Understanding the Role of Human Values in the Spread of Misinformation

Tracie Farrell
The Open University
tracie.farrell@open.ac.uk

Lara Piccolo
The Open University
lara.piccolo@open.ac.uk

Serena Coppolino Perfumi
Stockholm University
serena.perfumi@sociology.su.se

Harith Alani
The Open University
h.alani@open.ac.uk

Martino Mensio
The Open University
martino.mensio@open.ac.uk

Abstract

Social media platforms are often implicated in the spread of misinformation for encouraging the behaviour of rapid-sharing without adequate mechanisms for verifying information. To counter this phenomena, much related research in computer science has been focusing on developing tools to detect misinformation, to rank fact-check-worthy claims, and to understand their spread patterns, while psychosocial approaches have been focused on understanding information literacy, ideology and partisanship. In this paper, we demonstrate through a survey of nearly 100 people that the Human Values could have a significant influence on the way people perceive and share information. We argue that integrating a values-oriented perspective into computational approaches for handling misinformation could encourage misinformation prevention, and assist in predicting and ranking misinformation.

1 Introduction

Why do people share misinformation online and how can we prevent it? These questions have prompted an interdisciplinary response focusing on various aspects of the problem, such as how individuals become exposed to misinformation, who might be most vulnerable, and how to help news consumers make good judgements about what they see online.

Computer science approaches address a number of tasks that are important for fighting misinformation at scale, such as identifying deep fakes (Yang et al., 2018), analysing provenance and spread patterns (Wu et al., 2016), and assisting humans (e.g. fact-checkers and journalists) with **ranking and prioritisation** (Wiebe and Riloff, 2005; Cazalens et al., 2018). As a companion to technological and fact-checking approaches, **information literacy** is promoted as a mitigating force against misinformation and disinformation online (DiFranzo and Gloria-Garcia, 2017). The assumption is that, if the end-consumer or media professional understands how to view information critically, they will be less inclined to believe or transmit misinformation. However, research has shown that many people do not adequately fact-check, or even read past the headlines on social media (Allcott and Gentzkow, 2017). This signals that factors other than analytic reasoning or access to proper fact-checking resources may be involved in the spread of misinformation online. Furthermore, while research indicates that individuals with information literacy are less likely to share misinformation online and more likely to fact-check news items before sharing, studies show that information literate people are also less likely to share news items *in general* (Pennycook and Rand, 2018b). Sociological and psychological research has indicated that there are other motivations for engaging with misinformation, which can be emotional and ideological in nature (Jost et al., 2018; Bronstein et al., 2018). Several studies have emerged, in particular since the 2016 US Presidential Election, focusing on the role of ideology and belief in sharing misinformation online. However, this research is still somewhat contradictory and difficult to evaluate or to apply as strategy for interventions (Guess et al., 2019).

In our work, we aimed to investigate **Human Values** as psycho-social features that are possible to detect and predict computationally, and which provide a meaningful context for considering misinformation. In particular, we are interested in the following research questions:

- How are values visible in how individuals would share and respond to information?

- What kinds of patterns can be observed?

- How would consideration of human values complement other recommendations around information literacy or

computer science approaches to dealing with misinformation?

In this paper, we explore data gathered from 97 individuals recruited from within University and library networks. We discuss our collection and analysis of their statements relative to a selection of news headlines with ambiguous and unusual claims. Applying the *Theory of Basic Values* from Schwartz (Schwartz, 2012), we explore how human values may influence judgements about information and willingness to share. Finally, we argue that integrating human-values approach could encourage misinformation prevention, and assist computational approaches for predicting and ranking misinformation.

2   Background and Related Work

Social media platforms have provided a fertile ground for many different types of research questions about (mis)information. In the following section, we describe some of the work that has been done in different domains to get at **root causes** for misinformation.

Psycho-Social Approaches

Psycho-social approaches to understanding the root causes of misinformation generally focus on the **traits or characteristics** that may be correlated with sharing misinformation and **mechanisms for shaping behaviour**. For example, researchers have associated the characteristics of **extroversion** and **cooperativeness** with sharing misinformation (Zhu and Dong., 2010; Chen and Sin, 2013), as well as **dogmatism** and **religious belief** (Bronstein et al., 2018). Some of these traits are also believed to be demographic. Guess et al. (2019), for example, looked at the **age** and self-described **political affinity** of Facebook users and found that conservatives were more likely to share news from disreputable sources, but perhaps more significantly, that older users on Facebook were seven times more likely to share news from fake domains. Such studies, however, can be difficult to interpret. The same study by Guess et al. (2019), for example, determined that **older users were also more likely to share facts**.

Studies on the role of **partisan thinking** and misinformation have also had mixed results. Some studies show that conservatives share more misinformation (Grinberg et al., 2019; Ecker and Ang, 2019), while other studies have argued that this correlation may be related to other potentially confounding factors such as perceived bias in the media and in fact-checking organisations (Allcott and Gentzkow, 2017), or shared information processing tendencies of conservative *versus* liberal individuals (Harper and Baguley, 2019). Recent research by Harper and Baguley (2019) demonstrated that liberals and conservatives are equally vulnerable to believing misinformation, but for different reasons. The authors found that the greater the partisan attachment (on either side), the more willing individuals appear to be in engaging in "cognitive distortion" to protect their views.

In addition to contradictory ideas about vulnerability to misinformation, researchers also have different perspectives on how to tackle emotional or ideological factors in **misinformation correction**. Weeks (Weeks, 2015) found that certain emotional states can mitigate or exacerbate partisan thinking, but that ultimately receiving corrected information did improve "belief accuracy". Ecker and Ang (2019) explored this further in the context of so-called "backfire effects" and found that partisan thinking was most difficult to correct when it involved general misinformation, rather than information about a specific case, depending on how much the correction requires an adjustment of their attitude toward the issue at hand. Marietta and Barker (2019) suggested that **underlying values, rather than partisanship or domain knowledge**, are at the root of cognitive biases in the domain of misinformation. Jost et al. (2018) argue that these **values might make certain groups more susceptible to misinformation** and more likely to create the conditions under which it thrives (such as homogeneous networks and "echo chambers"). The study by Ecker and Ang (2019) largely supports this interpretation as well, though they also conclude that further research is necessary. Work of this kind makes an attempt to **understand intention in misinformation**, insofar as it relates to **preserving values**. If we can understand what an individual is trying to preserve in endorsing or refuting a particular claim, one can understand more about the motivations with which one engages in the production or spread of misinformation.

However, a **lack in analytic reasoning skills** is often suggested as the primary root cause for misinformation, even more than domain knowledge or other traits and characteristics. Analytic reasoning skills outperform other types of political or

domain knowledge in helping individuals to distinguish trustworthy from non-trustworthy content (Kahne and Bowyer, 2017), including when it supports or confronts a personal bias (Kahne and Bowyer, 2017; Pennycook and Rand, 2018a; Metaxas, 2018). However, analytic reasoning has been associated with a more general "unwillingness" to share news on social media (real or "fake") (Pennycook and Rand, 2018b). Our research allowed us to explore the roots of misinformation further, to see what other information processing activities might be taking place.

Computational Approaches

Computing has offered many contributions to battling misinformation, from **detection techniques** to studying **information cascades** (Fernandez and Alani, 2018). Reputation and credibility of news sources also plays a role in automatic detection of misinformation. Computational techniques have been developed for assigning **credibility scores** based on how often one produces or shares misinformation (Alrubaian et al., 2017), the **general credibility of publishers** (Ciampaglia et al., 2018) and the **persistence of that misinformation** within their network (Bradai, 2014). However, reputation-based classification is not sufficient for misinformation detection (Asr and Taboada, 2018). More recent work has explored detecting misinformation through **"content indicators"**, such as de-contextualised quotations, title representativeness and language use (avoiding 'clickbait'), logical fallacy and inference (Zhang et al., 2018).

One area in which computer science has had a fruitful partnership with human experts is with the process of **ranking 'claims' for fact-checking**. A claim is a statement that can be extracted from a document and fact-checked (Cazalens et al., 2018). Machine learning algorithms can now identify and sort claims within information, so that they can be prioritised for human review (Wiebe and Riloff, 2005; Cazalens et al., 2018). The ClaimReview[1] markup has become an open standard, making it possible for many fact-checking organisations to identify and publish, through structured data, what is actually being claimed, who made the claim and the verdict against the claim (according to fact-checkers). Stories can also be ranked according to their reach, such as in **rumour detection** (Zubiaga et al., 2016, 2018), as well as persistence in public

debate. For example, Dispute Finder[2] is a plug-in that identifies contested claims. UK fact-checking organisation, Full Fact[3], has also used data from the Open Data Institute[4] to extract topics of local interests that are regularly updated in the institute's databases. Improving tools that assist fact-checkers in this way can help them to anticipate early which stories might have the greatest impact on specific communities.

However, while automated techniques are an important aspect of computational approaches, there is some consensus that these techniques should be accompanied by **strong social science modelling** and human expertise to improve their value. On social media, facts are often modified with subjective statements in the commentary, which appear to have misinforming consequences, even when the information presented in the news article is correct (Anspach and Carlson, 2018). In addition, stories that have already been viewed many times (reach) over a considerable period of time (persistence) may have already had serious consequences (Cazalens et al., 2018). In her investigation of misinformation and disinformation detection approaches, Søe (2018) concluded that focusing on truth in misinformation detection may miss the mark. She argues that **it is not whether something is true or false, credible or not credible**, that is actually the distinguishing feature between information and misinformation. Rather, **intention and "misleadingness"** are the true values by which something can be judged as misinformation (Søe, 2018). This suggest the need for a stronger partnership between the computationals and social sciences to achieve a better understanding of how to identify and intervene in misinformation.

Human Values

Human values are understood as "the criteria people use to select and justify actions, to evaluate people (including the self) and events" (Schwartz, 1992). Influential studies on human values by Rokeach (1973), Hofstede (2009), Schwartz (1992, 2012), Kluckhohn and Strodtbeck (Hills, 2002) have all arrived at somewhat similar conclusions, that some values are shared universally, with varying degrees of influence over human understanding and behaviour. This suggests that it might be possible to find patterns in how individuals relate to in-

---

[1]https://schema.org/ClaimReview

[2]http://ennals.org/rob/archive/confront/Dispute$_F inder.html$
[3]https://fullfact.org/
[4]https://theodi.org/

formation and misinformation, by considering what might be important to them. Schwartz's theory, in particular, exposed *value relationships*, which highlight how clusters of values operate together. The "theory of basic human values" from Schwartz (see Figure 1) argued that there are 10 basic set of universal values guiding human behaviour, which can be represented in a circular structure, evidencing conflicts and compatibility among the ten values (Universalism, Benevolence, Conformity and Tradition, Security, Power, Achievement, Hedonism, Stimulation and Self-direction).
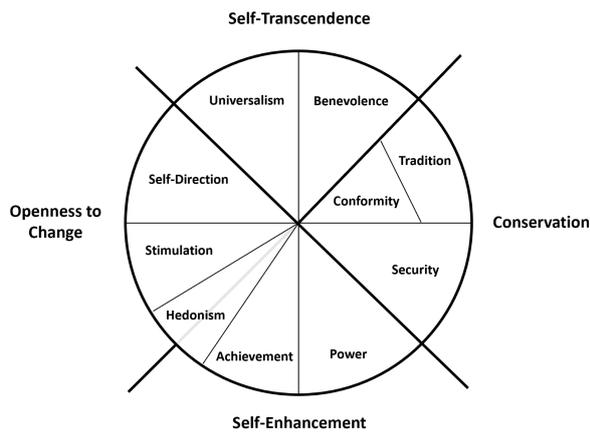
Figure 1: Human Values from Schwartz (Schwartz, 2012)

For Schwartz (2012) and Lyons et al. (2007), values can also be viewed on a dimension of **openness to change**. Universalism, Hedonism and Stimulation are very clearly on the open side of the wheel from Schwartz. Likewise, Conformity, Tradition, Security and Power are very clearly on the other.

Although the structure is universal, **individuals and social groups differ substantially in the relative importance they attribute to the values** (Schwartz, 2012). Although present in the judgement process, the impact of values in everyday decisions is rarely conscious, as values transcend specific actions and situations (Schwartz, 2012). As further described, in this study, we look at the human values as additional criteria beyond analytical reasoning for assessing information.

Computational detection of human values has been conducted typically through **"content analysis"**. This can be achieved at the word, phrase, sentence, multiple sentence, paragraph, or document level (Fleischmann et al., 2009). Takayama et al. (2014) applied a **word-scale probabilistic latent variable mode**l for detecting human values, using a combination of literature on human values

to annotate their training data, including a meta-inventory conducted by Cheng and Fleischmann (2010) for the purposes of easing automatic detection of human values in relation to Net Neutrality. Their resulting classifier achieves a level of accuracy similar to human annotators. Ishita et al. (2017) demonstrated how human values might "motivate sentiment" toward the Fukushima Nuclear Disaster. The authors annotated a corpus of 2,100 newspaper articles relating to the event, using a **handcrafted set of annotations related to human values**, and trained a classifier to detect those values. The authors do not indicate how the annotations were derived.

Only more recently, Human Values have also been explored **in conjunction with other demographic features** to understand how different individuals relate to misinformation. Verma et al. (2019), for example, found connections between valuing security, and trust in mainstream media. To our knowledge, detecting values has not yet become a primary part of the misinformation detection or prioritisation process. Still, the work above demonstrates that human values are possible to detect and relevant for the domain of misinformation. Moreover, it is possible to utilise this information to understand human reactions to complex topics, such as ethics, computationally. To this end, it is necessary the development of a strong set of training data reflecting scenarios in which quick value judgements might play a role.

3 Research Approach

For our research, we sought out a group of participants with similar experience and knowledge, to improve our ability to see what else (other than reasoning skills or domain knowledge) makes a difference in how individuals process and judge the information they see online. We also added a baseline exercise to check participants' reasoning skills with their ability to recognise a subjective from an objective claim. To simulate a social media environment, in which an individual might quickly browse headlines on a smart phone or other devices, we chose to present participants with short headlines and no additional data. We were less interested in participant success in identifying true stories, but rather their justification process. Therefore, we borrowed news headlines from the *Fake News Card Game*[5]. In the game, the objective is

---

[5]by The Takeover Game Ltd.

to convince others that fake news stories are true, so the stories are deliberately ambiguous to make assigning a truth value difficult. We chose 10 stories that involved a range of topics, from political themes to natural wonders.

Participants recruitment

We recruited participants from our network of library professionals, to ensure information literacy as described above. Participants were provided with a link in the recruitment notice to complete a questionnaire at any time within a one month period, after which we closed the questionnaire for future responses. They did not provide any personal or identifying information, except their perceived level of experience and confidence on information literacy. Participants were asked (and trusted) not to utilise the Internet in completing the questionnaire.

Research instrument

An online questionnaire[6] consisted of 2 main parts: the first one for assessing the veracity of a set of headlines and reflecting on their potential willingness to share the stories or not. The second part included a control test based on judgement *versus* observation to assess participants' analytic reasoning skills.

Part 1. Assessing Headlines

In part 1 of the questionnaire, participants were required to assign a truth value to each claim and justify their decisions. The stories provided to participants were:

1. Bearded London hipster mistaken for member of Isis and assaulted by nationalists

2. Man high on drugs rescues dog from imaginary house on fire

3. Nigerian restaurant serves human flesh

4. Fish survives six months without half its body

5. Neighbour from hell eats girl's guinea pig

6. Man allowed to board plane after bomb found in his baggage

7. NHS purchases gluten-free bread for £32.27 per loaf

8. In Switzerland, it is illegal to own only one guinea pig

9. The Bluegill fish is one of the most dangerous fish in North America. When the bluegill are feeding in a school, they can completely dismantle a human body in less than 15 minutes.

---

[6]The full questionnaire is available for view at: https://forms.gle/1yo3UexvNNM4ZQUF8

10. Britain has the highest rate of cocaine use among young adults in Europe. Their consumption is almost double that of other nations on the continent.

These particular stories were chosen because they did not have any overt political association, yet they do touch on themes such as drug use, national security, law and order, and the natural world. Five of the stories were true and five were false. The participants were informed of this to avoid the total scepticism in those with strong reasoning skills (Pennycook and Rand, 2018a), as noted earlier in this paper.

Participants were then asked to indicate: (i) which stories they would share if they were unable to fact-check them first; (ii) which they would share if they could fact-check them first. Once again, they provided a justification.

Part 2. Judgement *x* Observation

Part 2 of the questionnaire asked participants to determine from a set of 10 statements whether the statement is a judgement (subjective) or an observation (objective, fact-based).

Examples of these statements include:

- Cheryl is happy.

- Dominique is happy (with photo).

- President Trump lies a lot.

All of the statements provided were subjective judgements. We drew on definitions of objectivity and subjectivity from Letherby et al. (2012), in which judgements (subjectivity) are interpretations of facts we cannot know or determine with the information provided. 4 of the judgements were designed to resemble objective statements, but which included subjective cues, such as colours, imprecise measures of volume or size, and accompanying photographs. These items were the control items, intended to test whether the participant had carefully analysed each statement. A score >=7 on this exercise was considered a strong score, as it meant that a participant got at least one the control items correct.

Data Analysis

The first part of data analysis generated descriptive statistics about the population and how it performed on certain tasks. We identified some initial patterns in the data, which we then investigated further using qualitative methods, such as thematic analysis (Braun et al., 2019) of participant justifications and

comments. To add weight to this analysis, we calculated a 'demand for fact-checking', by looking at the increase in the number of participants who would share a story if it were fact-checked and determined to be true.

The analysis was centred around the *Theory of Basic Values* from Schwartz (2012) to annotate both the headlines we shared with participants and the participant's statements about those headlines. Annotations for headlines were provided by five raters (two of whom have a background in Schwartz's theory, and three with a background in social science and qualitative research). Annotations for comments were provided by two raters, both with a social science background. Categories were not mutually exclusive, meaning a rater could assign more than one value to a headline or comment.

To be able to calculate inter-rater agreement, it was necessary to simplify the categories after annotation (Kirilenko and Stepchenkova, 2016). Therefore, we used the dimension of openness to change, as per Schwartz (2012) and Lyons et al. (2007), to look at "groups" of values ("open", "closed" and "boundary"). Universalism, Hedonism and Stimulation were considered open. Conformity, Tradition, Security and Power are closed. Benevolence and Achievement were boundary values that we observed in connection with both open and closed value systems among our participants. We then calculated Krippendorf's Alpha (De Swert, 2012) with the resulting categories, as we used multiple different raters. Table 1 provides a description of the initial coding scheme for justifications and comments which reflects Schwartz's scale plus a code for references to 'plausibility'.

4 Results

97 participants took part in the study. Surprisingly, only 82.5% of participants answered that they felt information literate, regardless of their involvement in library services. However, we did not identify any notable differences (other than individual perceptions) in how these participants' responses.

Recognising Judgements vs. Observations

In part 2 of the questionnaire, most participants tended to struggle on items that involved determinations of size and volume (such as 'gigantic' or 'a lot'), and subjective judgements about emotion when provided with a picture. The statement

Code	Description	Example
Plausibility	statements expressing simple plausibility	"Sounds true." "Sounds plausible." "Probably true."
Self-Direction	statements implying independent thought	"I always thought…" "I remember this story…"
Achievement	statements implying respect for reasoning and being right	"The headline doesn't say when he was allowed to board…" "This relates to a contentious issue."
Universalism	statements implying care for social justice and nature	"Sounds like racist clickbait." "There is sometimes no limit to the ignorance of some groups."
Benevolence	statements about being kind to others and helping	"Guinea pigs get lonely. :(" [about guinea pig story]"I hope this is true!" "Scaremongering."
Security	statements implying concern with safety, danger and worry	"we have good security…" "I would have heard about this if it were true."
Conformity	statements about what is appropriate or usual behaviour	"[Not True Because]Time period (post 9/11)…" "False…due to science! And organs and stuff."
Tradition	statements about what is customary (handed down)	"Because it's [gluten free bread] too expensive!" "Sounds ridiculous." "Don't think anybody who is high on drugs would have the compassion to rescue a dog"
Hedonism	statements that imply an interest in enjoying life (also used for humorous comments)	"cause they tryna flex.." "With Freddie Starr's hamster as the dessert course?" "cocaine is not a ting boiii.."
Stimulation	statements about amazement and wonder in learning	"Nature is often astounding.." "Nature is amazing!"
Power	statements about social control, social power and trust in authorities	"He would have been arrested…" "wouldn't happen in these times…"

Table 1: Description of Values Coding Scheme

"Cheryl is happy" was correctly identified as a judgement by 88% of participants. The statement "Dominique is happy", accompanied by a photograph of a woman smiling was correctly identified as a judgement by only 45% of participants[7]. We did not find a strong relationship between performance on the judgement *versus* observation exercise and performance in assigning a correct truth value for a news item. However, participants who scored 90-100% on either exercise never scored below 70% on the other. This indicates that there may be some slight benefit, with more skill or experience, to having more **finely tuned reasoning skills** when confronted with limited information (such as a headline).

Justifications with Limited Context

In 5 cases, the number of participants assigning a value of true or false was noticeably unbalanced. Headline 3 received a correct assessment of 'false' by most participants (91%). When participants had to justify the reasons for their choice, participants expressly mentioned the perceptions of racism (n=15) and the motivations one might have for making controversial statements about a sub-Saharan African country (n=3). Other justifications included that the headline sounded sensationalist and unlikely, or hope that the story was not true (n=9). Only 6 participants said they would share this story without fact-checking it, and this number

---

[7]projection of internal states is subjective (Letherby et al., 2012)

increased to only 14 participants, who said they would share this story if they could fact-check it. Headline 4 was identified correctly by 72% of participants as true. The justifications provided by the participants were, for example, related to what the participants felt they already knew about the natural world (n=33) and the perceived plausibility of the claim (n=11). Compared to other stories, more participants would choose to share this story both before and after fact-checking. 72% of participants incorrectly identified Headline 6 as false. Participants overwhelmingly had difficulty imagining such a story could be true, according to their justifications, in light of security concerns worldwide. Their justifications included, for example, trust in secure borders (n=14), or that it simply wasn't likely given the events of the past 8-10 years (n=10). Some participants, who correctly identified the story as true, mentioned that they could imagine additional context that might make the story correct, such as a bomb that was not live or a bomb taken on by mistake (n=4). 61% of the participants correctly identified Headline 8 as true. This story was also among those most likely to be shared by participants both before (n=51) and after fact-checking (n=37). However, fewer participants would share it after fact-checking. Finally, 71% of participants correctly identified Headline 10 as true. This was the fourth story that was most likely to be shared both before (n = 33) and after fact-checking (n=33).

Figure 2 shows a breakdown of how many times certain justifications were given by participants in choosing why they might share a certain story.
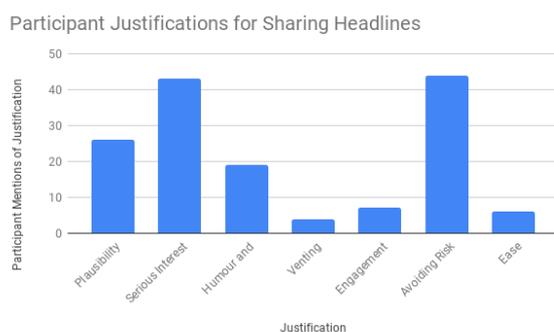
Figure 2: Participants' justifications for sharing or not sharing headlines

For our participants group, serious interest in the topic and avoidance of risk or consequences were the two most commonly given justifications for why a participant made their choice to share or not share a given story.

Values Analysis

Table 2 shows how we mapped participant justifications for sharing or not sharing a headline to the values scheme from Schwartz. In this table, FC represents the 'demand for fact-checking' on the story, as calculated by how fact-checking influences whether or not a participant would then share a given story. The items marked in grey have the greatest demand for a fact check. One interesting result from this is that the story of a Nigerian restaurant serving human flesh has a high demand for fact-checking, but a low interest in sharing, even after fact-checking. Desires to share the story about the neighbour who allegedly ate a pet guinea even reduce after fact-checking. This indicates that the importance of the truth of the story may be separate from a desire to share.

Headlines	Would share	Fact-check for sharing	FC	Values it challenges Schwartz [33]
Guinea pig	51	37	-14	Universalism (nature); Benevolence (caring); Power (control)
Fish survives	47	34	-13	Stimulation (learning, novelty)
Rescue dog	38	26	-12	Universalism (nature); Benevolence (caring); Conformity (compliance with social norms)
Cocaine	33	33	0	Universalism (societal concern)
Bluegill	29	29	0	Stimulation (learning, novelty)
NHS	22	35	13	Power (control of material resources); Universalism (societal concern); Conformity (violate social expectations)
Hipster	19	33	14	Conformity (inclinations likely to harm others); Security (personal, societal security) ; Tradition (respect, acceptance); Universalism (societal concern)
Bomb	15	24	9	Security; Conformity (inclinations likely to harm others); Universalism (societal concern)
Neighbour from hell	14	11	-3	Conformity (inclinations likely to harm others); Security; Universalism (nature)
Nigerian	6	14	8	Conformity (inclinations likely to harm others); Tradition (respect, acceptance); Universalism (Tolerance);

Table 2: Values Analysis

Stories that create a demand for fact-checking appear to be most connected to values of Conformity and Universalism, two values that are quite far apart on Schwartz's wheel. Conformity is one of the closed values that is related to "Conservation" in Schwartz's theory, while Universalism is a more open value related to "Self-Transcendence". Figure 3 shows the "value profile" of each headlines, as assigned by our 4 raters according to the values they felt could be triggered or challenged. For some headlines, the "value profile" is more stable, such as stories 4, 7 and 8. In others, raters saw the potential for many different values to play a role, for example, with story 6. This is partly to do with the non-exclusivity of the codes and also the different interpretations of values that are quite close together on Schwartz's wheel, such as Con-

formity and Tradition, or Universalism and Stimulation. This continued in the analysis of values present in participants' comments about headlines.
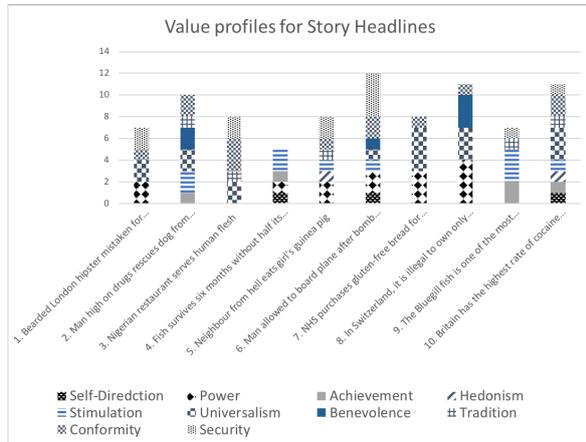
Figure 3: Values Profiles for Headlines

Figure 4 connects participant scores on the baseline, with the values annotators assigned to their statements about a given story. High performers on the analytic baseline made more statements related to Self-direction and Achievement than low performers. Low performers made more statements related to Security, Conformity, Tradition, and Hedonism. As described earlier, we also simplified the 10 values from Schwartz into categories of "open", "closed" and "boundary" values for a more general analysis of participant comments [8]. For this analysis, categories were exclusive and Krippendorf's alpha was calculated at .886 overall.

One can see from the simplified categories that the basic trend is preserved, in which the boundary value of Achievement (in particular with regard to the respect for intelligence and reasoning) is connected with better performance overall in both exercises in the questionnaire. This may explain the role of analytic reasoning from a values-based, rather than a competency-based approach.

We also compared the values assigned to participant statements based on those who believed a given story was true compared with participants who did not believe the story was true (see Figure 6 and 7). One can see how values of security are more present in participants that incorrectly identified the story about the bomb on the plane as false. Likewise, values of Universalism are more present
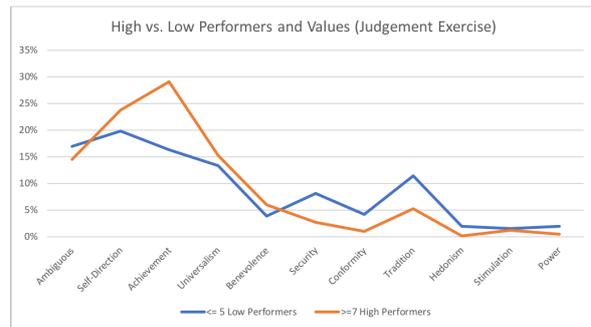
Figure 4: High and Low Performers on the judgement exercise
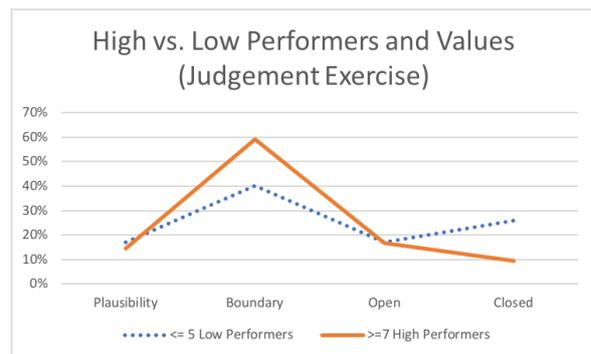
Figure 5: High and Low Performers on the judgement exercise (simplified)

in those that incorrectly identified the story about a hipster mistaken for ISIS as true.
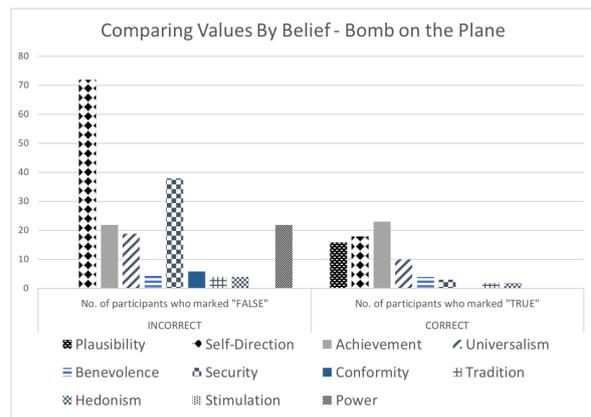
Figure 6: Comparing values by belief - Security

We can observe that certain value schemes do emerge in relationship to the stories, especially with respect to maintaining a generally open value system or a generally closed system, with the boundary values acting as modifying components, or as general ambiguity or cautiousness in making a judgement.
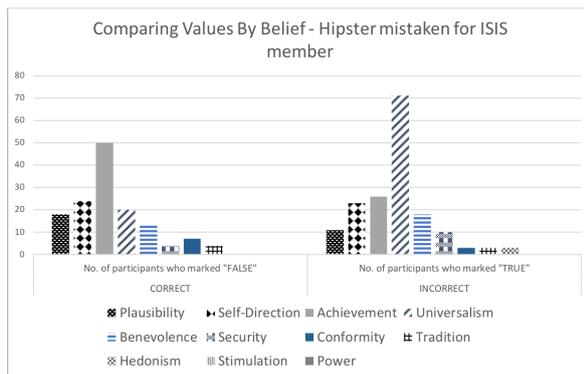
Figure 7: Comparing values by belief - Universalism

5 Discussion

Our findings have suggested that values, in addition to the truth, impact which information an individual will deem worthy of sharing. In some cases, values even outweigh the truth in the decision to share. As such, understanding how news stories signal values to readers and how readers process those signals is an important aspect of misinformation detection and correction.

Our study suggested that "closed" values of Security, Conformity and Tradition were associated with greater difficulty both in assigning the correct truth value to a story and recognising the difference between subjective and objective statements. When progress and tradition are at loggerheads, as might be considered the case with topics such social justice or climate change, one could expect values to confound our judgement over the "facts". This is important because much of what we might encounter online, especially on social media, may include subjective claims that frame facts in misleading ways, or which include false information. Our findings indicate that individuals guided by more "closed" values may be more vulnerable to misinformation. More research is needed to understand the interplay between competency, interest, and emotion in the development and expression of values. This finding does, however, correlate with other studies (Ecker and Ang (2019) and Harper and Baguley (2019), for example) that explore liberal and conservative responses to misinformation.

Limitations and Future Work

We recognise several limitations that impact our research, some of which present additional opportunities for future work. For example, we could

not verify that participants did not use the internet (though their responses do not indicate that they did). In addition, our participant group was fairly homogeneous, by design, to be information literate. However, they are all library professionals, which may attract participants with a certain type of value system. We saw a high representation of values associated with Self-transcendence, such as Universalism and Benevolence. In our study, we also found that our participants tended to be more keen to share claims that do not challenge their values or carry a lot of risk. Such findings fit the participant pool, but would be interesting to test with other types of professional groups, or groups that are more interested in conflict (e.g. trolls, conspiracy theorists).

In addition, this work is descriptive and qualitative. We plan to take this work further to examine our results in more detail. In our related work, we have already described some approaches for detecting human values. In the immediate future, we will conduct a larger scale human annotation exercise to get more precise results for both images and text. We will also develop and train a classifier. The completion of this exercise would allow us to compare our results with those of studies using the meta-inventory analysis conducted by Cheng and Fleischmann (2010).

6 Conclusions

In this paper, we have argued that values play an important role in the way (mis)information is perceived and the choices people make about sharing information online. Outlining various approaches from psychology, sociology and computer science, we demonstrated how the analysis of human values in misinformation detection could be beneficial. Our research indicates that strategies to fight misinformation should be informed by the current practices, expectations and values of the consumers of information, to help understand what matters to them and what drives their decision-making. Knowing what values are being protected in contentious topics could help provide clues for resolving the **root causes of misinformation**. In addition, it can help to **anticipate surges in misinformation** related to issues that reflect deeply ingrained values.

References

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of*

Economic Perspectives, 31(2):211–36.

Majed Alrubaian, Muhammad Al-Qurishi, Mabrook Al-Rakhami, Mohammad Mehedi Hassan, and Atif Alamri. 2017. Reputation-based credibility analysis of twitter social network users. *Concurrency and Computation: Practice and Experience*, 29(7):e3873.

Nicolas M Anspach and Taylor N Carlson. 2018. What to believe? social media commentary and belief in misinformation. *Political Behavior*, pages 1–22.

Fatemeh Torabi Asr and Maite Taboada. 2018. The data challenge in misinformation detection: Source reputation vs. content veracity. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 10–15.

Amira Bradai. 2014. *Secured trust and reputation system: analysis of malicious behaviors and optimization*. Ph.D. thesis, Institut National des Télécommunications.

Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic analysis. *Handbook of Research Methods in Health Social Sciences*, pages 843–860.

Michael V Bronstein, Gordon Pennycook, Adam Bear, David G Rand, and Tyrone D Cannon. 2018. Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*.

Sylvie Cazalens, Philippe Lamarre, Julien Leblay, Ioana Manolescu, and Xavier Tannier. 2018. A content management perspective on fact-checking. In *" Journalism, Misinformation and Fact Checking" alternate paper track of" The Web Conference"*.

Xinran Chen and Sei-Ching Joanna Sin. 2013. 'misinformation? what of it?'motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–4.

An-Shou Cheng and Kenneth R Fleischmann. 2010. Developing a meta-inventory of human values. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*, page 3. American Society for Information Science.

Giovanni Luca Ciampaglia, Alexios Mantzarlis, Gregory Maus, and Filippo Menczer. 2018. Research challenges of digital misinformation: Toward a trustworthy web. *AI Magazine*, 39(1).

Knut De Swert. 2012. Calculating inter-coder reliability in media content analysis using krippendorff's alpha. *Center for Politics and Communication*, pages 1–15.

Dominic DiFranzo and Kristine Gloria-Garcia. 2017. Filter bubbles and fake news. *XRDS*, 23(3):32–35.

Ullrich KH Ecker and Li Chang Ang. 2019. Political attitudes and the processing of misinformation corrections. *Political Psychology*, 40(2):241–260.

Miriam Fernandez and Harith Alani. 2018. Online misinformation: Challenges and future directions. In *Companion Proc of the The Web Conf 2018*, WWW '18, pages 595–602, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Kenneth R Fleischmann, Douglas W Oard, An-Shou Cheng, Ping Wang, and Emi Ishita. 2009. Automatic classification of human values: Applying computational thinking to information ethics. In *Proceedings of the 72nd Annual Meeting of the American Society for Information Science and Technology*, pages 6–11. Vancouver, Canada.

Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586.

Craig A Harper and Thom Baguley. 2019. "you are fake news": Ideological (a)symmetries in perceptions of media legitimacy.

Michael D Hills. 2002. Kluckhohn and strodtbeck's values orientation theory. *Online readings in psychology and culture*, 4(4):3.

Geert Hofstede. 2009. Geert hofstede cultural dimensions.

Emi Ishita, Toru Oga, An-Shou Cheng, Kenneth R Fleischmann, Takayama Yasuhiro, Douglas W Oard, and Yoichi Tomiura. 2017. Toward automating detection of human values in the nuclear power debate. *Proceedings of the Association for Information Science and Technology*, 54(1):714–715.

John T Jost, Sander van der Linden, Costas Panagopoulos, and Curtis D Hardin. 2018. Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, 23:77–83.

Joseph Kahne and Benjamin Bowyer. 2017. Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, 54(1):3–34.

Andrei P Kirilenko and Svetlana Stepchenkova. 2016. Inter-coder agreement in one-to-many classification: fuzzy kappa. *PloS one*, 11(3):e0149787.

Gayle Letherby, John Scott, and Malcolm Williams. 2012. *Objectivity and subjectivity in social research*. Sage.

Sean T Lyons, Linda Duxbury, and Christopher Higgins. 2007. An empirical assessment of generational differences in basic human values. *Psychological reports*, 101(2):339–352.

Morgan Marietta and David C Barker. 2019. *One Nation, Two Realities: Dueling Facts in American Democracy*. Oxford University Press.

Panagiotis Metaxas. 2018. Technology, propaganda, and the limits of human intellect. *arXiv preprint arXiv:1806.09541*.

Gordon Pennycook and David G Rand. 2018a. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*.

Gordon Pennycook and David G Rand. 2018b. Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking.

Milton Rokeach. 1973. *The nature of human values*. Free press.

Shalom H Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier.

Shalom H Schwartz. 2012. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):11.

Sille Obelitz Søe. 2018. Algorithmic detection of misinformation and disinformation: Gricean perspectives. *Journal of Documentation*, 74(2):309–332.

Yasuhiro Takayama, Yoichi Tomiura, Emi Ishita, Douglas W Oard, Kenneth R Fleischmann, and An-Shou Cheng. 2014. A word-scale probabilistic latent variable model for detecting human values. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1489–1498. ACM.

Nitin Verma, Kenneth R Fleischmann, and Kolina S Koltai. 2019. Understanding online trust and information behavior using demographics and human values. In *International Conference on Information*, pages 654–665. Springer.

Brian E Weeks. 2015. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, 65(4):699–719.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 486–497. Springer.

Liang Wu, Fred Morstatter, Xia Hu, and Huan Liu. 2016. Mining misinformation in social media. *Big Data in Complex and Social Networks*, pages 123–152.

Xin Yang, Yuezun Li, and Siwei Lyu. 2018. Exposing deep fakes using inconsistent head poses. *arXiv preprint arXiv:1811.00661*.

Amy X Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, et al. 2018. A structured response to misinformation: Defining and annotating credibility indicators in news articles. In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 603–612. International World Wide Web Conferences Steering Committee.

Chuansheng Chen Elizabeth F. Loftus Chongde Lin Qinghua He Chunhui Chen He Li Robert K. Moyzis Jared Lessard Zhu, Bi and Qi Dong. 2010. Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, 48(8):889–894.

Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2):32.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tomie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989.