LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Department of Psychological and Behavioural Science

COURSEWORK SUBMISSION FORM AND PLAGIARISM/GENERATIVE AI/ACADEMIC HONESTY DECLARATION

Please ensure that a completed copy of this form is uploaded as part of your coursework submission.

Candidate Number: 46341

Course code: PB4D3: Behavioural Science in an Age of New Technology

Word-count: 10,970

Date: 7th August 2023

The Department wishes to draw your attention to the School Calendar Regulations on Assessment Offences and Plagiarism:

https://info.lse.ac.uk/Staff/Divisions/Academic-Registrars-Division/Teaching-Quality-Assurance-and-Review-Office/Assets/Documents/Calendar/RegulationsAssessmentOffences-Plagiarism.pdf

All work submitted as part of the requirements for any assessment of the School (e.g., examinations, essays, dissertations, and any other work, including computer programs), whether submitted for formative or summative assessment, must be expressed in your own words and incorporate your own ideas and judgments. Plagiarism must be avoided in all such work.

Plagiarism can involve the presentation of another person's thoughts or words as if they were your own. However, please note that plagiarism also includes self-plagiarism, which is where you as the author re-use your own previously submitted work or data in a "new" written piece of work without letting the reader know that this material has appeared elsewhere. The definition of "your own work" also includes work produced by collaboration or group-work expressly permitted by the Department.

Furthermore, it is prohibited to use any form of generative artificial intelligence in an unauthorised way when working on a summative assessment.

Please also note that plagiarism as defined by the School above, need not be deliberate or intentional for it to constitute an assessment offence.

Declaration (without signature, to preserve anonymity): Having read and understood LSE's guidelines on plagiarism/academic honesty, I hereby confirm by completing and attaching this form that the work submitted is my own. By submitting this form I hereby confirm I understand the Department's policy on summative assessment, and that I have read the relevant parts of the MSc programme handbook.

Exploring the Impact of Sentiment, Partisan Bias, and Likes on Truth Judgements of Misinformation through a Conjoint Analysis

Abstract

Understanding the dynamics of misinformation has far-reaching implications for societal trust, informed discourse, and the functioning of democratic societies. Within the mechanisms that drive misinformation spread, the factors which influence people's susceptibility to believing misinformation has become a topic of paramount importance. This study investigates how three key aspects of misinformation on social media platforms – headline sentiment, partisan bias, and like counts – influence individuals' judgments of headline accuracy. To do so, a pairwise choice-based conjoint analysis was employed (N = 500) wherein two machine learning models to measure sentiment and partisan bias were utilised. The results suggest the existence of a negativity bias within misinformation truth judgements, along with information-avoidance-based motivated reasoning. No notable effect of like counts is found, however heterogeneity of their effect between participant partisanships and types of headlines is suggested. Further research is suggested to explore the effect these attributes have with a larger sample size, allowing for a more granular investigation.

Table of Contents

Introduction

Throughout history, misinformation has played a significant role in shaping societies, from the days of Roman emperor Octavian who employed it to slander his political opponent, Mark Antony (Posetti & Matthews, 2018), to its modern utilisation in political campaigns as a means of waging ideological warfare (Strömbäck et al., 2022). In recent years, the prevalence of misinformation has become increasingly evident, manifesting in various contexts with significant consequences such as during the 2016 US presidential elections, where misinformation played a role in influencing voter behaviour (Allcott & Gentzkow, 2017). Most notably, the spread of misinformation about the COVID-19 virus led to the formation of anti-vaccination groups, posing a threat to public health by undermining vaccination efforts (Loomba et al., 2021).

Considering its ever-growing relevance, academics from several disciplines have undertaken initiatives to understand the causes, symptoms, and consequences of the spread of misinformation. Notably, significant attention has been directed toward exploring individual susceptibility to misinformation, particularly from a psychological perspective. Gaining insights into the cognitive factors that influence people's vulnerability to misinformation is of paramount importance, as it holds the key to preventing its continuous dissemination.

In academics' various efforts to understand the individual factors which affect one's susceptibility to misinformation, the field of emotion and misinformation has been mostly ignored. This is surprising given the abundance of research which suggests that emotion plays a crucial role in our truth judgements. This investigation aimed to explore this field by investigating how sentiment within news headlines affects people's truth judgements of misinformation. All the while, this investigation also explored how motivated reasoning and social norms affect perceptions of misinformation through the investigation of how partisan biases and like counts affect peoples' truth judgements of misinformation.

To do so, I employed a pairwise conjoint experiment, wherein participants were shown numerous pairs of headlines with varying attributes and were asked to choose the most truthful of the pair. The headlines shown varied in their veracity (True or False), sentiment (Very Negative → Very Positive), partisan bias (Left Extreme → Right Extreme), and the inclusion of a like count (no like count, like count with friends mentioned, like count without friends). From the results that this experimental design yielded, Average Marginal Component Effects (AMCE) were calculated, which show the change in the probability that a headline is chosen when participants are presented with different attributes.

To measure and change the partisan bias and sentiment of the headlines, two machine learning models were employed from the field of natural language processing. Python library VADER (Hutto & Gilbert, 2014) was employed to measure sentiment, and the machine learning model from the "thebipartisanpress" (Wang, 2019) was used to measure partisan bias. The inclusion of both models to investigate misinformation is another novelty of the investigation.

Literature Review

Since this study tested for the effect of sentiment, partisan biases, and the inclusion of a like count on headline belief, the following literature review will briefly cover the relevant literature of all three of these attributes. Beforehand however, it is important to clarify what I mean when I refer to "misinformation".

Defining Misinformation

There have been several recent attempts at defining misinformation (Vraga & Bode, 2020). In the early 2000s, Kuklinski et al. (2000) definition of misinformation was the most popular, who defined misinformation as when "people hold inaccurate beliefs, and do so confidently". As the field progressed, it became clear that this definition lacked an important distinction between "misperception" and "misinformation", eventually giving rise to Nyhan & Reifler's (2010) definition, which drew this distinction, defining "misperception" as the beliefs that people hold, and "misinformation" as false information in itself. Nyhan and Reifler (2010) also attempted to establish a degree of objectivity within the definition of misinformation, declaring that this false information must not be "supported by clear evidence and expert opinion" (Nyhan & Reifler, 2010). To minimise the subjectivity of what "expert opinion" means, I will define it as the supported opinion of credible, non-partisan fact-checking institutions.

As the field has progressed, eight distinct types of misinformation have been identified. These are: satire, propaganda, hoaxes, biased, fabricated news, conspiracy theories, rumours, and clickbait (Zannettou et al., 2019). Some academics have also emphasised the difference between misinformation and *disinformation*, wherein the former entails the unintentional spread of false information, and the latter refers to the intentional spread of misinformation (Pérez-Escolar et al., 2023). However, this investigation did not distinguish between misinformation's various types or between disinformation and misinformation, as previously done by Zannettou et al. (2019). I did this due to the limited real-world headlines available on the topic which the headlines focus on, as well as the difficulty in distinguishing between disinformation and misinformation in the collected material.

Emotion and Misinformation

In this study, the measure of sentiment is solely representative of how the field of natural language processing has defined it. In contrast to the definitions of sentiment within psychology (see Munezero et al., 2014), sentiment within natural language processing has been loosely defined as "a negative or positive opinion" (Mejova, 2019; Melville et al., 2009). Thus, sentiment is often referred to as a *polarity* of emotion, frequently measured on a single continuum from negative to positive. Hence, measuring sentiment offers a proxy of the **valence** of emotion within the text. For example, happiness would generally be seen as a positive sentiment, whereas sadness or anger is seen as a negative sentiment (Sailunaz & Alhajj, 2019). Though the use of sentiment has shown to be very prominent within misinformation (Ghanem et al., 2020; J. Lee et al., 2021; Zaeem et al., 2020) research on how the sentiment of headlines affects perceptions of misinformation is almost non-existent. Thus, to form testable hypotheses it becomes of great importance to look at the wider literature on emotion and truth judgements of misinformation. Consequently, the primary research question of this investigation arises:

RQ1: How does Sentiment affect participants' truth judgements of misinformation?

To establish testable hypotheses for this research question, I will delve into two of the most popular and relevant topics in the literature: negativity bias, and affect and processing styles.

Negativity Bias

Though only formally coined in Rozin & Royzman (2001), the initial evidence of negativity bias comes from Tversky & Kahneman's (1974), seminal work on loss aversion, which demonstrated our tendency to generally be loss-averse, even in situations where traditional economic theory would predict otherwise (Tversky & Kahneman, 1991). Negativity bias is often defined as a tendency to place greater value towards negative information, which in turn has been shown to have various effects, such as greater memory, attention, and perceived complexion of such information (Norris, 2021). This bias has been hypothesised to exist by evolutionary psychologists to heighten our chances of survival, as placing greater value on negative information results in a greater awareness of threats and risks (Rozin & Royzman, 2001). The existence of negativity bias has been extensively supported by both developmental studies of negativity bias in infants (Vaish et al., 2008), as well as neuroscientific studies, which have both confirmed the existence of the negativity bias (Cunningham et al., 2004) and attributed certain brain areas to its activation (Schupp et al., 2004).

The role of negativity bias in truth judgements has been explored by altering the framing in which information is presented and measuring the potential differences this has on people's belief. For example, In Hilbig (2009), participants were divided into two groups. One was presented with a negative frame, where they were informed that 85% of attempts of rape were successful. The other group received a positive frame, stating that 15% were unsuccessful. In post-hoc analysis, it was found that the negative framing condition was perceived as significantly more truthful than the positive framing condition. These results have been replicated in several other studies (e.g., Baumeister et al., 2001; Hilbig, 2012). While the precise mechanisms underlying the negativity bias in truth judgments remain a topic of ongoing debate, a notable proposition by Pennycook & Rand, (2019) suggests that a negative frame could trigger heightened attention and arousal, potentially facilitating the easier retrieval of pertinent knowledge and the generation of supporting evidence. Though negativity bias has not been explicitly studied within the context of misinformation, this research would suggest the following results within this investigation:

H1.1: Headlines with a negative sentiment will be believed significantly more than other sentiments.

However, the literature also seems to suggest that this negativity bias will not be as pronounced within very negative headlines, in contrast to negative headlines. Though not investigated within misinformation, several findings from various fields such as persuasion, public speaking, and the broader field of emotion and credibility judgements repeatedly report that a high level of emotionality within information consistently lowers perceptions of credibility. This has been investigated and supported within various contexts such as product reviews (Vendemia, 2017), news articles (Karduni et al., 2021), and even public speaking (Stephens et al., 2019). Thus, the following is also to be expected within the experiment:

H1.2: Participants will believe headlines with very positive, or very negative sentiment less than their moderate counterparts (negative and positive sentiment).

Affect and Processing Styles

While research on negativity bias explores the impact of information **presentation** on truth judgments, it is equally pertinent to explore how different affective states could influence truth judgments, considering headlines may evoke diverse affective states in participants. Within this literature, there is some work on how affective states may affect truth judgements of news and misinformation. The earliest example of this is Staats & Staats (1958), where political messages were found to be more effective when participants were in a positive mood (induced by being given free lunch) rather than a negative mood (induced by exposure to “noxious gases”). Some contemporary studies take similar approaches, such as Koch & Forgas (2012), where participants were exposed to emotionally evocative films before being shown “doubtful statements”. Nonetheless, the dearth of these studies warrants a wider look at the literature on affect and truth judgements.

Within the wider literature, academics have dedicated a substantial amount of work to understanding the relationship between affect and processing styles. This is of particular importance to this study, given processing styles are increasingly being posited as one of the primary determinants of misinformation susceptibility (Pennycook & Rand, 2019, 2020, 2021). In this literature, the majority of evidence suggests that a positive mood tends to increase ones reliance on heuristics in decision-making, whereas a negative mood induces a more reflexive, data-oriented processing style (Forgas & Eich, 2012; Fredrickson, 2001). As one would expect, this has consequently been shown to impact truth judgements in domains like misinformation. One of these domains is that of gullibility, where a positive affective state has been repeatedly shown to increase people’s susceptibility to misleading information (Forgas, 2019). Given gullibility and misinformation susceptibility are very closely related concepts (Shen et al., 2019), one would expect a parallel effect on one’s truth judgements of misinformation. However, this assumes that sentiment directly induces affect (i.e., negative sentiment will make people feel negative & positive sentiment will make people feel positive). Thus, the following exploratory hypothesis arises:

*EH1: Positive news will significantly worsen **truth judgements** (i.e., correct distinction between misinformation and real news) in comparison with other sentiments.*

This hypothesis may sound contradictory to the negativity bias hypothesis, however, there is a crucial distinction between the two. Whereas negativity bias would be manifested in the experiment as an increase in the probability to believe headlines with a negative sentiment, the current hypothesis predicts a worsened ability to **discern** between misinformation and true news when exposed to a positive sentiment.

Partisan-Bias and Misinformation Perceptions

Given I also investigate the effect of partisan bias within this experiment, the following research question also arises:

RQ2: How does partisan bias affect participants' truth judgements of misinformation?

To formulate testable hypotheses, I will cover the two dominant accounts which predict how partisan bias would affect misinformation truth judgements. Afterwards, the evidence supporting these accounts will be re-evaluated to create hypotheses for the experiment.

Motivated Reasoning Account of Misinformation

One of the most well-established accounts of misinformation spread is the motivated reasoning account of misinformation (Kahne & Bowyer, 2017). The motivated reasoning account originates from motivated reasoning theory, which posits that one's reasoning is often guided by pre-existing beliefs, or desired outcomes (Kunda, 1990), oftentimes shadowing rationality. This theory explains why, for example, dieters claim that a couple of scoops of ice cream won't harm their dietary goals, smokers are quick to rationalise their habits even in consideration of evidence of its health effects, and a majority of parents believe their children are "unusually gifted" (Epley & Gilovich, 2016).

Put simply, the motivated reasoning account of misinformation suggests that people are more likely to believe misinformation which aligns with their pre-existing political beliefs (ideologically concordant), than that which does not (ideologically discordant). This claim was, until recently, the dominant narrative of political misinformation due to its substantial amount of experimental evidence. For example, in Redlawsk et al. (2010), it was shown that when voters were exposed to negative information on their preferred political candidates, their support for them increased. Yet another example is partisans' refusal to believe fact checks of politically discordant misinformation (Reinero et al., 2023), and people's tendency to engage in heated debates against arguments inconsistent with their political ideology, while at the same time passively and uncritically accepting arguments that align with their political beliefs (Strickland et al., 2011).

With regards to this investigation, the motivated reasoning account of misinformation would predict that partisan bias within the headlines would increase belief in misinformation when it is concordant with the respondents' partisanship, and would do the opposite with politically discordant headlines (Vegetti & Mancosu, 2020).

Classical Reasoning Account of Misinformation

However, in the past couple of years, the motivated reasoning account of misinformation has been under scrutiny in light of the ever-growing popularity of the classical reasoning account of misinformation (Borukhson et al., 2022).

The classical reasoning account stems from Daniel Kahneman's dual-process framework of cognition, in which he broadly divides our decision-making process into two types: automatic (which he refers to as system 1) and reflective (system 2) (Kahneman, 2013). Kahneman states that given our limited cognitive ability and oftentimes need for quick decisions, we mostly rely on system 1 thinking, as it is less cognitively demanding, and works much quicker than the more precautious system 2 (Samuels, 2009). Building on Herbert Simon's work on bounded rationality (Simon & Newell, 1958) Kahneman claims that within system 1 we rely more on heuristics (i.e., mental short-cuts) when making decisions, which, though mostly lead us to the correct decisions, can also result in the over-reliance on them, resulting in the formation of behavioural biases.

Within the context of misinformation, the classical reasoning account states that one's susceptibility to misinformation is influenced by one's ability to engage in reflective (system 2) thinking, not by motivated reasoning (Bago et al., 2020). This has been supported by the various findings correlating cognitive reflection test scores (CRTs) with a variety of behaviour related to susceptibility to misinformation such as paranormal beliefs, the belief of conspiracy theories, and the detection of "pseudo-profound nonsense" (Bago et al., 2020; Pennycook et al., 2012; Pennycook & Rand, 2021). Recently, this correlation has been directly proven within misinformation, as participants' ability to engage in deliberation was shown to be correlated to their ability to discern true from false news (see Bago et al., 2020; Fazio, 2020).

Whilst claiming that misinformation susceptibility is predicted by deliberation, articles which support this account also contradict the motivated reasoning account by repeatedly finding no statistically significant relationship between motivated reasoning and one's accuracy discernment when controlling for deliberation (e.g., Pennycook & Rand, 2019). As Gordon Pennycook and David Rand (the main adversaries of the motivated reasoning account of misinformation) themselves state:

> "People fall for fake news because they *fail* to think; not because they think in a motivated or identity-protective way" (Pennycook & Rand, 2019, p. 9)

Predicting the Impact of Partisan Bias on Misinformation Belief

Considering this literature, it becomes difficult to create testable hypotheses on the potential effect of partisan bias on headline credibility. Though the motivated reasoning account has recently declined in popularity, there remains a continuous stream of work which supports it (e.g., Savolainen, 2022). Similarly, whilst the classical reasoning account of misinformation is currently the dominant narrative of misinformation spread, it too is consistently challenged (see Gawronski, 2021). Some studies altogether reject both accounts (Ceylan et al., 2023).

Looking in more detail at the literature, however, two crucial distinctions emerge. Studies supporting the classical reasoning account of misinformation measure how **truth discernment** is influenced by motivated reasoning, while those supporting the motivated reasoning account measure its impact on **overall belief** (Gawronski, 2021). These two concepts are different, as truth discernment refers to one's ability to accurately judge news, whereas overall belief refers to participants' belief in headlines, regardless of their veracity.

This is problematic, as the existence of motivated reasoning, theoretically speaking, should have a null effect on an individual's discernment abilities. An increase in the belief in ideologically concordant news will result in a decrease in accuracy in discerning false news and an increase in true news, whereas a decrease in the belief of ideologically discordant news would result in an increase in accuracy for false news and a decrease in accuracy for true news.[1]

[1] See Appendix A figure 1 for table explaining null effect

Thus, it seems that the classical reasoning account's dismissal of the motivated reasoning account is founded on a problematic conceptualisation of its manifestation. In fact, within articles which dismiss the existence of motivated reasoning, one can oftentimes see they support the motivated reasoning account when exclusively looking at their results for overall belief (e.g., Pennycook & Rand, 2021). Thus, it seems probable that these accounts are not mutually exclusive, and susceptibility to misinformation could be explained by both a lack of deliberation and motivated reasoning.

Therefore, the following Hypothesis can be formed:

H2.1: Participants will believe politically concordant news ***more*** *than politically discordant news.*

Within the literature, there are also discrepancies regarding how motivated reasoning manifests itself as a partisan bias within misinformation. Some papers such as Pennycook & Rand (2021), define partisan bias as when ideologically concordant news is believed **more than ideologically discordant news**. Other papers, such as Gawronski (2021), define partisan bias as when ideologically concordant news is believed **more than news without a bias**, and ideologically incongruent news is believed **less than news without a bias**. Thus, the following sub-research question arises:

RQ2.1: Does motivated reasoning manifest itself relatively between ideologically discordant and concordant news, or as a bias relative to news with no partisan-leanings?

A wider look at the literature on partisan bias and motivated reasoning also seems to suggest that the effect of motivated reasoning may be most pronounced within animosity towards the out-party (Abramowitz & Webster, 2018). This is suggested by the negative partisanship theory, which posits that partisanship is predominantly founded on a dislike towards the out-party rather than a liking of the in-party (Lee et al., 2022). This has been empirically supported in the US, as affective evaluations of the out-party among US citizens have only worsened, whereas evaluations of the in-party have remained mostly stable (Bankert, 2022). Though negative partisanship has not yet been investigated within motivated reasoning and misinformation, its existence would suggest the following:

EH2. – Participants will believe politically concordant news more when it is negative than when it is not.

Social Norms and Misinformation Judgements

Finally, in investigating the potential effect that like counts may have on participants' accuracy judgments of misinformation, the following research question arises:

RQ3: How does the inclusion of a like count affect participants' truth judgements of misinformation?

Within the context of social media, the inclusion of engagement statistics such as likes on Facebook, or upvotes on Reddit, has often been referred to as aggregated user representations (AURs) (Walther & Jang, 2012). By showing the degree of social endorsement to the user of the social media platform, AURs harness the power of social norms to influence behaviour within these platforms (Eranti & Lonkila, 2015; Sumner et al., 2018). Thus, to understand and predict the impact of the inclusion of likes within this experiment, reviewing the social norms literature is of paramount importance.

Within the literature on social norms, there have been two widely recognised types of norms: descriptive and injunctive social norms (Cialdini et al., 1990). Whereas descriptive social norms simply describe what the social norm is, injunctive norms describe what people should do by insinuating rewards or punishments. Within this experiment, the inclusion of a standard like count could be termed a descriptive social norm, whereas the inclusion of a like count with the mention of friends could be termed a descriptive and injunctive social norm, given that aside from it purely describing what people are more likely to do, it may make participants consider the consequences of disagreeing with their "friends", such as feeling ostracised or excluded.

AURs which act as descriptive norms have been extensively investigated. Within the majority of this research, they have repeatedly been found to increase people's credibility of information in various contexts, such as in news headlines, which are perceived to be more credible with higher votes (Xu, 2013), as well as online health forums, which are perceived to be more credible in the presence of five-star ratings (Jucks & Thon, 2017). These results are shared with the narrower pool of research which focuses on like counts and misinformation, which consistently reports that a greater number of likes increases the credibility of both true and false headlines (Luo et al., 2022). Within this literature, the effect that descriptive AURs have is often termed the *bandwagon heuristic* (Bryanov & Vziatysheva, 2021; Luo et al., 2022).

Nonetheless, this effect seems to disappear when descriptive AURs are investigated within the context of political misinformation (see Luo et al., 2022). This has been speculated to be because the effects of partisanship and pre-existing beliefs on political information have a greater influence than that of the bandwagon heuristic (Guess et al., 2018). Thus, given this investigation focuses on political misinformation, the following is to be expected within the experiment:

H3.1 – The Inclusion of a Like count without friends will not significantly influence partisans.

Taking a closer look, one finds that these findings agree with the broader literature on descriptive social norms, which claims that they will only affect decision-making in situations where the person does not have an explicit preference between the alternatives (Legros & Cislaghi, 2020). Thus, assuming independents have no preference between each party, one would also expect the following in this experiment:

H3.2 - The inclusion of a like count will significantly increase the probability independents choose a headline as the most truthful.

There is a lack of research which directly addresses the impact that like counts with the inclusion of friends, or personalised AURs have on decision-making. However, looking at the wider literature on injunctive social norms seems to suggest that, unlike descriptive social norms, they can influence decision-making regardless of the individual's pre-existing beliefs (Anderson & Dunning, 2014; Legros & Cislaghi, 2020). Within this experiment, this would suggest the following:

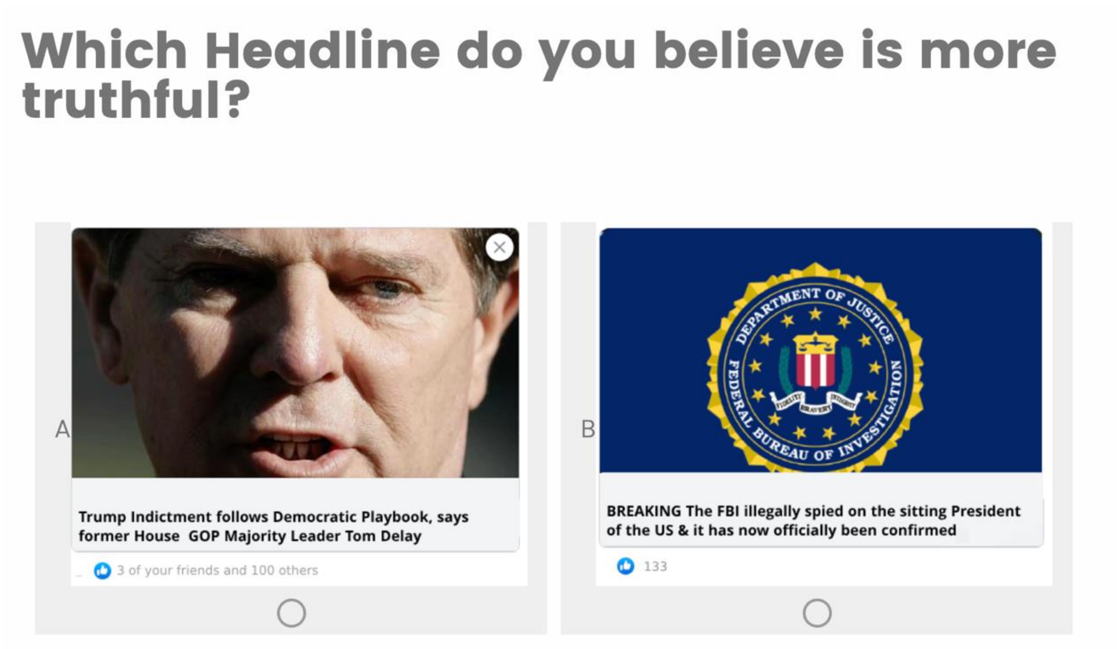
H3.3 – The Inclusion of a like count with friends will significantly increase the probability a headline is chosen for all respondents, regardless of partisanship.

Method

Participant Procedure

Before beginning the task, the survey asked participants to disclose their education level and partisanship through the employment of a 7-level partisanship scale.[2] Participants were then asked to choose which headline they believe "is most truthful" between 18 pairs of screenshots of Facebook news headlines. All these varied in their veracity, sentiment, partisan bias, and their inclusion of a like count. An example question can be seen in Figure 1. As I will explain later, the specific combinations of these attributes and the number of pairs shown were chosen in a way that both maximises the D-efficiency of the design, as well as minimises the costs and cognitive fatigue amongst participants. In the middle of these 18 questions, participants were also asked to complete an attention check.

Figure 1 - Example Question from Survey



[2] See Appendix A Figure 2 for full experiment

Experimental Design

I employ a choice-based conjoint analysis with a pairwise design in this experiment, as recommended by Hainmueller et al. (2015). Choice-based conjoint analyses are a variation of discrete choice experiments, which are based on the theoretical foundations of canonical random utility models. In a conjoint analysis, participants are presented with variations of an item (headlines) whose attributes randomly vary and are asked to choose between the pairs according to certain criteria (in this case truthfulness). This design then allows for a post-hoc analysis of how the attributes of each item affected the participants choices. In this study, the attributes which vary between headlines are the truthfulness of the headline, the sentiment, partisan bias, and the inclusion and type of a like count. The specific combination of attributes, number of pairs, and order in which they are presented were all designed to optimise the D-efficiency of the experiment by using the Stata package dcreate (see Hole, 2017), where it was also calculated that 18 choice sets (pairs) are the lowest number which would allow for a D-efficiency above 0.8. Most importantly, the experiment was approved by a departmental ethics review prior to its release.

I chose a conjoint analysis for various reasons. First, in contrast to other experimental methods, conjoint experiments provide high statistical power with relatively small sample sizes (Hauber et al., 2016). This design was also optimal for the investigation given conjoint analysis' ability to investigate the relative impact of different attributes simultaneously (Carnahan et al., 2022). Moreover, choice-based conjoint analyses allow for the exploration of subgroup effects, such as investigating the effect of the attributes depending on the veracity of the headline, or the participants' partisanship (Eggers et al., 2018).

Materials

I collected a total of 50 (25 True and 25 False) headlines for the experiment. To control for variability in knowledge, motivations, and other extraneous variables, these were all on Donald Trump's Indictments. Out of other potential topics, I chose Donald Trump's indictment for three main reasons: its recency, extensive coverage in the news, and political nature. Its recency holds significance since potential familiarity with the headlines or topic may fail to capture the dynamics of how fake news spreads in real-life scenarios (Pennycook et al., 2021). For example, focusing on COVID-19 misinformation in 2023 may not be an externally valid representation of how misinformation spread during the pandemic given much of the misinformation has been disproven. Second, given headlines had to be collected that fit very specific criteria, choosing a less extensively covered topic would have complicated the collection of headlines which was undesirable given time constraints. Finally, the political nature of the topic was important as it facilitated the observation of pre-existing preferences by measuring partisanship. Whereas with other non-political topics, there is no simple measure of pre-existing beliefs.

Following guidance from Pennycook et al., (2021), all of the misinformation headlines collected are real headlines which have been posted online. Here, by "real" I refer to them having existed in online environments. I primarily did this to preserve the ecological validity of the experiment. Following guidance from Pennycook et al. (2021), the subjects the misinformation touched on were proven to be incorrect by several non-partisan fact-checkers such as Snopes, ATP fact checker, and PolitiFact. I found the headlines from either these same fact-checking organisations (which tend to reference original misinformation posts) or from either Twitter or Facebook by searching for the false subject. In cases where these were not sufficient, I also collected misinformation headlines from news websites with the lowest credibility ratings in iffynews, an extensively used database (e.g., Jiang et al., 2022; Pierri et al., 2023) which evaluates the credibility of news websites with multiple reliability indexes. Given research suggests typographical errors can increase one's detection of misinformation (Sitaula et al., 2020), I also corrected instances where headlines contained misspellings or grammatical mistakes to avoid backdoor variation in their believability.

I collected the true headlines from mainstream media sources, as recommended by Pennycook et al. (2021). Importantly, these were not exclusively fact-checked topics, as with the false headlines. This is because news which has been fact-checked tends to involve dubious claims. Therefore, exclusively collecting true news which has been fact-checked would grey the line between false news and true news, thus obstructing the validity of the experiment.

Once collected, I formatted these headlines to look as if they were posted on Facebook, as extensively done in the literature (e.g. Dias et al., 2020; Pennycook et al., 2018), to further increase the ecological and external validity of the investigation, given much of misinformation spread occurs within social media platforms (Allcott et al., 2019). Finally, given the influence of source credibility as a heuristic to evaluate the veracity of information (Mena et al., 2020), I also removed the source of the headlines to further ensure the results were being directly caused by the independent variables.

Variables

Outcome/Dependent Variable

In the Conjoint Experiment, Participants were asked: “Which headline do you believe is most truthful?”. Thus, the outcome variable is which articles the participants chose, or did not choose within the choice set.

Independent Variables

There was a total of 5 independent variables within the experiment, whose respective levels can be seen in Table 1.

Table 1 - Attributes & Levels

Attributes	Levels
Veracity	True
	False
Sentiment	N+ (Very Negative)
	N (Negative)
	Neu (Neutral)
	P (Positive)
	P+ (Very Positive)
Partisan Bias	L+ (Left Extreme)
	L (Left Moderate)
	Neu (Neutral)
	R (Right Moderate)
	R+ (Right Extreme)
Like Count	No like Count
	Like Count w/o Likes
	Like Count w/ Likes

Veracity – In each choice set, there was always one true and one false headline. Even though the focus of this experiment is misinformation, I included veracity to investigate how the potential effects of the other independent variables may vary between real headlines and misinformation headlines. Moreover, as will be touched on later, the inclusion of veracity as an attribute also allowed me to investigate how these attributes affected the truth discernment of participants.

Sentiment – I measured the sentiment of the headiness using natural language processing python library VADER (short for Valence Aware Dictionary and sEntiment reasoner) (see Hutto & Gilbert, 2014). From other possible sentiment analysis libraries, such as TextBlob and CoreNLP, I chose VADER for two reasons. Firstly, VADER is trained on short-form social media content, whereas other libraries such as TextBlob are trained on long-form articles (see Elbagir & Yang, 2019; C. Hutto & Gilbert, 2014). Given the experiment focuses on headlines, which typically don't exceed one sentence, this model seemed like the most appropriate for the experiment. Moreover, unlike other libraries, VADER is capable of interpreting and analysing informal language, emoticons, and acronyms, all of which are commonly used in misinformation (Iswara & Bisena, 2020). When given text, VADER outputs a score in a continuum from -1 to 1, where -1 represents 100% negative sentiment, and 1 represents 100% positive sentiment. This score is a compound of the positivity, negativity, and neutrality that the program rates each word of the headline. To simplify the analysis process, I created five categories of sentiment as shown in Table 2.

Table 2 - Categorisation of Vader Scores for Sentiment

Sentiment	*Lower Range*	*Upper Range*
Very Negative	-1	-0.5
Negative	-0.5	0
Neutral	0	0
Positive	0	0.5
Very Positive	0.5	1

Partisan-Bias – I measured the partisan bias of the headlines using the political bias AI provided by the bipartisan press (see Wang, 2019). This AI is built upon several natural language processing libraries and is trained on two datasets: AdfontesMedia's list of articles with prelabelled article biases, as well as 10,000 other articles extracted using webhose.io (advanced web data-as-a-service provider). In comparison to other comparable models, such as “readacrosstheaisle” I chose the bipartisan media’s model since it can take specific pieces of texts as input rather than requiring entire articles, allowing for a more accurate evaluation of a headline’s partisan bias. I used the built-in categories with which the model labels bias in the experiment, resulting in a total of five levels: left-extreme, left-moderate, neutral, right-moderate and right-extreme.

Like Counts – I photoshopped Facebook article thumbnails to either not have a like count, include a like count, or include a like count with the mention of friends. Given I cannot personalise the like count for each user and their Facebook friends, rather than mentioning friends in the “like count with friends” level of the attribute, the like count states “3 of your friends and [number of likes] others”. Importantly, I also varied the number of like counts across headlines to increase their ecological validity.

Participants

A total of 500 participants finished the survey (510 pre-attrition). I calculated the number of participants required to achieve sufficient statistical power for the conjoint analysis by running the power analysis recommended by Stefanelli & Lukac (2020), and using an effect size of 0.05 (see Appendix A Figure 3). The participants were hired from the online survey platform Prolific and were required to reside in the United States and speak English to take the survey. The average age of these participants was 25 years old, and most of the participants were American (78%). The participants were evenly distributed across sex, with 50% of them identifying as males, and 50% as females. 8% of the participants were students. Regarding their partisanship, 56.2% of participants identified as democrats, 24.6% identified as republicans, and 19.2% identified as independents. For most participants, the maximum level of education completed was an undergraduate degree (46.6%), followed by high school (31.2%), a graduate degree (18.6%), a professional degree (e.g., JD/MD) (3.4%) and no formal education (0.2%).

Analysis

To analyse the effect that each attribute and its respective level has on participants' truth judgements of an article, I computed the average marginal component effects (AMCEs) for each attribute and its level. Following guidance from Hainmueller et al. (2013), this was calculated by regressing the article choice on all attribute levels using a multiple OLS regression with clustered standard errors on the participants' id to account for the fact that the choices are nested (Christensen, 2021; Hainmueller et al., 2013, 2015). Thus, the linear OLS regression model is:

$$Y_j = \beta_1 Veracity_j + \beta_2 Sentiment_j + \beta_3 Partisan\ Bias_j + \beta_4 Like\ Count_j + \varepsilon_j$$

Where Y = Indicator of whether an article was chosen (dummy variable = 1 if chosen, 0 if not), j = Article.

Veracity, Sentiment, Partisan Bias, Like Count, are all categorical dummied variables representative of their different levels. [3]

In running these regressions, the "baseline level" from which the AMCEs were calculated relative to was set to false news in veracity, neutral in both sentiment and partisan bias, and "no like counts" for the likes.

In more detail, the AMCEs I calculated represent the average change in the probability of selecting an article as the most truthful when comparing the attribute level to the chosen baseline level (Hainmueller et al., 2013). This average is calculated by considering all other possible combinations of the other attributes (Hainmueller et al., 2015). I also calculated AMCEs within subgroups to analyse how the impact of these attributes varies depending on participant characteristics such as partisanship, or on other attributes, such as the veracity of the news. The calculation of AMCEs by subgroups is also oftentimes referred to as average interaction component effects (AICEs) (Christensen, 2021). As recommended by Hainmueller et al. (2013) I visualised all AMCEs through coefficient plots with lines representing confidence intervals.

[3] The full regression model with all dummied categorical variables can be seen in Appendix B Equation 1

Results[4]

To review the results of the experiment, I will first summarise the overall results, followed by the results when analysing true headlines and misinformation headlines separately and when dividing participants by their partisanship. Finally, I will review what the findings suggested concerning the exploratory hypotheses. The overall results of this experiment can also be seen below in Table 3.

Table 3 - Hypotheses and Results

Attribute	*Hypotheses*	*Results*
Sentiment	***H1.1*** *- Headlines with a negative sentiment will be believed significantly more than other sentiments.*	PS
	H1.2*: Headlines with "very" positive, or negative sentiment will be believed less than their moderate counterparts (negative and positive sentiment).*	PS
	EH1: *Positive news will significantly worsen t**ruth judgements** (i.e., correct distinction between misinformation and real news) in comparison with other sentiments.*	PS
Partisan-Bias	***H2.1****: Participants will believe politically concordant news **more** than politically discordant news.*	S
	EH2.*– Participants will believe politically concordant news more when it is negative than when it is not.*	NS
Like count	***H3.1*** *– The Inclusion of a Like count without friends will not significantly influence partisans.*	S

[4] Several sensitivity tests were also implemented by testing the assumptions underlying discrete choice conjoint analyses, as well as the assumptions underlying the use of multiple linear OLS regressions to compute AMCEs (Hainmueller et al., 2013). More specifically, the assumption of heterogeneity of preferences between other respondent characteristics was tested by calculating the AMCEs separately for different respondent characteristics such as age, sex, and education level (see Appendix B figures 1 - 3), all rarely showed any statistically significant difference. Moreover, a leave-one-out sensitivity analysis was conducted by iteratively computing the AMCEs leaving out one choice set at a time, to test for any choice sets having a significant effect on the AMCEs. As shown in Appendix B figure(s) 4, this test only showed significant differences in AMCEs when excluding the 2nd and 18th choice set.

	H3.2 *- The inclusion of a like count will significantly increase the probability independents choose a headline as the most truthful.*	NS
	H3.3 *– The Inclusion of a like count with friends will significantly increase the probability a headline is chosen for all respondents, regardless of partisanship.*	NS

*S = Supported, NS = Not supported, PS = Partly supported

General Overview

Firstly, I analysed the results of the entire sample to provide a broad overview of how attributes and their levels influence the credibility of news. The coefficients in Table 4 and Figure 2 represent the calculated AMCEs for each attribute level, with the lines representing the 95% confidence intervals. Thus, if these lines intersect with the dotted line at zero, the attribute did not have a statistically significant effect ($p>0.05$). Similarly, if the confidence intervals overlap between coefficients, it cannot be said that the coefficients are significantly different. It is important to note that this initial analysis does not differentiate respondents by their partisanship or the veracity of the headlines, thus limiting the hypotheses which can be addressed.

Table 4 – General AMCE Results[5]

Change in Pr(Article chosen as most truthful) for different attribute levels

	AMCEs
Veracity = False	0.00
	(.)
Veracity = True	0.56^{***}
	(0.01)
Sentiment = Neu	0.00
	(.)
Sentiment = N+	-0.03^{**}
	(0.01)
Sentiment = N	0.04^{***}
	(0.01)
Sentiment = P	-0.11^{***}

[5] Table with confidence intervals in Appendix C Table 1

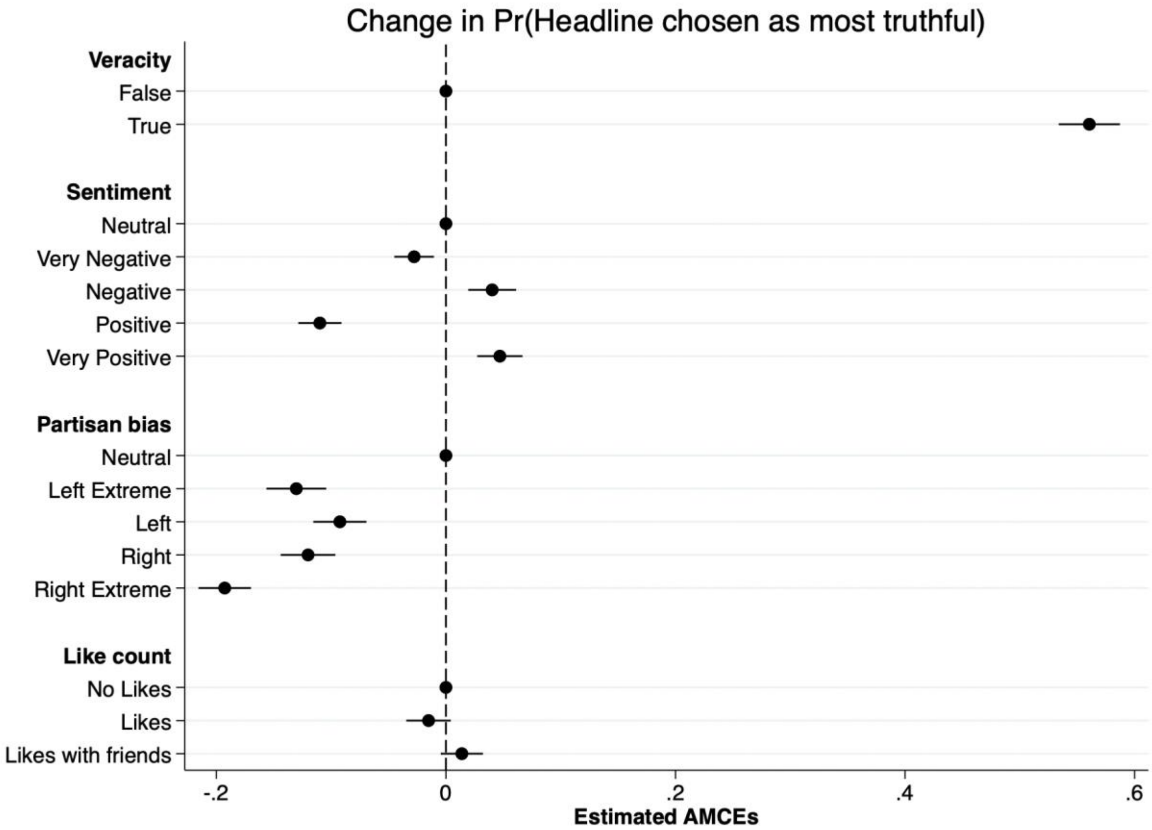
	(0.01)
Sentiment = P+	0.05***
	(0.01)
Partisan-bias = Neu	0.00
	(.)
Partisan-bias = L+	-0.13***
	(0.01)
Partisan-bias = L	-0.09***
	(0.01)
Partisan-bias = R	-0.12***
	(0.01)
Partisan-bias = R+	-0.19***
	(0.01)
No Likes	0.00
	(.)
Likes	-0.02
	(0.01)
Likes with friends	0.01
	(0.01)
Constant	0.34***
	(0.01)
Observations	18000

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Most prominently, the probability an article was chosen as the most truthful of the pair increased by 56 percentage points when such article was a true headline rather than a misinformation headline. This is to be expected, given most participants correctly guessed more than half of the headlines.

Moving on to the sentiment, negative and very positive sentiments seem to be the only ones which increased the probability an article was chosen as the most truthful significantly by 4 and 5 percentage points respectively ($p<0.001$), whereas very negative and positive sentiments decreased this probability by 3 and 11 percentage points respectively in comparison to a neutral sentiment. These results conflict with hypothesis 1.1, which predicted headlines with a negative sentiment would be believed the most. Moreover, hypothesis 1.2. seems to only be partly supported within this analysis, as though headlines with a very negative sentiment are believed less than those with a negative sentiment, the same is not true for positive headlines.

Figure 2 - Plotted AMCEs for General Results



With regards to partisan bias, the results suggest that all forms of partisan bias decreased participants perceived truthfulness of headlines. Right extreme bias had the strongest reduction on the probability participants chose a headline as the most truthful, which reduced this probability by 19 percentage points compared to neutral headlines ($p<0.01$). This is followed by left extreme (-13 percentage points), right moderate (-12 percentage points) and left moderate (-9 percentage points). One can also see a preference within both the extreme and moderate biases for left-leaning biases, though this may only be representative of the greater number of democrats than republicans who partook in the study.

Finally, the inclusion of any form of a like count seems to have no statistically significant influence on the probability a headline is chosen as the most truthful of the pair ($p > 0.05$). This directly disproves hypothesis 3.3., which predicted a significant positive increase in the probability a participant chooses a headline as the most truthful with the inclusion of a like count with friends, regardless of partisanship.

Differences in Attribute Effects between Misinformation and True Headlines

By separating the analysis between true and false headlines, I then investigated potential differences in how these attributes affect the perceived truthfulness between both types of information. These results can be seen in Table 5 and Figure 3. Since this analysis does not yet look at the effects of each participant's partisanship, I will not touch on the effects of partisan bias within this sub-section. However, I included them in the following figures as their potential effects were still controlled for. Overall, the results suggest heterogeneity in the way sentiment and like counts affected truth judgements between true, and misinformation headlines.

Table 5 – AMCE Results splitting by veracity[6]

Change Pr(Article chosen as most truthful) for different attribute levels for Misinformation & True News

	Misinformation	Real News
Sentiment = Neu	0.00	0.00
	(.)	(.)
Sentiment = N+	-0.19^{***}	0.03^{*}
	(0.01)	(0.01)
Sentiment = N	**0.11^{***}**	0.06^{***}
	(0.01)	(0.01)
Sentiment = P	-0.14^{***}	-0.16^{***}
	(0.01)	(0.01)
Sentiment = P+	-0.08^{***}	0.06^{***}
	(0.01)	(0.01)
No Likes	0.00	0.00
	(.)	(.)
Likes	-0.09^{***}	-0.13^{***}
	(0.01)	(0.02)
Likes with friends	0.05^{***}	-0.13^{***}
	(0.01)	(0.02)
Observations	9000	9000

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

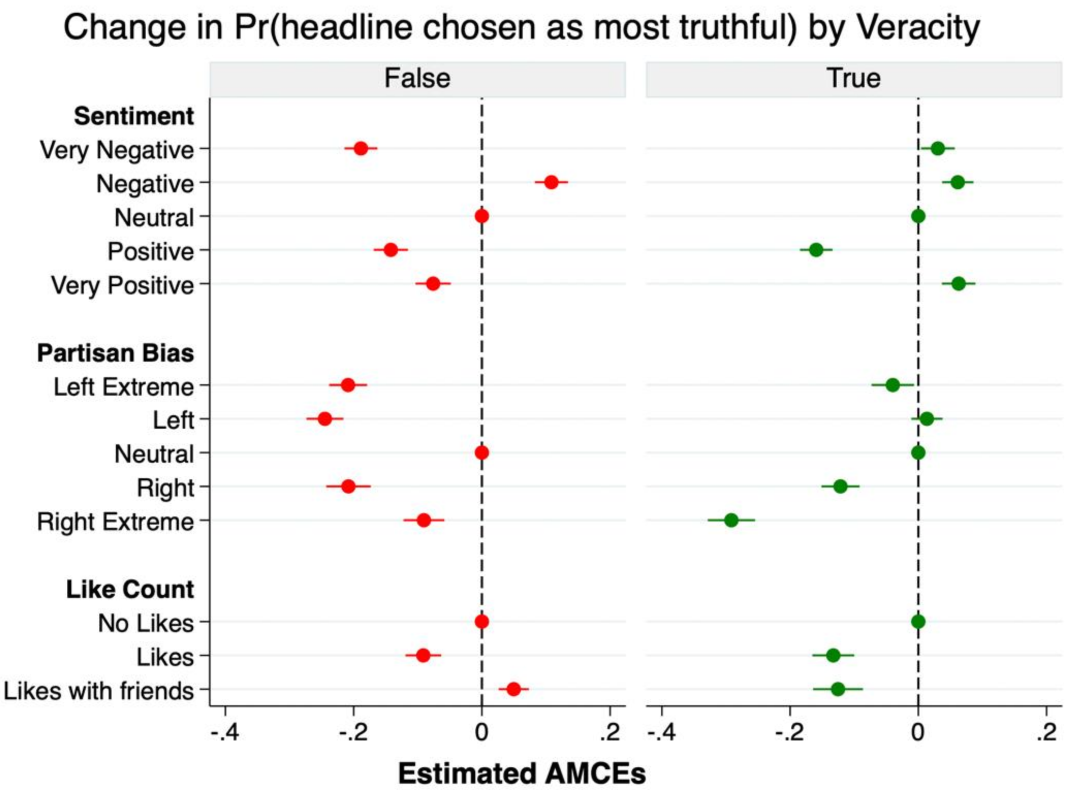
Firstly, the results suggest the existence of a negativity bias within truth judgements of misinformation, as headlines with a negative sentiment (N) were 11 percentage points more likely to be chosen as the most truthful of the pair in comparison to misinformation headlines

[6] Results with confidence intervals in Appendix C Table 2 & 3

with a neutral sentiment ($p < 0.001$). Thus, hypothesis 1.1 is supported, yet only within misinformation headlines. All other sentiments within misinformation headlines led to a decrease in the perceived truthfulness of the headlines compared to those with a neutral sentiment. A very negative sentiment (N+) decreased the probability a headline was chosen as the most truthful by 19 percentage points, a positive sentiment (P) by -14 percentage points, and a very positive sentiment (P+) by -8 percentage points (all $p < 0.001$). These results again only partly support hypothesis 1.2, as though headlines with very negative sentiment are believed less than those with a negative sentiment, the same is not true for their positive counterparts.

Within true news headlines, only a positive sentiment decreased the probability of a headline being chosen, whereas all other sentiments increased the likelihood of the headline being selected as the most truthful. However, these effects were quite small, with very positive and negative sentiments increasing the probability by 6 percentage points, and very negative increasing the probability by 3 percentage points (all $p < 0.05$). Moreover, no support is found for hypothesis 1.2, as neither very negative nor very positive sentiments significantly lowered truth judgements more compared to their moderate counterparts (negative & positive).

Figure 3 - Plotted AMCE Results by Veracity



With regards to like counts, a significant negative effect of the inclusion of a like count without friends on the probability a headline was chosen as most truthful can be seen in both misinformation (-9 percentage points p<.001) and true news (-13 percentage points p<0.001). A like count with friends increased the probability the misinformation headline was chosen by 5 percentage points (p<0.001). However, the opposite is observed within true headlines, where a like count with friends reduced the probability a headline was chosen as the most truthful by 13 percentage points. These results suggest that a like count without friends lowered the truth judgements of both misinformation headlines and true news, whereas a like count with friends marginally increased the probability a misinformation headline was believed, thus partly supporting hypothesis 3.3.

Differences in Attribute Effects by Partisanship

Next, I calculated the AMCEs for each attribute based on the participant's partisanship, to allow for the direct evaluation of how ideology may moderate the effect of the attributes. Since the literature did not suggest any differences in sensitivity to sentiments across ideologies, I will not touch on the effect that these have on each partisan, given the smaller statistical power within these subgroups. The results for this regression using partisanship as subgroups can be seen in both Table 6 and Figure 4.

Table 6 - AMCE Results per Partisanship[7]

Change in the Pr(Article chosen as most truthful) for different levels of each attribute by partisanship

Respondent Partisanship	Independents	Democrats	Republicans
False	0.00	0.00	0.00
	(.)	(.)	(.)
True	0.58^{***}	0.57^{***}	0.53^{***}
	(0.03)	(0.02)	(0.03)
Partisan-Bias = Neu	0.00	0.00	0.00
	(.)	(.)	(.)
Partisan-Bias = LE	-0.14^{***}	-0.07^{***}	-0.27^{***}
	(0.03)	(0.02)	(0.03)
Partisan-Bias = LM	-0.08^{**}	-0.05^{***}	-0.19^{***}
	(0.03)	(0.02)	(0.02)
Partisan-Bias = RM	-0.09^{**}	-0.14^{***}	-0.11^{***}
	(0.03)	(0.01)	(0.03)
Partisan-Bias = RE	-0.19^{***}	-0.20^{***}	-0.17^{***}
	(0.03)	(0.01)	(0.02)
No Likes	0.00	0.00	0.00
	(.)	(.)	(.)
Likes	-0.05^{*}	-0.02	0.02
	(0.02)	(0.01)	(0.02)
Likes with friends	-0.04	0.04^{**}	0.00
	(0.02)	(0.01)	(0.02)
Observations	3456	10116	4428

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Again, concerning partisan bias, a general trend of avoidance of biased news can be seen within all partisanships, as all biases significantly reduced the probability of an article being chosen as the most truthful. With regards to sub-research question 2.1, this suggests that motivated

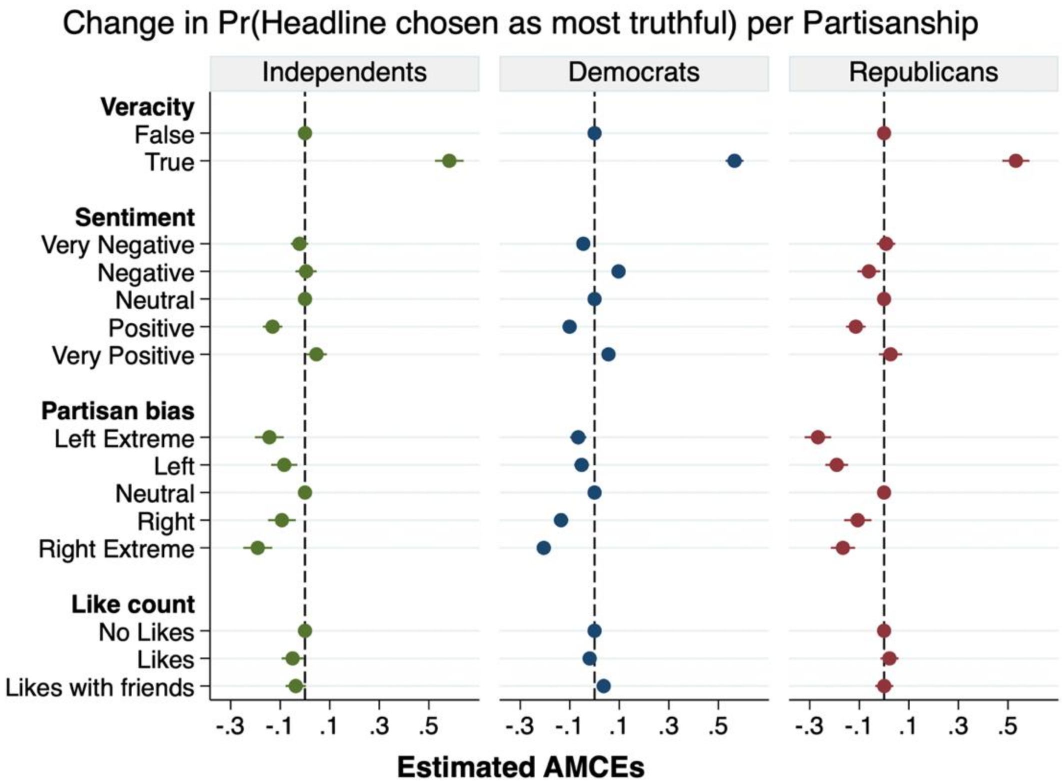
[7] Table with confidence intervals in

Appendix C Table 4, 5 and 6

reasoning does not exist when comparing ideologically concordant/discordant biased news with non-biased news.

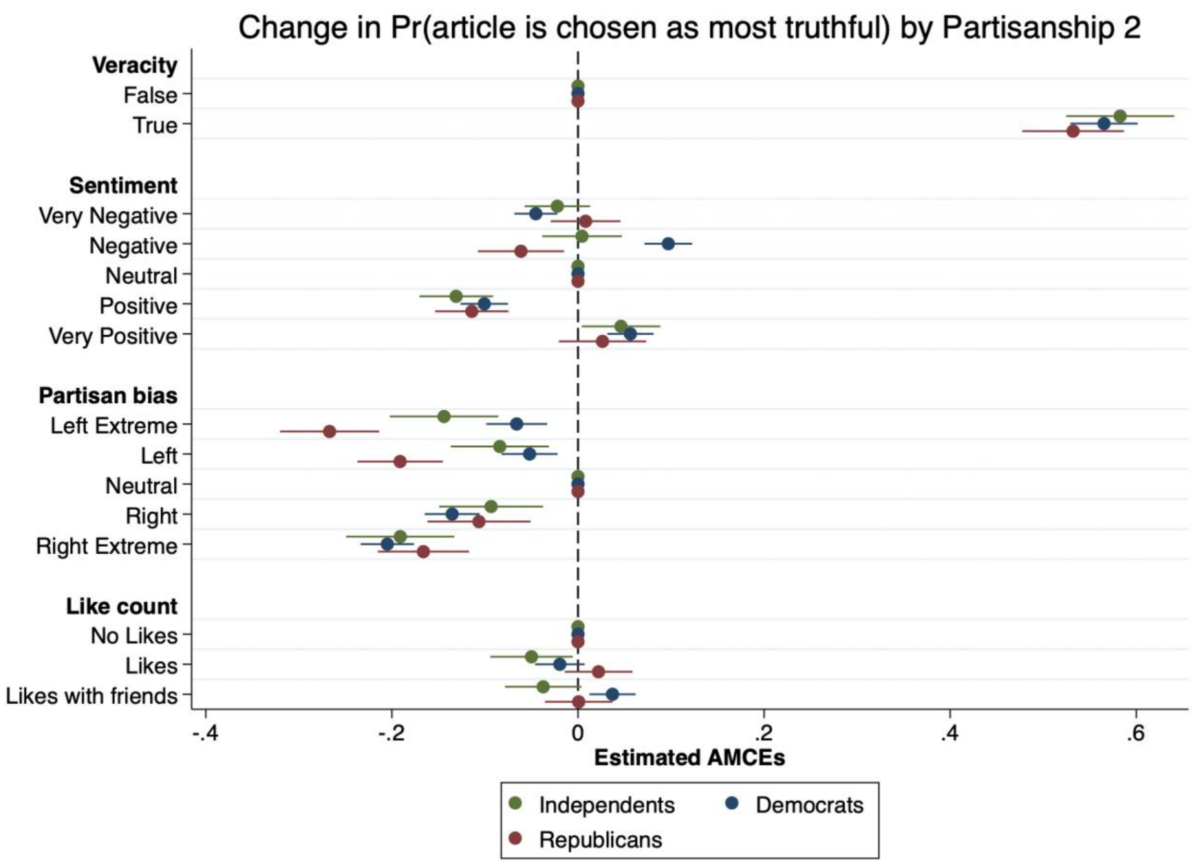
Within democrats, there is a significantly greater avoidance of right-leaning biases, regardless of their magnitude (i.e., moderate, or extreme), as right extreme and right moderate are both the biases which decreased the probability a participant chose a headline as the most truthful the most (-20 and -14 percentage points respectively, p<.001). Left-extreme and left-moderate biases decreased the probability a headline was chosen as most truthful the least, at -5 percentage points and -7 percentage points respectively. However, democrats, on average, show no statistically significant preference between left-extreme news and left-moderate news.

Figure 4 - Plotted AMCEs per Partisanship



Partisan bias seemed to have a different effect on republicans. Though both right moderate and right extreme biases lowered the probability an article was chosen as the most truthful the least (-11 and -17 percentage points respectively, p<0.001), and their left-leaning counterparts lowered it the most, there is only a statistically significant preference of right moderate over left extreme biased news, given these are the only two effects whose confidence intervals don't overlap. That is, unlike democrats, republicans showed less ingroup preference, with only left-extreme bias being statistically significantly preferred over right-moderate bias. It must be noted, however, that this insignificance is mostly a result of republicans' wider confidence intervals, which may have been caused by their underrepresentation in the study. Nonetheless, these results suggest hypothesis H2.1 can only be party supported, as though democrats have a significant preference for ideologically concordant news over discordant news regardless of the extent to which they are biased, republicans only show a statistically significant preference for ideologically concordant news over discordant news when comparing left-extreme biased news to right-moderate biased news. These results can also be seen in Figure 5, where the AMCEs for each partisanship are stacked on top of each other.

Figure 5 - Plotted AMCEs per Partisanship (stacked)



Independents' probability of choosing a headline as the most truthful was reduced the most by right extreme followed by left extreme biases (-19 and -14 percentage points respectively, p<.001), and is least negatively affected by left moderate followed by right moderate biases (-8 and -9 percentage points respectively, p<0.001). However, between these, there seems to be no statistically significant difference. Therefore, Independents show no statistically significant preference towards either party.

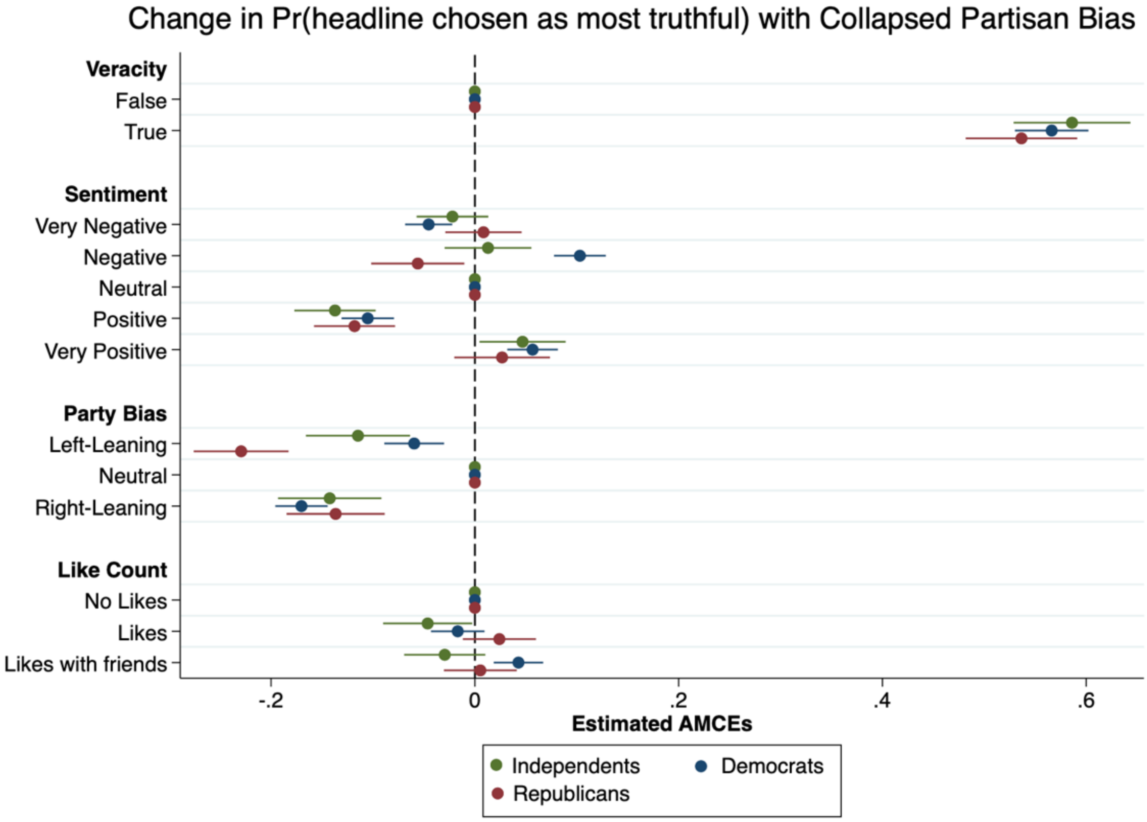
I also used this separation by partisanship to test hypothesis 3.2, which posited that a like count without friends would significantly increase the probability that independents chose a headline as most truthful. As shown in Figure 5 and Table 6, this hypothesis was not supported by the results, as though a like count without friends only had a statistically significant influence on independents, this effect was in the opposite direction to that which was predicted, as its inclusion decreased the probability a headline was selected as the most truthful by 5 percentage points ($p < 0.05$). Nonetheless, the results do support hypothesis 3.1, as a like count without friends had no statistically significant effect on partisans (i.e., republicans or democrats).

To evaluate the overall impact of partisan bias regardless of its strength, I estimated the AMCEs with a collapsed partisan-bias variable, which solely represents the political leaning of the bias, as shown in Figure 6[8]. These results again suggest that democrats were more affected by partisan bias than republicans, as they are 11 percentage points more likely to choose a headline as most truthful when it is left-leaning rather than right-leaning ($p<0.05$). Republicans, on the other hand, showed the greatest sensitivity to partisan bias, as they were more negatively affected by left-leaning and right-leaning bias (-23 and -14 percentage points, $p<0.001$). However there was no statistically significant preference for ingroup versus outgroup biased news, as the confidence intervals for both effects overlap. Independents again show no statistically significant preference for either left or right-leaning news. Thus, hypothesis 2.1 is partly supported when employing a collapsed partisan bias variable, as only democrats show a statistically significant preference for ideological concordant over discordant news.[9]

[8] AMCE results shown in Appendix C Table *7*

[9] Due to concerns regarding the power of the design, I did not further stratify this analysis to see how the effects of partisan bias changed between real, and misinformation headlines. Nonetheless, the results of doing so seem to suggest differences between both types of misinformation as can be seen in Appendix C Figure 1.

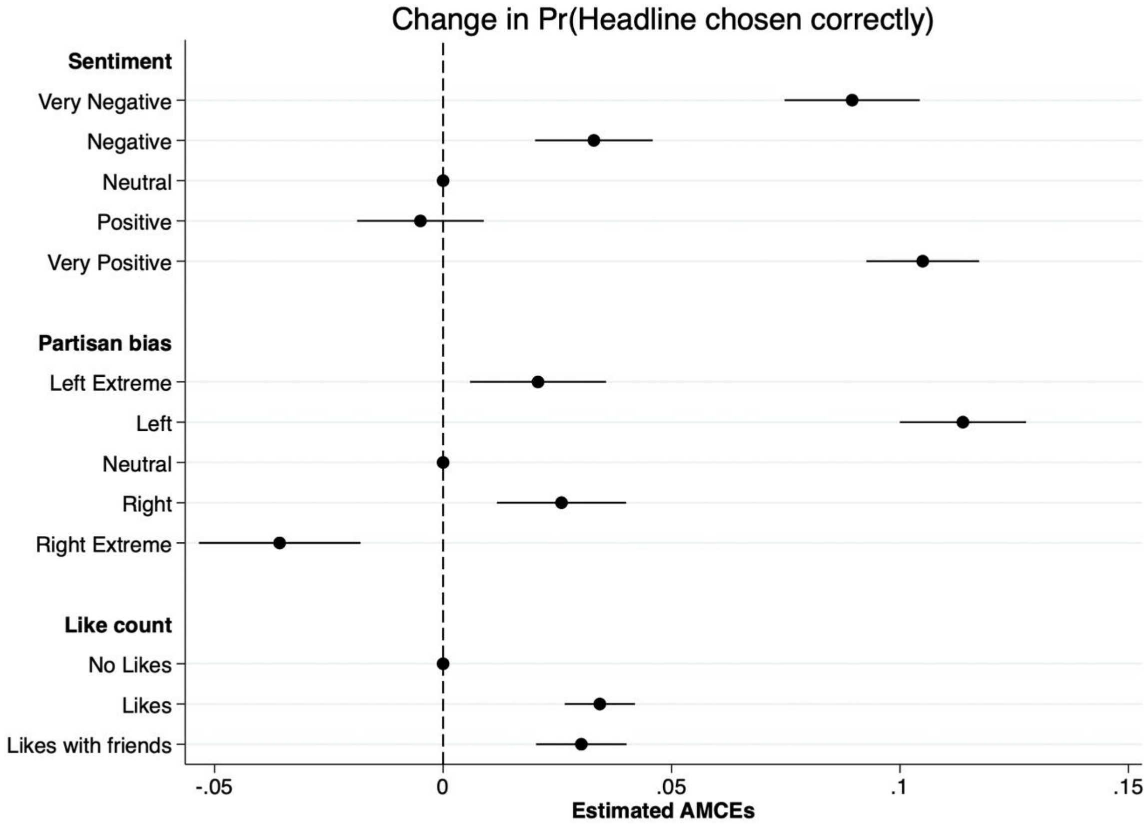
Figure 6 - Plotted AMCEs per partisanship with collapsed Partisan Bias



Exploratory Hypotheses

I tested exploratory hypothesis 1 by introducing a dummy truth discernment variable and calculating the average marginal component effects on the probability a headline was chosen correctly. This variable takes the value of 1 when a headline is correctly chosen (i.e., true headline chosen and misinformation headline not chosen) and 0 when a headline is wrongly chosen (i.e., misinformation headline chosen, and true headline not chosen). The plotted coefficients for this analysis can be seen below in Figure 7. Only partial evidence was found for exploratory hypothesis 1 (EH1), as though positive sentiment decreases the probability of choosing the correct headline less than all non-neutral sentiments, it is not significantly different to the effect neutral sentiment had on the probability a headline is chosen correctly.

Figure 7 - Plotted AMCEs for Discernment



Lastly, I also investigated exploratory hypothesis 2 (EH2) by estimating the AMCEs of left, or right-leaning news (using the collapsed partisan bias variable) for negative, positive, or neutral headlines separately, as shown in Appendix C Table 8 and Appendix C Figure 2. Here, I also created a collapsed sentiment variable to increase the statistical power of the findings. Moreover, I only conducted this analysis for democrats, given their higher sample size in the experiment. The results don't seem to point towards the presence of negative partisanship within motivated reasoning, as democrats did not have a statistically significant greater belief of politically concordant news in headlines with a negative sentiment relative to other sentiments. In fact, the results seem to suggest the opposite effect, as democrats showed the greatest increase in belief in politically concordant news within positive headlines (increase of 15 percentage points, $p<0.05$), thus potentially suggesting the existence of "positive partisanship". Nonetheless, these findings are merely suggestive, as they suffer from a proportionally much smaller sample size and thus hold very limited external validity.

Discussion

Theoretical Implications

This study has found a statistically significant negativity bias within misinformation, as respondents were 11 percentage points more likely to believe a misinformation headline when the headline was presented with a negative sentiment. This effect however did not hold for true news, wherein all sentiments but positive sentiments had a similar positive effect on truth judgements, thus suggesting heterogeneity in our reliance on the negativity bias when judging misinformation and true news. As expected, there was no negativity bias within misinformation headlines with a very negative sentiment, which instead significantly reduced the probability participants chose a misinformation headline as most truthful by 19 percentage points. However, the probability of choosing a headline as most truthful was not significantly lower for those with a very positive compared to those with a positive sentiment. This calls for further investigation, as the existing literature does not provide any contextualisation for these findings.

Furthermore, a greater belief in politically concordant news compared to discordant news was discovered. However, no preference towards ideologically concordant news was found when comparing them to non-biased headlines. These results suggest that motivated reasoning plays an important role in information avoidance within truth judgements of headlines rather than information seeking, as partisans were unlikely to believe any form of biased news, yet even less likely if it is leaning towards the out-group. This effect was more pronounced among participants who identified as democrats rather than republicans, as democrats had a significantly greater probability of believing any left-leaning headlines in comparison to any right-leaning headlines, regardless of the strength of the bias (i.e., moderate, or extreme). However, this may have been due to the underrepresentation of republicans among the respondents. As expected, independents showed no statistically significant partisan bias preference.

A like count with no friends had, as suggested in the literature, no significant effect on the choices that partisans made. However, in contrast to the hypothesised effect, its inclusion caused an average decrease in the probability independent participants chose a headline as most truthful of 5 percentage points. A possible explanation of this is that independents do have preferences, in that their preference is to not support either party. Consequently, this may have resulted in the inclusion of like counts having had a negative effect on them. Yet another possible explanation arises from the literature suggesting that “real” independents don’t actually exist, and most actually are leaners towards a party (see Lascher & Korey, 2011). This may have resulted in independents being a varied pool of partisans, resulting in their subgroup estimation of AMCEs being faulty given it failed to control for their different partisanships.

Contrary to what I hypothesised, a like count with friends did not have a significant effect when considering all participants collectively. Though not directly addressed within the results, this could be due to variability in its impact within different subgroups. This is suggested by the heterogeneity of its influence between true and misinformation headlines, as well as individual differences in how different partisans are affected by it as one can see in Figure 3 and Figure 4. However, this warrants future investigation, given the lack of statistical power to explore this further within this study, as well as the lack of evidence which would suggest this within the literature.

Finally, the findings suggest the existence of positive partisanship within motivated reasoning, wherein participants showed greater belief of politically concordant news within headlines with a positive sentiment. This conflicts with the expected negative partisanship, and as suggested by recent studies within other domains (Lee et al., 2022), posits that positive partisanship may be more prevalent within misinformation. However, this requires further research given the limited sample size with which I conducted this subgroup analysis. Furthermore, though I found no direct evidence supporting positive sentiment worsening participants ability to discern misinformation, this may have been due to variability in the extent to which headlines invoked positive affect within the participants, thus also warranting future research.

Practical Implications

The findings of this study hold significant practical implications. Firstly, this study has provided new insights on how sentiment may affect truth judgements of misinformation through the discovery of a negativity bias within misinformation headlines, as well as partisan bias’s role within misinformation avoidance. In doing so, this study will have hopefully contributed to a greater understanding of the causes of misinformation spread, and consequently, the development of new strategies and interventions to combat the dissemination of misinformation. Perhaps most importantly, this study will have served as an incentive for academics to further investigate the potentially crucial role that sentiment and emotion play within misinformation spread.

Moreover, this investigation suggests new, exciting research areas which may further our understanding of misinformation. Particularly, this investigation advocates future investigation of partisan differences in social norm susceptibility and positive partisanship within motivated reasoning. The future investigation of these may also lead to significant advances in the prevention of misinformation spread.

On a purely methodological basis, through the employment of two distinct machine learning models to measure partisan bias and sentiment, this investigation has also shed light on more replicable ways to measure and evaluate the effect of seemingly subjective variables on human behaviour within Behavioural Sciences. Aside from potentially motivating other researchers to do the same, this may also further incentivise researchers to explore how similar models can be used to measure other behaviours.

Limitations

Firstly, though I calculated the sample size a-priori, the statistical power of some analyses done was low when dividing the participants into subgroups. This was especially troubling when I divided the participants by their partisanship, where there was an underrepresentation of both republicans and independents, given they only consisted of 24.6% and 19.2% of the total sample. Given in many instances the AMCEs were relatively small, this may have resulted in some Type II errors within the independent and republican subgroups, as the lower sample may have led to larger confidence intervals than those within the democrat subgroup. Moreover, the limited sample size also restricted the granularity of the investigation. For example, given the effect of attributes such as sentiment and like counts varied depending on both the veracity of the headline, as well as the partisanship of the participant, it would have been ideal to further calculate the AMCEs between veracity and partisanships simultaneously. However, this would have resulted in an extremely small sample size with very limited external validity.

Moreover, though all headlines were on Donald Trump's Indictment, it is important to acknowledge the variability in the believability of these headlines. Aside from the variance in sentiment and partisan bias, the sub-topic of the indictment which the headlines touched on varied, and some topics may have been much more believable than others. In turn, this may have introduced some noise within the results. The leave-one-out sensitivity analysis seems to suggest this may have been the case within the 2nd and 18th pair participants were presented with[10]. However, this was inevitable for two reasons. First, presenting headlines on the same sub-topic throughout the conjoint analysis may have potentially led to carryover effects, potentially heightening participants' ability to discern the veracity of the headline (Pennycook et al., 2018; Smelter & Calvillo, 2020). Second, since I exclusively used real-world headlines, there was a limited number of headlines which were available for selection for the experiment. When narrowing the existing pool down further to meet the criteria of sentiment and partisan bias for the experiment, it was impossible to find headlines that all touched on similar sub-topics. Nonetheless, these all seemed like appropriate trade-offs considering the greater ecological validity that using real headlines results in.

Not including a "neither" option within the conjoint analysis could also be seen as a limitation of the study. Many academics have pointed towards the value of allowing participants to not make a choice when presented with a choice set in a conjoint analysis (see Haaijer et al., 2001). One could argue that obligating participants to choose either of the profiles fails to mimic how misinformation is encountered in the real world and may therefore not capture how the attributes affect the participants' preferences accurately. However, studies which investigate how the inclusion of a no-choice option in a D-efficient conjoint analysis affects results compared to those which don't include this option find only marginal effects on the estimation, and prediction accuracy of the results (Vermeulen et al., 2008).

Expanding on the ecological validity of the study, an important limitation lies in its attempt at replicating a social media platform rather than directly conducting the research within one. This approach impacts the generalisability of the findings to real social media settings. Aside from potential variations in participant behaviour between the experiment and their actual social

[10] See Appendix B Figure(s)

media usage, this discrepancy might considerably affect the estimated impact of a like count with friends on truth judgments. In a real social media context, such likes would have included the participants' friends, likely intensifying their influence on decision-making processes through the creation of a more pronounced external obligation as suggested by the literature on personalised social norm nudges (Marley et al., 2016) and injunctive social norms (Legros & Cislaghi, 2020).

Future Directions

Firstly, a replication of the current study with a much larger participant pool and equally represented partisanships would yield insights with greater confidence, as well as allow for a more granular investigation of how the effects of the attributes may vary depending on the veracity of the headline and the partisanship of the respondents. More specifically, this would also allow for a deeper exploration of how the effect of partisan bias on different partisanships may change depending on the veracity of the news. All the while, this greater granularity enabled by the larger sample size may also find distinct effects within other sentiments aside from a negative sentiment which may have gone undiscovered or unmentioned within the current study.

Further studies could also look at how an individual's tendency to engage in motivated reasoning changes depending on their partisan strength as well as their affective polarisation. This may be especially interesting given the ever-growing literature indicating the importance of both characteristics within political behaviour (Luttig, 2018). Insofar as a large enough sample is used, this could also be done by replicating the current experiment and using a folded variable of partisanship for partisan strength, and accordingly creating subgroups to investigate how the effects of each attribute vary between them.

Moreover, a more ecologically valid investigation of the role of like counts within misinformation perceptions could also be undergone, potentially with the employment of experiments within social media platforms (see Mosleh & Pennycook, 2021).

Finally, several investigations could be conducted to further explore the exploratory hypotheses within this study. More specifically, experiments could investigate how headlines with a positive sentiment induce a positive affect within participants of the experiment, and consequently, how this positive affect may affect truth judgements of headlines. Also, the aforementioned replication of this investigation with a larger sample size would allow for a more in-depth investigation of the suggested existence of positive partisanship.

Conclusion

This study investigates the role that sentiment within headlines plays within misinformation spread, whilst also adding valuable new insights into the role that motivated reasoning and like counts play. To do so, this investigation employed a conjoint analysis with considerable ecological validity, given its use of real-world headlines. This study finds the existence of a negativity bias within misinformation truth judgements, as well as information avoidance motivated reasoning, wherein partisans are unlikely to believe any form of politically biased news, yet even less likely to believe out-group biased news. Nonetheless, its several limitations suggest the need to further investigate these attributes, potentially allowing for a more granular and comprehensive investigation of the role that all these attributes play, and how their effect differs between partisanships and information types.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), 205316801984855. https://doi.org/10.1177/2053168019848554

Anderson, J. E., & Dunning, D. (2014). Behavioral Norms: Variants and Their Identification. *Social and Personality Psychology Compass*, *8*(12), 721–738. https://doi.org/10.1111/spc3.12146

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Bankert, A. (2022). The Personality Origins of Positive and Negative Partisanship. *Politics and Governance*, *10*(4), 299–310. https://doi.org/10.17645/pag.v10i4.5719

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Borukhson, D., Lorenz-Spreen, P., & Ragni, M. (2022). When Does an Individual Accept Misinformation? An Extended Investigation Through Cognitive Modeling. *Computational Brain & Behavior*, *5*(2), 244–260. https://doi.org/10.1007/s42113-022-00136-3

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLOS ONE*, *16*(6), e0253717. https://doi.org/10.1371/journal.pone.0253717

Carnahan, D., Ulusoy, E., Barry, R., McGraw, J., Virtue, I., & Bergan, D. E. (2022). What should I believe? A conjoint analysis of the influence of message characteristics on belief in, perceived credibility of, and intent to share political posts. *Journal of Communication*, *72*(5), 592–603. https://doi.org/10.1093/joc/jqac023

Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, *120*(4), e2216614120. https://doi.org/10.1073/pnas.2216614120

Christensen, H. S. (2021). A conjoint experiment of how design features affect evaluations of participatory platforms. *Government Information Quarterly*, *38*(1), 101538. https://doi.org/10.1016/j.giq.2020.101538

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and Explicit Evaluation: FMRI Correlates of Valence, Emotional Intensity, and Control in the Processing of Attitudes. *Journal of Cognitive Neuroscience*, *16*(10), 1717–1729. https://doi.org/10.1162/0898929042947919

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-001

Eggers, F., Sattler, H., Teichert, T., & Völckner, F. (2018). Choice-Based Conjoint Analysis. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of Market Research* (pp. 1–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-05542-8_23-1

Elbagir, S., & Yang, J. (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. *Hong Kong*.

Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, *30*(3), 133–140. https://doi.org/10.1257/jep.30.3.133

Eranti, V., & Lonkila, M. (2015). The social significance of the Facebook Like button. *First Monday*, *20*(6). https://doi.org/10.5210/fm.v20i6.5505

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-009

Forgas, J. P. (2019). Happy Believers and Sad Skeptics? Affective Influences on Gullibility. *Current Directions in Psychological Science*, *28*(3), 306–313. https://doi.org/10.1177/0963721419834543

Forgas, J. P., & Eich, E. (2012). Affective Influences on Cognition. In *Handbook of Psychology, Second Edition*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118133880.hop204003

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Gawronski, B. (2021). Partisan bias in the identification of fake news. *Trends in Cognitive Sciences*, *25*(9), 723–724. https://doi.org/10.1016/j.tics.2021.05.001

Ghanem, B., Rosso, P., & Rangel, F. (2020). An Emotional Analysis of False Information in Social Media and News Articles. *ACM Transactions on Internet Technology*, *20*(2), 19:1-19:18. https://doi.org/10.1145/3381750

Guess, A., Nyhan, B., & Reifler, J. (2018). *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign* (United States of America) [Report]. Dartmouth College. https://apo.org.au/node/126961

Haaijer, R., Kamakura, W. A., & Wedel, M. (2001). *The 'No-Choice' Alternative in Conjoint Choice Experiments* (SSRN Scholarly Paper 2395381). https://papers.ssrn.com/abstract=2395381

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400. https://doi.org/10.1073/pnas.1416587112

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2013). *Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments* (SSRN Scholarly Paper 2231687). https://doi.org/10.2139/ssrn.2231687

Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G. M., Prior, T., Marshall, D. A., Cunningham, C., IJzerman, M. J., & Bridges, J. F. P. (2016). Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health*, *19*(4), 300–315. https://doi.org/10.1016/j.jval.2016.04.004

Hilbig, B. E. (2012). Good Things Don't Come Easy (to Mind). *Experimental Psychology*, *59*(1), 38–46. https://doi.org/10.1027/1618-3169/a000124

Hole, A. (2017). *DCREATE: Stata module to create efficient designs for discrete choice experiments* [Computer software]. https://econpapers.repec.org/software/bocbocode/S458059.htm

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14550

Iswara, A. A., & Bisena, K. A. (2020). Manipulation And Persuasion Through Language Features In Fake News. *RETORIKA: Jurnal Ilmu Bahasa*, *6*(1), Article 1. https://doi.org/10.22225/jr.6.1.1338.26-32

Jiang, B., Sheth, P., Li, B., & Liu, H. (2022). *CoVaxNet: An Online-Offline Data Repository for COVID-19 Vaccine Research*.

Jucks, R., & Thon, F. M. (2017). Better to have many opinions than one from an expert? Social validation by one trustworthy source versus the masses in online health forums. *Computers in Human Behavior*, *70*, 375–381. https://doi.org/10.1016/j.chb.2017.01.019

Kahne, J., & Bowyer, B. (2017). Educating for Democracy in a Partisan Age: Confronting the Challenges of Motivated Reasoning and Misinformation. *American Educational Research Journal*, *54*(1), 3–34. https://doi.org/10.3102/0002831216679817

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st ed.).

Karduni, A., Wesslen, R., Markant, D., & Dou, W. (2021). Images, Emotions, and Credibility: Effect of Emotional Facial Images on Perceptions of News Content Bias and Source Credibility in Social Media. *ArXiv*. https://www.semanticscholar.org/paper/Images%2C-Emotions%2C-and-Credibility%3A-Effect-of-Facial-Karduni-Wesslen/2de46b740613722db6b0ac9916eb06fa75af98b0

Koch, A. S., & Forgas, J. P. (2012). Feeling good and feeling truth: The interactive effects of mood and processing fluency on truth judgments. *Journal of Experimental Social Psychology*, *48*(2), 481–485. https://doi.org/10.1016/j.jesp.2011.10.006

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *Journal of Politics*, *62*(3), 790–816. https://doi.org/10.1111/0022-3816.00033

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lascher, E. L., & Korey, J. L. (2011). The Myth of the Independent Voter, California Style. *California Journal of Politics and Policy*, *3*(1), 1–20. https://doi.org/10.5070/P20W2R

Lee, A. H.-Y., Lelkes, Y., Hawkins, C. B., & Theodoridis, A. G. (2022). Negative partisanship is not more prevalent than positive partisanship. *Nature Human Behaviour*, *6*(7), Article 7. https://doi.org/10.1038/s41562-022-01348-0

Lee, J., Kanthawala, S., Britt, B. C., Deavours, D. F., & Ott-Fulmore, T. (2021). Prevalence of anger, engaged in sadness: Engagement in misinformation, correction, and emotional tweets during mass shootings. *Online Information Review*, *46*(3), 422–440. https://doi.org/10.1108/OIR-03-2021-0121

Legros, S., & Cislaghi, B. (2020a). Mapping the Social-Norms Literature: An Overview of Reviews. *Perspectives on Psychological Science*, *15*(1), 62–80. https://doi.org/10.1177/1745691619866455

Legros, S., & Cislaghi, B. (2020b). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, *15*(1), 62–80. https://doi.org/10.1177/1745691619866455

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), Article 3. https://doi.org/10.1038/s41562-021-01056-1

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*, *49*(2), 171–195. https://doi.org/10.1177/0093650220921321

Luttig, M. D. (2018). The “Prejudiced Personality” and the Origins of Partisan Strength, Affective Polarization, and Partisan Sorting. *Political Psychology*, *39*(S1), 239–256. https://doi.org/10.1111/pops.12484

Marley, S., Bekker, H. L., & Bewick, B. M. (2016). Responding to personalised social norms feedback from a web-based alcohol reduction intervention for students: Analysis of think-aloud verbal protocols. *Psychology & Health*, *31*(9), 1007–1024. https://doi.org/10.1080/08870446.2016.1161192

Mejova, Y. (2019). *Sentiment Analysis: An Overview*.

Melville, P., Gryc, W., & Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1275–1284. https://doi.org/10.1145/1557019.1557156

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *Social Media + Society*, *6*(2), 2056305120935102. https://doi.org/10.1177/2056305120935102

Mosleh, M., & Pennycook, G. (2021). Field Experiments on Social Media. *Current Directions in Psychological Science*, *31*, 096372142110547. https://doi.org/10.1177/09637214211054761

Munezero, M., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text. *IEEE Transactions on Affective Computing*, *5*(2), 101–111. https://doi.org/10.1109/TAFFC.2014.2317187

Norris, C. J. (2021). The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach. *Social Neuroscience*, *16*(1), 68–82. https://doi.org/10.1080/17470919.2019.1696225

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

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

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335–346. https://doi.org/10.1016/j.cognition.2012.03.003

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

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

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pérez-Escolar, M., Lilleker, D., & Tapia-Frade, A. (2023). A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation. *Media and Communication*, *11*(2), 76–87. https://doi.org/10.17645/mac.v11i2.6453

Pierri, F., Luceri, L., Jindal, N., & Ferrara, E. (2023). Propaganda and Misinformation on Facebook and Twitter during the Russian Invasion of Ukraine. *Proceedings of the 15th ACM Web Science Conference 2023*, 65–74. https://doi.org/10.1145/3578503.3583597

Posetti, J., & Matthews, A. (2018). *A short guide to the history of 'fake news' and disinformation*.

Redlawsk, D. P., Civettini, A. J. W., & Emmerson, K. M. (2010). The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"? *Political Psychology*, *31*(4), 563–593. https://doi.org/10.1111/j.1467-9221.2010.00772.x

Reinero, D. A., Harris, E. A., Rathje, S., Duke, A., & Bavel, J. J. V. (2023). *Partisans are more likely to entrench their beliefs in misinformation when political outgroup members fact-check claims*. PsyArXiv. https://doi.org/10.31234/osf.io/z4df3

Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, *36*, 101003. https://doi.org/10.1016/j.jocs.2019.05.009

Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (1st ed., pp. 129–146). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199230167.003.0006

Savolainen, R. (2022, June 15). *What drives people to prefer health-related misinformation? The viewpoint of motivated reasoning.* [Text]. University of Borås. https://doi.org/10.47989/irpaper927

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The Facilitated Processing of Threatening Faces: An ERP Analysis. *Emotion*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Shen, T. J., Cowell, R., Gupta, A., Le, T., Yadav, A., & Lee, D. (2019). How Gullible Are You? Predicting Susceptibility to Fake News. *Proceedings of the 10th ACM Conference on Web Science*, 287–288. https://doi.org/10.1145/3292522.3326055

Simon, H. A., & Newell, A. (1958). Heuristic Problem Solving: The Next Advance in Operations Research. *Operations Research*, *6*(1), 1–10. https://doi.org/10.1287/opre.6.1.1

Sitaula, N., Mohan, C. K., Grygiel, J., Zhou, X., & Zafarani, R. (2020). Credibility-Based Fake News Detection. In K. Shu, S. Wang, D. Lee, & H. Liu (Eds.), *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (pp. 163–182). Springer International Publishing. https://doi.org/10.1007/978-3-030-42699-6_9

Smelter, T. J., & Calvillo, D. P. (2020). Pictures and repeated exposure increase perceived accuracy of news headlines. *Applied Cognitive Psychology*, *34*(5), 1061–1071. https://doi.org/10.1002/acp.3684

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology*, *57*(1), 37–40. https://doi.org/10.1037/h0042782

Stefanelli, A., & Lukac, M. (2020). *Subjects, Trials, and Levels: Statistical Power in Conjoint Experiments*. SocArXiv. https://doi.org/10.31235/osf.io/spkcy

Stephens, K. K., Waller, M. J., & Sohrab, S. G. (2019). Over-emoting and perceptions of sincerity: Effects of nuanced displays of emotions and chosen words on credibility perceptions during a crisis. *Public Relations Review*, *45*(5), 101841. https://doi.org/10.1016/j.pubrev.2019.101841

Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T., & Oscarsson, H. (2022). *Knowledge Resistance in High-Choice Information Environments* (1st ed.). Routledge. https://doi.org/10.4324/9781003111474

Sumner, E. M., Ruge-Jones, L., & Alcorn, D. (2018). A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society*, *20*(4), 1451–1469. https://doi.org/10.1177/1461444817697917

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *New Series*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Vegetti, F., & Mancosu, M. (2020). The Impact of Political Sophistication and Motivated Reasoning on Misinformation. *Political Communication*, *37*(5), 678–695. https://doi.org/10.1080/10584609.2020.1744778

Vendemia, M. A. (2017). (Re)Viewing Reviews: Effects of Emotionality and Valence on Credibility Perceptions in Online Consumer Reviews. *Communication Research Reports*, *34*(3), 230–238. https://doi.org/10.1080/08824096.2017.1286470

Vermeulen, B., Goos, P., & Vandebroek, M. (2008). Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing*, *25*(2), 94–103. https://doi.org/10.1016/j.ijresmar.2007.12.004

Vraga, E. K., & Bode, L. (2020). Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication*, *37*(1), 136–144. https://doi.org/10.1080/10584609.2020.1716500

Walther, J. B., & Jang, J. (2012). Communication Processes in Participatory Websites. *Journal of Computer-Mediated Communication*, *18*(1), 2–15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

Wang, W. (2019, December 22). Calculating Political Bias and Fighting Partisanship with AI. *The Bipartisan Press*. https://www.thebipartisanpress.com/politics/calculating-political-bias-and-fighting-partisanship-with-ai/

Xu, Q. (2013). Social Recommendation, Source Credibility, and Recency: Effects of News Cues in a Social Bookmarking Website. *Journalism & Mass Communication Quarterly*, *90*(4), 757–775. https://doi.org/10.1177/1077699013503158

Zaeem, R. N., Li, C., & Barber, K. S. (2020). On Sentiment of Online Fake News. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 760–767. https://doi.org/10.1109/ASONAM49781.2020.9381323

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *Journal of Data and Information Quality*, *11*(3), 1–37. https://doi.org/10.1145/3309699

Appendices

Appendix A: Experiment Material and explanatory table(s)

Appendix A Figure 1

	Real news	Fake news
Truth discernment		
Ideology congruent	Judged as 'real'	Judged as 'fake'
Ideology incongruent	Judged as 'real'	Judged as 'fake'
Partisan bias		
Ideology congruent	Judged as 'real'	Judged as 'real'
Ideology incongruent	Judged as 'fake'	Judged as 'fake'

From Gawronski (2021)

Appendix A Figure 2

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Introduction and Ethics form

Participant Information Sheet

Thanks for your interest in this project about misinformation. In this page, I give you information about the project and ask for your consent to participate. If you agree, please sign the consent form below stating your name and that you agree to the statements in the table below to give your consent.

What is the study about?

This research will look at the role of different factors in the spread of misinformation.

What will my involvement be?

You will be asked to disclose your level of education and political orientation after whi you will be asked to choose what headline you think is most accurate between 18 sets of headlines.

Do I have to take part?

Participation is **voluntary.** There are no consequences if you decide not to take part in this study. If you decide to take part but then later on you change your mind, you can exit the survey at any time without penalty.

What will my information be used for?

The information you supply will only be used for a Master's Dissertation.

Will my information be anonymous?

Your participation will anonymous - your name will not be used in any reports or publications resulting from the study.

By choosing to Participate, you acknowledge that you are at least 18 years of age, and reside in the United States.

Consent to Survey

- Yes, I wish to participate
- No, I do not wish to participate

Demographic questions & Prolific ID

Please enter your unique Prolific ID

Please note that this response should auto-fill with the correct ID'

${e://Field/PROLIFIC_PID}

How do you politically identify yourself?

Strong Democrat			Neutral			Strong Republican
-3	-2	-1	0	1	2	3

What is the highest level of education that you have completed?

- No formal education
- High School or equivalent
- Undergraduate degree (BA/BSc/Other)
- Graduate degree (MA/MSc/PhD/Other)
- Professional degree (JD/MD/Other)

MAIN EXPERIMENT

In the following questions, you will be presented with pairs of news headlines, and will be asked to choose which you believe is more accurate.

The Study consists of 18 Pairs

Which Headline do you believe is more truthful?

A

Maga, who think that Trump being indicted is going to put him back in the WH, must not realize that prisoners can't run for office

3 of your friends and 330 others

B

Trump's criminal indictment unleashes a bitter new phase in American politics

440

Which Headline do you believe is more truthful?

A

New Yorkers react to Trump indictment with celebration and relief; A great day in American History

120

B

Donald Trump is a Spy

Which Headline do you believe is more truthful?

A

B

Which Headline do you believe is more truthful?

A

Trump indictment turns 2024 election upside down, but not the way Democrats hope

B

Donald Trump Crying because he got caught

170

Which Headline do you believe is more truthful?

A

Trump indictment 'is not good for the country': Mollie Hemingway

B

Donald Trump is not paying a single $1 for his legal team, but he did offer money at a dinner last night to the lawyer that told on him and lost his law license!!!!!!

3 of your friends and 300 others

Which Headline do you believe is more truthful?

A

Trump Indictment follows Democratic Playbook, says former House GOP Majority Leader Tom Delay

3 of your friends and 100 others

B

BREAKING The FBI illegally spied on the sitting President of the US & it has now officially been confirmed

133

Which Headline do you believe is more truthful?

A

He's facing 400 years. The DOJ would never allow Trump to walk free. Equal Justice, Under the Law!

3 of your friends and 165 others

B

Former AG Barr slams 'pathetically weak' legal theory behind Trump Indictment: 'An abomination'

220

Which Headline do you believe is more truthful?

A

B

Which Headline do you believe is more truthful?

A

B

Please select 'Strongly Disagree" in the question below

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Which Headline do you believe is more truthful?

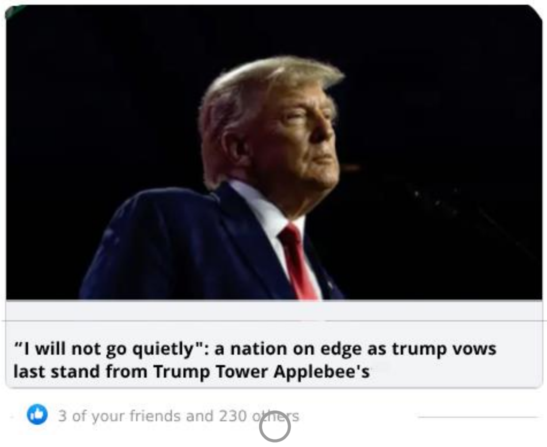
A

B

Which Headline do you believe is more truthful?

A

B



Which Headline do you believe is more truthful?

A

B

Which Headline do you believe is more truthful?

A

B

Which Headline do you believe is more truthful?

A

B

Which Headline do you believe is more truthful?

A	B
 Trump Will Take a Mug Shot on Tuesday and it's Possible He Will Sit in a Jail Cell for '4 to 5 Hours' Before Arraingment	 "Trump Indictment may have given him the 'kiss of life' for 2024", veteran pollster says: 'Like oxygen to fire'
	520

Which Headline do you believe is more truthful?

A	B
 New Yorkers react to Trump indictment with celebration and relief; A great day in American History	 Hillary Clinton INDICTED as Rand Paul REVEALS shocking Evidence EXPOSING her
120	3 of your friends and 210 others

Which Headline do you believe is more truthful?

A

GEORGE SOROS IS LINKED TO TRUMP INDICTMENT, DESPITE NYT LIES

400

B

Trump wont be defeated through Indictments Alone

3 of your friends and 300 others

Which Headline do you believe is more truthful?

A

Donald Trump just announced he will not be paying any of his lawyers a single cent until they convince the DOJ to drop the coming indictment

3 of your friends and 323 others

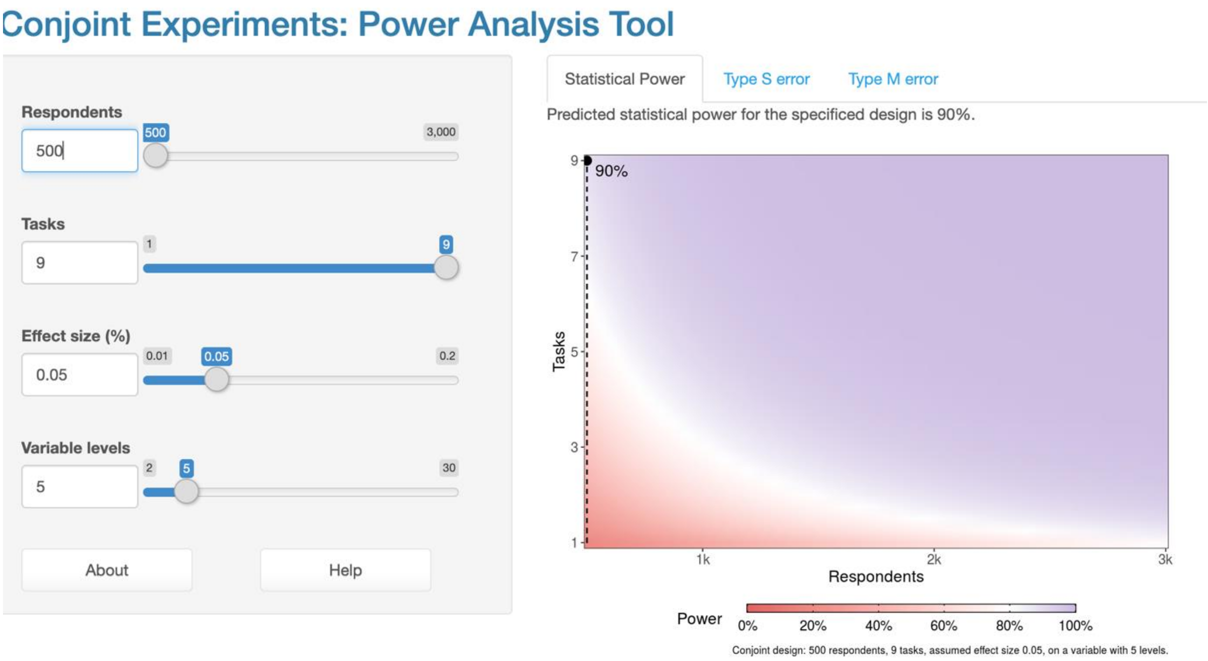
B

Even the Liberal Media Aren't Buying Alvin Bragg's Bogus Trump Indictment

170

Powered by Qualtrics

Appendix A Figure 3



From Stefanelli & Lukac (2020)

from

Appendix B: Sensitivity Analyses

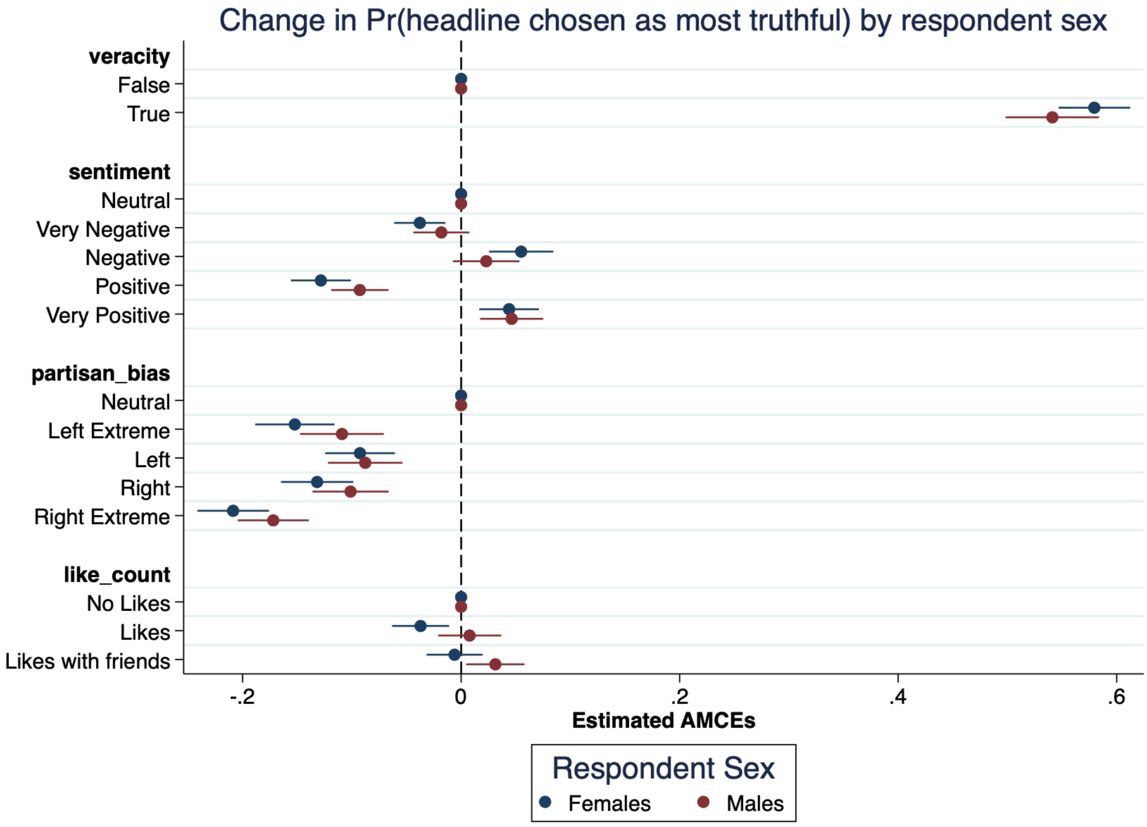
*Appendix B Equation 1**

$$Y = \beta_1 V^1 + \beta_2 S_j^1 + \beta_3 S_j^2 + \beta_4 S_j^3 + \beta_5 S_j^4 + \beta_6 P_j^1 + \beta_7 P_j^2 + \beta_8 P_j^3 + \beta_9 P_j^4 + \beta_{10} L_j^1 + \beta_{11} L_j^2 + \varepsilon_j$$

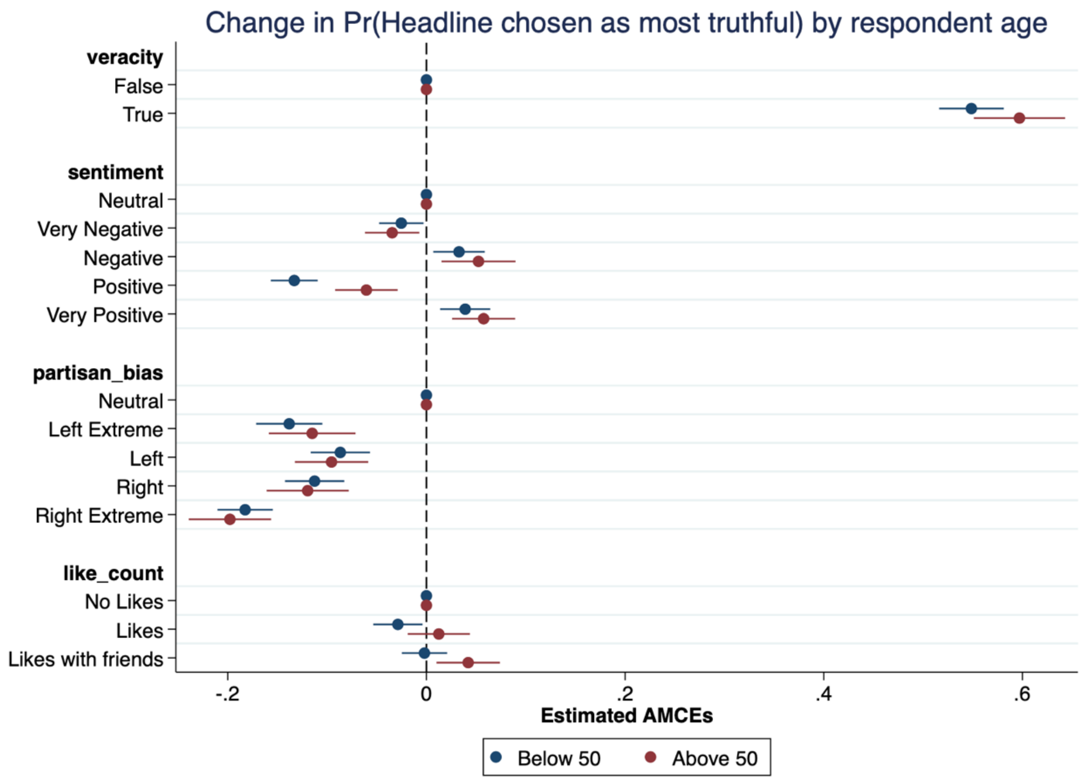
Where V^1= True coded dummy variable, S_j^n = Sentiment levels dummy variables, P_j^n = Partisan-bias levels dummy variables, L_j^n = like count levels dummy variables

*Note: reference level omitted from regression.

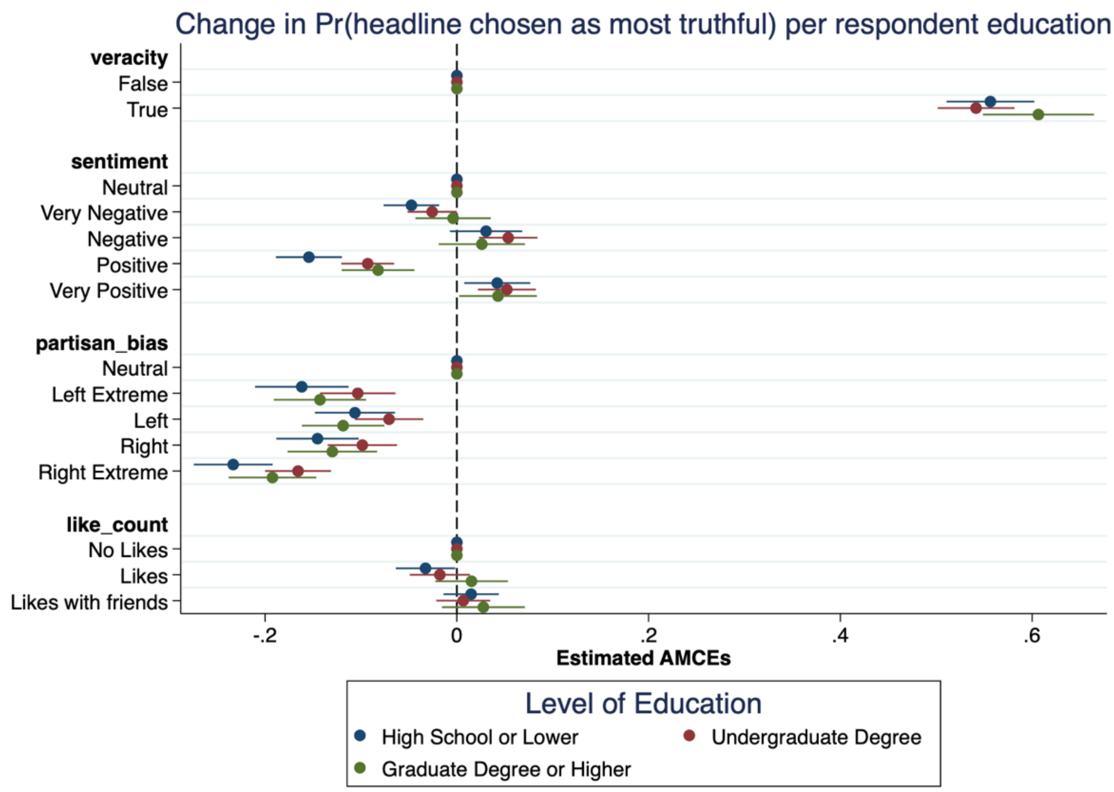
Appendix B Figure 1


Change in Pr(headline chosen as most truthful) by respondent sex
veracity
False
True
sentiment
Neutral
Very Negative
Negative
Positive
Very Positive
partisan_bias
Neutral
Left Extreme
Left
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2
0
.2
.4
.6
Estimated AMCEs
Respondent Sex
Females
Males

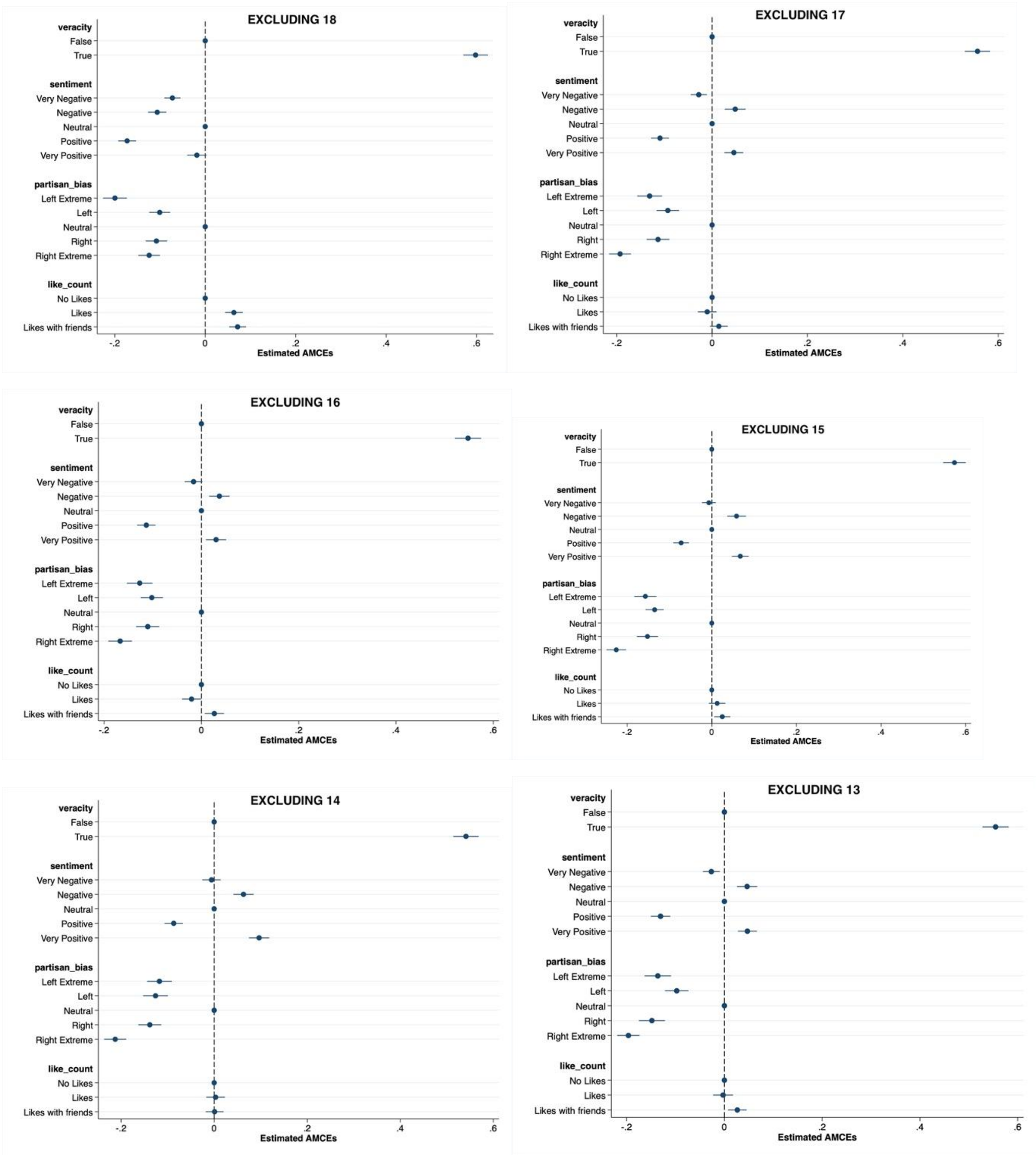
Appendix B Figure 2

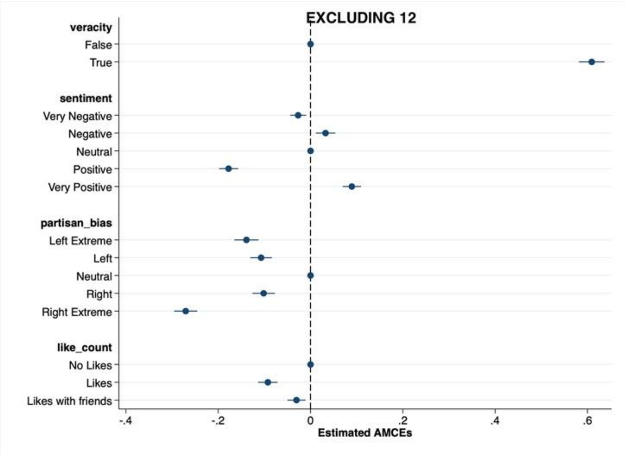


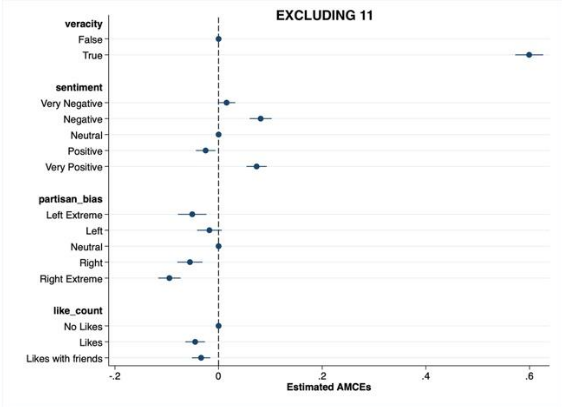
Appendix B Figure 3

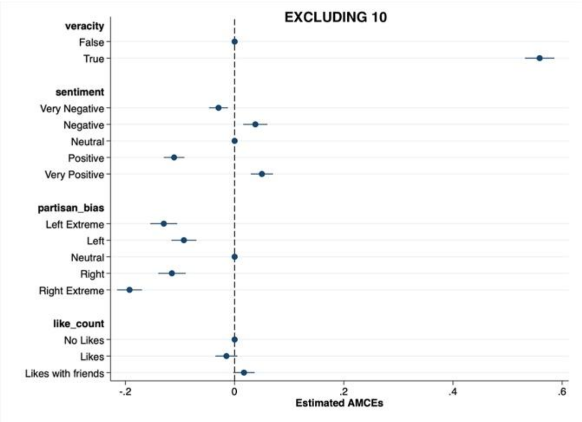


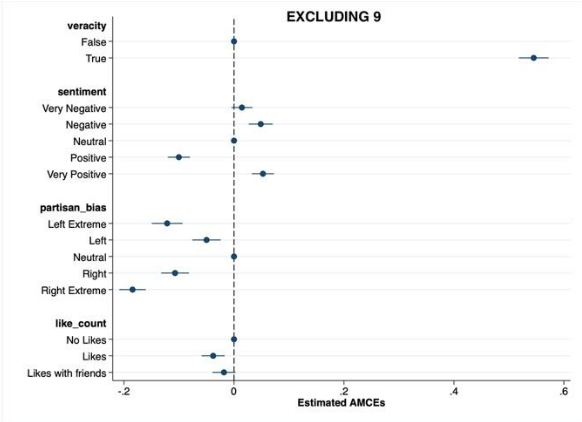
Appendix B Figure(s) 4 - Leave-one-out analysis

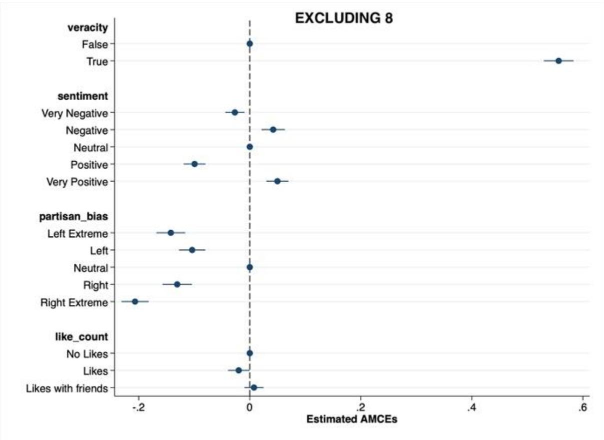


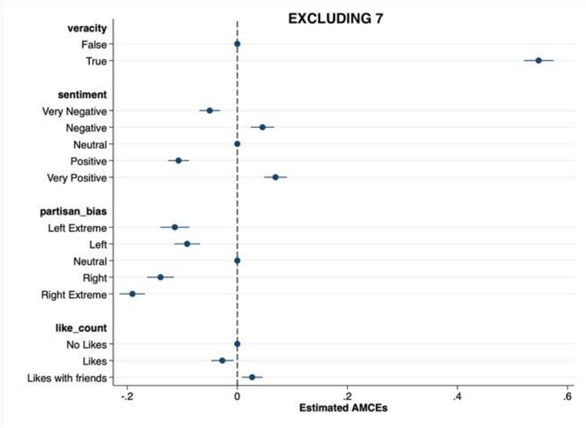

EXCLUDING 12
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.4 -.2 0 .2 .4 .6
Estimated AMCEs

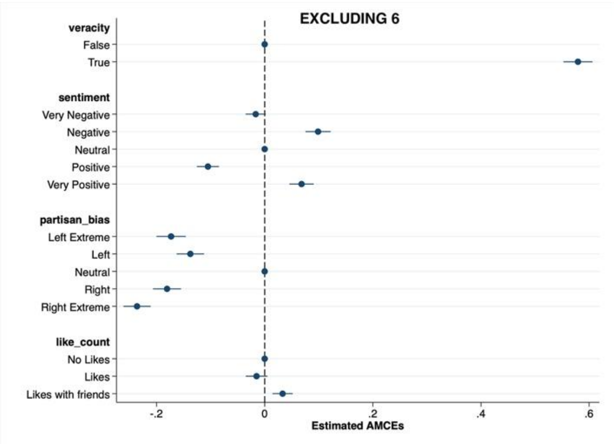

EXCLUDING 11
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs

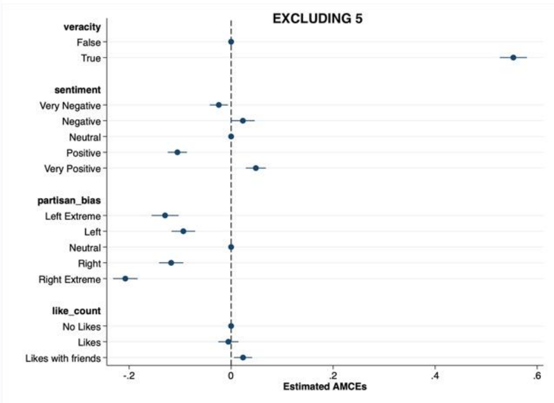

EXCLUDING 10
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs

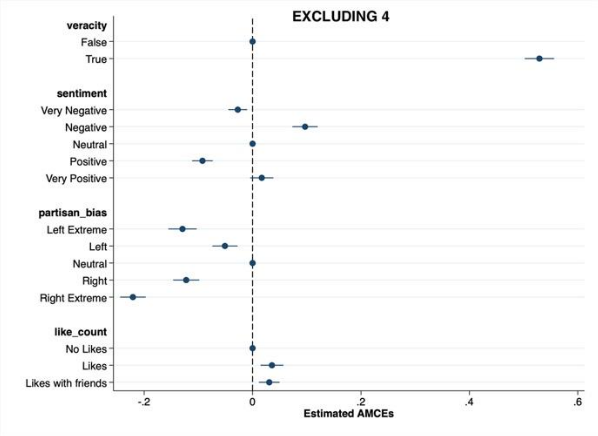

EXCLUDING 9
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs

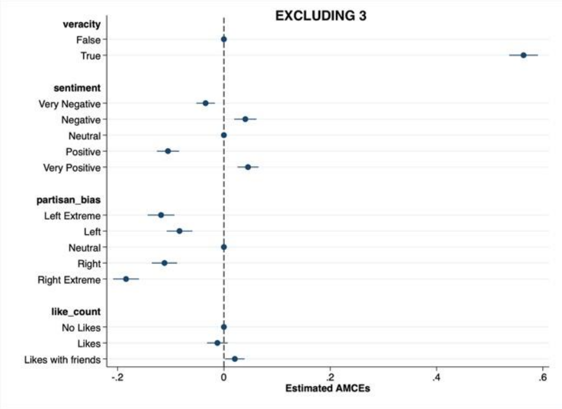

EXCLUDING 8
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs

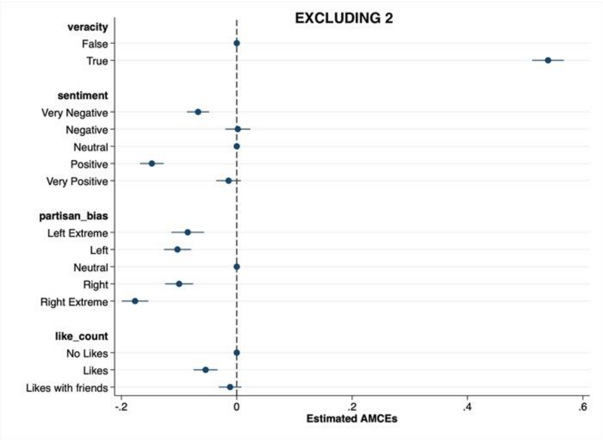

EXCLUDING 7
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs

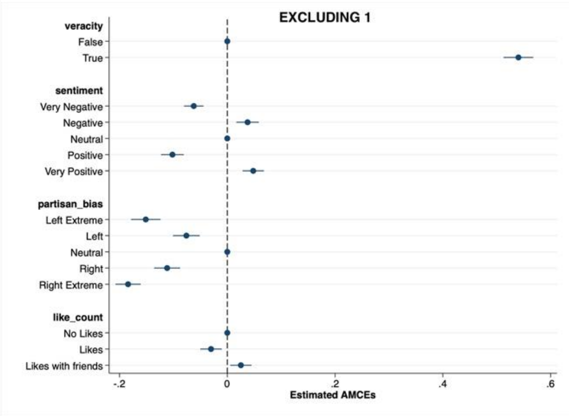

EXCLUDING 6
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs


EXCLUDING 5
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2 0 .2 .4 .6
Estimated AMCEs


EXCLUDING 4
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2
0
.2
.4
.6
Estimated AMCEs


EXCLUDING 3
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2
0
.2
.4
.6
Estimated AMCEs


EXCLUDING 2
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2
0
.2
.4
.6
Estimated AMCEs


EXCLUDING 1
veracity
False
True
sentiment
Very Negative
Negative
Neutral
Positive
Very Positive
partisan_bias
Left Extreme
Left
Neutral
Right
Right Extreme
like_count
No Likes
Likes
Likes with friends
-.2
0
.2
.4
.6
Estimated AMCEs

Appendix C: Additional Figures & Tables

Appendix C Table 1 – Main results with confidence intervals

Change in Pr(Article chosen as most truthful)

	AMCEs	LCI	UC
Veracity = True	0.56^{***}	0.53	0.59
	(0.01)		
Sentiment = N+	-0.03^{**}	-0.0450	-0.0105
	(0.01)		
Sentiment = N	0.04^{***}	0.0194	0.0611
	(0.01)		
Sentiment = P	-0.11^{***}	-0.1286	-0.0910
	(0.01)		
Sentiment = P+	0.05^{***}	0.0272	0.0667
	(0.01)		
Partisan-bias = L+	-0.13^{***}	-0.1564	-0.1043
	(0.01)		
Partisan-bias = L	-0.09^{***}	-0.1155	-0.0693
	(0.01)		
Partisan-bias = R	-0.12^{***}	-0.1449	-0.0963
	(0.01)		
Partisan-bias = R+	-0.19^{***}	-0.2156	-0.1698
	(0.01)		
Likes	-0.02	-0.0345	0.0042
	(0.01)		
Likes with friends	0.01	-0.0046	0.0323
	(0.01)		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 2 – Main results by veracity with CIs - Misinformation

Change in Pr(Article chosen as most truthful) for Misinformation

	Misinformation	LCI	UCI
Sentiment = N+	-0.19***	-0.2144	-0.1630
	(0.01)		
Sentiment = N	**0.11*****	0.0826	0.1344
	(0.01)		
Sentiment = P	-0.14***	-0.1687	-0.1153
	(0.01)		
Sentiment = P+	-0.08***	-0.1035	-0.0485
	(0.01)		
Partisan-Bias = LE	-0.21***	-0.2382	-0.1792
	(0.02)		
Partisan-Bias = LM	-0.24***	-0.2737	-0.2160
	(0.01)		
Partisan-Bias = RM	-0.21***	-0.2430	-0.1733
	(0.02)		
Partisan-Bias = RE	-0.09***	-0.1224	-0.0584
	(0.02)		
Likes	-0.09***	-0.1192	-0.0635
	(0.01)		
Likes with friends	0.05***	0.0261	0.0733
	(0.01)		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 3 – Main results by Veracity – True News

Change in Pr(Article chosen as most truthful) for True News

	Real News	LCI	UCI
Sentiment = N+	0.03*	0.0041	0.0568
	(0.01)		
Sentiment = N	0.06***	0.0371	0.0861
	(0.01)		
Sentiment = P	-0.16***	-0.1847	-0.1338
	(0.01)		
Sentiment = P+	0.06***	0.0367	0.0891
	(0.01)		
Partisan-Bias = LE	-0.04*	-0.0731	-0.0066
	(0.02)		
Partisan-Bias = LM	0.01	-0.0109	0.0378
	(0.01)		
Partisan-Bias = RM	-0.12***	-0.1512	-0.0916
	(0.02)		
Partisan-Bias = RE	-0.29***	-0.3285	-0.2547
	(0.02)		
Likes	-0.13***	-0.1653	-0.1000
	(0.02)		
Likes with friends	-0.13***	-0.1642	-0.0863
	(0.02)		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 4 - MAIN RESULTS BY PARTISANSHIP WITH Cis (Independents)

Change in Pr(Article chosen as most truthful) for Independents

Respondent Partisanship	Independents	LCI	UCI
False	0.00		
	(.)		
True	0.58^{***}	0.5245	0.6409
	(0.03)		
Partisan-Bias = Neu	0.00		
	(.)		
Partisan-Bias = LE	-0.14^{***}	-0.2024	-0.0856
	(0.03)		
Partisan-Bias = LM	-0.08^{**}	-0.1368	-0.0313
	(0.03)		
Partisan-Bias = RM	-0.09^{**}	-0.1492	-0.0375
	(0.03)		
Partisan-Bias = RE	-0.19^{***}	-0.2492	-0.1328
	(0.03)		
No Likes	0.00	-0.2024	-0.0856
	(.)		
Likes	-0.05^{*}	-0.1368	-0.0313
	(0.02)		
Likes with friends	-0.04	-0.1492	-0.0375
	(0.02)		
Constant	0.36^{***}	-0.2492	-0.1328
	(0.03)		
Observations	3456		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 5 - MAIN RESULTS BY PARTISANSHIP WITH Cis (Democrats)

Change in Pr(Article chosen as most truthful) for Democrats

Respondent Partisanship	Democrats	LCI	UCI
False	0.00		
	(.)		
True	0.57^{***}	0.5292	0.6016
	(0.02)		
Partisan-Bias = Neu	0.00		
	(.)		
Partisan-Bias = LE	-0.07^{***}	-0.0986	-0.0332
	(0.02)		
Partisan-Bias = LM	-0.05^{***}	-0.0822	-0.0219
	(0.02)		
Partisan-Bias = RM	-0.14^{***}	-0.1648	-0.1058
	(0.01)		
Partisan-Bias = RE	-0.20^{***}	-0.2336	-0.1763
	(0.01)		
No Likes	0.00		
	(.)		
Likes	-0.02	-0.0462	0.0071
	(0.01)		
Likes with friends	0.04^{**}	0.0122	0.0620
	(0.01)		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 6 - MAIN RESULTS BY PARTISANSHIP WITH Cis (Republicans)

Change in Pr(Article chosen as most truthful) for Republicans

Respondent Partisanship	Republicans	LCI	UCI
False	0.00		
	(.)		
True	0.53***	0.4774	0.5869
	(0.03)		
Partisan-Bias = Neu	0.00		
	(.)		
Partisan-Bias = LE	-0.27***	-0.3202	-0.2137
	(0.03)		
Partisan-Bias = LM	-0.19***	-0.2371	-0.1453
	(0.02)		
Partisan-Bias = RM	-0.11***	-0.1618	-0.0511
	(0.03)		
Partisan-Bias = RE	-0.17***	-0.2152	-0.1171
	(0.02)		
No Likes	0.00		
	(.)		
Likes	0.02	-0.0144	-0.0585
	(0.02)		
Likes with friends	0.00	-0.0356	-0.0371
	(0.02)		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 7 – Change in Pr(Article chosen as most truthful) for Misinformation & True News using collapsed partisan bias value

Pr(Article chosen as most truthful)		
Partisanship	Democrats	Republicans
party_bias = Neutral	0.00	0.00
	(.)	(.)
party_bias = Left-Leaning	-0.06***	-0.23***
	(0.01)	(0.02)
party_bias = Right-Leaning	-0.17***	-0.14***
	(0.01)	(0.02)
Observations	10116	4428

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Table 8 – Results for Exploratory Hypothesis 2

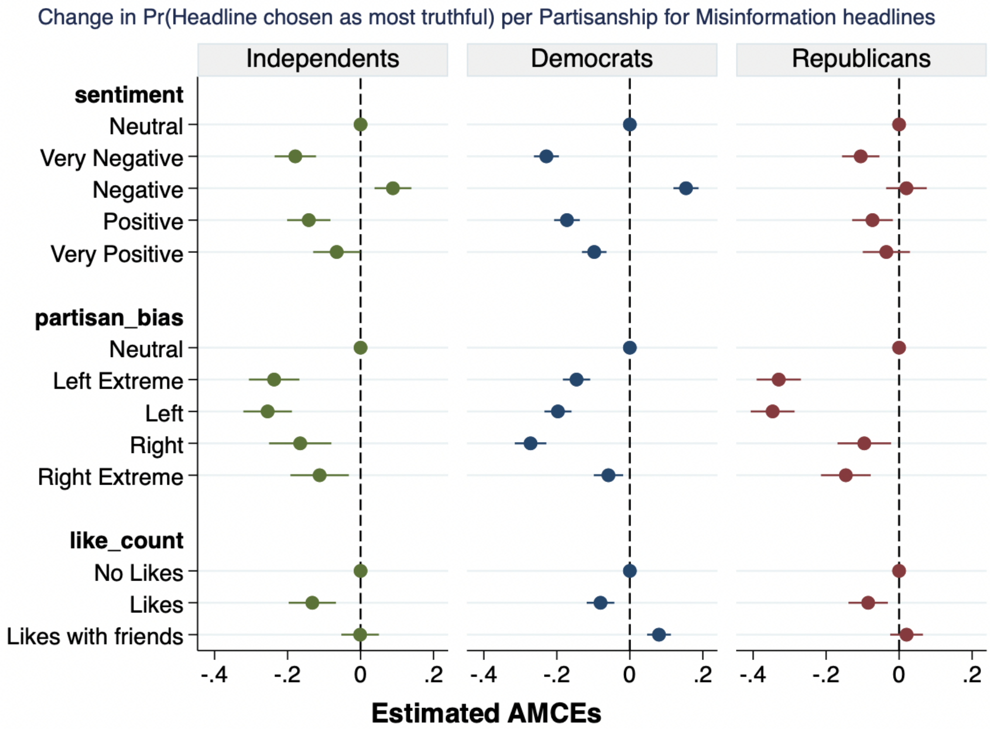
Change in Pr(Article chosen as most truthful) for Democrats per Headline Sentiment

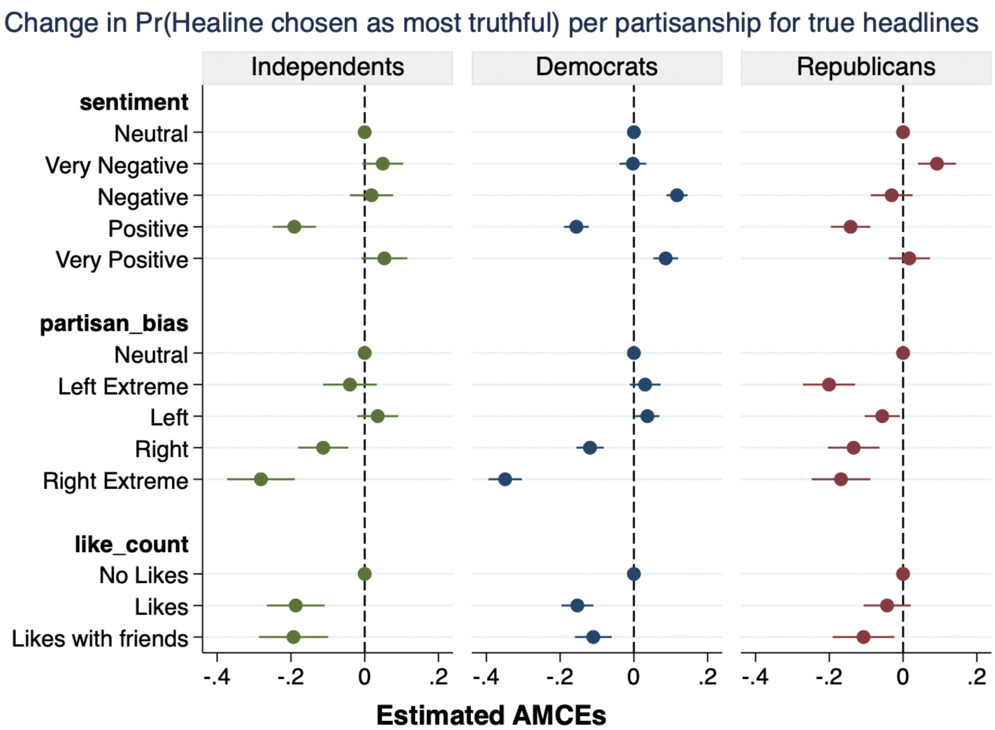
	Neutral Sentiment	Negative Sentiment	Positive Sentiment
False	0.00	0.00	0.00
	(.)	(.)	(.)
True	0.27***	0.63***	0.58***
	(0.03)	(0.02)	(0.02)
Neutral bias	0.00	0.00	0.00
	(.)	(.)	(.)
Left-leaning bias	-0.03	0.01	0.15***
	(0.02)	(0.02)	(0.02)
Right-leaning bias	-0.50***	-0.07***	0.09***
	(0.03)	(0.02)	(0.02)
No Likes	0.00	0.00	0.00
	(.)	(.)	(.)
Likes	0.38***	-0.10***	0.11***
	(0.03)	(0.02)	(0.02)
Likes with friends	0.43***	-0.08***	0.02
	(0.03)	(0.02)	(0.02)
Constant	0.23***	0.30***	0.03
	(0.03)	(0.02)	(0.02)
Observations	2248	3653	4215
Adjusted R^2	0.415	0.361	0.336

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Figure 1





Appendix C Figure 2 – AMCEs for Exploratory Hypothesis 2