. We thank Fabio Melliger, Darya Zare, and Sarvenaz Ebrahimi for outstanding research assistance.

Author contributions. MA conceived the study. FG secured funding. All authors contributed to survey design. EH conducted Study 1. All authors conducted Study 2. MA drafted the initial manuscript. All authors revised the paper.

Data and code availability. All data and codes to reproduce the results are available on a Open Science Foundation (OSF) repository.

References

[1] Guidi, B.: When blockchain meets online social networks. Pervasive and Mobile Computing **62**, 101131 (2020)

[2] Freni, P., Ferro, E., Ceci, G.: Fixing social media with the blockchain. In: Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good, pp. 175–180 (2020)

[3] Li, C., Palanisamy, B.: Incentivized blockchain-based social media platforms: A case study of steemit. In: Proceedings of the 10th ACM Conference on Web Science, pp. 145–154 (2019)

[4] Sockin, M., Xiong, W.: Decentralization through tokenization. Technical report, National Bureau of Economic Research (2022)

[5] Bail, C.: Breaking the social media prism. In: Breaking the Social Media Prism. Princeton University Press, ??? (2021)

[6] Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., *et al.*: The science of fake news. Science **359**(6380), 1094–1096 (2018)

[7] Zuckerberg, M.: A blueprint for content governance and enforcement. Retrieved from Facebook Newsroom website: https://www. facebook. com/notes/mark-zuckerberg/ablueprint-for-content-governance-andenforcement/10156443129621634 (2018)

[8] Raman, A., Joglekar, S., Cristofaro, E.D., Sastry, N., Tyson, G.: Challenges in the decentralised web: The mastodon case. In: Proceedings of the Internet Measurement Conference, pp. 217–229 (2019)

[9] Kuperschmidt, K.: As Musk reshapes Twitter, academics ponder taking flight Many researchers are setting up profiles on social media site Mastodon. Science (2022). https://doi.org/10.1126/science.adf6617

[10] Ba, C.T., Zignani, M., Gaito, S.: The role of cryptocurrency in the dynamics of blockchain-based social networks: The case of steemit. PloS one **17**(6), 0267612 (2022)

[11] Guidi, B., Michienzi, A.: Users and bots behaviour analysis in blockchain social media. In: 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 1–8 (2020). IEEE

[12] Freni, P., Ferro, E., Moncada, R.: Tokenomics and blockchain tokens: A design-oriented morphological framework. Blockchain: Research and Applications **3**(1), 100069 (2022)

[13] Hasell, A.: Shared emotion: the social amplification of partisan news on twitter. Digital Journalism **9**(8), 1085–1102 (2021)

[14] Rathje, S., Van Bavel, J.J., van der Linden, S.: Out-group animosity drives engagement on social media. Proceedings of the National Academy of Sciences **118**(26) (2021)

[15] Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)

[16] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., Rand, D.G.: Shifting attention to accuracy can reduce misinformation online. Nature **592**(7855), 590–595 (2021). https://doi.org/10.1038/s41586-021-03344-2. Accessed 2021-11-26

[17] Prior, M., Sood, G., Khanna, K.: You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions. Quarterly Journal of Political Science **10**(4), 489–518 (2015). https://doi.org/10.1561/100.00014127. Accessed 2021-11-26

[18] Bak-Coleman, J.B., Kennedy, I., Wack, M., Beers, A., Schafer, J.S., Spiro, E.S., Starbird, K., West, J.D.: Combining interventions to reduce the spread of viral misinformation. Nature Human Behaviour, 1–9 (2022)

[19] Loomba, S., de Figueiredo, A., Piatek, S.J., de Graaf, K., Larson, H.J.: Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. Nature human behaviour **5**(3), 337–348 (2021)

[20] Alizadeh, M., Shapiro, J.N., Buntain, C., Tucker, J.A.: Content-based features predict social media influence operations. Science advances **6**(30), 5824 (2020)

[21] Guess, A.M., Lerner, M., Lyons, B., Montgomery, J.M., Nyhan, B., Reifler, J., Sircar, N.: A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences **117**(27), 15536–15545 (2020). https://doi.org/10.1073/pnas.1920498117. Accessed 2021-04-29

[22] Pennycook, G., Binnendyk, J., Newton, C., Rand, D.G.: A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. Collabra: Psychology **7**(1), 25293 (2021). https://doi.org/10.1525/collabra.25293. Accessed 2022-02-11

[23] Dafoe, A., Zhang, B., Caughey, D.: Information equivalence in survey experiments. Political Analysis **26**(4), 399–416 (2018)

[24] Alizadeh, M., Gilardi, F., Hoes, E., Klüser, K.J., Kubli, M., Marchal, N.: Content moderation as a political issue: The twitter discourse around trump's ban. Journal of Quantitative Description: Digital Media (2022)

[25] Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., Rand, D.G.: Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological science **31**(7), 770–780 (2020)

[26] Batzdorfer, V., Steinmetz, H., Biella, M., Alizadeh, M.: Conspiracy theories on twitter: emerging motifs and temporal dynamics during the covid-19 pandemic. International journal of data science and analytics **13**(4), 315–333 (2022)

[27] Pennycook, G., Bear, A., Collins, E.T., Rand, D.G.: The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. Management Science **66**(11), 4944–4957 (2020)

[28] Mosleh, M., Pennycook, G., Rand, D.G.: Self-reported willingness to share political news articles in online surveys correlates with actual sharing on twitter. Plos one **15**(2), 0228882 (2020)

[29] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) **57**(1), 289–300 (1995)

S1 Decentralized Social Media: Background Information

S1.1 What is Cryptocurrency?

A cryptocurrency is a digital asset that is designed to work as a currency through a decentralized network of computers in a way that is nearly impossible to counterfeit or double-spend [1]. The technology that secures the transactions and store the history in a digital ledger is called the *blockchain* technology [2]. Therefore, we can define a cryptocurrency as a digital currency that is built on blockchain technology. Bitcoin was the first cryptocurrency released in 2009. Since then many other cryptocurrency have been introduced. For example, Litecoin and Ethereum were launched on October 2011 and August 2015, respectively.

Although initially designed to be a peer-to-peer electronic cash system [3], Bitcoin, and other cryptocurrencies created afterward (a.k.a., alternative coins or altcoins), rapidly gained the reputation of being pure speculative assets [4]. Their prices are mostly idiosyncratic [5], subject to market conditions [6], mainly driven by behavioral factors [7], and are correlated with some major classes of financial assets (e.g. US Dollar index) and uncorrelated with some others (e.g. gold prices) [8].

S1.2 Where does the Monetary Value of Cryptocurrency Come From?

The monetary value of some cryptocurrencies such as bitcoin and etherium comes from the fact that implementing the blockchain technology needs a computing infrastructure, which is costly (e.g. building facilities, electricity bills, and high-performance computers) [9]. The process of maintaining and securing the ledger through the blockchain technology is called *mining* [9]. Other cryptocurrencies earn their value based on a business behind them. For example, a new company can emerge for decentralized money exchange, in which users can send money to anyone in the world without relying on banks, and introduce a new cryptocurrency (i.e. coin) for a blockchain technology that supports its business. The process of converting a physical asset (e.g. a building or land) or a business to a cryptocurrency is called *tokenization* [9].

S1.3 What is NFT and How Does it Create Value?

A new form of tokenization, called non-fungible token (NFT), has recently emerged that tokenizes digital data such photos, videos, and audio. An NFT is a unit of data stored on a blockchain that certifies a digital asset to be unique and therefore not interchangeable, while offering a unique digital certificate of ownership for the NFT [10]. While initial NFTs were all part of the Ethereum (a decentralized open-source blockchain with smart contract functionality), more blockchains are implementing their own version of NFTs [11].

NFTs has been embraced by high-profile musicians, singers, and athletes, as well as digital art creators [10]. Each week, hundreds of millions of dollars of NFTs are being traded on online marketplaces such as OpenSea [3]. Though complex to grasp for many, NFT brings many advantages for digital art market: 1) solves the problem of ownership by storing the digital identity of the owner on a public-facing digital ledger (i.e. blockchain); 2) facilitates the transferring of the rights of use, 3) makes the market transparent by keeping the history of all trades; 4) and facilitates development of a highly-engaged community for digital artists (think it as a membership card).

In summary, NFTs are creating new markets based on new forms of ownership. The value of NFTs, in essence, is coming from the same core dynamics of cryptocurrencies, which is users' shared agreement [12]. In other words, an NFT value comes from the users community one builds around it, and the more this community attracts engagements from users and become part of their everyday life, the more the value increases.

S2 Existing Decentralized Social Media Platforms

Table S.1 shows a list of major existing decentralized social media platforms along with their release year and the major platform they are alternative to them.

No.	Name	Alternative To	Release Year
1	Diaspora	Facebook	2010
2	LBRY	Youtube	2015
3	Minds	Facebook/Twitter	2015
4	Mastodon	Twitter	2016
5	Element	Whatsapp/Slack	2016
6	Dtube	Youtube	2016
7	Steemit	Medium	2016
8	Indorse	Linkedin	2017
9	Emanate	Spotify	2017
10	Peertube	Youtube	2018
11	Peepeth	Twitter	2018
12	Entre	Twitter/Linkedin	2019
13	Subsocial	Any platform	2020

Table S.1 List of the some of the well-know existing decentralized social media platforms.

S3 Recent Plans of Platforms to Financially Reward User Engagements

- Twitter: Tips (https://blog.twitter.com/en_us/topics/product/2021/bringing-tips-to-everyone, accessed August 12, 2022)

[3] https://opensea.io/

- Twitter: Creator monetization for all forms of content (https://twitter.com/elonmusk/status/1589010272341340160?s=20&t=vtBPJ4kqdVoHmfKf2Yj28w, accessed November 8, 2022)
- Medium: Partner Program (https://help.medium.com/hc/en-us/articles/115011694187-Getting-started-with-the-Partner-Program, accessed August 12, 2022)
- Reddit: Community Points (https://www.reddit.com/community-points/, accessed November 8, 2022)
- Reddit: tokenization of *Karma* points (https://cointelegraph.com/news/reddit-to-reportedly-tokenize-karma-points-and-onbo accessed August 12, 2022)

S4 Token Classification Framework

As discussed in the main text, Freni et al (2022) proposed a framework for classification of tokens. Below we provide defintions of the five categories mentioned in the 'Introduction' part of the paper:

1. Technology domain: refers to all technical characteristics of a token, including the level of integration, the blockchain technology, and protocol. The Technology Domain comprises five dimensions: Chain, Permission, Number of Blockchains, Representation Type.
2. Behavior domain: includes the intrinsic characteristics of a token which determines the actions that can be performed with the token. The behaviour domain includes six dimensions: Burnability, Expirability, Spendability, Fungibility, Divisibility, Tradability.
3. Inherent Value domain: refers to the characteristics of a token that define the economic value of the token, how it is originated, and what factors affect its price and fluctuations.
4. Coordination domain: consists the characteristics of a token which enable coordination among the holders of that token. The coordination domain includes four dimensions: Underlying Value, Supply Strategy, Incentive Enablers, Incentive Drivers.
5. Pseudo-archetypes: describes the characteristics of a token "that anticipate a set of token archetypes since they implicitly combine different token characteristics" (Freni et al 2022).

S5 Regression Analysis of Study 1

In study 1, n = 1,022 American individuals from Prolific were presented with five statements about cryptocurrencies of which two were correct and three incorrect which they had to rate as either as 'right' or 'wrong'. In general, the results show that the more participants think they know about cryptocurrency or have more positive opinion about it, the less likely they answer correctly to questions.

Table S.2 reports the results of our first study, which focused on knowledge regarding cryptocurrencies. In the survey, respondents were presented with five statements of which two were correct and three incorrect which they had to rate as either 'right' or 'wrong' . In the first column, the dependent variable is the share of correct answers, whereas in the second it is the number. In both models, the intercept shows the outcome for the reference groups. On average, respondents in the reference groups answered about 46% of the questions correctly, that is about 2.3 questions. We observe that there is little variation in correct response rates. The only significant differences, in a negative direction, are found for older respondents (65+, relative to 18-24 years old) and, interestingly, respondents who reported to be "extremely knowledgeable" about crypto (relative to those reporting to be "not at all knowledgeable"). Based on our questions, both groups have a poorer understanding of crypto than the corresponding reference categories. These results suggest that the understanding of crypto is, on average, mediocre, with little variation across most groups.

Table S.2: *Correlates of understanding of crypto (linear regression models)*

	Correct answers: Percent	Correct answers: Number
(Intercept)	0.46***	2.31***
	(0.11)	(0.57)
genderMale	0.01	0.07
	(0.01)	(0.07)
genderOther	0.09	0.43
	(0.07)	(0.36)
age25-34	0.00	0.01
	(0.02)	(0.12)
age35-44	−0.03	−0.17
	(0.03)	(0.13)
age45-54	−0.03	−0.16
	(0.03)	(0.13)
age55-64	0.00	0.01
	(0.03)	(0.13)
age65+	−0.06*	−0.29*
	(0.03)	(0.14)
raceAmerican Indian or Alaskan Native	−0.10	−0.52
	(0.11)	(0.55)
raceAsian or Asian American	−0.01	−0.03
	(0.06)	(0.30)
raceBlack or African American	−0.05	−0.25
	(0.06)	(0.28)
raceHawaiian or other Pacific Islander	0.15	0.77
	(0.16)	(0.78)
raceHispanic or Latino	−0.04	−0.20
	(0.06)	(0.30)
raceNon-Hispanic White	−0.03	−0.13

Table S.2: *Correlates of understanding of crypto (linear regression models)*

	Correct answers: Percent	Correct answers: Number
	(0.05)	(0.27)
educationCollege graduate	0.04	0.20
	(0.07)	(0.37)
educationHigh school graduate	0.02	0.08
	(0.08)	(0.38)
educationPost-graduate training/Professional school	0.03	0.17
	(0.08)	(0.38)
educationSome college, no 4-year degree	0.01	0.04
	(0.07)	(0.37)
educationTechnical, trade or vocational school	−0.00	−0.00
	(0.08)	(0.40)
ideologySomewhat conservative	0.02	0.09
	(0.03)	(0.13)
ideologySomewhat liberal	0.01	0.03
	(0.02)	(0.10)
ideologyVery conservative	−0.00	−0.02
	(0.03)	(0.17)
ideologyVery liberal	0.03	0.17
	(0.02)	(0.11)
pid_forcedRepublican Party	0.02	0.11
	(0.02)	(0.12)
news>1 hour	−0.01	−0.06
	(0.02)	(0.12)
news 1 hour	0.00	0.01
	(0.02)	(0.12)
news 30 minutes	0.01	0.04
	(0.02)	(0.11)
news10-20 minutes	−0.01	−0.03
	(0.02)	(0.10)
N social media accounts	−0.00	−0.02
	(0.01)	(0.03)
Investment experience	0.00	0.00
	(0.00)	(0.00)
crypto_knowExtremely Knowledgeable	−0.13*	−0.67*
	(0.06)	(0.31)
crypto_knowKnowledgeable	−0.01	−0.05
	(0.02)	(0.12)
crypto_knowSlightly knowledgeable	−0.03	−0.13
	(0.02)	(0.08)
crypto_knowVery Knowledgeable	−0.05	−0.26
	(0.04)	(0.19)
crypto_realAgree	−0.03	−0.15
	(0.04)	(0.22)
crypto_realDisagree	−0.03	−0.14
	(0.04)	(0.21)

Table S.2: *Correlates of understanding of crypto (linear regression models)*

	Correct answers: Percent	Correct answers: Number
crypto_realNeither agree/disagree	0.01	0.05
	(0.04)	(0.21)
crypto_realStrongly agree	−0.01	−0.05
	(0.05)	(0.26)
crypto_justanotherAgree	0.01	0.05
	(0.04)	(0.19)
crypto_justanotherDisagree	0.02	0.12
	(0.04)	(0.19)
crypto_justanotherNeither agree/disagree	−0.01	−0.03
	(0.04)	(0.19)
crypto_justanotherStrongly agree	−0.06	−0.28
	(0.04)	(0.22)
crypto_scamAgree	0.05	0.25
	(0.03)	(0.16)
crypto_scamDisagree	0.03	0.13
	(0.02)	(0.12)
crypto_scamNeither agree/disagree	0.01	0.05
	(0.03)	(0.13)
crypto_scamStrongly agree	−0.01	−0.03
	(0.05)	(0.23)
R^2	0.06	0.06
Adj. R^2	0.02	0.02
Num. obs.	1022	1022

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

S6 Power Analysis

We preregistered a target sample of 1,500 participants based on a power analysis aiming to obtain 0.80 power to detect a small effect size of 0.02 at the standard 0.05 alpha error probability (Fig S.1). We also used a Bonferroni adjustment by dividing the nominal alpha level, 0.05, by the number of models tested, 15 (5 hypotheses, 3 outcomes), yielding an alpha of 0.003. The adjusted alpha level was used in the power analysis. The analysis revealed that we needed to recruit at least 730 participants.

S7 Pre-Survey

1. GENDER: How do you describe yourself?
   (a) Male
   (b) Female
   (c) Other
2. AGE: Which of the following categories includes your current age?
   (a) 17 or younger
   (b) 18 to 24

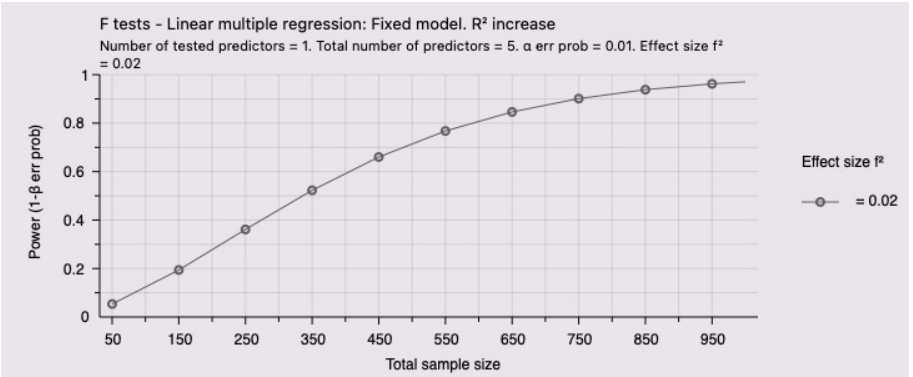

Fig. S.1 Power analysis to aiming to obtain 0.80 power to detect a small effect size of 0.02 at the standard 0.05 alpha error probability. We also used a Bonferroni adjustment by dividing the nominal alpha level, 0.05, by the number of models tested, 15 (5 hypotheses, 3 outcomes), yielding an alpha of 0.003. The adjusted alpha level was used in the power analysis. The analysis revealed that we needed to recruit at least 730 participants.

(c) 25 to 34
(d) 35 to 44
(e) 45 to 54
(f) 55 to 64
(g) 65+

3. RACE: How do you describe yourself? (Please check the one option that best describes you)
   (a) American Indian or Alaskan Native
   (b) Hawaiian or other Pacific islander
   (c) Asian or Asian American
   (d) Black or African American
   (e) Hispanic or Latino
   (f) Non-Hispanic White
   (g) Other
4. EDUCATION: What is the highest level of education you have completed?
   (a) None, or grades 1-8
   (b) High school incomplete (grades 9-11)
   (c) High school graduate (grade 12 or GED certificate)
   (d) Technical, trade or vocational school AFTER high school
   (e) Some college, no 4-year degree (includes associate degree)
   (f) College graduate (B.S., B.A., or other 4-year degree)
   (g) Post-graduate training/professional school after college (toward a Master's degree or Ph.D., Law or Medical school)
5. IDEOLOGY: In politics, people sometimes talk about liberal and conservative. In general, would you describe yourself as:
   (a) Very liberal
   (b) Somewhat liberal
   (c) Moderate

(d) Somewhat conservative
(e) Very conservative

6. PID: Generally speaking, do you think of yourself as a Democrat, a Republican, an Independent, or something else?
   (a) Strong Democrat
   (b) Weak Democrat
   (c) Lean Democrat
   (d) Independent
   (e) Lean Republican
   (f) Weak Republican
   (g) Strong Republican
7. PID - FORCED: If you absolutely had to choose between only the Democratic and Republican Party, which would you prefer?
   (a) Democratic Party
   (b) Republican Party
8. NEWS: On a typical day, how much time would you say you spend reading the news (either on a mobile phone or a computer?
   (a) Less than 10 minutes
   (b) Between 10-20 minutes
   (c) About 30 minutes
   (d) About 1 hour
   (e) More than 1 hour
9. POLITICAL INTEREST: How interested are you in politics?
   (a) Not at all interested
   (b) Not very interested
   (c) Somewhat interested
   (d) Very interested
   (e) Extremely interested
10. SOCIAL MEDIA ACCOUNT 1: Do you have a social media profile (e.g., Facebook, Twitter, Instagram, Reddit, TikTok)?
    (a) No
    (b) Yes
11. SOCIAL MEDIA ACCOUNT 2: Do you have accounts on any of the following social media sites? Please select all that apply
    (a) Twitter
    (b) Instagram
    (c) Reddit
    (d) TikTok
    (e) None
12. SHARING BEHAVIOR: Would you ever consider sharing something political on social media?
    (a) No
    (b) Yes

13. SOCIAL MEDIA USE: On a typical day, how much time would you say you spend on social media (such as Facebook and Twitter; either on a mobile or a computer)?
    (a) Less than 30 minutes
    (b) 30-60 minutes
    (c) 1-2 hours
    (d) 2-3 hours
    (e) 3+ hours
14. SHARING POST: When deciding whether to share a piece of content on social media, how important is it to you that the content is (Likert scale):
    (a) Surprising
    (b) Politically aligned
    (c) Funny
    (d) Interesting
    (e) Accurate
15. INVESTMENT EXPERIENCE: Do you have any experience in financial trading markets such as stocks, other investments or cryptocurrency (e.g. Bitcoin, Ethereum, NFTs)? Please check all that apply.
    (a) Stocks
    (b) Cryptocurrency
    (c) Other investments
    (d) None
16. KNOWLEDGE OF CRYPTOCURRENCY: How do you evaluate your knowledge of cryptocurrencies (e.g., Bitcoin, Ethereum, NFTs)? (Likert scale)
17. ATTITUDE ON CRYPTOCURRENCY. How much do you agree with the following statements? (Likert scale)
    (a) Cryptocurrencies are real thing with high growth potential in the future
    (b) Cryptocurrencies are just another investment opportunity like the stock market
    (c) Cryptocurrencies are a scam

S8 Treatments

Each participant is randomly assigned to either the control group (20% of the sample) or one of four treatment conditions (20% of the sample each). The four treatment conditions combine two factors: (1) whether users can only be rewarded for the content they post (Group 3 and 4), or also penalized (Group 1 and 2); and (2) whether the link between their behavior and rewards/penalties is clear (Group 1 and 3) or not (Group 2 and 4). We use simple randomization, not conditional on any observable characteristics of participants.

S8.1 Background Information

Participants in the treatment groups will be asked to read the following text, explaining that social media platforms are increasingly considering implementing a system

which provides an opportunity for their users to gain monetary rewards for the engagements that their content attracts.

"[An image of a news headline titled 'Twitter Fully Incorporates Ethereum Tipping and Wallet Support'.] There have been some recent developments in social media companies to provide an opportunity for their users to gain monetary rewards for the user engagements (number of likes, shares, comments, and video views) they receive. For example, Reddit announced its plan to convert "community points" to digital credits, which users would be able to sell in an online market. When social media platforms such as Facebook, Instagram, TikTok and Twitter will introduce this new feature, users will be able to convert their engagement points to actual money with a one-click button in their profile.

To make these features more concrete, imagine a scenario in which you can earn money as a result of the amount of likes, shares, comments, and video views you receive from your activity on a social media platform (e.g. Facebook or Reddit). Here is what is going to happen in that hypothetical world for you:

1. First, you have to be patient and build a reputation for yourself by posting content and attracting likes, shares, comments, or video views from other users. This may take a few months.

2. After the above warm up period, you start earning "reputation points" from that platform based on the engagements (e.g. likes, shares, comments, or video views) you received. Generally, the more engagement, the more points you receive.

3. Your reputation points will be shown to others in your profile. The more points you have the more trustable and attractive you will appear to other users. Also, the more points you have, the higher the probability that your posts will be shown to other users by the platform's ranking algorithms.

4. Once you have enough reputation points, there will be an online market where you can sell all or part of them. Like in the stock market, the price for points in that market is variable: it might go up in the future, go down, or stay the same."

S8.2 Four Treatment Groups

Moreover, participants in the treatment groups will asked to read one of following texts:

S8.2.0.1 Group 1:

"Imagine a scenario in which you have the possibility to post a link to a news article on your social media timeline. In this hypothetical scenario, you also have the opportunity to earn "reputation points" which can be converted into a digital currency (cryptocurrency). More particularly, you will receive one reputation point for every ten likes, shares, comments, or video views you get. This means that the more engagement your post attracts, the more reputation points you will receive, and therefore the more money you can make. However, posting a link to a news article containing misinformation or hate speech, will cost you half a reputation point.

On the following pages, you will be shown a number of headlines. Keeping in mind that the popularity and content of your post determine how much money you can make, for each headline, please indicate whether you would consider posting it on your timeline. First, however, answer the question on the next page."

S8.2.0.2 Group 2:

"Imagine a scenario in which you have the possibility to post a news headline on your social media timeline. In this hypothetical scenario, you also have the opportunity to earn "reputation points" which can be converted into a digital currency (cryptocurrency). The number of reputation points you receive depends on the amount of likes, shares, comments or video views you get. This means that the more engagement your post attracts, the more points you will receive, and therefore the more money you can make. However, the exact formula that the platform use to compute your reputation points based on the user engagements you received in unknown. In addition, posting a news headline containing misinformation or hate speech will decrease your reputation points.

On the following pages, you will be shown a number of headlines. Keeping in mind that the popularity and content of your post determine how much money you can make, for each headline, please indicate whether you would consider posting it on your timeline. First, however, answer the question on the next page."

S8.2.0.3 Group 3:

"Imagine a scenario in which you have the possibility to post a link to a news article on your social media timeline. In this hypothetical scenario, you also have the opportunity to earn "reputation points" which can be converted into a digital currency (cryptocurrency). You will receive one reputation point for every ten likes, shares, comments, or video views you get. This means that the more engagement your post attracts, the more points you will receive, and therefore the more money you can make.

On the following pages, you will be shown a number of headlines. Keeping in mind that the popularity of your post determines how much money you can make, for each headline, please indicate whether you would consider posting it on your timeline. First, however, answer the question on the next page."

S8.2.0.4 Group 4:

"Imagine a scenario in which you have the possibility to post a news headline on your social media timeline. In this hypothetical scenario, you also have the opportunity to earn "reputation points", which can be converted into a digital currency (cryptocurrency). The number of reputation points you receive depends on the amount of likes, shares, comments or video views you get. This means that the more engagement your post attracts, the more points you will receive, and therefore the more money you can make. However, the exact formula that the platform use to compute your reputation points based on the user engagements you received in unknown.

On the following pages, you will be shown a number of headlines. Keeping in mind that the popularity of your post determines how much money you can make, for each headline, please indicate whether you would consider posting it on your timeline. First, however, answer the question on the next page."

S9 Debriefing

All participants will be presented with the following debriefing note:

"Thank you for your participation in this research study. Now that you completed or have ended your participation, we will provide you with some additional information about the purposes of this study.

What you should know about this study

The main purpose of this study was for us to observe how likely participants were to share various kinds of news headlines, depending on the rewards and penalties implied by the implementation of crypto technology by social media platforms. The news headlines were real, and were selected to be either false, hyperpartisan, or neutral. It is thus important for you to keep in mind that these were not personally targeted towards you and that they were purposefully misleading and/or hyperpartisan (when they were not neutral).

If you have questions

The main researchers conducting this study are Dr. Meysam Alizadeh (meysam.alizadeh@uzh.ch), Prof. Fabrizio Gilardi (gilardi@ipz.uzh.ch) and Dr. Emma Hoes (hoes@ipz.uzh.ch) (University of Zurich, Department of Political Science). If you have questions, you may contact one of the main researchers. If you have any questions or concerns regarding your rights as a research participant in this study, you may contact the Ethics Committee: Prof. Dr. Lilly Shanahan Universität Zürich Jacobs Center for Productive Youth Development Andreasstrasse 15, P.O. Box 12 CH-8050 Zürich""

S10 Robustness Check

As mentioned in Section 4.1, 30% of participants did not pass the manipulation test. To check for the robusteness of our results, we remove those who failed the manipulation check and report the OLS estimates in Fig S.2.

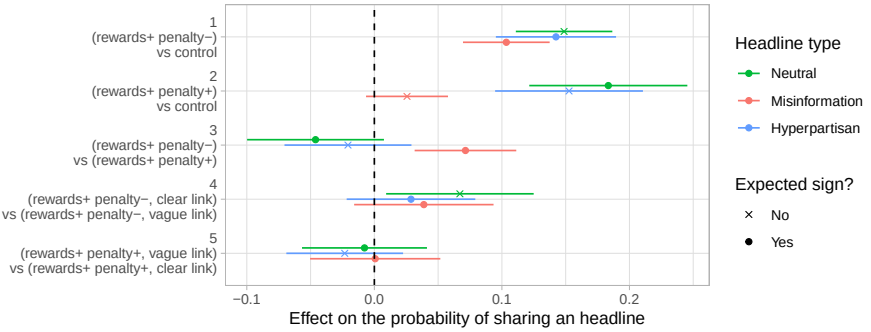

Fig. S.2 Point estimates from OLS regressions with 95% CI for participants who passed the robustness check. Effect of monetary rewards for user engagements and penalties for problematic sharing behavior, as well as of the clarity of the link between rewards/penalties and engagements/behavior, on the willingness to share neutral and hyperpartisan news, and misinformation. Without penalties, rewards increase the willingness to share misinformation by about nine percentage points, compared to the control group. With penalties, the effect decreases to about four percentage points. The clarity of the link has no effect.

Appendix References

[A1] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A Kroll, and Edward W Felten. Sok: Research perspectives and

challenges for bitcoin and cryptocurrencies. In *2015 IEEE symposium on security and privacy*, pages 104–121. IEEE, 2015.

[A2] Dylan Yaga, Peter Mell, Nik Roby, and Karen Scarfone. Blockchain technology overview. *arXiv preprint arXiv:1906.11078*, 2019.

[A3] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260, 2008.

[A4] David Yermack. Is bitcoin a real currency? an economic appraisal. In *Handbook of digital currency*, pages 31–43. Elsevier, 2015.

[A5] Helder Sebastião and Pedro Godinho. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, 7(1):1–30, 2021.

[A6] Fan Fang, Carmine Ventre, Michail Basios, Leslie Kanthan, David Martinez-Rego, Fan Wu, and Lingbo Li. Cryptocurrency trading: a comprehensive survey. *Financial Innovation*, 8(1):1–59, 2022.

[A7] Mehmet Balcilar, Elie Bouri, Rangan Gupta, and David Roubaud. Can volume predict bitcoin returns and volatility? a quantiles-based approach. *Economic Modelling*, 64:74–81, 2017.

[A8] Yechen Zhu, David Dickinson, and Jianjun Li. Analysis on the influence factors of bitcoin's price based on vec model. *Financial Innovation*, 3(1):1–13, 2017.

[A9] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press, 2016.

[A10] Matthieu Nadini, Laura Alessandretti, Flavio Di Giacinto, Mauro Martino, Luca Maria Aiello, and Andrea Baronchelli. Mapping the nft revolution: market trends, trade networks, and visual features. *Scientific reports*, 11(1):1–11, 2021.

[A11] The Writer's Lounge. Choosing the right blockchain for your nft. Technical report, Medium, 2020.

[A12] Steve Kaczynski and Duke Kominers. How nfts create value. Technical report, Harvard Business Review, 2021.