MDPI

Article

Assessment of Consumer Perception of Online Content Label Efficacy by Income Level, Party Affiliation and Online Use Levels

Jeremy Straub [1],*, Matthew Spradling [2] and Bob Fedor [1]

1   Department of Computer Science, North Dakota State University, Fargo, ND 58105, USA; robert.fedor@ndsu.edu
2   Department of Mathematics and Applied Sciences, University of Michigan Flint, Flint, MI 48502, USA; mjspra@umich.edu
*   Correspondence: jeremy.straub@ndsu.edu; Tel.: +1-701-231-8196

Abstract: Deceptive online content represents a potentially severe threat to society. This content has shown to have the capability to manipulate individuals' beliefs, voting and activities. It is a demonstrably effective way for foreign adversaries to create domestic strife in open societies. It is also, by virtue of the magnitude of content, very difficult to combat. Solutions ranging from censorship to inaction have been proposed. One solution that has been suggested is labeling content to indicate its accuracy or characteristics. This would provide an indication or even warning regarding content that may be deceptive in nature, helping content consumers make informed decisions. If successful, this approach would avoid limitations on content creators' freedom of speech while also mitigating the problems caused by deceptive content. To determine whether this approach could be effective, this paper presents the results of a national survey aimed at understanding how content labeling impacts online content consumption decision making. To ascertain the impact of potential labeling techniques on different portions of the population, it analyzes labels' efficacy in terms of income level, political party affiliation and online usage time. This, thus, facilitates determining whether the labeling may be effective and also aids in understating whether its effectiveness may vary by demographic group.

Keywords: online content labeling; fake news; trust; income level; party affiliation; online time

1. Introduction

Deceptive online content in the form of misinformation, mal-information, and disinformation, which is commonly referred to as "fake news", is a growing problem [1]. Each form of fake news shares a commonality of including inaccurate, incomplete, or misleading information posing as accurate news. Misinformation includes falsehoods that were not necessarily created with the intention to misinform. Disinformation, by contrast, is false information created to intentionally deceive. Mal-information, similarly, is created with the intention of deception, but it uses selective facts while omitting important details in order to form the deception. Fake news is sometimes targeted at specific demographic groups to enhance and target its effect. Tong et al. [2] contend that some current uses actually represent a "weaponization of fake news".

Fake news has been blamed for election interference in multiple countries worldwide [1] and was identified as a driving force in the United Kingdom's departure from the European Union [3]. Allegations of its use as part of foreign influence campaigns [4] abound. The problem is so pronounced that Lee [5] labeled it a "sinister force" that threatens democracy itself.

Due to the magnitude and impact of the problem, a variety of potential solutions have been proposed. These have included restricting internet access [6], content filtering [7], and detecting and removing content [8]. Problematically, little consensus on which standards

should be used for these exists. It is also likely that these techniques would run afoul of many democracies' speech freedom protections. Another possible solution, content labeling [9], has been proposed. Unlike the earlier solutions, labeling does not prevent (or remove) speech. Instead, it provides additional information as context and, in some cases, warns information consumers about particularly problematic content.

While content labeling does not have the speech restriction issues of other techniques, it also does not prevent access to the content. Thus, its effectiveness as a solution to the issues of fake news' spread and impact depend on the impact of labeling on consumers' consumption decisions and post-consumption activities.

This paper aimed to determine what the impact of different label types will be on these behaviors. It presents and analyzes the results of a national survey in the United States on consumers' preferences and their beliefs regarding the effectiveness of different labeling approaches. From these results, this paper drew conclusions regarding the effectiveness of content labeling. The impact and effectiveness were analyzed based on respondents' income levels, political party affiliations and online usage time to ascertain whether labels' impact would be fairly consistent across the population or whether its effectiveness would differ by group. This data will inform the development and prospective deployment of a content labeling system. In addition to suggesting whether it will be effective overall, it also facilitates determining whether targeted training and other roll-out activities would be needed to encourage system adoption and use and to improve its effectiveness among various demographic groups.

This paper continues with a review of prior work in Section 2. Section 3 presents an overview of the survey administration process and demographic details about the respondents. Sections 4–6 present and analyze data on several different types of labeling. Then, Section 7 assesses the implications of the data presented in the previous three sections. Finally, Section 8 presents the key conclusions of the paper and discusses needed areas of future work.

2. Background

This section presents prior work in several areas that provide a foundation upon which the current work draws. First, prior work related to fake news and deceptive online content is presented. Next, content labeling in other areas is discussed. Finally, prior work on online content labeling is reviewed.

2.1. Fake News and Deceptive Content Online

Early "fake news" referred to satirical content that viewers and readers knew was false and comedic [10]. This content, while having similarities to news in formatting, was not designed to fool people (though it occasionally did [11]). In the mid-2010s, though, the use of the term changed and deliberately deceptive—often political—content, designed to manipulate readers grew in prevalence on the Internet [12]. By 2016, deceptive content fake news had become a notable part of the online news content. In the United States, approximately 6% of all news content was fake during the presidential election [13], and it is estimated that each American had, on average, consumed one to three fake articles [14]. In the United Kingdom, fake content was helping to drive the Brexit movement [3,4]. On Twitter, Bovet and Makse [15] calculated that 25% of tweets during this period were "fake or extremely biased news". Cunha et al. [1] showed that this extended well beyond the United States and the United Kingdom to at least 20 other countries.

Beyond the effects on elections, the societal impacts are pronounced. The impact of fake news content on youth is one area of concern. College students surveyed indicated that they expected news on social media to be inaccurate [16]; however, 18–29 year olds generally use social media more frequently than other age groups and trust those sources more than average [16,17].

Fake news causes actual harm. It has been identified as confusing the public [18] and was even involved in starting an armed standoff [19]. It has also been blamed for

circulating damaging health-related information [20]. In response to these issues, a variety of techniques have been developed for mitigating its efficacy and spread. This has led to the development of a number of attempted detection methods with various degrees of success [21]. Tandoc, Lim and Ling [22] even went so far as to develop a classification scheme for fake news content.

While both automated and manual labeling can be performed, a key challenge is how to present the results of this labeling to the prospective reader to maximize its notability and safety impact. Several types of labeling have been proposed [23] as a technique to mitigate deceptive content while not infringing upon content creators' speech rights.

2.2. Content Labeling for Other Applications

Product warning labels, such as those displayed on alcohol and tobacco products, are designed to promote public health. The goal of a warning label in this space is to limit consumption of the product, either by the entire target market or by a vulnerable subgroup. Tobacco product warning labels have a low cost of implementation, but they have been shown to be highly effective at communicating the dangers of tobacco and discouraging young people from taking up the habit [24]. While tobacco companies have shown willingness to implement warning labels of a sort, they have fought strongly against implementation of the most effective forms of warning labels. Current cigarette packaging regulations in the United States date back to 1984 [25] and carry only a text-based Surgeon General's warning [26], using one of four required statements [27].

Graphic health warning labels (containing images such as lungs afflicted with cancer) have been shown to have a great deal more impact than simple text warnings; however, their adoption has been slow. Initially, the FDA proposed "graphic" labels [28], such as those shown in Figure 1, which the tobacco industry contended [29] forced "cigarette makers to display government anti-smoking advocacy more prominently than their own branding". The requirement to use this packaging was not upheld by the courts [29,30], as it was found to violate the First Amendment [31].

Figure 1. The FDA's proposed cigarette labels from 2011 [32].

In 2019, the FDA proposed new labels which were "based on–and within the limits of–both science and the law" [31]. These labels, shown in Figure 2, were planned to launch in June of 2021; however, they have been delayed at least seven times and are currently planned to be required as of 9 April 2023 [33].
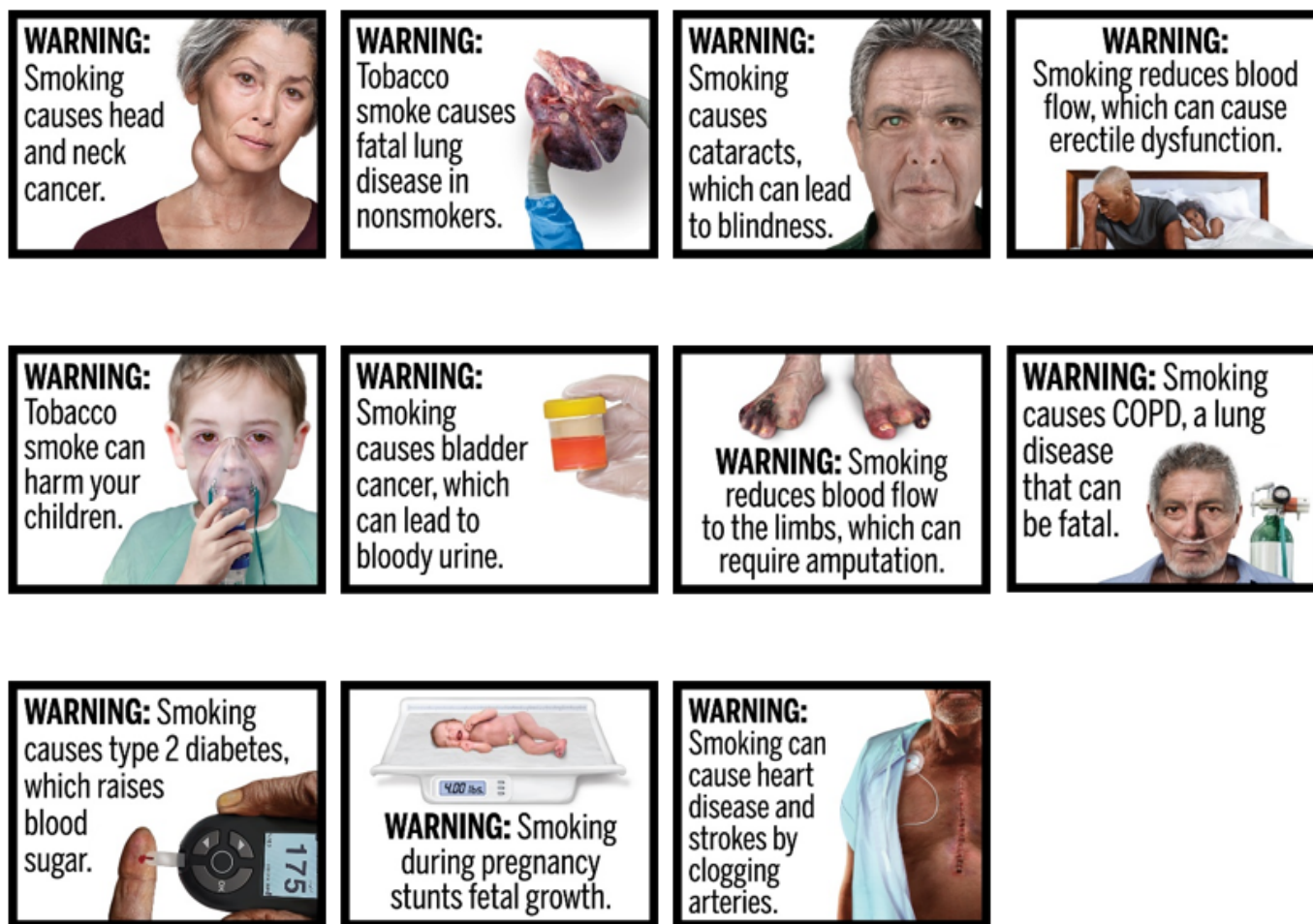
Figure 2. Cigarette labels proposed in 2019 [34].

Websites that present intentional news-style misinformation may be uninterested in self-regulation and, similar to cigarette manufacturers, may be resistant to industry and government labeling requirements that deter users from visiting their sites, reading the misinformation and producing ideological goals [35] or generating advertising revenue [36]. In the case of tobacco, efforts to implement graphic health warning labels have been successful in some nations through government regulation. In New Zealand, the Smoke-Free Environments Regulations of 1999 required that tobacco products include graphic health warnings. This legislation faced legal challenges by the tobacco industry, just as similar legislation has in the United States, but ultimately it was adopted and enjoyed strong public support [37].

Online content labeling may be more legally problematic than cigarette labeling, if required by a government. As previously noted, a federal court prevented the FDA from requiring its 2011 graphic health cigarette warnings due to the presence of free speech concerns [31]. Notably, the cigarette manufacturers were not trying to engage in pure speech but instead speech related to selling their product (which has been held, in some cases, to be less protected [38]). Even with this lower standard of projection than online content would likely enjoy, the labeling requirement was proscribed.

In the United States, thus, labeling may be most effectively implemented by industry self-regulation or collaborative industry–government cooperation. A variety of examples of effective content regulation, developed by or in conjunction with relevant industries, exist. The MPAA movie and V-Chip television ratings [39–41] and explicit lyric warning labels [42] for music are several such examples. In the case of anti-piracy warnings, the US Federal Bureau of Investigation created a voluntary program that allows content creators

to warn consumers about the legal risks of piracy activities [43]. Arguably, in this case, the interests of the content producers were well aligned with the agency's, which is not the case in many other areas where labeling is used.

Deceptive online content is inherently an international challenge. The regulations that may impact the implementation of labeling vary considerably. The freedom of speech guaranteed by the United States Constitution serves as an argument against government required content labeling, which could be taken to infringe upon the speech rights of the publisher [44]. Other countries, though, have regulations with different focuses. The People's Republic of China's Computer Information Network and Internet Security, Protection and Management Regulations of 1997, for example, prohibits internet users from "making falsehoods or distorting the truth, spreading rumors, destroying the order of society". Under these regulations, content labeling may be unnecessary as misinformation should be removed entirely rather than simply to be labeled as such [45].

Other countries have similar laws. Ethiopian law, for example, prohibits certain types of "false accusations", Cote d'Ivoire prohibits "'false information' that could harm the reputation of institutions" and Malawi's laws proscribe the "publication of false statements that may 'cause fear and alarm to the public or do disturb the public peace'" [46]. Bangladesh, while having constitutional protections for free expression, created a law "to control the spread of online misinformation" that has, according to Haque et al. [47], been used to jail journalists and close publications. Indonesia has laws that can jail those convicted of "spreading false information or news that intentionally causes public disorder" for up to a decade [48]. The European Union, on the other hand, has taken actions to "facilitate digital platforms' self-regulation to tackle misinformation and disinformation", which have been met with, at least, partial success [48]. Other countries' laws vary. Yadav et al.'s work [49] demonstrates the diversity of regulation. They identified and analyzed over 100 national laws with conflicting purposes, varying scopes and which met with different levels of success.

2.3. Online Misinformation Detection and Content Labeling

Online content labeling can draw from labeling for television, movies and video games as well as from product labeling. However, it presents several challenges. First, a source for the label's content must be identified. Second, the design of the label itself must be acceptable to content consumers so that they are willing to use it.

Fake news labeling begins with its identification and classification. Identification drives label display, while classification is key to the information that is provided on the label. A variety of forms of identification are possible (see [50,51] for an extended discussion). Manually curated, automatically generated or hybrid manual/automated approaches can be utilized. Zhou and Zafarani [21] describe identification approaches based on style, network analysis and distributing users. Wang [52] demonstrated an automated approach using machine learning with manual annotations. A variety of other automated techniques exist including those that use graph-attention neural networks [53], natural language processing [54], neural stacking [55] and deep neural networks [56]. The social sciences have also contributed through the consideration of emotion cognizance [57] and the use of signal detection approaches [58].

Multiple industry-implemented examples also exist. Twitter's Birdwatch service utilizes manual curation of Twitter posts by users [59]. Wikipedia, similarly, maintains a manually curated list of news sources that is annotated with details regarding their reliability (https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources/Perennial_sources).

Several taxonomies for deceptive online content have been proposed. Tandoc, Lim and Ling [22] developed a classification system that included the categories "satire", "parody", "fabrication", "manipulation", "propaganda" and "advertising". Bakir and McStay [60] had one more category in their system which had the groupings of:

- "False connection (where headlines, visuals or captions do not support the content)";
- "False context (genuine content shared with false contextual information)";

- "Manipulated content (genuine imagery/information manipulated to deceive)";
- "Misleading content (misleading use of information to frame an issue or individual)";
- "Imposter content (genuine sources are impersonated)";
- "Fabricated content (100 per cent false, designed to deceive and harm)";
- "Satire/parody (with potential to fool but no intention to cause harm)".

Perhaps the most important information to label, for consumer protection, is deliberate mal-information and disinformation. One approach to this is through influence analysis. Identifying influential nodes in social networks has been well studied. While the optimization problem is NP-hard, identification can be approximated such as by using Monte Carlo simulation [61,62]. Early work on the spread of misinformation in social media built upon this notion of influential nodes, modeling the problem in terms of what Budak, Agrawal and Abbadi [63] refer to as "competing cascades dissipating in a network". The multi-campaign independent cascade model (MCICM) considers the diffusion of two competing information campaigns, such as the case where one campaign represents the truth, and the second campaign represents misinformation. The core problem, eventual influence limitation (EIL), is to minimize the number of nodes that will adopt the message of the misinformation campaign given a limited budget for the counterinfluence campaign. While the EIL problem is NP-hard to optimize, the authors found initial success with a degree centrality heuristic [63].

More recently, real-time detection of misinformation, disinformation and mal-information spreading on Twitter has been accomplished with some success. Suchia et al. [64] proposed an early algorithm that identifies actively propagated "rumors", defined by the authors as information "many people believe to be true" but that diverges from the facts available on "verified news channels". As an example, if a trending headline observes the fact that the CEO of Corporation X is stepping down and an unverified claim that Corporation X will declare bankruptcy is trending along with it, the supposed bankruptcy story would be the "rumor". The notion is that a rumor is detected as piggybacking alongside one or more legitimate news stories, adding misinformation, mal-information or disinformation to the real narrative.

Trends in the spread of misinformation related to the COVID-19 pandemic [65] have been a particular focus of investigation. One study [66], conducted using data available through the Twitter Streaming API from January to March 2020, reviewed 2,792,513 tweets; 18,168,161 retweets; 456,878 quoted retweets on the subject in over 30 languages with 55.2% in English. Approximately 40.5% of the original tweets contained links to external sources, and these sources were investigated by the researchers. This research showed that only 0.6% of tweets were sharing one of five common myths about the spread, treatment and origins of COVID-19. However, the study also found that just as few, only 0.51%, linked to "reputable health sources", such as the Centers for Disease Control or the World Health Organization, preferring instead to link to popular news media (13%) or other less reliable sources.

The identification of misinformation, mal-information and disinformation is only a portion of the challenge, though. Once potentially harmful content is identified, the next question becomes what to do about it.

The United States' 2016 and 2020 presidential elections were instructive in this regard. Trends regarding the spread of misinformation surrounding the election on Twitter and other social media sites were analyzed [13,67]. Perhaps in response to this, numerous social media platforms, such as Facebook, Twitter, Instagram and YouTube, began labeling social media posts and videos in the run-up to the 2020 United States Presidential Election to combat the spread of misinformation surrounding the candidates, rules about in-person and mail-in voting and the election results themselves [68,69]. A public dataset of over one billion "tweets" (Twitter posts) was released, initially from 1 December 2020 through 22 January 2021 but later expanded to include earlier and later tweets [70]. A separate study [71] of tweets from 1 November 2020 through 8 January 2021 from the Twitter account of former United States President Donald Trump was conducted, covering the days

before the 2020 election up until the date that Donald Trump's account was suspended by Twitter. During this period, Twitter flagged numerous tweets from the former President as containing misinformation, disinformation or mal-information related to the 2020 election. These flags appeared as a content label viewable to Twitter users. In some cases, called "hard intervention" by the authors, the original tweet was unable to be liked or replied to and was concealed behind the label, requiring the user to click a button to view the content after reviewing the warning regarding the misinformation. Other "soft interventions" provided a warning as to the content being misleading but did not prevent the tweet from being interacted with or covered behind the "view" button.

This study [71] found that while hard interventions were successful in preventing the spread of individual tweets, tweets having a soft intervention spread over social media further and longer than tweets that were not flagged at all. This does not necessarily indicate that a soft flag caused the tweet to spread more frequently. Given that the content of a "soft flagged" tweet may have been more interesting to users than a typical "unflagged" tweet, the "soft flagged" tweets may have naturally spread more than those that did not receive a flag, simply due to the quality of the content. Indeed, those tweets that received a soft flag may have spread even further without soft intervention than with it. To know whether a label design has a causal effect on a social media post's propagation, a more controlled study would need to compare the circulation of identical posts, some of which are "soft labeled" and some of which are not, in separate study groups. The study did show that this style of soft intervention used by Twitter was not sufficient to fully deter the spread of Donald Trump's tweets regarding the 2020 election results. Better methods with clear analysis of the causal effects on propagation are still needed to address the issue.

While identification and classification are a key step, the primary focus of this work was on labeling and built upon the labels proposed by Fuhr et al. [23]. This work proposes several media labeling categories including factuality vs. opinion statements; readability and reading level; the current level of virality of the topic's spread online; the usage of emotionally charged words and phrases; the level of public controversy surrounding the topic; the authority and credibility of the source; the degree of field-specific technical knowledge required to interpret the paper; how topical is the document. It also discusses numerous detection methods already available for generating information regarding each category, though future work may improve further upon these methods for specific application to informational labeling techniques.

An addition to this model was proposed by Lespagnol et al. [72] to include "information check-worthiness", while Vincentius et al. [73] also suggested an expansion to include source, article popularity and political bias categories.

Political bias has been a source of significant concern [74], particularly given the use of the term "fake news" as a way of attacking political adversaries' content. Fairbanks et al. [75] created a technique that perhaps offers a partial solution by classifying text as "liberal words", "conservative words" and "fake news words"; however, the fake news category was unreliable.

In prior work, the benefits and approaches to labeling have been discussed [9]. This included presenting a comparison of online media labeling technology to product labeling such as the "nutrition facts" labeling utilized in the United States, the ESRB rating system for video games and FDA warning labels for cigarette packaging. Based on this, it considered multiple paths for developing online media labeling techniques and their potential consequences.

Additionally, the perspectives of university community members [76] and all Americans [77] with regard to content labeling were assessed. University community members' label preferences were also analyzed [76].

The study of the perceptions of university members on online media labeling compared two demographically diverse university communities within the United States. It included questions relating to participants' views regarding multiple label categories (including those proposed by Fuhr et al. [23]). These categories included the article's title,

publisher, publication date, author, sponsors, author's political affiliations, the sponsor's political affiliations, the publisher's political affiliations, writing quality, topic virality, topic controversy level, the reading level and the use of field-specific technical statements. It also gauged respondents' perspectives on and receptiveness to several labeling styles that were previously [9] developed.

This label preference analysis [76] utilized questions similar to the ones for which data were analyzed herein. Each respondent was asked how much they used a particular category of information, how much they believed other people used that category and how much they believe the category ought to be ideally used. This work showed that respondents tended to prefer informational labels over blocking labels, and that they would prefer to have information that they can then use to make decisions from rather than to be told simply whether a media source is trustworthy. This suggests that a "nutrition facts"-style informational label may be preferred by these communities.

3. Survey Administration and Respondent Characteristics

This section describes the methods used for this study. Specifically, it provides details regarding the survey that was used to collect the data that are analyzed in Sections 4–6. First, the survey instrument is discussed. Then, the process of survey administration is reviewed. Finally, the demographic characteristics of the respondents are presented.

3.1. Survey Instrument

A survey instrument was created that was based on the one utilized in [76]. The survey was edited for brevity with a target completion time of 15 min (or less). The new survey combined the questions from the three separate surveys utilized in [76]. Much of the editing involved the removal of questions. Some were removed due to the fact of being redundant among the three surveys (e.g., the demographic questions); others were removed for brevity. Limited editing for consistency and clarity of the newly combined survey was also performed. The final survey was reviewed by the authors and the Qualtrics survey staff before it was put to use. The survey administration plan started with a short-term pilot study, which served to validate the revised instrument before the full-scale survey. No issues were detected during this pilot phase; therefore, these responses were utilized as part of the dataset and applied to the applicable quotas based on Qualtrics' survey administration procedures.

3.2. Survey Administration

A quota-based stratified sampling technique was utilized to collect the data presented and analyzed herein. Data were collected by Qualtrics International Inc., through the use of the survey instrument that was described previously and based on the instrument used in [76].

Qualtrics recruited respondents based on providing a population proportionate representation in terms of the key demographics of income level, age, gender and political affiliation. The survey was administered in October of 2021 and approximately 550 responses were collected. Of these, 500 were part of the population's representative sample; however, all responses that included an answer to the applicable demographic and response question being analyzed were considered herein. Respondents were given an incentive based on complete survey submission. Thus, the vast majority of responses were complete.

3.3. Respondent Demographic Details

Respondent demographics are presented in this section. Due to the quota-based stratified technique used (which was described in the previous section), the respondents were well distributed across the key demographics. In terms of gender, approximately 49% of respondents were male and 51% were female. A small number of respondents (less than 1%) indicated a non-binary gender; however, due to the limited number of responses and small sample size, the perspectives of non-binary respondents could not be analyzed.

In terms of the age demographic, approximately 11% of respondents were in the 18–24, 25–29 and 30–34 age groups (each). Approximately 10% of respondents were aged 35–39, and 9% of respondents were aged 40–44. The 45–49 age range composed 7% of respondents, and 6% of respondents were aged 50–54. The 55–59 age group included 14% of respondents, and 12% of respondents were aged 60–64. Finally, 11% of respondents were in the 65 and older age group.

Respondents also had various levels of educational attainment. High school graduates (without college attendance) constituted 26% of respondents. An additional 23% had completed some college but not a degree. Associate degree graduates (without higher degrees) constituted 12% of respondents, and bachelor's degree graduates (without higher degrees) constituted 22% of respondents. Approximately 13% of respondents held graduate degrees with 10% holding master's degrees and 3% holding doctorates. Finally, 5% of respondents reported that they had not completed high school.

4. Informational Labels

This section considers the impact of informational labels on Americans' news content consumption behaviors. It presents and discusses the results from several survey questions relevant to informational labels. Respondents were presented with multiple potential labels and asked five questions about each regarding whether they would find it helpful or annoying, whether they and others would use the label and whether they found it useful for judging content's trustworthiness.

4.1. Informational Label with Article Summary

The first of these labels is presented in Figure 3. This informational label presents the title of the article, a brief text overview and an image from the article. It then presents ten pieces of information about the article that could be used by prospective readers to assess it and to determine whether they will choose to read it or not. Figures A1–A15, in Appendix A, present respondents' views about the utility of this type of a label. Error bars are included on these and all data figures to depict the standard error level for each type of response.
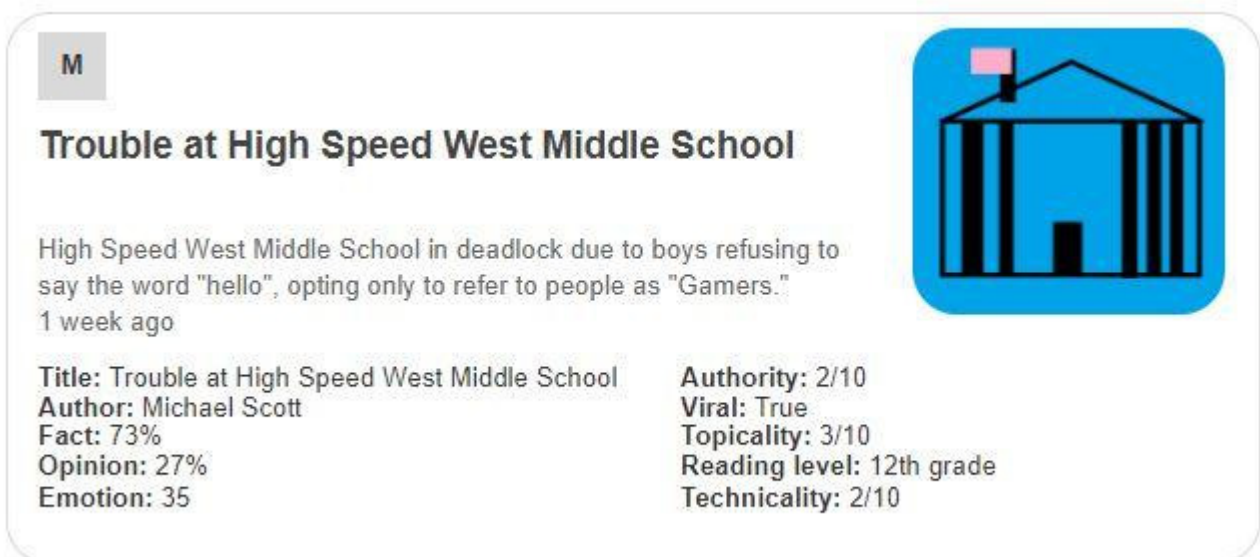
Figure 3. Informational label with article summary [76].

Respondents were first asked whether they would find this label helpful. Responses to this question, analyzed by income level, political party affiliation and internet usage levels, are presented in Figures A1–A3, respectively. Each figure shows the responses including (left) and excluding (right) "unsure" responses.

Figure A1 shows that most respondents, who had an opinion, found the labels to be helpful. The level of uncertainty fluctuated notably by income level, with the greatest numbers of individuals reporting being unsure at the lowest, highest and USD 75,000–99,999 income levels. When considering only "yes" and "no" responses, a trend of a slight decline in perceived helpfulness with increasing income level is shown in Figure A1 (right). Overall, though, there was not a tremendous difference between perceived helpfulness across different income levels. It is notable that there was an increase in perceived helpfulness between the USD 50,000 and USD 75,000 income levels. This trend was also present in the data for some other questions related to other labels.

Figure A2 considers respondents' perceptions of helpfulness by political party affiliation. Democrats were more certain about their responses than Republicans and independents/other party members and also found the label to be more helpful. Notably, the difference between Democrats and Republicans was smaller than the difference between Democrats and independents/other party members for both uncertainty and helpfulness. Similar with the age data, there was no tremendous variation in usefulness perception across the different political affiliations.

Next, perceived helpfulness was analyzed in terms of internet usage level. Notably, the level of uncertainty did not correlate with either greater or lower levels of internet usage (as the greatest uncertainty levels are reported in the highest and lowest usage level categories). There was a slight trend with respondents indicating greater helpfulness of the label with higher levels of internet usage.

Respondents were also asked about the annoyingness of the label. Figures 5, A4 and A6 present the data related to annoyingness.

Figure A4 presents the levels of annoyingness of the label reported by income levels. Notably, several levels (comparing Figures A1 and A4) had reduced uncertainty reported as compared to helpfulness. Additionally, while most respondents did not find the labels annoying (i.e., a no answer), this was only slightly above half. There was also a trend of declining level of annoyance from the USD 25,000–49,999 to the USD 100,000–124,999 income levels, though the USD 125,000 income level had the highest level of annoyance reported, overall, and the USD 24,999 and less group had more individuals reporting the label being annoying then the next group up.

Next, annoyingness was analyzed in terms of political party affiliation. While the levels of uncertainty reported mirrored those reported for the helpfulness question for this label, the three affiliation groups had minimal differences between the number of respondents reporting the label as being annoying.

The responses for annoyingness by internet usage level also mirrored those for helpfulness. This was the case both for the level of uncertainty reported and for the number of individuals indicating that the labels were annoying (and not annoying).

Respondents were next asked whether they would be willing to review the label presented in Figure 3. Data related to respondents' willingness are presented in Figures A7–A9. In Figure A7, there was a slight trend in reduced uncertainty with higher income levels and a similar trend in being less willing to review, which also correlated with higher income levels. In both cases, the USD 75,000 and USD 125,000 and higher income levels bucked the trend, reporting greater uncertainty than the next lower income level and more willingness to review.

Willingness to review was next assessed by political affiliation. Notably, Republicans showed a significantly lower level of willingness to review than Democrats and less than that of independent/other party members. This is notable, as more Republican respondents had indicated the label to be helpful and fewer had indicated it being annoying compared to independents/other party members. Thus, it appeared that Republicans' wiliness to use the labeling was notably influenced by factors other than the label itself. Despite the differences between helpfulness, annoyingness and willingness to use levels, the uncertainty levels by affiliation mirrored the helpfulness and annoyingness ones.

The data related to willingness to use by internet use level are also interestingly different from the helpfulness and annoyingness data. While a similar pattern of uncertainty response was still present (albeit less pronounced and fluctuating), increased willingness to use the label had a very strong correlation with increased internet usage levels.

Respondents were next asked about others' willingness to review. These data are shown in Figure A10. While higher levels of uncertainty were reported by those with higher income levels, no clear trend in the perception of others' willingness to use the label was present. Over half of all groups indicated believing that others would use the labels, and two groups had over 70% who reported being willing. It is notable that, in this data, there was a sharp decline at the USD 50,000 income level followed by an increase at the USD 75,000 income level.

Figure A11 presents data regarding the willingness of others to use the label based on political affiliation. The same patterns in individuals' own willingness to use the label were present in the willingness of others to use data, albeit they were somewhat less pronounced. Again, Democrats had the highest belief in the willingness of others to use and the lowest uncertainty. Independents/other party affiliates had the highest uncertainty and Republications reported the lowest willingness of others to use the label. This may be indicative of individuals associating with those that share their political views but being moderated by the fact that they also associate with those that do not share their political views.

The data for willingness of others to use the label correlated with internet usage level are presented in Figure A12. There was no consistent pattern in this area. This is likely indicative of individuals associating with others with all levels of internet usage (as opposed to principally associating with those with similar internet usage levels as themselves). This is notably different from the political party affiliation responses discussed immediately above (shown in Figure A11) as well as the trend of increasing willingness to review by internet usage (shown in Figure A9).

The final question regarding the first label indicates its level of usefulness in judging trustworthiness. These data are presented in Figures A13–A15. Figure A13 shows the relationship between perceived trustworthiness judgement usefulness and income level. It shows that uncertainty decreased with income level (with the highest income level not following this trend), and perceived usefulness decreased with income level (with the highest income group, again, not following this trend). There was, again, an increase in perceived usefulness between the USD 50,000 and USD 75,000 income levels.

Trustworthiness judgement usefulness was next assessed by political affiliation. As with the willingness to use data, Democrats reported the lowest uncertainty and highest perceived judgement usefulness. Independents/other party members reported the highest uncertainty and Republicans indicated the lowest level of perceived utility for assessing trustworthiness. Notably, over half of all three political parties indicated that they believed this label would be useful for assessing trustworthiness.

Perceived trustworthiness assessment utility was also analyzed by internet usage level. Mirroring the data presented in Figure A9, there was a general trend of decreased uncertainty and increased perception of usefulness associated with higher internet usage levels. Notably, the highest usage group was lower, in both categories, than the usage level below it; however, it still reported lower uncertainty and higher perceived utility than the lowest internet usage level.

4.2. Informational Label without Article Summary

A second label was assessed, which is shown in Figure 4. This label presents the article's characteristic data that the first label (Figure 3) provided. However, it lacks the article overview and article image. Figures A16–A30, in Appendix A, present the relevant respondent perception data regarding this label design.
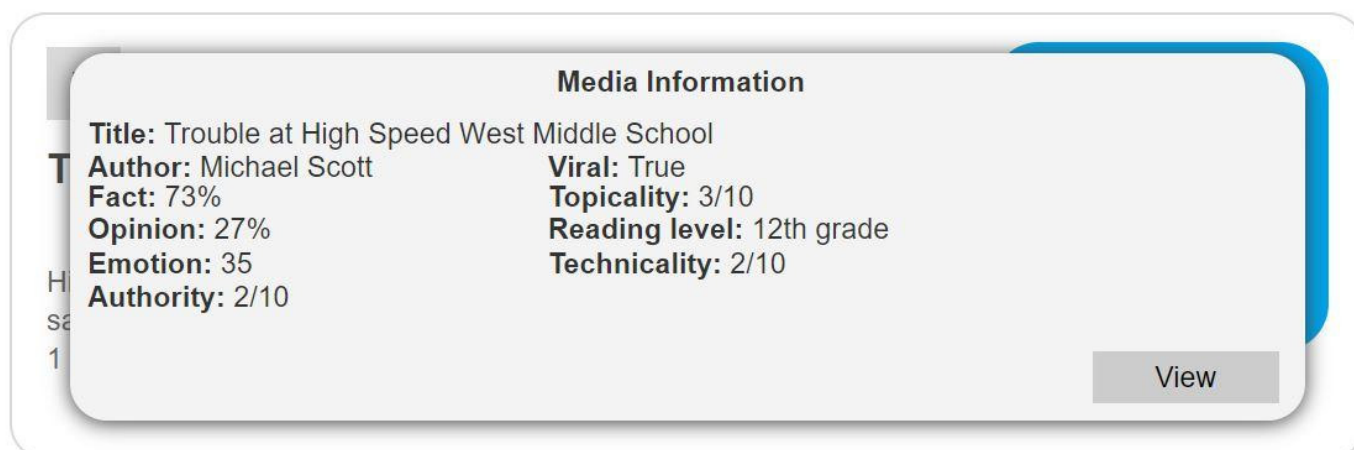
Figure 4. Informational label without article summary [76].

Figures A16–A19 present data regarding the perceived helpfulness of the second informational label. Figure A16 characterizes the impact of income level on respondents' perception of the helpfulness of the label. There was a partial trend of reduced uncertainty with increased income level at the lowest three income levels; however, no clear trend was present after this. The perception of helpfulness showed two downward trends. One existed at the three lowest income levels. It reset at USD 75,000 in annual income, and a second downward trend started at this point. There was a similar increase at the USD 75,000 annual income level for information label 1 (Figure A1), though that increase was less pronounced than what was seen for information label 2.

Figure A17 shows the impact of political affiliation on respondents' helpfulness perceptions. As is common with many responses, Democrats had the lowest level of uncertainty and believed the label to be helpful the most often. Republicans had the second lowest uncertainty level and the second highest helpfulness perception for label 2.

Figure A18 shows the impact of internet usage on respondents' perception of the second informational label's helpfulness. Two slight trends were present. A slight reduction in respondents' levels of uncertainty and an increase in respondents' helpfulness perceptions were present with increased online usage time.

Next, Figures A19–A21 present data regarding the perceived annoyingness of the second label. Figure A19 shows a correlation between increased income levels and decreased uncertainty regarding annoyingness. There was also a slight trend of increased annoyingness being reported by individuals with higher income levels.

Figure A20 shows the perception of annoyingness by political affiliation. The pattern shown in several previous questions of Democrats having the lowest uncertainty and greatest favorability toward the label was also present here. However, it was much less pronounced than with several of the other question responses. Democrats and Republicans were nearly tied in terms of finding the label annoying, while notably more independents/other party members found the label annoying than either Democrats or Republicans.

Figure A21 shows the annoyingness perception of label 2 by internet usage level. There was no notable pattern of association between respondents' level of internet usage and annoyingness perception, either in terms of the level of uncertainty reported or in terms of finding the label annoying or not.

Figures A22–A24 characterize respondents' indication of their own willingness to use informational label 2. Figure A22 characterizes respondents' willingness to use the label in terms of income level. A general trend of reduced uncertainty with increased income level was present. In addition, a trend of reduced willingness with higher income level was present up to the USD 75,000 point. As with the question regarding helpfulness, there was

a similar increase in willingness to review from the USD 50,000 to the USD 75,000 income levels for both labels, though it was, again, more pronounced for the second label.

Figure A23 presents respondents' willingness to use the second informational label in terms of party affiliation. While Democrats indicated the highest willingness to use the label (followed by independents/other party members, with just slightly more willingness than Republicans), it is notable that Republicans had the lowest uncertainty for this label. The difference between the uncertainty levels was limited, albeit.

Figure A24 presents respondents' willingness to use the label in terms of online usage levels. There was a slight trend of increasing willingness to use the label with increased online usage. There was also a trend of reducing uncertainty with increased online usage, which was present at all but the highest usage level.

Figures A25–A27 present respondents' perceptions of others' willingness to review the label. Figure A25 presents this in terms of income level. There was no clear trend present in terms of a correlation between income level and willingness to review or uncertainty about this question. There was a notable increase between the USD 50,000 and USD 75,000 income levels, after a decline leading up to the USD 50,000 income level.

Figure A26 considers respondents' perceptions of others' willingness to use the second label by political party. As in many other cases, Democrats had the highest willingness to use the label (followed by independents/other party members) and the lowest uncertainty (followed by Republicans).

Figure A27 considers respondents' perceptions of others' willingness to review the label in terms of internet usage level. A trend of declining uncertainty with increasing usage level was shown (however, the highest usage level bucked this trend). There was no notable level of difference between willingness to use responses and usage level, though.

Finally, for this label, respondents were asked whether they would find the label useful for judging article trustworthiness. These data are presented in Figures A28–A30. Figure A28 presents respondents' perceptions of the utility of label 2 for judging trustworthiness by income level. While a trend existed regarding reduced uncertainty with increasing income level (which the highest income level, again, bucked), no clear trend was present in the actual willingness responses by income level. Again, there was an increase in perceived usefulness between the USD 50,000 and USD 75,000 income levels.

Figure A29 considers willingness to use the second label by political affiliation. Again, Democrats perceived the label the most positively, with the highest willingness to use percentage and lowest uncertainty. Republicans had the second lowest uncertainty and second highest willingness to use, followed (with limited difference) by independents/other party affiliates. This was, of course, somewhat different from the trend shown in previous political party-related data.

Figure A30 presents data regarding the judging usefulness of label 2 in terms of internet usage levels. This data show a trend of decreasing uncertainty and increasing perceived usefulness with increasing internet usage level; however, the level of difference between the usage level groups was limited.

4.3. Informational Label including Article Summary and Component Descriptions

Respondent perceptions were also solicited regarding a third type of informational label. This label presents all of the information from labels 1 and 2 (including the description and graphic from Figure 3). It also includes a brief description of what each piece of article metric data means to aid in the interpretation of the data. This third label is presented in Figure 5. Data related to this label are presented in Figures A31–A35, in Appendix A.

Figure 5. Informational label with article summary and component score descriptions [76].

Data regarding the helpfulness of the third informational label is presented in Figures A31–A33. Figure A31 shows the number of respondents indicating perceived helpfulness of the label by income level. While there was no overarching trend in these data, the uncertainty and perceived helpfulness decreased with increased income across the three lowest levels of the data. The perceived helpfulness then increased for the next two levels, though a similar trend with uncertainty was not present. The highest income level did not participate in either of these trends. As with the previous two labels, there was an increase in perceived helpfulness between the USD 50,000 and USD 75,000 income levels, though it was less pronounced than for the second label (Figure A16). In this instance, the trend differed, as there was a continued increase in perceived helpfulness for the third label at the USD 100,000 income level; however, it showed a similar pattern between the USD 50,000–75,000 income levels as was present in previous label styles.

Figure A32 shows the perceived helpfulness of the third label by political affiliation. As was true in many cases, Democrats reported the lowest uncertainty and highest perceived helpfulness. This was followed by the Republicans (for both uncertainty and helpfulness) and then independents/other party members.

Data presenting the helpfulness of label 3 by internet usage level are shown in Figure A33. These data show no clear trend related to uncertainty; however, a trend of increased perception of usefulness with higher internet usage was present at the lower three levels of internet usage. This trend was bucked by the highest internet usage level.

Figures A34–A36 characterize the perceptions of the annoyingness of label 3. Figure A34 presents the perceptions of annoyingness in terms of income level. There was a general trend of declining uncertainty with increasing income level, with some deviations. There was also a trend of increasing perceived annoyingness up to the USD 50,000–74,999 income group followed by declining perceived annoyingness. The highest income level bucked this trend, with annoyingness reported at a similar level to the USD 50,000–74,999 group. Once again, between the USD 50,000 and USD 75,000 income levels, a trend was present. In

the case of annoyingness, this trend presented itself as a decrease, rather than the increases generally found for helpfulness and usefulness.

Figure A35 characterizes the annoyingness of the label by political affiliation. Once again, Democrats perceived the label the most positively, having the lowest uncertainty and lowest annoyance reported. Republicans and independent/other party members had similar higher uncertainty and annoyingness levels.

Figure A36 shows perception of annoyingness of the third informational label in terms of internet usage level. While there was a very slight trend regarding reduced uncertainty with increased internet usage, no trend was present in the internet-usage-associated annoyingness data itself.

Figures A37–A39 present data regarding respondents' willingness to use informational label 3. Figure A37 shows the willingness level correlated with income level. There were no notable trends present in either the uncertainty level or income level associated with the willingness level data itself. It is notable, though, that respondents' willingness to review, once again, increased from the USD 50,000 to the USD 75,000 income levels.

Figure A38 presents the willingness to review data correlated with party affiliation. As was the case with some frequency, Democrats had the lowest uncertainty and highest positive perception of the label. Republicans had slightly less uncertainty than independents/other party members and indicated willingness to use with slightly less frequency.

Figure A39 presents the willingness to use data related to label 3 correlated with internet usage levels. A trend of declining uncertainty was present across the lowest three levels, and a trend of increased willingness to use was present (with slight a deviation at the 3–5 h level) across all four levels.

Figures A40–A42 present data regarding respondents' perceptions of others' willingness to review the third informational label. Figure A40 presents this data by income level. As shown in Figure A40, there are no notable trends in the correlation between income level and others' willingness to review the third informational label. Perceived willingness of others to review increases when between the USD 50,000 and USD 75,000 income levels, as was observed for the first (Figure A7) and second (Figure A25) labels.

Figure A41 shows respondents' perceptions of others' willingness to review in terms of political affiliation. Notably, as has been common, Democrats have the lowest uncertainty and highest willingness to use the label. This is followed by Republicans and then independents/other party members.

Figure A42 presents respondents' perceptions of others' willingness to review the third informational label by internet usage level. There are slight trends of reducing uncertainty and increasing willingness to review the label (with a deviation at the 3–5 h level) with greater internet usage time.

Focus now turns to the utility of the third informational label for assessing the trustworthiness of an article. Figures A43–A45 present data related to this. Figure A45, in particular, presents data regarding the association between income level and perceived usefulness for trustworthiness assessment. There is a slight trend of decreasing uncertainty and increasing perception of trustworthiness assessment utility with increased income level, for this label. Notably, the highest income bracket does not participate in either of these trends. Once again, there is a decrease in perceived usefulness from the USD 25,000 to USD 50,000 income levels followed by an increase at the USD 75,000 income level.

Figure A44 presents data regarding trustworthiness determination utility associated with party affiliation. As was frequently the case, Democrats had the lowest uncertainty level and highest trustworthiness determination utility level. For uncertainty, Republicans had only slightly more uncertainty than Democrats did for this label. The difference was more pronounced for the difference in perceived trustworthiness determination, though the Republicans also found there to be more utility from this label than independents/other party members.

Finally, Figure A45 presents the trustworthiness judging utility data correlated with internet usage levels. There was a clear trend in this data of reducing uncertainty with in-

creasing internet usage levels. There was also a trend of slightly increasing trustworthiness judging utility with increasing internet usage levels.

5. Warning Labels

Focus now turns to warning labels. This section presents and discusses the results from several survey questions relevant to warning labels (paralleling the discussion of information labels in Section 4). Respondents were presented with three warning labels to review and then indicated their perspectives.

5.1. Warning Label including Description

Figure 6 presents the first warning label. This label pops up in front of the content, has a warning icon and explains the rationale for the warning.
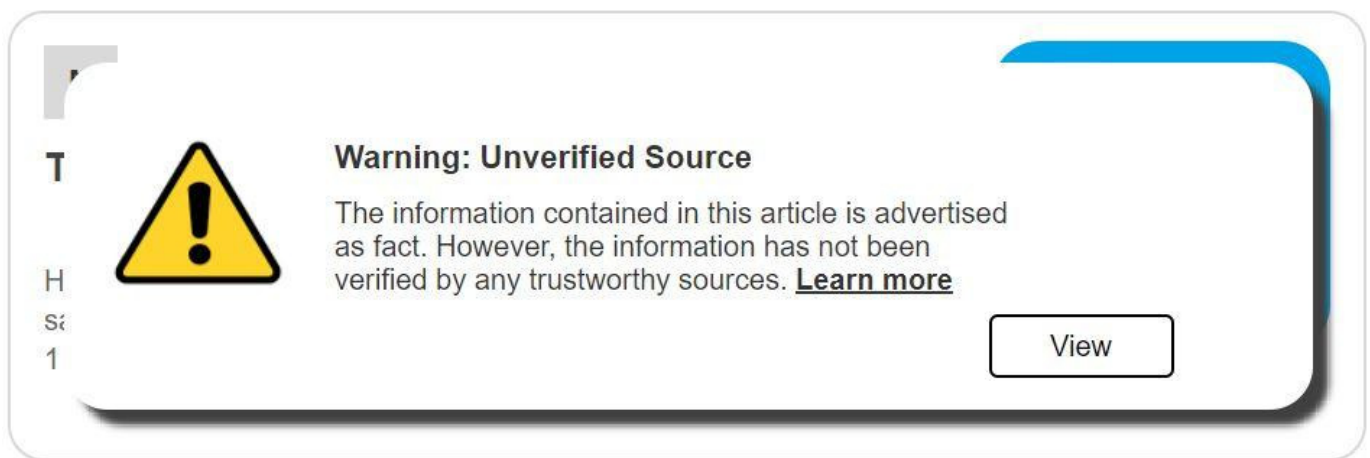
Figure 6. Warning label with description [76].

Figures 7–9 present data related to respondents' perception of the helpfulness of warning label 1 (Figure 6). Figure 7 presents perceptions of helpfulness in terms of income level. Notably, there was comparatively small (to the informational labels) uncertainty, which had a slight trend of reduction with increasing income level. There was also a high level of individuals reporting helpfulness (above 70% for all income levels), with a slight trend of reduction in helpfulness perception with increasing income level.
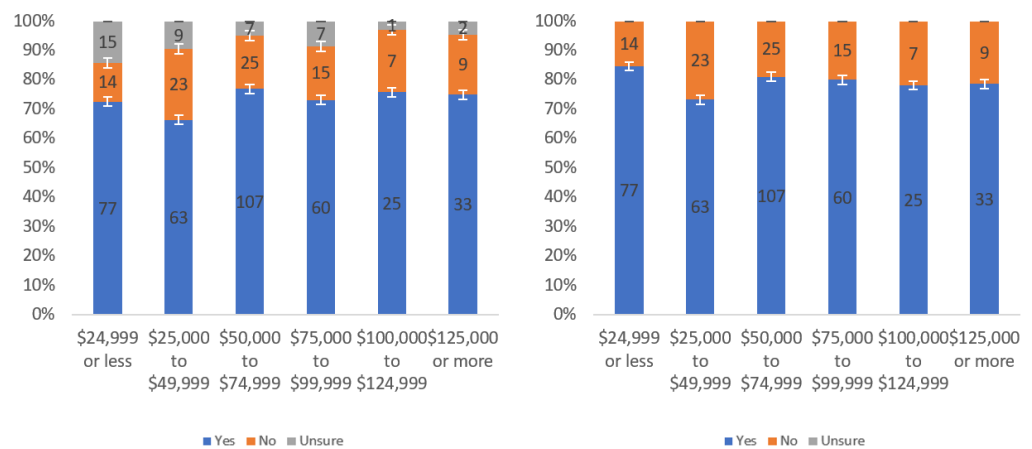

Figure 7. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.
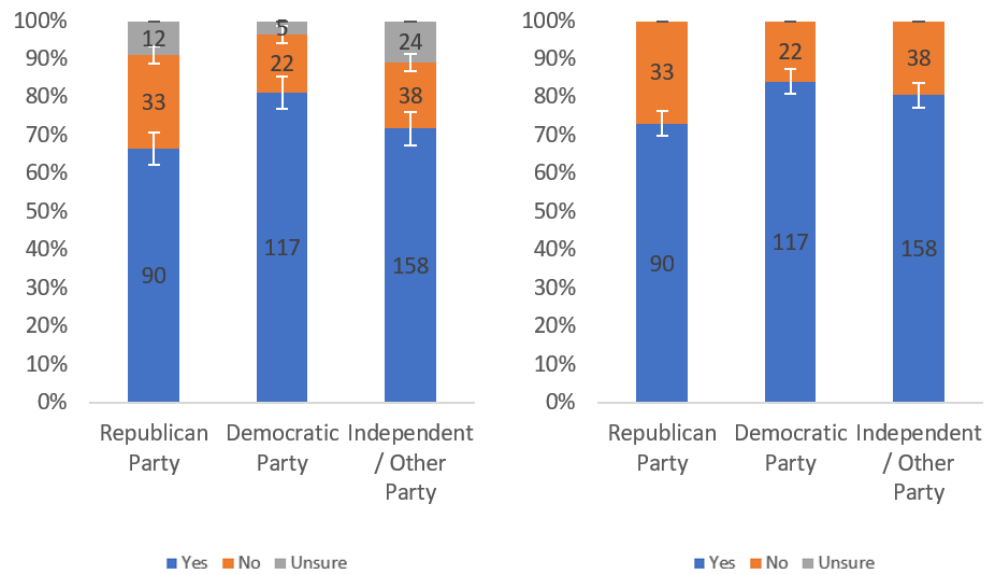
Figure 8. Responses regarding label helpfulness by political affiliation including (**left**) and excluding (**right**) unsure responses.
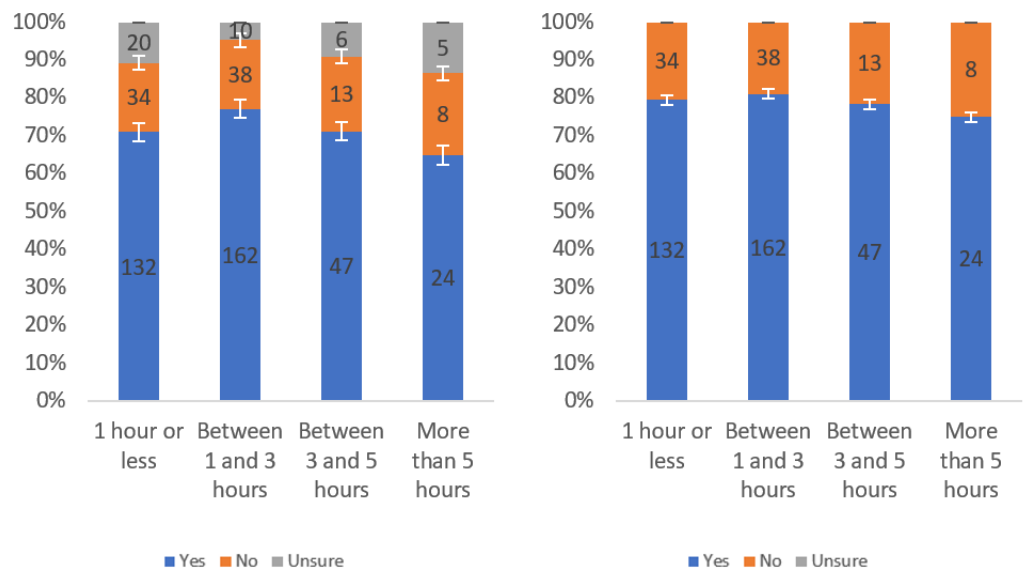
Figure 9. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure 8 presents the helpfulness perceptions by respondents' political affiliations. As with the informational labels, Democrats had the best perceptions of the label with the lowest uncertainty and highest levels of finding it helpful. Republicans had the second lowest uncertainty, while independents/other party members had the second highest level of perception of helpfulness.

Figure 9 considers helpfulness of the first warning label by internet usage level. There was a slight downward trend at the three highest levels of usage for perception of helpfulness. This was coupled with a growing level of reported uncertainty.

Respondents were asked about the annoyingness of the first warning label, and these results are presented in Figures 10–12. Figure 10 presents the annoyingness perception based on respondents' income levels. While there was a trend toward a slight decrease in uncertainty with increased income levels, no trend existed in the annoyingness perceptions themselves.
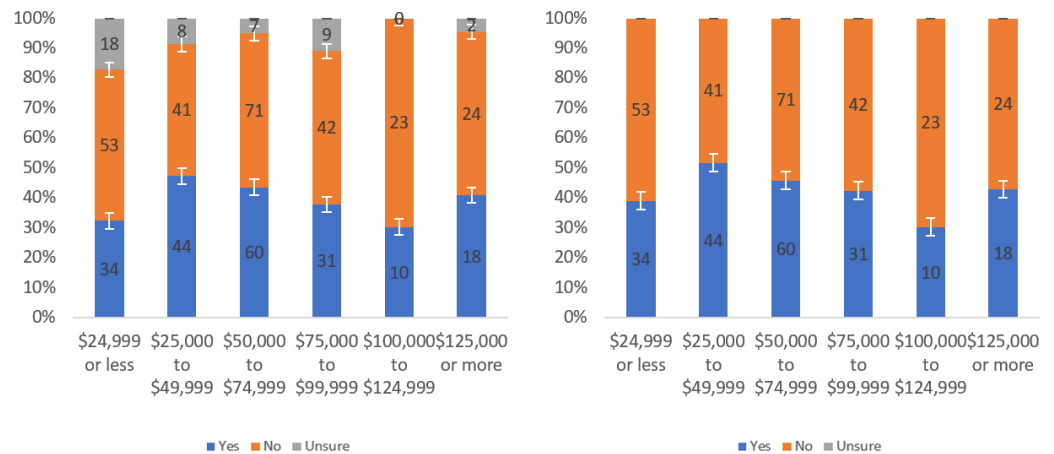
Figure 10. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.
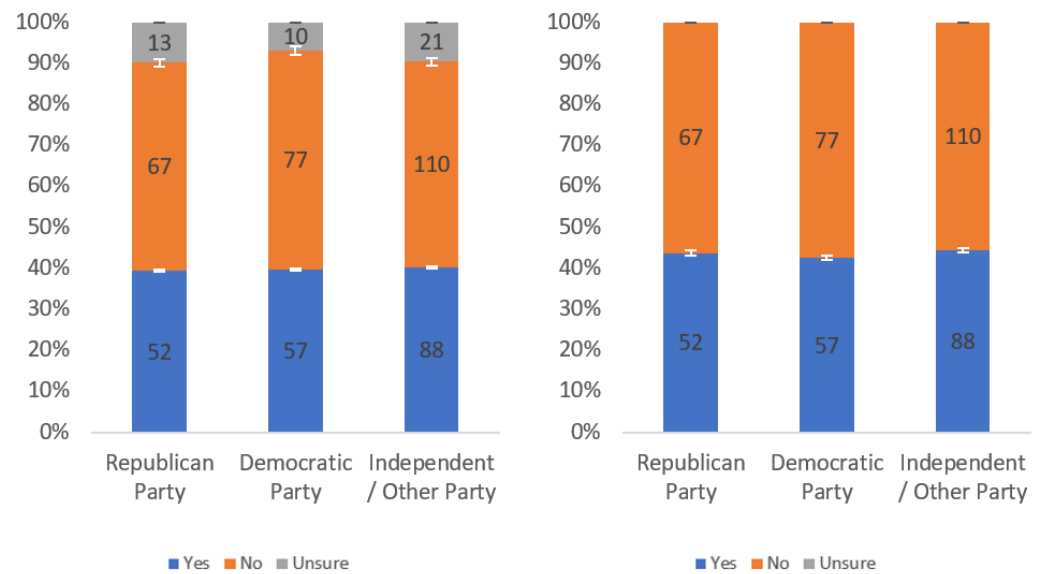
Figure 11. Responses regarding label annoyingness by political affiliation including (**left**) and excluding (**right**) unsure responses.

Figure 11 presents the annoyingness perceptions in terms of respondents' political affiliations. Notably there was very little deviation between the political affiliations' responses. The three groups had nearly identical annoyingness perception levels, and the difference among the uncertainty levels was not practically significant.

There were also no notable trends present in the annoyingness data presented in terms of internet usage level in Figure 12. There was minor and practically insignificant variation present.

Focus now turns to respondents' responses regarding their willingness to use the first warning label. Data related to this are presented in Figures 13–15. Figure 15 shows willingness to use in terms of respondents' income levels. There was a trend present regarding reduced uncertainty, generally, with increased income levels. There was also a slight reduction in willingness to use with increasing income levels; however, the reduction was not practically significant.
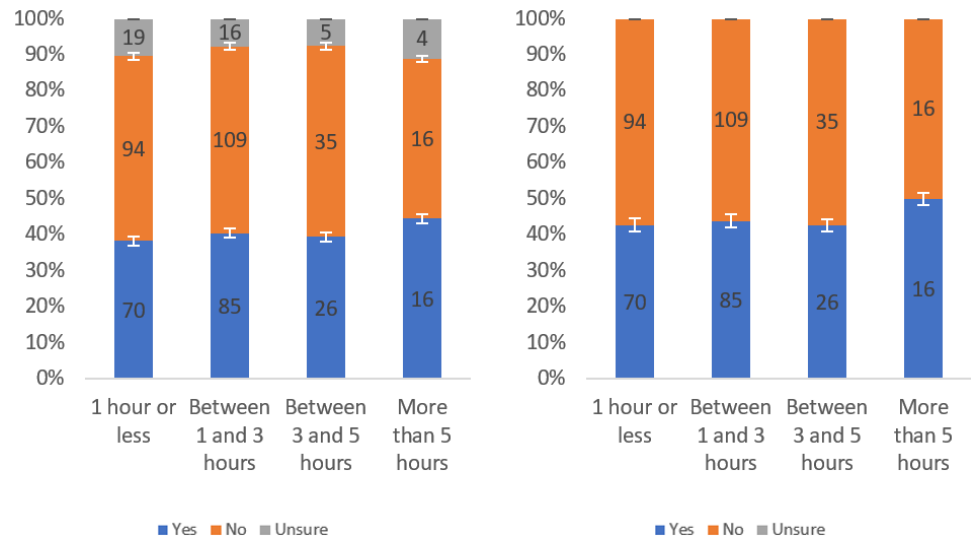
Figure 12. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.
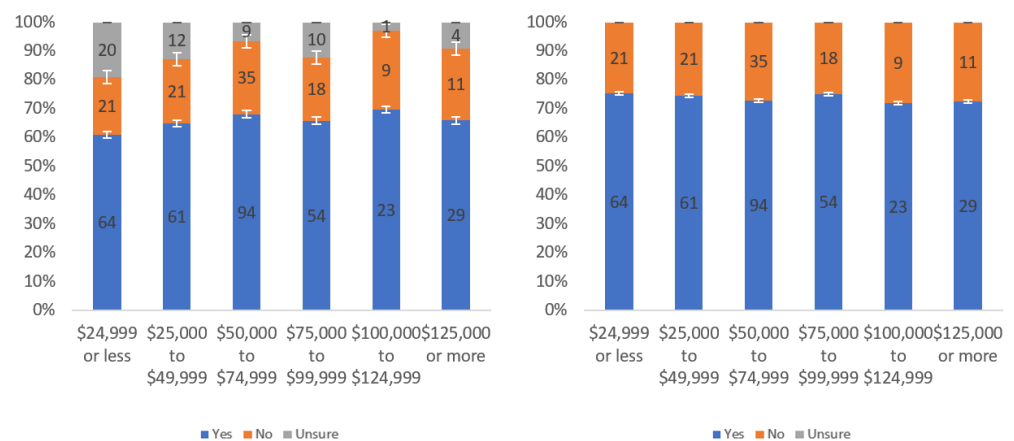
Figure 13. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
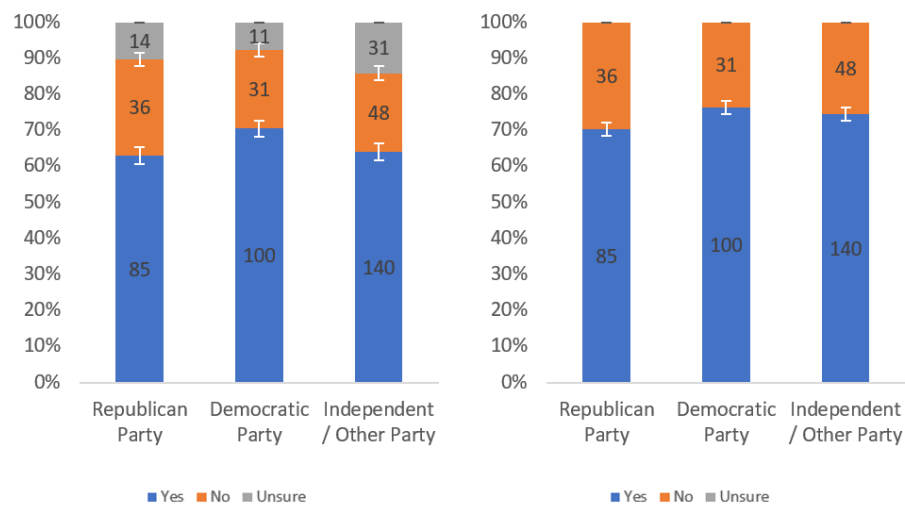
Figure 14. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
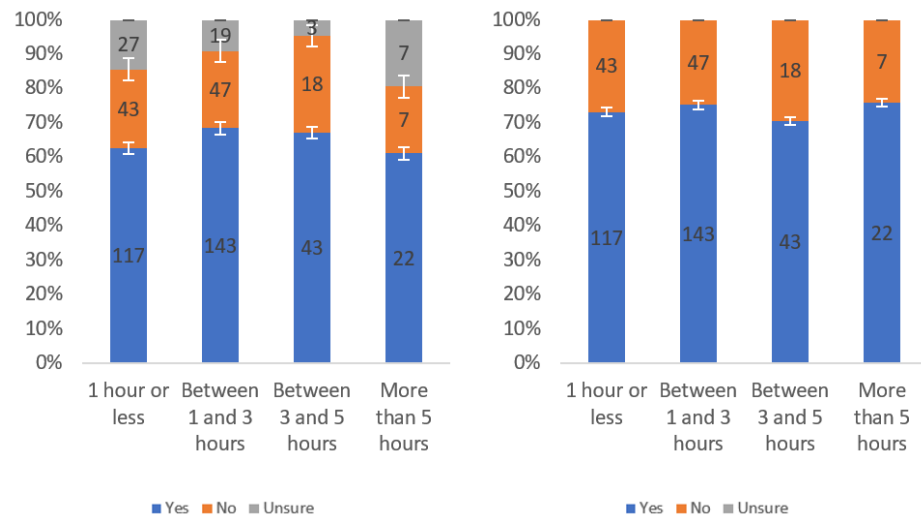
Figure 15. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

The willingness to use data is now considered in terms of respondents' political affiliations. Again, the variations among the different groups were relatively small. The Democrats had the most favorable views of the label with the lowest uncertainty and highest levels of willing to use responses. Republicans had the second lowest uncertainty level, while independents/other party members had the second highest level willingness to use. Notably, the willingness to use (among those not indicating uncertainty) was at approximately 70% or above for all three political affiliation groups.

Figure 15 presents the willingness to use data in terms of online usage time. There was neither a clear trend in the willingness to use levels nor differences among the groups that were practically significant. There was a trend of reducing uncertainty levels with increased online usage among the three lowest levels; however, the highest online usage level also had the highest uncertainty, bucking this trend.

Figures 16–18 characterize respondents' perceptions of others willingness to use the first warning label. Figure 16 presents the willingness to use data in terms of income levels. No clear trend existed with regard to willingness to use the label and income level. A trend of reduced uncertainty with increased income level was present at the three highest levels. It is notable that for the first warning label, this was the only question where a notable increase was observed between the USD 50,000 and USD 75,000 income levels.
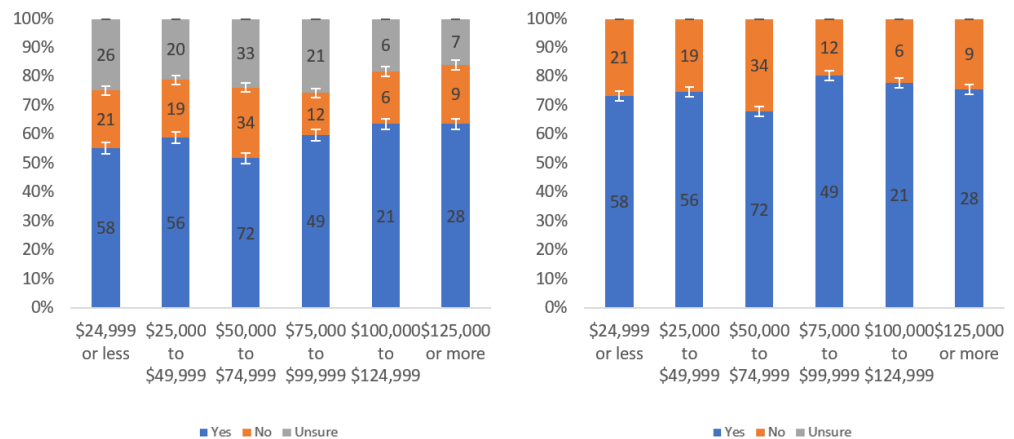
Figure 16. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
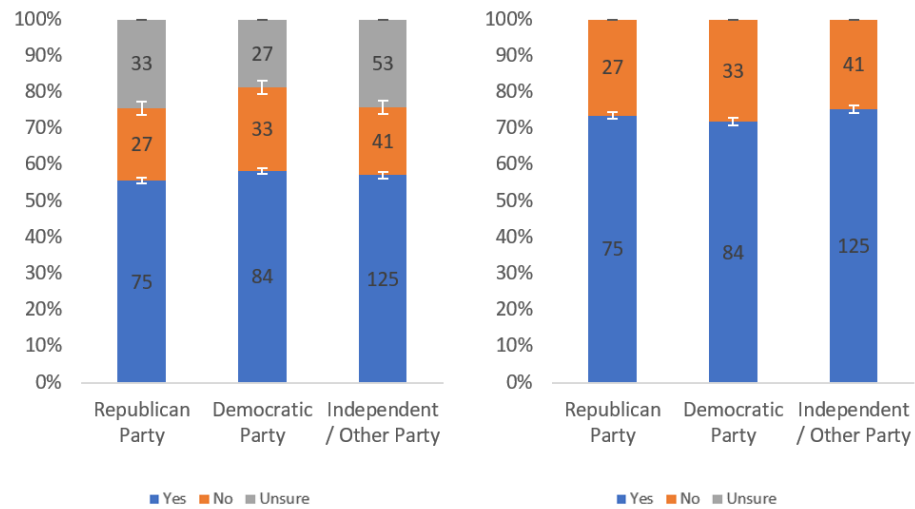
Figure 17. Responses regarding others' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
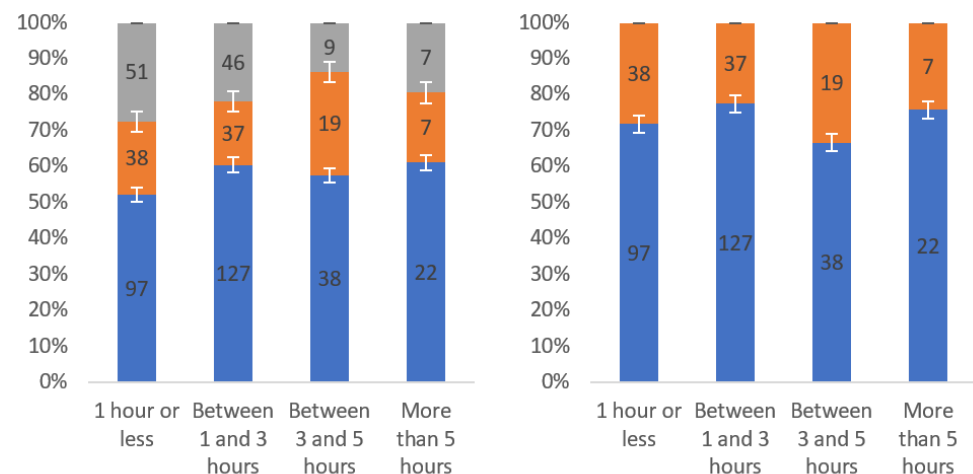
Figure 18. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure 17 presents respondents' perception of others' willingness to use the first warning label in terms of political affiliation. Notably, while Democrats had the lowest uncertainty, they also had the lowest willingness level, with independents/other party members and Republicans reporting higher levels of belief in others being willing to use the label.

Finally, Figure 18 presents respondents' perception of others' willingness to use the first warning label in terms of respondents' online usage levels. There was reduced uncertainty with increased online usage at all but the top level of usage. No trend, however, was present in the willingness data correlated with online usage levels.

Figures 19–21 present data regarding the utility of the label for judging the trustworthiness of the article. Figure 21 presents this in terms of income level. In this figure, there is a clear trend of reducing uncertainty with increased income levels in this data. However, no trends were obvious in the judging utility responses themselves.
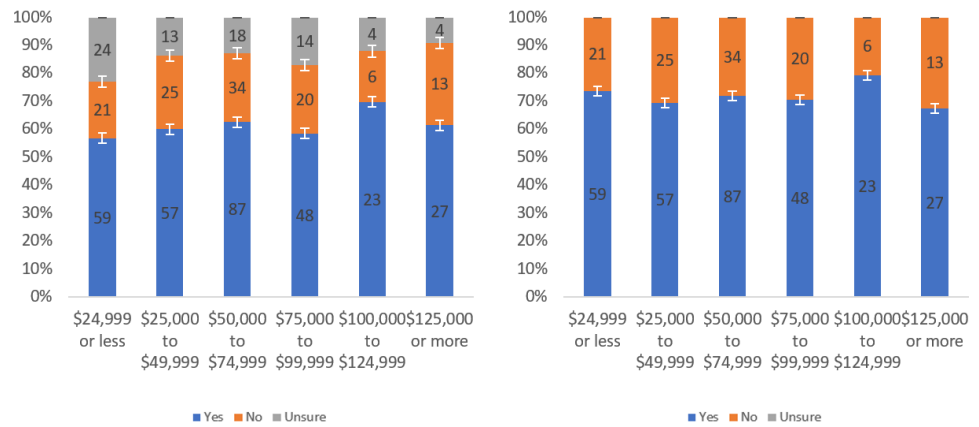
Figure 19. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.
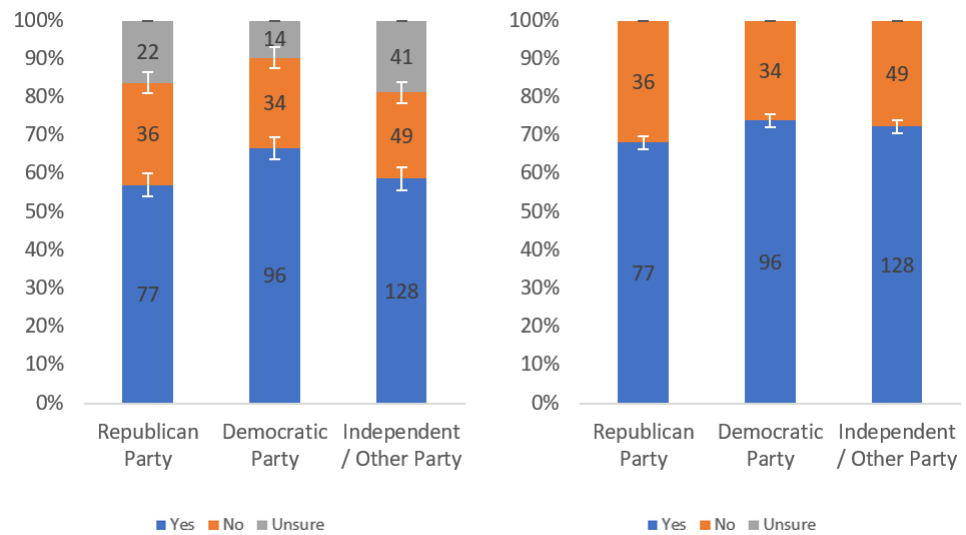
Figure 20. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
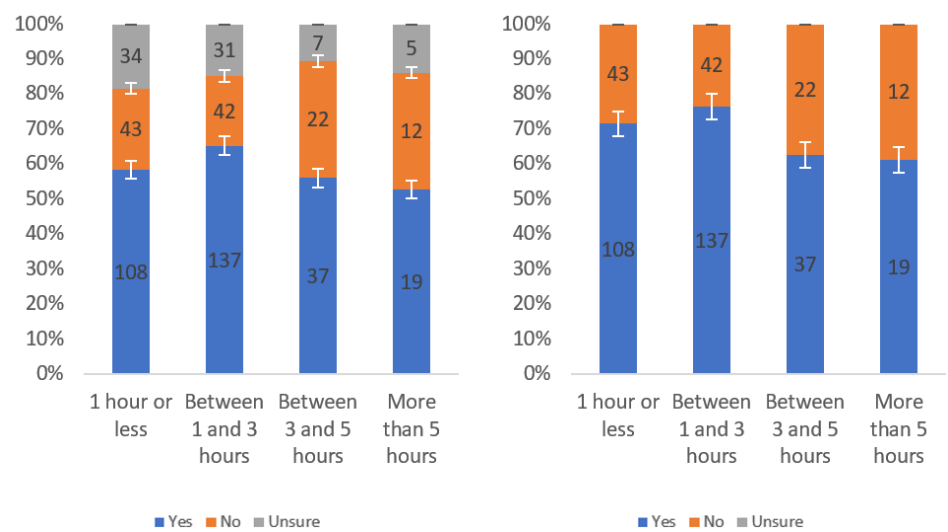
Figure 21. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure 20 presents the trustworthiness judging utility data perceptions in terms of respondents' political affiliation. While there were larger differences in uncertainty level, the difference between the levels of judging utility perceptions, themselves, were limited. As in many cases, Democrat respondents indicated the lowest uncertainty and highest perceived judging utility. The difference in perceived utility levels were quite low, though, and not practically significant.

Figure 21 presents the judging utility data in terms of respondents' online usage levels. There was a general trend of decreasing uncertainty with increased online usage. There was also a general trend of decreased perceptions of judging effectiveness with increased online usage levels.

5.2. Warning Label including Summary

A second type of warning label is presented in Figure 22. This label has a less prominent warning icon and explanation of why the warning is being displayed. It also provides a brief summary of the article and a picture from the article. Data related to this second article are presented in Figures 23–38.
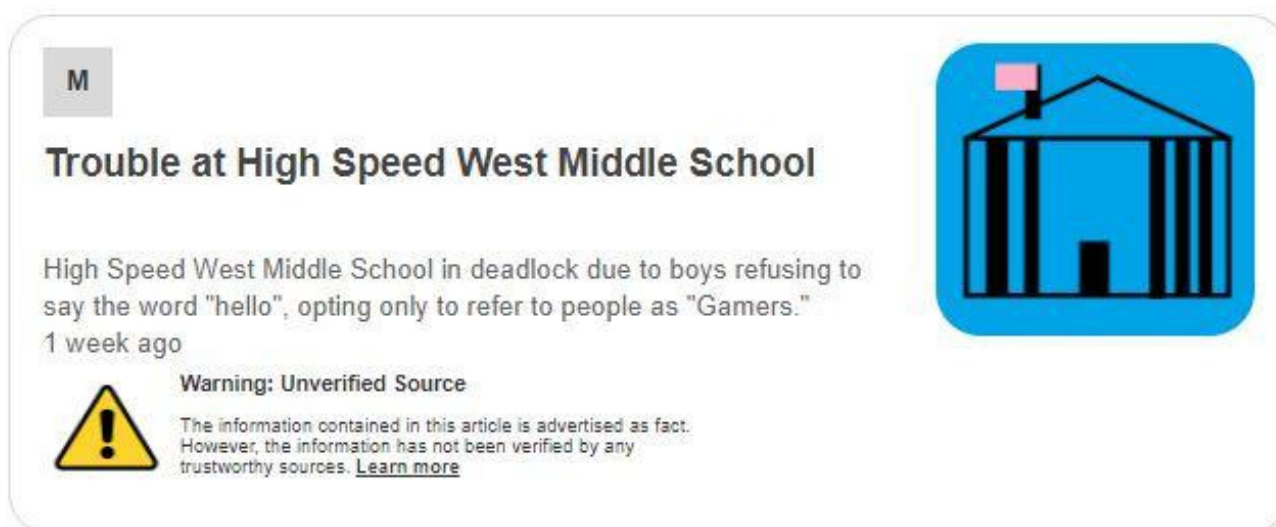
Figure 22. Warning label with article summary [76].

Figures 23–25 present data regarding the perceived helpfulness of the second warning label. Figure 23 presents the helpfulness data in terms of respondents' income level. While there was a reduction in uncertainty with growing income level, in all but the highest income level group, there was no notable trend shown in terms of an association between helpfulness perception and income level.

Figure 24 depicts respondents' perceptions regarding the helpfulness of the second warning label in terms of political affiliation. As in many cases, Democrat respondents had the lowest uncertainty level and highest perception of helpfulness. Republicans had both the second lowest uncertainty and second highest perceived helpfulness responses for this label. Figure 25 presents respondents' perceptions of label helpfulness in terms of their internet usage levels. There were no clear trends present for either the uncertainty levels or the helpfulness perception levels. The differences in helpfulness perception levels were also not practically significant.
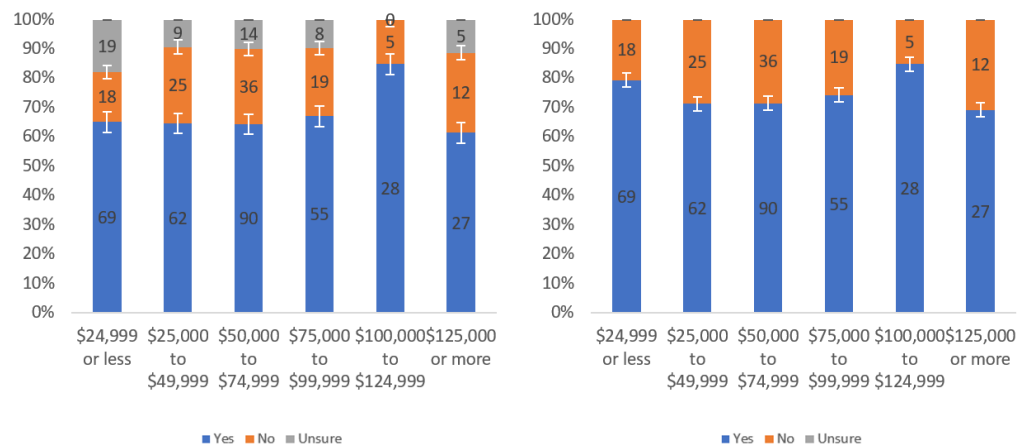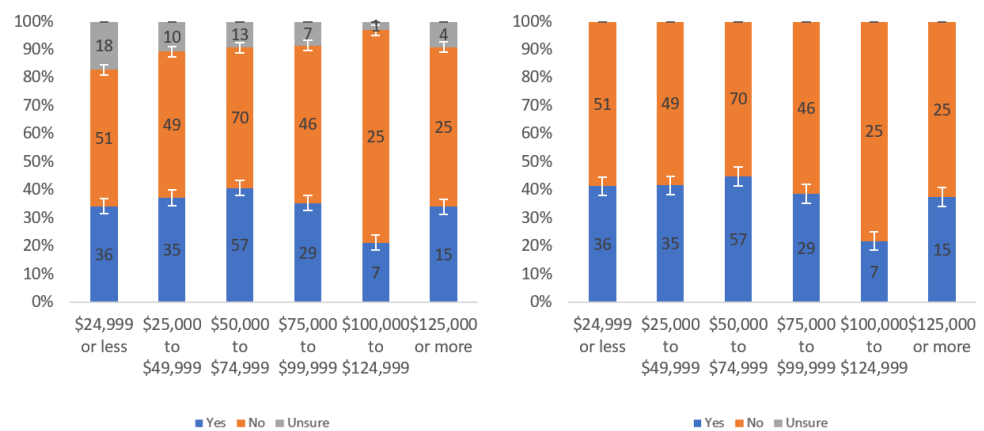
Figure 23. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.
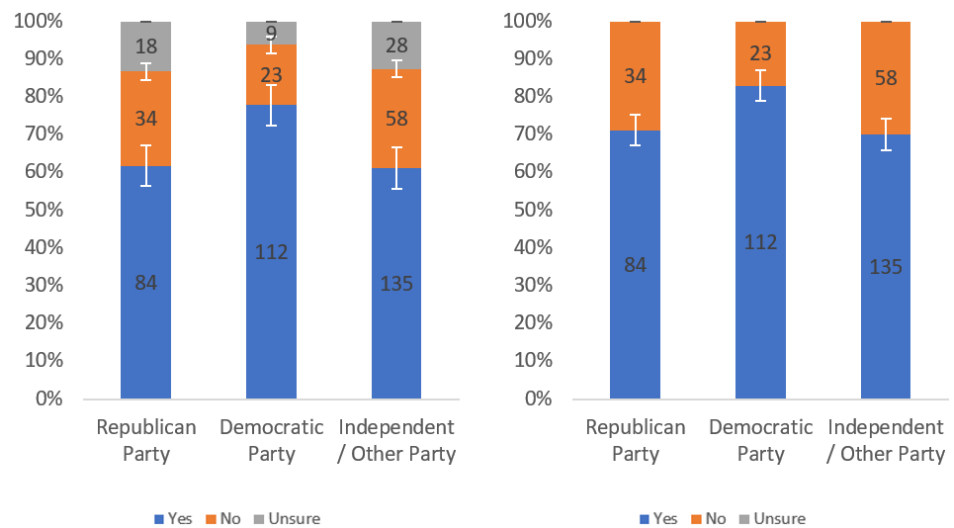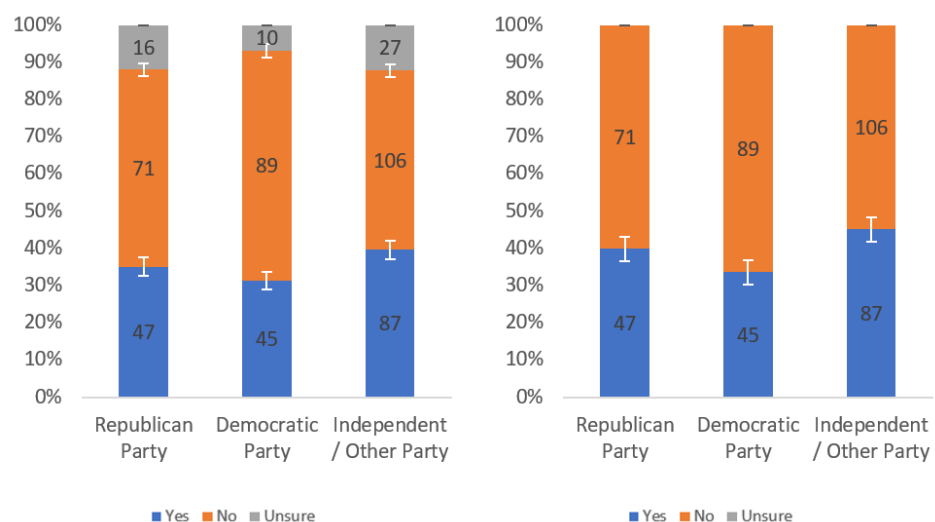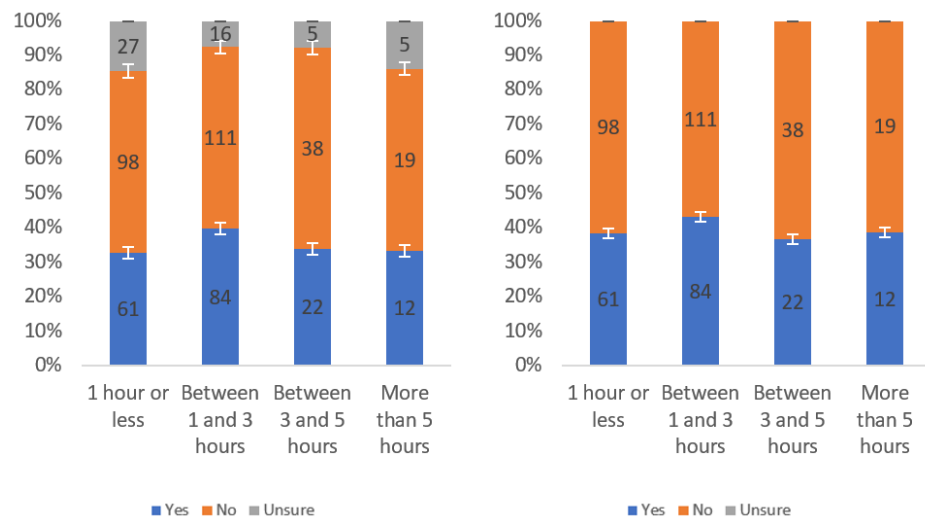

Figure 24. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
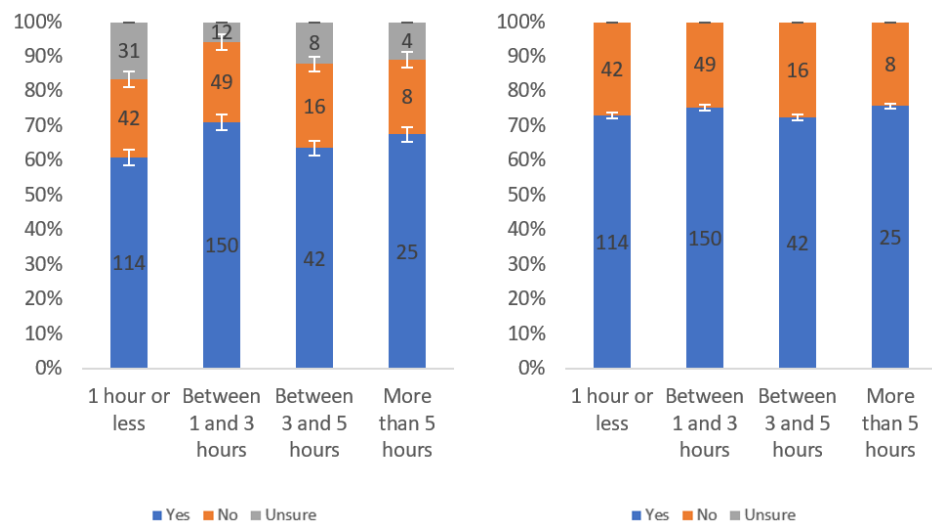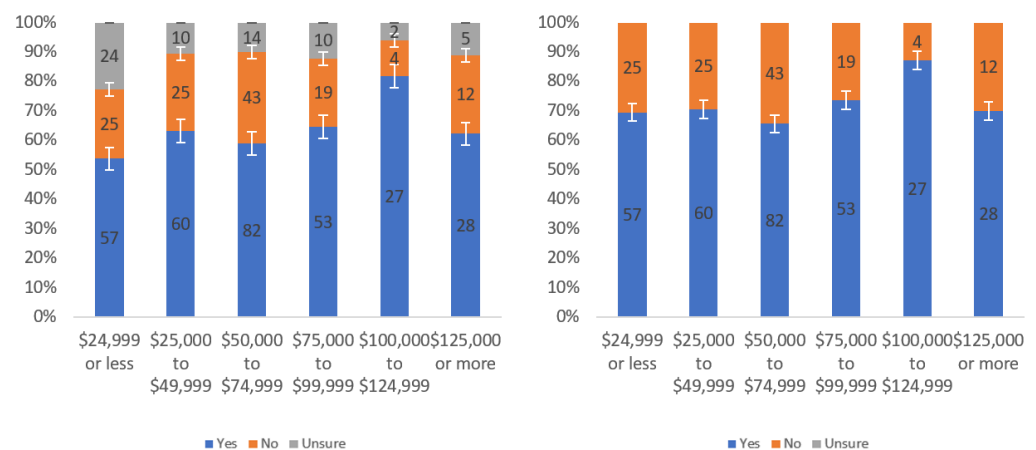

Figure 25. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

The focus now turns to respondents' perception of the second warning label's annoyingness. These data are presented in Figures 26–28. Figure 26 presents this data in terms of respondents' income levels. While Figure 26 shows a noticeable trend of reducing uncertainty with income level (excluding the highest income level group), there was no notable correlation between annoyingness perception and income level.

Figure 27 presents the correlation between annoyingness perception of the second warning level and political affiliation. As usual, the Democratic respondents had the highest perceptions of the label, and they had the lowest uncertainty and the lowest annoyingness levels. Republicans had the second lowest uncertainty and annoyingness levels.

Figure 28 presents the annoyingness level data in terms of online usage time. There was a noticeable reduction in uncertainty with increasing online usage amongst the three lowest usage levels; however, no notable trend was present among the annoyingness level data itself.

Figure 26. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.

Figure 27. Responses regarding label annoyingness by political affiliation including (**left**) and excluding (**right**) unsure responses.

Figure 28. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Next, respondents' willingness to use the second warning label is considered. These data are presented in Figures 29–31. Figure 29 presents this in terms of respondents' income levels. No clear trend was present in this data for either the uncertainty or willingness in terms of income level, though there was, once again, a decrease from the USD 25,000 to USD 50,000 income levels followed by an increase at the USD 75,000 income level.

Figure 30 presents the willingness data in terms of political affiliation. As typical, Democrats had less uncertainty and the highest willingness to use the label. Independents/other party members had the second lowest uncertainty, while Republicans had the second highest willingness to use. While there were noticeable differences in uncertainty levels, the differences in willingness to use levels were less pronounced.

Figure 31 presents the data regarding willingness to use for the second warning label in terms of respondents' online usage levels. The figure shows (with deviation at the highest usage level) a downward trend in uncertainty and a positive trend in willingness to use the label with increasing levels of internet usage.
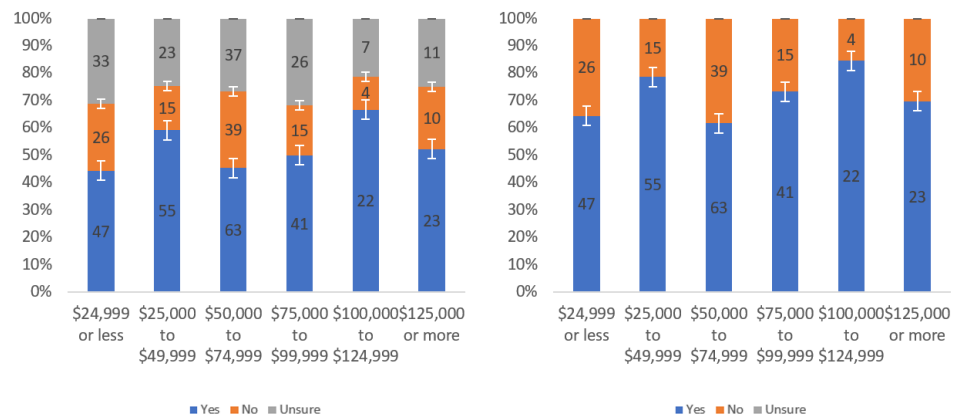

Figure 29. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
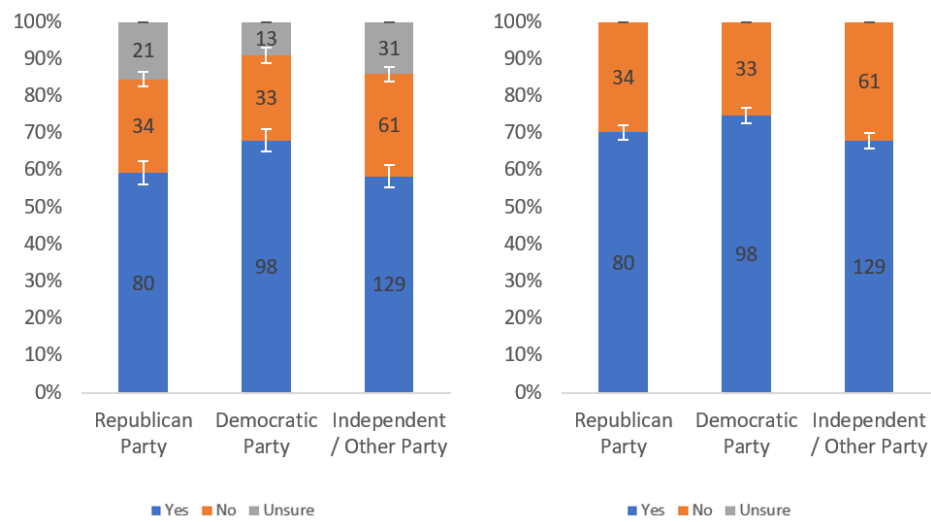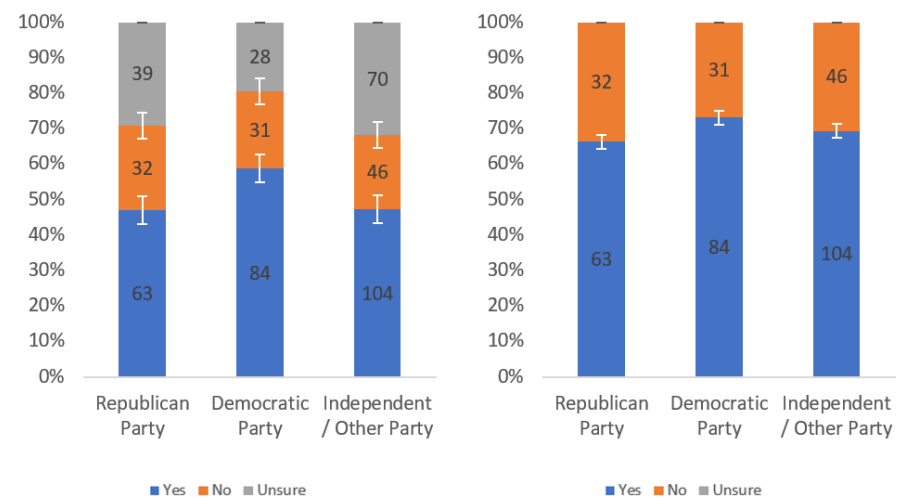
Figure 30. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
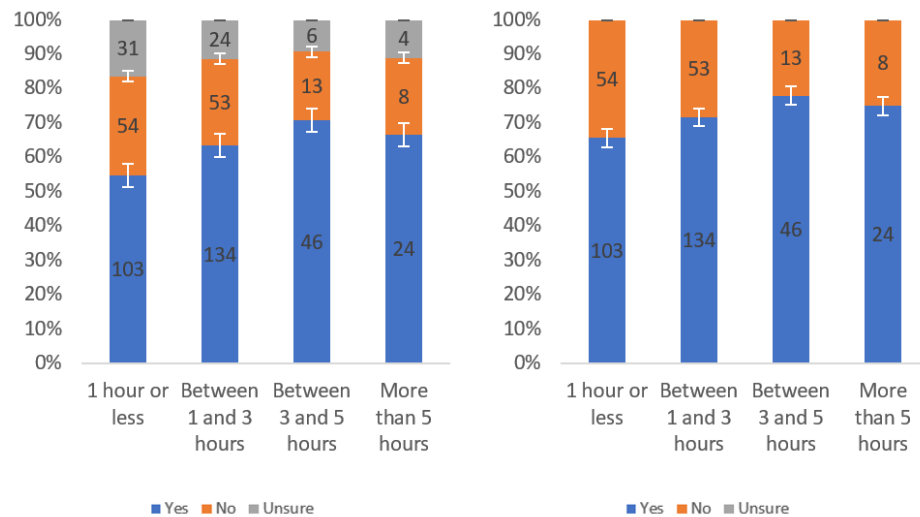
Figure 31. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figures 32–34 present data related to respondents' perception of others' willingness to use the second warning label. Figure 32 presents this data in terms of respondents' income levels. There was no notable trend present in terms of either uncertainty or willingness correlated with income levels with one exception: the previously observed trend at the USD 50,000 and USD 75,000 income levels was again present. There was a sharp decline at the USD 50,000 income level followed by a sharp increase at the USD 75,000 income level.

Figure 33 shows data regarding respondents' perceptions of others' willingness to use the second warning label in terms of political affiliation. As typical, Democrats reported the lowest uncertainty and highest wiliness to use. Republications had the second lowest uncertainty and independents/other party members had the second highest willingness to use levels. Notably, the difference among the three groups' willingness to use levels was quite small.

Figure 34 presents respondents' perceptions of others' willingness to use the second warning label in terms of online usage level. There were trends of decreasing uncertainty and increasing willingness to use the label with increasing online usage levels among the lowest three usage levels. Notably, the differences between the willingness levels were quite small and much smaller than the difference between the uncertainty levels.

Figure 32. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.

Figure 33. Responses regarding others' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
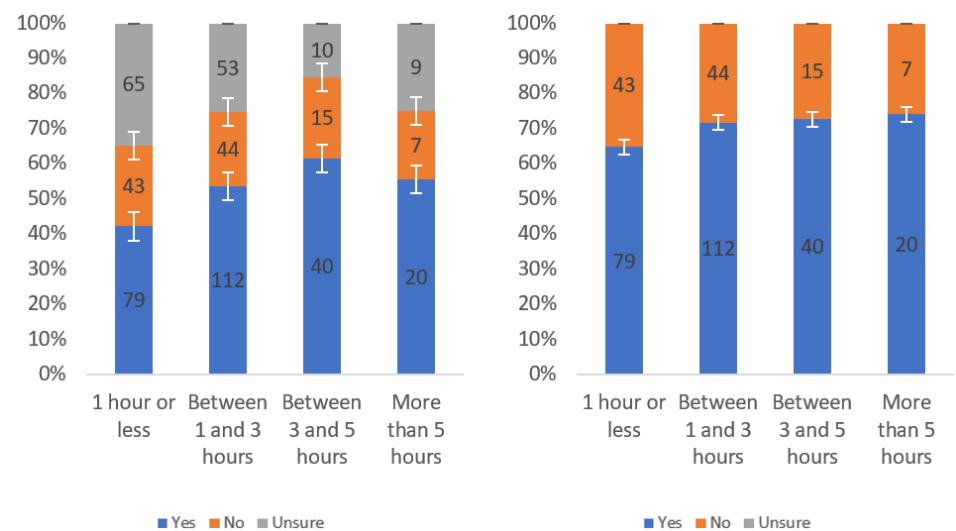

Figure 34. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figures 35–37 present data regarding respondents' perception of the use of the second warning label for judging the trustworthiness of online content. Figure 35 presents this data in terms of income level. There was a noticeable trend amongst all but the highest income level of declining uncertainty and increased perception of utility for judging trustworthiness. Once more, perceived usefulness declined from the USD 25,000 to USD 50,000 income levels and then increased again at the USD 75,000 income level.

Figure 36 presents respondents' perceptions of utility for assessing article trustworthiness in terms of political affiliation. As typical, Democrat respondents reported the lowest uncertainty and highest trustworthiness determination utility. Republicans reported the second lowest uncertainty level and second highest utility levels.

Finally, Figure 37 presents respondents' perceptions of the usefulness of the second warning label for judging trustworthiness in terms of internet usage level. While there was a trend present of declining uncertainty with increased usage level, there was no noticeable trend in the utility perception levels themselves.
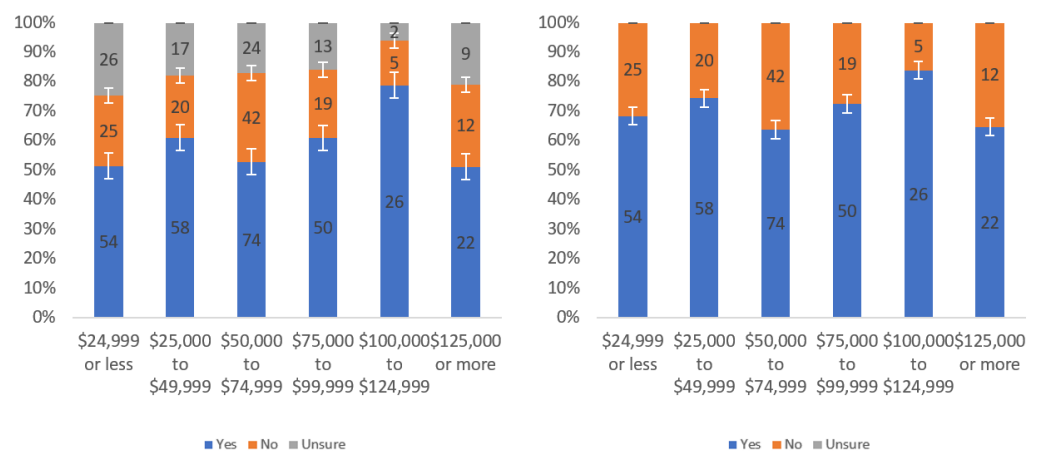
Figure 35. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.
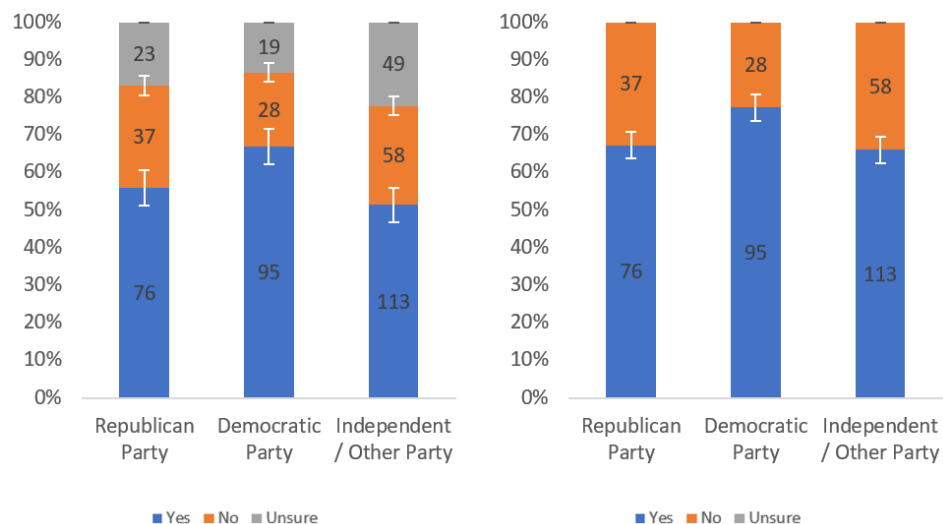
Figure 36. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
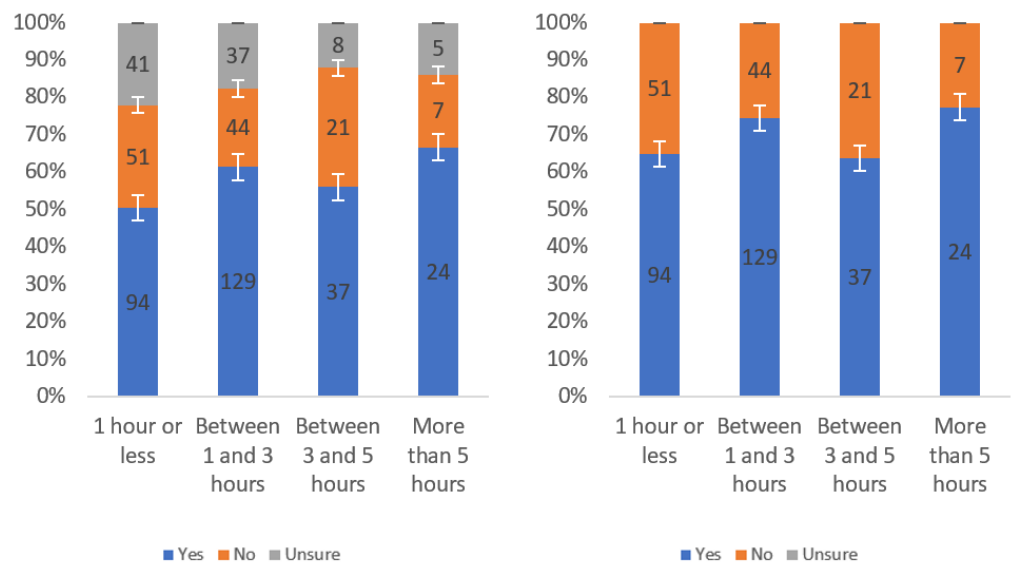
Figure 37. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.
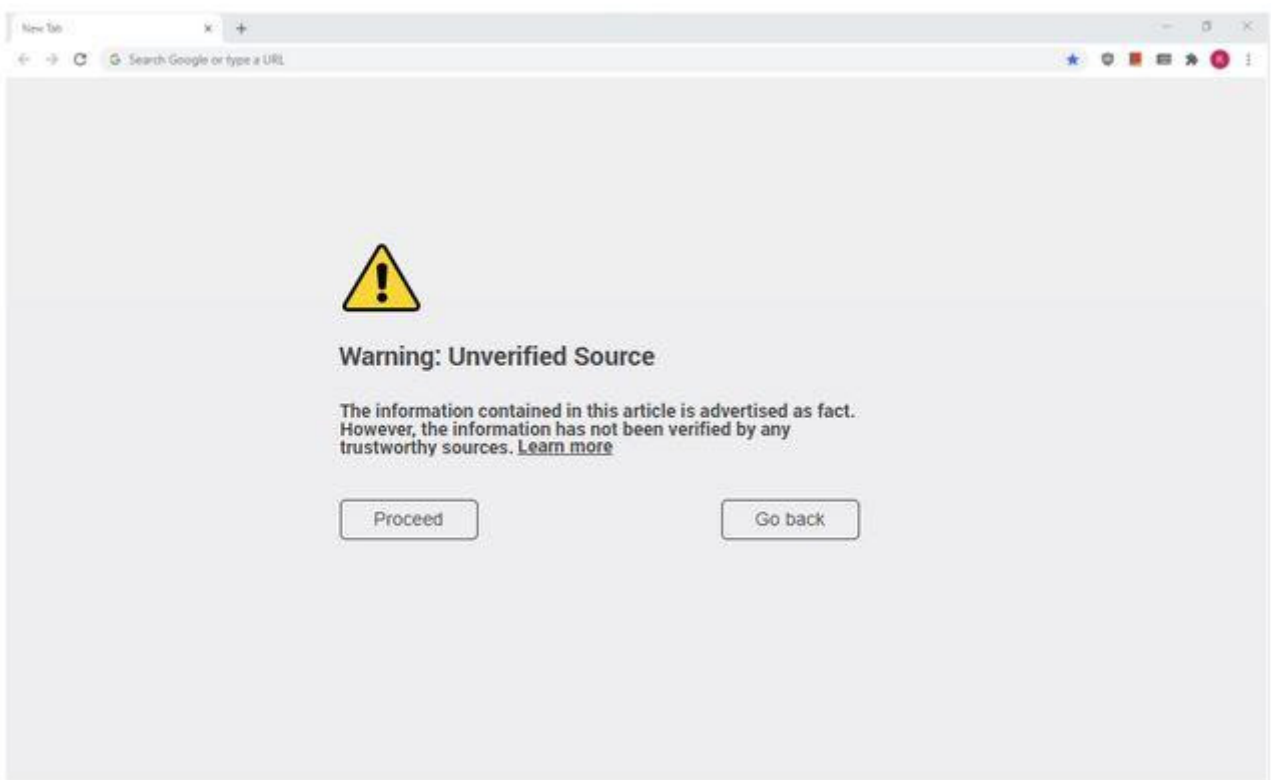

Figure 38. Blocking warning label [76].

5.3. Blocking Warning Label

Focus now turns to the third type of warning label, a blocking warning label. This label is presented on a separate page that loads before the content page loads and must be clicked through to view the content. It has a prominent warning icon and explanation of why a warning is being issued for the content. This third warning label is presented in Figure 38. Data related to respondents' perceptions of it are presented in Figures 39–53.

Figures 39–41 present data regarding the helpfulness of the blocking label. Figure 39 presents this data in terms of income level. Not notable trend was present in the uncertainty

or helpfulness data. A blocky variation of the trend, which peaked at the USD 25,000 and USD 75,000 income levels, was observed again in these data.

Figure 40 presents helpfulness data for the blocking label in terms of political affiliation. As typical, Democrats had lower levels of uncertainty and higher levels of helpfulness. Republicans had the second lowest uncertainty and second highest perceptions of helpfulness.

Figure 41 presents the helpfulness data in terms of respondents' internet usage levels. There was a downward trend present between higher usage levels and greater uncertainty. The level of usefulness also exhibited a slight trend upwards with increased internet usage. The latter is in contrast to other labels which had trends that were almost the opposite.
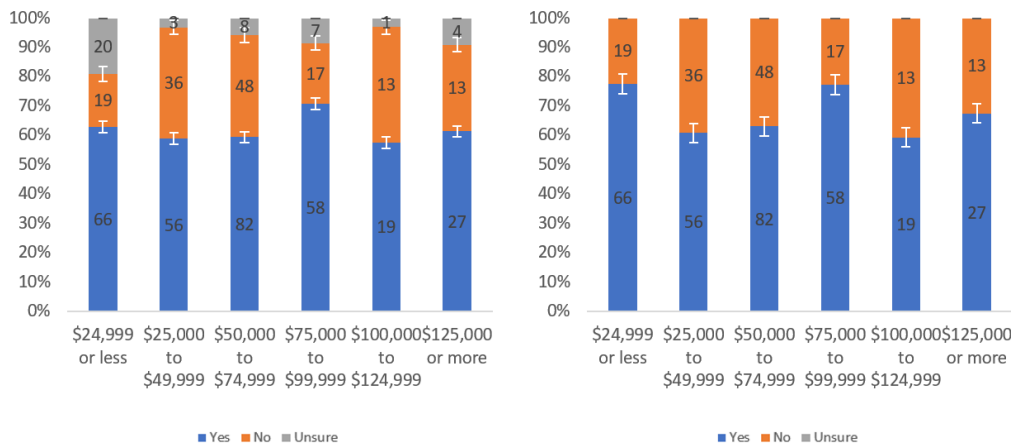
Figure 39. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.
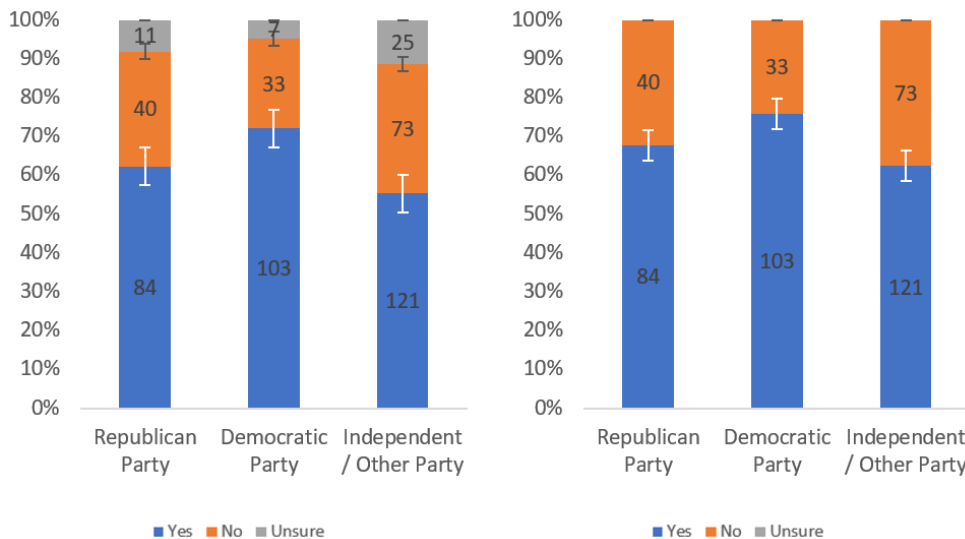
Figure 40. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
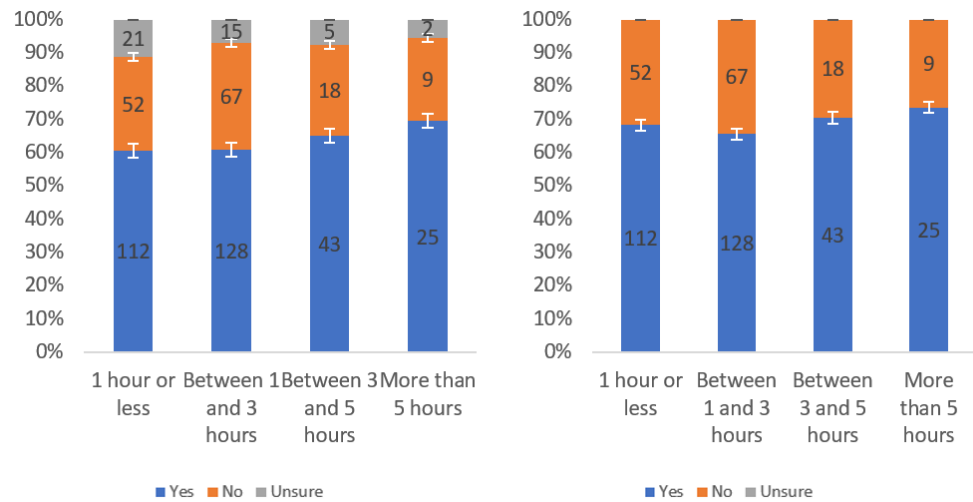
Figure 41. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figures 42–44 present data regarding the annoyingness of the blocking label. Figure 42 presents these data in terms of respondents' income levels. There was no clear trend in either the uncertainty level or annoyingness perception across all levels, though smaller trends existed in each. Most notably, there was a trend between the USD 25,000 and USD 99,999 income levels of decreased annoyingness perception associated with increasing income levels.

Figure 43 presents data regarding annoyingness in terms of political affiliation. While the recurring trend of Democrats having the lowest uncertainty was present, there was only a slight difference in the levels of annoyingness reported by political affiliation. Republicans also reported the lowest level of annoyingness. This differs from the typical situation of Democrats reacting more favorably towards most labels.

Figure 44 presents the annoyingness data for the blocking label in terms of online usage time. These data had a slight trough in annoyingness at the 1–3 h level (which was also the trough for uncertainty levels). The differences in annoyingness among the different internet usage levels were limited.

Now, focus turns to respondents' willingness to use the blocking label. There was a noticeable trend, shown in Figure 45, of declining uncertainty with increased income throughout the income levels. No clear trend was present in the willingness data itself. There was, again, a small increase from the USD 50,000 to USD 75,000 income levels.

Figure 46 presents the willingness to use data by political affiliation. This data mirrors that for other labels. Again, Democrats had the most favorable view of the label and the lowest uncertainty. Nearly 70% of Democrats indicated willingness to use as opposed to just under 65% of Republicans. Approximately 60% of independents/other party members indicated willingness, placing them at the lowest level of willingness. Republicans also had the second lowest uncertainly levels, behind Democrats.

Figure 47 presents the willingness to use levels in terms of online usage level. A trend of decline with increased usage was present in the lowest three of the four uncertainty levels. The willingness to use the blocking label had a negative correlation with increased online usage at these same three levels.
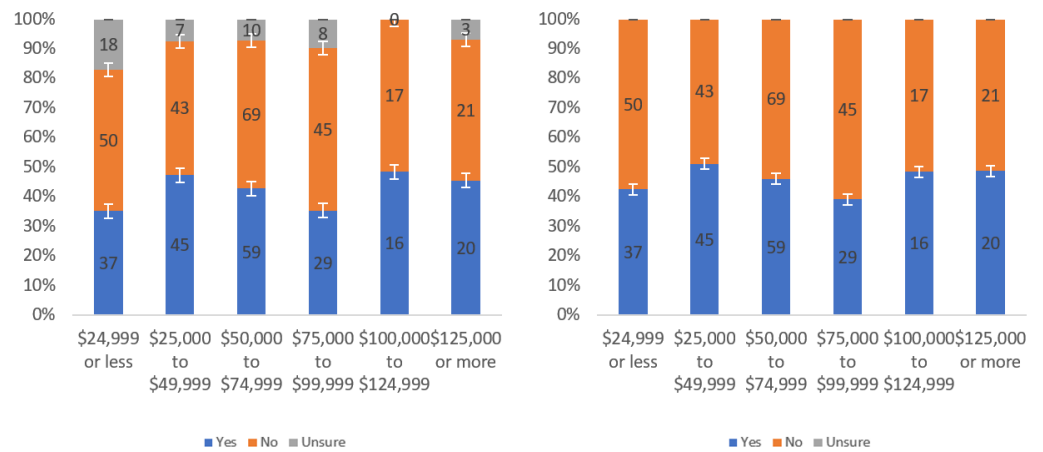
Figure 42. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.
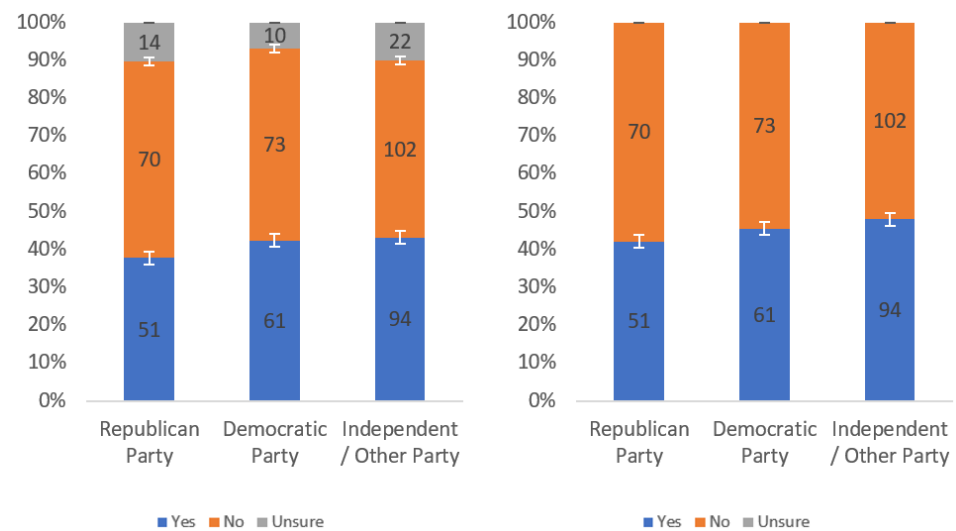

Figure 43. Responses regarding label annoyingness by party affiliation including (**left**) and excluding (**right**) unsure responses.
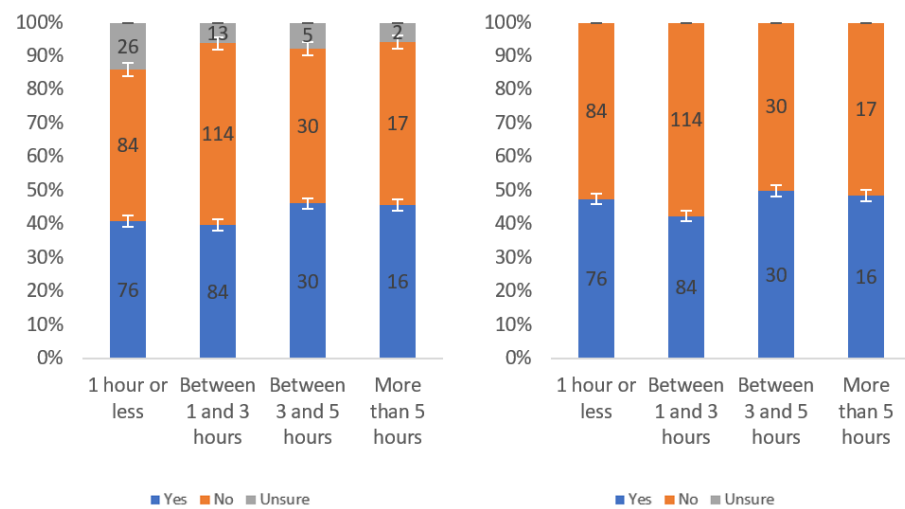

Figure 44. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.
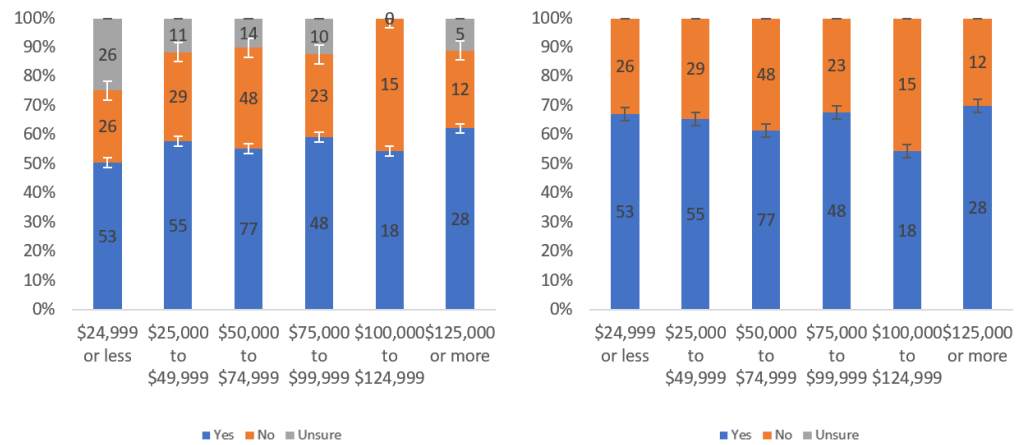
Figure 45. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
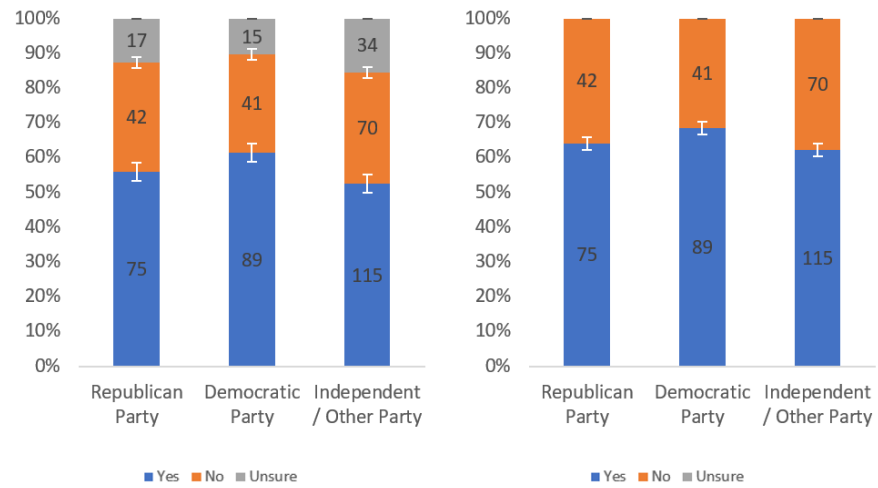
Figure 46. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
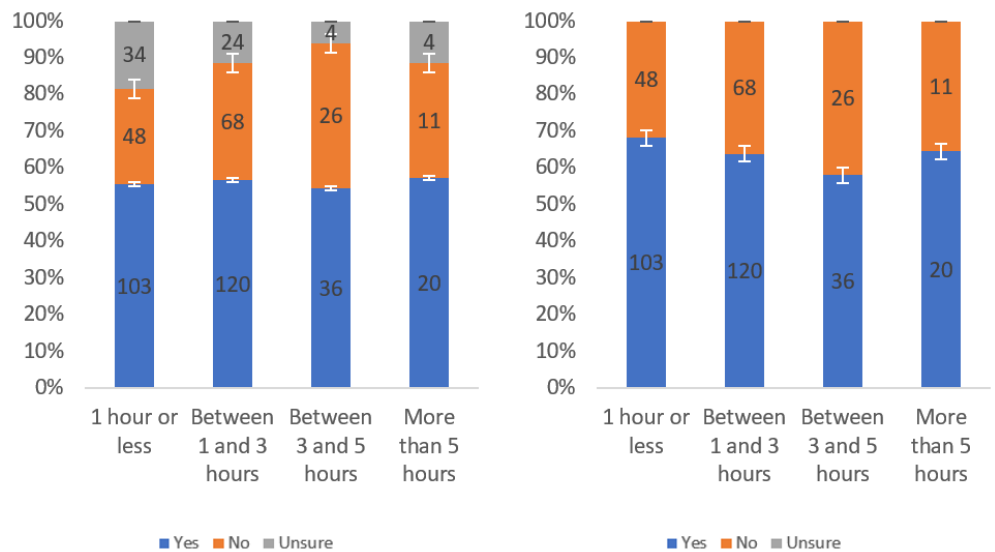
Figure 47. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figures 48–50 present data regarding respondents' perceptions of others' willingness to use the label. Figure 48 presents this by income levels. The data show a trend of decreasing uncertainty with increasing income level and a less noticeable trend of decreasing perception of others' willingness to use the label with income level. There was a small decline from the USD 24,999 or less to USD 50,000 income levels followed by a sharp increase at the USD 75,000 income level, with a more pronounced decline following this.

Figure 49 presents respondents' perceptions of others' willingness to use the labels in terms of political affiliation. Democrats and independents indicated others to be more likely to use the label, for this particular label. Democrats had the lowest uncertainty, followed by Republicans. Independent/other party members indicated uncertainty more than Democrats and Republicans; however, Republicans indicated the lowest level of thinking others would use the label.

Figure 50 presents the willingness to use data in terms of online usage time. A very slight decline with increased usage levels was present in the willingness data. Decreasing uncertainty with increased income levels was shown for the three lowest income brackets.
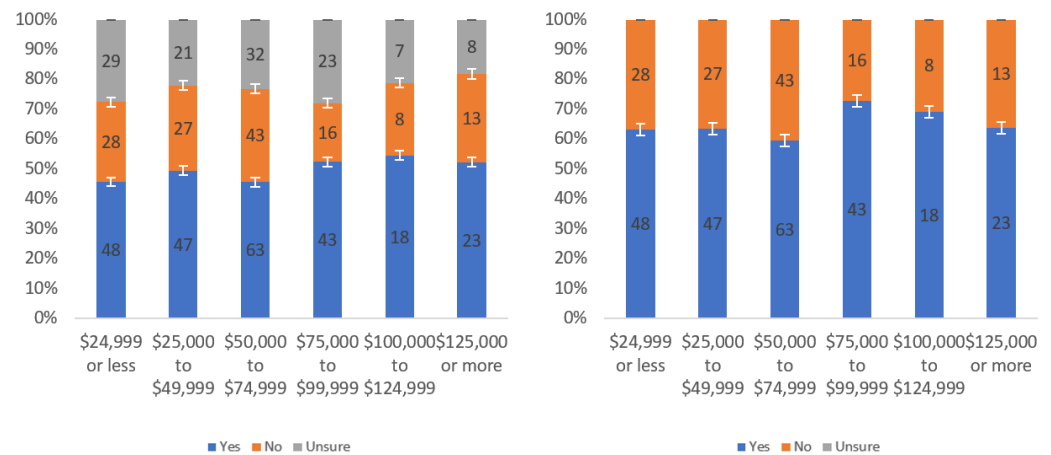

Figure 48. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
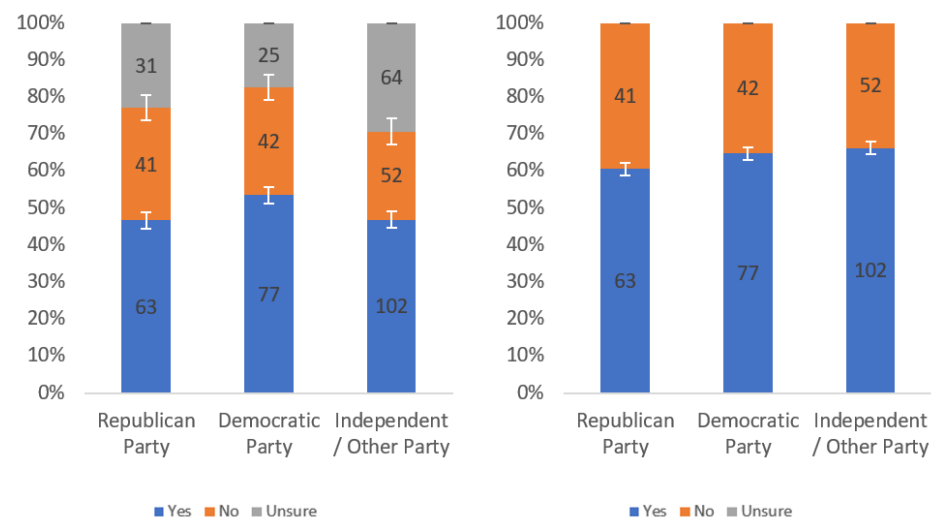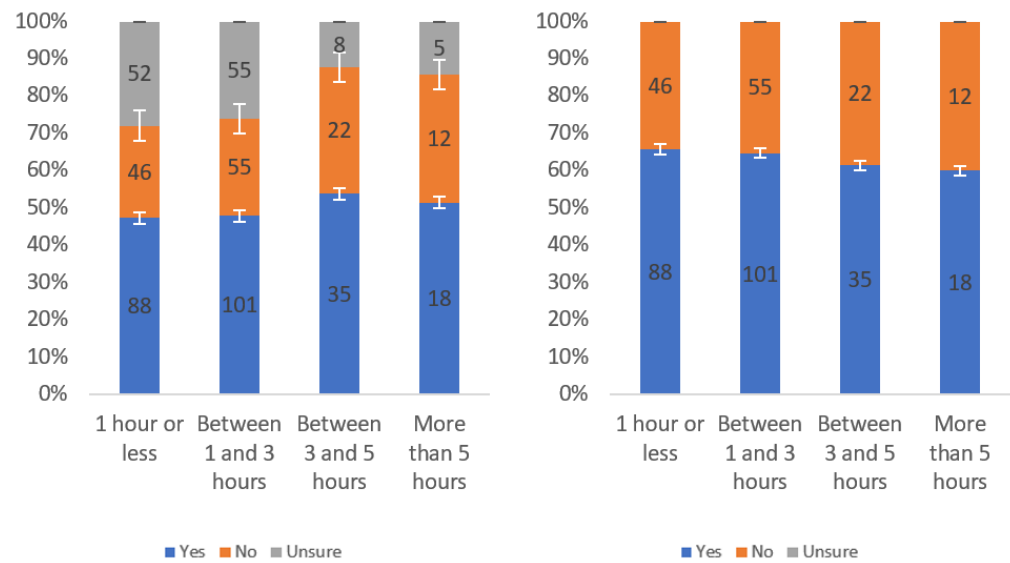

Figure 49. Responses regarding others' willingness to review by political affiliation including (**left**) and excluding (**right**) unsure responses.

Figure 50. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Focus now turns to the blocking label's utility for judging trustworthiness. Figures 51–53 present data related to this. Figure 51 presents data related to income levels. While there was no clear trend regarding uncertainty, there was a slight positive correlation, with deviations at the USD 50,000 and USD 100,000 levels between increased income levels and perception of utility for judging trustworthiness.

In Figure 52, which presents trustworthiness judging utility by political affiliation, the recurring pattern of Democrats having the lowest (though only marginally, in this case) uncertainty levels and highest utility levels was again present. In this instance, Republicans had the second least uncertainty (just slightly less than the Democrats), while the independents/other party members and Republicans had similar utility levels.

Finally, Figure 53 presents the trustworthiness judging utility data in terms of internet usage level. No notable correlation between greater internet usage and utility was shown. Uncertainty had a clear trend of declining with increased internet usage levels.
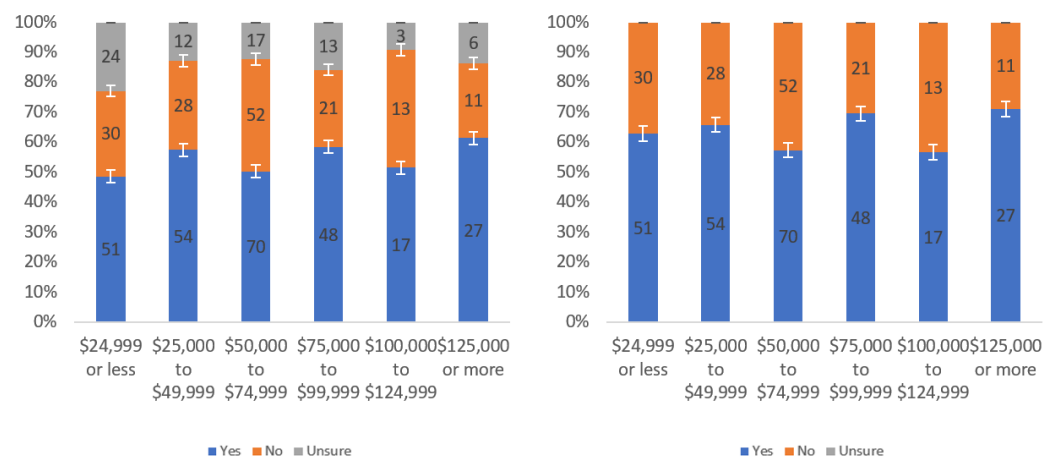
Figure 51. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.
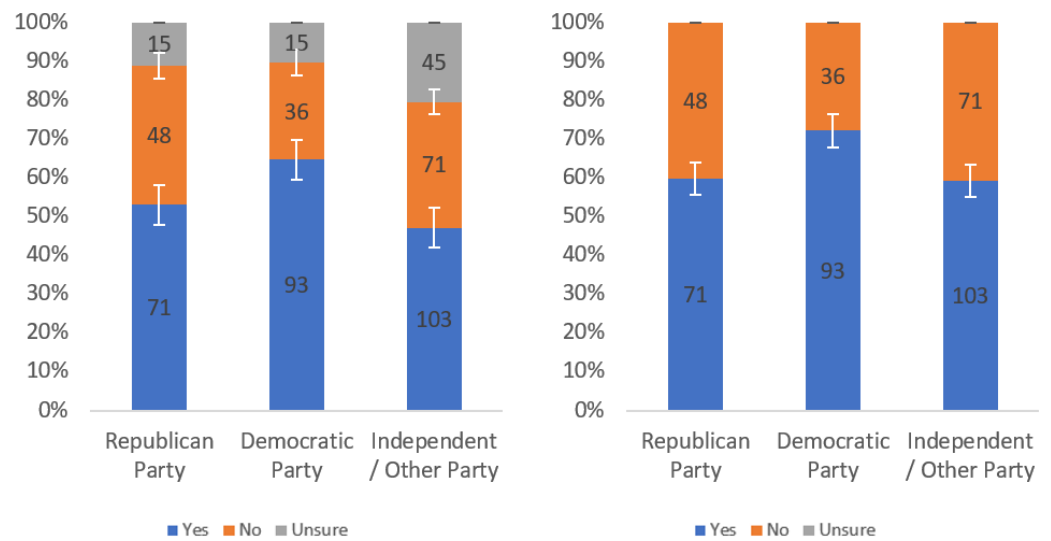
Figure 52. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
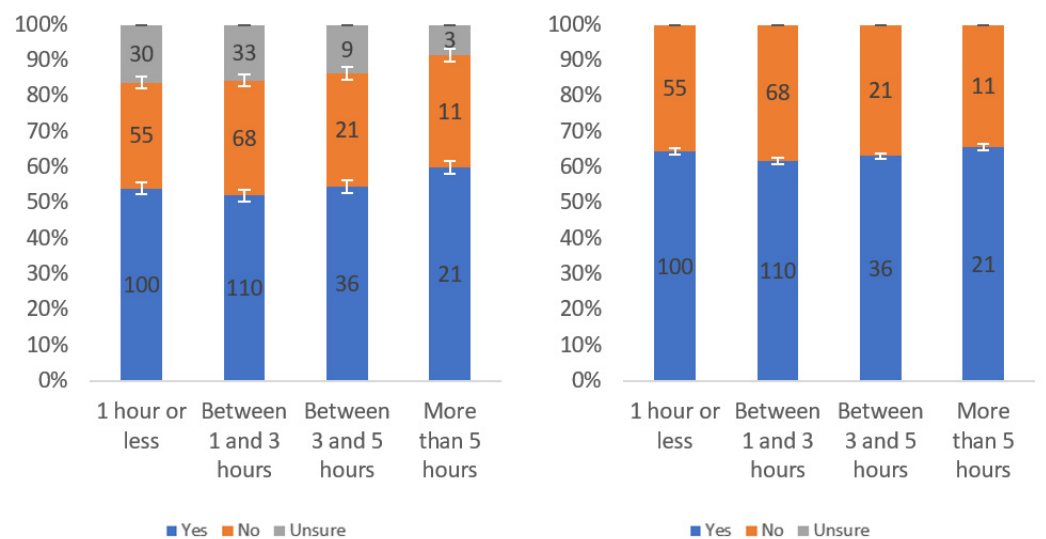
Figure 53. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.

6. Supplemental Information Labels

Finally, a third type of label—one that provides supplemental information—is assessed. As in Sections 4 and 5 for informational and warning labels, respectively, this section presents and discusses the results from several survey questions relevant to supplemental information labels. Figure 54 presents the example of this type of figure that was presented to respondents in the survey. As the figure shows, this label provides the title of the article, a brief summary of the article and a picture from the article. It then has a "learn more" box that provides a link to factual details relevant to claims made in the article.

Figure 54. "Learn more" additional information label [76].

Figures 55–57 present data regarding respondents' perceptions of the helpfulness of this label. Figure 55 presents data in terms of income levels. While there was no notable trend related to the uncertainty level, the label had a negative correlation between increased income levels and perception of helpfulness at the lowest income levels, followed by increasing perception of helpfulness at higher ones.
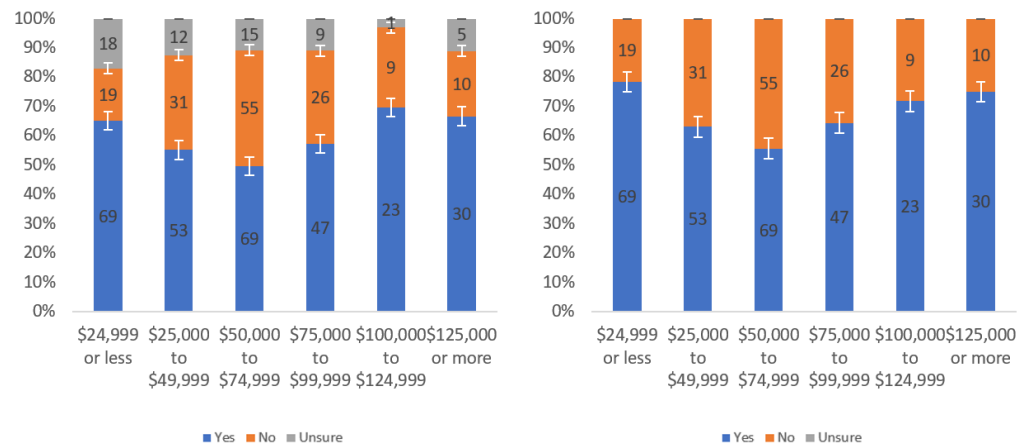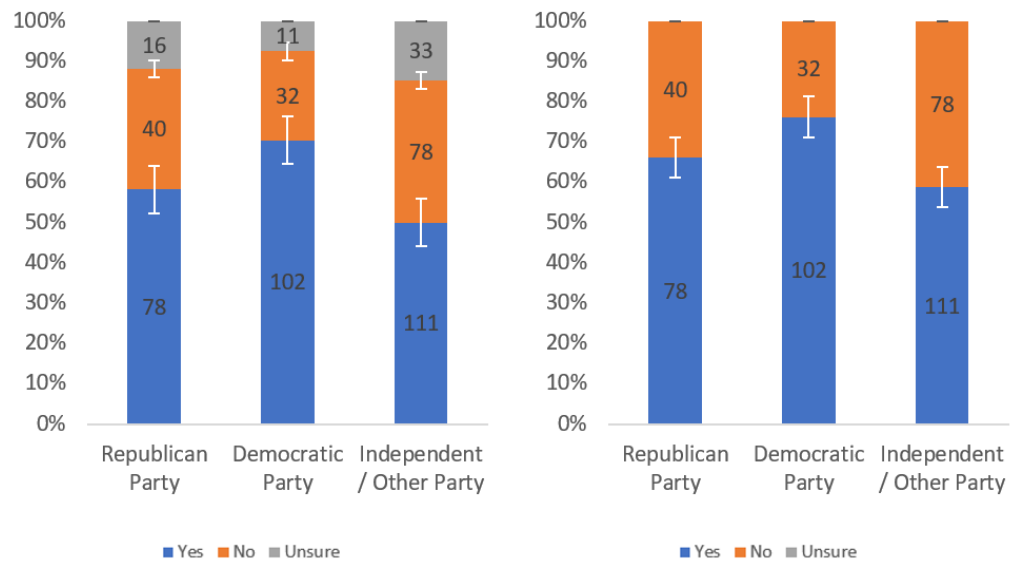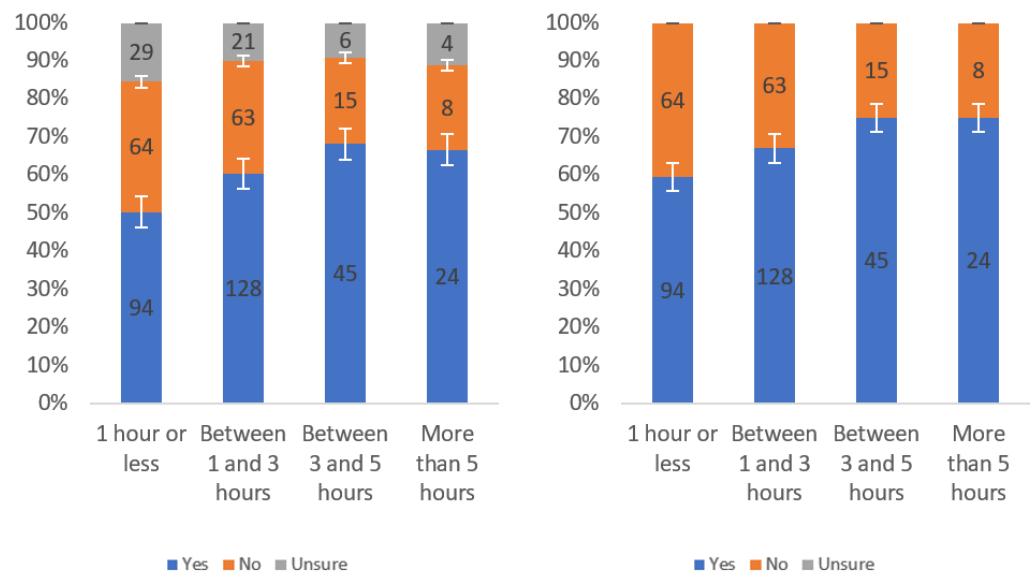

Figure 55. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.

Figure 56 shows the helpfulness data in association with political affiliation. As usual, Democrat respondents found the label to be the most helpful. They also had the lowest level of uncertainty. Republican respondents had slightly more uncertainty and slightly less perceptions of helpfulness. Independents/other party members had higher uncertainty and lower helpfulness perceptions.

The helpfulness data are also presented in terms of correlation with online usage levels. While there was a limited decline in uncertainty shown in the lower three levels of internet usage, a positive trend was present in the actual helpfulness data associated with these internet usage levels.

Figure 56. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
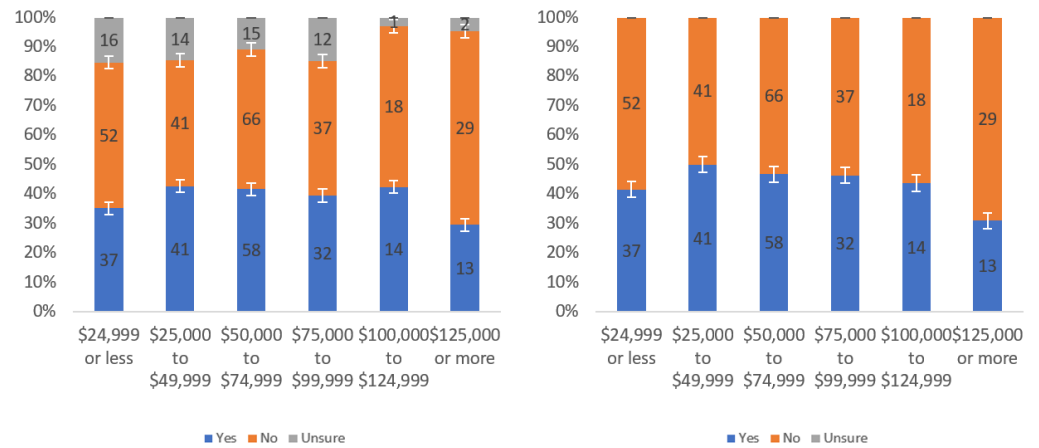

Figure 57. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Next, the annoyingness of the label is assessed. Figures 58–60 present data related to the annoyingness of the supplemental information label. Figure 58 presents data in association with income level. There was a general minor positive association in the uncertainty data, and there was a limited negative correlation between increased income level and perception of annoyingness, excluding the lowest income bracket.

The annoyingness data are presented associated with political affiliation in Figure 59. As typical, Democrats had the lowest uncertainly level. Democrats also found the label the least annoying, followed—with only a slight difference between each—by Republicans and independents.

The annoyingness data are presented in Figure 60 in terms of online usage levels. In this data, the uncertainty values decreased with increased internet usage levels. The annoyingness level values also declined with increased levels of internet usage, except for a very slight increase between the highest two levels.

Figure 58. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.
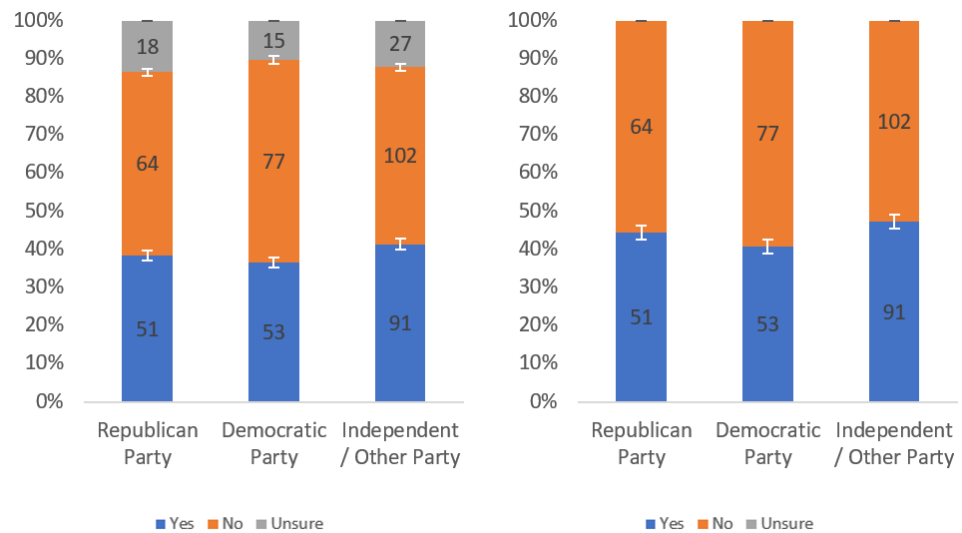
Figure 59. Responses regarding label annoyingness by party affiliation including (**left**) and excluding (**right**) unsure responses.
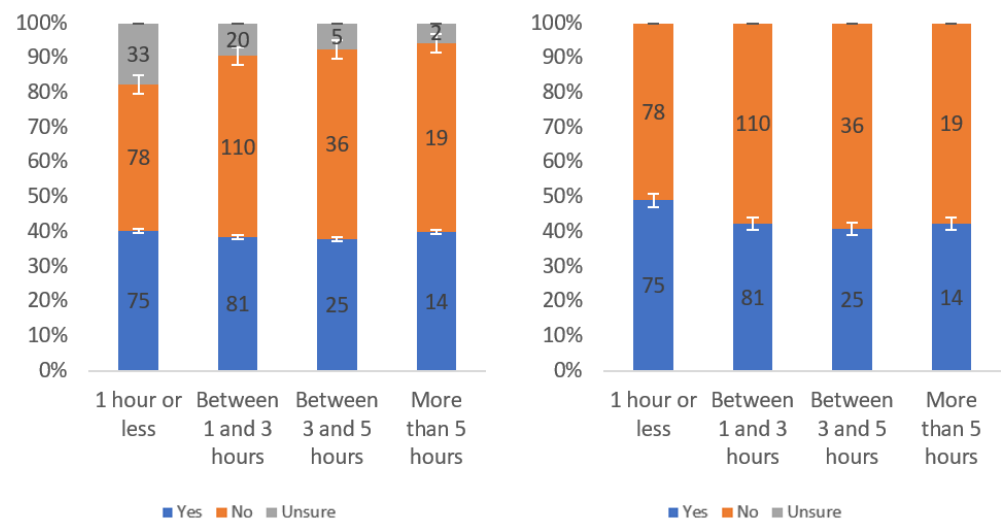
Figure 60. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Focus now turns to respondents' willingness to use the supplemental information label. Figures 61–63 present data related to this. Figure 61 presents data in terms of income level. There was no notable correlation trend between income level and willingness to use. The level of uncertainty declined relatively steadily with increased income level at the lowest three levels; however, no trend was present at higher levels.

Next, in Figure 62, willingness to use data for the supplemental information label is presented in terms of association with political affiliation. As typical, Democrat respondents indicated the lowest uncertainty and also the highest level of usage willingness. Republicans had the second lowest uncertainty and the second highest level of willingness to use.

Figure 63 characterizes respondents' willingness to use the label correlated with internet usage levels. There was a general trend of positive correlation between increased levels of online usage and increased levels of willingness to use the supplemental information label. There was also a small correlation between decreased levels of uncertainty and increased internet usage levels.
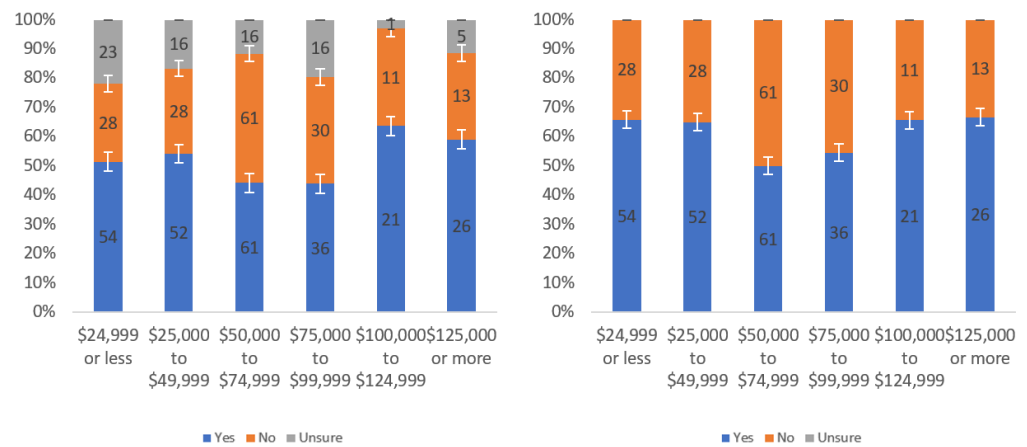

Figure 61. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
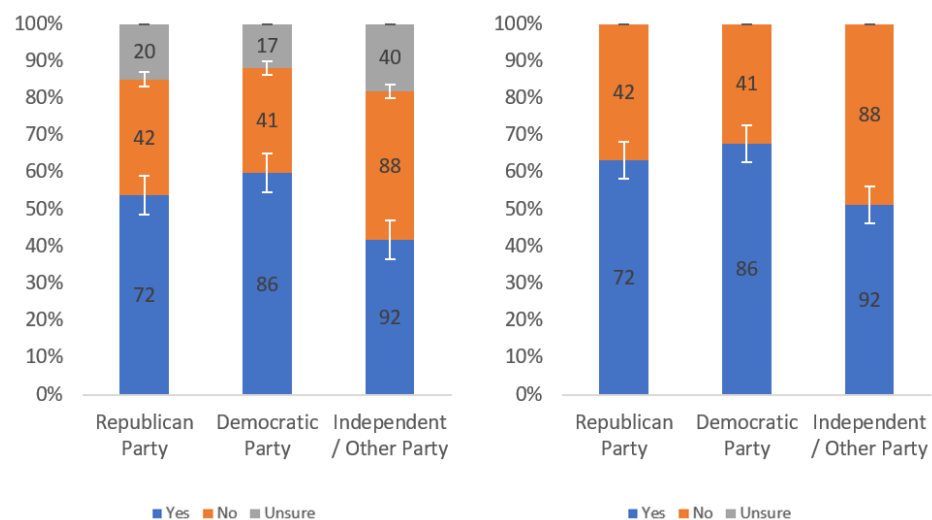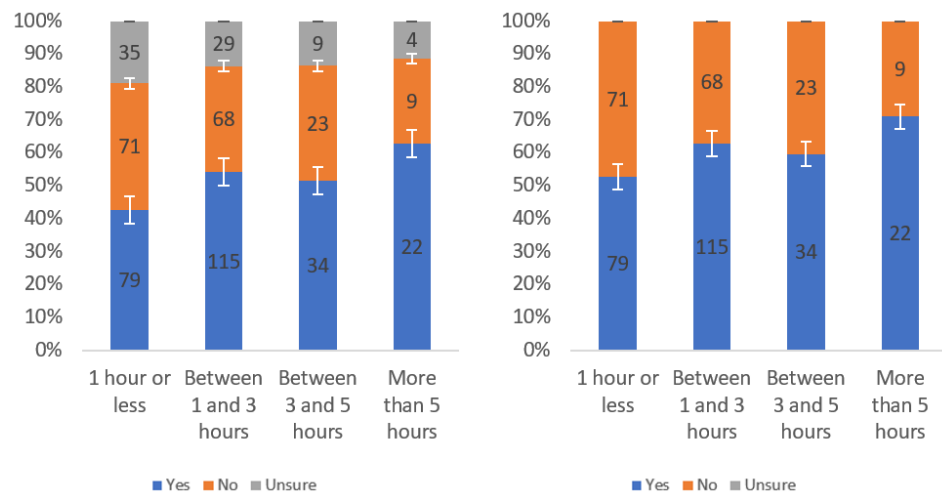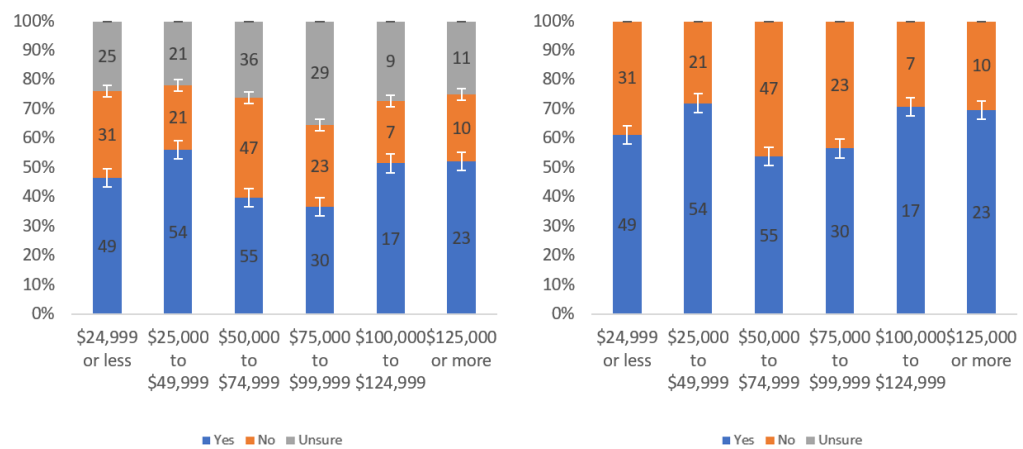

Figure 62. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.

Figure 63. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Now, focus turns to respondents' perceptions of others' willingness to use the supplemental information label. Data related to this is presented in Figures 64–66. Figure 64 presents this in terms of income level. The data show no clear trend between uncertainty and increased income levels. There was also no clear overarching trend visible in the others' willingness level data.

Figure 64. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.

Figure 65 shows the others' willingness data correlated with political affiliation. As is typical, Democrats had the lowest level of uncertainty and the highest level of perceived willingness of others to use the labels. Republicans had the second lowest uncertainty levels and the second highest others' willingness levels.

Figure 66 presents the others' willingness to use data in terms of internet usage level. A clear trend of increased perception of others' willingness correlated with increased internet usage was present. The lowest three levels of internet usage also had a trend of decreased uncertainty with increased internet usage.
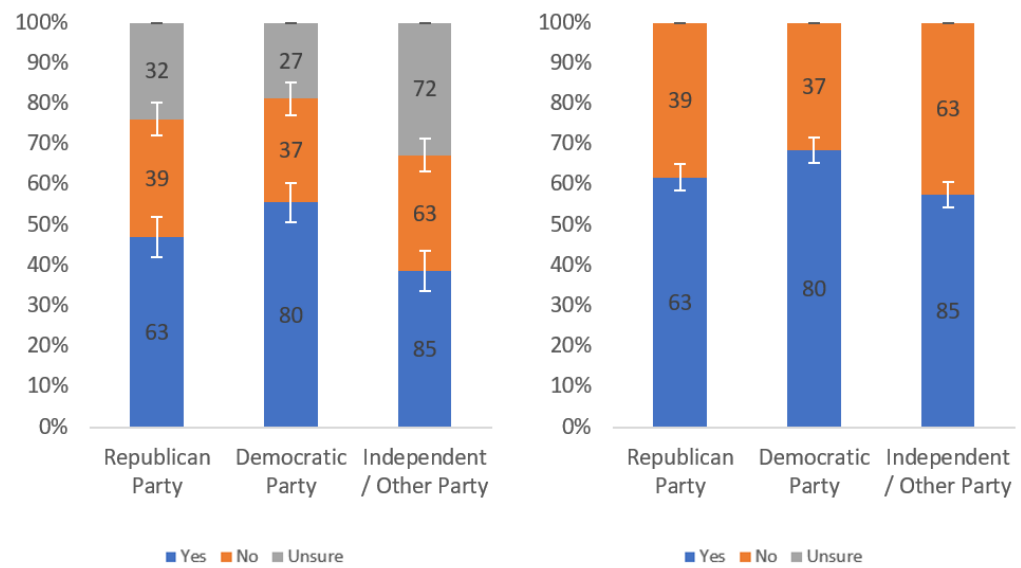
Figure 65. Responses regarding others' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
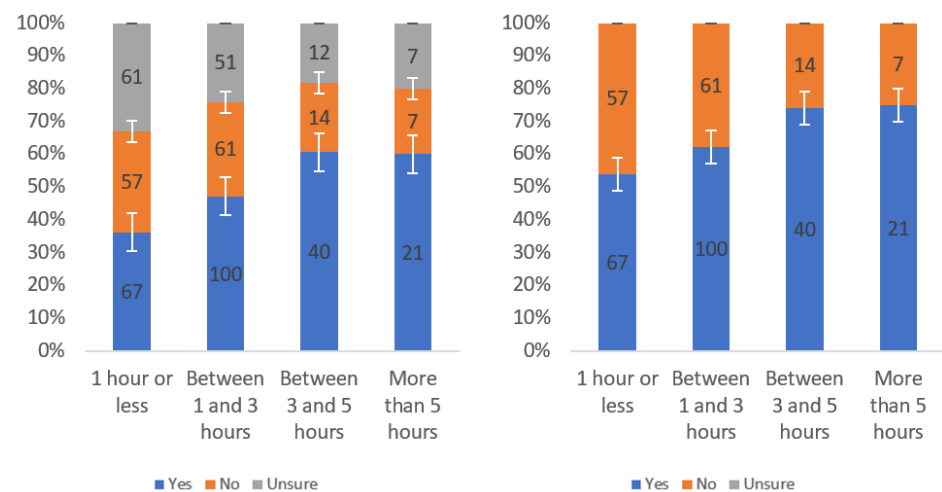

Figure 66. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Finally, the label's efficacy for judging trustworthiness is assessed. Figures 67–69 present data related to this. Figure 67 presents this data associated with income level. No clear trends were present in either the uncertainty level or usefulness data.

The trustworthiness judging efficacy data are presented in terms of political affiliation in Figure 68. As typical, Democrats reported lower uncertainty and higher levels of perceived trustworthiness judging efficacy. Republicans had the second highest uncertainty and the second highest level of trustworthiness judging efficacy belief. Independents reported the greatest uncertainty and lowest usefulness among the three groups.

Finally, the trustworthiness judging efficacy of the supplemental information label is considered relative to online usage levels. For both the uncertainty and usefulness values, a trend was present. The first trend was decreasing uncertainty with increased usage, and the second trend was an increasing perceived utility for judging trustworthiness that was positively associated with increased internet usage.

Figure 67. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.

Figure 68. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
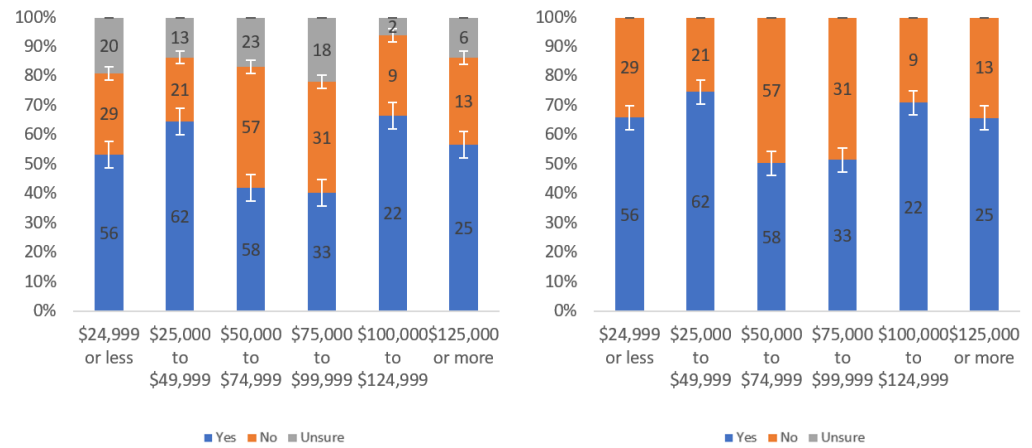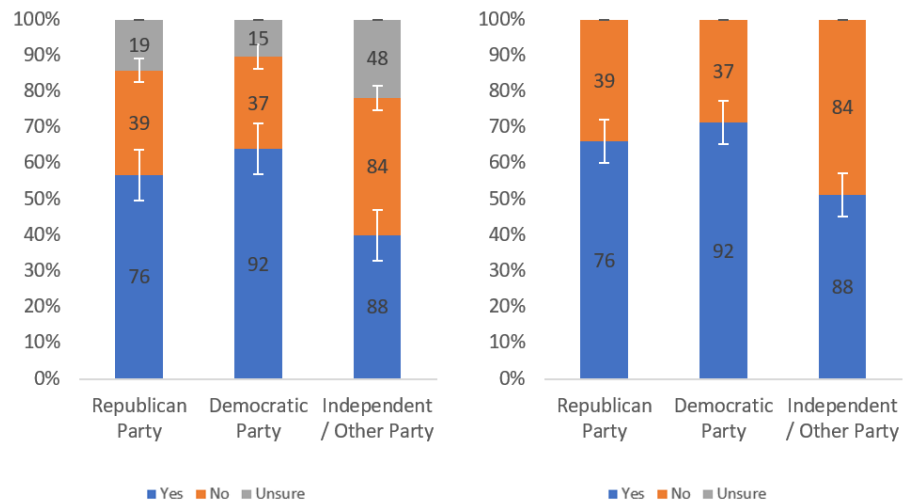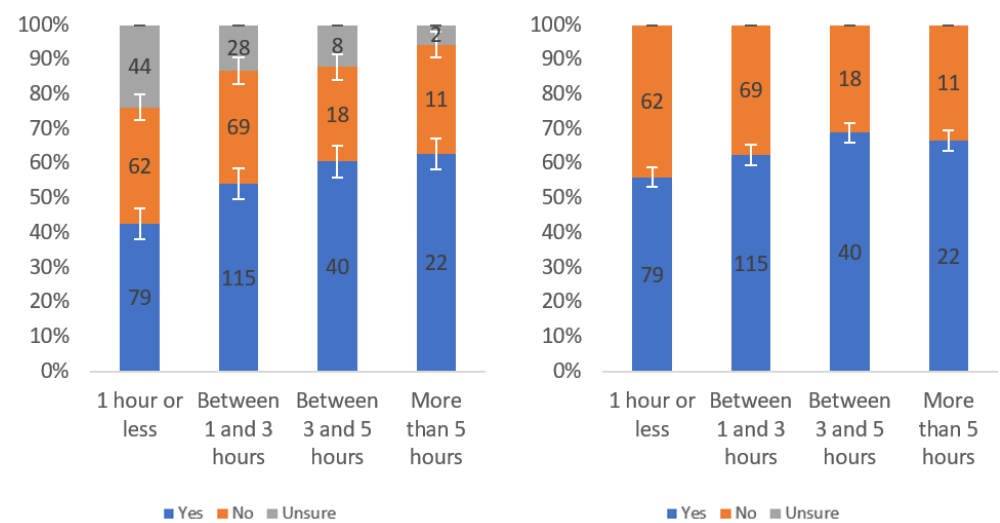
Figure 69. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.

7. Analysis and Implications

This section presents comparative analysis of the data presented in the previous sections. First, the trends present in the data, which were previously discussed, are briefly summarized and key patterns are identified. Then, the level of willingness of respondents and respondents' perceived willingness of others to use each label is compared between labels and by income level, political affiliation and online usage levels. Finally, respondents' willingness to use each label is compared to their perceived willingness of others for each label and based on income level, political affiliation and online usage levels.

Table 1 presents the trends present throughout the data for the seven labels in terms of income level. Several overarching trends are present. First, in general, metrics regarding the perception of the labels decreased with increased income. There were two limited exceptions to this. The first was annoyingness, which showed a slight positive association with income level in one instance, no notable trend in two instances, conflicting trends in one instance and a partial negative association in two instances. The second was trustworthiness, which had a marginal positive association in one instance and a partial positive association in another instance (it also had one negative, one partial negative and two no notable associations). Generally, though, the labels were seen less positively with increased income levels (and as more annoying, in at least one case). The number of respondents indicating uncertainty also had a strong negative correlation with increased income. Thus, respondents can be taken to be less supportive of the labels and more certain of this belief with increased income levels. An exception to this rule was seen at the USD 75,000 income level, which, for some labels, showed an increase in support for positive question categories (i.e., helpfulness, self-willingness, others'-willingness and usefulness) relative to the USD 50,000 income level. Similarly, there was a decrease, in some cases, in "annoyance" when shifting from the USD 50,000 to the USD 75,000 income level.

The three labels that showed an at least somewhat positive trend in perceived trust with increasing income were the labels with the most information. Informational label 3 has the extended description of each metric presented and warning label 2 presents the article summary along with the rationale for the warning. The supplemental information label provides a link to a relevant source. This may be indicative of those with higher income levels seeking more information for their decision making. It may be that a label with more information on it (or perhaps linked to from it) than any of the ones analyzed herein would perform the best for higher income groups. This is a potential topic for future work to analyze.

Table 2 compares the trends in the perception of labels based on respondents' political affiliations. Most notable from reviewing these data was that in almost all cases, Democratic Party-affiliated respondents found the labels the most helpful and useful for judging trustworthiness, and they believed that they and others were most likely to use them. In most cases, Democrats found the labels the least annoying or there was a tie for finding them the least annoying. Because of this, the most notable data elements were those where this pattern did not hold. In terms of the metric, there were only two instances of this: independents/other party affiliates indicated others' willingness to use warning labels 1 and 3 at a higher level than Democrats. Similarly, on the uncertainty side, there was a single exception: Republicans had higher certainty for self-willingness for informational label 2. However, even with these two deviations, the pattern of higher support for labels by Democrats is very clear and pronounced.

Table 3, similar to Tables 1 and 2, compares the trends in the perception of labels based on respondents' online usage levels. This data were far more varied than the data in the previous table and were, in some cases, somewhat contradictory in potential meaning. Looking at the metric data, informational label 1 had positive correlations (of various levels) to all metrics except others' willingness to use the label. Notably, the only strong positive correlation was for self-willingness, which increased with online usage levels. However, the slight and partial positives for helpfulness and judging trustworthiness (which aligned with

the positive for self-willingness) were at odds with the partial positive for annoyingness. However, these differences are marginal.

Table 1. Trends for label types for income level data.

	Metric	**Uncertainty**
Informational Label 1		
Helpfulness	Two negatives	Partial negative
Annoyingness	Partial negative	Fluctuates
Self-Willingness	Slightly negative	Slightly negative
Others' Willingness	Partial negative	Partial positive
Judging Trustworthiness	Negative	Partial negative
Informational Label 2		
Helpfulness	Two negatives	Not notable
Annoyingness	Slightly positive	Negative
Self-Willingness	Partial negative	Partial negative
Others' Willingness	Not notable	Not notable
Judging Trustworthiness	Partial negative	Not notable
Informational Label 3		
Helpfulness	Conflicting trends	Partial negative
Annoyingness	Conflicting trends	Negative
Self-Willingness	Not notable	Not notable
Others' Willingness	Not notable	Not notable
Judging Trustworthiness	Partial positive	Partial slight negative
Warning Label 1		
Helpfulness	Slightly negative	Slightly negative
Annoyingness	Not notable	Slightly negative
Self-Willingness	Slightly negative	Negative
Others' Willingness	Partial negative	Partial negative
Judging Trustworthiness	Not notable	Negative
Warning Label 2		
Helpfulness	Not notable	Partial negative
Annoyingness	Not notable	Partial negative
Self-Willingness	Not notable	Not notable
Others' Willingness	Not notable	Not notable
Judging Trustworthiness	Marginally Positive	Negative
Warning Label 3 (Blocking)		
Helpfulness	Not notable	Not notable
Annoyingness	Partial negative	Partial negative
Self-Willingness	Partial slight negative	Negative
Others' Willingness	Partial negative	Slightly negative
Judging Trustworthiness	Not notable	Not notable
Supplemental Info. Label		
Helpfulness	Conflicting trends	Partial slight negative
Annoyingness	Partial negative	Partial negative
Self-Willingness	Partial positive	Partial negative
Others' Willingness	Not notable	Partial negative
Judging Trustworthiness	Not notable	Not notable

Table 2. Trends for label types for political affiliation data.

	Metric	**Uncertainty**
Informational Label 1		
Helpfulness	Democrats most helpful	Democrats most sure
Annoyingness	Effective tie	Democrats slightly surer
Self-Willingness	Democrats most willing	Democrats most sure
Others' Willingness	Democrats most willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats most sure
Informational Label 2		
Helpfulness	Democrats most helpful	Democrats most sure
Annoyingness	Democrats slightly less annoying	Democrats slightly surer
Self-Willingness	Democrats most willing	Republicans slightly surer
Others' Willingness	Democrats most willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats most sure
Informational Label 3		
Helpfulness	Democrats most helpful	Democrats slightly surer
Annoyingness	Democrats less annoying	Democrats most sure
Self-Willingness	Democrats most willing	Democrats most sure
Others' Willingness	Democrats most willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats slightly surer
Warning Label 1		
Helpfulness	Democrats most helpful	Democrats slightly surer
Annoyingness	Effective tie	Democrats very slightly surer
Self-Willingness	Democrats slightly more willing	Democrats slightly surer
Others' Willingness	Independents slightly more willing	Democrats slightly surer
Judging Trustworthiness	Democrats slightly more useful	Democrats most sure
Warning Label 2		
Helpfulness	Democrats most helpful	Democrats most sure
Annoyingness	Democrats less annoying	Democrats most sure
Self-Willingness	Democrats slightly more willing	Democrats most sure
Others' Willingness	Democrats slightly more willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats slightly surer
Warning Label 3 (Blocking)		
Helpfulness	Democrats most helpful	Democrats very slightly surer
Annoyingness	Effective tie	Democrats very slightly surer
Self-Willingness	Democrats slightly more willing	Democrats slightly surer
Others' Willingness	Independents most willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats very slightly surer
Supplemental Information Label		
Helpfulness	Democrats most helpful	Democrats surer
Annoyingness	Democrats slightly less annoying	Democrats slightly surer
Self-Willingness	Democrats most willing	Democrats most sure
Others' Willingness	Democrats most willing	Democrats most sure
Judging Trustworthiness	Democrats most useful	Democrats most sure

Table 3. Trends for label types for online usage level data.

	Metric	Uncertainty
Informational Label 1		
Helpfulness	Slightly positive	Not notable
Annoyingness	Partial positive	Partial negative
Self-Willingness	Positive	Not notable
Others' Willingness	Not notable	Partial negative
Judging Trustworthiness	Partial positive	Partial negative
Informational Label 2		
Helpfulness	Slightly positive	Not notable
Annoyingness	Not notable	Not notable
Self-Willingness	Slightly positive	Partial negative
Others' Willingness	Not notable	Partial negative
Judging Trustworthiness	Slightly negative	Slightly positive
Informational Label 3		
Helpfulness	Partial positive	Not notable
Annoyingness	Slightly negative	Not notable
Self-Willingness	Partial negative	Positive
Others' Willingness	Slightly negative	Positive
Judging Trustworthiness	Slightly positive	Negative
Warning Label 1		
Helpfulness	Partial negative	Partial positive
Annoyingness	Not notable	Not notable
Self-Willingness	Not notable	Partial negative
Others' Willingness	Not notable	Slightly negative
Judging Trustworthiness	Negative	Slightly negative
Warning Label 2		
Helpfulness	Not notable	Not notable
Annoyingness	Not notable	Partial negative
Self-Willingness	Partial negative	Partial positive
Others' Willingness	Partial negative	Partial positive
Judging Trustworthiness	Not notable	Negative
Warning Label 3 (Blocking)		
Helpfulness	Partial slight positive	Slightly negative
Annoyingness	Not notable	Negative
Self-Willingness	Partial negative	Partial negative
Others' Willingness	Slightly negative	Negative
Judging Trustworthiness	Not notable	Slightly negative
Supplemental Info. Label		
Helpfulness	Partial positive	Partial negative
Annoyingness	Partial slight negative	Negative
Self-Willingness	Positive	Slightly negative
Others' Willingness	Positive	Partial negative
Judging Trustworthiness	Partial positive	Negative

Informational label 2 is similarly confusing with slight positive correlations between online usage time and helpfulness and self-willingness and a slight negative correlation with judging trustworthiness. Informational label three has similarly conflicting trends with the helpfulness, annoyingness and judging trustworthiness trends suggesting one pattern while both willingness metrics suggest a conflicting one. Again, though, these differences are somewhat marginal due to the slight and partial nature of these trends.

Warning labels 1 and 2 had more consistent trends. In cases where trends were present, label support had a negative correlation with increased online usage. Warning label 3 (the blocking label) returned to conflicting trends, with one of the metrics having a limited positive correlation, two having limited negative correlations and two not exhibiting a trend.

Finally, the supplemental information label showed consistency. All five metrics had trends that aligned.

Looking at the uncertainty levels, most of the labels (except for informational label 1, warning label 3 and the supplemental information label) showed conflicting trends of growing or reducing uncertainty with increased online usage. Note that, for uncertainty, positive for annoyingness was not seen to be at odds with other positive associations, as the uncertainty growth (or decline) would have a similar meaning for all five metrics. Focus now moves on to comparing the different levels of respondents' and respondents' perceptions of others' willingness to use the different labels. Tables 4–6 present these data in terms of income level, political affiliation and online usage levels.

Table 4. Comparison of self and others' willingness to use labels by income level.

	USD 24,999 or Less	USD 25,000–49,999	USD 50,000–74,999	USD 75,000–99,999	USD 100,000–124,999	USD 125,000 or Higher
Informational Label 1						
Self-Willingness	67.9%	62.8%	61.7%	67.6%	56.7%	64.1%
Others' Willingness	67.9%	73.9%	59.2%	71.4%	66.7%	63.3%
Informational Label 2						
Self-Willingness	70.7%	63.4%	59.1%	69.6%	64.5%	70.0%
Others' Willingness	74.0%	71.2%	63.0%	70.2%	86.4%	66.7%
Informational Label 3						
Self-Willingness	65.1%	67.5%	60.8%	68.2%	71.9%	64.1%
Others' Willingness	63.2%	74.0%	52.8%	72.4%	79.2%	61.3%
Warning Label 1						
Self-Willingness	75.3%	74.4%	72.9%	75.0%	71.9%	72.5%
Others' Willingness	73.4%	74.7%	67.9%	80.3%	77.8%	75.7%
Warning Label 2						
Self-Willingness	69.5%	70.6%	65.6%	73.6%	87.1%	70.0%
Others' Willingness	64.4%	78.6%	61.8%	73.2%	84.6%	69.7%
Warning Label 3 (Blocking)						
Self-Willingness	67.1%	65.5%	61.6%	67.6%	54.5%	70.0%
Others' Willingness	63.2%	63.5%	59.4%	72.9%	69.2%	63.9%
Supplemental Information Label						
Self-Willingness	65.9%	65.0%	50.0%	54.5%	65.6%	66.7%
Others' Willingness	61.3%	72.0%	53.9%	56.6%	70.8%	69.7%

Table 4 presents respondents' and respondents' perceptions of others' willingness to use the labels based on respondents' income level. Based on the data presented, warning label 1 was a clear favorite across income levels, with four of the five income levels indicating the highest respondents' wiliness to use this label and three of the five levels having the highest level of others' willingness for this label as well. Notably, the three income levels with respondents indicating that others would prefer an alternate label and the one with respondents themselves indicating that they would prefer an alternate label, use only two other labels: informational label 2 and warning label 2.

While these labels perform the best for given groups, there is also a notable difference, in most cases, between the better performing labels and the underperforming ones. Informational label 1, for example, underperforms warning label 1 by at least 7%, in all cases, and over 10% in several. For the USD 100,000–124,999 income level (where informational label 2 performed the best for others' willingness), it underperformed the best performing label by 20%. Warning label 3 and the supplemental information label also appeared to underperform the best performing label significantly, in most cases (though warning label 3 only slightly underperforms for respondents' willingness for the USD 125,000 or higher group). Notably, while the worst performing labels for each group were less consistent than the best, only four of the labels had worst performing statuses, and all were worst

performing for at least two groups. In addition, no label was both a best performer for one group and simultaneously a worst performer for another group.

Given the foregoing, it appears that, in terms of respondents' willingness to use a label, there are clearly preferred labels to select. Despite this, it may still be desirable to support multiple labels to maximize the number of individuals who are willing to use a label (as some may not be willing to use the generally preferred labels).

Table 5. Comparison of self- and others' willingness to use labels by party affiliation.

	Republican Party	Democratic Party	Independent/Other Party
Informational Label 1			
Self-Willingness	51.8%	76.2%	62.8%
Others' Willingness	59.4%	75.2%	64.3%
Informational Label 2			
Self-Willingness	58.8%	75.2%	62.6%
Others' Willingness	64.0%	79.5%	66.7%
Informational Label 3			
Self-Willingness	60.7%	72.2%	63.1%
Others' Willingness	64.4%	71.2%	59.7%
Warning Label 1			
Self-Willingness	70.2%	76.3%	74.5%
Others' Willingness	73.5%	71.8%	75.3%
Warning Label 2			
Self-Willingness	70.2%	74.8%	67.9%
Others' Willingness	66.3%	73.0%	69.3%
Warning Label 3 (Blocking)			
Self-Willingness	64.1%	68.5%	62.2%
Others' Willingness	60.6%	64.7%	66.2%
Supplemental Information Label			
Self-Willingness	63.2%	67.7%	51.1%
Others' Willingness	61.8%	68.4%	57.4%

Table 5 presents respondents' and respondents' perceptions of others' willingness to use the labels based on respondents' political party affiliation. The results for the different political parties are very similar to the income level groups. Warning label 1 is a clear preference, with it being the preferred self-use label for all three groups (tied with warning label 2 for Republicans) and the preferred others' use label for two of the three groups. As with the income level groups, the other two labels that performed best for a group were informational label 2 and warning label 2.

The largest difference between the income level and political affiliation date was the performance of informational label 1 for Democrats. The label was only 0.1% less popular than the best performing label for self-willingness for use. Notably, this label performed well for Democrats and was the worst performing for both self- and others' willingness for use by Republicans. This label clearly had a demonstrable difference in political party affiliation-based perception.

Table 6 presents respondents' willingness and their perceptions of others' willingness, to use the labels based on respondents' online usage level. The results by online usage level have some key similarities to those by income level and party affiliation. Warning label 1, again, performed the best. However, it was principally preferred by those with lower levels of online usage time. The supplemental information label performed, consistently, the worst for users with less than three hours of daily internet usage. For higher levels of use respondents, though, the results were quite different. Informational label 1 was the preferred label for those with more than five hours of daily usage and warning label 2 was the preferred label for those with three to five hours of usage. Warning label 3 was the least preferred for the higher usage level respondents. Notably, the supplemental information label, which was consistently the worst for the lower-usage level respondents, was the

best performing for others for one group and just slightly (0.9% lower) less than the best performing for another group.

The data show that there was a demonstrable difference in preference between low-usage and higher-usage level respondents. The higher usage level respondents clearly did not prefer the blocking label. More research will be needed to ascertain whether commonalities between informational label 1, warning labels 1 and 2 and the supplemental information label were responsible for their superior performance for higher-usage level users or if the differences in preference were indicative of true differences in preference between the higher-usage level groups.

Table 6. Comparison of self- and others' willingness to use labels by online usage level.

	1 h or Less	Between 1 and 3 h	Between 3 and 5 h	More than 5 h
Informational Label 1				
Self-Willingness	58.1%	64.5%	69.1%	80.0%
Others' Willingness	64.6%	69.0%	62.0%	72.0%
Informational Label 2				
Self-Willingness	60.8%	67.6%	66.7%	71.0%
Others' Willingness	68.9%	71.1%	69.1%	70.4%
Informational Label 3				
Self-Willingness	62.7%	66.3%	63.8%	73.3%
Others' Willingness	59.8%	69.6%	57.7%	71.4%
Warning Label 1				
Self-Willingness	73.1%	75.3%	70.5%	75.9%
Others' Willingness	71.9%	77.4%	66.7%	75.9%
Warning Label 2				
Self-Willingness	65.6%	71.7%	78.0%	75.0%
Others' Willingness	64.8%	71.8%	72.7%	74.1%
Warning Label 3 (Blocking)				
Self-Willingness	68.2%	63.8%	58.1%	64.5%
Others' Willingness	65.7%	64.7%	61.4%	60.0%
Supplemental Information Label				
Self-Willingness	52.7%	62.8%	59.6%	71.0%
Others' Willingness	54.0%	62.1%	74.1%	75.0%

Finally, focus turns to respondents' comparative perceptions of each label and whether they saw it as most useful for themselves or others. This data are presented in Tables 7–9.

Table 7 presents the data in terms of respondents' income level. There are few patterns in this data, and perhaps the most notable pattern was that the respondents' can be effectively grouped into two groups: those with incomes above and below USD 75,000. Those with incomes below USD 75,000 found informational label 2 most useful for others, consistently, and warning label 3 most useful for themselves. One of the two was a higher-performing label and one was a lower-performing one based on the data in Table 4. However, neither was the highest or lowest performing. The higher income respondents had two labels that were consistently identified as better for others than respondents themselves: warning label 1 and the supplemental informational label, and warning label 2 was identified, consistently, as best for respondents. Warning label 1 was the best performing label overall, and warning label 2 was towards the better performing end of the spectrum. The supplemental information label, conversely, was the poorest performing label.

Given the juxtaposition of the data between Tables 4 and 7, the status of a label as being preferred for respondents own use or others' use is of limited utility. However, the pattern of change at USD 75,000 is an interesting outcome.

Table 7. Do respondents see labels as more valuable to self or others, by income level.

	USD 24,999 or Less	USD 25,000–49,999	USD 50,000–74,999	USD 75,000–99,999	USD 100,000–124,999	USD 125,000 or More
Informational Label 1	Others	Others	Self	Others	Others	Self
Informational Label 2	Others	Others	Others	Others	Others	Self
Informational Label 3	Self	Others	Self	Others	Others	Self
Warning Label 1	Self	Others	Self	Others	Others	Others
Warning Label 2	Self	Others	Self	Self	Self	Self
Warning Label 3 (Blocking)	Self	Self	Self	Others	Others	Self
Supplemental Information Label	Tie	Others	Others	Others	Others	Others

Table 8. Whether respondents saw labels as more valuable to self or others by party affiliation.

	Republican Party	Democratic Party	Independent/Other Party
Informational Label 1	Others	Self	Others
Informational Label 2	Others	Others	Others
Informational Label 3	Others	Self	Self
Warning Label 1	Others	Self	Others
Warning Label 2	Self	Self	Others
Warning Label 3 (Blocking)	Self	Self	Others
Supplemental Informational Label	Self	Others	Others

Table 9. Do respondents see labels as more valuable to self or others, by online usage level.

	1 h or Less	Between 1 and 3 h	Between 3 and 5 h	More than 5 h
Informational Label 1	Others	Others	Self	Self
Informational Label 2	Others	Others	Others	Self
Informational Label 3	Self	Others	Self	Self
Warning Label 1	Self	Others	Self	Others
Warning Label 2	Self	Others	Self	Self
Warning Label 3 (Blocking)	Self	Others	Others	Self
Supplemental Information Label	Others	Self	Others	Others

Table 8 presents the self-versus-others' preference data in terms of respondents' party affiliation. A few interesting patterns are present in this data. Republicans tended to see informational labels as more useful for others, while seeing the warning labels (two of the three) as most useful for themselves. Additionally, they were the only group that found the supplemental information label more useful for their own use. Democrats, on the other hand, seemed to find most labels more useful for their own use. They only identified one informational label and the supplemental information label as being more useful for others. Independent/other party affiliates, on the other hand, had nearly the opposite perspective. They identified six of the seven label types (all except informational label 3) as being more useful for others. This difference in perception may be important when considering how to introduce labels and be indicative of differences in willingness to learn about labels and participate by party affiliation.

Finally, Table 9 presents the self-versus-others' preference data in terms of respondents' online usage level. The patterns in this data are less pronounced than the party affiliation date. Those with the lowest online usage time find the informational labels (excepting number 3) most useful to others and the warning labels most useful to themselves. This may be indicative of the group feeling that they need more explicit guidance due to their lower familiarity with and exposure to the Internet. Those with one to three hours of online usage per day found all labels, except the supplemental information label, to be more useful to others. This group was the only group to find the supplemental information label more useful to themselves than others. An explanation for this could be the group believing that they have enough exposure and experience with internet usage to make their own decisions but requiring the additional support of the extended information to do so.

The higher usage level respondents' responses are more inconsistent. There are three labels that both groups find to be more useful for their own use and one that they find to be more useful to others. Notably, there is a strong transition back to finding labels more useful for self-use between the one to three and 3–5 groups which persists with the more than 5 h group. One interpretation of this is that the lower usage level respondents found the labels useful, as they need the support of them for decision making (and, thus, focus on the warning labels), while the higher usage level respondents see the value in some of the labels for their own use, though there is not a clear label type of preference notable in this data.

8. Conclusions and Future Work

Online content labels are an approach to protecting individuals against the harmful impacts of intentionally deceptive online content without censorship of content creators. They may help prevent the spread of online misinformation and may even increase users' awareness of problematic content through their ongoing use.

To advance the potential future use of online content labeling, this paper presented and analyzed the results and implications of a national survey of consumer perceptions regarding online content labels to ascertain their efficacy. Data were collected from over 500 respondents and analyzed in terms of their key demographics including income level, political party affiliation and online usage level.

The analysis of this data demonstrated a great deal of support for labels, both in general and across various demographic groups. It also demonstrated that some groups had clear preferences for and against certain types of labels. Individuals were also shown to be more or less likely to utilize labels based on key demographic characteristics.

Labels were generally shown to be less well received with increasing income levels. Respondents also indicated greater certainty about their usage decision with higher income levels. The labels that were the best received for those with higher income levels were the ones that made more information available.

Democrats were shown to have a stronger willingness to use labels, in general, and to believe that others would use labels more than Republicans and Independents/other party members. This was true overall and across virtually all label types. However, the first and third warning labels were better received by Independents/other party members than Democrats.

The trends were less clear with regards to online usage levels. The principal trend that was detected, though, for two warning labels was a decline in use with increased online usage.

When looking at preferences, warning label 1 was a strong preference across all income levels for both individuals' own use and use by others. Informational label 2 and warning label 2, though not performing as well as warning label 1, were also well received across income levels. Warning label 1 was also the most popular for use across political affiliations for individuals and others.

Unlike the consistency shown across income levels and political affiliations, differences in a label preference were clear by online usage level. The first warning label was preferred by lower usage groups. However, the label of choice varied among the higher usage groups. The higher usage level groups demonstrated a trend towards being more receptive of labels with extended information on them.

Finally, while individuals' indication of greater usage preference for each level, for both themselves and others, varied by income level and online usage level, notable differences are present between political affiliation groups. Democrats indicated greater preference for all but two of the labels for self-use, while Republicans and Independents/other party members indicated greater preference for the use of labels by others, in most cases.

Knowledge about the demographic groups' preferences and trends can be useful in a number of ways. At the most basic level, computer system operators, users themselves and others can pick the labeling that is most preferred based on the demographics of typical

system users. This could also be customized on a per-user basis, informed by the user's demographic characteristics.

Demographic preferences also provide a potential starting point for additional user-specific (or group-specific) customization. They may also provide a starting point for further research into the more specific preferences of particular groups and the development of new and modified types of labels.

The data can also inform decisions regarding the training of groups of user and individual users, allowing focus to be placed on areas that demographic preferences indicate are important. For example, some groups have shown a preference for warning style labels while others have shown a preference for labels with more information (presumably to facilitate the individual making their own informed decision. This type of a preference not only suggests what labels should be presented and suggested to a group of users, but it also has implications regarding what features and capabilities of a labeling system particular demographic groups find most valuable and, thus, how labeling should be presented overall and how training should be presented and focused.

This same knowledge is, thus, inherently useful for product development and feature decision making. Labels that perform well across several groups can be prioritized for implementation over those that are not as well received or as broadly supported.

Appendix A

This section presents supporting figures for Section 4. Each figure includes error bars showing the standard error range for the threshold values. Figures A1–A15 present data for the article summary information label (shown in Figure 3). Figures A16–A30 present data for the informational label without article summary (shown in Figure 4). Finally, Figures A31–A45 present data for the informational label with article summary and component score descriptions (shown in Figure 5).
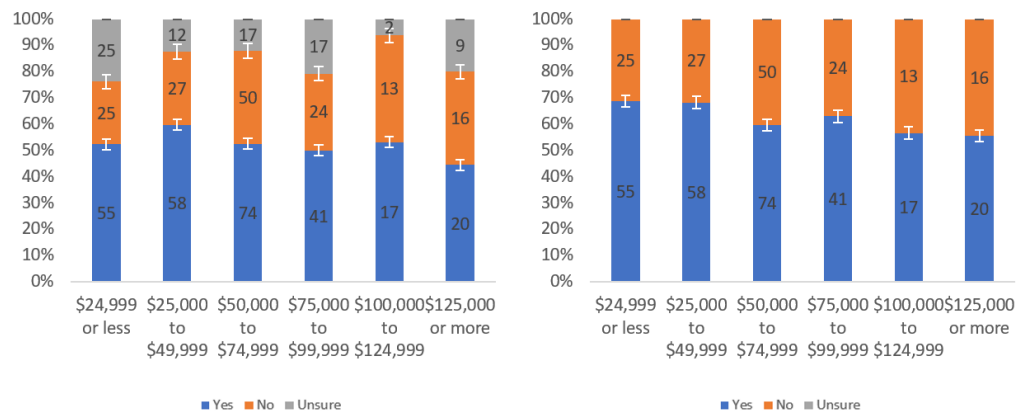
Figure A1. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.
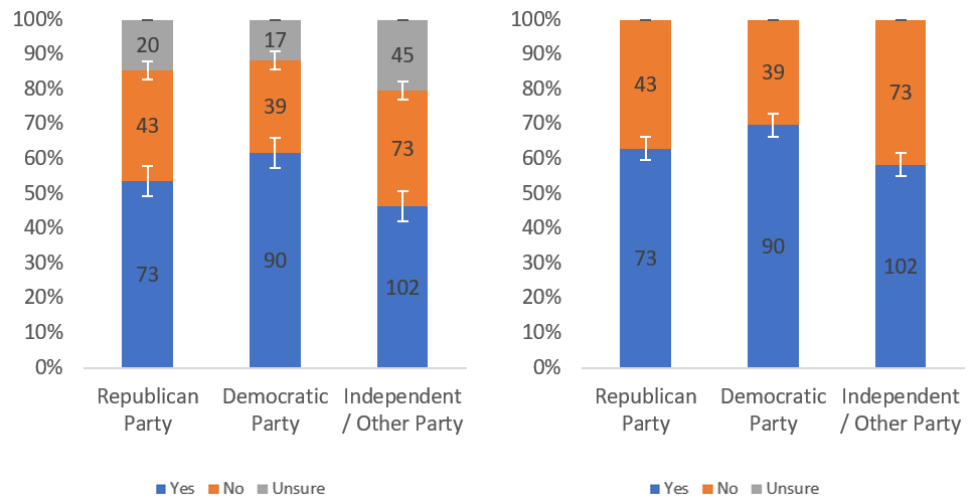
Figure A2. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
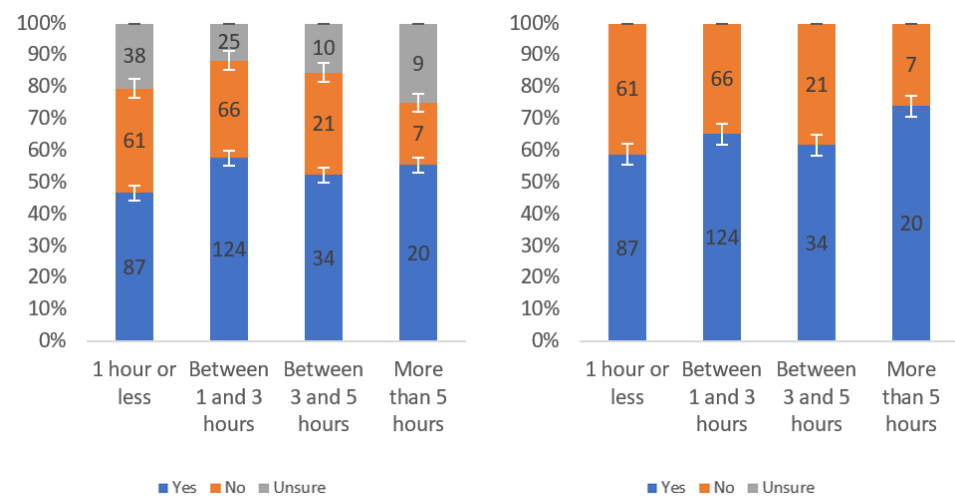
Figure A3. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.
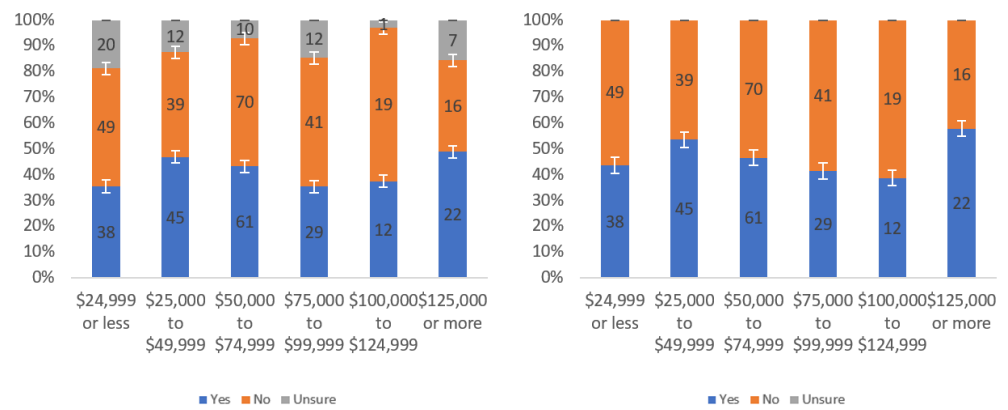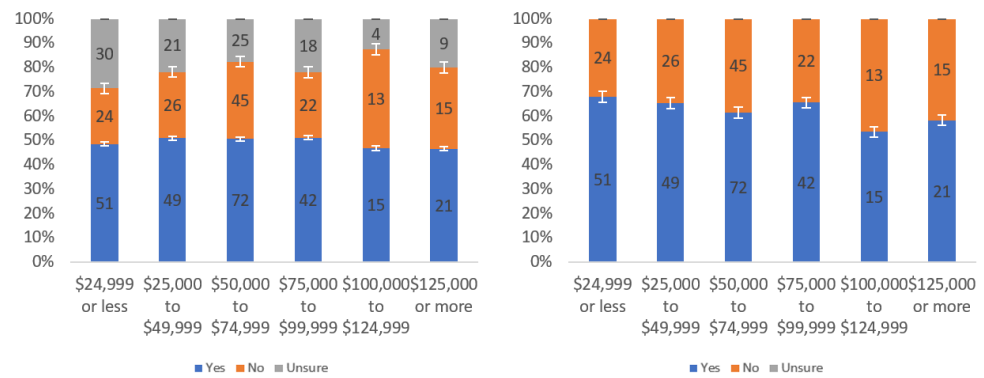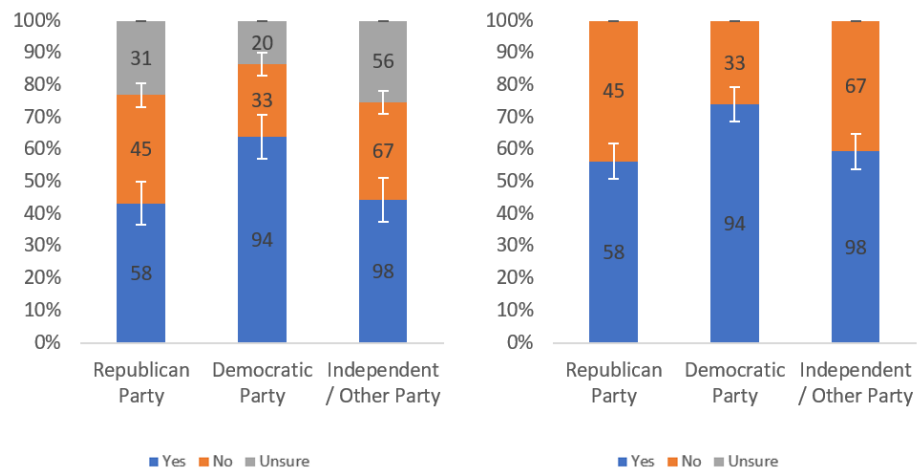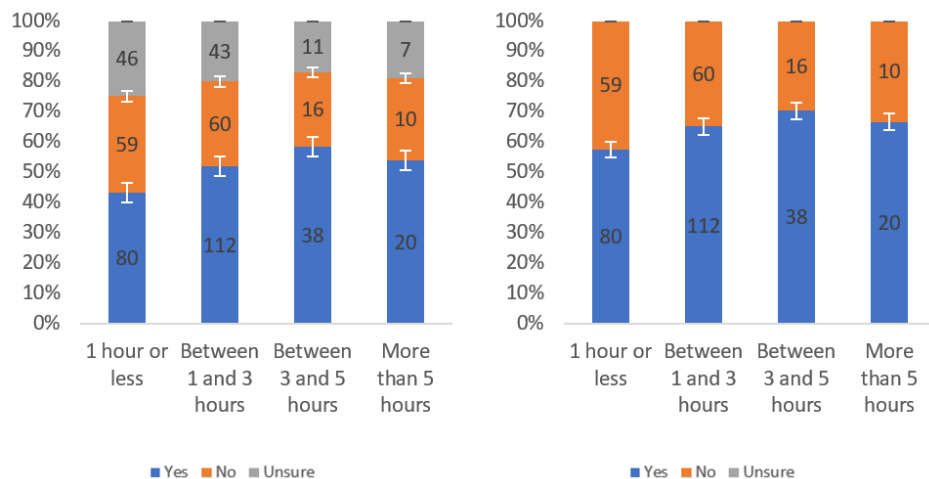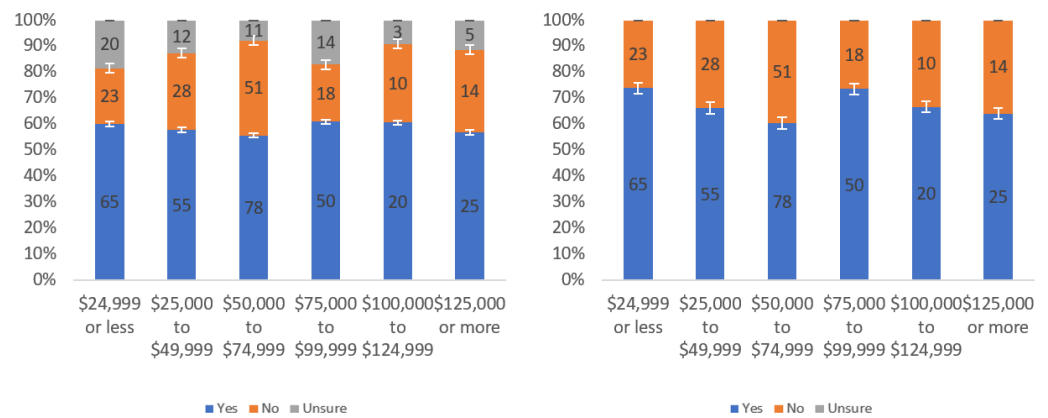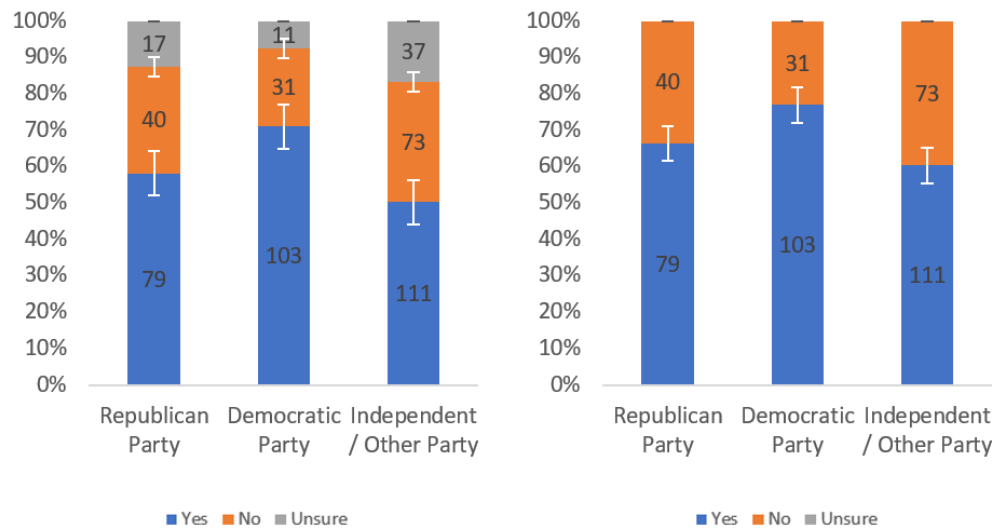
Figure A4. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.
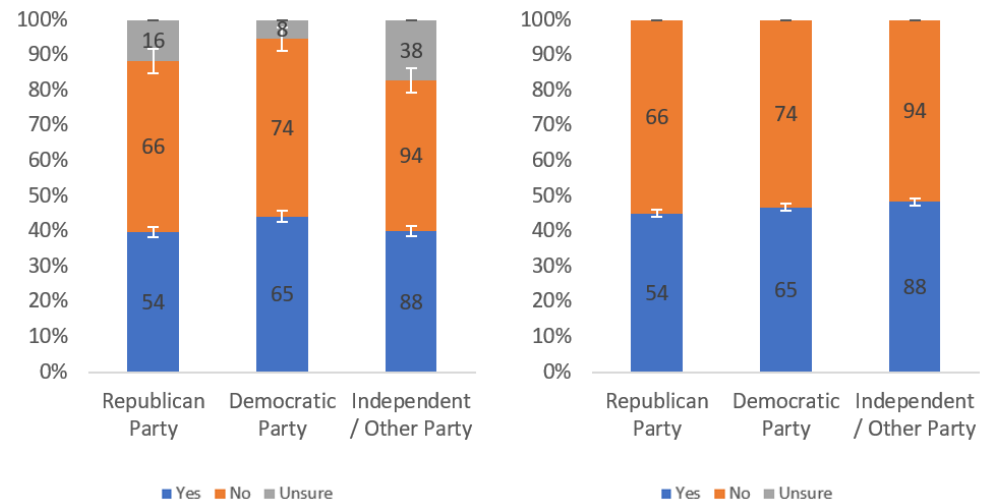
Figure A5. Responses regarding label annoyingness by party affiliation including (**left**) and excluding (**right**) unsure responses.
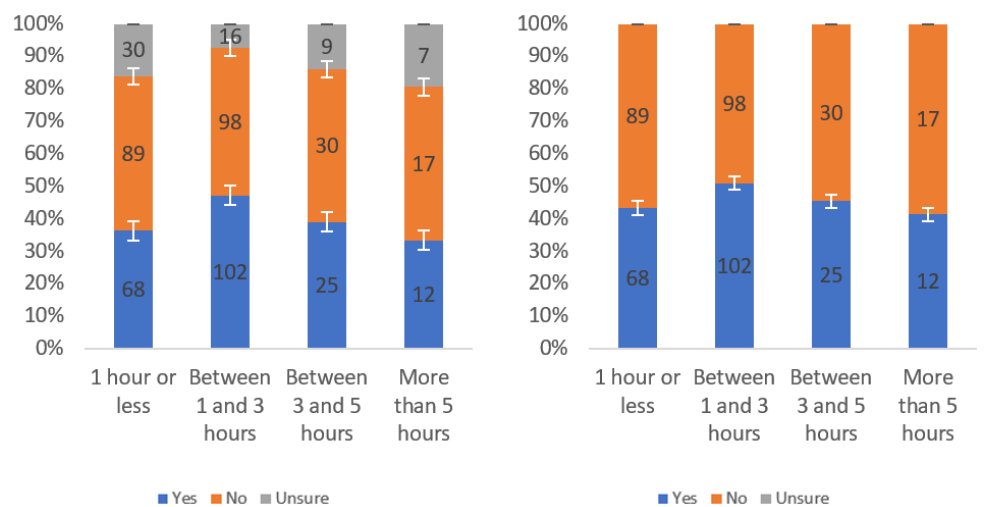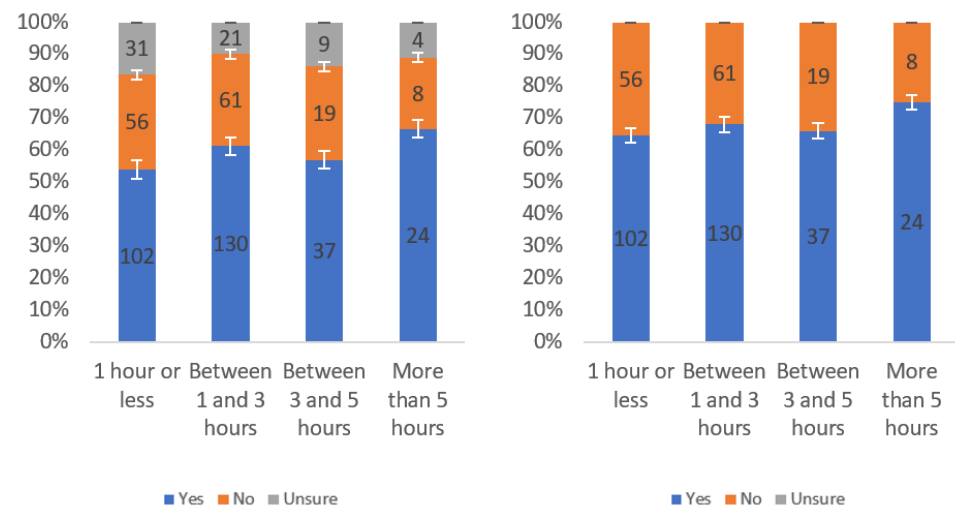

Figure A6. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.
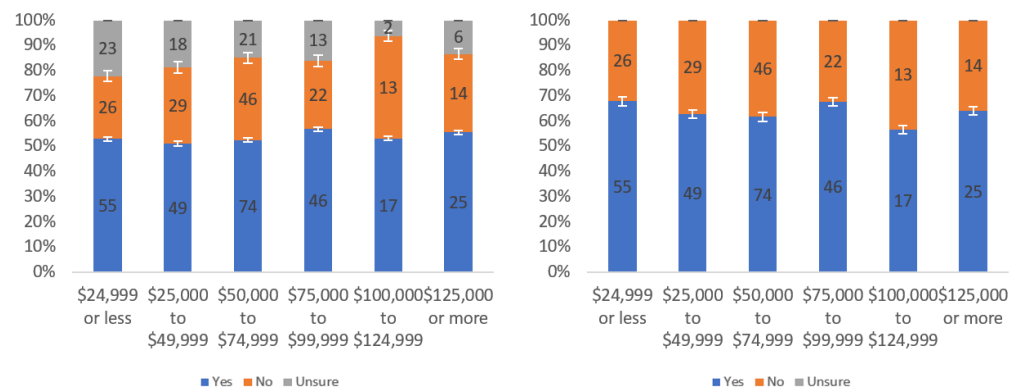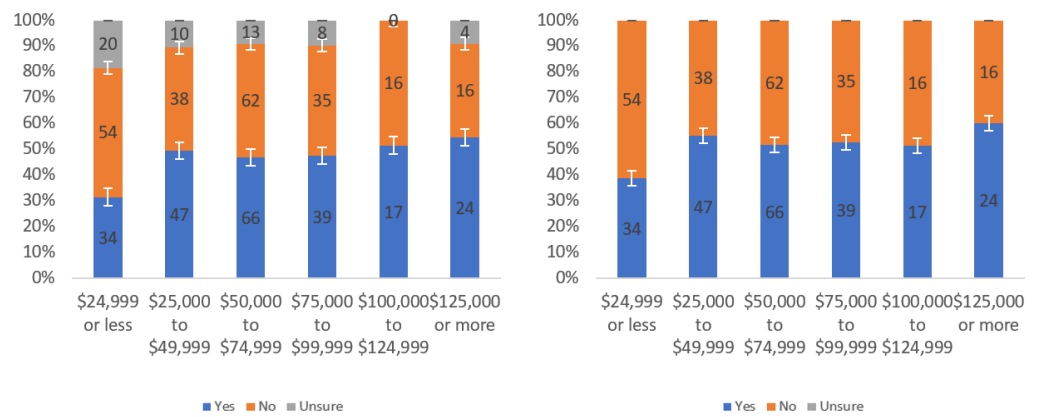
Figure A7. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
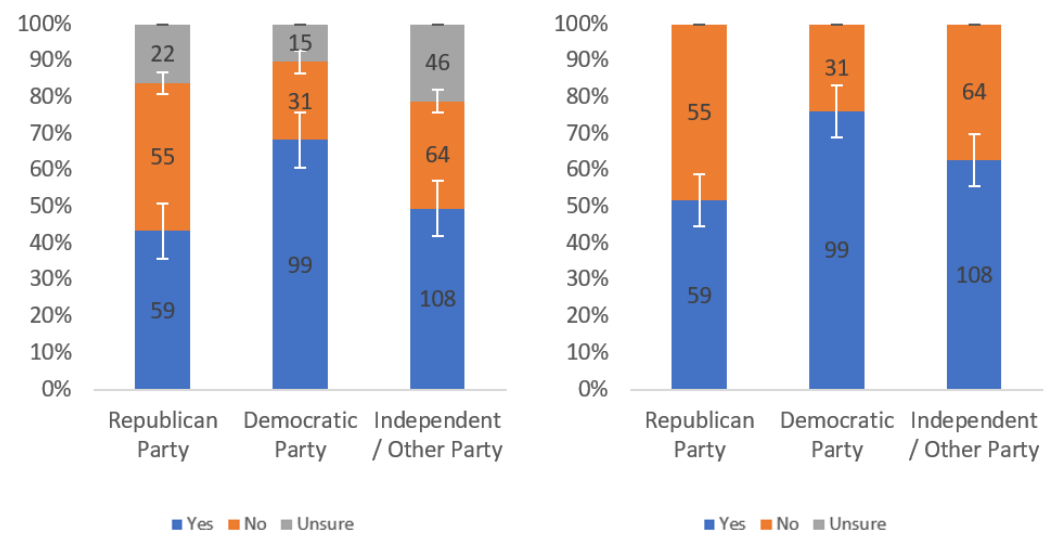
Figure A8. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
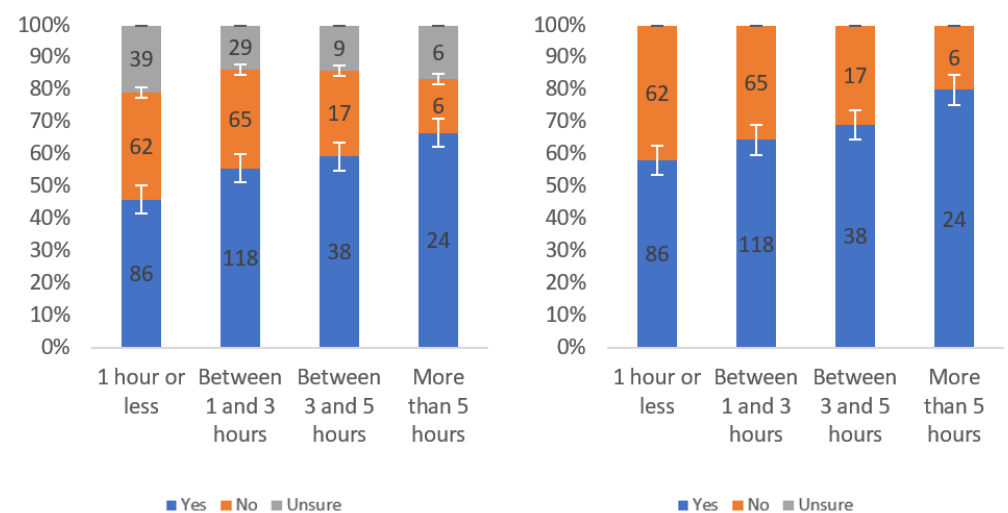
Figure A9. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.
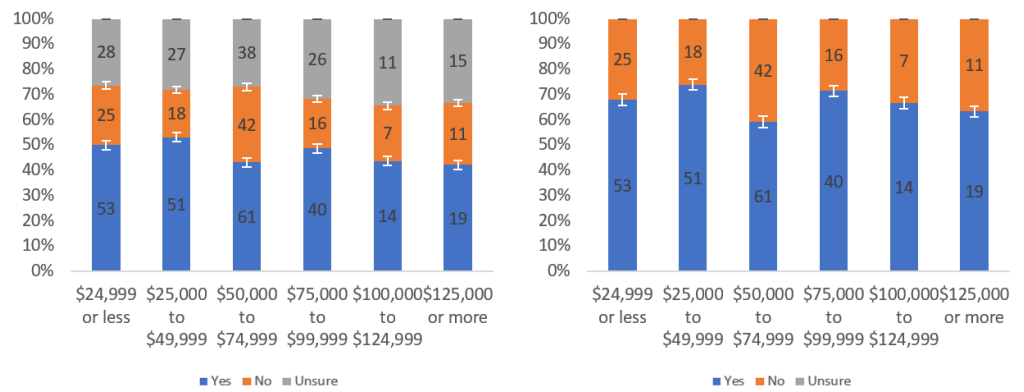
Figure A10. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
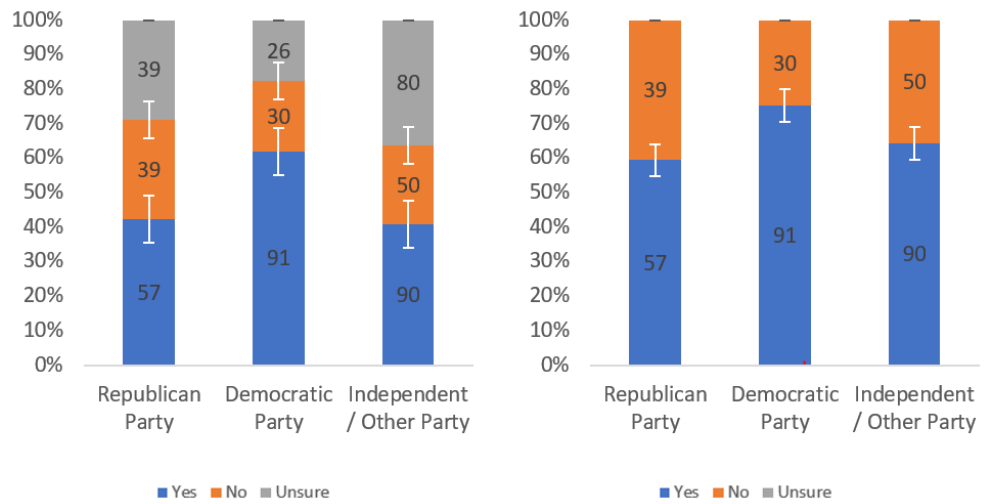
Figure A11. Responses regarding others' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
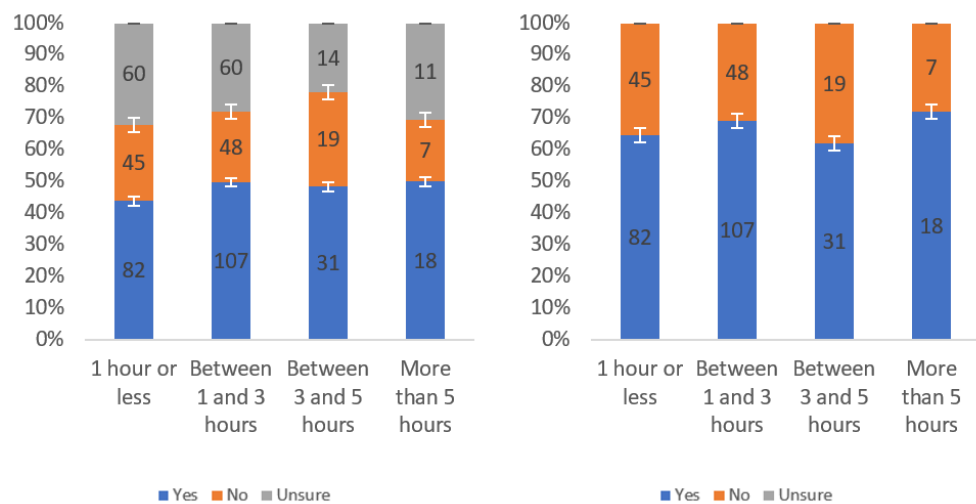
Figure A12. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A13. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.

Figure A14. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.

Figure A15. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A16. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.

Figure A17. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
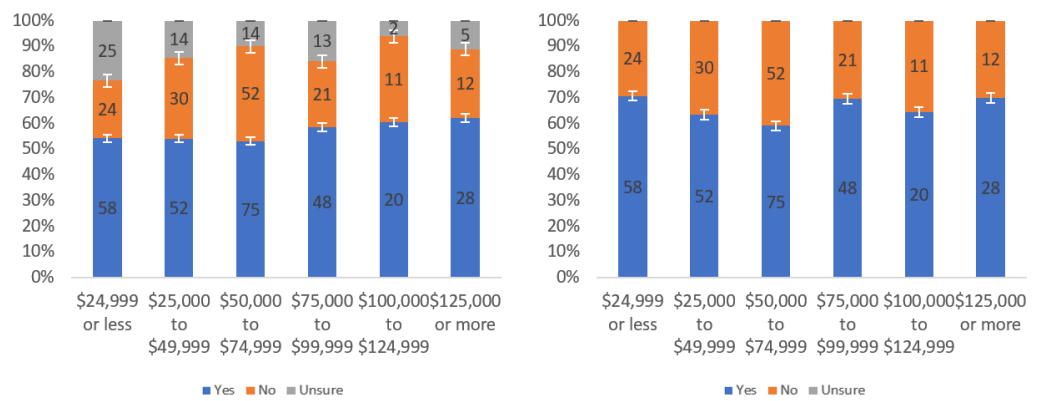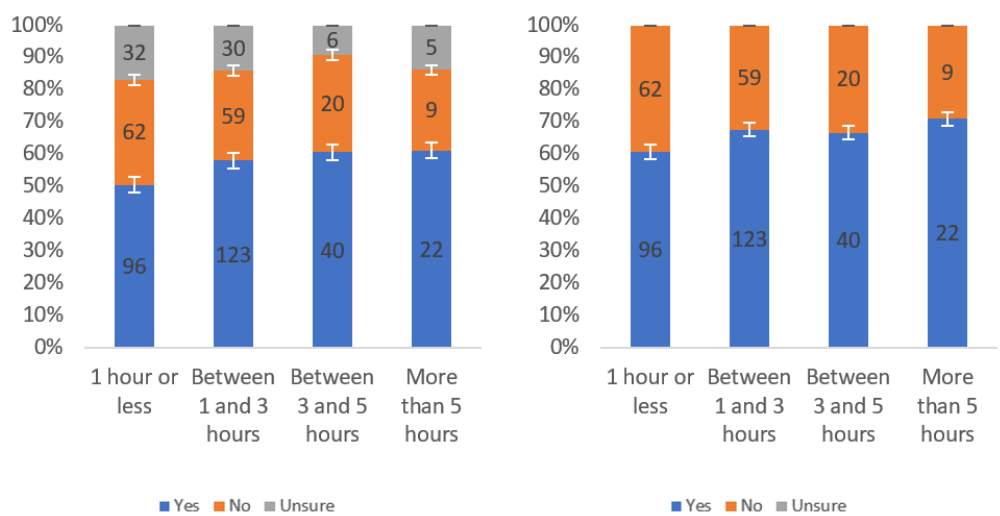

Figure A18. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A19. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.
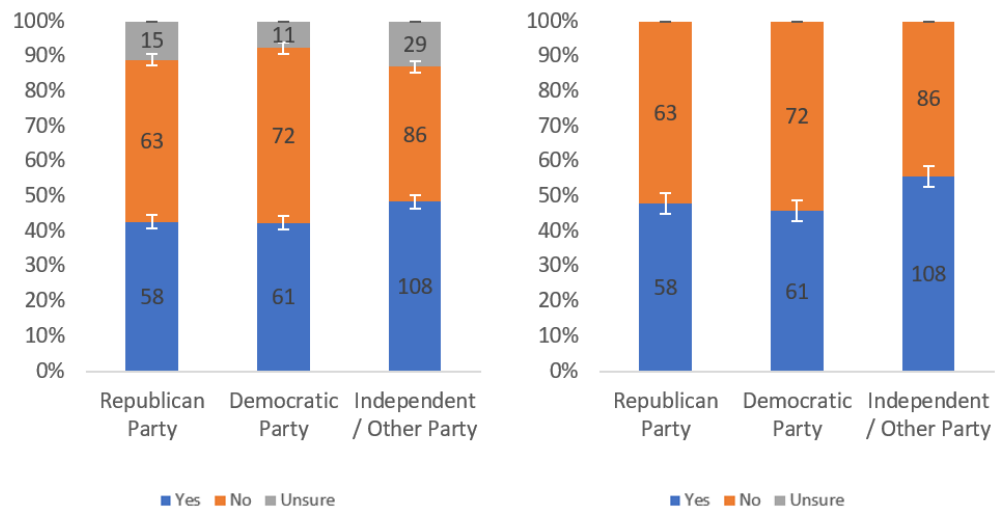
Figure A20. Responses regarding label annoyingness by party affiliation including (**left**) and excluding (**right**) unsure responses.
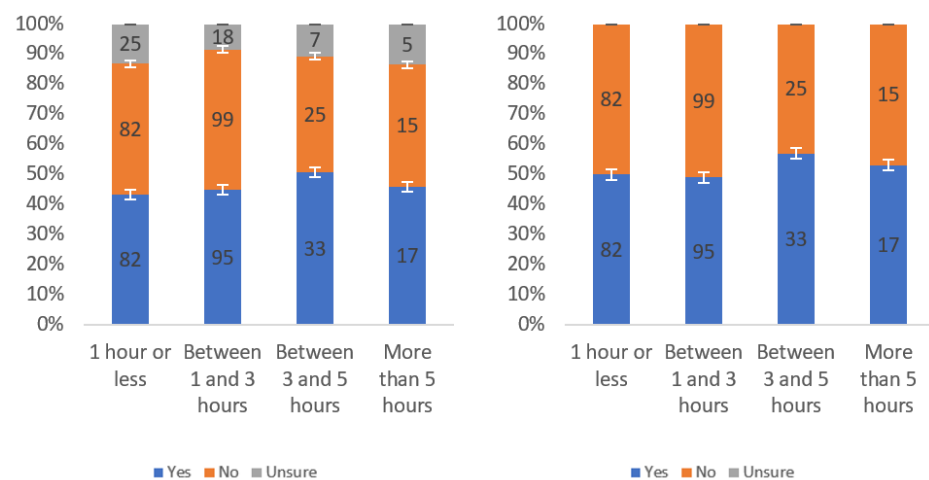
Figure A21. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A22. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
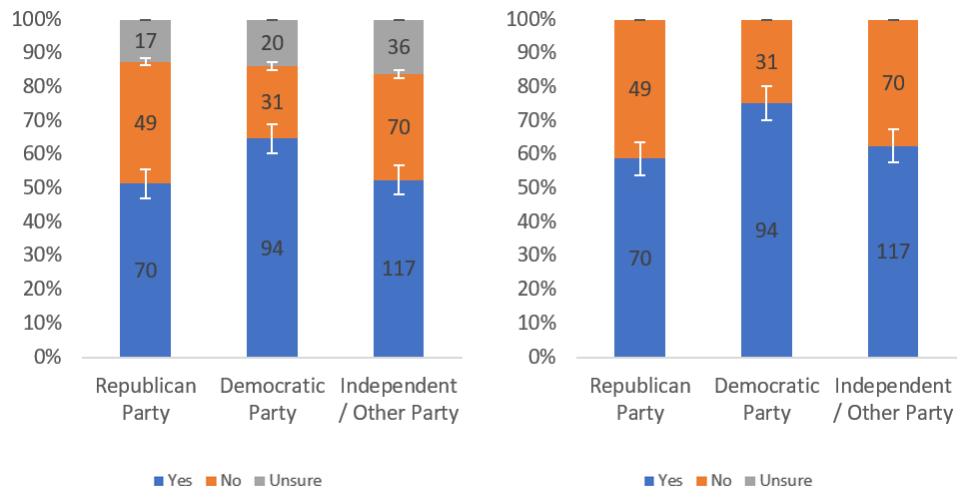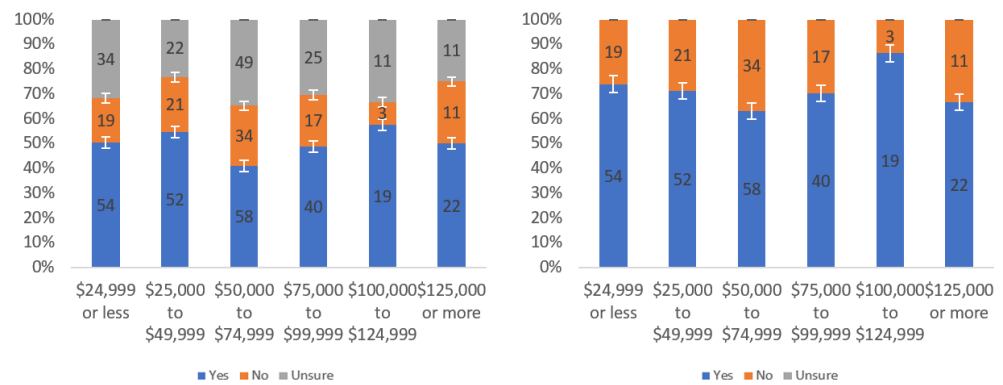
Figure A23. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.

Figure A24. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A25. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.
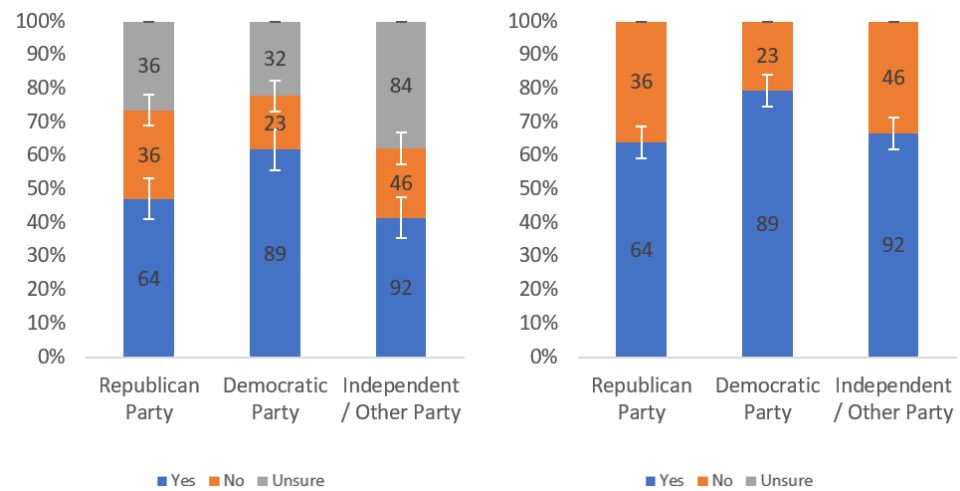

Figure A26. Responses regarding others' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
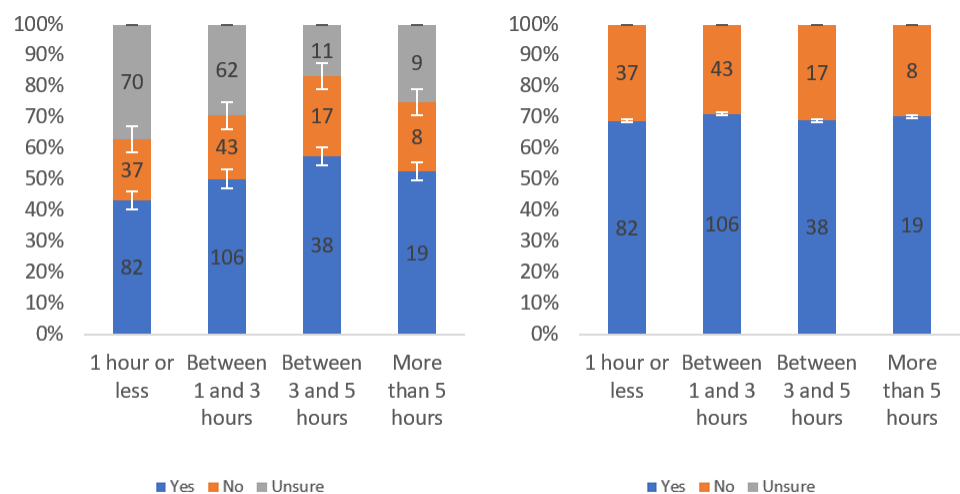

Figure A27. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.
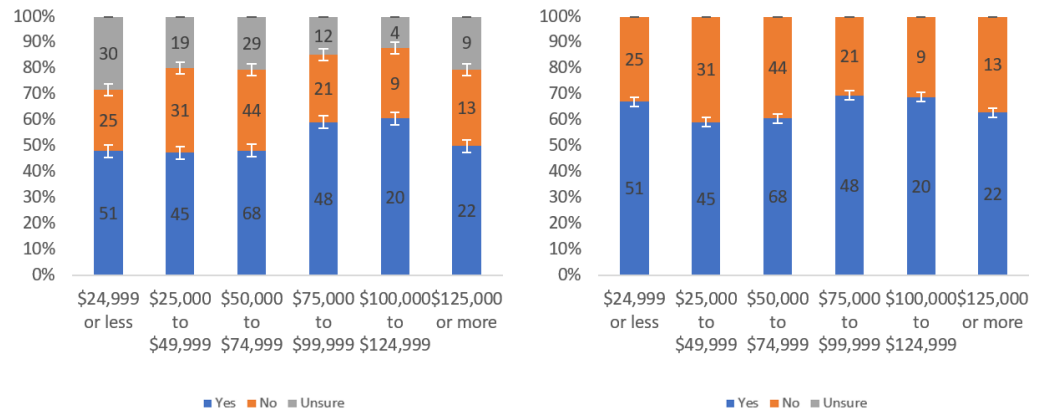
Figure A28. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.
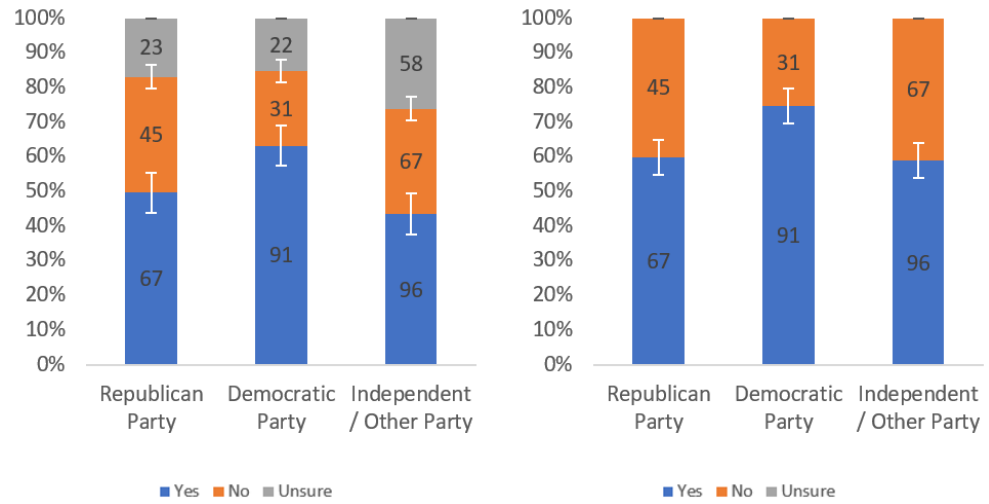

Figure A29. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
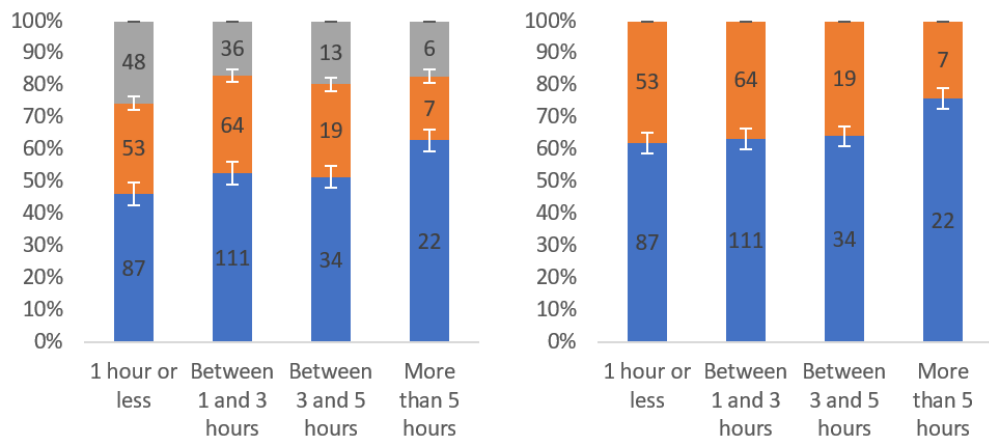

Figure A30. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.
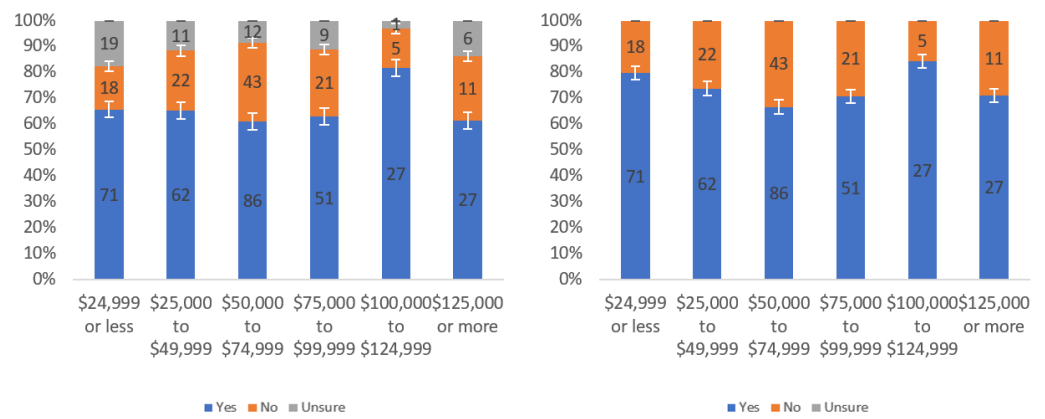
Figure A31. Responses regarding label helpfulness by income level including (**left**) and excluding (**right**) unsure responses.
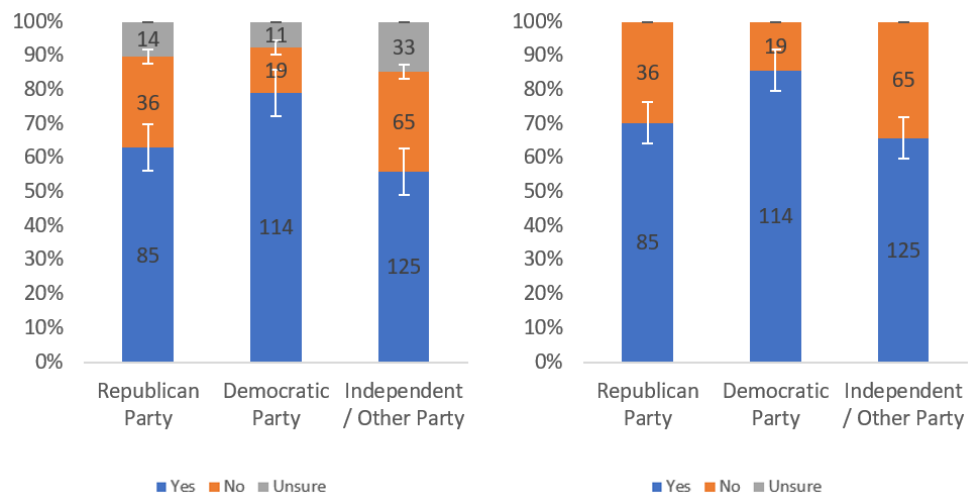

Figure A32. Responses regarding label helpfulness by party affiliation including (**left**) and excluding (**right**) unsure responses.
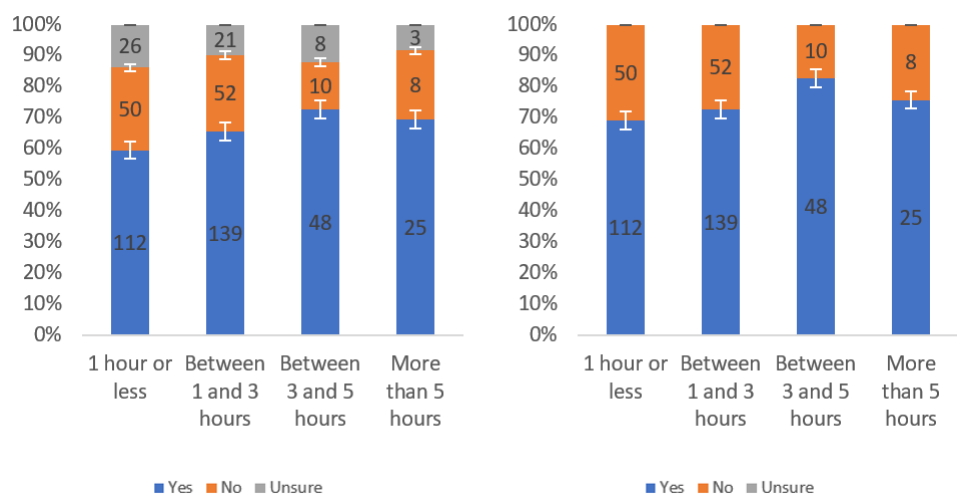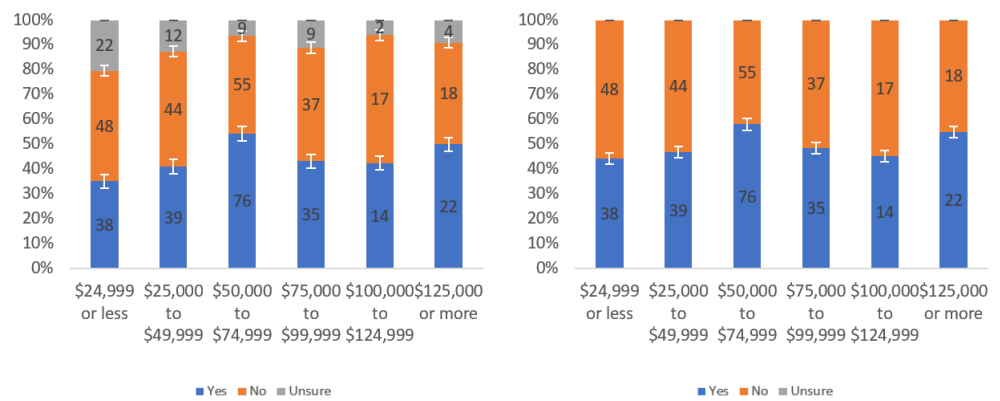

Figure A33. Responses regarding label helpfulness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A34. Responses regarding label annoyingness by income level including (**left**) and excluding (**right**) unsure responses.

Figure A35. Responses regarding label annoyingness by party affiliation including (**left**) and excluding (**right**) unsure responses.
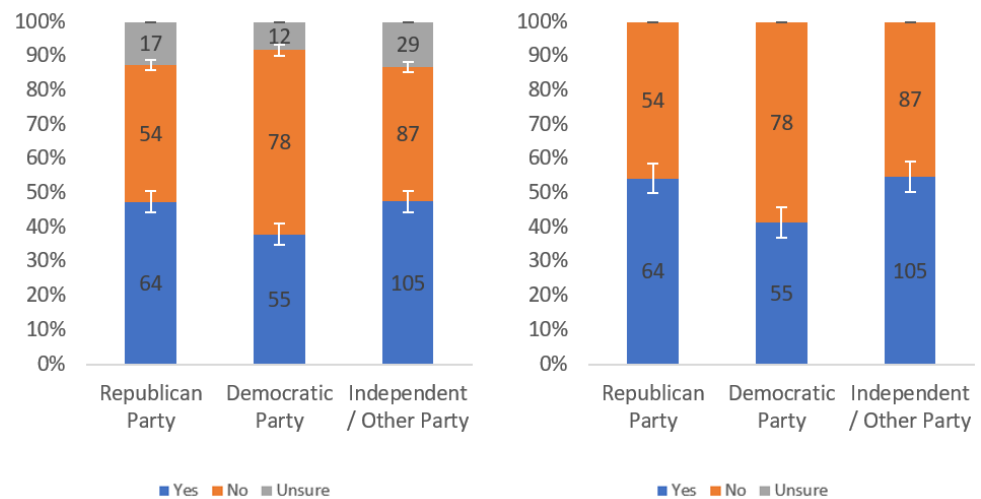

Figure A36. Responses regarding label annoyingness by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A37. Responses regarding respondents' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.

Figure A38. Responses regarding respondents' willingness to review by party affiliation including (**left**) and excluding (**right**) unsure responses.
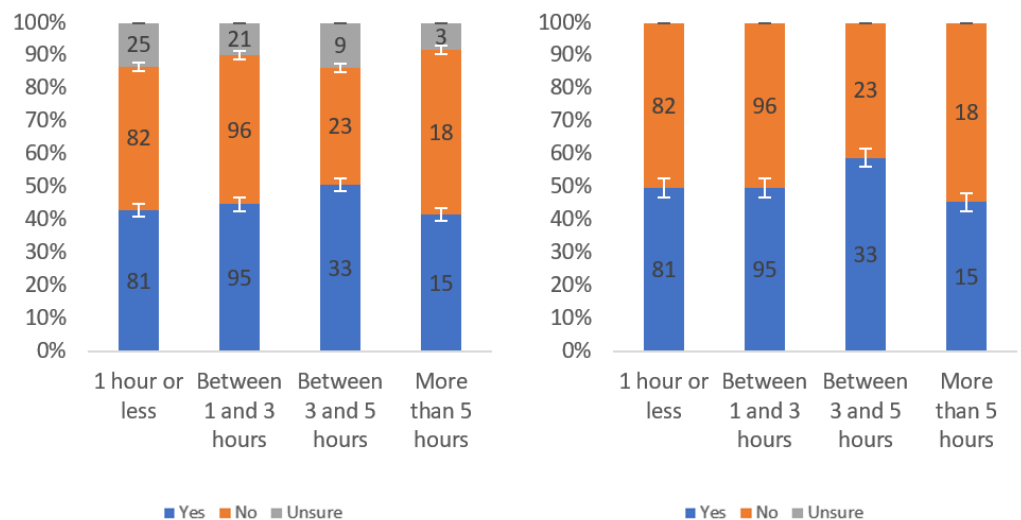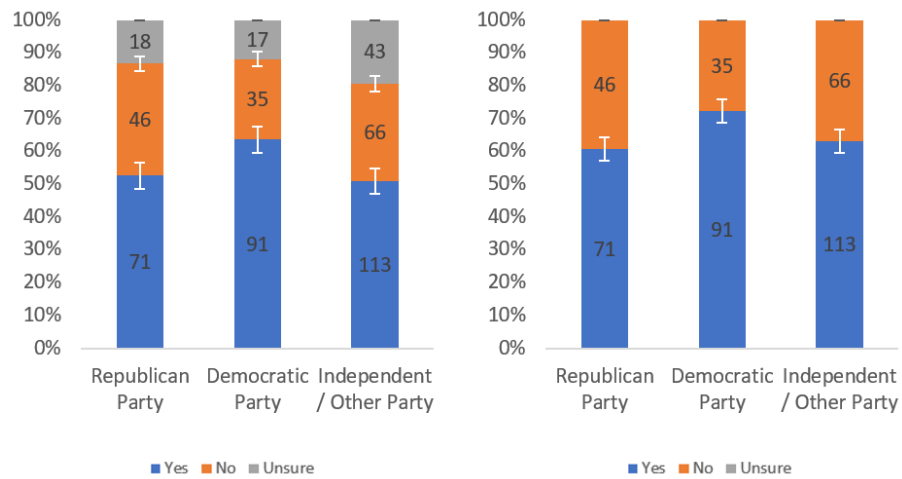

Figure A39. Responses regarding respondents' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.

Figure A40. Responses regarding others' willingness to review by income level including (**left**) and excluding (**right**) unsure responses.

Figure A41. Responses regarding others' willingness to review by political affiliation including (**left**) and excluding (**right**) unsure responses.
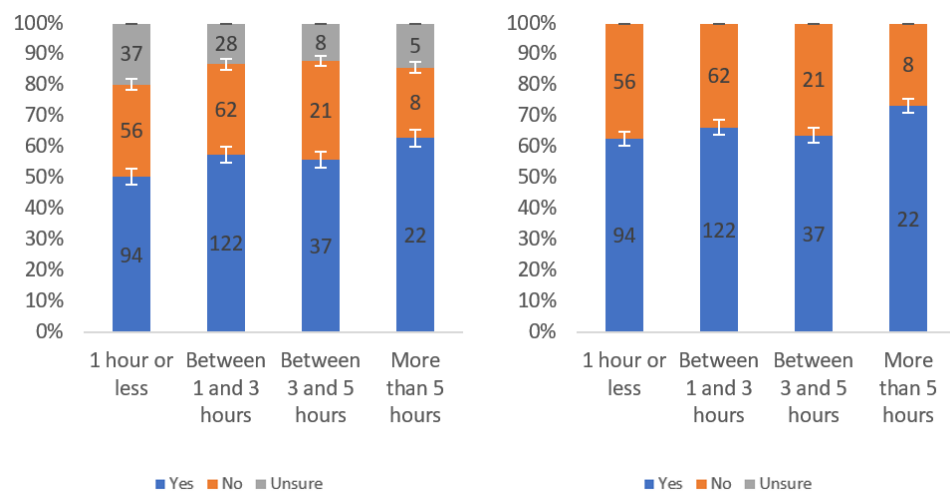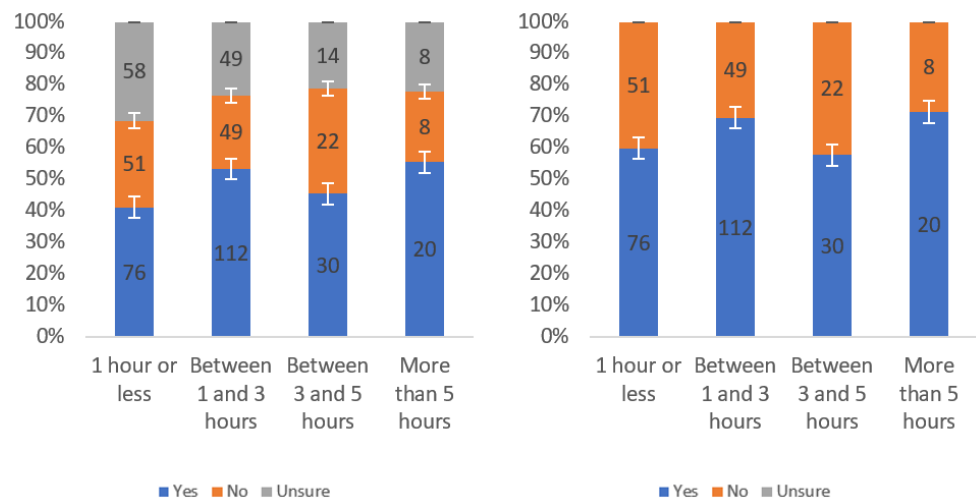

Figure A42. Responses regarding others' willingness to review by internet usage level including (**left**) and excluding (**right**) unsure responses.
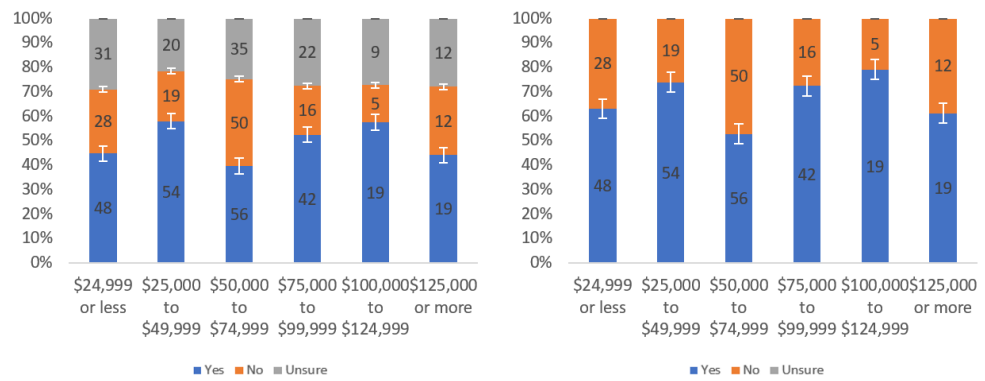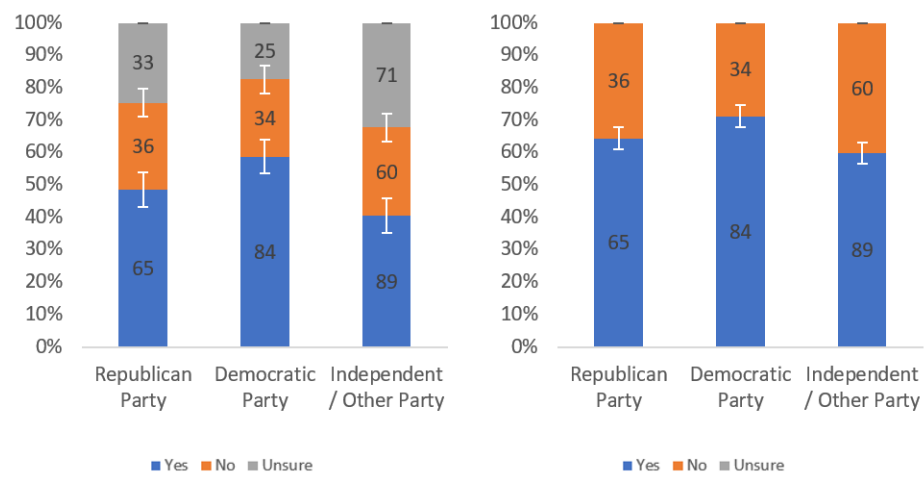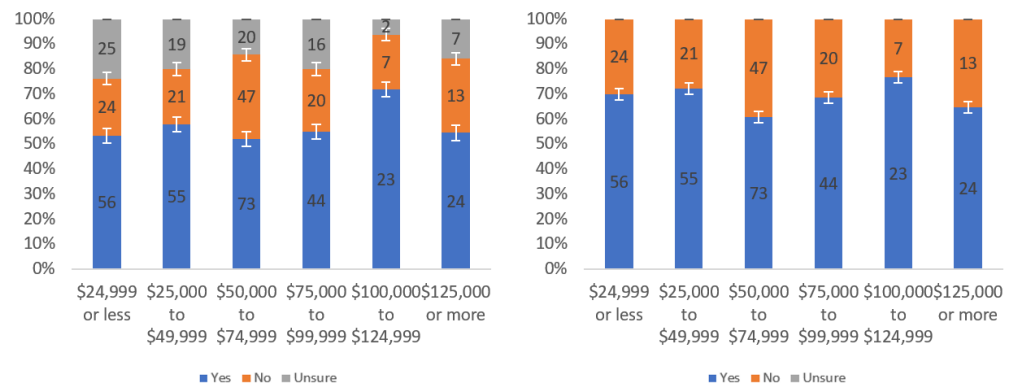
Figure A43. Responses regarding usefulness in judging trustworthiness by income level including (**left**) and excluding (**right**) unsure responses.
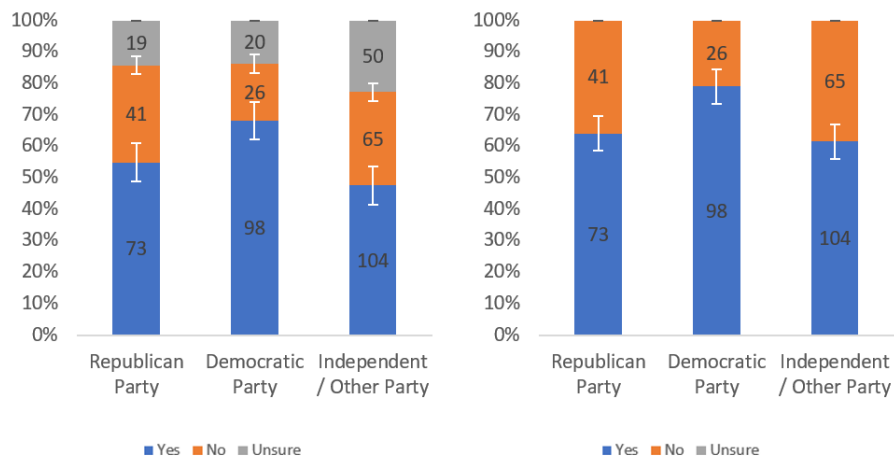
Figure A44. Responses regarding usefulness in judging trustworthiness by party affiliation including (**left**) and excluding (**right**) unsure responses.
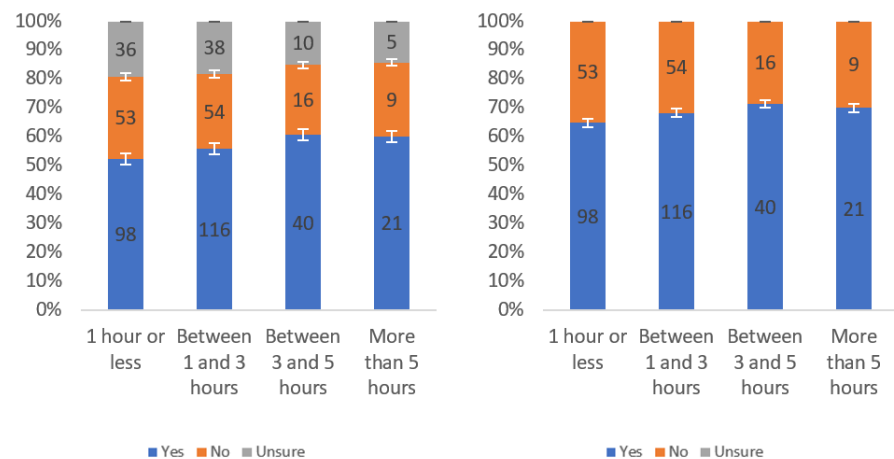
Figure A45. Responses regarding usefulness in judging trustworthiness by internet usage level including (**left**) and excluding (**right**) unsure responses.

References

1. Cunha, E.; Magno, G.; Caetano, J.; Teixeira, D.; Almeida, V. Fake News as We Feel It: Perception and Conceptualization of the Term "Fake News" in the Media. In *Social Informatics*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11185, pp. 151–166. [CrossRef]
2. Tong, C.; Gill, H.; Li, J.; Valenzuela, S.; Rojas, H. "Fake News Is Anything They Say!"—Conceptualization and Weaponization of Fake News among the American Public. *Mass Commun. Soc.* **2020**, *23*, 755–778. [CrossRef]

3. Bastos, M.T.; Mercea, D. The Brexit Botnet and User-Generated Hyperpartisan News. *Soc. Sci. Comput. Rev.* **2017**, *37*, 38–54. [CrossRef]

4. McGaughey, E. Could Brexit be Void? *King's Law J.* **2018**, *29*, 331–343. [CrossRef]

5. Lee, T. The global rise of "fake news" and the threat to democratic elections in the USA. *Public Adm. Policy* **2019**, *22*, 15–24. [CrossRef]

6. Griffiths, J. *The Great Firewall of China: How to Build and Control an Alternative Version of the Internet*, 2nd ed.; Zed Books: London, UK, 2021.

7. Silva, R.M.; Santos, R.L.S.; Almeida, T.A.; Pardo, T.A.S. Towards automatically filtering fake news in Portuguese. *Expert Syst. Appl.* **2020**, *146*, 113199. [CrossRef]

8. Collins, B.; Hoang, D.T.; Nguyen, N.T.; Hwang, D. Trends in combating fake news on social media—A survey. *J. Inf. Telecommun.* **2020**, *5*, 247–266. [CrossRef]

9. Spradling, M.; Straub, J.; Strong, J. Protection from 'Fake News': The Need for Descriptive Factual Labeling for Online Content. *Futur. Internet* **2021**, *13*, 142. [CrossRef]

10. Ott, B. Some Good News about the News: 5 Reasons Why 'Fake' News is Better than Fox 'News'—Flow. *Flow* **2005**, *2*. Available online: https://www.flowjournal.org/2005/06/news-fox-television-reception-the-daily-show-jon-stewart/ (accessed on 8 May 2022).

11. Kim, S. All the Times People Were Fooled by The Onion. Available online: https://abcnews.go.com/International/times-people-fooled-onion/story?id=31444478 (accessed on 4 February 2022).

12. Saez-Trumper, D. Fake Tweet Buster: A Webtool to Identify Users Promoting Fake News on Twitter. In Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT'14), Santiago, Chile, 1–4 September 2014.

13. Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; Lazer, D. Fake news on Twitter during the 2016 U.S. presidential election. *Science* **2019**, *363*, 374–378. [CrossRef]

14. Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The science of fake news. *Science* **2018**, *3*, 1094–1096. [CrossRef]

15. Bovet, A.; Makse, H.A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **2019**, *10*, 7. [CrossRef]

16. Shearer, E.; Matsa, K.E. News Use Across Social Media Platforms 2018. Available online: https://www.pewresearch.org/journalism/2018/09/10/news-use-across-social-media-platforms-2018/ (accessed on 21 September 2021).

17. Fatilua, J. Who trusts social media? *Comput. Hum. Behav.* **2018**, *81*, 303–315. [CrossRef]

18. Balmas, M. When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism. *Commun. Res.* **2014**, *41*, 430–454. [CrossRef]

19. Kang, C.; Goldman, A. In Washington Pizzeria Attack, Fake News Brought Real Guns. *New York Times* **2016**, *5*, A1.

20. Haithcox-Dennis, M. Reject, Correct, Redirect: Using Web Annotation to Combat Fake Health Information—A Commentary. *Am. J. Health Educ.* **2018**, *49*, 206–209. [CrossRef]

21. Zhou, X.; Zafarani, R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Comput. Surv.* **2020**, *53*, 109. [CrossRef]

22. Tandoc, E.C.; Lim, W.; Ling, R. Digital Journalism Defining "Fake News" A typology of scholarly definitions. *Digit. Journal.* **2018**, *6*, 137–153. [CrossRef]

23. Fuhr, N.; Giachanou, A.; Grefenstette, G.; Gurevych, I.; Hanselowski, A.; Jarvelin, K.; Jones, R.; Liu, Y.; Mothe, J.; Nejdl, W.; et al. An Information Nutritional Label for Online Documents. *ACM SIGIR Forum* **2018**, *51*, 46–66. [CrossRef]

24. Hammond, D. Health warning messages on tobacco products: A review. *Tob. Control* **2011**, *20*, 327–337. [CrossRef]

25. Lomeli, N.; Funke, D. Fact check: Cigarette warning labels in US haven't changed since 1984. *USA Today*. 2022. Available online: https://www.usatoday.com/story/news/factcheck/2022/01/27/fact-check-australias-cigarette-package-warning-labels-not-new/6513681001/ (accessed on 8 May 2022).

26. Hiilamo, H.; Crosbie, E.; Glantz, S.A. The evolution of health warning labels on cigarette packs: The role of precedents, and tobacco industry strategies to block diffusion. *Tob. Control* **2014**, *23*, e2. [CrossRef]

27. U.S. Federal Trade Commission Federal Cigarette Labeling and Advertising Act. Available online: https://www.ftc.gov/enforcement/statutes/federal-cigarette-labeling-advertising-act (accessed on 1 March 2022).

28. Hensley, S. FDA Unveils Graphic Cigarette Labels. NPR Website. 2011. Available online: https://www.npr.org/sections/health-shots/2011/06/21/137316580/be-warned-fda-unveils-graphic-cigarette-labels (accessed on 8 May 2022).

29. CBS News Judge Blocks FDA Requirement for Graphic Tobacco Warning Labels. Available online: https://www.cbsnews.com/news/judge-blocks-fda-requirement-for-graphic-tobacco-warning-labels/ (accessed on 1 March 2022).

30. Ingram, D.; Yukhananov, A.U.S. Court Strikes Down Graphic Warnings on Cigarettes. Available online: https://www.reuters.com/article/us-usa-cigarettes-labels/u-s-court-strikes-down-graphic-warnings-on-cigarettes-idUSBRE87N0NL20120824 (accessed on 1 March 2022).

31. U.S. Food & Drug Administration. FDA Proposes New Required Health Warnings with Color Images for Cigarette Packages and Advertisements to Promote Greater Public Understanding of Negative Health Consequences of Smoking. Available online: https://www.fda.gov/news-events/press-announcements/fda-proposes-new-required-health-warnings-color-images-cigarette-packages-and-advertisements-promote (accessed on 1 March 2022).

32. FDA Label Imaegs. Available online: https://web.archive.org/web/20120302084657/http://www.fda.gov/downloads/TobaccoProducts/Labeling/CigaretteWarningLabels/UCM259974.zip (accessed on 1 March 2022).

33. Craver, R. Tobacco Manufacturers Gain Three More Months before Graphic-Warning Labels Required on Cigarette Packs | Local | Journalnow.com. Available online: https://journalnow.com/business/local/tobacco-manufacturers-gain-three-more-months-before-graphic-warning-labels-required-on-cigarette-packs/article_fd8915b6-8f43-11ec-aad6-2f790b9bdb5a.html (accessed on 1 March 2022).

34. U.S. Food & Drug Administration. Cigarette Labeling and Health Warning Requirements | FDA. Available online: https://www.fda.gov/tobacco-products/labeling-and-warning-statements-tobacco-products/cigarette-labeling-and-health-warning-requirements (accessed on 1 March 2022).

35. Baptista, J.P.; Gradim, A. Understanding Fake News Consumption: A Review. *Soc. Sci.* **2020**, *9*, 185. [CrossRef]

36. Braun, J.A.; Eklund, J.L. Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism. *Digit. J.* **2019**, *7*, 1–21. [CrossRef]

37. Hoek, J.; Wilson, N.; Allen, M.; Edwards, R.; Thomson, G.; Li, J. Lessons from New Zealand's introduction of pictorial health warnings on tobacco packaging. *Bull. World Health Organ.* **2010**, *88*, 861–866. [CrossRef] [PubMed]

38. Rostron, A. Pragmatism, Paternalism, and the Constitutional Protection of Commercial Speech. *Vt. Law Rev.* **2012**, *37*, 527.

39. Motion Picture Association Inc.; National Association of Theatre Owners Inc. *Classification and Rating Rules*; Sherman Oaks, CA, USA, 2020. Available online: https://www.filmratings.com/Content/Downloads/rating_rules.pdf (accessed on 8 May 2022).

40. WELCOME TO FilmRatings.com. Available online: https://www.filmratings.com/ (accessed on 1 February 2020).

41. The V-Chip: Options to Restrict What Your Children Watch on TV | Federal Communications Commission. Available online: https://www.fcc.gov/consumers/guides/v-chip-putting-restrictions-what-your-children-watch (accessed on 1 February 2020).

42. Harrington, R. Record Industry Unveils Lyrics Warning Label. Available online: https://www.washingtonpost.com/archive/lifestyle/1990/05/10/record-industry-unveils-lyrics-warning-label/6fc30515-ac8a-4e5d-9abd-a06a34cb54f2/ (accessed on 28 February 2022).

43. U.S. Federal Bureau of Investigation FBI Anti-Piracy Warning Seal. Available online: https://www.fbi.gov/investigate/white-collar-crime/piracy-ip-theft/fbi-anti-piracy-warning-seal (accessed on 1 March 2022).

44. United States of America. United States Constitution, First Amendment. In Proceedings of the First Congress of the United States, New York City, NY, USA,, 15 December 1791.

45. U.S. Embassy Beijing New PRC Internet Regulation. Available online: https://irp.fas.org/world/china/netreg.htm (accessed on 28 February 2022).

46. Diagne, A.; Finlay, A.; Gaye, S.; Gichunge, W.; Pretorius, C.; Schiffrin, A.; Cunliffe-Jones, P.; Onumah, C. Misinformation Policy in Sub-Saharan Africa. In *Misinformation Policy in Sub-Saharan Africa*; University of Westminster Press: London, UK, 2021; p. 224. [CrossRef]

47. Haque, M.M.; Yousuf, M.; Alam, A.S.; Saha, P.; Ahmed, S.I.; Hassan, N. Combating Misinformation in Bangladesh. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 130. [CrossRef]

48. Carson, A.; Fallon, L. Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific. *Trobe Rep.* **2021**. Available online: https://opal.latrobe.edu.au/articles/report/Fighting_Fake_News_A_Study_of_Online_Misinformation_Regulation_in_the_Asia_Pacific/14038340/1/files/26480915.pdf (accessed on 8 May 2022). [CrossRef]

49. Yadav, K.; Erdoğdu, U.; Siwakoti, S.; Shapiro, J.N.; Wanless, A. Countries have more than 100 laws on the books to combat misinformation. How well do they work? *Bull. At. Sci.* **2021**, *77*, 124–128. [CrossRef]

50. Kumar, P.J.S.; Devi, P.R.; Sai, N.R.; Kumar, S.S.; Benarji, T. Battling Fake News: A Survey on Mitigation Techniques and Identification. In Proceedings of the 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 829–835. [CrossRef]

51. Sharma, K.; Qian, F.; Jiang, H.; Ruchansky, N.; Zhang, M.; Liu, Y. Combating fake news: A survey on identification and mitigation techniques. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 21. [CrossRef]

52. Wang, W.Y. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *arXiv* **2017**, arXiv:1705.00648.

53. Yuan, H.; Zheng, J.; Ye, Q.; Qian, Y.; Zhang, Y. Improving fake news detection with domain-adversarial and graph-attention neural network. *Decis. Support Syst.* **2021**, *151*, 113633. [CrossRef]

54. De Oliveira, N.R.; Pisa, P.S.; Lopez, M.A.; de Medeiros, D.S.V.; Mattos, D.M.F. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. *Information* **2021**, *12*, 38. [CrossRef]

55. Koloski, B.; Stepišnik-Perdih, T.; Pollak, S.; Škrlj, B. Identification of COVID-19 Related Fake News via Neural Stacking. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2021; Volume 1402, pp. 177–188. [CrossRef]

56. Deepak, S.; Chitturi, B. Deep neural approach to Fake-News identification. *Procedia Comput. Sci.* **2020**, *167*, 2236–2243. [CrossRef]

57. Anoop, K.; Deepak, P.; Lajish, L.V. Emotion cognizance improves health fake news identification. In Proceedings of the 24th International Database Engineering & Applications Symposium (IDEAS 2020), Incheon, Korea, 12–14 August 2020; Association for Computing Machinery: New York, NY, USA, 2020.

58. Batailler, C.; Brannon, S.M.; Teas, P.E.; Gawronski, B. A Signal Detection Approach to Understanding the Identification of Fake News. *Perspect. Psychol. Sci.* **2022**, *17*, 78–98. [CrossRef] [PubMed]

59. Pröllochs, N. Community-Based Fact-Checking on Twitter's Birdwatch Platform. *arXiv* **2021**, arXiv:2104.07175. [CrossRef]

60. Bakir, V.; McStay, A. Fake News and The Economy of Emotions. *Digit. Journal.* **2018**, *6*, 154–175. [CrossRef]

61.   Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 199–207. [CrossRef]

62.   Chen, W.; Yuan, Y.; Zhang, L. Scalable influence maximization in social networks under the linear threshold model. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, NSW, Australia, 13–17 December 2010; pp. 88–97. [CrossRef]

63.   Budak, C.; Agrawal, D.; Abbadi, A. El Limiting the spread of misinformation in social networks. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 665–674. [CrossRef]

64.   Jain, S.; Sharma, V.; Kaushal, R. Towards automated real-time detection of misinformation on Twitter. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 2015–2020. [CrossRef]

65.   World Health Organization Coronavirus Disease 2019 (COVID-19): Situation Report-55. Available online: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200315-sitrep-55-covid-19.pdf?sfvrsn=33daa5cb_8&download=true (accessed on 5 November 2021).

66.   Singh, L.; Bansal, S.; Bode, L.; Budak, C.; Chi, G.; Kawintiranon, K.; Padden, C.; Vanarsdall, R.; Vraga, E.; Wang, Y. A first look at COVID-19 information and misinformation sharing on Twitter. *arXiv* **2020**, *arxiv:2003.13907*. [CrossRef]

67.   Scott, M.; Overly, S. Silicon Valley is Losing the Battle against Election Misinformation—POLITICO. Available online: https://www.politico.com/news/2020/08/04/silicon-valley-election-misinformation-383092 (accessed on 4 March 2022).

68.   Conger, K.; Isaac, M.; Wakabayashi, D. Twitter and Facebook worked to crack down on election disinformation, but challenges loom. *New York Times*, 2020. Available online: https://www.nytimes.com/2020/11/04/us/politics/twitter-and-facebook-worked-to-crack-down-on-election-disinformation-but-challenges-loom.html (accessed on 8 May 2022).

69.   Buntain, C.; Bonneau, R.; Nagler, J.; Tucker, J.A. YouTube Recommendations and Effects on Sharing Across Online Social Platforms. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 11. [CrossRef]

70.   Chen, E.; Deb, A.; Ferrara, E. #Election2020: The first public Twitter dataset on the 2020 US Presidential election. *J. Comput. Soc. Sci.* **2021**. [CrossRef]

71.   Sanderson, Z.; Brown, M.A.; Bonneau, R.; Nagler, J.; Tucker, J.A. Twitter flagged Donald Trump's tweets with election misinformation: They continued to spread both on and off the platform. *Harv. Kennedy Sch. Misinformation Rev.* **2021**, *2*. [CrossRef]

72.   Lespagnol, C.; Mothe, J.; Ullah, M.Z. Information Nutritional Label and Word Embedding to Estimate Information Check-Worthiness. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019; pp. 941–944.

73.   Vincentius, K.; Aggarwal, P.; Sahan, A.; Högden, B.; Madan, N.; Bangaru, A.; Schwenger, C.; Muradov, F.; Aker, A. Information Nutrition Labels: A Plugin for Online News Evaluation. In Proceedings of the First Workshop on Fact Extraction and VERification, Brussels, Belgium, 1 November 2018; pp. 28–33.

74.   Gawronski, B. Partisan bias in the identification of fake news. *Trends Cogn. Sci.* **2021**, *25*, 723–724. [CrossRef]

75.   Fairbanks, J.; Fitch, N.; Knauf, N.; Briscoe, E. Credibility Assessment in the News: Do we need to read? In Proceedings of the MIS2 Workshop Held in Conjuction with 11th Int'l Conference on Web Search and Data Mining, Los Angeles, CA, USA, 9 February 2018.

76.   Suttle, R.; Hogan, S.; Aumaugher, R.; Spradling, M.; Merrigan, Z.; Straub, J. University Community Members' Perceptions of Labels for Online Media. *Futur. Internet* **2021**, *13*, 281. [CrossRef]

77.   Straub, J.; Spradling, M. Americans' Perspectives on Online Media Warning Labels. *Behav. Sci.* **2022**, *12*, 59. [CrossRef] [PubMed]