2020

How disinformation becomes misinformation under our ‘watchful’ eye

A research into the mitigation of the effect of illusory truth using a warning tag in visual and written news items

Selma Ali
S1516809
6th July 2020
First reader: Tommy van Steen
Second reader: Yannick Veilleux-Lepage

Mitigating the effect of illusory truth using a warning in visual and written news items

Introduction

"War, tech and politics have blurred into a new kind of battle space that plays out on our smartphones".

This quote by Singer and Brooking (2019: 408) illustrates the ever-increasing influence of social media and the Internet over the past years. Not only our personal but also global networks between states and international organizations depend on this technological structure. The internet has facilitated the interconnection of global society by enabling the tools to a rapid flow of information between an endless number of sources and users (Hameleers et al. 2020). This development has also enabled users, companies and states to use the internet to broaden their influence on a large number of people by emphasizing and distributing certain information, advocating in favor of a certain societal narrative or political opinion (Hamiti, 2016, Zannettou et al. 2019).

1.1 Problem definition

Despite the removal of time and space barriers and creating possibilities to connect global communities, the rise of the internet has enabled the emergence of a new phenomenon namely, the spread of disinformation and misinformation via deceptive or manipulated media content (Hameleers et al. 2020). Manipulation of media content can occur for a variety of reasons by several different actors with different purposes. It can be distributed using textual content, images or for instance, video content (Marwick and Lewis 2017). It is important to address that disinformation can also comprise an authentic image which is altered or distributed without context or background information. Thereby deceiving the audience's perspective of reality (Di Marco 2019).

Messages containing fake news, incorrectly distributed information or even intentionally manipulated media content lead to a crisis in trust towards the media and the accuracy of information (Bradshaw and Howard 2019, Marwick and Lewis 2017, Zannettou et al. 2019). The multitude of discourses used by media platforms and media fragmentation further complicate this process, as individuals are confronted with contradictions and content inconsistent with their world view. This results in a global dilemma on the 'truthfulness' of information depending on the narrative of individuals and possibly a hidden objective pursued

by the source spreading the information which affects not only individuals but also interstate and corporate relations (Barkun 2017, Marwick and Lewis 2017). Is the information I receive true? And who might be interested in making me believe differently or distribute deceptive information this way, and why?

Despite the increasing international efforts to create think tanks used for detection and removal of disinformation, aiming to create more awareness on the importance to assess all information one comes across, the amount of disinformation on the Internet is still great (EU East Stratcom Task force 2020).

1.2 Relevance

Legitimizing and spreading information which turns out to be unreliable or even has the intention to deceive the public not only causes a breach in trust towards all sources of information, the media and peers, but can also be a danger to global society and interstate relations (Marwick and Lewis 2017).

It would allow for a breach in trust towards the media to grow, which consequently undermines the gate keeper position the media holds in all democratic societies. As the media provide citizens with a mean to control and assess behavior and actions by the government which becomes endangered when citizens no longer trust the media to represent this interest, as it is unclear whether the information provided by the media can be trusted (Bradshaw and Howard 2019). Secondly, undermining another state's credibility and stimulating polarization among citizens using disinformation campaigns can influence inter- and intrastate relations, to the point of causing distrust and potential (internal) conflict (Marwick and Lewis 2017).

When referring to society, this research might be relevant for a dual reason. Firstly, the creation and spread of dis- and misinformation, as well as the reduction and removal by thinktanks and international organizations is a highly concealed subject (EU Stratcom Task Force, Bradshaw and Howard 2019). Actors aiming to manipulate content and influence individuals for their own interests will not expose their purpose while disseminating the information, nor will the organizations reveal where or how the information was collected (Marwick and Lewis 2017). In fact, transparency conditions or criteria on which piece of information is considered disinformation is lacking. In addition, there could be asked whether it is desirable for international organization to determine which piece of information is 'true' and which is not.

This would undermine the primary function of the media, as it should function as a mean to be critical and assess one's opinions and stances (Merloe 2017). Therefore, educating and creating awareness among the public on the presence, influence and mitigation of disinformation is important to decrease the negative effects of disinformation and decreased trust in society as it is individual members of society that, mostly, are the object and mean to spread disinformation (Brashier et al. 2020).

1.3 Aim of the research

One of the mechanisms, supposedly enabling disinformation to become part of the dominant discourse and spread is the illusory truth effect, which explains how information that is (vagely) familiar to the receiver, will automatically be labeled more 'truthful', even if this is not necessarily accurate (Brashier et al. 2020). For instance, one reads a news article containing several errors on news facts which the reader is aware of. The reader will most likely assess the information as unreliable. However, the illusory truth effect, increases the possibility for an individual, if one would read a second, similar news article, containing the same facts from a different source or style of writing, to assess the article as credible. Therefore, if one would find out the information was inaccurate after all, the spread of inaccurate information will most likely be regarded to as misinformation as it was not the purpose of the individual spreading this information to mislead people (Dechene et al. 2010, Fazio et al. 2015). However, the initial source of the disinformation, has achieved its goal, as the information has, even if temporarily, become part of the dominant discourse and has reached many individuals (Barkun 2017).

Previous research projects have explored possibilities on how to counter this effect by conducting an experiment where respondents would be exposed to disputed information online with a warning tag. This warning tag would have caused individuals to pay closer attention to the source of information and their previous knowledge on the subject, after which they were able to provide a more accurate evaluation of the truthfulness of the information and were less willing to share the items with others.

The aim of this study would be to build on these results and examine whether the effect of a warning tag is equally effective for both visual and written content, as most previous research has focused mainly on the exposure to written content. Considering the contemporary increase of the use of memes, videos and images by the media and on social media, it could be

valuable to assess whether dis-or misinforming visual content should be countered with different means.

1.4 RQ and sub-questions

Therefore, the following research question will be asked:

Does adding a warning to dis- or misinforming news items have a different effect on visual or written content in mitigating the illusory truth effect?

This question will be disseminated into four sub-questions. Firstly, the illusory truth effect and its relation to the spread of dis- and misinformation will be explained to provide support for the study. As argued before, the illusory truth effect would facilitate disinformation to transform to misinformation as the process concerning the dissemination of information via a third party eliminated the 'deliberate' characteristics present in the definition of disinformation (Karlova and Fisher 2013, Hameleers et al. 2020). Therefore, the following sub-question will be asked: How does the illusory truth effect affect the spread and dissemination of dis- and misinformation?

Incorporating a 'warning', in the form of a small banner at the bottom or top of the manipulated content, would, according to Brashier et al. (2016) and Marsh et al. (2016) mitigate the illusory truth effect and result in a more reliable assessment of information. Therefore, in order to examine the possible difference in effect between visual and written content, it is necessary to first establish the initial positive effect of a warning banner on the mitigation of the illusory truth effect. Leading to the following sub-question: Can the illusory truth be mitigated by adding a warning for the viewers of the content? Secondly, there will be examined whether a differentiation between visual and written content can be detected. This results in the following question: Is there a different in the effect of a warning to mitigate the illusory truth effect between visual and written content?

Finally, the spread of disinformation is dependent on whether individuals share the content with their networks, regardless of it being spread deliberately or accidentally. Therefore, the final aim of the research is to answer under what conditions and to what extent individuals are willing to share (false or true) news items (Chen et al. 2015). If the results would expose the motivations

for sharing dis- or misinformation, whether conditioned with a warning or not, it could provide insight into the process of the spread of dis- or misinformation (Khan and Idris 2019).

1.5 Reader guide

Firstly, the increasing trend concerning the presence of manipulated content in the news and media will be outlined in order to create an overview of the reach of this problem. This will include a literature review on previous research into the creation and purpose of manipulated content by different actors. Secondly, the theoretical framework will provide the conceptualization of the concept disinformation and misinformation, including the relation of these to concepts to the illusory truth effect. The illusory truth effect and previous research into the effects of this cognitive mechanism will be further explained using previous research by Brashier et al. (2020), Pennycook et al. (2017) and Marsh et al. (2016). Next, a definition for visual content, and an explanation on the difference in means of communication between visual and written content will be provided. Finally, using previous research by Chen et al. (2015), a set of motivations to share mis- or disinformation online will be outlined to determine which of these motivations might be relevant to the respondents in the sample. The methodology section will provide an outline on the experimental research design, relevant variables, operationalization of the concepts used, measurements, procedure, analysis and the possible limitations of the research.

Theoretical framework

2.1 Global increase of mis- and disinformation

To examine how disinformation and misinformation spreads, it is important to first look at what these concepts entail and how they are situated in contemporary global society. Since the presidential elections in the US in 2016, there has been increased attention for the dissemination of manipulated or fake news via traditional- or social media portals (Hameleers et al. 2020, Kumar and Geethakumari 2014). This is mainly because Trump as a presidential candidate has used these concepts during his campaign as an argument against the established order. Barkun (2017) describes the concept of 'mainstreaming the fringe', which refers to a global tendency concerning the normalization of ideological opinions which previously were not part of the dominant societal discourse, which potentially cause polarization and societal cleavages. This can also be referred to as the development of a post-truth community. This post-truth community is characterized by a lack of consensus on which piece of information or information source is reliable, as there is no longer a difference between 'mainstream' and 'fringe' discourse. Therefore, the reliability and truthfulness of information becomes highly dependent on individual ideological and social background. Hence, this could explain the growing support for populist, nationalist or even extremist ideological beliefs (Marwick and Lewis 2017).

In addition, the emergence of social media platforms such as Twitter and Facebook seem to have created a platform that is used to share (fake) news which reaches a large global population as the internet is not bound to geographical borders (Hameleers et al. 2020, Kumar and Geethakunari 2014). Research conducted by Kennedy and Prat (2019) shows that the Internet and namely social media platforms form are increasingly important tools for individuals to gather information and read news articles. However, research conducted by Tsfati, and Ariely (2014) argues that news gathered from social media is considered less credible.

The simplicity, anonymity and reach the internet provides enables sources spreading manipulated content to remain unknown and hide their underlying purpose for spreading a certain narrative or ideological content. This increases credibility of the manipulated content, as if one would be aware news is spread by a malicious source it would immediately cause distrust and unreliability (Di Marco 2019, Marwick and Lewis 2017).

2.2 Information, misinformation and disinformation

Echoing the growing importance of skepticism and confusion on the reliability and veracity of the news, the following section will outline the differences between three types of information: information, misinformation and disinformation. To differentiate between these types, Karlova and Fisher (2013) used five criteria which could be appointed to one or more of the three types. The following question would be asked to assess which type of information is discussed. Is the content: (1) true, (2) complete, (3) current, (4) informative and/ or (5) deceptive?

Starting with the concept of information, which seems simple to explain. Madden (2000: 343) defines information as "a fact or circumstance of which one is told", indicating information is characterized by its veracity and the transmission of the content from sender to receiver. As displayed in table 1, information is characterized by content that is true, could be both complete or incomplete, is current, informative and most importantly is not deceptive.

When relating information to the spread of (manipulated) news articles, the transmission and even more important, the interpretation of the content transmitted by a (social) media source or user defines the adoption of the content and impacts the credibility assessment of the receiver. Therefore, even if the news content is categorized as information, this could easily be changed in this process of dissemination between sender and receiver as the interpretation and potential sharing of the content with a different interpretation from the initial source, cause facts to become untrue or are used for a deceptive purpose (Karlova and Fisher 2013, Metzger and Flanagin 2013).

Table 1: summary of features information, misinformation and disinformation

	Information	**Misinformation**	**Disinformation**
True	Y	Y/N	Y/N
Complete	Y/N	Y/N	Y/N
Current	Y	Y/N	Y/N
Informative	Y	Y	Y
Deceptive	N	N	Y
Y = Yes; N = No; Y/N = Could be Yes and No, depending on context & time			

Table 1. A summary of features of information, misinformation, and disinformation.

Misinformation can be true, complete and current although these are not criteria, however it is not deceptive, that is, not with the intention to deceive. This underlines the difference between mis- and disinformation as disinformation is deceptive, that is, created and spread with the intention to deceive the receiver.

Kumar and Geethakumari (2014) have appointed five characteristics to the concept of disinformation. Firstly, (1) disinformation is usually planned and involves a technical though-through process, which intends to target a specific group of people who the recipient intends to influence (2). The information can also be distributed randomly (3) and does not necessarily only reaches the targeted group (4) as the information may be distributed via third parties or between social networks. Disinformation includes textual or verbal content including, manipulated images and/ or video material (5). However, the most important feature of disinformation is in the purpose to distribute the content: it involves deliberately manipulated content which is meant to deceive the targeted group or person.

The fourth point argued by Kumar and Geethakumari (2014), concerning the distribution of disinformation through third parties forms the bridge between dis- and misinformation. Misinformation involves the spread of misleading or manipulated content, without the intent to deceive or mislead the recipient. The source, spreading this material, might accidentally assess this information to be truthful, resulting in the spread of misinformation which would initially be categorized as disinformation. Therefore, the mechanisms enabling disinformation to become misinformation, through an information dissemination process where disinformation is spread accidentally is a vital factor in the strategic planning of the spread of disinformation (Karlova and Fisher 2013, Marwick and Lewis 2017). Traditional and namely, social media has an important role in this process. As the media distributed the news, misleading content, when picked up by the media, could be distrusted widely and will most likely be assumed to be truthful as it derived from a source people will find credible (Bradshaw and Howard 2019).

2.3 Credibility of information

Despite the categorization of the three types of information, it is difficult to assess to which type of information certain content belongs as there might be overlap and ambiguity on the source, intent and factuality of information (Karlova and Fisher 2013). Therefore, it is important to examine how individuals would assess the credibility of the information they come across. Research conducted by Gigerenzer and Todd (1999) and Metzger and Flanagin (2013) on the use of cognitive heuristics in the assessment of credibility of information has produced a model containing six factors that would impact and explain the credibility of information for an individual, two of which will be explored more extensively in this research.

Firstly, credibility is dependent on the concept of endorsement. The assessment of the information by peers and their reliability impacts the credibility assessment. For instance, if a

news article is liked or shared by a known person, the individual exposed to this information is inclined to assess it as more credible. As describes, this research aims to examine the sharing and liking behavior of respondents when confronted with manipulated content, whether mis- or disinformation. According to this condition of credibility, information shared by peers, would automatically be more trustworthy, even if this would be mis- or disinformation. This will be explained more extensively in the paragraph on motivations to share and like misinformation amongst individuals (Gigerenzer and Todd 1999, Metzeger and Flanagin 2013.

Secondly, the consistency of the messages across different sources. If the content of the information is echoed by multiple sources, it will increase the credibility of the message, regardless of the factuality of the content. Third, the expectancy violation, which refers to the behavior and presentation of the information content. If for instance, one comes across a website that differs greatly from 'trusted' news websites and uses for instance, informal language, one is inclined to assess the content or source as less credible. Fourth, affecting the credibility of information is self- confirmation, which refers to whether the message confirms one's preexisting worldview and discourse. Again, this factor endorses the irrelevance of the factuality of the content, as individuals will automatically tend to find sources echoing their discourse more credible. Fifth, the intent of the source creating the content is important when assessing the credibility. However, this factor proves to be difficult when for instance, applying to social media networks as it is often unclear who the author of the content is and what intent they might have in diffusing the message (Metzger and Flanagin 2013).

Finally, the reputation or recognition of information. Gigerenzer and Todd (1999) describe how individuals tend to assess information or sources of information they are familiar with as more valuable and trustworthy, regardless of whether the content and argumentation is indeed correct and factual. This is directly related to the effect examined in this research, namely the illusory truth effect, which relies on this same principle. If an individual is already familiar or has been exposed to a certain subject or claim, they will most likely assess the content as more credible.

2.4 Illusory Truth effect

Marsh et al. (2016) have attempted to describe the effect of illusory truth on the basis of retrieval fluency. Individuals who have gained knowledge often will not explicitly remember where they have retrieved that knowledge, it has just become part of their memory. For instance, from a student's perspective, you 'just know' that Amsterdam is the capital of the Netherlands and the parliament is located in Den Hague, but you will not necessarily remember where you learned this. However, the location of for instance, the French parliament would be more difficult to retrieve from memory and would rely less on a process of fluency. People judge their knowledge based on whether it is easy or difficult to retract that knowledge from memory, which is defined as retrieval fluency. This subsequently conditions the of confidence in one's answer, as more fluency is related to more confidence. As illustrated in the model below, reliance on retrieval fluency will most likely cause the individual to answer a question of 'truth' affirmatively (Brashier et al. 2020, Fazio et al. 2015, Pennycook et al. 2018).

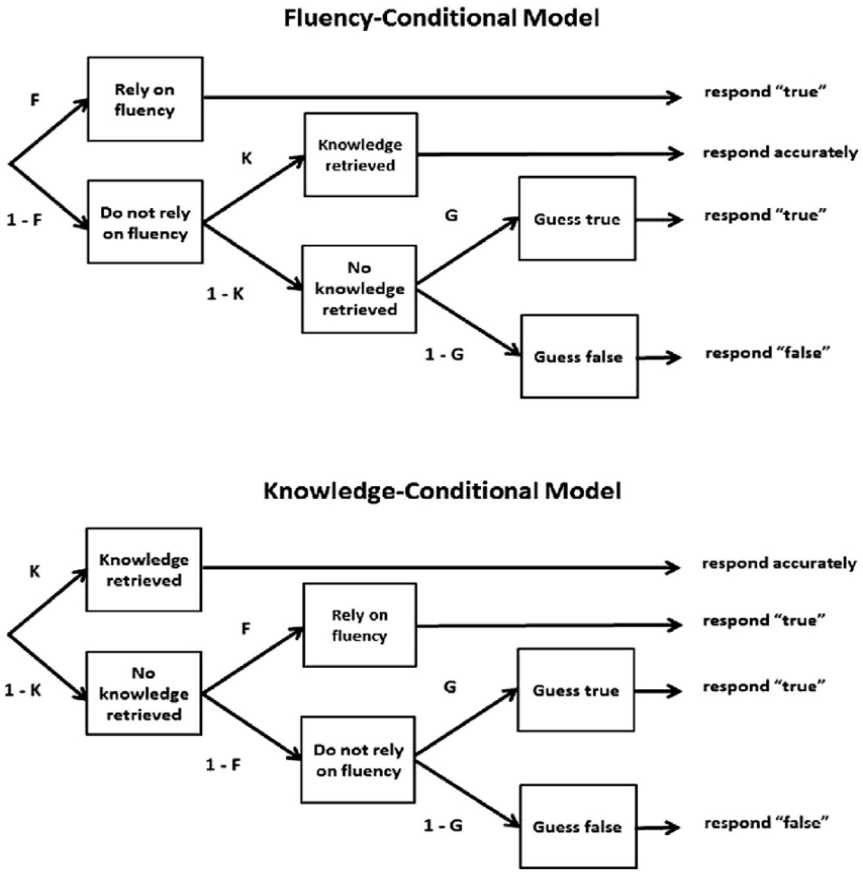


Table 2: fluency- and knowledge-conditioned model

The primary effect caused by illusory truth is that repetition or prior exposure would increase the likelihood that one would assess content to be accurate, even if this is not the case. Familiarity would, through repetition, simplify processing content and would increase retrieval fluency, which subsequently is used to verify the accuracy of the content. (Brashier et al. 2020, Dechene et al. 2010, Brashier and Marsh 2016). Research by Marsh et al. (2016) has shown that this effect even occurs if an individual does remember the source of knowledge. Their sample of respondent has been asked to read misinforming stories and later answer questions on general knowledge. After the test, many respondents first of all, made errors in their general knowledge based on the misinformation they were exposed to and secondly, indicated they has answered the question based on pre-experimental knowledge, while the results show that the questions were answered based on the misinformation. This indicates that the effect of illusory truth via repetition and prior exposure causes individual to rely on a fluency-conditioned process even if they can remember the source of the information.

The applicability and presence of this effect will be assumed in this research given the fact that multiple research projects have confirmed the effect of previous exposure on familiarity of information and the assessment of credibility.

2.5 Warning effectiveness

However, research conducted by Ecker, Lewandowsky and Tang (2010) shows that the incorporation of a specific warning, would reduce reliance on misinformation. According to these authors, a warning reduces the reliance on misinformation when it specifically attempts to activate a 'source retrieval mode' in memory. This would increase awareness and the importance of taking note of the source of information. However, this research notes that the effect is only reduced but not eliminated. Despite the conclusion drawn by Marsh et al. (2016), the results from Ecker et al. (2010) could provide a potential condition to counter the effect of illusory truth, as the incorporation of an element of a specific warning on the presence of misinformation, combined with a focus on source retrieval might mitigate the effect describes by Marsh et al. (2016) as this research has not specifically emphasized the 'presence' of misinformation before conducting the survey.

In imitation of Ecker et al. (2010), Brashier et al. (2020) have conducted a survey on the effectivity of a specific warning if individuals are explicitly asked to take the role of 'fact-checkers', hence activating the source retrieval mode. In addition, a second condition for reliance on misinformation was added to the research, namely, the illusory truth effect. The

results of this research show that even if the illusory truth effect would be present, the incorporation of an 'accuracy focus' via a warning and activating the role of 'fact checker' would reduce the strength of this effect. Building on these results, the following hypothesis has been be formulated:

H1: Incorporating an 'warning' on the possibility of misleading information will trigger individuals to pay close attention to the source of information, hence attend to a knowledge-based path to assess the veracity of content instead of relying on familiarity.

If the H1 will be confirmed, this will automatically confirm the second hypothesis:

H2: Therefore, if individuals are made aware of the possibility of misleading information in news reports, they will likely counteract the effect of illusory truth and result in a more accurate assessment of factuality of news items.

However, a second research by Pennycook et al. (2020) emphasizes the potential negative consequences of the use of a warning to counter the spread of misinformation and reliance on familiarity to assess information. The effect of a warning on perceived accuracy of information is, according to this research, limited, while it causes a second problem concerning the verification of items without a warning. The 'implied truth effect', is described in this research as a mechanism that could cause unintentional verification of items that have not been tagged with a warning. If not all items containing mis- or disinformation are tagged with a warning, they risk automatic verification of the authenticity of the content, as individuals rely on the accuracy of the warning to indicate any false or deceiving information.

2.5 Visual and written communication

A distinction that has not yet been made in previous research is whether there could be a different effect of the incorporation of a specific warning between visual and written content.

Visual content containing dis- or misinformation can be defined as an image or video that contains information that is not real or how are not representative of reality. This includes content that causes erroneous associations or are contextual deceptive that are not in line with the reality presented in the picture. For instance, a photoshopped image, deep fake or a picture

that only shows part of the actual image, creating the possibility for someone to interpret the content from a deceptive context (Di Marco 2019). Hameleers et al. (2020) argue that visual content provides a direct relation to reality, hence causing them to be perceived as more credible by the receiver. In short, visual content lowers awareness and suspicion on the possibility of disinformation and is therefore automatically perceived as more trustworthy.

If, as Di Marco (2019) and Hameleers et al. (2020) stated, visual content is more credible to individuals as it represents reality in a direct way, there could be hypothesized that visual content would prove less receptive to a specific warning to mitigate the effect of illusory truth. Therefore, the following hypothesis has been formulated:

H3: Visual content containing disinformation conditioned with a warning might show less effect in mitigating the illusory truth effect.

2.6 U&G theory

In addition to the importance of recognizing mis- and disinformation, the distribution and sharing of mis- and disinformation, namely on the Internet is of equal importance. Social media focusses on user-generated content which is spread globally via social networks between these users. This simplifies and enabled the spread of disinformation not only due to the reach but also due to the lack of control and filter means to apply to the content that is shared (Hameleers et al. 2020, Al-Rawi 2019).

Lee and Ma (2012) and Chen et al. (2015) have examined the motivation for individuals to share information online. Their research focused on testing a model of four motivations which are: information seeking, socializing, entertainment and status seeking. This model is based on the theory of uses and gratifications (or U&G theory). This theory attempts to explain what motivates a specific audience to use a specific media channel and what conditions the choice of content of the individual. In relation to this, motivation, attitudes and behavior concerning sharing this content could be formulated.

Firstly, information seeking. This motivation is related to media consumption behavior in general and online news use (Kennedy and Prat 2019). Reading the news online and the usage of social media networks helps individuals to gain information and to share this information with others who have similar interests. Individuals would be stimulated to share content if they

have anticipated on others potential information needs and share it to facilitate information to others (Lee and Ma 2012, Chen et al. 2015).

Secondly, socializing and creating social relations on the internet and social networks could form an alternative to communication in reality. Maintaining online relations and interaction with others could motivate individuals to share content because they want to share experiences and therefore want their social network to be informed of the same subjects.

Third, entertainment and emotional relief could be a motivation to share content as discussing and for instance joking on the content of news articles could provide enjoyment to both the sender and receiver (Chen et al. 2015).

Fourth, status seeking is an important motivation to share content online, as it provides the sender recognition of their peers and achieves a knowledgeable status. Therefore, status seeking is closely related to socializing, as the social consequences of sharing certain content could provide the sender with a higher status among their social network.

Finally, prior experiences with sharing content online could be more of a catalysator instead of a motivation. If one has had a positive experience in sharing content, the familiarity of the content and website increases, hence increasing the perception of the capability and competence to share. This results in a high probability that the individual will share their content (Chen et al. 2015, Lee and Ma 2012).

In addition, Chen et al. (2015) have conducted a research into the motivations to share misinformation, in specific on social media platforms. The results of this research show all four motivations (information seeking socializing, entertainment and status seeking) are related to the users' behavior concerning sharing misinformation, in specific news articles, online. Respondents indicated the information could be a good topic of conversation and sharing would help to get other people's opinions regarding the information in the article.

Strikingly, the motivations to share misinformation among social networks appear to be independent of the credibility and veracity of the content of the news article. Chen et al. (2015) indicated that accuracy and the authority of the information source scored very low in their scale indicating the motivation to share certain content. Therefore, there might be concluded that credibility and accuracy of information is not a salient motivation included in the consideration to share content.

In short, it seems as creating awareness on the presence of misinformation does not limit sharing content, as it seems that content is shared for a variety of reasons despite any knowledge on the

presence of mis- or disinformation in the content. However, it could be valuable to examine whether this attitude would change if individual were exposed to an explicit warning on the presence of mis- or disinformation. Therefore, the final hypothesis has been formulated as following:

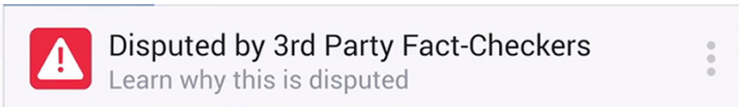
H4: If individuals are made aware of the presence of mis- or disinformation via a warning, it might activate a source retrieval and knowledge-conditioned processing of information and cause a change in sharing behavior.

Methods

3.1 Research design

To examine whether there is a difference in the effect of a warning on mis- or disinformation between visual or written content, an experimental survey will be conducted using quantitative research methods. This research design seems fit to examine this effect as conducting an experiment provides the opportunity to test correlations expected concerning credibility assessments and allows to add a manipulation to further test and complement the robustness of the model explaining the effect of illusory truth (Bryman 2012).

The research sample is divided into four groups. Manipulation groups 1 and 2 will receive no warning and either 6 written or 6 visual news items. Manipulation group 3 and 4 are conditioned with a warning and receive either 6 written or visual news items. The warning condition will only be present in news items that are false, as confirmed by the misinformation database. This results in four groups: (1) written without a warning, (1) visual and headline without a warning, (3) written with a warning, (4) and visual and headline with a warning. The survey is divided into three phases to test the hypothesis: the familiarization phase, where the respondents are exposed to 8 subjects to impose prior exposure, the distraction phase containing questions on demographics, the third phase, containing questions on social media use and news consumption and finally, fourth phase containing questions on the veracity of news items complemented by questions on the willingness and motivation to share a certain item (Brashier and Marsh 2016, Brashier et al. 2020, Ecker et al. 2010, Pennycook et al. 2018).



Warning that will present at the bottom of an item containing mis-or disinformation (ABS CBN news 2016).

To test the hypothesis on the effect of a warning on visual or written content, the distribution of the news items among all groups will be as following: all respondents will, during the fourth phase, receive either 6 visual news items (an image or photo) with a headline or 6 written items with an additional explanation of the topic.

Finally, not all news items will contain mis-or disinformation, 3 news items all four groups are exposed to will derive from a mis-or disinformation database. However, considering the fact

that implausibility of a news item could mitigate the effect of illusory truth, news items that are not entirely implausible are sorted out to control for this effect (Pennycook et al. 2018). All respondents will receive the same 6 news items, to limit the possibility of external effects of the content of the news items. As for instance, some i*f*tems might cause an emotional effect as the subject is currently debated, which could, if only one group is exposed to this article, impact the internal validity of the research (Di Marco 2019).

Research methods

3.2 Dependent and independent variables

The first independent variable is 'prior exposure' which is presumed to influence the dependent variable of 'truth rating'. If prior exposure to a subject would indeed cause a higher truth rating in the number of questions asked, this mechanism proves the effect of the illusory truth. Prior exposure will be conditioned in the survey by adding questions on 'interestingness' of certain subjects. Exposure to these subjects will be an indicator for prior exposure (Ecker et al. 2010). For this research there has been chosen to condition all groups with prior expose, as the illusory truth effect is presumed to be applicable considering the robust academic research and confirmation of this effect.

The second independent variable is the 'warning' condition, which will be indicated by the Third-Party fact checkers tag situated at the bottom of the news articles provided to manipulation group 2 (Brashier et al. 2020). If this variable indeed causes the manipulation groups conditioned with this variable, to have a more accurate assessment of the truthfulness of the items compared to the control groups, this would confirm the hypothesis on the effectiveness of a warning to stimulate a source retrieval mode.

The third independent variable is the condition on either visual or written news items, which will be indicated by the different types of news items respondents will be exposed to.

The dependent variable is 'truth rating' as a warning and/ or the differentiation between visual or written items are expected to affect the participants rating concerning accuracy of the news articles they are exposed to. This variable serves as an indicator for the illusory truth effect, as prior expose (which all groups have been conditioned with) is expected, based on the hypothesis, to be affected by either of the two independent variables.

In the analysis and data processing this variable will be divided into three variables: the mean of all truth rating of respondents, the mean of the truth rating of false news items and the mean of the truth rating of factual news items. This, to examine whether the difference in accuracy and assessment of credibility of false or true item differ between the four manipulation groups.

Finally, the second dependent variable is 'willingness to share', which will be analyzed and compared between all four groups. This to see whether these is a difference between the groups in terms of their sharing behavior and motivation to share. If the results would indicate a difference between either the independent variable 'warning' or 'visual/ written items', the final hypothesis could be confirmed.

3.3 Data collection

The data to examine the hypothesis will be extracted from the survey as described above. Qualtrics XM will be used to construct and extrapolate the survey data, after which it will be analyzed in SPSS. The visual and written items containing mis-or disinformation will be extracted from a multitude for sources providing examples on items containing disinformation, such as the EUvsDisinfo database, local news broadcast, Nieuwscheckers.nl, Facebook and Twitter (BBC news 2020, EU Stratcom Task Force 2020, Newton 2016). During the construction of the survey, all news items will be provided with a source. This to condition the respondent with a minor 'accuracy focus' (Brashier et al. 2020) and the option to use the source in their credibility assessment. Due to the fact that many websites already attempt to remove as many disinforming items as possible, it proves difficult to find a source that provides more than one news item to use in the survey. However, the website EUvsDisinfo by the European External Action Service's East StratCom Task Force (2020) and the website Nieuwscheckers from the University of Leiden provide a fact-checked database with useful news items on disinformation. Both websites have categorized disinformation by subject and provides both articles containing disinformation and accurate information on the subject. The content on these websites have provided reliable data as all content has been fact-checked and the assessment of the accuracy of the articles is substantiated by multiple sources (Brashier et al. 2020).

An important note to clarify why there has been decided to use news items both deriving from social media and traditional media portals would be that in the contemporary society, the

dividing line between social media and news websites has become ambiguous. This refers to the fact that news websites (almost) always have multiple direct links and options to share the content gathered on the website on one or more social networks, such as Twitter and Facebook. Subsequently, these news portals also use social media to spread the same items that have also been posted on the website. Therefore, it seems logical to assume the difference between news items on social media and news items posted on websites could fall under the same category. However, a limitation of this approach could be that social media is not limited to one news source, it provides a platform for all types of news sources, while websites usually host only one news portal. However, all items have been selected based on the fact that they have been published online, as examining the spread of mis- or disinformation from a TV episode for instance, would be difficult to test in an online survey.

3.4 Sampling

This research will use convenience and snowball sampling to select participants for the experimental survey. Participating to the survey will be on a voluntarily basis, as participants will be approached via social media and verbal communication. The sample will consist of individuals that are conveniently available or encounter the survey via another participants (Bryman 2012).

Due to time limitations it is not feasible to use random sampling as this would require gathering data on the entire population the research aims to study, which consists of the entire global population. However, due to the fact that disinformation and the assessment of information via traditional and online channels is used by almost all groups of society, regardless of age, nationality, education level or economic status, using a convenience sampling method does not necessarily results in limitations concerning the reliability and validity of the research (Fazio et al. 2015, Pennycook et al. 2018).

3.5 Respondents

In order to study a representative sample of the population, the sample consist of 152 respondents, between the age 16 and 74. After cleaning the data and sorting out missing data, (most) analyses have been conducted with 149 respondents. Even though convenience sampling usually caused a specific age group to be overrepresented, this does not seem to be the case in this study. A large N will increase the reliability of the research and will increase the probability that the results are generalizable for the population (Bryman 2012). However, due to the current covid-19 pandemic that is still progressing while this research is conducted, it proved to be

more difficult to find a large sample of respondents willing to participate in the research. Dependence on online channels to recruit participants have possibly limited the reliability of the study as the four groups each only consist of 37 to 40 participants.

3.6 Measurements and operationalization

Prior exposure will be measured using a number of questions focusing on asking participants whether they are interested in a certain subject. By asking these questions, the subject present in the questions will be familiarized and therefore can be an indicator of prior exposure. Participants will be asked to rate on a scale from 1 to 10 (1= very uninteresting to 10 = very interesting) whether they are interested in the presented subject (Ecker et al. 2010, Marsh and Brashier 2016). The outcome of these questions will not be processed during the analysis, as the only purpose of these questions is to make sure the respondents have been exposed to a certain set of subjects to condition prior exposure.

The truth rating will be measured using a second set of questions on the truthfulness of 6 news articles. Depending on the manipulation group the respondent had been assigned to, respondents will receive either 6 visual news items with a headline or 6 written news items with or without a warning.

All of the subjects have also been present in the familiarization phase. Participants will be asked to rate the accuracy of the news article on a scale of 0 to 100 (0 = definitely false to 100 = definitely true). In order to create a reliable variable to measure the truth rating given by respondents, a mean variable, based on the 6 answers provided per respondent will be constructed with SPSS (Brashier et al. 2020, Ecker et al. 2010). This scale will range from 0 to 100 and provides an overview of the possible difference between the 4 groups.

The images and written news items will consist of a variety of 6 subjects: politics, digital innovation, art and cultural heritage, food and health, conflict resolution in the middle east and law enforcement and democracy. A variety of subjects has been chosen for a dual reason. Firstly, selecting only one subject, might cause a bias in the results as some respondent might know more on the particular subjects than other respondents. Which could result in a result that is not representative for the population (Pennycook et al. 2018). Secondly, disinformation emerges in all types of news items, therefore, using a multitude of subjects might broaden the

research on disinformation in the media as the results could be generalizable (Marwick and Lewis 2017, Di Marco 2019). At the bottom of all images and written items the source will be provided. An example of a visual or written news item respondents will be exposed will be provided below.

Example 1: visual content containing disinformation (Watt 2019)

The image on the left gives the impression of a child lying between two graves. However, the photographer has indicated to have taken this picture of his nephew who was just playing in the sand (source). This would be an example of an authentic image that has been manipulated by erasing the context. Below versions of this example will be provided, each will represent a news item that will be provided to a different condition group of respondents. The item provides below is an example of a false news item. The first group who will receive only a headline and written statement with additional information on the subject.

Orphaned Syrian Boy Sleeping Between his Parents' Graves.

During the winter of 2017 a shocking news item of a young Syrian orphan lying between the graves of his parents, decorated with stones and flowers, has been spread across the world. This would be a result of the war Syria, which causes thousands of deaths and refugees every year.

Published by multiple sources on Twitter and Harald Doornbos World Press

Image 1: news item manipulation group 1: text without a warning

The second group, who will receive only an image and headline without a warning of the veracity of the item would be exposed the image below. On the right side the source of the image has been provided.

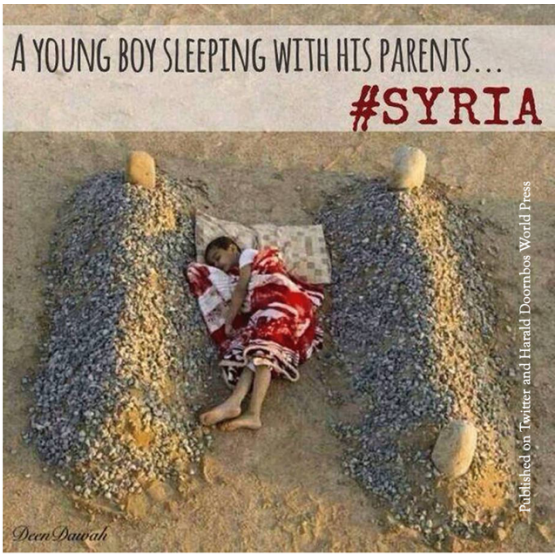


Image 2 manipulation group 2: image and headline without a warning

The third group will receive the same text and additional information as the first manipulation group; however, a warning tag has been added to make the respondent aware of possibly disputed information or falsified content.

Image 3: manipulation group 3: text with a warning

The same has been conducted for the last group, however, they will receive a headline and image instead of a written statement.

Image 4: manipulation group 4: image and headline with a warning

The willingness to share news items will be measured with a similar scale to the truth rating scale described above. Respondents will be asked to indicate on a scale of 0 (= would definitely not share) to 100 (= would definitely share) whether they would be willing to share a certain news item. This scale will subsequently be used in the analysis to compare the means of the 4 groups and assess whether the willingness to share would be less when one is made aware of the presence of false information.

The motivations to share news items containing mis-or disinformation will be measured according to the scale provided by Chen et al. (2015). During this study, they have developed a number of statements respondents could mark as a reason to share an item, which would each indicate one out of four motivations to share content online (information seeking, socializing, entertainment and status seeking). After rating the truthfulness of a news item, the respondent will be asked whether they would share this particular item (whether on social media such as Facebook, Twitter, WhatsApp or in person). If they would respond with 'yes', they will receive a number of statements indicating their motivation to share the item.

The following statement will be incorporated in the survey:

1. “Sharing helps me bookmark useful information” (1) and “Sharing helps me get other people's opinions regarding the information/event” (2) would indicate the motivation ‘information seeking’.
2. “Sharing helps me interact with people” and “The information can be a good topic for conversation.” would indicate the motivation ‘socializing’.
3. “Sharing is good for keeping boredom away” and “I feel enjoyment while sharing.” would indicate the motivation entertainment.
4. “I can express my opinion by sharing that information.” and “Sharing makes me feel influential” would indicate the motivation status seeking or self-expression.

The final two statements although not significant in the study conducted by Chen et al. (2015) will be incorporated, “The information seems accurate” (1) and “The information comes from authoritative sources” (2). These statements, representing the motivations accuracy and authority of the source, could be valuable to measure as, if they would be significant in this study, confirm the significance of stimulating a ‘source retrieval mode’ in individuals when assessing information. If accuracy and authority of the source would for instance be significant in manipulation group 2, conditioned with a warning, there might be concluded that a warning causing individuals to assess the information they share more thoroughly and change their attitude and motivations to share information.

3.7 Procedure

After the respondents have given their consent to participate in this study, Qualtrics XM will randomly divide the respondents into manipulation group 1, 2, 3 or 4. All groups will proceed to the first phase and answer 8 questions on their interest in a certain subject to condition prior exposure. Subsequently, all groups will continue to fill in questions on their demographics, social media use and news gathering behavior. After the initial phases the respondents will move on the final phase, where they will be asked to rate the truthfulness, willingness to share and motivation to share of 6 news items. Finally, the respondents will be asked if they have any comments on remarks regarding the survey or research. They can optionally leave their email address if they are interested in receiving the results of the research.

3.8 Data analysis

For the analysis firstly, the demographics and characteristics of the sample will be provided in bar charts, diving their age, education level and gender. Secondly, frequency tables and descriptive statistics will be provided to analyze the results of the social media use and news consumption behavior, in order to examine whether individuals are engaged in social media networks and consult the same channels to gather their daily news. To examine the first hypothesis the dependent variable of 'truth rating' will be compared to the independent variables of the 'warning' and 'visual/written content' using ANOVA and independent t-tests to conclude on whether there is a difference in effect between the 4 groups. Subsequently, the same analysis will be conducted for the two separate variables differentiating between the false and true news items. If these results would indicate an effect that is significant in the groups who have received a warning, the first and second hypothesis could be confirmed. In addition, if the result would show a difference between the groups who have received either visual or written content, the third hypothesis could be confirmed, if indeed, the groups who have received visual content show have a higher truth rating despite receiving a warning compared to the other groups. In order to conclude on the correlation between the willingness to share and the truth rating provided by respondents a bivariate correlation test will be conducted.

Followed by similar analyses to examine the second dependent variable on willingness to share, by comparing between the groups using ANOVA and providing descriptives on the mean value on the willingness to share by respondents. This to examine the fourth hypothesis, as this could be confirmed if indeed there is a difference between the groups who have received a warning. Subsequently, there will also be examined whether the willingness to share items that are false, is equal compared to the items that were true.

In addition, frequency tables will be constructed to examine and structure the motivations respondents have given to share the content they have been exposed to.

3.9 Limitations

However, this research could be limited by a number of factors. Firstly, if the number of respondents willing to fill out the survey would be low (below 100 for instance), the reliability of the results of the research could be reduced (Bryman 2012). Meaning that if this research would be repeated, the probability that the results would be similar to the outcomes of the first research are limited.

Secondly, despite the attempt to support the choice of news items as precise as possible, the sources providing dis-or misinformation and theory on which type of news items should be selected for this type of research is very limited. Therefore, the content chosen could be subjective, as the researcher could have biased the choice of content. This risks a reduction in the internal validity, as it is possible that there has not been measured what has been claimed to.

Finally, the scale aiming to measure the dependent variable 'truth rating', should be controlled for a mediating effect of previous knowledge of the respondent. If a respondent has previous knowledge on a subject, the likelihood of rating an item as truthful due to the effect of illusory truth reduces (Pennycook et al. 2018). According to Fazio et al. (2015) however, knowledge on a subject would not mitigate the effect of illusory truth entirely, as the multitude of sources and frames used to address the subject could cause an individual to still rely on fluency of retrieval. Still, controlling for previous knowledge at least to some extent is necessary but proves difficult as all respondents are likely to have at least some knowledge on a subject. This results in a potential reduced internal validity, as there cannot be stated with certainty that it is indeed the prior exposure manipulated in the survey, which causes the respondent to have a high truth rating.

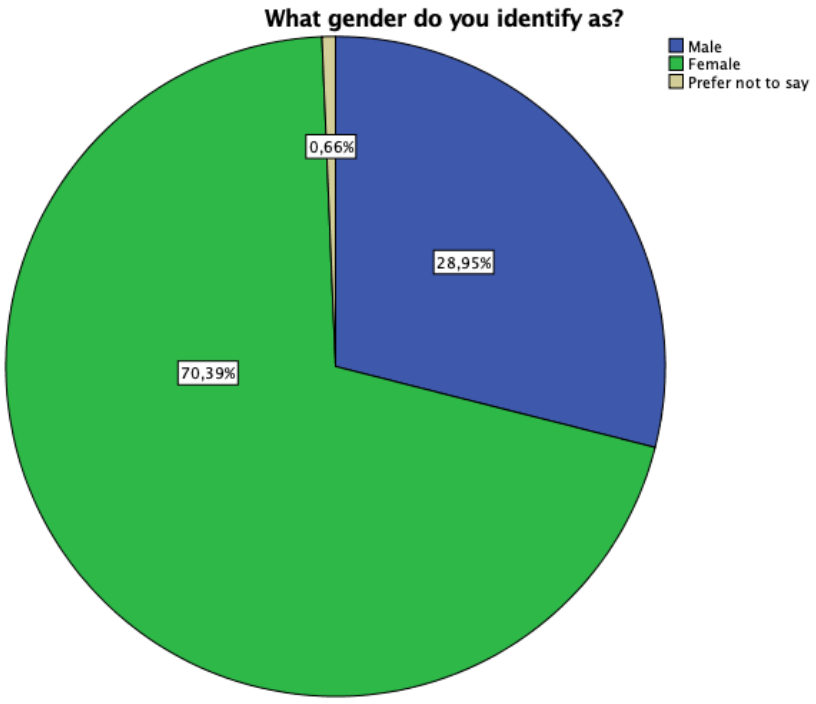
Results

4.1 Demographics and sample description

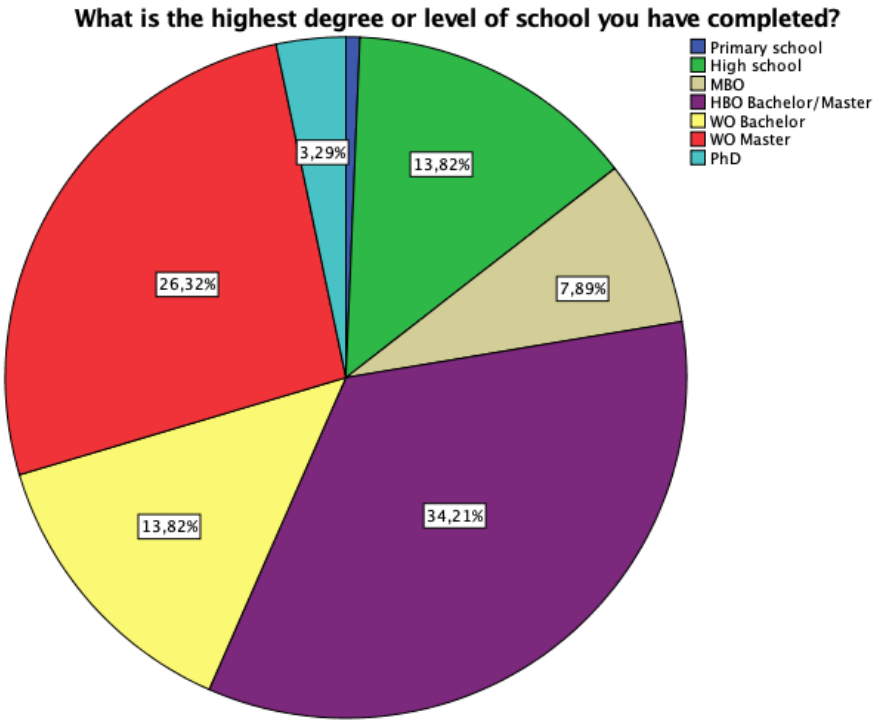
To interpret the results of the survey it is necessary to first examine the characteristics of the sample in terms of age, gender and education.

The age of the respondents ranges between the age of 16 and 74 in a sample of 152 respondents. 41, 4% of the sample is between the age of 16 and 25, which includes the mode of 23 years. Only 17,8% is between the age of 26 and 50 years, making the mean of 40,22 less relevant. 40,8% percent of the sample is between the age of 51 and 75 years old. This can be explained by the sample method that has been chosen, namely convenience- and snowball sampling. These age categories have been created to illustrate the distribution of respondents between age groups.

Second, the sample shows a gender distribution of 28,9 percent male and 70,4 percent female respondents, with only one respondent indicating not wanting to disclose their gender. The overrepresentation of female respondents could have influenced the reliability of the results. However, the aim of the research is to examine sharing behavior and the influence of the illusory truth effect, which applies to all individuals regardless of their gender, therefore, the potential bias is limited.



Pie chart 1: gender distribution



Pie chart 2: education level

The results in the pie chart show that 43,43% of the respondents have completed a WO education or PhD, whereas 34,21% percent has completed an HBO Bachelor or Master. Therefore, the vast majority of the sample has completed a higher education (77,64%). The overrepresentation of individuals with a higher education could potentially bias the results of the research as it is possible that the access, knowledge and interpretation on different news source differs between these groups.

4.2 Interest questions

In the survey, respondents were first asked to rate their interest in 8 subjects on a scale from 1 to 10. The results of these questions, however, are not of statistical value as these questions were only used to condition the illusory truth effect. All news item questions answered by respondents contained one or more of the subjects they had been exposed to in the interest section. If someone has been exposed to a certain subject beforehand, the probability that the news item with a similar subject is evaluated as truthful increases.

4.3 Social media use

Below a frequency table on the communication and social media channels used by respondents has been provided. Two observations are interesting considering the objective of this research.

First, only 34,2% of the respondents has indicated to spend 1 hour or less per day on social media. This indicates that two third of this sample spend more than an hour per day on social media or communication channels, which increases the change for these individuals to be exposed to mis- or disinformation. In addition, 95,4% of the respondents has indicated to use WhatsApp, 80,3% uses Facebook and 61, 8% uses Instagram. TikTok, Google+ and Pinterest were only used on approximately 10% of the sample.

Table 3: communication channels and social media use

	Frequency	**Percentage**
Facebook	122	80,3
Twitter	35	23,0
Instagram	94	61,8
WhatsApp	145	95,4
LinkedIn	76	50,0
Snapchat	30	19,7
Pinterest	26	17,1
TikTok	11	7,2
YouTube	93	61,2
Google+	23	15,1
Other	4	2,6

Table 4: Time spent on social media per day

	Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Less that 30 minutes	8	5,3	5,3	5,3
30 – 60 minutes	44	28,9	28,9	34,2
1 – 2 hours	58	38,2	38,2	72,4
2 – 3 hours	34	22,4	22,4	94,7
3 hours +	8	5,3	5,3	100,0
Total	152	100,0	100,0	

4.4 News consumption and sharing behavior

In terms of sharing news items on any of the social media channels used by respondents, 24,3% has indicated to share on a daily basis, 44,7% shares content 1 to multiple times per week, whereas 5,3% has indicated they never share content on social media.

Table 5: Sharing behavior on Social Media

	Frequency	Percent	Valid Percent	Cumulative Percent
Daily	37	24,3	24,3	24,3
4-6 times a week	23	15,1	15,1	39,5
2-3 times a week	26	17,1	17,1	56,6
Once a week	19	12,5	12,5	69,1
Several times a month	16	10,5	10,5	79,6
Once a month	11	7,2	7,2	86,8
Less than once a month	8	5,3	5,3	92,1
Never	12	7,9	7,9	100,0
Total	152	100,0	100,0	

Strikingly, however not surprising considering the growing importance of the internet and online news channels, apps and websites are the channels mostly used to gather the news. Newspapers still are used by a significant amount of the sample; however, this could also be the online version of the newspaper as respondents were not asked to indicate which channel of the newspaper they use. Considering the fact that online channels have been indicated to be used more frequently compared to offline means (41,12% offline compared to 58,88% online), the findings not only confirms the contemporary transition and importance of online sources to gather the news, but also possibly explains why the spread and creation of mis-and disinformation has been expanding over the past years. When using online sources or apps to gather the daily news, it enables the individual to reach out to sources that fit their societal narrative and information needs, thereby reciprocating the narrative they already had. Secondly, the gatekeeper position usually in the hands of media channels, evaluating the veracity and authenticity of the information, is absent. This increases and simplifies the possibility for disinforming sources to expose individuals to deceptive information. According to Tsfati and

Ariely (2014), trust in news items gathered online or via social media is less compared to traditional media sources, which could explain the increase of the global crisis in trust towards the media, when assessing the results of this sample.

Table 6: News consumption channels used most frequently

	Frequency	Percentage
Newspapers	110	72,4
TV	97	63,8
Websites	110	72,4
Radio	43	28,3
Apps	116	76,3
Social Media	96	63,2
Podcasts	36	23,7

Table 7: News sources consulted mostly

	Frequency	Percentage
NOS	133	87,5
Nu.nl	68	44,7
De Volkskrant	86	56,6
De Telegraaf	16	10,5
Het AD	31	20,4
Facebook	54	35,5
Twitter	23	15,1
Trouw	26	17,1
Rtl Nieuws	24	15,8
Geen Stijl	4	2,6
NRC Handelsblad	59	38,8
Other	53	34,9

In addition, respondents were asked to indicate which news sources they use mostly. The results above show that the Dutch news channel NOS and the newspaper De Volkskrant are consulted by 87,5 and 56,6% of the sample. Constructing an overview of which news sources are used mostly, is necessary as it allows to review the social narrative and political perspective the sample of respondents hold. For instance, the Telegraaf is known to be a right-wing oriented newspaper, echoing news items from a specific perspective, whereas the NRC is known to be emphasizing the economic or financial aspect of societal issues. However, respondents were

asked to provide all answers that apply to their situation. Therefore, it is possible that multiple answers were chosen as individuals have consulted multiple sources to construct a balanced image of a news item.

4.5 Correlation

In the next section the results of the analysis concerning the dependent variable of truth rating and willingness to share variable will be provided. All test will be conducted and assessed based on a p-value of 0.05, meaning that if the significance level is <0.05, a difference can be confirmed, subsequently rejecting the H0. Firstly, the mean of the truth rating variable (1) and willingness to share variable will be provided, followed by a bivariate correlation analysis to explore the possible relation between the variables.

The results show a mean score of respondents on the truthfulness of the six news items of 56.22 (M) out of 100 (SD = 14.07). Secondly, the chance respondents would share any of the news items is 17.21 (M) out of 100 (SD = 17.94), which is seemingly low. Possibly this is a result of the sample of the research, as many people have indicated to be reluctant to share items online.

In addition, after running a bivariate correlation test the results show a weak positive relation between the truth rating and willingness to share news items ($p= 0.019$, $r=0.191$). This indicates that the higher the truth rating provided by the respondent, the higher the chance would be that they share the item.

To examine whether the incorporation of a warning affects the truth rating or willingness to share a news item (whether true or false), it is necessary to create a condition variable, by dividing the four experimental and control groups to; text without a warning (1), image and headline without a warning (2), text with a warning (3) and image and headline with a warning (4). In addition, two extra variables have been created to divide the effect of a warning (1) and the effect of visual or written content (2).

The sample contains of 37 respondents who have received only text without a warning, 37 who have received an image and headline without a warning, 40 respondents who have received text and a warning banner with false news items. 37 respondents have received an image and headline with a warning banner. The inequality between the groups, mainly the 3 additional respondents in the third group, potentially caused an issue for the reliability of the research. As this condition is overrepresented, it might cause the truth rating mean of this group to be higher or lower, which might result in a spurious relation.

4.6 Truth rating

Firstly, the results of the independent t-tests with on the warning (1) and type of content condition (2) compared to the truth rating mean will be provided, followed by the results of the one- and two-way ANOVA tests.

Surprisingly enough, the independent t-test comparing the groups with (3,4) or without a warning (1,2) shows a minor difference in the truth rating means. According to the research results of both Ecker et al. (2010) and Brashier et al. (2020) a different outcome would be expected, as a warning could stimulate the individual to rely less on fluency and pay more attention to the source and credibility of the item. The group conditioned with a warning has a mean of 54.64 (SD = 12.073), whereas the group without a warning has a mean of 57.87 out of 100 (SD = 15.793). Levene's test shows equal variances are assumed; F=2.349, p= 0.127. With a t (149) of -1.416 and a p-value of 0.159 the difference between the groups with or without a warning is not statistically significant, therefore the H0 cannot be dismissed.

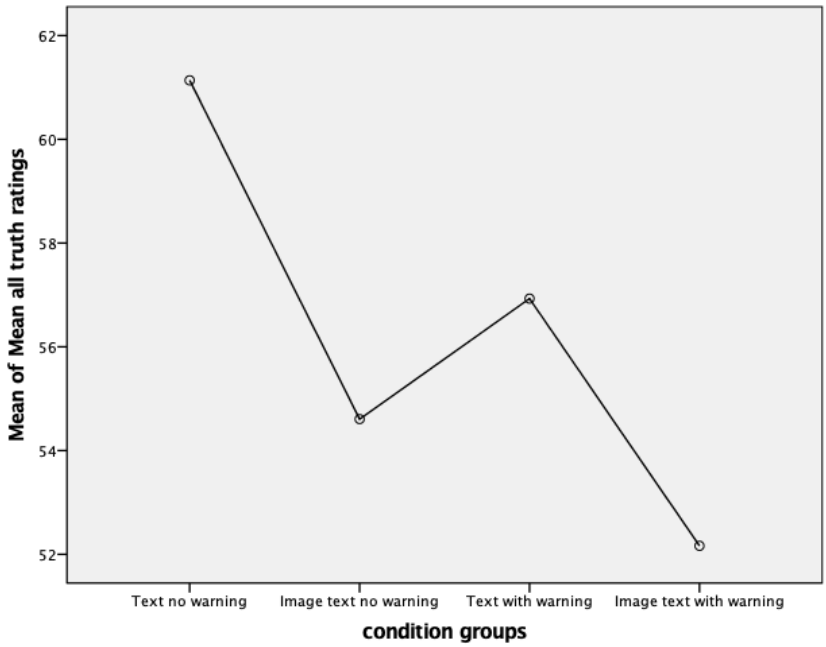
However, when the same tests are conducted to see whether there is a difference between the groups who have received an image and headline or text only, the results show a significance level of 0.015 (p); (F= 0.342, p=0.560, t (149) =2.472). Therefore, it can be concluded that in this case the H0 can be dismissed, meaning that there is a difference between these two groups that is statistically significant. This indicates that the truth rating of the group who has received the news items with visual content is lower, (M= 53.38, SD = 13.165) compared to the group who has received only text (M= 58.95, SD = 14.445).

The next analyses conducted to examine the effect of both independent variables on the truth rating is a one- and two-way ANOVA, testing for the main effects of the image and text and the warning variables compared to the dependent truth rating variable and the interaction effect between both independent variables. Firstly, the table below provides an overview of the means of the 4 groups, showing a minor difference between the truth rating means of all groups. The group who has received only text has the highest mean of 61,14 whereas the group conditioned with a warning and both image and headline has the lowest mean of 52,16. Secondly, the mean of the groups who have received a warning (Mtotal=54.64) is lower compared to the mean of the groups who have not received a warning (Mtotal=57.87), despite the difference in written or visual content they have received.

Table 8: Truth rating mean per group

Dependent Variable: Mean Truth Rating				
No Warning	Image and text	Mean	Std. Deviation	N
Warning	Text only	56,93	12,063	40
	Image and text	52,16	11,745	37
	Total	54,64	12,073	77
No Warning	Text only	61,14	16,534	37
	Image and text	54,61	14,507	37
	Total	57,87	15,793	74
Total	Text only	58,95	14,445	77
	Image and text	53,38	13,165	74
	Total	56,22	14,066	151

Before conducting the one-way ANOVA test, Levene's test shows that equal variances are assumed ($F = 1.052$, $p = 0.372$). The test will be conducted to compare the truth rating mean with the independent variables on the warning and content type conditions. The results of the one-way ANOVA test show a difference between the truth rating means of the four groups ($F(3, 147) = 2.829$, $p = 0.041$). A Tukey post hoc test conducted afterwards revealed that this difference is only significant for the group who have received text without a warning ($M=61.14$, $SD=16.53$) and the group who have received both an image, headline and a warning ($M=52.16$, $SD=11.75$) with a mean difference of 8.974 ($SE= 3.212$, $p=0.030$).



Graph 1: Tukey HSD: mean of all truth ratings with condition groups

Second, the two-way ANOVA test. First, Levene's test shows that homogeneity of variances between groups can be assumed as F = 1.052 (sig.= 0.372). When observing the results of the test between subjects, the main effect of the variable on the content type has a significant effect (F (1,147) =6.303, p=0.013), confirming the initial results of the second t-test. This implies that this variable (partially) explains the difference in effect on the truth rating of respondents. However, the interaction effect between image and text * warning or no warning (F (1,147) =0.154, p=0.695), and the main effect of the warning are not significant (F (1,147) =2.184, p=0.142). With an adjusted r^2 value of 0.35, 35 percent of the difference in effect on the truth rating can be explained by the variables adopted in this test, mainly the difference between text and image as only this effect is significant.

From the results of the t-tests and both one- and two-way ANOVA tests it can be concluded that only the independent variable on visual or written content partially explains the difference in the truth rating mean between the 4 groups. Surprisingly, the incorporation of a warning seems to have no effect nor can it explain the effect, as the results are not significant.

Therefore, the first and second hypothesis cannot be confirmed, as the warning condition does not provide an explanation for the differences. As a result, the illusory truth effect or credibility based on familiarity cannot be mitigated by the incorporation of a warning according to the results of this survey. However, the third hypothesis can be partially confirmed,

as the difference between groups who have received either visual or written content is significant. As it was expected for the condition group receiving visual content and a warning to show less mitigation of the incorporation of the warning compared to the group who would receive only text and a warning. This hypothesis cannot be confirmed, to the contrary, the effect seems to be the other way around. Exposure to visual content with a warning causes a lower, more accurate truth rating of the news items.

4.6.1 True and false news items

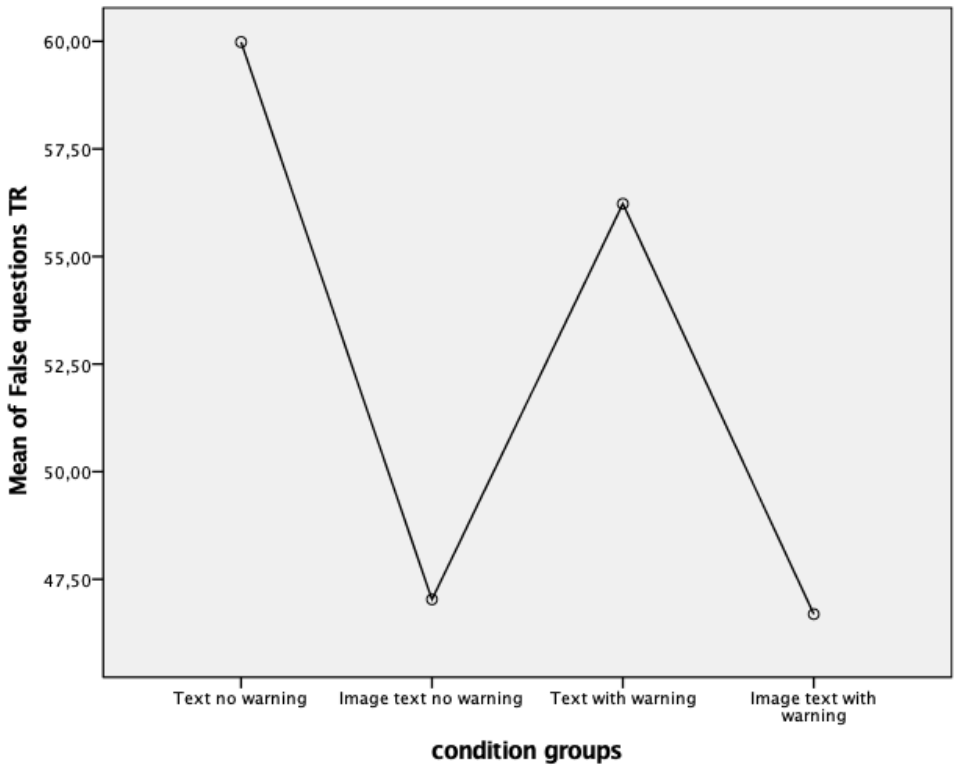
However, considering the fact that 3 of the questions asked in the survey were true and 3 were false, it is important to examine the difference between the groups in terms of their truth rating and accuracy between these two types of questions. If, for instance, the analysis shows a difference in the truth rating mean of the false questions that is significant between the groups and dependent variables, it could be concluded that it is indeed the dependent variable causing an effect on the assessment of information.

Therefore, for the initial tests, two variables have been created, categorizing the questions in the survey to either true or false. These variables have then been used to conduct an one- and two-way ANOVA analysis with the truth rating mean of the condition groups (1) and independent variables (2) and two independent t-tests with the variables differentiating between the group with or without a warning and between the groups who have received only text or an image with a headline.

4.6.2 False news items

After confirming the assumption of equal variances by conducting Levene's test ($F(3,146) = 0.124$, $p = 0.946$), A one-way ANOVA test reveals that there is a statistically significant difference between all four groups ($F(3,146) = 5.264$, $p = 0.002$). Meaning that the H0, on the absence of a difference between the groups in terms of their truth rating on the false questions, can be dismissed. A Turkey post hoc test only reveals a significant effect between two pairs of groups. Firstly, the group who received only text without a warning ($M = 59.98$, $SD = 17.70$) has a higher mean compared to the group who have received an image and headline also without a warning ($M = 47.03$, $SD = 18.40$) with a significance level of 0.011 ($p = 0.05$) with a mean difference of 12.95 ($SE = 4.123$). Secondly, the group 'image and headline with a warning' ($M = 46.96$, $SD = 18.48$) has a lower mean compared to the group 'text without a warning' with a significance level of 0.008 ($p = 0.05$) and a mean difference of -13.29 ($SE = 4.123$). Therefore, it

could be concluded that the independent variable of either written or visual content has an effect on the truth rating appointed by respondents on the questions that were false. The graph below indicated that the truth rating score of the respondents who have received only text is higher than the rating given by respondents who have received an image, headline and a warning. This might indicate that the additional image could have caused respondents to evaluate the content of the news item more thoroughly.



Graph 2: Tukey HSD with independent variable: mean of truth rating of false news items

Prior to conducting a two-way ANOVA, homogeneity of variances was assumed (F (3,147) = 0.277, p=0.877). Second, both main effects, warning (F (1,147) =4.549, p=0.035), content type (F (1,147) =6.278, p=0.013) and the interaction effect (F(1,147) =4.054, p=0.046) are significant. With an adjusted r^2 level of 0.72, these results show that the independent variables or main effects, explain the difference in the truth ratings of respondents in false news items for 72 percent. The interaction between the effect of a warning and the type of content one is exposed to, causes a difference in the level of truthfulness the respondent would rate the (false) news item. Therefore, both variables seem to cause respondents to assess the accuracy of the news items more thoroughly, as individuals who have received items with a warning or with an image have a lower truth rating mean on the false news items. Remarkably, the manipulation groups who have received only text (with or without a warning) have a higher mean overall of

M=59.98 (SD= 17.70, text no warning) and M=56.24 (SD=16.46, text with warning) compared to the groups who have received an image with (M=46.69, SD=18.37) or without a warning (M=47.03, SD= 18.39). In this case, when one would have assessed all the items correctly, the truth rating should be 0, therefore, the lower the truth rating, the more accurate one has assessed the items.

4.6.3 True news items

To compare the truth rating mean of false and true news items the table below will provide an overview of the means per category. This table reveals a difference in the minimum mean between the false and true items, as the mean of the false items is 0.00 whereas the minimum mean of the true items is 20.00. Secondly, the mean of the false items is almost 10 points lower (M=52.63, SD-18.46) compared to the mean of the true items (M=62.41, SD=15,21). This indicates that respondents have assessed the 'true' items with a higher truth rating (overall) compared to the false items.

After confirming the assumption on homogeneity of variances with Levene's test (F(3, 145)=0.987, p=0.401), a one- an two-way ANOVA has been conducted for the variable addressing the 'true' questions, comparing to the independent variables addressing the effect of a warning of content type. The one-way ANOVA reveals no difference between the four groups (F (3,145) =1.956, p=0.123). A two-way ANOVA test conducted next, reveals that no statistically significant effect can be detected of independent variable of the warning (F (1,145)=0.107, p=0.744) or content type (F (1,145)=3.647, p=0.058) nor is there an interaction effect detectable between the two main effects (F (1,145)=2.032, p=0.156). The adjusted r^2 of .19 shows that only 19% of the difference in mean between the groups can be explained with the main effect incorporated in this analysis, which is seemingly low (Bryman 2012).

Based on these observations, it can be concluded that the effect of the independent variable(s) on the truth rating score of the respondents is only applicable to the false news items. This indicates that the independent variable addressing written or visual content, causes the groups conditioned with only written content, to have a higher truth rating.

In conclusion, the results concerning the assessment of the items and the effect of a warning or content type, show a difference between two pairs of groups, namely, the group who have received text without a warning and the groups who have received visual items with or without the warning. Additional tests show a difference that is significant between the groups who have either received written or visual content, however this effect is not significant for the warning

condition. Therefore, it can be concluded that the truth rating score of respondents to the false news items differs depending on the type of content they have received, meaning the truth rating in visual items has been more accurate. The warning has had no effect. This confirms the first conclusion on the dismissal of the first two hypotheses and the partial confirmation of the third hypothesis.

4.7 Willingness to share

The fourth hypothesis in this thesis focuses on the willingness to share either true or false news items and whether this behavior could be mitigated or stimulated by the incorporation of a warning or visual content.

Despite the fact that the bivariate correlation analysis addressing the correlation between the truth rating and willingness to share, indicated a weak positive relation which was significant, the results of this analysis show a different result. Firstly, the M on willingness to share is 17.21 (SD=17.94), whereas surprisingly, the groups 'image without warning' (M=20.56, SD=20.55) and 'image with warning' (M=17.94, SD=17.05) have a higher M compared to the groups who have received text only (without warning M=14.91, SD=18.08 and with warning M=15.38, SD=16). However, the standard deviation is very high in all groups, meaning the spread of the sample is great which could indicate the reduced reliability of the mean. If the sample consists of many 'outliers', very low or very high scores, the mean value is less representative for the sample and greater population.

Before conducting one- and two-way ANOVA tests, the assumption on homogeneity of variances in confirmed by conducting Levene's test ($F(3,147) = 1.403$, $p = 0.244$). The one-way ANOVA test reveals no statistically significant difference between either of the four groups in terms of their willingness to share any of the items they were exposed to ($F(3,147) = 0.849$, $p = 0.469$). A two-way ANOVA using a univariate linear model reveals similar results, as the main effects of the warning ($F(1,147) = 0.164$, $p = 0.686$) and content type ($F(1,147) = 2.078$, $p = 0.152$) tested for in this model do not explain variation in the mean of the dependent variable of the willingness to share nor do they have an effect on each other ($F(1,147) = 0.323$, $p = 0.571$). With a negative r^2 of -0.03, it can be confirmed that the independent variables do not provide an explanation for the differentiating in respondents' willingness to share, to the contrary.

Therefore, these results lead to the conclusion that the willingness of the respondents to share any of the news items cannot be explained with either of the two independent variables, nor is there any difference in the outcomes of the sharing rate between the four groups. In conclusion,

the fourth hypothesis cannot be confirmed, as it seems that the respondent's willingness is not affected by the presence of a warning nor by the additional visual or written content.

4.7.1 True and false news items

Similar analyses have been conducted to explore the difference and correlation between the true and false news items. When comparing the means on the willingness to share between the true and false news items, it is evident that the willingness to share is increased in the truthful news items.

	False items (M=)	**SD**	**True items (=M)**	**SD**
Text without warning	14.83	18.86	15.37	19.05
Image and headline without warning	17.05	18.80	24.51	24.90
Text with warning	11.58	15.84	19.55	20.17
Image and headline with warning	15.99	16.81	19.88	20.04
Total	14.80	17.54	19.85	21.19

Table 9: Willingness to share for false and true news items

However, the total means of both the false (M=14.8) and true news items (M=19.85) are very low, indicating a lack of willingness to share any of the items, whether true or false. However, to see whether the difference in the willingness to share can be explained by the incorporation of a warning or content type, and whether this differs between items that are false or true, a one- and two-way ANOVA will be conducted in the next section.

Before conducting the first one-way ANOVA for the false news items compared to the willingness to share, results of Levene's test has allowed to confirm the assumption on equal variances of the variables ($F(3,147) = 0.853$, $p = 0.467$). The test reveals that there is no statistically significant difference between the mean on the willingness to share false items between any of the groups ($F(3,147) = 0.705$, $p = 0.550$). A univariate linear model with an adjusted r^2 of -0.06 confirms these results as this ANOVA test does not indicate a statistically significant effect of either of the two independent variables on the willingness to share false items ($F(1,147) = 0.046$, $p = 0.703$). In conclusion, a warning or visual/written content does not

affect nor explain the difference between the willingness to share among respondents in either of the four groups.

Secondly, when similar ANOVA test are conducted to assess the 'true' news items, again Levene's test confirms the assumption of equal variances between variables ($F(3,145) = 2.063$, $p=0.108$. A one-way ANOVA reveals no statistically significant difference between the willingness to share any of the true items ($F(3,145) = 1.139$, $p=0.335$). In addition, a factorial ANOVA test shows no statistically significant effect of either of the two independent variables, meaning these variables cannot provide an explanation for the differentiation in the willingness to share ($F(1, 145) = 1.613$, $p=0.206$). With an r^2 of 0.03, this univariate model has very little explanatory strength, as only 3% of the differentiation in the dependent variable can be explained with the independent variables adopted in the model.

This analysis confirms the first conclusion on the absence of a difference in the willingness to share if respondents are conditioned with either a warning or visual/written content. Sharing behavior is not affected by the incorporation of a warning, therefore, the results of this research confirm the conclusions drawn by Chen et al. (2015) on the limited effect of raising awareness on disputed content.

4.8 Sharing motivation(s)

Finally, in the survey respondents were asked to identify their motivation to share a certain news item, if they indicated wanting to share the item. During the process of data collection, feedback provided by respondents after which it was decided to condition this question, only individuals who have indicated that they want to share the item are asked to answer the question. However, already 50 people had filled out the survey, with the forced response to this question, therefore, the results of this variable could be biased and unequal between the four groups of analysis due to the changes in the question lay-out and conditionality. As a result, only descriptive statistics and the mode per question will be analyzed.

The table below provides an overview of the questions indicating the motivation for individuals to share an item, accompanied by the number of answers to these questions per block and the corresponding mode.

Block	**Questions**	**Number of answers per block**	**Mode(s) per question**
Text no warning	B1.2 to B1.12	129	4 and 6
Image and headline, no warning	B2.2 to B2.12	131	6
Text with warning	B3.2 to B3.12	126	4 and 6
Image and headline, with warning	B4.2 to B4.12	143	4, 6 and 9

Table 10: mode of motivation to share per group

After the model of Chen et al. (2015), the respondents were exposed to 10 possible motivations to share an item, of which the motivation: "I can express my opinion by sharing that information" (4) and "The information can be a good topic for conversation" (6) have been indicated most frequently in all four groups. These sentences are indicators for two motivations, status seeking, which would be indicated by number 4 and socializing, which could be indicated by the second line (6). These results show a different effect compared to the research conducted by Chen et al. (2015), as previous results indicate the main motivation for individuals to share items to be, to provide others with information and socialize.

Remarkably, the last experimental group who received an image and warning, seems to have an extra mode which indicating the motivation "The information seems accurate". This motivation, although not significant in the study conducted by Chen et al. was hypothesized to be an indicator for the activation of a source retrieval mode, meaning that the individual would have been stimulated to evaluate the source and accuracy of the item more thoroughly. Even though with this study it is not possible to analyze the statistic correlation due to the small N, the results do show a potential tendency and effect of a warning and visual content on the stimulation of an accuracy focus.

Conclusion

To examine the mitigating effect of a warning on the effect of illusory truth, all respondents were exposed to this effect in an experimental survey, dividing respondents into four categories. Respondents would receive either a written item without a warning, an image and headline without a warning, written items with a warning or headlines with additional image with a warning.

Based on theoretical and academic research it was hypothesized (H1 and H2) to find a difference in the accuracy of the truth rating between the groups who received a warning and those who did not. Secondly, it was expected to find a difference in the mitigating effect of the warning between visual and written items (H3). Visual items would show less mitigation and be less effected by the presence of a warning.

Finally, the willingness to share items was expected to be affected by the presence of a warning (H4). Awareness on the presence of disinformation was expected to result in less willingness to share.

When revising the results, the following observations can be made. First, the presence of a warning shows no significant effect. No difference in either the willingness to share nor the truth rating of false or true news items between the groups can be observed. Therefore, the first and second hypothesis can be dismissed.

Second, a significant effect between exposure to visual or written content can be observed. The truth rating score of respondents exposed to visual content is more accurate compared to the other groups, either with or without a warning. However, the third hypothesis will be dismissed, as the effect of the warning is not significant nor does exposure to visual content seem to mitigate the effect less compared to written items.

Finally, no difference can be observed in the willingness to share either true or false news items between the groups. Therefore, the fourth hypothesis can be dismissed, as awareness on the presence of disinformation has not affected sharing behavior. However, although not confirmed by statistical correlation, the motivation to share items seems to be different for the group who has received visual content and a warning. The focus seems to be on the accuracy of the content as a motivation to share, instead of emphasizing socialization or status seeking motives as were observed in the other three groups.

Limitations

To assess the reliability and generalizability of the results it must be taken into account that the research has a small N. The sample consists of only 152 people, which affects the representativeness for the larger population and reduced the reliability (Bryman 2012). If this research were to be repeated with a larger sample using random sampling the generalizability of the effects would could be better.

Second, measuring the effect of illusory truth is partly dependent on controlling for any previous knowledge respondents might have on a subject (Pennycook et al. 2018). This possible moderating variable has only minorly been taken into account, as it seemed difficult to control for this effect using an experimental online survey. There has been attempted to limit the effect of previous knowledge by selecting an assessing less well-known news items on less contemporary subjects, according to Fazio et al. (2015) their research method and building on their conclusions. If respondents had previous knowledge on the majority of the subjects, for instance on the presence of disinforming images or news items, the effect of the warning could be mitigated. Therefore, the internal validity might still be limited as a result.

Discussion

When examining the results of the research in the light of the academic debate concerning the spread of disinformation and creating tools for individuals to identify trustworthy sources of information, three observations can be made.

First, the accuracy of the truth rating of false or true news items is not affected by the presence of a warning tag. This could be a result of the small N in the research sample, as previous research has confirmed the effect of the warning on the mitigation of the illusory truth. Secondly, the linguistic characteristics of the instructions provided with the survey could have caused uncertainty about the purpose of the research. If individuals thought they were judged on the number of items they rated accurately and assumed the presence of trick questions, then the warning has exceeded its purpose. Disregarding the veracity and authenticity of the warning due to the illusory truth effect and reliance on fluency of retrieval could also have caused the effect, in this case, the illusory truth effect was not mitigated by the warning.

Second, the willingness to share is similar for true and false items, despite the applied warning. This could be explained by examining the motivation for individuals to share. Most respondents have indicated either wanting to share items to express their opinion or to socialize with others. Due to the emergence of memes, hoaxes and satirical media platforms it is possible,

individuals intentionally share items that seem unreliable, not to inform themselves or others but to socialize (Chen et al. 2015, Marwick and Lewis 2017). This would explain why a warning would make no difference in the willingness to share, as the motive to share is not focused on authenticity or informativeness. The risk of this new tendency is that disinforming items are still spread without the proper context, as one individual might recognize a piece of information as satirical, the other might not (idem 2017).

Third, a new effect was observed in this research although this was not hypothesized. Exposure to visual content in news items has caused individuals to rate the truthfulness of the item more accurately compared to written items. According to Hameleers et al. (2020), visual content would lower awareness and suspicion on mis-or disinformation, hence will result in a higher truth rating. However, the contradictory results in this research could be caused by the combination of an image and headline. This potentially stimulates individuals to assess whether the image matches the headline, hence this stimulates a more thorough assessment of the credibility of the content.

Recommendations

The conclusions of this research propose one important note for future research into this subject, namely, still very little is known about the effects of measures aiming to counter the dissemination and identification of mis-or disinformation. The emphasis should not only be on academic research into other tools than warning tags, but also on creating awareness on the functioning and effects of sharing disinforming items. As overreliance on the presumed functioning of the warning tag, which has been disproved in this study, could have negative consequences (Pennycook et al. 2020). Therefore, agency-based protocols in the assessment of information should be prioritized instead of relying on institutional authorities to verify the veracity of information. Examining the use of more visual content to create awareness and educate individuals on how to identify and assess of information could be a starting point for further research.

Bibliography

ABS-CBN news (2016). Facebook to flag 'disputed' stories to address fake news, ABS-CBN news: Quezon City, Philippines. Retrieved on 10th April 2020, https://news.abs-cbn.com/news/12/15/16/facebook-to-flag-disputed-stories-to-address-fake-news

Al-Rawi, A. (2019). Gatekeeping fake news discourses on mainstream media versus social media. *Social Science Computer Review*, *37*(6), 687-704.

Annemans, N. (2019) "*Waarom bevroren fruit en groenten vaak beter zijn dan verse"*, Algemeen Dagblad, retrieved on 24th April 2020, https://www.ad.nl/koken-en-eten/waarom-bevroren-fruit-en-groenten-vaak-beter-zijn-dan-verse~a603a502/

Armstrong, A. (2019). *Following Destructive Fire, Notre-Dame Cathedral Labeled 'At-Risk' by World Monuments Fund,* Art news: New York City, retrieved on 10th April 2020, https://www.artnews.com/art-news/news/notre-dame-at-risk-world-monuments-fund-13482/

Barkun, M. (2017). President Trump and the "fringe". *Terrorism and Political Violence*, *29*(3), 437-443.

Bradshaw, S., & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organized social media manipulation*. Working Paper 2019.3. Retrieved from The Computational Propaganda Project, Oxford University website: https://comprop. oii. ox. ac. uk/research/cybertroops2019.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054.

Bryman, A. (2012). *Social research methods*. Oxford University Press.

Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship*, *41*(5), 583-592.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238-257.

Di Marco, G. (2019). The illusory truth effect in the visual context: an exploration of potential applications.

Doornbos, H. (2014). "*Saudi photographer tells me: This is not Syria, it's not a grave and the boy's parents are not dead*", retrieved on 25th April 2020, https://haralddoornbos.wordpress.com/2014/01/17/saudi-photographer-telle-me-this-is-not-syria-its-not-a-grave-and-the-boys-parents-are-not-dead/

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & cognition*, *38*(8), 1087-1100.

EU East Stratcom Task Force (2020), "EU vs Disinfo database: The MH17 was shot down by a Ukrainian fighter", EUvsDisinfo, Retrieved from: https://euvsdisinfo.eu/report/an-australian-jurist-told-the-mh17-was-shot-down-by-a-ukrainian-fighter/

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Gigerenzer, G., and Todd, P.M., (1999). Simple Heuristics that make us Smart. Oxford University Press, New York.

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, *37*(2), 281-301.

Hamiti, U. (2016). The War of US against ISIS in Psychological Warfare and Internet as the New Frontline. *European Journal of Interdisciplinary Studies*, *2*(2), 129-134.

Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour.

Kennedy, P. J., & Prat, A. (2019). Where do people get their news? *Economic Policy*, *34* (97), 5-47.

Khan, M. L., & Idris, I. K. (2019). Recognise misinformation and verify before sharing: a reasoned action and information literacy perspective. *Behaviour & Information Technology*, *38*(12), 1194-1212.

Kumar, K. K., & Geethakumari, G. (2014). Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, *4*(1), 1-22.

Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in human behavior*, *28*(2), 331-339.

Londoño, E. and Andreoni, M. (2019). *'They Came to Kill.' Almost 5 Die Daily at Hands of Rio Police*. New York Times, retrieved on 24th April 2020, https://www.nytimes.com/2019/05/26/world/americas/brazil-rio-police-kill.html

Madden, A. D. (2000). A definition of information. In *Aslib Proceedings: new information perspectives* (Vol. 52, No. 9, pp. 343-349). Emerald Group Publishing Limited.

Marsh, E. J., Cantor, A. D., & Brashier, N. M. (2016). Believing that humans swallow spiders in their sleep: False beliefs as side effects of the processes that support accurate knowledge. In *Psychology of learning and motivation* (Vol. 64, pp. 93-132). Academic Press.

Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. *New York: Data & Society Research Institute*.

Medium (2018) *Chinese Railway station police are now wearing face recognition glasses to catch criminals*, retrieved on 24th April 2020, https://medium.com/shanghaiist/chinese-railway-station-police-are-now-wearing-face-recognition-glasses-to-catch-criminals-1d62c7686024

Merloe, P. (2015). Authoritarianism Goes Global: Election Monitoring vs. Disinformation. *Journal of Democracy*, *26*(3), 79-93.

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of pragmatics*, *59*, 210-220.

Newton, C. (2016). *Facebook partners with fact-checking organizations to begin flagging fake news*, New York: The Verge. Retrieved on the 10th of April 2020, https://www.theverge.com/2016/12/15/13960062/facebook-fact-check-partnerships-fake-news.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, *147*(12), 1865.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*.

Singer, P and Brooking, E. (2018). *LikeWar: The Weaponization of Social Media*, Boston: HMH Books.

Tsfati, Y., & Ariely, G. (2014). Individual and contextual correlates of trust in media across 44 countries. *Communication Research*, *41*(6), 760-782.

Watt, L. (2016), *Media manipulation*, World Press: Melbourne. Retrieved on the 10th of April 2020, http://www.mediafactory.org.au/lydia-watt/2016/05/05/media-manipulation/

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 218-226.

Appendix

Survey

Disinformation - illusory truth and accuracy focus

Start of Block: Intro

Q1.1
Thank you very much for participating in this study!

My name is Selma and for my master's program Crisis and Security Management at the University of Leiden I am conducting a survey on news consumption and behavior on social media.

The more responses I can gather, the more reliable the outcomes of the research will be, the more difference it can make. Naturally, all responses will be anonymized and kept strictly confidential. The responses will not be shared with third parties and will only be used for the purpose of this research. None of the information provided can be tracked back to your identity.

If you have any questions concerning this survey, the research purpose or any other topic, I can be contacted on the email address: s.ali.7@umail.leidenuniv.nl. Below you will be asked to provide your informed consent to participate in this study in the question below. You will be able to withdraw from this study at any time without consequences.

Thank you so much in advance!

I agree to participate in the research study. I understand the purpose and nature of this study and I am participating voluntarily. I understand that I can withdraw from the study at any time, without any penalty or consequences.

- ○ Yes (1)
- ○ No (2)

Skip To: End of Survey If Q1.1 = No

End of Block: Intro

Start of Block: Interests

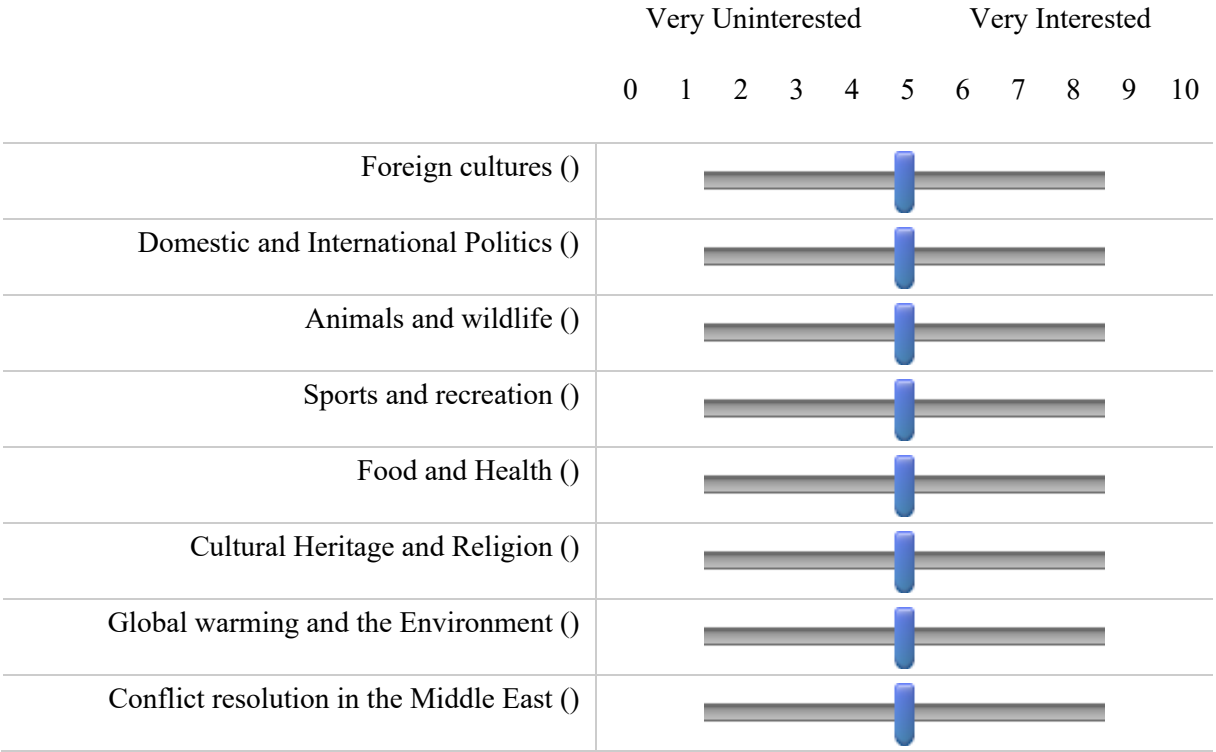
Mitigating the effect of illusory truth using a warning in visual and written news items

Q1.2

In the first part of the survey I would like to ask you some questions on your interests.

Please think carefully before you answer the question and enter the answer that applies to you the most.

How interested are you in...



Page Break

End of Block: Interests

Start of Block: Demographics news and social media

intro In the next section, a few questions will be asked about your demographic characteristics, news consumption and social media use.

Please enter the answer that applies to you the most.

Q1.3 What gender do you identify as?

- Male (1)
- Female (2)
- Other (3)
- Prefer not to say (4)

Q1.4 What is your age in years?

_____

Q1.5 What is the highest degree or level of school you have completed?

- Primary school (1)
- High school (2)
- MBO (3)
- HBO Bachelor/Master (4)
- WO Bachelor (5)
- WO Master (6)
- PhD (7)

Page Break

Q1.6 Which communication platforms or social media do you use most?

You can enter multiple answers.
Please enter **all** answers that apply to you.

- [ ] Facebook (1)
- [ ] Twitter (2)
- [ ] Instagram (3)
- [ ] WhatsApp (4)
- [ ] Snapchat (5)
- [ ] LinkedIn (6)
- [ ] Pinterest (7)
- [ ] TikTok (8)
- [ ] YouTube (9)
- [ ] Google+ (10)
- [ ] Other (11)

Q1.7 On an average day how much time do you spend on social media?

- ○ Less than 30 minutes (1)
- ○ 30 - 60 minutes (2)
- ○ 1 -2 hours (3)
- ○ 2 - 3 hours (4)
- ○ 3 hours + (5)

Q1.8 How often do you share news items on social media? (including WhatsApp or private messages)

- ○ Daily (1)
- ○ 4-6 times a week (2)
- ○ 2-3 times a week (3)
- ○ Once a week (4)
- ○ Several times a month (5)
- ○ Once a month (6)
- ○ Less than once a month (7)
- ○ Never (8)

Q1.9 Where do you get your news from?

You can enter multiple answers.
Please enter **all** answers that apply to you.

- ☐ Newspapers (1)
- ☐ TV (2)
- ☐ Websites (3)
- ☐ Radio (4)
- ☐ Apps (NOS, nu.nl etc.) (5)
- ☐ Social Media (6)
- ☐ Podcasts (7)

Q1.10 Which media sources do you use to gather the daily news?
You can enter multiple answers.
Please enter **all** answers that apply to you.

- ☐ NOS (1)
- ☐ Nu.nl (2)
- ☐ De Volkskrant (3)
- ☐ De Telegraaf (4)
- ☐ AD (5)
- ☐ Facebook (6)
- ☐ Twitter (7)
- ☐ De Trouw (8)

- [ ] RTL nieuws (10)
- [ ] GeenStijl (12)
- [ ] NRC Handelsblad (13)
- [ ] Other (14)

Page Break

End of Block: Demographics news and social media

Start of Block: Text no warning

intro
In this final section, you will be provided with a series of news items from different sources and countries. For each news item you will be asked three questions.

Page Break

text

Orphaned Syrian Boy Sleeping Between his Parents' Graves.

During the winter of 2017 a shocking news item of a young Syrian orphan lying between the graves of his parents, decorated with stones and flowers, has been spread across the world. This would be a result of the war Syria, which causes thousands of deaths and refugees every year.

Published by multiple sources on Twitter and Harald Doornbos World Press

QB1.1 On a scale of 0 to 100...

0 100

What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB1.1 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.2 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

‘They Came to Kill.’ Almost 5 Die Daily at Hands of Rio Police.

RIO DE JANEIRO — Shooting from helicopters, armored personnel carriers or at close range, police officers in Rio de Janeiro have gunned down 558 people during the first four months of the year — the highest number in this period since the state began keeping records more than two decades ago.

Published by Reuters and the New York Times

QB1.3 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB1.3 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.4 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

Mitigating the effect of illusory truth using a warning in visual and written news items

intro

Chinese railway station police are now wearing face recognition glasses to catch criminals

Just in case you were wondering why that railway cop wouldn't stop staring at you

Railway police in Zhengzhou are the first in China to use facial recognition glasses to screen passengers. Security personnel at Zhengzhou East Railway Station donned the new accessories ahead of the Chinese New Year travel rush to help verify passengers' identities, spot impostors—and even catch suspected criminals.

Published by the Huffington post and Medium

QB1.5 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB1.5 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.6 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

During the 2017 G20 meeting in Hamburg a serious-faced Putin was surrounded by other world leaders. This would have been the first time Trump and Putin met since Trump took office.

Published by the Daily Mail and on Twitter by multiple sources

QB1.7 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB1.7 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.8 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

Why frozen fruit and veggies may be better for you than fresh

By Lisa Drayer, CNN

Updated 0436 GMT (1236 HKT) May 31, 2019

Research has revealed that frozen fruits and vegetables can have just as many vitamins -- and sometimes more -- as compared to fresh. When you compare fresh string beans in a store versus frozen, frozen will almost always be higher in nutrient content.

Published by AD, CNN, Sciencetimes, posted on Facebook and Twitter by multiple sources

QB1.9 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB1.9 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.10 What would generally be the most important reason to share the item?

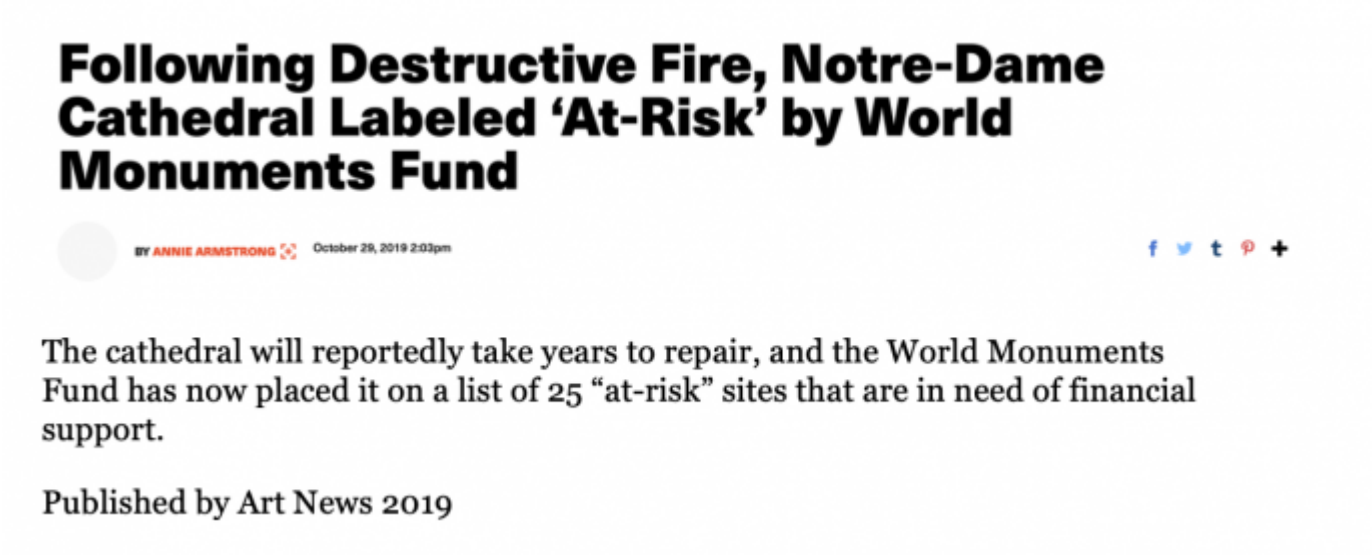
Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro



Mitigating the effect of illusory truth using a warning in visual and written news items

QB1.11 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: End of Block If QB1.11 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB1.12 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

End of Block: Text no warning

Start of Block: Text image no warning

intro
In this final section, you will be provided with a series of news items from different sources and countries. For each news item you will be asked three questions.

Page Break

intro

QB2.1 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB2.1 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.2 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB2.3 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB2.3 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.4 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB2.5 On a scale of 0 to 100...

	0	100
What is the chance that this news item is **true**? ()		
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()		

Skip To: intro If QB2.5 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.6 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB2.7 On a scale of 0 to 100...

	0 — 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB2.7 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.8 What would generally be the most important reason to share the item?

Please select **one** of the options below.

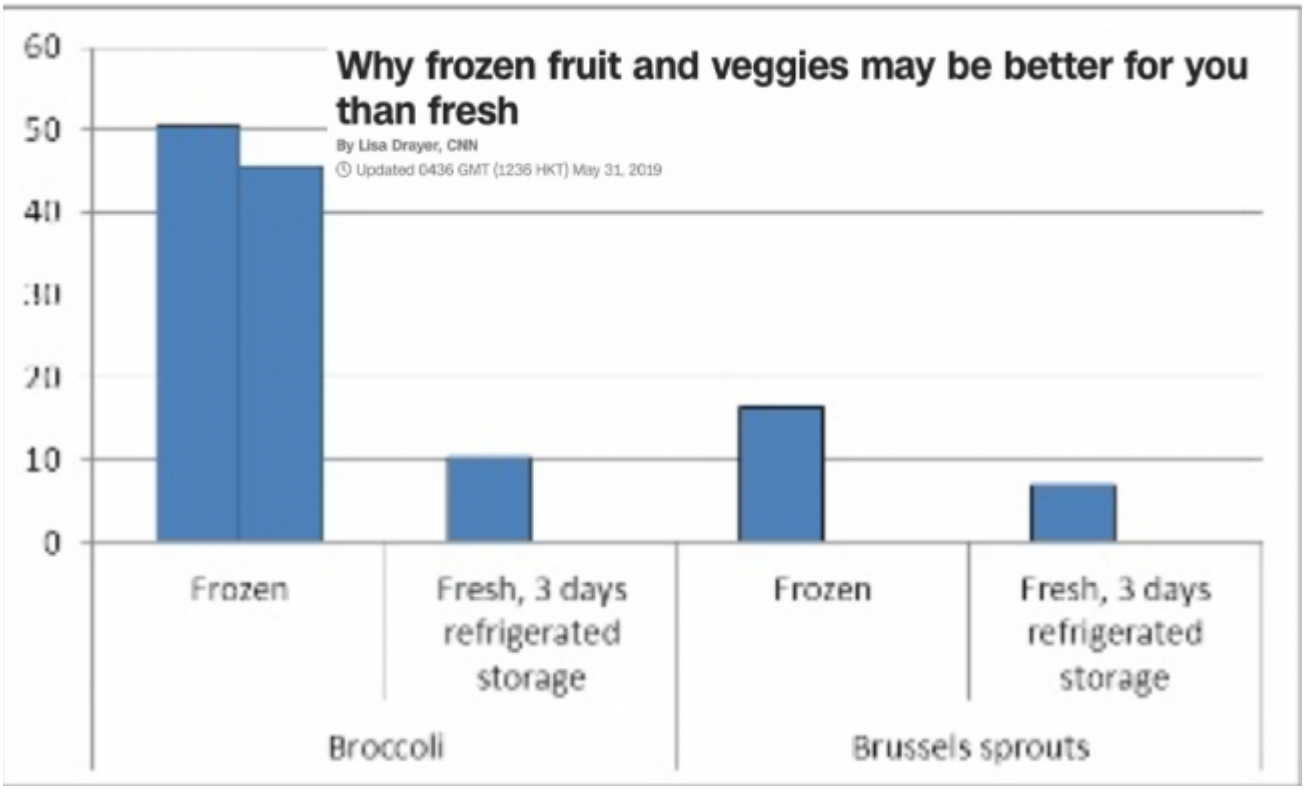
Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro



QB2.9 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB2.9 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.10 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB2.11 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: End of Block If QB2.11 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB2.12 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

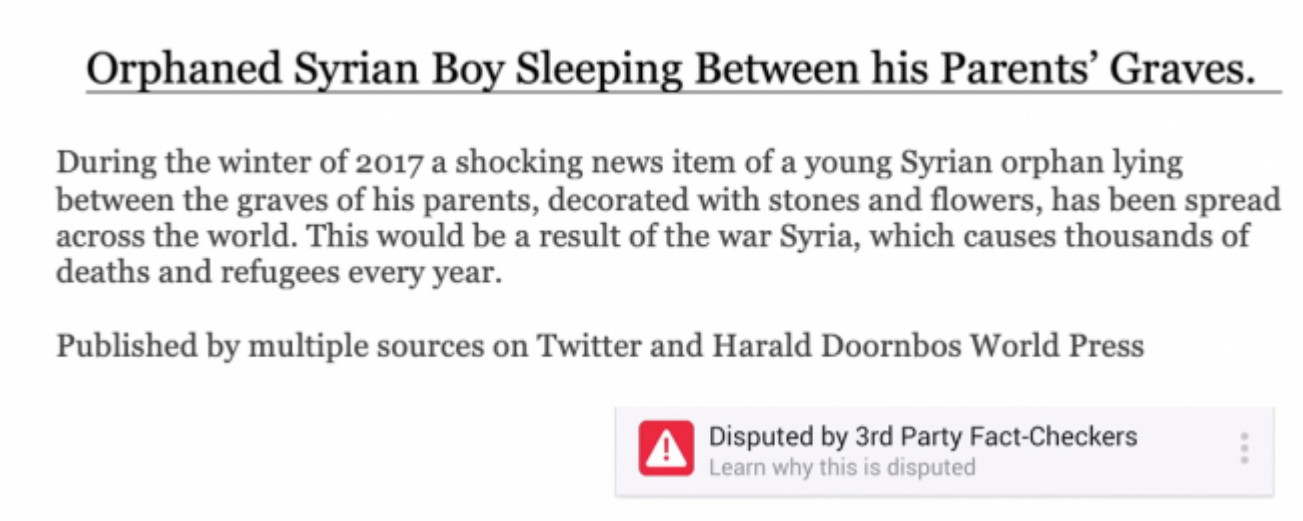
End of Block: Text image no warning

Start of Block: Text warning

intro
In this final section, you will be provided with a series of news items from different sources and countries. For each news item you will be asked three questions.

Page Break

intro



QB3.1 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB3.1 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.2 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

'They Came to Kill.' Almost 5 Die Daily at Hands of Rio Police.

RIO DE JANEIRO — Shooting from helicopters, armored personnel carriers or at close range, police officers in Rio de Janeiro have gunned down 558 people during the first four months of the year — the highest number in this period since the state began keeping records more than two decades ago.

Published by Reuters and the New York Times

QB3.3 On a scale of 0 to 100...

0 100

What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB3.3 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.4 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

Chinese railway station police are now wearing face recognition glasses to catch criminals

Just in case you were wondering why that railway cop wouldn't stop staring at you

Railway police in Zhengzhou are the first in China to use facial recognition glasses to screen passengers. Security personnel at Zhengzhou East Railway Station donned the new accessories ahead of the Chinese New Year travel rush to help verify passengers' identities, spot impostors—and even catch suspected criminals.

Published by the Huffington post and Medium

QB3.5 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB3.5 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.6 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

During the 2017 G20 meeting in Hamburg a serious-faced Putin was surrounded by other world leaders. This would have been the first time Trump and Putin met since Trump took office.

Published by the Daily Mail and on Twitter by multiple sources

QB3.7 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB3.7 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.8 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB3.9 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB3.9 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.10 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

Following Destructive Fire, Notre-Dame Cathedral Labeled 'At-Risk' by World Monuments Fund

BY ANNIE ARMSTRONG October 29, 2019 2:03pm

The cathedral will reportedly take years to repair, and the World Monuments Fund has now placed it on a list of 25 "at-risk" sites that are in need of financial support.

Published by Art News 2019

QB3.11 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: End of Block If QB3.11 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB3.12 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

End of Block: Text warning

Start of Block: Text image warning

intro
In this final section, you will be provided with a series of news items from different sources and countries. For each news item you will be asked three questions.

Page Break

intro

QB4.1 On a scale of 0 to 100...

	0 — 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB4.1 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.2 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB4.3 On a scale of 0 to 100...

	0 — 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB4.3 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.4 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB4.5 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB4.5 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.6 What would generally be the most important reason to share the item?

Please select **one** of the options below.

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB4.7 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB4.7 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.8 What would generally be the most important reason to share the item?

Please select **one** of the options below.

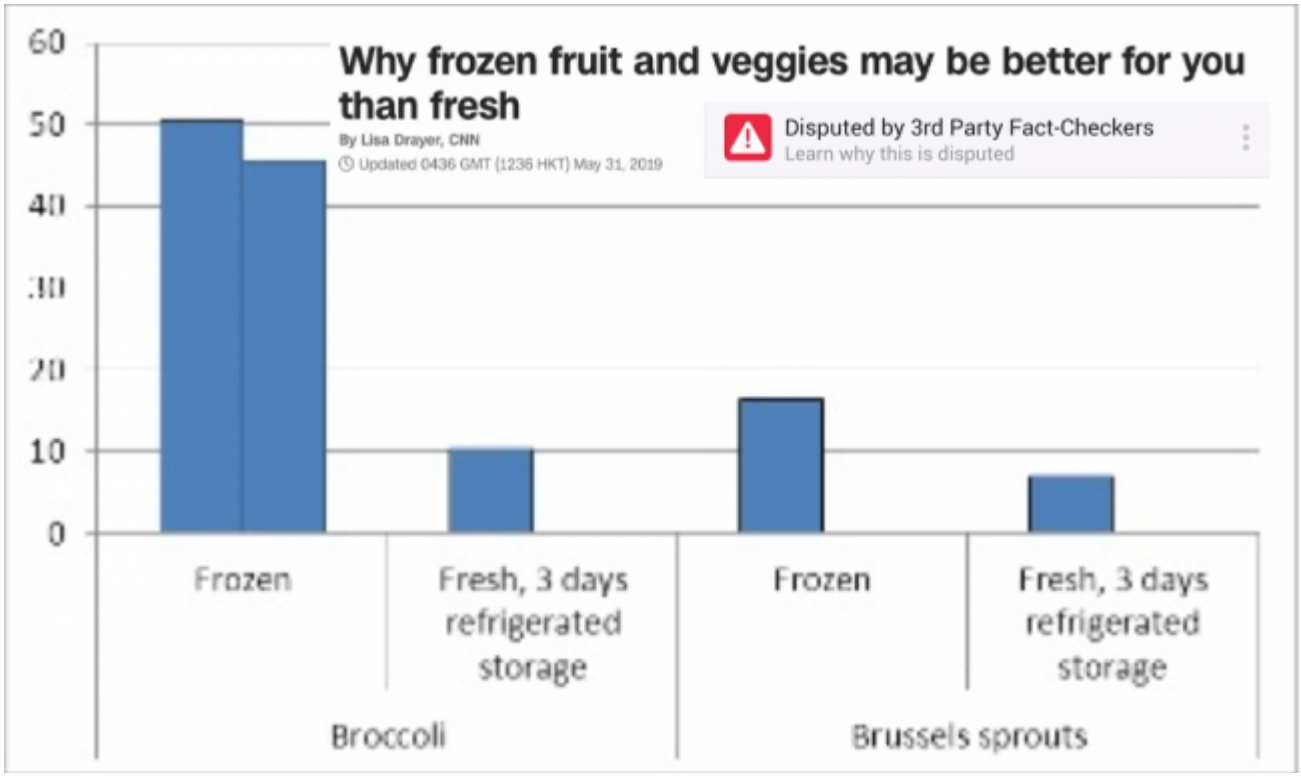
Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro



QB4.9 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: intro If QB4.9 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.10 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

Page Break

intro

QB4.11 On a scale of 0 to 100...

	0 100
What is the chance that this news item is **true**? ()	
How likely are you to **share** this item on social media? (including WhatsApp and private messages) ()	

Skip To: End of Block If QB4.11 [How likely are you to share this item on social media? (including WhatsApp and private messages)] <=

QB4.12 What would generally be the most important reason to share the item?

Please select **one** of the options below.

Mitigating the effect of illusory truth using a warning in visual and written news items

I would share this item because...

- ○ Sharing helps me **bookmark useful information** (1)
- ○ Sharing **helps me interact** with people (2)
- ○ Sharing is good for **keeping boredom away** (3)
- ○ I can **express my opinion** by sharing information (4)
- ○ Sharing **helps me get other people's opinions** regarding the information or event (5)
- ○ The information can be a **good topic for conversation** (6)
- ○ I feel **enjoyment** while sharing (7)
- ○ Sharing makes me feel **influential** (8)
- ○ The information seems **accurate** (9)
- ○ The information comes from **authoratitive sources** (10)

End of Block: Text image warning

Start of Block: End questions

end1 You have almost reached the end of this survey!

I would very much like to know if you have any comments, suggestions or complaints on the format or questions asked in the survey. If so, please enter your response below. If you do not have any comments you can click the button below after which you will be redirected to the end of the survey.

_____

end2 If you are interested in the results of this research you can provide your email address in below.

Your email address will not be used for purposes other than sharing the research results.

_____

Mitigating the effect of illusory truth using a warning in visual and written news items

Page Break

end3 Thank you very much for participating in this survey!

Before you finish the survey I would like to inform you that the purpose of this study is to examine how people react to authentic and fake news items. Therefore, the items you have just seen are a combination of fake and true news items. All items have been strongly adapted or even self-fabricated by adding or removing elements from the original news items and lack context.

Please don't forget to click the button below to save your response.

End of Block: End questions