

MASTER THESIS
INFORMATION SCIENCES



RADBOD UNIVERSITY

How do authenticity methods affect perceptions of tweets?

Design, execution, and qualitative analysis of a case study on user perceptions of microblog posts under different authenticity methods

Author:

K. Verdenius (Koen)
koen.verdenius@ru.nl

First supervisor:

dr. B.E. van Gastel (Bernard)
b.vangastel@cs.ru.nl

Second supervisor:

dr. H.K. Schraffenberger (Hanna)
hanna.schraffenberger@ru.nl

March 19, 2023

Abstract

The prevalence of online falsehoods has led to a multitude of misinformation mitigation strategies. The development of these strategies often lacks input on user experience. Authenticity methods are such a strategy to allow users to confirm an author's identity on social media. An example is Twitter's pre-subscription Verification, which hopes to prevent impersonation. Another is a new proposed method, Twid, that allows users to sign messages with author attributes. Twid seeks to achieve goals akin to Verification and provide additional author information. By designing, executing and analyzing an exploratory case study on user perceptions of authenticity methods, we contribute to addressing the shortcoming in understanding user comprehension of authenticity methods. Fourteen participants provided perceptions on messages through the think-aloud method and semi-structured interviews. We analyzed broad perceptions, and those on perceived credibility, authenticity and interactions using qualitative thematic analysis. In our findings, we try to encompass the underlying concepts and functionality of the two tested methods to produce concluding hypotheses generalizing away from the specifics of the study. Participants were mostly modestly positive about both methods tested, and both methods more so than not positively impacted the perceived credibility and authenticity of authors. However, both methods risk not achieving their intended usage advantages, may be vulnerable to abuse by authors, and pose a risk of unwittingly misleading users. Our further findings were a perceived positive impact of the notability of an author on their credibility, subjective phrasing in messaging positively increasing author authenticity, and the interrelated experienced nature of decreased method usability and increased experienced method informativity.

Contents

1	Introduction	1
2	Preliminaries	3
2.1	Social media, microblogs, Twitter, and interactions	3
2.2	Misinformation	3
2.3	Fact-Checking	3
2.4	Credibility	3
2.5	Authenticity	4
2.6	Authenticity methods	5
2.7	Verification	5
2.8	IRMA	6
2.9	Twid	6
3	Related Work	8
3.1	On Verification	8
3.2	On Twid	8
3.3	On authenticity	9
3.4	On credibility	10
3.5	On interactions	11
4	Methodology	13
4.1	Philosophical position and paradigm	13
4.2	Approach	13
4.3	Strategy	13
4.4	Methods	13
4.5	Time span and group design	14
4.6	Data collection and analysis	14
4.7	Sample and recruitment	15
4.8	Procedure and materials	15
4.9	Ethics	19
4.10	Position of the researcher	20
5	Results	22
5.1	Authenticity	23
5.2	Credibility	24
5.3	Interactions	28
5.4	Quality concern	29
5.5	Scrutiny	30
6	Discussion	32
6.1	The novelty and prevalence of themes and perceptions in literature	32
6.2	Key findings	35
6.3	Critical notes and limitations	38
6.4	Future work	41
7	Conclusions	44
A	Appendix A	53
B	Appendix B	55
B.1	Code tables	64
C	Appendix C	72
C.1	Twid project proposal	72
C.2	Interviewguide (Dutch)	77
C.3	Interview Guide (English)	81
C.4	Study variant 1	85
C.5	Study variant 2	87
C.6	Study variant 3	89
C.7	Study variant 4	91

1 Introduction

With the continued development of social media over the last decade and a half, we also saw the co-evolution of misinformation online. Misinformation, which is false or inaccurate information (Wu, Morstatter, Carley, & Liu, 2019), has been a prominent presence on online platforms. A well-known example of the potential consequences of misinformation is its likely influence on the outcome of the 2016 American presidential elections (Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019). If the outcome of an election indeed shifted as a result of foreign intervention, it undermines democracies' central tenet of citizen autonomy as their ability to choose is guided under false pretences. It also affects a nation's ability to exert sovereignty over internal affairs, in disregard of international law. Another example is the prevalence of medical misinformation during the COVID-19 pandemic (Kouzy et al., 2020). Medical misinformation contributed to unrest, mental anguish, and death due to improper precaution and has links to suicide (Rosenberg, Syed, & Rezaie, 2020). The consequences of online misinformation have been severe enough that the UN has adopted motions to condemn it (United Nations Human Rights Council, 2022), the European Union has devised a strategy to address it (European Commission, 2022), and in Dutch national politics, it is a subject of ample debate (Ministerie van algemene zaken, 2023).

On social media platforms such as Twitter, misinformation diffuses more quickly than truthful messages (Vosoughi, Roy, & Aral, 2018). This directly relates to falsehood often being novel, which is attractive for online gossip (Talwar, Dhir, Kaur, Zafar, & Alrasheedy, 2019; Vosoughi et al., 2018). Some further factors contributing to the prevalence of misinformation are there being no sanity check on information, limited information on author credibility (Morris, Counts, Roseway, Hoff, & Schwarz, 2012), ease of impersonation (Tsikerdekis & Zeadally, 2014), and users willfully spreading information they know to be false (Chadwick & Vaccari, 2019). To address the prevalence of misinformation, a number of mitigating strategies have been proposed. A well-known example of this is fact-checking (Vlachos & Riedel, 2014), which seeks to correct statements of a factual nature. Social media companies also implement their own mitigation methods. Verification is a well-known example of this (2.7). Verification on most social media platforms prevents impersonation by providing users with a badge signifying their legitimacy. Most implementations of Verification limit it to notable accounts and require some proof of user identity. Although Twitter recently began changing Verification to a paid subscription service without a notability requirement (Twitter, 2022a), the aforementioned functionality is mostly still in use on their platform as well as on most other platforms. The Twid project (van Gastel, Bernard, Schraffenberg, Hanna, Bor, Dennis, & Vervoort, Lian, 2021) suggests an alternative approach reminiscent of digital signatures. This approach addresses the issue by using the attribute-based credential system IRMA (Alpár, Jacobs, Lueks, & Ringers, 2017) to sign tweets with labels. Labels are attributes provided by certain third parties and are meant to showcase an author's relation to their message content. Twid also provides some guarantees to counter impersonation akin to Verification. Twid further distinguishes itself by being freely available, and scalable, and providing users insight into how message authors relate to posted content.

Both Twid and Verification are meant to positively impact the user experience of Twitter and the quality of content by using authenticity to combat impersonation and disinformation respectively. However, knowledge of the manner users experience these methods is somewhat lacking. Authenticity methods often give proof of an author's identity (2.7), which is a measure of being authentic or genuine. On the other hand, research has focused more on analyzing the user effect of authenticity methods as a credibility perception, which is a measure of a subjective truth perception (Morris et al., 2012; Vaidya, Votipka, Mazurek, & Sherr, 2019). There is also a gap in research when it comes to understanding how users perceive the credibility of social media and microblog posts such as Twitter (Choi & Stvilia, 2015). Research such as Morris et al. (2012) and Vaidya et al. (2019), more so concerns quantitative hypothesis proving rather than exploring user experience. Although this does allow researchers to prove a pre-determined cause-effect relationship, it performs poorly at exploring what is attributed to the

success or failure of a tested method and encompassing experience beyond the pre-determined scope. In summary, we took note of the understudied nature of credibility on social media, the narrow scope of analysis usually chosen to examine authenticity methods, and a poor understanding of the manner users experience authenticity methods. We, therefore, believe that a broadly scoped inductive study to examine user perceptions of microblog posts provided with authenticity methods can give information to assist in examining if and how methods influence the user experience on social media.

Our research goal is the exploration, description, and explanation of user perceptions of authenticity methods. We also want to provide future researchers with similar interests pathways for continued work, as well as a research design process they may use and improve upon. Furthermore, we strive for contributions that may aid in developing authenticity methods on a basis of understanding to combat misinformation, counter impersonation, and positively impact user experience. To achieve this, we design and conduct a primary (Lowe, Zemliansky, Driscoll, Stewart, & Vetter, 2010) within-subject design (Charness, Gneezy, & Kuhn, 2012) case-study (Gerring, 2004). Tweets crafted for the study are provided with Twid attributes, Twitter's Verification, and no method of authenticity. Participants' perceptions are provided by reacting to the different tweets with the think-aloud method (Barnard, Someren, Barnard, & Sandberg, 1994) and through semi-structured interviews (Boeije, 2009). Fourteen participants were selected using purposive sampling (Boeije, 2009). Data analysis was performed by qualitative mono-method data analysis (Boeije, 2009), specifically thematic analysis of transcripts (Braun & Clarke, 2006).

2 Preliminaries

Before we delve deeper into the materials, we provide definitions and information to create a shared understanding of concepts and ideas for later sections. Displaying the frictions and shortcomings in scientific literature is mostly reserved to the related works section.

2.1 Social media, microblogs, Twitter, and interactions

As Carr and Hayes (2015), p. 50 defines: “Social media are Internet-based channels that allow users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others.”. A subsection of social media is microblogs, where users can share their thoughts in strictly short messages (Java, Song, Finin, & Tseng, 2007). Twitter (Twitter, 2022a) is a specific instance of a microblogging service. On Twitter, users follow other users, seeing the activity of followings in a personalized feed. Following another user is not a reciprocal relationship. Twitter has an authenticity method (2.5) called Verification (2.7). Liking, commenting, retweeting, and sharing a tweet are interactive behaviour users can have with tweets of others and will collectively be known as interactions throughout this thesis. They are a subset of possible engagements (3.5) chosen for their relevance within the planned study environment (4.8).

2.2 Misinformation

Misinformation generally refers to the creation and propagation of false and inaccurate information regardless of intent (Lazer et al., 2018; Wu et al., 2019). The influence of falsehoods is nothing new. For example, Socrates was put to death based on spurious accusations (Ryan K. Balot, 2009) and propaganda used to be, and sometimes still is, a government posting (DOOB, 1950; Micheal Madden, 2015). However, misinformation is mostly used to refer to falsehoods on social media (Wu et al., 2019). Two sub-categories of misinformation we want to highlight are disinformation and fake news. In disinformation, the falsehood is intentional (Wu et al., 2019). Fake news (Lazer et al., 2018) is misinformation through lacklustre editorial standards or disinformation made to look like traditional news content.

2.3 Fact-Checking

Fact-checking is determining the truthfulness of a claim (Vlachos & Riedel, 2014). The method involves reducing a claim to a true or false statement. This is done by compounding multiple claims to a single fact, by representing the claim on a true or false spectrum, or by picking and choosing which claims to verify (Hughes et al., 2014). Additionally, some explanation may be provided to allow insight into the resulting truth value. Although fact-checking as a method could refer to any form of hypotheses testing, it mostly refers to a more recent development of journalists and organizations to do so with public statements. Claims often originate on social media. A subcategory of fact-checking is automated fact-checking. This is the algorithmic Verification of claims rather than manual Verification (Graves, 2018).

2.4 Credibility

Credibility means “the quality that somebody/something has that makes people believe or trust them” (Oxford learners dictionary, 2021) or “the quality or power of inspiring belief” (Merriam-Webster Dictionary, 2022). These textbook definitions serve well in exploring academic usage. Credibility is a subjective perception of someone or something that is under judgment, not an actual objective quality (Choi & Stvilia, 2015). This subjectivity is why throughout this thesis we refer to perceived credibility when talking about the observations of a person instead of just credibility. Moreover, “the quality” aspect of credibility sets up that there are actual dimensions to these judgments. Originally and still commonly this is described using two dimensions (Hovland, Janis, & Kelley, 1953), perceived expertise (originally

expertness) and perceived trustworthiness. The perceived trustworthiness captures the goodness, lack of bias, and honesty of the source (Fogg, Soohoo, Danielson, & Marable, 2003; Pornpitakpan, 2004). The perceived expertise is more a reference to the knowledge, skill, and experience of the source in the domain relevant to the message (Choi & Stvilia, 2015). Referring to “someone/something” highlights that being the object of a credibility perception is not limited to people. A publication, corporation, social media platform, or even information without a clear source can receive a credibility perception.

Over the years, researchers identified numerous factors likely to influence credibility perceptions. A complete list of potential factors is truly extensive and would be impossible to cover, although some are quite relevant to our study. Therefore, we created a custom selection of factors we found that are significant to this thesis and subdivided them for readability. Keep in mind that factors are derived from different sources and literature reviews, resulting in the possible correlation and causation of factors not always being fully explored. We always mention if a factor is specifically researched for websites or Twitter. The first list of factors relates to the recipient of information. In our study, we refer to this as the participant. They are the ones making a credibility judgment and the factors mentioned are inherent to them. These can be found in table 1 of Appendix B. The second list of factors relates to the content of the message. Content factors influence recipients and can be found 2 of Appendix B. Thirdly, some factors related to the source conveying the message. In our study, we refer to this as the author. Author factors are inherent to them. Despite this, like all factors mentioned, the author’s factors influence participants. Author factors can be found in 3 of Appendix B. Finally, the few remaining factors were more difficult to categorize under a common denominator. They can be found in 4 of Appendix B. We will regularly refer to factors from these tables throughout the rest of this Thesis.

2.5 Authenticity

From a semantic and philosophical viewpoint, authenticity can be used to define sincerity in reflecting one’s values and ideas, autonomy over one’s own actions, or awareness over the distinction between inherent values of the self and those outside the self (Varga & Guignon, 2014). We use authenticity mostly in reference to sincerity and autonomy in this thesis. To understand this further, we examine authenticity’s use in information security. Authenticity in information security is defined and used most in line with what is meant in this thesis and directly relates to the functionality of authenticity methods (2.6). The definition of authenticity in the CIA triad, an early information security model, is given as: “Authenticity denotes the quality of being original and genuine, and therefore authentication is the process of verifying, to some desired level of confidence, that a claimed identifier is valid and is actually associated with a particular item or person.” Samonas and Coss (2014), p.34. In the more recent Parkerian hexad model it is defined as: “Authenticity is assurance that a message, transaction, or other exchange of information is from the source it claims to be from. Authenticity involves proof of identity.” L. Clemmer (2009), p.14. Regardless of what model we are using, authenticity is used regarding the quality of knowing the person or system you are communicating with is actually who you think they are. Authenticity is determined during communication between systems, people, or a combination of the two. In this communication, some sort of evidence of identity is given. A password can be seen as proof of knowledge over a secret sequence meant to be private to the account holder, and therefore serving as evidence of your identity. Biometrics as proof of biology. Authenticity provided by a system is a guarantee, with a system giving you a binary outcome that according to its specifications, something, or someone is authentic or not. However, in some cases, judgement over the authenticity of others is not made by a system but by people. For example, when a user has to determine if an e-mail was sent by a scammer or when a user sees a Verification (2.7) badge on social media. Determining authenticity in that context becomes a subjective perception of the true value of the authenticity of an author or account owner.

2.6 Authenticity methods

Social media has peer-to-peer methods in place to allow users to confirm the authenticity of other accounts. For example, a government official providing a link to a government website listing their account fills this function. Specific social media elements, such as a Verification badge (2.7), are fully catered to this. These methods fully catered to this is a big part of what we will examine for this Thesis. We will refer to them throughout this document as authenticity methods, which we use to refer generically to any implementation with peer-to-peer authenticity as a main intent. The methods examined specifically for this study will be Verification (2.7) and Twid (2.9). We deviate slightly from past works by preferring the term "authenticity method" over Verification and other synonyms. We do this because authenticity reflects the actual functionality provided (3.3). To define this further in the context of social media, we provide the following two definitions based on the Twid project proposal (C.1) and Verification official designation as "account verification" (3.1). The definitions will delineate two types of possible authenticity guarantees that can be provided by authenticity methods. The definitions are more so relevant for understanding the authenticity methods used in this thesis than they present a strict or complete coverage of authenticity methods.

Account authenticity is when an authenticity method provides guarantees over the authenticity of the account. The underlying idea is that only the owner, the author, of the account, has access to the account. Author access on social media usually occurs through knowledge of a password. Account authenticity is obtained by providing evidence of the truth of your identity in relation to the account. Methods solely implementing account authenticity provide no checks during specific actions like posting a message. An example of a method implementing this would be Verification (2.7).

Message authenticity is when an authenticity method provides guarantees over the authenticity of the message. For example, continued use of a digital signature by a person would imply that each continuous message is sent by the same person, but provides no evidence of whom this person or account is. An example of a method implementing this would be Twid (2.9).

It would be possible to have a single method or combination of methods that allow for both forms of authenticity at the same time.

2.7 Verification

Figure 1: *The world's first tweet (jack [@Jack], 2006) is from a verified account, the blue badge next to the account name signifies this*



Verification is a Twitter service providing accounts that meet certain requirements a blue badge when authoring a message to signify account authenticity (Twitter, 2022b). Recently, this system was changed to a subscription service available to accounts that pay for the service. Paid subscribers need to meet requirements for being active, providing sufficient account information, and being non-deceptive. As of writing this, the rollout of the service is limited to the US, Canada, Australia, New Zealand, and the UK. This new system will not be examined in this study. When we reference the new subscription

system, it will always be explicitly mentioned for either its subscription functionality or under its new name "Twitter blue".

In this thesis, we examine the so-called "legacy system". The legacy Verification service is still used even in those countries that now also offer Twitter blue, and is still the only system available in other countries. When we refer to Verification in this thesis, this "legacy" system is what we refer to. This "legacy" implementation of Verification is a Twitter system available to all accounts of notable users who are rewarded with additional legitimacy based on some proof of identity provided by the corresponding account owner (Twitter, 2022b). Proof of identity requires the owner of an account to provide evidence that they are who they claim to be (L. Clemmer, 2009). Providing a copy of your identification papers if you are a person or linking to some sort of verifiable official website in the case of an organization are examples of proofs Twitter allows for. Furthermore, Twitter requires you to be a notable user: "Your account must represent or otherwise be associated with a prominently recognized individual or brand, in line with the notability criteria described below." (Twitter, 2022b). The criteria leave some room for interpretation, but in short, require one to have some level of a verifiable impact online. This means that Verification under this implementation is not achievable for everyone. Among others, Facebook and Facebook Messenger (Facebook, 2023), Instagram (Instagram, 2023), TikTok (TikTok, 2023), Telegram (Telegram, 2020), WhatsApp (WhatsApp, 2023), Snapchat (Snapchat, 2023) and Pinterest (Pinterest, 2023) provide an almost identical Verification service. Some other social media companies also provide a service under the same name, although actual implementation may differ. Verification is sometimes also referred to as Account Verification.

2.8 IRMA

I Reveal My Attributes (IRMA) (Alpár et al., 2017) is an attribute-based credential system first developed at the Radboud university and managed by the spinoff non-profit Privacy by Design Foundation (Privacy by Design Foundation, 2023b). Attribute-based credentials are a way of proving elements of your identity to a third-party service such as Twid (2.9). IRMA's intended use is providing the minimum amount of information context requires, while limiting access to all other information about the user, and includes guarantees over provided information. Specific organizations can provide users with their attributes for local storage through a one-time authentication. These organizations could theoretically be anyone meeting IRMA requirements and are called attributed issuers. An example of a user getting an attribute from an attribute issuer would be a national government issuing a date of birth attribute to a citizen after authentication. Users can store and access all their IRMA attributes on their IRMA app. IRMA attributes are only stored locally. The Privacy by Design Foundation mentions this decentralized nature as a key feature setting it apart from similar systems. Users maintain control over which attributes they wish to share. Organizations wishing to implement features necessitating attribute-based identification are called verifiers. A verifier could theoretically also be anyone implementing IRMA's method. An example would be an online alcohol store. If implemented as intended, such a store checks if someone is of legal age by requiring users to only disclose their date of birth and not other identifying information. Verifier can check the integrity and origin of attributes. Users maintain a complete log and cannot provide expired attributes. Detailed information on available attributes, documentation, and provided guarantees is freely available on the Privacy by Design foundation website (Privacy by Design Foundation, 2023b).

2.9 Twid

Twid (van Gastel, Bernard et al., 2021) is a project from the iHub (Radboud's Interdisciplinary Hub for Security, Privacy, and Data Governance). The full project proposal for Twid (van Gastel et al., 2021) can be found in the appendix (C.1). The ambition of the Twid project is to mitigate disinformation by message labelling of author attributes through IRMA. Initial development is for Twitter, but could feasibly be extended to other social media platforms. When signing, Twid verifies provided IRMA

Figure 2: An example of what a signed tweet using Twid could look like.



attributes, allowing others to judge message authenticity. Twid users who choose not to or fail to sign their message will have a “not signed” indicator displayed under their message. Although not obligatory to be used any which way, it is geared toward displaying domain knowledge or some other author relation over the information contained in your message. advantages this hopes to bring as compared to Verification are :

1. Availability: Twid is available to all Twitter users with the freely available IRMA app. Verification only to notable individuals or paying users in the subscription system (Twitter, 2022b).
2. Shift from whom to what: providing an informative attribute instead of a verified badge shifts attention towards its relevance. The intended use of Twid is to provide provable proficiency over message content. Contrary to Verification, authors with relevant affiliations can signal this with their attributes. This allows users, especially those who previously did not know the authoring account, to judge their suitability in providing credible content.
3. Scalability: Verification and Twitter blue are provided through human review (Twitter, 2022b). Twid relies on IRMA to solve the authentication problem, removing the need for human reviews.
4. Individual message authenticity: authenticity guarantees are extended to individual messages. Sending a message on a hacked account requires access to their IRMA account as well. Moreover, attributes can be chosen by the author for their relevance to the message. Attributes are timestamped to disallow discredited experts from signing and recent experts from signing past messages.

3 Related Work

We reviewed and examined literature relevant to our topic. We did this for the most part in the preparation for designing and analyzing our study. We want to present works that help contextualize our study. We present our influences and highlight academic consensus and points of discussion. We also feature the knowledge gaps in these works we chose as focal points for our study.

3.1 On Verification

One of the authenticity methods we will include in our study is Twitter’s Verification (2.7). This method and its potential effects have received some academic attention, although we shall also showcase that we can not yet speak of any definitive understanding.

Morris et al. (2012) used a combination of a survey and two experiments to analyze several factors influencing user perceptions of tweets. They then mapped this to how credibility was affected. Among the features analyzed was Verification. They found that Verification was one of the features most enhancing a tweet’s perceived credibility. They also found that users lean on heuristics over content to make credibility perceptions. Morris et al. (2012) provides a broad and detailed account of credibility perceptions on Twitter, although less so specifically of Verification. Morris et al. (2012) did not study Verifications effect isolated, instead opting to group it together with if users heard of the author of a tweet before. Their main means for results also do not reflect actual behaviour as its findings originate from a survey in which users are asked to self-report on behaviour. Furthermore, Morris et al. (2012) focused more on if users were influenced than on how and why users were influenced. Morris et al. (2012) also primed users on factors such as Verification and credibility and did not verify the behaviour purported by self-reporting. This lack of unconscious impact is also an important criticism of Vaidya et al. (2019). In a series of experiments more meant to measure unprimed behaviour than explicit perceptions, Vaidya et al. (2019) concluded that even though users can effectively recognize an authenticity indicator such as Verification, they do not conflate it with credibility. In a separate experiment, focused solely on the relevance of Verification on Twitter, Edgerly and Vraga (2019) came to similar conclusions. In their findings, they saw no impact of Verification on perceived credibility.

Although the more recent studies we found suggest a limited impact of Verification, there is no scientific consensus to speak of yet on the issue. Studies focussed on variable isolation and hypotheses proving. This means the setup often was to prove what effect Verification had on a chosen outcome. Their chosen outcome was always credibility. Although some studies also covered if participants understand it is Verifications intent to prevent impersonation, this was less the focus than credibility. Overall, the setup of these studies allowed for minimal insight into the manner in which participants perceived to be influenced. We find it difficult, given this lack of consensus and minimal description of user experience, to attribute what may or may not contribute to the outcome of these studies. In extension, this makes it difficult to understand how this widely used innovation affects social media users.

3.2 On Twid

The second authenticity method in our study is the proposed method of Twid (van Gastel, Bernard et al., 2021). The project proposal (C.1) by van Gastel et al. (2021) is our best-written source for Twid’s intended impact. The supposition in the proposal is that shifting attention from reputation, who a person is, to an author attribute, what a person is, is a better heuristic than Verification currently provides. The proposal outlines how this assists in judging content from unknown authors and allows users to judge the relevance an author has in making a specific statement more easily. The idea is well-reasoned but remains unproven. The proposal pays little attention to potential pitfalls. A first exploratory analysis of Twid is given by Simon (2022). They provided evidence of Twid having a positive influence on credibility and sharing behaviour, with a possible relation between the two. They found no decrease in credibility

for a not signed indicator compared to the complete absence of any authenticity method. [Simon \(2022\)](#) provided a relatively small-scale analysis. It utilized only one relatively factual tweet paired with one label high in expertise. This makes this an excellent setup for producing evidence of Twid being capable of providing any effect but is not a reflection of Twids' actual effect if it were to be adopted by Twitter. [Simon \(2022\)](#) also gives relatively little descriptive data for understanding how recipients utilize an attribute to form a judgement.

We are interested in understanding the perceptions of participants in a more diverse and less clinical setup. We want to examine this as Twid is sure to be used in a less optimal way, so understanding it under somewhat less ideal circumstances would be beneficial in understanding how it is experienced if it were to be rolled out on Twitter. Furthermore, we would also want more descriptive information to better understand how elements of Twid potentially contribute to its success or hold it back.

3.3 On authenticity

We set out to employ a broad scope in exploring user perceptions of authenticity methods. Nevertheless, some conceptual baselines will help us gauge participant observations. Twitter's own goal with Verification is: "the blue Verified badge on Twitter lets people know that an account of public interest is authentic" - ([Twitter, 2022b](#)). Being authentic is therefore the proof actually being provided by Verification, and the "public interest" requirement contained in their definition of notability is the main barrier to being eligible for proof. Only [Vaidya et al. \(2019\)](#) considered authenticity as an important concept to cover in their analysis of Verification. Although it was valid for their purposes, they simplified experienced authenticity in their analysis to the presence of Verification on a tweet. Twid's proposed method also includes a form of authenticity. Twid's intended implementation extends the same guarantees over Twid attributes as IRMA provides over its attributes. Twid, therefore, does not guarantee "being authentic", which we would describe as the authenticity of the self. Twid does guarantee authenticity over the truth of an attribute and its relation with an author. For both, the authenticity method provides reflections of reality, be it personhood or a certain attribute of personhood. Based on understanding these implementations, we can now conceptualize authenticity for authenticity methods in line with views from [Varga and Guignon \(2014\)](#) on authenticity. In both Twid and Verification authenticity can be seen as autonomously displayed externally verifiable elements reflective of the sincere self. We do want to note that from a strict conceptual viewpoint, methods always provide a fallible application of authenticity. With this, we do not wish to say either method is badly implemented. Only any security measure can be bypassed given enough motivation. It would be possible through malcontent or consent to posit as something or someone you are not. We mean that another user can allow you to secretly post in their name, you could hack an account, you could be unjustly verified or obtain Twid attributes you ought not to. Authenticity provided by methods is therefore not completely guaranteed, although we mostly mention this so as to not create a semantic issue. The security implementation of the methods will not be covered further in this thesis and can therefore be mostly disregarded.

As ([Vaidya et al., 2019](#)) lays out, authenticity is a difficult concept for users to understand and often conflated with integrity. Integrity here relates to the quality and objective truthfulness of the information and is a guarantee more akin to those provided by fact-checking. [Felt et al. \(2016\)](#) describes how some users believed the content of websites to be true simply because the website security connection icon indicated it was secure. Conceptually, applying authenticity might also be more difficult than intended. Authenticity guarantees may be relatively well-known within the niche of people with knowledge about digital security and computing science but have repeatedly been shown to be difficult to understand by the public ([Vaidya et al., 2019](#)). Ultimately [Vaidya et al. \(2019\)](#) found no proof of a similar effect for Verification, although they deemed this a legitimate risk. The public interest element of Verification's requirements risks leading users to believe that a level of importance and authority is provided by it. To add to the confusion, the literal definition of Verification is about the truth or accuracy of something.

Although it is meant to refer to the verified nature of the account, it can be imagined how this may be misconstrued as verified content. Presumably, for this reason, social media companies sometimes do in their documentation refer to Verification as Account Verification. This use is not super consistent with "Verification" is widely used and commonly known. Both Verification's confusing name and the risk of conflating it with integrity, users may falsely expect fact-checking to be applied as part of the method. This is also why we give preference to the term authenticity method over Verification as it more so reflects the provided guarantee (2.6). Use of the term authenticity methods should be reserved for academic and contextualized use though, since as we just mentioned, it may be difficult to comprehend for the layman. User interpretation of a measure may further lead to unexpected behaviour. [Felt et al. \(2016\)](#) describes how and why Google changed its website security connection iconography. Outside our previous coverage of authenticity, the icon was not intuitive, did not translate across cultural borders and did not sufficiently emphasize insecure websites. The original exact functionality of the connection security icon was publicly documented. This is more than can be said of Verification. Although Twid is very well documented, the foundation managing Twid's underlying method IRMA mentions knowledge and understanding required from users as a main disadvantage. If the simple and well-documented connection icon caused such issues, we feel it is valid to at least question the same for Twid and Verification.

In conclusion, we view authenticity as conceptually important to explore when studying authenticity methods. Authenticity has a direct connection to provided guarantees by both Twid and Verification. How methods contribute to authenticity has not yet been examined. However, it may be difficult to understand authenticity for users, and seems understudied as a concept in the context of authenticity methods. Outside a functional and logical explanation, its connection to Twid is yet to be explored.

3.4 On credibility

In contrast to the study on user comprehension of authenticity, the extent to which messaging convinces users has been widely studied through the lens of credibility research. The Ancient philosopher Aristotle already documented ideas on the subject with his examination of persuasion. Persuasion is the act of increasing credibility which he examined with his study of rhetoric ([Rapp, 2022](#)). The more modern seminal work on credibility is [Hovland et al. \(1953\)](#). Although by now the study is almost 70 years old, [Hovland et al. \(1953\)](#) is foundational in works on credibility. He was the first to suggest dividing credibility into trustworthiness and expertise, which is still almost universally done today. [Pornpitakpan \(2004\)](#) did a literature study giving a detailed account of research done in the five decades since [Hovland et al. \(1953\)](#). It reaffirms the original division into expertise and trustworthiness and expands on it by mentioning dozens of factors that have been proven to influence credibility perception. [Pornpitakpan \(2004\)](#) also explores how users generally tend to tie more value to trustworthiness. The examination is over twenty years old and mostly pre-date studies done on credibility perception on the web. It even mentions the possibility of its shortcoming, seeing as how it puts forward that the medium hugely affects how credibility is perceived. This does not mean its contribution and findings should be completely disregarded, more that it has shortcomings on the potentially updated details of the effects of existing factors and lacks perspective specific to the internet.

Expanding on medium-related shortcomings, [Choi and Stvilia \(2015\)](#) reviewed literature specifically relating to credibility on the web. It offers an operationalization of web credibility separating operator, content, and design to better analyse specific elements of credibility on the new medium. For each element, expertise, and trustworthiness can be examined. It extends the influencing factors from previous research with factors specific to the web. [Choi and Stvilia \(2015\)](#) also covers how online credibility research has understudied areas. Research mostly focused on the credibility of websites as singular entities. We refer with this to websites more reminiscent of the earlier internet, where those posting information and the owner of the website are more or completely interrelated. [Choi and Stvilia \(2015\)](#) states that research has less so focussed on user-generated content, which includes microblogs. This

means that credibility research so far has a somewhat limited understanding of social media platforms. This shortcoming extends to [Choi and Stvilia \(2015\)](#) operationalization for online messaging. It is somewhat lacklustre in allowing for an author of a microblog message and the medium, Twitter as a company, to be treated as separate entities. We detailed further specifics on the credibility of Verification and Twid in their earlier subsections ([3.1](#), [3.2](#)).

Research on credibility, or at least the research we found, is primarily market research. This leads to a very specific way of looking at credibility. Studies often analyze factors to maximize persuasion. This differs in the way the researchers behind Twid seemingly look at it since their goal is more so to make individuals credible to the extent they deserve it. Although definitive knowledge about its workings on the web and Twitter is somewhat lacking, credibility is a natural and well-documented way of examining user perceptions of persuasion and therefore a measure we deem useful for utilizing. It is, however, not a definitive measure for determining the inherent quality or impact of authenticity methods.

3.5 On interactions

The pursuit of an additional measure of understanding the impact of authenticity methods on social media leads us to the final area of literature we will examine. We will now explore interactive behaviours. For our study, we will mostly focus on Twitter's Like, Share, Comment, and Retweet features and refer to specifically these as interactions. On social media, messages can potentially reach an incredibly wide audience. It would be possible for a piece of not credible misinformation to be spread on social media. Shared in a small group, this misinforming message may not be impactful or likely to be deemed credible. However, if this now gets spread widely enough, it might be very impactful. With a large enough reach, it now may convince more people, resulting in more impact. As a dimension of analysis, interactions are therefore relevant for their direct relation to the group success of a message.

We want to move from theory to evidence on this idea of information somewhat akin to pathogenic infectivity. It is difficult to pinpoint though how popular social media prioritizes messages. Underlying algorithms are mostly not public. Twitter's own activity dashboard suggests impressions, how often tweets are viewed, and engagement, an extended number of interactions, as important metrics for audience engagement ([Twitter, 2023](#)). Furthermore, we know that we can see the interactive behaviour of accounts we follow on our feed. These factors combined suggest engagements have a strong influence on the overall reach of tweets. According to a literature review, [Dora-Olivia Vicol \(2020\)](#) up to 43% of people admit to sharing information containing falsehoods. 25% admits to knowing it was false at the time of sharing. This provides evidence that credibility alone cannot determine sharing behaviour and diffusion of disinformation. [Vosoughi et al. \(2018\)](#) even showed using data analysis of over 100,000 stories across roughly three million users that falsehoods diffuse quicker and deeper on Twitter across all subjects tested. [Lazer et al. \(2018\)](#) describes that false information is retweeted more, and this is exacerbated when it is political. Some interaction predictors, such as those found in cognitive dissonance ([Vosoughi et al., 2018](#)), and emotional response, ([Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012](#); [Stieglitz & Dang-Xuan, 2013](#); [Vosoughi et al., 2018](#)) show overlap with credibility predictors ([2.4](#)). When looking at predictors of interactive behaviour that do not overlap with predictors for credibility, [Vosoughi et al. \(2018\)](#), [Lewandowsky et al. \(2012\)](#), [Talwar et al. \(2019\)](#), and [Thompson, Wang, and Daya \(2020\)](#) all provide evidence of interaction through novelty. In short, their combined findings are as follows; Novelty triggers human interest because of our need to gossip and engage in status-enhancing behaviour for self-gratification. In an attempt to address this, ([Pennycook et al., 2021](#)) achieved some success with an intervention targeting novelty. It did so by subtly shifting attention towards information accuracy. In examining misinformation-sharing interventions, ([Lewandowsky et al., 2012](#)) identified ways they can be successful. It helps to incorporate real-world affirmation, informed scepticism and integrating misinformation into interventions as counterarguments. However, they also provide a warning for a backfire effect interventions have been known to have, leading to misinformation becoming more

entrenched in users. Although not entirely understood, interventions that, threaten a user's worldview, refute without an alternative account, overemphasize misinformation as part of refutation or provide complex true information to replace simple disinformation are covered as risking backfire.

The sources cited exploring the effects of sharing and interactions are mostly a combination of literature studies applying older theories on social media and quantitative analysis setting out to prove specific hypotheses. This leads to limited insight into the underlying thought processes users have when deciding to interact. Relating interactions to authenticity techniques seems entirely under-researched. We, therefore, justify further studying users' underlying decision-making processes, perceptions, and experiences involved in deciding to interact.

4 Methodology

We use this section to explain our choices relating to the study design, applied research methods, and our positioning a priori conducting the participant study.

4.1 Philosophical position and paradigm

Although our research goal is pragmatic, the paradigm we choose to achieve this is interpretivism. Interpretivism is a philosophical inclination and research paradigm that emphasizes the subjective and socially constructed nature of reality (Boeije, 2009). Pragmatism is a paradigm more focused on action-based research and allows more flexible use of objective and subjective research methods if they can positively change ever-changing reality (Goldkuhl, 2012). Pragmatists posit objective reality may or may not exist, but cannot be separated from humanity experiencing it in our interpretation. Our study seeks to better explore and describe the subjective experience of authenticity methods. The goal is for this to result in an improved understanding serving future research and development. We believe that positivist studies can be appropriate for this or any purpose. A lack of understanding of the user perspective on authenticity methods means that a positivist study seems less applicable as we feel additional exploration of authenticity methods experiences is first required. In this study, therefore: “interpretivism is seen as instrumental for a pragmatist study” - Goldkuhl (2012), p.144.

4.2 Approach

Our study primarily takes an inductive approach. This means that we move from disconnected observations towards connected theories based on observed patterns that serve as our final conclusions (Katz, 2001). We chose inductive reasoning to study the perceived user experience of authenticity methods to allow us to encompass a broad scope and produce detailed descriptions (Boeije, 2009). The main consequence of this approach is that our conclusions are to be seen as generalized hypotheses arising from the best available data and not as provable conclusions. It further means that cause-effect cannot be definitively established. To create our setup and guide, it was necessary to retrieve and use previous knowledge (Kallio, Pietilä, Johnson, & Kangasniemi, 2016). As a consequence, our study design and analysis do present some deductive elements. We mention this to not deny the influence of previous work on ours, and not to suggest we set out to fit our data into pre-existing frameworks (Braun & Clarke, 2006).

4.3 Strategy

For our purposes, we have designed a case study. In a case study, a phenomenon is studied in a limited and descriptive scope. A case study usually involves tracking participant observations in a carefully documented context to ascribe meaning to them (Boeije, 2009). We set out to achieve a broader understanding of the user experience of authenticity methods and microblogs. A case study should allow for this and allow us to derive more generalized hypotheses from the analysis of Twid and Verification (Baxter & Jack, 2015). We do note that the case is of an entirely constructed nature due to Twid being still under development and because of perceived limitations in using real tweets (6.3.4).

4.4 Methods

Our chosen method of analysis was a mono-method qualitative thematic analysis (Braun & Clarke, 2006). This process involves a researcher or research team interpreting and structuring the data along observed themes, producing a framework of the data meant to aid them in understanding and presenting interpreted relations in the data and noteworthy elements of interest. We chose it for its flexible nature and because a thematic analysis is relatively easy to apply by a novice researcher. We were further interested in somewhat isolating and highlighting how participants’ experiences are shaped, especially regarding authenticity methods. Choosing themes surrounding shared concepts should allow for this.

Thematic analysis as it is applied by us is a subjective process. Subjective does not just apply to the capture of the subjective experience of the participants. Our choices as researchers in constructing and highlighting themes and perceptions are also subjective and personal. Although we may repeat this too often, the preceding was just an elongated way of stating we do not posit some absolute truth in our resulting findings. We do provide a description of the context of our study and the research process that led to our findings. We use existing research methods and present what can conclude from them. We also explain how and why we chose to deviate from established methods and highlight some of our failings in applying existing ideas. We highlight our subjectivity and positioning to allow the reader to see how it may have influenced us. We do all of this in hopes of allowing readers to judge the quality of our work and with it the validity of our findings (Korstjens & Moser, 2018).

We did familiarize ourselves with the literature to facilitate study design before data collection and analysis. Our goal is not deductive thematic analysis, and we took precautions to not limit our analysis to a projection of existing work onto our corpus. We tried to remain open to the previously described phenomena and avoided literature while collecting, processing and analyzing the data. Our process is noted here to contextualize potential effect and avoid presenting the analysis as grounded (Harris, 2014).

4.5 Time span and group design

The study is cross-sectional, gathering participant data during a single session (Levin, 2006). A longitudinal study encompassing Twid is not yet possible. Twid is not yet available and an extended timespan would require participants to continuously be able to engage with it. Such a timespan also does not fit the boundaries of a Thesis project. A cross-sectional time span limits scale while allowing a broad interpretation of observations. A consequence is that results present a snapshot.

The design of the study is a within-subject design (Charness et al., 2012). Participants make use of all the different methods available in a within-subject design. This group design minimizes random noise caused by differences in participants and requires few participants. We chose it because a smaller group fit our limited labour resources. A potential drawback to consider is that subjects may transfer knowledge throughout the study and can be primed.

4.6 Data collection and analysis

We collected primary data from semi-structured interviews and think-aloud reactions to the case study. Primary means the data were collected as part of this study instead of a secondary data set made available by others (Boeije, 2009). We chose this as secondary data encompassing Twid is not available. We could also not find data with Verification that suited our purposes.

A semi-structured interview is a flexible method of data collection, allowing a researcher to lightly delineate areas of interest to talk about, whilst allowing for the discussion of previously undeveloped ideas (Boeije, 2009). We picked this interview method because we wanted participant input on authenticity methods and means for understanding. We also wanted to encompass unexpected observations in an underexplored field. Therefore, we could use prepared questions incrementally more specific to our pursuit, while also discussing and exploring new ideas.

The think-aloud method produces data on participant observations and processing of tasks (Barnard et al., 1994). We included it to gain data on the initial perceptions of participants. We specifically were interested in seeing what elements were brought up as well as hearing things they found noteworthy regardless of influences caused by continued study participation.

Audio recordings were made of each study participant. Audio recordings are fed to Azure text-to-speech services (eric urban, 2022). When using this service, no data is retained or stored by Microsoft. Transcribing continues manually, including some pre-processing like structuring transcriptions and removing long-form explanations included in the interview. Atlas.ti 22 was used for qualitative analysis

(*ATLAS.ti | The Qualitative Data Analysis & Research Software*, 2022). The analysis follows the six-step framework for thematic analysis (Braun & Clarke, 2006).

4.7 Sample and recruitment

The selection of participants happened employing heterogeneous purposive sampling over gender, age, education levels, and domain expertise. Purposive sampling (Boeije, 2009) is a non-probability sampling technique used to meet specific requirements, commonly applied for studies with a smaller participant base as well as those that require specific participants. Heterogeneous is a measure aimed at achieving a spread over characteristics likely to cause noise. We chose purposive sampling since our study setup would mean that each participant would generate a lot of data, but also that each participant would cost a lot of time to process. Therefore, we needed to limit ourselves to a smaller sample of participants which would make probability sampling sub-par. We also did not have access to a large sample to pull from. We chose a heterogeneous measure as we were more interested in generating a broad perspective on our presented case than we were in gaining insight into specific influences of user characteristics.

Sampling characteristics were chosen from user factors described to influence credibility perceptions (1), as it was our best source for determining potential influences. Age was contentiously and unclearly mentioned as being influential for "older" individuals. We ended up deciding to characterize ages under 50 and of 50 and higher based on Choi and Stvilia (2015) using Zulman, Kirch, Zheng, and An (2011)'s work to do the same. Spread in domain expertise was only required over the two non-political topics, as politics characterize a unique reaction. Characteristics we found in the literature (1) were not considered in sampling if we perceived the time investment as too high or the complexity in accurately obtaining them beyond our ability. For example, reliably testing ability would have taken time, required understanding of how to test this for our purposes and meant that participation would have lengthened to boot.

The participant study is in Dutch. The study materials were also created in Dutch, with translations available in the appendix. We did this as we are based in the Netherlands. This means Dutch participants are more available to us. We gave preference to participants expressing themselves in their original language.

We recruited potential participants by using electronic messaging to reach out to people in social circles surrounding the main researcher. The message we sent contained a short explanation of the study, time commitment, and some research ethics and made clear people should feel free to not participate without the need to dispense a reason. We started by openly asking in groups for volunteers to participate. Some people also offered to share this solicitation in groups we were not a part of with acquaintances. On a case-by-case basis, we sometimes consented to this if it helped us reach potential participants with characteristics that we had difficulty recruiting. We retained the final say over messages sent in our name and made sure this did not jeopardize research ethics (4.9). Older and less educated individuals were initially more difficult to reach for us. To a lesser extent, this also holds for women. We resolved this by specifically asking around in groups with people that fulfil those requirements and personally reaching out to a few people who filled those characteristics. We aimed to interview between ten and fifteen people at the outset of the study. We finally ended up with fourteen participants. At that point, we already experienced more difficulty in finding new volunteers and noticed less unexpected and completely novel participant experiences.

4.8 Procedure and materials

(See [Appendix C](#) for all materials)

Our study procedure was created with input from involved researchers, which should limit bias presented throughout (Boeije, 2009). Other students involved in similar research projects were also asked for feedback for similar reasons. Furthermore, we performed a two-person pilot study (Boeije, 2009) to test, improve and fine-tune the procedure and materials.

4.8.1 Study procedure

Sessions with potential participants took a little over an hour. This was a bit more than we anticipated originally. Most of the study procedure is documented in the interview guide (C.2, C.3), but we give a general rundown of the procedure and some of our choices. We will also highlight our choice of questions and tweets a bit later in this subsection.

Our procedure can be summarized as follows: we asked potential participants to react to the displayed tweets we provided. Tweets had different authenticity methods. They would answer questions about what they perceived, how they experienced authenticity methods, and how concepts such as credibility came into play. Over the course of three rounds, we would incrementally increase their understanding of the concepts and methods involved and question them again.

We will now describe a more detailed chronological overview of our procedure. During our rundown, we will differentiate potential participants from participants by if they chronologically have given final consent for study inclusion. Before starting, potential participants recruited through the messages would arrive. Most sessions took place in an empty private office. Some sessions were only possible through a video call. We would start with an ethical explanation as well as a short rundown of the procedure (C.2, 4.9). After consenting to participation, we would start audio recording and begin. We first asked potential participants some background questions. We would follow up by explaining the think-aloud method and asking potential participants specifically to include if, how and why they would interact. We also explained that the round would be followed by questions. Further, we told them that we could revisit the tweets later. We would not yet explain anything about authenticity methods, credibility, or authenticity. Then we would show the first round of tweets each with a different authenticity method or lack thereof. Tweets were displayed one by one on a screen for potential participants to react to. We followed up with a round of questions concerning the tweets. We would inquire how potential participants perceived tweet credibility, tweet authenticity, their willingness to interact with tweets, and how authenticity methods played into this. Unique to the first round, we would also explore a potential participant's conceptual understanding of authenticity and credibility. Outside of asking them to define these concepts, we would ask them how they determine authenticity and credibility on social media and the value they tie to it. We would then start round two. We now provide a conceptual explanation of authenticity and credibility. We also asked participants to include an authenticity and credibility perception as part of their think-aloud response to tweets. Round two would then follow a similar procedure as round one. Questions exploring conceptual understanding were now excluded from the list of questions. After round two, we would give a functional explanation of both Twid and Verification and highlight what guarantees the methods provide. Then the rest of the third round would follow the same procedure as the second round. We would follow this with a few final questions concerning the experiment and Twid. Finally, we would close off with a detailed explanation of our goals and setup. If potential participants would then still consent, we would stop the recording, and thank participants for their time. We would further still offer those that made it this far the chance to discuss the materials in more detail and also correct some false information included in our materials.

The main design choice we want to highlight is the reactions to tweets being split over three rounds. In the first round, potential participants react to tweets without further explanation about the authenticity methods conceptual ideas such as credibility and authenticity. We reason that in the first round, reactions are as free of priming and influence as possible, allowing for perceptions the least bounded by theoretical conceptions relayed through questions or interviewer bias to arise. Furthermore, it allows input on the intuitiveness of unfamiliar methods and preconceived ideas about familiar methods. The second round is preceded by a conceptual explanation of authenticity and credibility. We reasoned this should obtain specific reactions within the conceptual lenses, minimally bounded by knowledge about functionality presented later. The final round is preceded by an explanation of functionality in order to obtain participant perceptions given knowledge of functionality. This was done to get perceptions specifically

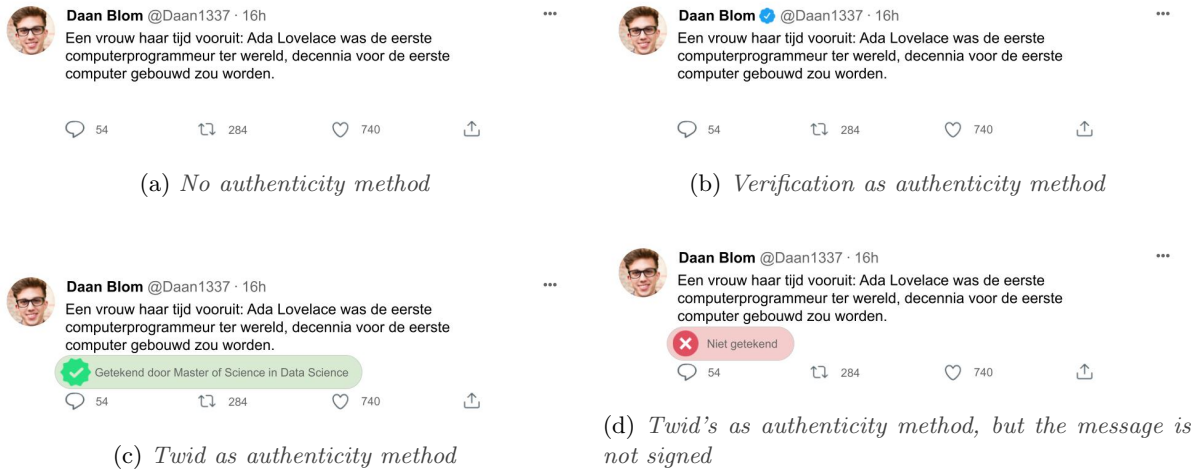


Figure 3: This figure shows all four variations of a tweet contained in the experiment (5). In Dutch, the message states that Ada Lovelace was a woman ahead of her time as she was the first computer programmer decades before a computer was built. The Twid attribute displays that the author has obtained a Master's in Data Science.

relating to that. The reason we do not initiate the study with an explanation of the functionality is that past works have taught us that users often navigate functionality without understanding and might even misinterpret functionality when they do try to understand it (3.3). Therefore, we also wanted to obtain information about a usage situation all but certain to arise.

For some sessions, we opted for a video call. This was because public health reasons, time constraints, and transportation costs would have otherwise disallowed participation for some included in the study. We already had some trouble finding potential participants. Both parties would still be in a private space for video call sessions. Given that the experiment only involved reacting to a picture on a screen, we deemed video call sessions acceptable.

4.8.2 Tweets

(See table 5 of appendix B)

Our design choices for the tweets generally strived for heterogeneity. This was done so as to not overemphasize any specific element, but also to allow us to document potential participant perceptions in a broad set of circumstances. It further offered them tools to abstractly discuss the study environment given some varied experiences.

Although we set no hard requirements for it, we strived for heterogeneity in created author characteristics over gender, names, and usernames, and looks to minimize these characteristics potentially skewing results. Pictures used for authors were obtained Pexels Pexels (2023) and could be freely used in our context without the need for further attribution

Tweets fall within three topics: politics, computer science, and food and health. Past work showed that participants tend to react differently on topics within domain expertise and politics (1). Our recruitment was likely to yield potential participants with domain knowledge of computing science and food and health because of their social proximity to us. We chose two topics to avoid the risk of reactions being skewed by some topical influence we were unaware of. We also were interested in including tweets on politics and within domain expertise to explore participant reactions given that broadened scope. Therefore, so as to not elongate study participation, our choice came to these three topics.

The tweet text was a short temporally relevant statement, ranging in tone from very subjective and personal to rather dry and objective. There is also some diversity in the truthfulness of statements. These ranges were chosen to not limit ourselves to just factual statements as in some other research (Morris et al., 2012; Simon, 2022), as our research methods allow for us to gain descriptive data from it. Topics each have messages that present variety in tone and truthfulness.

Attributes used are already included in IRMA (Irma, 2023a) or can be implemented in the future (Irma, 2023b). Some are based on ideas mentioned in the Twid proposal (van Gastel et al., 2021). Attributes were all topically relevant to the message. The difficulty of obtaining an attribute, the skill presented by an attribute, and how dependable attributes suggested authors were varied.

We randomly generated the number of interactions. All numbers stayed consistent per tweet across experiment variants. All numbers were chosen from a range in a ratio between interactions based on an observational approximation of distribution on Twitter. We chose this method to have all tweets feel more natural as they vary a bit in the number of interactions. By the number of interactions being somewhat similar across tweets, we tried to avoid a certain tweet being viewed much more favourably because of it. We chose a baseline such that it neither suggests someone who is extremely famous nor someone whose message goes mostly ignored.

Four variants are made of the experiment (C.4, C.5, C.6, C.7). Each tweet is represented in all of them, ranging over the four different authenticity states. These are a Twid attribute, the Twid not-signed indicator, Verification, and nothing. Our intent with multiple variants is to limit the effect expressed by certain tweet-authenticity method combinations. The tweet ordering is shuffled across variants. This is done to limit the effect the round or even the placement of a tweet within a round has on the results.

We used tweets to obtain direct reactions and to discuss elements that participants liked. However, our study is not about how participants perceived our tweets specifically. Therefore, our tweets were often also a way to start a conversation about what, how and why they viewed as limits of methods being used. This is also expressed in our interview guide questions asking more so questions about credibility, authenticity, interactions, and authenticity methods in a more generic sense. It is also something we used follow-up questions for. Specific reactions were more what we obtained from the think-aloud reaction.

4.8.3 Interview guide

(See C.2 or C.3 of appendix C)

Our general design considerations for questions were to make them well-worded, open-ended and not leading (Boeije, 2009; Kallio et al., 2016). We wanted it to be possible for participants to address concepts and ideas new to us, but we also were interested in delving into concepts or ideas obtained from past works. Therefore, we initiated with open questions with no theoretical conception, followed by open questions surrounding specific concepts (Kallio et al., 2016). We wanted descriptive answers and therefore designed questions to always start with "what", and "how" or otherwise be open-ended (Kallio et al., 2016). Why questions were used sparsely and only as smaller follow-up questions (Kallio et al., 2016). Questioning followed a line of going over the main themes and was followed with spontaneous questions aimed at expanding specific points from participants (Kallio et al., 2016). We also wanted the interviews to bring more than a rehash of thoughts presented in reacting to the tweets. Therefore, questions encompassed a more general abstract position on authenticity, credibility, and authenticity method than presented by the specific instances presented in the experiment. Throughout the guide and during interviews, we refer to authenticity methods as Verification methods. We did this to avoid confusion by having to introduce a new and complex, albeit more nuanced and functionally correct term. We also already conceptually discussed authenticity with participants, and we deemed adding another similar term as a risk factor in causing confusion. Pre-experiment questions track participant factors as previously described (4.7).

Question Po1.2, Po1.3, Po1.7, Po1.8, Po1.9, Po1.11, Po1.12, Po1.13 are unique post round one. We

considered having these questions as part of the pre-experiment, but this would prime and nudge participants during round one on existing concepts. These questions allowed insight into the individual understanding of underlying concepts. In turn, this allowed us to take this into consideration during the analysis and the interview. Barring the additional questions in round one, each round is followed by a limited amount (Kallio et al., 2016) of post-experiment questions. In line with our previous argumentation for interview design, this always started extremely broad. We then catered for each question to widen our understanding of participant perceptions of authenticity, credibility, interactions, and the implementation of authenticity methods within the context of the study. The last five questions in the guide serve additional functionality. The first four are included to add additional specific reflections on Twid to aid in potential future development. The last is meant to allow reflexivity on the experimental experience of the participants.

4.9 Ethics

We did not want study participation to cause disproportionate risk or grief for our participants. We felt this was not difficult to achieve for our study, but would still like to share some of our considerations and choices. We base our ethical considerations on Boeije (2009). It outlines three ethical principles and mentions ideas about participation risk avoidance.

4.9.1 Informed consent

We notified potential participants of the lack of potential risk, the broad scope, the time investment, the research procedure, and the possibility to withdraw consent at any time before participation. Only if potential participants consented, would we initiate the study. After we finished a session, potential participants were given a full explanation of the study now the risk of priming was gone. We only used their data if they still consented.

4.9.2 Privacy

We made participants aware of what data was collected and how we intended to use their data. We did not access data without participant knowledge and consent. This allowed participants to remain in control over the access we had to data concerning their private lives.

4.9.3 Confidentiality and anonymity

We protected participant anonymity by not tracking data that could be used to directly identify them, such as addresses or names. We also tried to avoid allowing indirect identification by not picking quotes that could easily infer who a person is. For reasons of confidentiality, we made clear agreements about the data. Audio recordings will be destroyed within half a year after the results are finalized. Interview data is to be securely kept for up to ten years.

4.9.4 Participation risk avoidance

We did not discuss the sensitive life events of the participants. We avoided exploiting them because participation carries no risk of note to their well-being and involved a priorly agreed upon time investiture of about an hour. Once it became clear time investiture was slightly higher, we also informed consecutive participants. Furthermore, because the study findings may positively improve society and contribute to finishing the education program of one of us, the study is not for nothing. In an attempt to limit the spread of misinformation, we also correct and contextualize all tweets post-experiment. We avoid coercing potential participants by not having any duress to participate nor recruiting in vulnerable groups and allowing exclusion at any moment without the need to dispense a reason. Although participants over 50 were recruited, non showed or shared indication of deteriorated mental faculties. We avoid sanctioning our participants because there is a lack of social stigmas surrounding our research. We also recruited from diverse social groups and participation was anonymous, so participants being subdued or subduing others to social sanctioning because of participation seems unlikely.

4.10 Position of the researcher

This is a Thesis project. This means our main researcher, Koen Verdenius, wanted to do the lion's share of the work and was also required as part of his study program to do so. He researched the literature, designed the study, executed the study, prepared the data, analyzed the data, and presented and documented this work. The other researchers served as supervisors and assessors and provided supervision, guidance, feedback, and support. They will also determine the grade of this Master's thesis. As Koen Verdenius was most intimately involved with the work, we now present his positioning from his perspective for the rest of this subsection.

I do not have any vested interest in the success or failure of any authenticity method involved in the research, nor in the success or failure of Twid. It should be noted that both my supervisors are involved in the Twid project and that they determine my grade. From the moment I became interested in doing a study involving Twid, they clearly indicated to me that results critical of the Twid project or lauding the Twid project would be equally welcomed. It was clear to me that grading would concern my academic skills. The success or failure of Twid is also not likely to have a large impact on their career or private lives. Their guidance was focused more on my process than on them steering results any which way. I felt a lot of freedom and support to design the study in the way I felt was appropriate while receiving feedback on the process. I did not feel pressured by them to steer results to present a more positive or negative picture. I never felt like I needed to obscure or highlight something I did not want to.

At the outset, I did not have a strong position or opinion about the authenticity methods involved in the research. I was mildly positive and intrigued by the ambition of Twid's project proposal. When I started to delve into both methods and began formulating more concrete ideas about the direction of a potential study, both methods brought on some mild feelings of scepticism. I support the intent of both methods in preventing impersonation and contributing to misinformation mitigation. My scepticism stems from my feeling intent was poorly reflected in functionality, or at least found I found reasoning and evidence to be insufficient. Since I also did not have proof to the contrary, my scepticism served as a big driving factor in pursuing this study, as I itched for understanding and answers. I wanted to limit potentially influencing potential participants with my positioning. Therefore, I tried not to discuss my personal position with potential participants during and before the study about ideas and concepts that could come up during participation.

I did not have experience using Twitter before I became interested in conducting this research. I decided to use the platform actively for several months to gain some personal familiarity with the platform and allow this to inform the creation of study materials (Kallio et al., 2016). I did not really enjoy Twitter. However, I have nothing against Twitter specifically or social media generally. I just personally do not enjoy using social media but hold no strong position on those that do or social media companies. I do think familiarity with Twitter positively contributed to my ability to create an environment that more accurately reflects the look and feel of messages on the platform.

This being the final project of my master's program should be indicative of my modest research experience. I do not have noteworthy research experience outside what I obtained as part of my Bachelor's and Master's programs. The course I found most relevant for conducting this research was one on qualitative research methods (Radboud University, 2022). This is the first academic participant study I ever designed and executed.

I am aware that I bring my own cognitive biases and subjective viewpoint into this research. In both the selection of potential participants and their proximity to me, my subjectivity has an effect. Furthermore, it undoubtedly impacted the construction of study materials and the themes I selected as part of the analysis and findings I present later. The subjectivity of both the researcher and participants is an inherent part of the interpretivism philosophy. Similarly, thematic analysis allows for and thrives under

researcher subjectivity. I endorse this opened armed approach as through it, I experienced flexibility in setting up this research, and it makes the final product feel personal, scientific, and relevant. I do however often highlight the limitations of the methodology as it is important in qualitative research to contextualize your work and document your process. I do not try to present some objective truth about reality, more so a well-documented collection of context-induced participant experiences presented from my position. Nevertheless, I wanted to avoid overtly pushing potential participant perceptions and opinions in a certain direction as I felt it would not help in furthering our research goal, nor be fair to and reflective of participant opinions if I were to do so. The documentation of my process, including those on my subjectivity is meant to display my reflections and facilitate the contextualization of results. I intend this to aid in this research being transferable, dependable, and confirmable (Korstjens & Moser, 2018).

5 Results

(See figure 4 of appendix A)

In this section, we present and describe the thematic framework we generated per our analysis. We highly recommend consulting the appendix to fully appreciate the findings and to be able to view all individual codes. To not misconstrue our findings, we advise you to keep our methods and their limitations in mind. Among other things, this means that the quantity of a code occurring is not indicative of its relevance, nor is it irrelevant. Self-observed behaviours or cause-effect relations suggested by either the participants or these results do not present a proven conclusion. The thematic structure is not meant to present an absolute or singular true interpretation of the data. The themes represent a methodologically reasoned but subjective interpretation of the data (Braun & Clarke, 2006). Our results are meant to help us explore and detail the experience, perceptions, and self-observed behaviour of our participants given our study context to help us provide meaning.

While analyzing the data, it was our goal to broadly highlight things that were noteworthy without limiting ourselves to a single question. Therefore, we do not have a singular research question we tunnel visioned onto while analyzing our data as we opted to keep a somewhat open mind. That being said, our main interest is in how users are affected by and experience authenticity methods. This means a big implicit question we wanted to answer is: How do authenticity methods influence user perceptions and interactions? We also wanted to encompass the broader context, potentially new ideas, and existing concepts (3.4, 3.3). All of this means that we had the following line of reasoning in mind during the analysis. Generally, we were just interested in how people reacted and perceived the study as presented to them and if there were any noteworthy or unexpected things that were brought forth in participants. Mainly for our analysis though, we explored what elements influenced the user perceptions of credibility, authenticity and interactive behaviours and how these elements influenced experiences. Within those conceptual lenses, we had a specific interest in further understanding perceptions of authenticity methods. We wanted to know what elements contributed to perceptions and how they shaped experience. This way of looking at our goals is reflected both in the way our setup, interviews, and the results of our analysis are structured.

The thematic structure follows a hierarchy of themes, sub-themes and codes one ought to keep in mind when examining it. The network model of the complete structure can prove very helpful as a visual aid for this (4, and all codes can be viewed as part of the code occurrence tables (B.1). The five base Themes encompass the broadest encompassing ideas and concepts and are mostly disjunct. Continuous sub-themes are smaller ideas contained within larger themes but still present broader categorical concepts. Codes much more distinct, are closely linked to participant perceptions, and often include the positionality of the participants. Codes are therefore much more specified ideas than the categorical conceptual themes that envelop them. Codes encompass quotes chosen from transcriptions. Most codes are perceptions that loosely present a positive, negative, or neutral experience of concepts and elements in the theme. There are also some codes that are more of a contemplative nature. We highlight this subdivision of codes by a colour-coding explained in the appendix (B.1). This subdivision is meant to be a bit loose, serving more as an aid than a definitive position on the user experience within the code. To exemplify how this can be used to examine the codes and structure, let us look at the code “Determining the value of attributes is difficult” (5.2.1.1.3). The code encompasses quotes that reflect negative experiences participants had in determining the value of an attribute. This in turn reflected poorly on the “Objectiveness of the field of knowledge” that could be obtained from the attribute, the encompassing sub-theme. This meant the “Perceived author expertise” obtained from an authenticity method decreased, leading to a decrease in perceived “Expertise” resulting in lower overall experienced “Credibility”.

The rest of this section will be dedicated to fully exploring the thematic structure we generated over our

data. In the next section, we will discuss the main implications of our results (6.2).

5.1 Authenticity

This theme encompasses perceived author autonomy and sincerity in posting a tweet. When we explained participants' authenticity, we asked them to view a post authentic if a person posted a tweet themselves or allowed others to post on their behalf, which lines up with the goal of authenticity methods to prevent impersonation. Participant observations captured within this theme mostly reflect ideas presented by our definition, but sometimes strayed somewhat from this more strict view. Occasionally, they also used authenticity to refer to authors remaining truthful to themselves in tweets.

5.1.1 Assumption

(See table 7 of appendix B)

This theme encompasses if and how participants perceive themselves to make assumptions about author authenticity in lieu of other cues. The codes mostly encompass participants leaning a certain way in their assumption with them explicitly saying that they do not have a more profound explanation for this. Participants' making comments within this theme were fairly split, with a slight majority of comments leaning towards participants assuming authors to be authentic.

Code: Assumes posts are authentic unless informed otherwise

"Let me see, well what stood out to me is that I pretty much assume all tweets are authentic. Yeah, because I really don't have any reason to doubt they are not." - 4:30 ¶ 179 in 4 (6)

5.1.2 Authenticity method

(See table 8 of appendix B)

Numerous participants brought up how the presence of authenticity methods influences them in deciding on the authenticity of a tweet. Most people saw the presence of any method as very positive for its authenticity. Many people mentioned an authenticity method only has value in determining the identity and not for much else. A few mentioned they appreciated that it was not mandatory to have an authenticity method.

Code: Any authenticity method contributes to making a post feel more authentic

"B: Yes, then I think they both add something, so the standard method and the new method both add something.....A: So if we're just talking about authenticity, then that's equivalent? B: Yes, equivalent." - 3:14 ¶ 64-68 in 3 (6)

5.1.3 Content

(See table 9 of appendix B)

The content of messages was mentioned to affect how authentic participants perceived a tweet to be. The code that most stood out to us in this sub-theme is that numerous users made perceptions relating to more subjective posts. The code encompasses participants stating that an opinion always appears authentic. Participants seemingly reasoned that any number of subjective positions are valid opinions, as people have widely differing views. Therefore, someone expressing any type of opinion is authentic.

Code: Opinions always come across as authentic

"Euh in this case I wouldn't put the authenticity in doubt. This is mostly related to it being an opinion of a large-scale issue, then I don't care if you work in a mine or are a member of Parliament." - 7:68 ¶ 179 in 7 (6)

5.1.4 Design elements

(See table 10 of appendix B)

Participants self-reported the influence of design elements outside of authenticity methods. The sub-theme has relatively few codes with few perceptions. Elements such as profile pictures and Twitter handles are mentioned by some to factor into deciding on tweet authenticity. These elements contributed to participants obtaining clues about the formality or provocative nature of authors.

5.1.5 Familiarity

(See table 11 of appendix B)

Familiarity with a tweet author was always indicated as a positive on authenticity and unfamiliarity

as a negative. Participants mostly mentioned author familiarity to contrast the unknown authors represented in the study, as they felt the accounts of authors they already presented less need for controlling authenticity by authenticity methods and other heuristics.

Code: An unknown user comes across as less authentic

“I do not have any indication to determine the authenticity of this tweet. I would not know how to recognize it. I do not know Maxime Hendriks.” - 14:13 ¶ 93 in 14 (6)

5.2 Credibility

With this theme, we captured perceptions relating to the perceived truthfulness, persuasiveness, and believability of tweets. Generally, credibility was the most important element of discussion for participants. This is reflected in explicit comments made in this regard, and the theme has by far the most codes and sub-themes contained within.

5.2.1 Expertise

Perceptions we designated to this sub-theme are more about the objective quality of a tweet and the provable skill of its author.

5.2.1.1 Authenticity method perceived author expertise

With this sub-theme, we delineate several ways participants expressed that the authenticity method influenced their perception of the expertise of a tweet author.

5.2.1.1.1 Absence of authenticity method

(See table 12 of appendix B)

The absence of authenticity methods was seen and mentioned as a negative factor in the perceived expertise of an author. Participants felt they did not know if authors were in a position to pose a claim without a method. Especially messages containing the not signed indicator were perceived poorly. The not-signed indicator drew attention towards it with participants specifically commenting on why an author would choose not to sign their message.

Code: Absence of authenticity method: Not signed messages seem like their users lack expertise

“Yeah, but that not signed indicator thing. Yeah, I think that discounts the opinion of [author]” - 8:32 ¶ 129 in 8 (6)

5.2.1.1.2 Effect

(See table 13 of appendix B)

Some users specifically commented on the impact they felt authenticity methods had on their overall credibility perception. All mentioned that they felt it was of limited impact since the methods also added little additional information. The effect was still deemed positive by some. Generally, people self-reported more so that they believed message content to be most important in forming their credibility judgement.

5.2.1.1.3 Objectiveness of a field of knowledge

(See table 14 of appendix B)

With this theme, we captured participant perceptions on the extent to which attributes feel verifiable, clearly delineated and reflect an author’s ability to make claims with a level of certainty. All perceptions in this sub-theme relate directly to attribute-based authenticity, or at least Twid’s implementation of it. Generally, objective content is mentioned as working well in tandem with knowledge-related attributes. Protected titles were mentioned as being especially useful. Protected titles received this valuation as they were seen as more standardized, recognizable and less beyond reproach. Quotes are generally more so negative in this theme. Participants mentioned that it is difficult to determine the value of an attribute of a field of knowledge you know nothing about. Participants were also afraid that experts in fields with controvertible issues can use their provable proficiency to generate an aura of absolute truth.

Code: Determining the value of attributes is difficult

“With this you introduce a quality difference in Twid which makes it impossible to check credibility gain.... I do not know all attributes. If it says gardener or architect. Those aren’t protected titles..... With a doctor, or a policeman or a judge. If

those are Twid labels, that's something not anyone can call themselves, so I assume for a judge it checks if someone indeed is working for a judicial district. This does not hold true for a gardener, which means you are introducing something terrible with Twid. You give an illusion of credibility. Sometimes this holds, but sometimes it doesn't.... I get more out of the concept of credibility. And that becomes less, because I can no longer differentiate...." - 2:44 ¶ 130 in 2 (6)

5.2.1.1.4 Relevance

(See table 15 of appendix B)

With the theme relevance, we capture how participants perceived the relation of attributes to a message as well as intrinsic attribute quality as essential in determining attribute value. Generally, participants mentioned that they preferred it if attributes displayed expertise extremely specific to the text for them to be of value. Though one person disagreed, most who discussed attributes displaying educational achievement looked at it favorably. Education relevant to the tweet was seen positive as it showed an author to have spent several years studying the discussed subject. More broadly, participants almost universally mentioned that displaying relevant domain knowledge steers participants in valuing an author's words more than authors that do not.

Code: Displaying domain knowledge helps to make a judgement

"[Twid] gives a positive feeling. Because others who are more knowledgeable said that this account or more so this content that it is true. Then you trust that, because it is their expertise" - 1:51 ¶ 167 in 1 (6)

Code: Attributes need to be very specific to the tweet to be useful

"With the last two [authenticity methods]. Yeah. Employee of 2nd chamber of parliament, could also be a coffee lady. Does not mean anything." - 11:7 ¶ 51 in 11 (6)

5.2.1.2 Content quality

This theme contains perceptions relating to the quality of the content of tweets, or at least how the content measures against the participant's subjective and objective ideas of quality.

5.2.1.2.1 Agenda

(See table 16 of appendix B)

With the agenda theme, we encompass codes that make reference to the presence or absence of an author pushing an agenda. Participants noted that for some messages they perceived the content as mild and uncontroversial. Such content was perceived as not calling out to act or change views. Those messages were then deemed more credible because the motivation to lie about the content was absent. A code exemplifying this is displayed below, where the participant is reacting to the tweet from "Daan Blom" (5) and they do not see a reason to lie about him sharing a fact deemed to be fairly tame. The tweet by "Emma Laatsbloei" (5) was deemed negatively by multiple participants because they thought it looked like an advertisement as she specifically promoted a lifestyle product.

Code: The message instils no fear that the user has motivation to misinform

"I do not know why you would lie about this, so that's why I think it is correct" - 3:35 ¶ 128 in 3 (6)

5.2.1.2.2 Foreknowledge

(See table 17 of appendix B)

This theme concerns how the content mirrored participants' previously held knowledge and opinions. The closer a message was to past ideas, the more positive it was perceived and vice versa. Most perceptions just mentioned their past ideas as a measure against the knowledge presented in the tweets without further explanation. Sporadically participants specifically referenced how their own proficiency or the opinion of an authority figure helped shape their judgement of a message.

Code: The message is misinformation based on the participant's knowledge

"Yes, those check marks have no influence on me in this. It's really pure nonsense what they write. A guy with a blue tick writing about chia seeds being a better source of omega 3, and I really think that sounds like complete bullshit so let's look it up." - 10:19 ¶ 79 in 10 (6)

5.2.1.2.3 Importance

(See table 18 of appendix B)

Several participants explicitly mentioned that to them, the message content is what is of most importance in forming credibility judgments. It is also what they mostly seek to experience when viewing tweets.

5.2.1.2.4 Logical

(See table 19 of appendix B)

This theme envelops perceptions on the experienced internal sensibility of messaging. Perceptions relate to judgements on if the argumentation is deemed to line up. There is an emphasis in perceptions on the feeling of a message being right, and making sense.

Code: The message is nonsensical

“I’ve already read the statement twice, but that doesn’t sound credible at all. Just like, how can you be a computer programmer if the first computer wasn’t around yet?” - 10:41 ¶ 120 in 10 (6)

5.2.1.2.5 Sourcing

(See table 20 of appendix B)

Evidence being provided for claims impacts the quality of a message according to the participants. A message by itself was not enough to determine if a claim is credible according to most, as it was deemed difficult to judge the value of an isolated claim without further knowledge. More so, people specifically perceived the absence of sourcing as a negative in persuading them as it suggested a lack of actual evidence about claims and lacklustre knowledgeability of the author.

Code: Lack of sourcing

“I’m not one to take things for granted without having a way to check.” - 6:40 ¶ 110 in 6 (6)

5.2.1.2.6 Spelling

(See table 21 of appendix B)

A few mentions were made of how bad spelling was perceived as something that has an adverse effect on content quality as it suggested carelessness and lack of knowledge about the content that was written about.

5.2.2 Trustworthiness

The theme of trustworthiness encompasses perceptions relating to how reliability and reputation were seen as influential in forming a credibility judgment.

5.2.2.1 Authenticity method

Several sub-themes directly relate to remarks regarding the authenticity methods. Mainly observations contained here concern the overall functionality different methods provide rather than their specific use by authors.

5.2.2.1.1 Abuse sensitivity

(See table 22 of appendix B)

This relates to how vulnerable participants thought authenticity methods were to abuse and the perceived reliability of methods. Most participants had something to say about this sub-theme, and most mentions were more so negative. Perceptions were not widely shared. Some people brought up the possibility of abuse that anonymity brings as it avoids the consequences of messages for the authors. This view on anonymity is in slight contradiction to some other perceptions indicating methods not being mandatory also holds value (5.5.1). Specific implementations of authenticity like providing identity papers or two-factor authentication was seen as positive as participants had faith in these methods. A participant mentioned how Twid could be used to provide misleading attributes from meaningless or even constructed organizations, diminishing the value of the method. One person mentioned how Verification does not stop people from using someone else’s unlocked opened account.

Code: Verification is vulnerable to someone sending a message from your opened device

“I can log in. Then my boyfriend can then send something from my phone while still on my account, then it’s still not me

who put it up.” - 9:12 ¶ 38 in 9 (6)

5.2.2.1.2 Design discernibly

(See table 23 of appendix B)

Perceptions relating to the experienced intuitiveness of design. All but one user made one or more comments relating to the design of an authenticity method being unintuitive. This held equally true for Twid and for those for whom Verification was less well known. This meant until explicit knowledge about the working of a method was obtained, it was deemed essentially useless. Some users chose to ignore methods where explicit knowledge lacked as they were unable to evaluate their worth for formulating a judgement. Others used their interpretation of a method’s functionality to form a judgment. An example is presented in the first quote below, the users’ intuition is poorly supported by actual functionality. Furthermore, the not-signed indicator was seen as especially harsh as those not willing to sign were suspected of malcontent and generally deemed untrustworthy. Finally, a few participants mentioned how the obvious presence of an authenticity method strongly steers their judgment.

Code: Unintuitive design of the authenticity method

“Well I still believe the blue badge means that a certain check is done about the truth of it or the trustworthiness of the people” - 13:6 ¶ 64 in 13 (6)

Code: Unintuitive design of the authenticity method

“Not signed, I do not understand what that means here” - 12:4 ¶ 40 in 12 (6)

5.2.2.1.3 Reputability

(See table 24 of appendix B)

The theme encompasses perceptions about how the standing of an authenticity method influenced participants. Participants explicitly mentioned how familiarity with an authenticity method can aid its reputation through understanding its workings. However, explicit explanations about Verification’s exact functionality being presented to participants actually decreased its perceived reputability in some users. Specifically, several participants had held the belief that Verification represented some quality control or editorial responsibility from Twitter. To their disappointment, this was much less true than they thought.

Code: Understanding of the workings of verification

“B: Credibility of [Verification] even reduced. Quite a lot, actually. A: How come? B: Yeah, the fact that. There is just no control anymore, you do something once and then they don’t really check on you anymore ” - 5:116 ¶ 215–216 in 5 (6)

5.2.2.2 Author

This sub-theme clusters all perceptions having to do with the perceived trustworthiness and reputability of an author.

5.2.2.2.1 Authenticity method

(See table 25 of appendix B)

This theme is meant to capture attribute influence over the author’s trustworthiness. Almost all codes in this sub-theme were not broadly shared, despite most participants making perceptions contained in this theme. Generally, people saw authenticity methods as a positive adage for authors. Relevant attributes brought a feeling of trustworthiness and displaying an attribute suggested a sense of authority according to some. A participant also mentioned that because Twid requires more effort to use for authors, it adds to their trustworthiness. Verification was mentioned as making an author a bit more believable because of trust in the method. Not everyone agreed on the positive influence of authenticity methods. Some criticized them for failing to provide any formal guarantee over message content, leading to a sense of false authority and unjustified faith in a message presented by an authenticated author.

Code: Authenticity method: Any authenticity already contributes more than nothing at all

“Yeah, you do notice that as soon as any type of verification happens, I dispose it slightly more positive ” - 7:48 ¶ 132 in 7 (6)

5.2.2.2.2 Author message relation

(See table 26 of appendix B)

This theme bundles perceptions on the connection between the author and content. The theme is in a similar vein as how we described that content that does not call to action or is perceived as mild is deemed more credible in agenda (5.2.1.2.1). Perception in this theme more so reflects more on the authors' affiliation and how the choice of message is motivated instead of how the content presents or lacks a clear agenda. Participants commented on how an author's expertise does not disallow them from unethically pushing an agenda. Experts could still be bought, and politicians were additionally perceived as untrustworthy. Outside of that, mention was made of how an author attribute combination can be seen as the source if they also authored evidence quoted in the message. A theoretical example explaining this observation would be Einstein mentioning details of his works on special relativity without further reference.

Code: Expertise does not exclude the possibility for authors' to push an agenda

"Yeah, that does not do much for me. Carvan Cevitam supposedly was recommended by medical doctors in the past, which was absolute nonsense" - 13:7 ¶ 65 in 13 (6)

5.2.2.2.3 Fame

(See table 27 of appendix B)

Fame represents the perceived renown of an author. Perceptions relating to this theme are often related to the notability requirement contained within Verification. Some perceived that notable users must have been credible to get to the position they are in, perceiving Verification as a status-based credibility indicator. Quite different from this, several mentions were made of how notability increases the chance that others manage the author's account. This is perceived as decreasing the author's trustworthiness since it no longer directly reflects the author. One person further observed this as an advantage for authors lacking Verification as their tweets are deemed more likely to be sent by them personally.

5.2.2.2.4 Social proof

(See table 28 of appendix B)

This relates to how participants perceive they and other users are being influenced by the opinion and expected behaviour of others. The theme is closely related to the description of fame (5.2.2.2.3), but represents less so a valuation of renown and more of a wider trust in a functioning social system. One user made mention of low interactions of a post and saw that as a negative. Furthermore, several mentions were made of how the expected backlash of lying is enough reason for people to be truthful.

Code: The potential backlash of lying is reason enough for it to be true

"Yes, because of that badge, I would again think it's real. Again because I think otherwise you will be held accountable for it." - 5:25 ¶ 121 in 5 (6)

5.3 Interactions

Perceptions relating to the available interactive behaviours. The theme is meant to bundle reasoning behind the intent to or not to interact. Thematic separation was more so apparent along the different interactions due to data sparsity and differences in the outcome of available interactions. Retweeting was never explicitly mentioned as an interaction intention. For the interactions mentioned, it occurred frequently that people indicated that there was a likelihood for or conditional to interaction more than outright stating they were going to interact. A quick fact check proving the tweet would be an example of a commonly mentioned conditional.

5.3.1 Comment

(See table 29 of appendix B)

The reasons participants gave for commenting were to ask for sources and to correct something that was to their best knowledge wrong. Asking for sources usually stems from an inherent interest in the topic discussed in the message, and want to therefore learn more about how they came to their conclusions. A

correction was mostly mentioned for messages that frustrated users for their obvious perceived falsehood.

Code: Correcting a message that is wrong

“Yeah, hmmm. Here I would want to give a reaction. Because that, so if you say these weird things, which they say, than I am more inclined to react. And I think he is wrong.” - 1:4 ¶ 37 in 1 (6)

5.3.2 Interacting

(See table 30 of appendix B)

We bundled observations on interactions that did not specify an action in this theme. Almost universally mentioned by participants was that they did not want to interact because it was just not interesting. Abstaining from interacting was justified by participants mentioning a general proclivity to do so, a lack of credibility of messages, a lack of authenticity of authors, and an unwillingness to further debate. Some mentioned any authenticity method as positive in their inclination to interact as it did improve credibility.

Code: The message is not interesting

“Again, this is not something that makes me think, that appeals to me. So I would not share it, retweet or otherwise interact.” - 8:43 ¶ 172 in 8 (6)

5.3.3 Like

(See table 31 of appendix B)

Participants gave tweets a like if they perceived them as interesting or agreeable. Furthermore, one participant dissented from what two participants said earlier on their decision not to interact because of an unwillingness to further debate (5.3.2). This participant specifically liked to achieve dispute as they liked stirring up discussion from time to time.

5.3.4 Share

(See table 32 of appendix B)

Perceptions of sharing reflected the more private nature of the Share feature. People only gave the intent to use it if they were already discussing things with, protective over, or aware of the interest of their acquaintances. This allowed them to discuss the message with them in private without publicly doing anything with it.

5.4 Quality concern

Participants made several comments on the study and its encompassing topics that were more reflective and meta. Unlike the sometimes more implied nature of the previous themes, these were often direct comments on the quality of social media and authenticity methods provided details. Some perceptions are also more directly related to preferences in the functionality of certain authenticity methods.

5.4.1 General

(See table 33 of appendix B)

Many people mentioned how they found it important that a study such as this one was conducted. They found it important that misinformation on social media is being addressed. They also supported the continued development of new solutions to mitigate problems on online platforms. On the authenticity methods, people gave preference to an implementation that would be available to everyone, however else it functions. Some also mentioned specifically that they would prefer the perceived functionalities of both methods of authenticity to be combined. With this, participants meant to state that Verification confirmed account authenticity and Twid contributed to message credibility. This was not widely shared.

5.4.2 Twid

(See table 34 of appendix B)

This theme captures perceptions explicitly of Twid as well as perceptions of functionality only provided by Twid. A condition participants saw for Twid’s success was that expertise attributes have to be from a trusted authority and cannot be faked so as to not lose significance. It was also mentioned that a high adoption rate is required for it to be useful. Mobile availability, insight into author conflict of interests,

and some peer review ability were recommended as improvements. Several participants mentioned their desire to use Twid, as they felt like they had relevant attributes of their own they would like to display. On the negative, one user indicated Twid’s similarity with Twitter’s name was unnecessarily confusing. We previously covered how a participant perceived an increase in credibility because the author expanded the effort to use Twid (5.2.2.2.1). However, some participants also mentioned effort as a negative in using it themselves because of an unwillingness to expand additional energy in doing so. Arguably this is also relevant for the later theme of Scrutiny (5.5) Finally, some users mentioned that reliance on a heuristic measure such as Twid that provides no guarantees over content quality would only encourage a lazy user base that believes tweets too quickly.

Code: Twid encourages a lazy user base

“I still think there may be a risk that people will no longer fact check.” - 6:80 ¶ 173 in 6 (6)

5.4.3 Verification

(See table 35 of appendix B)

Explicit perceptions of Verification were few and far between. However, the focus on identity over content, the semi-permanent nature of Verification, and needing to provide Twitter with your identity papers to be verified were all mentioned as negatives. One participant mentioned that just extending Verification to all users would be the ideal solution for them.

Code: Authenticity methods focussed on identity over content are not helpful

“Nothing. That doesn’t mean anything,... he is, who he says he is. But that in itself says little about the quality of the contribution.... I do believe that they have been verified. So I believe that someone who says who is if they have a certain check mark, that they are. But that doesn’t automatically make his tweet worth more or less.” - 2:5 ¶ 44 in 2 (6)

5.5 Scrutiny

The theme of scrutiny encompasses the perceived effort participants have to expend to be able to believe, interact or confirm the authenticity of a tweet. It encompasses ideas similar to what we mentioned for the effort of using Twid (5.4.2). Scrutiny has more to do with perceptions of the evaluation of information rather than the prospect of posting and signing yourself. It is a seemingly contradictory yet understandable requirement. Decreasing the effort of evaluation seems to be what participants are looking for. However, if a heuristic measure lowers scrutiny without providing expected guarantees over quality, it is mentioned as undesirable.

5.5.1 Authenticity method

(See table 36 of appendix B)

Generally, participants indicated that authenticity methods lower the level of scrutiny, especially in the case of the attribute system displaying relevant expertise.

Code: Any method allows less scrutiny

“[In reference to presence of authenticity] So then I don’t have to do all those control steps I talked about earlier. I can always do it. But the incentive is removed for me.” - 7:75 ¶ 192 in 7 (6)

5.5.2 Foreknowledge

(See table 37 of appendix B)

Familiarity with an author and a tweet matching up with pre-existing beliefs were each mentioned by a pair of participants to lower the level of scrutiny.

5.5.3 Methodology

(See table 38 of appendix B)

Several users went into specific ways they scrutinize tweets, as well as ways they scrutinize information elsewhere on social media. Reading comments is specifically mentioned as a valid strategy for scrutinizing information. Doing additional research was often mentioned as a pre-cursor for believing interesting tweet information, or for perhaps deciding to interact. Furthermore, other users specifically mentioned they always fact-check relevant claims they cannot immediately validate because they know that information

on social media can be dubious.

Code: Warrants further research because of interest in tweet

"I noticed, as mentioned, I actually almost always check everything I see before interacting with it" - 7:27 ¶ 98 in 7 (6)

6 Discussion

In this section, we will go over the results in an effort to ascribe meaning to them. We first will set out to place our findings within existing work. We are aware that our thematic findings as presented in the results are large, broad, and come across as somewhat convoluted. Therefore, we will follow up on placement by highlighting a combination of findings that we found most noteworthy. This will be followed by a critique of our research process and findings. We do this to contextualize our findings, present our shortcomings, mention our failings, and delineate limitations. Finally, we will explore some points for continued work based on this research. This will also cover some ideas for future work less so based on our findings and more so on ideas we see as interesting or essential that came up as a result of our research process.

6.1 The novelty and prevalence of themes and perceptions in literature

We want to provide some broad placement of most themes, codes, and findings from our results within the literature. This is because not everything from the results will be included in the key findings. Nevertheless, we found many links to literature and were also aware of the novelty of other findings. Therefore, we wanted to place what we can in case someone is interested in elements not present in the key findings. Some of our discussions will connect small findings from our study to larger bodies of research. We, therefore, do not recommend using this as evidence outside the context we provide. The rest of this subsection will follow our thematic structure.

6.1.1 Authenticity

We did not find a lot of literature connections for the theme Authenticity. We attribute this to us highlighting authenticity as a separate concept more than similar works have done (Morris et al., 2012; Vaidya et al., 2019), making our results somewhat novel. Barring the finding about the authenticity of opinions (5.1.3), our results did not seem noteworthy or surprising. Our analysis showed how participants mention familiarity with authors as a common theme in determining authenticity (5.1.5), which has some overlap with the concept of corporate credibility and reputation from previous works (3). Participants in our study generally perceived that Verification only is useful in determining author identity. This is functionally correct and also what Vaidya et al. (2019) found. However, since there are also codes contradicting this, it is not a consistent result. Past works on credibility covered how bad spelling and shabby design can be detrimental (2), and we found similar participant influences on how they perceived authenticity (5.2.1.2.6).

6.1.2 Credibility

There is a lot more applicable work on perceived credibility and author credibility than on authenticity. As such, we found a lot more connections.

6.1.2.1 Expertise

Generally, participant observations concur with previous descriptions of expertise we examined (5.2) (3.4).

6.1.2.1.1 Authenticity method perceived author expertise

The participants mentioned the absence of an authenticity method in their credibility judgement (5.2.1.1.1), which is novel. We do not feel this is necessarily very representative of normal behaviour, as our study explicitly brings attention to it. Past works suggest operator expertise (3) to increase with author credentials and affiliation. We saw this reflected in participant perceptions of how attributes displaying expertise in a field deemed objective are perceived positively (5.2.1.1.3). Participants also mentioned how it can be very difficult to judge the value of a Twid attribute. This perceived difficulty in evaluating an attribute or expertise seems novel. This difficulty in determining attribute value further suggests the credibility issue just moved from the message to the attribute. Participant perceptions on the relevance of an attribute (5.2.1.1.4) concur with the positive impact of a high perceived source

expertise, topical expertise and affiliation to a trusted institution from past works (3).

6.1.2.1.2 Content quality

Generally, description from the literature on message content quality (2) matches those of participant observations within the content quality theme. Perceptions of authors having an agenda resulting in a conflict of interest (5.2.1.2.1) are very similar to the description of the negative effect of message congruity with the source's self-interests (2). Participant description of the positive effect of content matching foreknowledge and past beliefs (5.5.2), aligns with previous ideas on the credibility impact of preconceived knowledge (2). Participant description of the sensibility of argumentation (5.2.1.2.4) concurs with the previously explored idea of fluency (2). The presence or lack of evidence being perceived by participants as a factor in their judgement (5.2.1.2.5) is in line with past work on sourcing (2). The importance of good spelling being brought up by participants (5.2.1.2.6) was previously detailed in works on the influence of typographical errors and design (2).

6.1.2.2 Trustworthiness

Similar to what we described for Expertise, perceptions and themes in the Trustworthiness theme mostly concur with past ideas.

6.1.2.2.1 Authenticity methods

We found participants mentioning several ways they felt authenticity methods influenced their credibility judgement. The credibility of authenticity methods is a somewhat new subject, as authenticity methods are somewhat new. Therefore, description from literature can be a bit meagre and findings novel. Participants made mention of their worries about the potential of abuse and vulnerability of methods (5.2.2.1.1). This can be construed as a criticism of the ability of authenticity methods to provide accurate reinforcement of content expertise and method ability to allow insight into author credentials and affiliation. Choi and Stvilia (2015) previously described these factors as ways for users to judge the expertise presented by authors, platforms, and content (originally: operators & content) on the web. Observations indicated that unintuitive design was something that came into play for participants (5.2.2.1.2). Vaidya et al. (2019) based on similar ideas relating to google iconography (Felt et al., 2016), already indicated this risk with Verification. It was discussed quite extensively by participants, although for two-thirds of the study, they were also unaware of the underlying functionality of Twid. Therefore, the amount it was discussed for this study should not necessarily be seen as completely representative of the performance of the provided methods. Some participants falsely assumed they knew the functionality provided by Verification and ran with that. Some also just assumed a definition for Twid which they applied before obtaining an explanation. A few participants had difficulty understanding Twid after it was explained. This was expected as Twid implements IRMA, which is also considered difficult to comprehend by its managing foundation (Irma, 2023a). Perceptions in the theme method reputation (5.2.2.1.3) are in line with previous work on corporate credibility (3). Distinct from past work is how credibility here relates to the method, or medium, and not the author. Therefore, the impact we found is changed and its presence is perhaps more significant. Since the impact covered for our findings affects all usage of methods regardless of author, it is at least more present. Similarly, familiarity with an author as a factor (3) is expected based on past findings. Yet participants describing it as a factor of method reputability (5.2.2.1.3) changes its impact.

6.1.2.2.2 Author

Observations on the impact of authenticity methods on author credibility (5.5.1) are in line with past work. We knew an author's perceived integrity, transparency, decency and above all, perceived reputation and expertise to positively impact credibility (3). Given these works, credibility can be expected to be increased for authors who use Twid as intended. Intended here referring to usage to display relevant affiliation over message content. This was in line with the observations. Participants' perceptions of the fame theme contained an assumption that authors need to be truthful to reach a certain level of renown (5.2.2.2.3). This seems faulty logic, as if this assumption were true, famous individuals would not lie

or lose their fame as soon as they did. Nevertheless, some past works have shown similar effects for messages from authority figures. The lab coat effect (Shaw, 2013) describes how being perceived as a generic expert or authority figure increases your credibility. This effect is similar, although Verification does not necessarily showcase expertise or authority. Morris et al. (2012) described how reputation through a history of platform engagement can have a similar effect as what we found. Furthermore, Verger (2021) describes this for social media influencers, although their findings were more pronounced for hedonistic product placement than for an actual belief in author competence. Participants mentioned how the relationship between the author and the message is relevant in judging if someone is simply furthering their own goals (5.2.2.2.2). This is previously described as congruity with self-interest (3). Perceptions of participants on the number of interactions serving as a credibility indicator (5.2.2.2.4) match the description of how the perceived influence of an author may make them seem credible (3). Outside of that, perceptions in the social proof theme (5.2.2.2.4) appear to conform somewhat with past works on social proof (Cialdini, 2007), herd behaviour and expectations expectation (Luhmann, 1985). These works describe how individuals are influenced by perceived societal expectations and how we try to adapt our behaviour by anticipating the anticipations of others in a process of constant flux. Similarly, participants ascribe influence to how they expect the current social system on Twitter to function.

6.1.3 Interactions

Participants indicating that they want to comment for correction or to ask for sources has some relation to the novelty of presented information and emotional response brought on by a disagreeable and potentially dangerous message (1). Both reasons can also be attributed to the need to participate in online gossip (3.5). Since we can make an educated guess that any interaction boosts the reach tweets have, these reasons for commenting may inadvertently go against the participant's implied intent of identifying and limiting the spread of misinformation. If these perceptions of participant intent to comment are related to behaviour, it would also give some indirect, support to Vosoughi et al. (2018) findings of falsehoods diffusing quickly.

6.1.4 Quality concerns

We see participating in an experiment on social media and authenticity methods as the main reason for the prevalence of perceptions within this theme. There were also explicit questions geared towards perceptions of this theme. We did not see it prudent to search for a further connection to past work as we expected it to be both thinly supported and irrelevant.

6.1.4.1 Twid

Participants perceived trusted authorities providing Twid attributes as a necessity for Twids' success. These observations are similar to ideas of how we value affiliation with trusted institutions for our credibility judgements (3). Participants also discuss how they would like insight into author conflict of interest and would appreciate peer reviewing functionality. These suggestions would allow participants to better judge authors on their congruence with self-interest, an idea from the literature we mentioned earlier (2). Participants mentioned how mobile availability was a hard requirement for them. This was said by those who also stated they solely use mobile Twitter, so seems sensible from a more practical usability standpoint. We also see this as an expression of participants wanting to increase affordance through interactivity, as described in Sundar (2008)'s model used for understanding how technology may affect credibility. Participants demanding a high adoption rate for usability reasons would allow Twid to be consistently useful, as per Sundar (2008)'s and Choi and Stvilia (2015)'s description of credibility through guarantees over-reliance on technology.

6.1.5 Scrutiny

The theme scrutiny bundles perceptions to combine two existing ideas with plentiful past descriptions. Users tend to look for a path of least resistance, preferring online technology that is easy to use Sundar (2008). Features that force users to think more about social media posts increase time spent on checking

quality (Lewandowsky et al., 2012; Pennycook et al., 2021). Although ideally one would increase information quality and make it easier to use, this may be difficult to achieve.

6.2 Key findings

In this subsection, we present the findings that we found most essential to highlight. We try to tie together a combination of our results and existing literature to make points that are more coherent and overarching than our previous presentation of thematic structure.

6.2.1 The slightly positive perceived effect of Authenticity methods

In our study, participants reacted to messages with Twid attributes, a Verification badge, and the absence of authenticity methods. Participants perceived both authenticity methods more so than not to make a positive contribution towards tweet authenticity (5.5.1) and credibility (5.2.1.1.1, 5.2.1.1.2, 5.2.2.2.1). Perceptions highlighting familiarity with an author (5.1.5), foreknowledge over message content (5.5.2) or the fluency of messages (5.2.1.2.4) mentioned those factors to be more influential than authenticity methods. The credibility impact of authenticity methods was generally indicated to be limited (5.2.1.1.2). These findings are in line with previous findings on the impact of Verification (3.1).

Therefore, we surmise that authenticity methods could be developed as a measure to influence users. It can be used to aid in the perceived authenticity and credibility of tweets. It may also be developed as a measure to address misinformation. However, we expect the overall impact authenticity methods will have on users to be minor.

6.2.2 Risk for misinterpretation

In the theme design discernibly (5.2.2.1.2), almost everyone mentioned or made perceptions that showcased that the provided authenticity methods can be unintuitive. Participants sometimes also had perceptions of Verification regarding perceived guarantees over author expertise not supported by actual functionality (5.2.2.1.3, 5.2.2.2.1). Similarly, perceptions of platform moderation are not reflective of reality (5.2.2.2.3). Literature on similar innovations (3.3) has shown that it is unlikely that everyone will understand or even try to understand the basic functionality authenticity methods provide. Misinterpretation of design elements has previously led users to have been unwittingly exposed to unsafe websites (Felt et al., 2016). Our results suggested that misinterpretation of authenticity methods could lead users to believe undeserving authors. We also note that almost all major social media platforms use some form of authenticity method (2.7).

As authenticity methods are prevalent, potentially ill-understood by their user base, and risk an unwanted impact on users if interpreted incorrectly, authenticity methods should strive to be as intuitive as possible to avoid risking an adverse effect. Our results suggest that both methods tested still have room for improvement.

6.2.3 Verifications ambivalent impact

Despite recent changes to the functionality of Verification, perceptions of "legacy" Verification bear much relevance. Most countries still use the tested method (5.4.3) and most major social media companies have a very similar method of account authenticity (3.1). In our findings, participants appreciated how Verification was helpful for determining authenticity (5.5.1) and participants perceived it to slightly contribute to credibility also (5.2.2.2.1). Participants mentioned how the simplicity and abuse safety of needing to provide identity papers (5.2.2.1.1) is something they appreciate. However, its user base restriction was seen as a negative (5.4.1, 5.4.3). For some information, the gain was deemed so limited that it was questioned if it has a significant impact at all (2.7, 5.2.1.1.2). As part of our results, we also potentially find issues for Verification achieving its main purpose of being able to prevent impersonation (2.7). Namely, participants' perceptions indicated that unverified users can be assumed as authentic when using opinions (5.1.3). This would partially nullify Verifications' intent, as it would enable authors to falsely use someone's name for their account if they just use subjective language for their message.

Furthermore, some users saw a positive effect of credibility caused by their perceived fame (5.2.2.2.3) and elements of social proof (5.2.2.2.4) that can be attributed to how they perceived the functionality Verification provides. This concurs with Vaidya et al. (2019) worry that the notability and public interest element of Verification opens itself to misinterpretation.

Our results make it difficult for us to definitively state anything on Verification beyond the contention from past works. Verification may have a positive impact in line with what it intends to achieve, but it also risks misleading users beyond its provided guarantees.

6.2.4 Twid’s potential for risk and success

Twid’s provided functionality was often seen by participants as a positive influence on tweets (5.5.1, 5.2.2.2.1). Participants generally appreciated the extra author information provided, especially for displaying relevant domain expertise over objective content and providing insight into the relationship between an author and their message (5.2.1.1.4, 5.2.1.1.3, 5.2.2.2.2). Contradicting Simon (2022), participants suggested improvement of Twid through the removal of the not signed indicator (5.4.2) and they also made perceptions indicating its undeserved negative impact on some participants (5.2.1.1.1). We say undeserved here as a message with a not-signed indicator provides equal guarantees over a message as the absence of authenticity methods do, yet was perceived more negatively. Further suggestions included changing Twid’s name to something less similar to Twitter, insight into conflict of interest, and adding functionality for peer reviewing (5.4.2).

Participants would have preferred the method to be available on mobile devices (5.4.2). This current lack of mobile support reflects poorly on Twid’s goal of availability (2.9), as Twid intends to be available to all, yet users view its current implementation as a barrier to using it. It should be noted that this limitation may be attributed to Twid currently being more a proof of concept. Although not tested for this Thesis, implementation of Twid through IRMA can hardly be called user-friendly as the Privacy by Design foundation even lists effort and the requirement of user knowledge as a main disadvantage to IRMA (Privacy by Design Foundation, 2023a). This usability issue could further be an issue for availability as intended for Twid.

Participants also made mention of a high adoption rate is a condition of making it useful (5.4.2). This reflects poorly on Twid’s goal of scalability (2.9). Technically, Twid is indeed scalable. However, if a critical mass of users is required for it to be deemed useful, this advantage of scalability only comes into play if that can somehow be achieved.

Participants feared it would be easy to elicit false authority using readily available, seemingly impressive but mostly meaningless attributes (5.2.2.1.1, 5.2.1.1.3, 5.4.2). Given a field of knowledge, judging attribute expertise, relevance and overall absolutism afforded by an author was seen as something that requires a level of knowledge not afforded by a layman (5.2.1.1.3). To boot, Twid was correctly perceived as not guaranteeing content quality (5.2.2.2.1). Experts using Twid to send paid corporate endorsements would be able to do so and might achieve a gain in credibility compared to Verification, despite no additional formal guarantees being provided. Participants also shared worry about Twid encouraging a lazy user base (5.4.2). All these participant perceptions explicitly and implicitly highlighting the issues with Twid’s suggested implementation of attributes reflect poorly on Twid’s intent to shift focus towards domain knowledge (2.9), since if attributes are difficult to comprehend and vulnerable to abuse their functionality is either nullified or risks unintended consequences. Additionally, attributes being difficult to interpret may also mean that although Twid is indeed scalable (2.9), its impact may be limited if attributes are not interpretable by laymen.

We also wanted to provide a final critique of Twid’s reasoning that is more reflective of our own updated positioning post-analysis rather than it being entirely reflected by participants. The following should therefore not be seen as resulting from the thematic analysis. Twid’s method seeks to allow users the ability to judge based on self-reported attribution if they deem an author a relevant stakeholder to

their message. This is an approach that suggests self-reported attributes are a better source of online information quality than their absence or Verification. Attributes can be used to display your experience, proximity or expertise in a legitimate way. However, there is no control over how you actually use them. This means attributes displaying personal experience can be used for an argument from the anecdote fallacy and attributes displaying expertise can be used as an appeal to authority fallacy (Bennett, 2012). Therefore, Twid cannot be seen as a reliable measure of integrity or quality. It is somewhat questionable if a measure meant to limit disinformation should be so open to being used as a means for providing false arguments of a sort.

All of this does not allow us to definitively say that Twid will be of good or bad influence, especially because it is still in development. It does provide cause for concern because of the observed pitfalls, especially those in conflict with its own goals. Before Twid is considered for release, these negatives should be further investigated, lest the cure is worse than the disease.

6.2.5 Subjectivity’s apparent positive reception

When participants were making assumptions about authenticity (5.1.1), a message being an opinion was perceived by some as a strong indication of authenticity. Although participants were aware of the possibility of authors having an agenda (5.2.1.2.1), some also mentioned the positive credibility of messages with no obvious motivation to lie. A belief among participants was also prevalent that the potential backlash of lying kept authors in check (5.2.2.2.4). Misinformation campaigns use subjectivity and may have motives that are unclear to their recipients (Nemr & Gangware, 2019). For example, these ads used to influence the 2016 elections were almost all highly opinionated (Bhardwaj, 2018). The prevalence of misinformation also discredits notions of social control countering misinformation, with past work even suggesting the opposite to be true (Vosoughi et al., 2018). Although factors relating to subjectivity such as pre-existing beliefs and emotional response have been known to influence users (3.4) (3.5), we did not find evidence to support or contradict our results on the potential impact of subjective or opinionated content. We were also not able to find statistical information on the extent to which misinformation is presented subjectively.

If this finding proves to hold, it implies a big shift in the way we need to look at misinformation. In a large body of research on misinformation, credibility, impersonation, and Twitter, clinical and objective tweets and messages are used to provide evidence of an effect (3). It is unclear if these effects hold with more opinionated messages containing misinformation. Interventions in misinformation, including Twid, also focus on factual information. Subjective information may therefore fall out of scope for several interventions.

6.2.6 The credibility of platform confirmation

Although perceptions relating to the fame of authors (5.2.2.2.3) and social proof provided to them (5.2.2.2.4) were mixed, some perceived renowned authors as more credible by rights of being verified (5.2.2.2.3). Furthermore, some participants perceived authors to be truthful as social proof and platform correction would “automatically” resolve this otherwise (5.2.2.2.4). The main implication is that Verification and factors reflecting social proof can collectively be seen as a form of platform confirmation that provides credibility. The publicity around Twitter and the recent changes to Verification may reduce the relevance of this finding on the platform. Yet, as we already mentioned, in most countries’ Verification on Twitter remains unchanged and on most other platforms it functions almost identically to what we tested. Since perceptions related to the notoriety of users and general trust in the platforms’ ability to self-correct, and influencer leverage on public opinion is ever relevant, it is worthwhile to figure out the truth of the matter.

6.2.7 Scrutiny and usability as requirements

Perceptions captured in the scrutiny theme showed how users spend less time and energy checking tweets if more information is given by the authenticity method (5.5). At the same time, some users

worry about how a method such as Twid might encourage a lazy user base and mention that they do not want to expend effort themselves to use a method (5.4.2). We also know from past works that although a subtle shift of focus can improve information quality (Pennycook et al., 2021), social media rewards novelty over quality (3.5) and interventions can backfire (Lewandowsky et al., 2012). Furthermore, both methods tested provide no guarantees over message quality, despite being perceived in relation to quality. Therefore, treating information quality and platform usability as interrelated and relevant requirements might positively impact the development of methods.

6.3 Critical notes and limitations

For this subsection, we report on the constraints and limitations of our findings. We already laid out how our methods and the small scale of this study bound the conclusions that can be based on findings in methodology (4), but we will revisit some of that and expand upon it. We will also lay out some of our flaws in implementing the design and some further unexpected issues we ran into. In line with applied methods, our findings should be seen as exploratory stepping stones for future research. Although our findings have relevance within their context, presented findings are preliminary conclusions based on a subjective interpretation of the best available data rather than proven fact. Although we will generally be quite critical of our work for the rest of this subsection, we do not mean to diminish our findings or present the idea that our approach is inherently wrong. Our approach allowed us a uniquely flexible approach for obtaining a rich description of user experience within a field where such information is lacking. We merely wish to document a complete picture with the succeeding critique so that our conclusions can be appreciated within context.

6.3.1 Application of thematic analysis

Our application of thematic analysis is not in perfect alignment with how Braun and Clarke (2006) would ideally do so. Partially this has to do with inexperience in applying the method, as well as with choices to delineate from past ideas.

Our choice to present large categorical concepts that aligned with domains of questioning as themes are disputable. It could be argued that at least Credibility, Authenticity, and interactions and perhaps also expertise and trustworthiness are more reminiscent of domain summaries Braun and Clarke (2019) than they are of Themes. It was difficult for us to definitively distinguish them for our study. A domain is an area of the data, encompassing everything participants had to say about something and often related to certain questions. Themes represent concepts and are a collection of observations with shared meaning. Our difficulty was that questions and domains of discussion also adhered to thematic lines in our work. We do feel that our main themes encompass shared meaning. Some main themes were entirely new, while others encompassed ideas beyond what we questioned. Especially our sub-themes delineate relevant sub-concepts that were not directly related to lines of questioning. We, therefore, defend and stick with our chosen decision, but do want to acknowledge its ambivalence.

Another choice we made that deserves some explanation is our frequent use of sub-themes. Although we would usually agree that too many sub-themes should be avoided, we feel the size of our data allowed for it. Furthermore, we did clearly see the sub-themes we chose as separate concepts that could be described as part of larger themes yet contributing to their parent theme. We did this to highlight aspects of a large dataset and not to present a definitive separation of how all people experience elements encompassed in our study.

Although we avoided it, a few codes were more ambivalent and proved difficult to place. An example of this would be how the "Expanding extra effort to use an authenticity method yourself" was placed in the Quality concern theme (5.4), but would also be somewhat appropriate in Scrutiny (5.5). Where we were unable to resolve this, we chose to mention similarities and connections.

It is undeniable that the research we did before the analysis influenced the thematic structure. This

influence can be attributed not only to our knowledge when conducting the analysis but also to how the study design was already based on our knowledge of past work. An example of where this is clearly present is our division of credibility into expertise and trustworthiness (2.4). Although inductive research methodologies like grounded theory (Harris, 2014) champions not delving too much into existing work, designing a guide for a semi-structured interview required us to familiarize ourselves with the subject up for analysis (Kallio et al., 2016). Similarly, we felt some knowledge was required to design the study environment. It is nevertheless undeniable that some theoretical deductive elements are introduced into our study and the way we approached analysis as a result (Braun & Clarke, 2006). It is likely that this also influenced how participant perceptions presented certain concepts as we fed them with the materials and questions to react upon. We again bring attention to the themes of interaction, credibility, and authenticity. They present clear conceptual themes but were also part of a direct line of questioning in the interviews and their presence and perceived influence on participants should therefore not be seen as disjunct from our design. To further highlight some themes that might be represented a certain way because of our study design, user expertise (5.2.1.1) is something Twid specifically calls attention to. Calling to the limited effect (5.2.1.1.2) might only come to those that feel less affected, rather than those who do not mention effect but are affected. The Absence of a method (5.2.1.1.1) is something that the study is also set up to call attention to.

For this research, we attempted to bundle perceptions that shared some commonality to generate themes that were helpful in achieving our research goals and highlighting key findings. We wanted to build an understanding of how authenticity methods are experienced. We also wanted to better understand this in the context of certain elements building towards experienced authenticity, credibility, interactions, and other user experience. It was not our goal to present a definitive conceptual framework for subdividing perceptions or a strongly supported model for how participants formed them, and should also not be used as such.

6.3.2 Sample

The sample of our study was small. It consisted largely of participants we already knew. A choice we made in purposive sampling was to mainly do so based on factors obtained from credibility research rather than creating a representative Twitter audience. As a result, it is likely our sample was less active on Twitter than the average Twitter user. Although all of our participants had used Twitter before, most barely used the platform anymore. For some participants, it was therefore also difficult to provide answers to interview questions reliant on platform knowledge. Especially understanding Verification proved quite foreign to those less active on social media. We chose our sampling criteria because we were more so interested in understanding the impact on perceived credibility authenticity methods had. Therefore, we wanted heterogeneous sampling coverage over factors relating to credibility. We have come to believe that it would be better in the future with similar studies to more so focus more the active user base. We expect their experience to prove more relevant to the roll-out of a method, and they should also better understand what is presented to them in such a study. One could then later always extend their work to a more generic audience as we did.

6.3.3 Size of the data

The total research data encompasses 200 pages that had to be partially manually compiled. Errors can and probably did slip in. In the same vein, connections had to be made between 700 quotes by fourteen different participants. This size allowed us to explore a broad understanding of encompassed subjects. It also allows more room for error and more complexity in building connections. Therefore, conclusions are likely less valid than studies to examine specific elements covered by our study. Nevertheless, some such elements would not come to light without these types of studies. The size of the data and the complexity of generating meaningful findings let us abandon several additional elements we would have liked to highlight. This includes describing how tweet truth value influences perceptions, exploring how different authenticity methods influence perceptions of the same tweets and delving into participant

influencing factors as part of their overall perceptions. We decided to abandon these factors for the weakness of the conclusions we would be able to draw from them and because of further investigation we deemed variable inclusion inappropriate for our inductive approach.

6.3.4 Constructed environment

The legal and ethical status of taking someone’s words out of context, editing them with labelling, and presenting them with labelling or Verification was unclear to us (Metzger, 2019). We saw it as unethical to include authors without their consent (Boeije, 2009), but consent would be difficult to realize and skew results towards authors and messages that would want to partake. It was also impractical to get representative topically recent tweets. Therefore, we decided to opt for a study environment of a constructed nature. This includes everything about the tweets. Consequently, none of the tweets can be technically deemed authentic as their authors do not exist, and they were never truly tweets, to begin with. A few users did ask about this during the interview. Users did so in an inquisitive manner and not as to question the realism of the environment or study. We instructed them to assume the reality of the tweets for the duration of the study and they did so without further issues.

Because the environment is constructed, familiarity with authors was not possible (5.1.5). The theme does not encompass direct perceptions of authors they were actually familiar with, only participants discussing it as important. None of the tweets included sourcing (5.2.1.2.5), and the theme should be viewed in a similar context as familiarity. Users would normally receive tweets in a personalized feed (2.1), which presents differently from our pre-selected constructed tweets. Furthermore, despite our criticism of overreliance on objective data in past work, all our Twid attributes are focused on domain expertise rather than experience-based attributes. For example, proving you are an inhabitant of Nijmegen and sharing how a local building project personally affected you is a way of using attributes we simply did not explore. Real tweets may also be distinct in ways that were not apparent to either the researchers or the participants but still affect them.

6.3.5 Unfamiliar method

Twid’s relative unfamiliarity to participants means results regarding it are heavily influenced by the information presented in the study rather than outside knowledge. This is relevant since we provided participants with a functional explanation of both Twid and Verification. However, with Twid this explanation was all they could base their understanding on, whereas with Verification most already had some level of understanding and opinion on it. Literature also suggests novelty increases interactivity and potentially increases credibility (3.5), which could therefore apply to our inclusion of Twid in this study. Prolonged and repeated exposure to information also impacts credibility (4). This effect measured for information may also bear some relevance to an authenticity method like Twid as extended use would in essence also increase media literacy specific to it (1). Although these limitations presented by short exposure to an unfamiliar concept somewhat constrain our findings, we deem them unavoidable. As we described earlier, an extended time span would have not been feasible nor appropriate for us (4.5).

6.3.6 Interaction data

The setup of our study proved sub-par for getting insight into perceptions relating to interactions. We find this somewhat disappointing as we considered this an important factor in determining the success of authenticity methods as we laid out in our related work section (3.5). Given that our other ambitions of broad understanding made it difficult to cater specifically to getting perceptions on interactions, this outcome was perhaps unavoidable. Results on interactions were very sparse, and most results were perceptions of intentions not to interact. In hindsight, we do not find the lack of results very surprising. On social media total number of engagements only encompasses a small number compared to how many posts are viewed (Parsons, 2022) and our posts are potentially less appealing than a personalized feed. Total intent to interact was mentioned twenty-four times. If we treat that as actual interactions, it would still net a higher than average interaction rate (Parsons, 2022). This outcome is still too feeble for strong, detailed, descriptive results. This is why we, for the most part, stayed away from mentioning

it as part of our key findings.

6.3.7 Tools

Transcriptions produced by Azure were of mixed quality. We do not know if the fault in these lies with the quality of our audio recordings, using Azure for the Dutch language or with some participants having slight accents. Using Azure for our purposes may have cost more time than it saved, and we would not recommend doing so again given the same circumstances.

6.4 Future work

We believe there are several avenues of future research that could prove interesting. Some build on our work or are very related to it. Others are mentioned because they came up as interesting and necessary through the process of producing this research. For those wishing to embark on research applying similar methods as us, we highly recommend using a combination of the methodology and preceding subsection to assist in making educated choices and avoid potential mistakes.

6.4.1 User and problem centered research and design strategies

We found that the development of authenticity methods and other misinformation intervention techniques seems to follow a top-down technocratic approach with a preference for using existing innovation rather than catering innovation to a specific purpose and social environment. Let us sketch our point out using the information available on Twid. We would like to say that Twid arguably does better on this point of criticism than most, and would further note that Twid has not rolled out yet. We also can only use Twid as an example as we are allowed more insight into their process than other measures.

The intent of Twid is to serve as a mitigation to disinformation. The initial suggestion for and continued development of Twid’s method does not initiate from a study on the literature of how to deal with this issue. Furthermore, it initially also did not survey a Twitter user base or other stakeholders on their thoughts on innovations for this issue. More so it seems that in an undoubtably positive attempt to “fight disinformation”, it was chosen to rehash the existing in-house technological identity management innovation, IRMA (Alpár et al., 2017). This choice seems more based on availability than on strong knowledge of what might present itself as a suitable solution. To be clear, we make no argument that Twid is not a positive attempt to contribute to the fight against misinformation. Nor is there anything wrong with a well-informed choice for the best available technique to solve a problem at hand. More so that we found innovation applied to a new context without understanding the context of a sub-ideal approach. Verification’s developmental secrecy makes it much more difficult to formulate similar critiques on it. However, its contentious effect in literature (3.1) reflects poorly on the possibility of Twitter properly researching its effects. Similarly, Twitter announced its plans to adopt a subscription service only weeks after a leadership change, all but guarantees that current development practices have not improved on this.

We suggest an alternative approach. This can be applied to the research on the development of new authenticity methods, the continued development of Twid, or even similar mitigation strategies. We believe these should consider a user-centred design strategy (Chammas, Quaresma, & Mont’Alvão, 2015) and involvement of stakeholders (Owen & Pansera, 2019). For our purposes, this could mean starting off with an extended literature study on social media, user psychology and misinformation mechanisms. Then one could extend this with user and stakeholder sessions to better understand their needs and behaviour and ensure innovation is usable and understandable. Although these ideas are not entirely new, we still think it is good to highlight them as it appears this is not always the approach taken.

6.4.2 Vulnerability

A popular research subject in digital security is vulnerabilities and attack strategies. If despite our earlier critiques, Twid in its current form is deployed in any shape, research on the truth of its abuse sensitivity

should be done. An approach we would find interesting is figuring out how to maximize credibility while minimizing the effort in obtaining attributes and then testing the resulting effect on the credibility of an author compared to no attribute. A similar approach could be of interest to other authenticity methods and misinformation mitigation strategies. Another approach that could further test some vulnerabilities of Twid, is more specific to IRMA. You could test what types of legal shell organizations you can create that can issue misleading attributes, and see how you can influence users with that.

6.4.3 Theoretical dimension of analysis

Research on authenticity methods, this research included, has emphasized viewing user perceptions through credibility (3.1, 3.2). This works well for research centred on subjective truthfulness. However, the current state of knowledge is still lacking medium-specific proof for some underlying assumptions on credibility (3.4). Furthermore, credibility research is market research and emphasizes the human perception of truth over actual truth (3.4), which leaves it open to using fallacies and cognitive biases as valid strategies if we limit ourselves to credibility as a measure of success alone. Therefore, continued development of credibility research in social media is of interest to broaden understanding of how users might perceive and try to persuade others. We see credibility as a necessary lens to delve into human perception, which we deem essential in understanding these and similar purposes. However, it might also be worthwhile to choose additional means of analysis if the goal is the improvement of social media. An example would be to look at Twitter and authenticity methods from a quality of information system perspective (Fisher, Lauria, & Chengalur-Smith, 2012). This is a discipline to rank the strength of knowledge systems. With this, one could envision a study analyzing social media systems for the inherent quality of information they provide. Combined with understanding credibility perceptions, this could be used to see what elements cause friction or could be improved with minimal effort for maximum success.

6.4.4 Hypotheses for testing

The authenticity of opinions (6.2.5), the credibility of notable users (6.2.6), and social proof on social media (6.2.6) are all things that deserve further research of their own. Quantitative hypothesis testing surrounding this would be a clear point of further research. We personally especially feel that further research on the effect of opinions and subjectivity is worthwhile. As we mentioned earlier, this finding is relatively novel but could have big implications (6.2.5).

6.4.5 Experience centric messages

Fact-checking is perhaps the most well-known misinformation mitigation strategy, but its effect is entirely limited to objective content. Past works, our own included, similarly focused both research and development efforts on covering objective content. Not all messages on Twitter adhere to this. We believe that exploring the effects of authenticity methods and misinformation mitigation strategies on content more so based on experience and mental impact rather than objective content would provide novel and relevant information (5.4.2).

6.4.6 Qualitative interaction data

We would be interested in research focusing solely on getting qualitative data on interactions. Information on why and how people interact and how this is influenced by authenticity methods is still beyond us but could prove valuable. If you wanted to do this, we recommend taking careful note of the critical reflections (6.3) on why this research produced lacklustre results on interactions.

6.4.7 Qualitative authors and medium data

We were taught in physics that sound requires a source, a medium, and a receiver to be heard and therefore considered a sound at all. Similarly, information has a source, a medium, and a recipient. This study focussed efforts on the recipients of the information. Research that analyzes the source, authors, and the medium, Twitter, using similar qualitative methods could provide valuable input. For medium, the state would also be of interest. Authors need to be analyzed because a few authors provide a large amount of content (Wojcik, Stefan & Hughes, Adam, 2019), and understanding how they use innovations

is a key element in determining their success. Similarly, Twitter and the government's willingness to support innovation is an essential factor in the resulting impact. Studies from this perspective for Twid and Verification are very valid avenues, although these avenues hold equally true for other misinformation mitigation strategies as well as other social media innovations.

7 Conclusions

We set out to increase understanding of the user experience of authenticity methods. We led with an examination of the literature on authenticity, credibility, interactions, and authenticity methods (3). This was used to design a cross-sectional within-subject design case study and semi-structured interview guide geared towards documenting user perceptions of different authenticity methods (4). We documented several ways we felt our design had shortcomings over scope, sampling, and input on user interactions (6.3). Our work combined with this criticism can be used by future researchers investigating similar topics or looking to utilize similar methods.

We performed inductive qualitative thematic analysis (4) on the primary data we obtained from our case study. This resulted in a thematic framework (5) we used to highlight overarching findings (6.2). Participant perceptions contained in our study were mostly in line with our expectations based on previous works (6.1). Both tested authenticity methods were generally perceived by participants to have a slightly positive impact on message credibility and authenticity, although other factors probably contributed more (6.2.1). Authenticity methods were often perceived to be unintuitive and pose the risk of being functionally difficult to understand for users (6.2.2). Methods allowed participants to scrutinize messages less. This was perceived as positive because participants prefer minimizing effort but were also perceived as risky since methods provide no quality guarantees over messages (6.2.7). Verification was generally perceived as being straightforward and positively impacting credibility (6.2.3). However, its restricted user base and limited information gain were perceived less well. Some users also attributed notability and other social factors to a gain in credibility not intended nor supported by functionality (6.2.6). Twid was generally perceived positively. Among other reasons, for providing relevant domain expertise over objective content (6.2.4). Our study further suggests Twid risks failing its original intentions of availability, scalability, and positive impact by shifting attention towards attributable aspects of people (2.9, 3.2). This is because our findings suggested that Twid is not set up to be easily available or accessible, Twid requires high adoption to be helpful, and attributes may be difficult to interpret (6.2.4). Twid’s method was also seen by some to be both misleading and open to abuse (6.2.4). The impact of all methods may be undermined by the perceived authenticity of opinionated content (6.2.5). This finding would mean opinionated messages potentially bypass Verification’s main intent of disallowing impersonation (3.1).

The results of our analysis are bounded by our choice of methods (4) and limitations caused by our application (6.3). Our findings should be viewed as induced hypotheses, not proven conclusions. Our application of analysis methods also carries some imperfections (6.3.1), among other reasons, because of researcher inexperience and deviations in applied methods.

We contribute a position on the impact of authenticity methods with our key finds that suggest users have a slight positive disposition towards them and authors using them (6.2.1), whilst noting an underwhelming influence at best and potentially enabling the spread of misinformation at worst (6.2). Both methods tested are also likely unproductive in achieving their intended usage advantages (6.2.3, 6.2.4). We also highlighted elements of social proof present in people’s credibility judgement (6.2.6), found evidence for opinionated messages seeming more authentic (6.2.5), and examined perceptions on usability and the unwanted effort of scrutinizing information online (6.2.7).

We believe interesting avenues of future work could be social vulnerability testing of misinformation intervention methods (6.4.2), medium-specific credibility research and diversified analysis dimensions (6.4.3), hypothesis testing the authenticity of opinions (6.4.4), hypotheses testing the credibility of notability and social proof (6.4.4), method interaction with experiential messaging (6.4.5), qualitative interaction data (6.4.6), and gaining insight into author medium and government perspective on authenticity methods (6.4.7). We also think that researchers and developers should consider involving

stakeholders more and earlier, and develop or choose technology towards a goal instead of reworking existing available technology to fit a goal (6.4.1). This holds for Twid, other authenticity methods and even generally for misinformation mitigation strategies.

We would like to close off with a slightly speculative note. Humans are inclined to engage with misinformation (Vosoughi et al., 2018) and findings suggest a preference for not expending energy to change this (6.2.7). Social media companies are financially incentivized to facilitate cumulative platform usage as it relates to their main income stream, advertisements. For example, Twitter made 86.3% of their 2020 revenue this way (Twitter, 2020). If Twitter were to abandon this strategy, another company could simply use the same principle and the user base would follow suit. Therefore, mitigating digital misinformation through technological developments might be like filling a bucket filled with holes, unless some disruptive technology, significant social changes, educational changes, or regulatory changes address the underlying social mechanisms that allow for its success.

References

- Alpár, G., Jacobs, B., Lueks, W., & Ringers, S. (2017). IRMA: practical, decentralized and privacy-friendly identity management using smartphones. , 2.
- ATLAS.ti | *The Qualitative Data Analysis & Research Software*. (2022). Retrieved 2022-09-08, from <https://atlasti.com>
- Barnard, Y. F., Someren, M. W. V., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method*.
- Baxter, P., & Jack, S. (2015, January). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*. Retrieved 2022-02-16, from <https://nsuworks.nova.edu/tqr/vol13/iss4/2/> doi: 10.46743/2160-3715/2008.1573
- Bennett, B. (2012). *Logically Fallacious: The Ultimate Collection of Over 300 Logical Fallacies (Academic Edition)*. eBookIt.com. (Google-Books-ID: WFvhN9lSm5gC)
- Bhardwaj, P. (2018, May). *These 30 Facebook ads were shared by Russian trolls just days before the 2016 election. Some were so subtle, you probably didn't realize they were ads*. Retrieved 2023-01-29, from <https://www.businessinsider.nl/facebook-ads-russian-trolls-before-election-photos-2018-5/>
- Boeije, H. R. (2009). *Analysis in Qualitative Research*. SAGE. (Google-Books-ID: 9EFdBAAAQBAJ)
- Braun, V., & Clarke, V. (2006, January). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101. doi: 10.1191/1478088706qp063oa
- Braun, V., & Clarke, V. (2019). *Answers to frequently asked questions about thematic analysis April 2019.pdf*. Retrieved 2023-03-18, from <https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/Answers%20to%20frequently%20asked%20questions%20about%20thematic%20analysis%20April%202019.pdf>
- Carr, C. T., & Hayes, R. A. (2015, January). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 23(1), 46–65. Retrieved 2022-09-27, from <http://www.tandfonline.com/doi/abs/10.1080/15456870.2015.972282> doi: 10.1080/15456870.2015.972282
- Centrum, N. C. S. (2022, February). *Digitale aanvallen oorlog Oekraïne - Nieuwsbericht - Nationaal Cyber Security Centrum* [nieuwsbericht]. Retrieved 2023-03-17, from <https://www.ncsc.nl/actueel/nieuws/2022/februari/website-pagina-oekraïne/website-pagina-oekraïne> (Last Modified: 2022-02-27T15:41 Publisher: Nationaal Cyber Security Centrum)
- Chadwick, A., & Vaccari, C. (2019). News Sharing on UK Social Media: Misinformation, Disinformation, and Correction. , 32.
- Chammas, A., Quaresma, M., & Mont'Alvão, C. (2015, January). A Closer Look on the User Centred Design. *Procedia Manufacturing*, 3, 5397–5404. Retrieved 2023-02-05, from <https://www.sciencedirect.com/science/article/pii/S2351978915006575> doi: 10.1016/j.promfg.2015.07.656
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012, January). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. Retrieved 2022-02-21, from <https://www.sciencedirect.com/science/article/pii/S0167268111002289> doi: 10.1016/j.jebo.2011.08.009
- Choi, W., & Stvilia, B. (2015). Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology*, 66(12), 2399–2414. Retrieved 2021-11-16, from <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23543> (.eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23543>) doi: 10.1002/asi.23543
- Cialdini, R. B. (2007). *Influence: the psychology of persuasion* (Rev. ed. ; 1st Collins business essentials ed ed.). New York: Collins.
- DOOB, L. W. (1950, January). Goebbels' Principles of Propaganda. *Public Opinion Quarterly*, 14(3),

- 419–442. Retrieved 2022-10-06, from <https://doi.org/10.1086/266211> doi: 10.1086/266211
- Dora-Olivia Vicol. (2020, January). *who-believes-shares-misinformation.pdf*. Retrieved 2021-11-04, from <https://fullfact.org/media/uploads/who-believes-shares-misinformation.pdf>
- Edgerly, S., & Vraga, E. K. (2019, April). The Blue Check of Credibility: Does Account Verification Matter When Evaluating News on Twitter? *Cyberpsychology, Behavior, and Social Networking*, 22(4), 283–287. Retrieved 2022-10-27, from <https://www.liebertpub.com/doi/full/10.1089/cyber.2018.0475> (Publisher: Mary Ann Liebert, Inc., publishers) doi: 10.1089/cyber.2018.0475
- eric urban. (2022). *Speech-to-text documentation - Tutorials, API Reference - Azure Cognitive Services - Azure Cognitive Services*. Retrieved 2022-09-08, from <https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/index-speech-to-text>
- European Commission. (2022, June). *Tackling online disinformation | Shaping Europe's digital future*. Retrieved 2022-09-27, from <https://digital-strategy.ec.europa.eu/en/policies/online-disinformation>
- European Union. (2022, September). *Standard Eurobarometer 97 - Summer 2022 - Country Factsheets in English Netherlands - en*. Retrieved from <https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=83460>
- Facebook. (2023, February). *Request a verified badge on Facebook | Facebook Help Center*. Retrieved 2023-02-02, from <https://www.facebook.com/help/1288173394636262>
- Felt, A. P., Reeder, R. W., Ainslie, A., Harris, H., Walker, M., Thompson, C., ... Consolvo, S. (2016). Rethinking Connection Security Indicators. In (pp. 1–14). Retrieved 2022-10-27, from <https://www.usenix.org/conference/soups2016/technical-sessions/presentation/porter-felt>
- Fisher, C., Lauria, E., & Chengalur-Smith, S. (2012). *Introduction to Information Quality*. AuthorHouse. (Google-Books-ID: UP8EyqCywEkC)
- Fogg, B. J., Soohoo, C., Danielson, D. R., & Marable, L. (2003). How Do Users Evaluate the Credibility of Web Sites? A Study with Over 2,500 Participants. , 15.
- Füegi, J., & Francis, J. (2015, August). Lovelace & Babbage and the creation of the 1843 'notes'. *ACM Inroads*, 6(3), 78–86. Retrieved 2023-03-17, from <https://dl.acm.org/doi/10.1145/2810201> doi: 10.1145/2810201
- Gerring, J. (2004, May). What Is a Case Study and What Is It Good for? *American Political Science Review*, 98(2), 341–354. Retrieved 2022-10-03, from https://www.cambridge.org/core/product/identifier/S0003055404001182/type/journal_article doi: 10.1017/S0003055404001182
- Goldkuhl, G. (2012, March). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21(2), 135–146. Retrieved 2022-12-16, from <https://doi.org/10.1057/ejis.2011.54> doi: 10.1057/ejis.2011.54
- Graves, L. (2018). Understanding the Promise and Limits of Automated Fact-Checking. , 8.
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019, January). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378. Retrieved 2022-09-27, from <https://www.science.org/doi/10.1126/science.aau2706> doi: 10.1126/science.aau2706
- Harmon-Jones, E., & Mills, J. (n.d.). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. , 3. Retrieved 2023-02-02, from <https://psycnet.apa.org/fulltext/2019-11198-001.pdf> (Publisher: Washington, DC, US: American Psychological Association) doi: 10.1037/0000135-001
- Harris, T. (2014). *Grounded Theory*. Retrieved 2022-12-22, from <https://dora.dmu.ac.uk/bitstream/handle/2086/13045/Grounded%20Theory%20by%20Tina%20Harris%20upload%20for%20Dora.pdf?sequence=3>
- Hovland, C., Janis, I., & Kelley, H. (1953). *Communication and persuasion*. New Haven, CT, US: Yale University Press.
- Hughes, M. G., Griffith, J. A., Zeni, T. A., Arsenault, M. L., Cooper, O. D., Johnson, G., ... Mumford,

- M. D. (2014, April). Discrediting in a Message Board Forum: The Effects of Social Support and Attacks on Expertise and Trustworthiness*. *Journal of Computer-Mediated Communication*, 19(3), 325–341. Retrieved 2021-10-23, from <https://doi.org/10.1111/jcc4.12077> doi: 10.1111/jcc4.12077
- Instagram. (2023, February). *Verified Badges | Instagram Help Center*. Retrieved 2023-02-02, from <https://help.instagram.com/854227311295302>
- Irma. (2023a, February). *About - IRMA credentials*. Retrieved 2023-02-02, from <https://privacybydesign.foundation/attribute-index/en/>
- Irma. (2023b, February). *What is IRMA? · IRMA docs* [What is IRMA]. Retrieved 2023-02-02, from <https://irma.app/docs/>
- jack [@Jack]. (2006, March). *just setting up my twttr* [Tweet]. Retrieved 2022-04-17, from <https://twitter.com/Jack/status/20>
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis - WebKDD/SNA-KDD '07* (pp. 56–65). San Jose, California: ACM Press. Retrieved 2021-12-05, from <http://portal.acm.org/citation.cfm?doid=1348549.1348556> doi: 10.1145/1348549.1348556
- Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72(12), 2954–2965. Retrieved 2021-10-20, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/jan.13031> (eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jan.13031>) doi: 10.1111/jan.13031
- Katz, J. (2001, December). Analytic Induction. In *International Encyclopedia of the Social and Behavioral Sciences* (Vol. 1, pp. 480–484). (Journal Abbreviation: International Encyclopedia of the Social and Behavioral Sciences) doi: 10.1016/B0-08-043076-7/00774-9
- Koninkrijksrelaties, M. v. B. Z. e. (n.d.). *Drinkwaterbesluit* [AMvB]. Retrieved 2023-03-17, from <https://wetten.overheid.nl/BWBR0030111/2015-11-28> (Last Modified: 2023-03-09)
- Korstjens, I., & Moser, A. (2018, January). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120–124. Retrieved 2023-02-01, from <https://www.tandfonline.com/doi/full/10.1080/13814788.2017.1375092> doi: 10.1080/13814788.2017.1375092
- Kouzy, R., Jaoude, J. A., Kraitem, A., Alam, M. B. E., Karam, B., Adib, E., ... Baddour, K. (2020, March). Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*, 12(3). Retrieved 2022-09-27, from <https://www.cureus.com/articles/28976-coronavirus-goes-viral-quantifying-the-covid-19-misinformation-epidemic-on-twitter> (Publisher: Cureus) doi: 10.7759/cureus.7255
- L. Clemmer. (2009, April). *Information Security Concepts: Authenticity*. Retrieved 2021-11-01, from <https://www.brighthub.com/computing/smb-security/articles/31234/>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... Zittrain, J. L. (2018, March). The science of fake news. *Science*, 359(6380), 1094–1096. Retrieved 2021-11-08, from <https://www.science.org/doi/10.1126/science.aao2998> (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.aao2998
- Levin, K. A. (2006, March). Study design III: Cross-sectional studies. *Evidence-Based Dentistry*, 7(1), 24–25. Retrieved 2023-02-02, from <https://www.nature.com/articles/6400375> (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/sj.ebd.6400375
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012, December). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. Retrieved 2021-10-18, from <https://doi.org/10.1177/1529100612451018> (Publisher: SAGE Publications Inc) doi: 10.1177/1529100612451018

- Lowe, C., Zemliansky, P., Driscoll, D., Stewart, M., & Vetter, M. (Eds.). (2010). *Writing spaces: readings on writing*. West Lafayette, Indiana ; Anderson, South Carolina: Parlor Press.
- Luhmann, N. (1985). *A Sociological Theory of Law*. Routledge & Kegan Paul.
- McCallum, K. (2021, April). *UTI Home Remedies: Does Cranberry Juice Really Help?* Retrieved 2023-03-17, from <https://www.houstonmethodist.org/blog/articles/2021/nov/uti-home-remedies-does-cranberry-juice-really-help/>
- Merriam-Webster Dictionary. (2022, November). *Definition of CREDIBILITY*. Retrieved 2022-01-11, from <https://www.merriam-webster.com/dictionary/credibility>
- Metzger, D. D. J. P. M. C. S., Thomas E. Kadri. (2019). *The Legal, Ethical, and Efficacy Dimensions of Managing Synthetic and Manipulated Media*. Retrieved 2023-02-01, from <https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439>
- Micheal Madden. (2015, August). *North Korea's New Propagandist?* Retrieved 2022-10-06, from <https://www.38north.org/2015/08/mmadden081415/>
- Ministerie van algemene zaken. (2023, January). *Onderwerpen - Desinformatie en nepnieuws - Rijksoverheid.nl* [Documenten desinformatie en nepnieuws]. Retrieved 2023-02-15, from <https://www.rijksoverheid.nl/onderwerpen/desinformatie-nepnieuws/documenten> (Last Modified: 2023-02-15T14:33 Publisher: Ministerie van Algemene Zaken)
- Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012). Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12* (p. 441). Seattle, Washington, USA: ACM Press. Retrieved 2021-10-18, from <http://dl.acm.org/citation.cfm?doid=2145204.2145274> doi: 10.1145/2145204.2145274
- National Institutes of Health (NIH) Office of Dietary Supplements (ODS). (2023, March). *Office of Dietary Supplements - Omega-3 Fatty Acids*. Retrieved 2023-03-17, from <https://ods.od.nih.gov/factsheets/Omega3FattyAcids-HealthProfessional/>
- Nemr, C., & Gangware, W. (2019). *Weapons-of-Mass-Distractioin-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf*. Retrieved 2023-01-29, from <https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distractioin-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf>
- NOS. (2022, May). *Nog meer kinderen toeslagenaffaire uit huis geplaatst*. Retrieved 2023-03-17, from <https://nos.nl/artikel/2428355-nog-meer-kinderen-toeslagenaffaire-uit-huis-geplaatst>
- OpenAI. (2023, March). *OpenAI*. Retrieved 2023-03-17, from <https://openai.com/>
- Owen, R., & Pansera, M. (2019, June). Responsible Innovation and Responsible Research and Innovation. *Handbook on Science and Public Policy*, 26–48. Retrieved 2023-02-05, from <https://www.elgaronline.com/display/edcoll/9781784715939/9781784715939.00010.xml> (ISBN: 9781784715946 Publisher: Edward Elgar Publishing Section: Handbook on Science and Public Policy)
- Oxford learners dictionary. (2021, July). *credibility noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com*. Retrieved 2021-12-07, from <https://www.oxfordlearnersdictionaries.com/definition/english/credibility?q=credibility>
- Parsons, J. (2022, May). *What's Considered a Good Engagement Rate For Tweets?* Retrieved 2023-01-19, from <https://follows.com/blog/2022/05/good-engagement-rate-tweets>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021, April). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595. Retrieved 2022-01-24, from <https://www.nature.com/articles/s41586-021-03344-2> (Bandiera_abtest: a Cg.type: Nature Research Journals Number: 7855 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Communication;Decision making;Human behaviour;Technology)

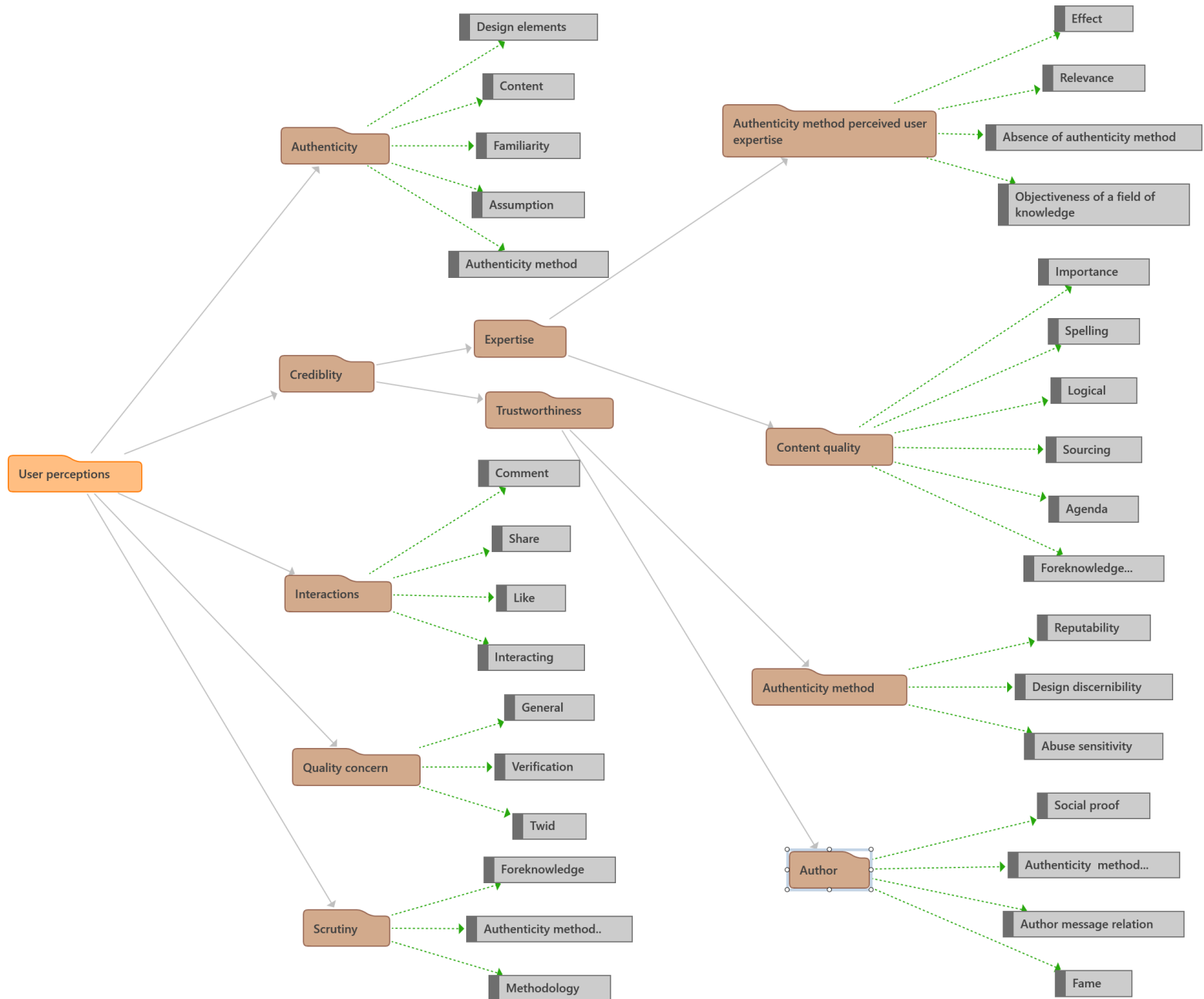
- Subject_term.id: communication;decision-making;human-behaviour;technology) doi: 10.1038/s41586-021-03344-2
- Pexels. (2023). *Gratis stockfoto's*. Retrieved 2023-03-18, from <https://www.pexels.com/nl-nl/>
- Pinterest. (2023, February). *Verified accounts*. Retrieved 2023-02-02, from <https://help.pinterest.com/en/business/article/verified-accounts>
- Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, 34(2), 243–281. Retrieved 2021-10-23, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.2004.tb02547.x> (_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.2004.tb02547.x>) doi: 10.1111/j.1559-1816.2004.tb02547.x
- Privacy by Design Foundation. (2023a). *IRMA explanation*. Retrieved 2023-02-16, from <https://privacybydesign.foundation/irma-explanation/>
- Privacy by Design Foundation. (2023b, August). *Privacy by Design Foundation*. Retrieved 2023-02-16, from <https://privacybydesign.foundation/en/>
- Radboud University. (2022). *Qualitative research course information*. Retrieved 2023-02-16, from <https://ru.osiris-student.nl/#/onderwijs/catalogus/extern/cursus?collegejaar=huidig&taal=en&cursuscode=NWI-I00152>
- Rapp, C. (2022). Aristotle's Rhetoric. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 ed.). Metaphysics Research Lab, Stanford University. Retrieved 2023-02-02, from <https://plato.stanford.edu/archives/spr2022/entries/aristotle-rhetoric/>
- Rosenberg, H., Syed, S., & Rezaie, S. (2020, July). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine*, 22(4), 418–421. Retrieved 2022-09-27, from <https://www.cambridge.org/core/journals/canadian-journal-of-emergency-medicine/article/twitter-pandemic-the-critical-role-of-twitter-in-the-dissemination-of-medical-information-and-misinformation-during-the-covid19-pandemic/9F42C2D99CA00FBAE50A66D107322211> (Publisher: Cambridge University Press) doi: 10.1017/cem.2020.361
- Ryan K. Balot. (2009). *A Companion to Greek and Roman Political Thought* (1st ed.). John Wiley & Sons, Ltd. Retrieved 2022-10-06, from <https://onlinelibrary.wiley.com/doi/10.1002/9781444310344#page=343> (_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781444310344>) doi: 10.1002/9781444310344
- Samonas, S., & Coss, D. (2014). THE CIA STRIKES BACK: REDEFINING CONFIDENTIALITY, INTEGRITY AND AVAILABILITY IN SECURITY. , 25.
- Shaw, C. (2013, October). *Dress For Success: The White Lab Coat Effect and the Subconscious Experience | CX Consulting*. Retrieved 2022-12-04, from <https://beyondphilosophy.com/dress-success-white-lab-coat-effect-subconscious-experience/> (Section: Subconscious Experience)
- Simon, M.-S. (2022, June). *Judging a tweets credibility - The effect of signature labels on perceived Tweet credibility* (Master thesis). Nijmegen, Netherlands: Radboud University. Retrieved 2022-10-24, from https://www.cs.ru.nl/bachelors-theses/2022/Marie-Sophie.Simon___1023848___Judging_a_Tweets_credibility_-_The_effect_of_signature_labels_on_perceived_Tweet_credibility.pdf
- Snapchat. (2023, February). *Verify Your Public Profile*. Retrieved 2023-02-02, from https://businesshelp.snapchat.com/s/article/public-profile-verify?language=en_US
- Stieglitz, S., & Dang-Xuan, L. (2013, April). Emotions and Information Diffusion in Social Media — Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29, 217–248. doi: 10.2753/MIS0742-1222290408
- Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital Media*, 29.

- Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019, November). Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services*, 51, 72–82. Retrieved 2022-01-13, from <https://www.sciencedirect.com/science/article/pii/S0969698919301407> doi: 10.1016/j.jretconser.2019.05.026
- Telegram. (2020, July). *How To Get Verified Telegram Account : Telegram Blue Tick Verification Badge*. Retrieved 2023-02-02, from <https://www.techscanner.in/how-to-get-verified-telegram-account/> (Section: How to)
- Thompson, N., Wang, X., & Daya, P. (2020, November). Determinants of News Sharing Behavior on Social Media. *Journal of Computer Information Systems*, 60(6), 593–601. Retrieved 2022-01-13, from <https://www.tandfonline.com/doi/full/10.1080/08874417.2019.1566803> doi: 10.1080/08874417.2019.1566803
- TikTok. (2023, February). *Verified accounts on TikTok | TikTok Help Center*. Retrieved 2023-02-02, from <https://support.tiktok.com/en/using-tiktok/growing-your-audience/how-to-tell-if-an-account-is-verified-on-tiktok>
- Tsikerdekis, M., & Zeadally, S. (2014, September). Online deception in social media. *Communications of the ACM*, 57(9), 72–80. Retrieved 2022-01-15, from <https://dl.acm.org/doi/10.1145/2629612> doi: 10.1145/2629612
- Twitter. (2020). *Twitter 2020 selected metrics and financial* (Tech. Rep.). Retrieved from https://s22.q4cdn.com/826641620/files/doc_financials/2021/q1/Q1'21-Selected-Metrics-and-Financials.pdf
- Twitter. (2022a). *About Twitter | Our company and priorities*. Retrieved 2022-02-06, from <https://about.twitter.com/en>
- Twitter. (2022b, September). *Twitter verification requirements - how to get the blue check*. Retrieved 2021-11-08, from <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>
- Twitter. (2023, February). *Twitter account activity analytics – engagement, impressions and more*. Retrieved 2023-02-02, from <https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard>
- United Nations Human Rights Council. (2022, April). *UN's rights council adopts 'fake news' resolution, States urged to tackle hate speech*. Retrieved 2022-09-27, from <https://news.un.org/en/story/2022/04/1115412>
- Vaidya, T., Votipka, D., Mazurek, M. L., & Sherr, M. (2019, May). Does Being Verified Make You More Credible?: Account Verification's Effect on Tweet Credibility. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Glasgow Scotland Uk: ACM. Retrieved 2021-10-15, from <https://dl.acm.org/doi/10.1145/3290605.3300755> doi: 10.1145/3290605.3300755
- van Gastel, Bernard, Schraffenberg, Hanna, Bor, Dennis, & Vervoort, Lian. (2021). *Twid: fighting disinformation on Twitter with authenticity - iHub*. Retrieved 2021-11-03, from <https://ihub.ru.nl/project/twid.page>
- van Gastel, B., Jacobs, B., Schraffenberger, H., Grassl, P., Botros, L., & Kleemans, M. (2021, April). *Twid: Fighting Fake News on Twitter*.
- Varga, S., & Guignon, C. (2014, September). Authenticity. Retrieved 2021-11-01, from <https://plato.stanford.edu/entries/authenticity/> (Last Modified: 2020-02-20)
- Verger, M. B. (2021). The effects of celebrities, macro-influencers and micro- influencers product endorsement on advertising effectiveness and credibility.
- Vlachos, A., & Riedel, S. (2014, June). Fact Checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science* (pp. 18–22). Baltimore, MD, USA: Association for Computational Linguistics. Retrieved 2021-11-08, from <https://aclanthology.org/W14-2508> doi: 10.3115/v1/W14-2508

- Vosoughi, S., Roy, D., & Aral, S. (2018, March). The spread of true and false news online. *Science*, 359(6380), 1146–1151. Retrieved 2021-10-13, from <https://www.science.org/doi/10.1126/science.aap9559> doi: 10.1126/science.aap9559
- WhatsApp. (2023, February). *Verified business account | WhatsApp Help Center*. Retrieved 2023-02-02, from <https://faq.whatsapp.com/794517045178057>
- Wojcik, Stefan, & Hughes, Adam. (2019, April). *How Twitter Users Compare to the General Public*. Retrieved 2022-01-31, from <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>
- Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019, November). Misinformation in Social Media: Definition, Manipulation, and Detection. *ACM SIGKDD Explorations Newsletter*, 21(2), 80–90. Retrieved 2021-11-08, from <https://dl.acm.org/doi/10.1145/3373464.3373475> doi: 10.1145/3373464.3373475
- Zulman, D. M., Kirch, M., Zheng, K., & An, L. C. (2011, February). Trust in the Internet as a Health Resource Among Older Adults: Analysis of Data from a Nationally Representative Survey. *Journal of Medical Internet Research*, 13(1), e1552. Retrieved 2023-02-02, from <https://www.jmir.org/2011/1/e19> (Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada) doi: 10.2196/jmir.1552

A Appendix A

Figure 4: *The hierarchical grouping of codes resulting from the thematic analysis.*



B Appendix B

Table 1: User characteristics influential in credibility perception

Factor:	Explanation:	Source:
Age	Older adults typically have more concern for credibility yet are prone to perceiving higher trustworthiness in an online environment.	(Choi & Stvilia, 2015)
Gender	Studies tend to find that females rate sources as more credible, while males are slightly worse at discerning indicators providing relevant information about the credibility of a message. These results are however not universally accepted, with some studies not finding any differences across genders. We could not find data concerning people identifying as transgender or other non-traditional gender identities.	(Choi & Stvilia, 2015)
Education level	Typically, people of a higher level of education were more capable of making a reasoned credibility judgment. Specifically, some sources mention that people tend to have problems interpreting things due to a lack of understanding of scientific methods. This includes things such as the proper interpretation of probabilities and statistics and how logical arguments are formulated.	(Choi & Stvilia, 2015; Vosoughi et al., 2018)
Motivation	Users who are highly motivated to assess credibility tend to look more into content-related features, as opposed to unmotivated users who are more likely to look at superficial design features.	(Choi & Stvilia, 2015)
Ability	Some users are simply less so capable than others to make proper credibility judgments.	(Choi & Stvilia, 2015)
Preconceived notions, motivated reasoning or cognitive dissonance	Expresses to what extent information is consistent with previously held ideas of the users. In psychology, combined with similar factors this is better known confirmation bias. Confirmation bias is a specific type of cognitive dissonance. Cognitive dissonance describes types of mental conflict occurring when your perceptions do not line up with your beliefs. Consistently, researchers mention cognitive dissonance and specifically confirmation bias to be the single most influencing factor, with statements closer to previously held ideas rated as more credible.	(Dora-Olivia Vicol, 2020; Harmon-Jones & Mills, n.d.; Pornpitakpan, 2004; Vosoughi et al., 2018)
Domain expertise	Relating to the level of knowledge over message content an evaluator has. Higher domain expertise typically allows better evaluation over argumentation as opposed to heuristic elements. The influence of domain expertise is disputed. Furthermore, some studies show the opposite effect of domain knowledge in the area of politics	(Choi & Stvilia, 2015; Morris et al., 2012)
Emotional response	Messages that evoke high-intensity emotions, specifically of anger, amusement, anxiety, or disgust, are far more likely to be deemed credible and shared with others. This holds especially true for stories that already had some credibility.	(Dora-Olivia Vicol, 2020)
Information literacy	Specifically relates to online channels. People who are more trained in finding information on the web can better identify factors communicating the credibility of information.	(Choi & Stvilia, 2015)
Media reliance	Relates to the relative influence a certain medium has over the user as a factor of how many different types of media the users frequently consumes. If a user gets most of their information from one place, like a specific news app, they are more likely to see the information presented there as credible as opposed to when they consume a range of different media. In general, more experienced internet users are also more likely to view information on the internet as credible, although they are better at differentiating and contextualizing different sources of online information.	(Choi & Stvilia, 2015)

Table 2: Message characteristics influential in credibility perception

Factor:	Explanation:	Source:
Argumentation	Argument quality, presence of supporting arguments, and lack of discrepancy in the message were a huge positive influence on the perceived credibility. This effect was even stronger if the perceived expertise of the author was also high. Properly using and refuting counterarguments to one’s own point can increase credibility.	(Fogg et al., 2003; Pornpitakpan, 2004)
Congruence with self-interest	Relating to an interaction between the content and the author of a message. Information that is self-serving to the author is generally deemed less credible	(Pornpitakpan, 2004)
Fluency	Information that looks right by right of being easy to process is easier to believe.	(Dora-Olivia Vicol, 2020)
Sourcing	With the sourcing, we refer to the presence of evidence to support the message. For example: providing a reference to a study. This is positive on overall credibility. The author as a source conveying a message and their influence on the credibility of a message will be delved into in the next paragraph.	(Pornpitakpan, 2004)
Design	How information is presented. This has to do with the structural elements of the website, how information is presented, and technical design like search function and algorithm. Good design leads to information being more convincing. The exact meaning of good design is a little vague, but known elements of influence are things like real-world feel, professionalism, the website rarely being down, etc. Typographical errors can be seen as a problem with intrinsic content quality. Good grammar and appropriate profile pictures are shown to positively impact credibility specifically on Twitter.	(Choi & Stvilia, 2015; Fogg et al., 2003; Morris et al., 2012)

Table 3: Author characteristics influential in credibility perception

Factor:	Explanation:	Source:
Corporate credibility	The credibility of a brand or corporation, high credibility positively affects the attitude users have towards a brand even if an endorser or single ad is not perceived as credible.	(Pornpitakpan, 2004)
Reputation	Someone a user follows, has heard of, or has been verified by Twitter has a positive impact on their overall perceived credibility.	(Morris et al., 2012)
Influence	Specific to Twitter, this relates to followers’ mention counts and retweets. It is positively correlated with credibility.	(Morris et al., 2012):
Operator expertise or topical expertise	Refers to the apparent expertise an author has on a subject, communicated by credentials, history of communication, or on Twitter specifically a bio description. Positively affects overall credibility perception.	(Choi & Stvilia, 2015; Morris et al., 2012)
Affiliation	Showing affiliation with a trustworthy organization or institution, such as having a degree from a respected university, positively affects perceived credibility.	(Fogg et al., 2003)

Table 4: Other credibility influence factors

Factor:	Explanation:	Source:
Repetition	We are more prone to believe the information the more we hear it. This can even hold true for things we know to be incorrect.	(Dora-Olivia Vicol, 2020)
Passage of time after exposure	Often called the sleeper effect, a low credible source becomes more persuasive the longer it has been since exposure.	(Pornpitakpan, 2004)
Medium, media modality, or technological affordance	The medium or type of medium information is communicated on impacts credibility. On Television it was found that Trustworthiness gains in emphasis over expertise for example. Similarly recent media models studied changes in credibility on websites as effects of website affordance in allowing users access to heuristic cues	(Pornpitakpan, 2004; Sundar, 2008)

Table 5: Tweet information

Topic:	Politics			
Display name:	Jan-Kees Overmeer	Truus Kort	Ton Heijdem	Maxime Hendriks
Tweet text:	Uit een recente poll blijkt een meerderheid van ondervraagden terug naar de gulden te willen. Waarom gaan we nog door met de euro?	Enkel het opstappen van Rutte zal leiden tot de verbetering die de Nederlander wil.	Ouders verwikkeld in de toeslagen affaire hebben bijna 1700 uithuisplaatsingen van hun kroost meegemaakt, leed dat geen mens gegund is.	Het zwaarder straffen van zedendelinquenten is niet een slimme oplossing om recidivisme te voorkomen.
Tweet text (translated):	In a recent poll a majority indicated they wanted to return to the guilder. Why would we continue with the euro?	Only Rutte stepping down will result in the improvement the Dutch want.	Parent involved in the childcare benefits scandal have experienced the removal of 1700 of their children from their care, suffering no person should experience.	Punishing sex offenders more severely is not a smart solution for preventing recidivism.
Username:	@Kamerlid_Overmeer	@Truus73	@Ton_NieuwigNieuws	@MHendriks
Attribute:	Getekend door medewerker Tweede Kamer	Getekend door medewerker Tweede Kamer	Getekend door medewerker Nieuwig Nieuws	Getekend door Master of Science Criminaliteit en Rechtshandhaving
Attribute (translated):	Signed by an employee of Dutch parliament	Signed by an employee of Dutch parliament	Signed by employee Newish News	Signed by Master of Science Crime and Law enforcement
Hours since posted (1,16):	12	2	3	14
Number of comments (50,100):	99	74	91	63
Number of retweets (150,300):	253	221	262	162
Number of likes (500,800):	774	660	535	586
Truth value:	False or misleading. 85% of Dutch inhabitants don't want this (European Union, 2022), even if recent poll showed other results.	Mostly an opinion, unclear what "the improvement" refers to	Mostly true. The number is accurate (NOS, 2022). The part on suffering is an opinion	Mostly an opinion. It is unclear what smart refers to. Research to support or counter any part of this claim is lacking.
Topic:	Food and Health			

Display name:	Friso Veringa	Emma Laatbloei	Daniëlle Wobstra	Yigit Demir
Tweet text:	Chia zaad is een betere bron van Omega-3 dan enige vis, en ook nog eens helemaal vegan.	Rauw water. Sinds kort ook in Nederland te koop. Het komt direct uit de natuur, zonder industriële verwerking, en is een stuk gezonder.	Als men niet meer fruit gaat eten verwacht ik binnenkort veel meer mensen met overgewicht.	De aanwezigheid van Proanthocyanidine zorgt ervoor dat cranberrysap een blaasontsteking kan genezen.
Tweet text (translated):	Chia seed is a better source of Omega-3 than any fish, and completely vegan.	Raw water. Now also available in the Netherlands. It originates directly from nature, without industrial processes, and is a lot healthier.	If we do not eat more fruit and vegetables I expect a lot more people who are overweight.	The presence of Proanthocyanidine means cranberry juice cures UTI's.
Username:	@Friso_MD	@Dokter_Emma	@VoedselcentDaniëlle	@YigitDemir
Attribute:	Getekend door geregistreerd arts	Getekend door geregistreerd arts	Getekend door medewerker Voedselcentrum	Getekend door medewerker Dieet- en voedingsadviesbureau Norma
Attribute (translated):	Signed by registered Doctor of Medicine	Signed by registered Doctor of Medicine	Signed by employee Foodcentre	Signed by employee Dietary and food advisory bureau Norma
Hours since posted (1,16):	1	11	14	7
Number of comments (50,100):	100	58	80	98
Number of retweets (150-300):	276	299	269	151
Number of likes (500-800):	714	503	705	761
Truth value:	Mostly true. Better is subjective, but it has a higher concentration without a clear disadvantage (National Institutes of Health (NIH) Office of Dietary Supplements (ODS), 2023)	Mostly false. You can only buy it in the US. It is also dangerous, and dutch filtering is not an industrial proces but filtering (Koninkrijksrelaties, n.d.).	Mostly an opinion. Regardless of its basis, she states a personal expectation.	False, but misleading. Proanthocyanidine is found to prevent not cure UTI's in some studies (McCallum, 2021).
Topic:	Computing science			
Display name:	Daan Blom	Robin Q. te Bruggen	Irene Pardoes	ir. Marloes Bakker
Tweet text:	Een vrouw haar tijd vooruit: Ada Lovelace was de eerste computer programmeur ter wereld, decennia voor de eerste computer gebouwd zou worden.	Deze week werd bekend dat een zwakte die stilzwijgend gefixt is in de populaire library Node.js er voor zorgde dat er jarenlang miljoenen gegevens gestolen konden worden.	Een quota voor vrouwen binnen de ICT zou de sector veel goed doen.	Ondanks eerdere angst heeft het Nationaal Cyber Security Centrum tot op heden geen aan de oorlog gerelateerde digitale aanvallen op Nederlandse belangen waargenomen.

Tweet text (translated):	A woman beyond her time: Ada Lovelace was the first computer programmer in the world, decades before the first computer was built.	This week it became clear that a vulnerability was quietly fixed in the popular library Node.js that for years allowed the personal data of millions to be stolen.	A quota for women in ICT would be good for the sector.	Despite earlier fears, the National Cyber Security Center has not detected any war related digital attacks on Dutch interests.
Username:	@Daan1337	@RobinQ	@Irene1989	@ir.Bakker
Attribute:	Getekend door Master of Science in Datawetenschap	Getekend door Master of Science in Software Science	Getekend door medewerker IBM	Getekend door medewerker ICT Nieuws
Attribute (translated):	Signed by Master of Science in DataScience	Signed by Master of Science in Software Science	Signed by employee IBM	Signed by employee ICT News
Hours since posted (1,16):	16	5	7	5
Number of comments (50,100):	54	52	90	67
Number of retweets (150-300):	284	292	235	184
Number of likes (500-800):	740	755	716	748
Truth value:	Mostly true. "Ahead of her time" is subjective, the rest is factual (Füegi & Francis, 2015).	Almost certainly false, unless they did it secretly. Defenitly baseles.	To opinionated and unclear what "good" means to fact check.	Mostly true. "fear" is subjective. No digital attack in relation to the war is true (Centrum, 2022)

Table 6: Quote translation table

Author, code, line and document informatio	Code(s):	Original (Dutch):	Translation (English):
1:4 ¶ 37 in 1	Interactions - Comment: Correcting a message that is wrong	Ja. Hmmm. Hier zou ik wel een reactie onder willen zetten. Want dat, dus dat is van die een rare dingen, wat die zegt, ben je sneller geneigd om een mening te geven. En ik vind dat hij ongelijk heeft.	Yeah, hmmm. Here I would want to give an reaction. Because that, so if you say these weird things, which they say, than I am more inclined to react. And I think he is wrong.
1:51 ¶ 167 in 1	Credibility - Expertise - Authenticity method percieved user expertise - Relevance: Displaying domain knowledge helps making a judgement	Ja, Het is een gevoel, krijg je daarbij. Omdat andere mensen die er meer verstand van hebben hebben gezegd van deze dit account klopt of deze inhoud die klopt. Dan ga je ervan uit, want das hun expertise.	[Twid] gives a positive feeling. Because others who are more knowledgeable said that this account or more so this content that it is true. Then you

2:5 ¶ 44 in 2	Credibility - Expertise - Authenticity - method perceived user expertise - Objectiveness of knowledge field	Niks. Dat betekent niks, want nee, volgens mij betekent dat niet zoveel. Dat betekent dat de de persoon die de zender is als het gaat om een bericht van zichzelf dat hij dat hij is, wie hij zegt dat hij is. Maar dat dat in zichzelf zegt weinig over de kwaliteit van van de bijdrage. Dus ik denk,ilk heb niet zoveel vertrouwen in de in de kwaliteit van wat dat betekent. Oké, ik zeg vertrouwen. Ik denk dat ik daar niet veel status aan toe ken. Dat is iets anders dan vertrouwen, want ik geloof wel best dat ze geverifieerd zijn. Dus ik geloof dat iemand die zegt wie is als die een bepaald vinkje heeft, dat hij dat ook wel is. Maar daarmee wordt zijn tweet niet automatisch meer of minder waard.	Nothing. That doesn't mean anything, because no, I don't think that means much. That means the person who is the sender when it comes to a message from himself that he is, who he says he is. But that in itself says little about the quality of the contribution. So I guess, I dont trust in the quality of what that means. OK, I say trust. I guess I don't attribute much status to that. That is different from trust, because I do believe that they have been verified. So I believe that someone who says who is if they have a certain check mark, that they are. But that doesn't automatically make his tweet worth more or less.
2:44 ¶ 130 in 2	Credibility - Expertise - Authenticity - method perceived author expertise - Objectiveness of field of knowledge: Determining the value of attributes is difficult	Daarmee heb je dus nu een intro een kwaliteitsverschil geïntroduceerd in de Twid betekenis, Die het voor mij, onmogelijk maakt om te checken of ik of het aan geloofwaardigheid wint. Wel aan oké hij, hij is degene die hij misschien beweert te zijn. Maar aan de geloofwaardigheids kant doet het helemaal niks. Want soms zegt het wat en soms niet, Maar ik ken niet alle verificatiemethoden als er staat tuinman. Of als er staat architect. Dat zijn geen beschermde titels. Ik mag mijn tuinman noemen, ondanks dat ik gras niet van een dennenboom weet te onderscheiden. Maar er staat wel dat ik het ben. Bij een arts en waarschijnlijk ook bij een rechter en bij een politiemann. Als dat labels in Twid zijn. Dat zijn dingen niet iedereen kan zich rechter noemen, dus ik neem aan dat daar dan wel een verificatiemethode achter zit die dat checkt en dat hij inderdaad werkt bij een rechtbank of bij een kanton gerecht. Maar bij de tuinman is dat niet zo, dus nu introduceer je eigenlijk doe je iets heel ergs met Twid. Ja verstrekt een schijn van betrouwbaarheid. Die er soms wel. en soms niet is. Want het is de authenticiteit, die zal je wel checken dat kloppen. Geloof ik wel. Maar ja, je hebt al gemerkt. Ik haak meer aan op het begrip, geloofwaardigheid en die wordt minder, want Ik kan nu niet meer onderscheiden. Zit er een solide betekenis achter die ondertekening of is het bagger?	With this you introduce a quality difference in Twid which makes it impossible to check credibility gain. He may be who he claims to be, but it does nothing for credibility. Sometimes it is meaningful and sometimes it isn't. I do not know all attributes. If it says gardener or architect. Those aren't protected titles. I can call myself a gardener, despite being unable to differentiate grass from a tree. But it says I'm a gardener anyway. With a doctor, or a policeman or a judge. If those are Twid labels, that's something not anyone can call themselves so I assume for a judge it checks if someone indeed is working for a judicial district. This does not hold true for a gardener, which means you are introducing something terrible with Twid. You give an illusion of credibility. Sometimes this holds but sometimes it doesn't. The authenticity, that will be checked if it holds. I believe that. But as you have noticed, I get more out of the concept of credibility. And that becomes less, because I can no longer differentiate. Is there a solid meaning behind the signature or is it garbage.

3:14 ¶ 64-68 in 3	Authenticity - Authenticity method: Any authenticity method contributes to making a post feel more authentic	B: Ja, dan vind ik ze allebei wel wat toevoegen, dus de standaard methode en de nieuwe methode voegen allebei wat toe. Waarbij die nieuwe methode misschien ook nog iets meer inhoud geeft aan die authenticiteit, dus maakt echt wel dat je wat meer achtergrond én wat meer informatie hebt over over de post wat daarachter zit, wie daar achter zit.A: Dus als ik het goed begrijp dan zorgen ze allebei voor een gevoel van authenticiteit. Maar de Twid methode die voegt daarnaast ook nog informatie toe? B: Precies die andere geef ik meer een soort authenticiteit van: Het is echt een oorspronkelijke persoon die geverifieerd bestaat, heeft zijn account geverifieerd gemaakt en die ander geeft nog wat meer achtergrond.A: Dus als het enkel over authenticiteit hebben, dan is dat wel gelijkwaardig?B: Ja gelijkwaardig.	B: Yes, then I think they both add something, so the standard method and the new method both add something. Whereby that new method may also give a little more information to that authenticity, so it really gives you a little more background and some more information about the post, what is behind it, who is behind it. A: So if I understand correctly, they both provide a sense of authenticity. But the Twid method also adds information? B: Exactly the other one I give more authenticity of: It's really an original person who exists and is verified, made their account verified and the other one gives some more background. A: So if we're just talking about authenticity, then that's equivalent? B: Yes, equivalent.
3:35 ¶ 128 in 3	Credibility - Expertise - Content quality - Agenda: The message instils no fear that the user has motivation to misinform	Ik weet niet waarom je hierover zou liegen, dus vandaar dat ik denk van nou, het zal wel kloppen	I do not know why you would lie about this, so that's why I think it is correct
4:30 ¶ 179 in 4	Authenticity - Assumption: Assumes posts are authentic unless informed otherwise	Even kijken nou wat me opviel is dat ik eigenlijk alle tweets wil als authentiek beschouw. Ja, eigenlijk omdat ik geen reden heb om aan de echtheid te twijfelen.	Let me see, well what stood out to me is that I pretty much assume all tweets are authentic. Yeah, because I really don't have any reason to doubt they are not.
5:25 ¶ 121 in 5	Credibility - Trustworthiness - Author - Social proof: The potential backlash of lying is reason enough for it to be true	Ja kijk door dat vinkje zou ik dus weer eerder denken dat het echt is. Gewoon weer omdat ik denk van ja. anders wordt je daar op aangesproken.	Yes, because of that badge I would again think it's real. Again because I think otherwise you will be held accountable for it.
5:116 ¶ 215-216 in 5	Credibility - Trustworthiness - Authenticity method - Reputability: Understanding of the workings of verification	B: Geloofwaardigheid van die van Twitter zelf is wel wat verminderd. Best wel wat denk ik.A: waar ligt dat aan?B: Ja, het feit dat er gewoon. Er is geen controle meer of zo weet je, je moet gewoon een keer iets doen en dan is er gewoon niet meere echt controle.	B: Credibility of [Verification] even reduced. Quite a lot actually. A: How come? B: Yeah, the fact that. There is just no control anymore, you do something wants and then they don't really check on you anymore
6:40 ¶ 110 in 6	Credibility - Expertise - Content quality - Sourcing: Lack of sourcing	Ik ga niet ja, ik ben niet iemand die dingen zomaar aanneemt zonder dat ik een manier heb om dat te checken.	I'm not one to take things for granted without having a way to check.
6:80 ¶ 173 in 6	Quality concern - Twid: Twid encourages a lazy userbase	Ik denk nog steeds wel dat daar een gevaar achter kan zitten dat mensen niet meer gaan factchecken.	I still think there may be a risk that people will no longer fact check.

7:27 ¶ 98 in 7	Scrutiny - Methodology: Warrants further research because of interest in tweet	Ik merkte sowieso, heb ik net al gezegd, ik controleer eigenlijk bijna altijd alles wat ik zie voordat ik interactie mee	I noticed, as mentioned , I actually almost always check everything I see before interacting with it
7:48 ¶ 132 in 7	Credibility - Trustworthiness - Author - Authenticity method: Any authenticity already contributes more than nothing at all	Ja, je merkt toch wel dat dat als er enige vorm van verificatie optreed, ik er wel wat meer positiever tegenover sta	Yeah, you do notice that as soon as any type of verificaiton happens, I dispose it slightly more positive
7:68 ¶ 179 in 7	Authenticity - Content: Opinions always come across as authentic	Eeuh in dit geval zou ik niet heel veel twijfels trekken aan de authenticiteit. Dit heeft vooral te maken met het feit dat dit een mening over een grootschalige issue is, dan maakt het mij niet uit of je mijnwerker uit Limburg of een tweede Kamerlid bent	Euh in this case I wouldn't put the authenticity in doubt. This is mostly related to it being an opinion of a large-scale issue, then I don't care if you work in a mine or are a member of Parliament.
7:75 ¶ 192 in 7	Scrutiny - Authenticity method: Any method allows less scrutiny	[in referentie tot authenticatie methode]Dus dan hoef ik al die controle stappen waar ik het eerder over heb gehad, hoef ik dan bijna zelf niet meer uit te voeren. Ik kan het altijd doen, maar de incentive wordt voor mij weggenomen.	[In reference to presence of authenticity] So then I don't have to do all those control steps I talked about earlier. I can always do it. But the incentive is removed for me.
8:32 ¶ 129 in 8	Credibility - Expertise - Authenticity method perceived author expertise - Absence of authenticity method: Not signed messages seem like their users lack expertise	Maar dat niet getekend dingetje. Ja, Ik denk wel dat het een beetje afdoet aan de mening van het Truus kort.	Yeah, but that not signed indicator thing. Yeah, I think that discounts the opinion of [author]
8:43 ¶ 172 in 8	Interactions - Interacting: The message is not interesting	Het weer ook niet echt iets wat Ik denk van, dit spreekt me aan, dus ik zou het ook niet delen. Retweet ofzo of verder interactie.	Again, this is not something that makes me think, that appeals to me. So I would not share it, retweet or otherwise interact.
9:12 ¶ 38 in 9	Credibility - Trustworthiness - Authenticity method - Abuse sensitivity: Verification is vulnerable to someone sending a message from your opened device	Ik kan inloggen op dat ding. Dan kan mijn vriend vervolgens op mijn telefoon daar iets opzetten nog steeds op mijn account, dan ben ik het nog steeds niet zelf die het erop zet.	I can log in. Then my boyfriend can then send something from my phone while still on my account, then it's still not me who put it up.

10:19 ¶ 79 in 10	Credibility - Expertise - Content quality - Foreknowledge: The message is misinformation based on the participants knowledge	Ja, die vinkjes hebben op mij geen invloed hierin. Het is echt puur onzin wat ze schrijven. Een man met een blauwe vinkje die schrijft over dat chia zaad een betere bron is van omega 3 en ik denk echt van dat klinkt als complete onzin laten we het opzoeken dus.	Yes, those check marks have no influence on me in this. It's really pure nonsense what they write. A guy with a blue tick writing about chia seeds being a better source of omega 3 and I really think that sounds like complete bullshit so let's look it up.
10:41 ¶ 120 in 10	Credibility - Expertise - Content quality - Logical: The message is non-sensical	Statement zelf heb ik al twee keer gelezen, maar dat klinkt all niet geloofwaardig. Gewoon van, hoe kun je een computer programmeur zijn als de eerste computer er nog niet zou zijn?	I've already read the statement twice, but that doesn't sound credible at all. Just like, how can you be a computer programmer if the first computer wasn't around yet?
11:7 ¶ 51 in 11	Credibility - Expertise - Authenticity method percieved user expertise - Relevance: Attributes need to be very specific to the tweet to be useful	Bij de laatste twee verificaties. Ja medewerker van de Tweede Kamer, dat kan ook de koffie juffrouw zijn. Zegt ook niks.	With the last two [authenticity methods]. Yeah. Employee of 2nd chamber of parliament, could also be a coffee lady. Does not mean anything
12:4 ¶ 40 in 12	Credibility - Trustworthiness - Authenticity method - Design discernibility: Unintuitive design of the authenticity method	Niet getekend, ik weet niet wat dat betekent in dit geval.	Not signed, I do not understand what that means here
13:6 ¶ 64 in 13	Credibility - Trustworthiness - Authenticity method - Design discernibility: Unintuitive design of the authenticity method	Nou, ik denk als als zo'n blauw vinkje inderdaad zou betekenen dat er een zekere check is gedaan om de waarheid ervan of op de betrouwbaarheid van de personen	Well I still believe the blue badge means that a certain check is done about the truth of it or the trustworthiness of the people
13:7 ¶ 65 in 13	Credibility - Trustworthiness - Authenticity method - Author message relation: Expertise does not exclude the possibility for authors' to push an agenda	Ja, daar heb ik niet zoveel mee. Vroeger op de carvan cevitam stond ook dat het aanbevolen werd door een art, ook absolute kul natuurlijk	Yeah that does not do much for me. Carvan Cevitam supposedly was recommended by medical doctors in the past, which was absolute nonsense
14:13 ¶ 93 in 14	Authenticity - Familiarity: An unknown user comes across as less authentic	Ik heb geen enkele indicatie om te kijken om de authenticiteit van deze tweet te duiden. Ik zou niet weten waar ik dat aan moet kennen. Maxime Hendriks ken ik niet.	I do not have any indication to determine the authenticity of this Tweet. I would not know how to recognize it. I do not know Maxime Hendrixks.

B.1 Code tables

Most codes represent perceptions that loosely give a positive, negative, or neutral experience of concepts and elements in the theme. There are also some codes that are more of a contemplative nature or escape clear positioning. Positive codes are green, negative red, neutral yellow and blue is those that present an unclear positioning. This subdivision is meant to be a bit loose, serving more as an aid than a definitive position on the user experience within the code.

Table 12: Absence of authenticity method

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
No authenticity seems like users lack expertise	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2
Not signed messages seem like their users lack expertise	0	0	0	0	0	1	0	2	3	0	0	0	1	1	8
Total	0	0	0	0	0	1	0	2	5	0	0	0	1	1	10

Table 13: Effect

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
The effect of authenticity is limited, but noticeable	0	0	0	0	1	0	1	0	1	0	0	0	0	1	4
The effect of authenticity is too limited to be of any use	0	1	0	0	1	0	1	1	1	0	0	0	0	0	5
Total	0	1	0	0	2	0	2	1	2	0	0	0	0	1	9

Table 14: Objectiveness of a field of knowledge

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Determining value of attributes is difficult	0	4	1	0	0	1	0	0	0	0	1	1	0	0	8
Field of knowledge with controvertible issues may give Twid users an aura of absolute truth	0	0	0	0	0	1	0	0	0	1	0	0	0	0	2
Objective content is well suited to receiving expertise labelling	0	0	0	0	1	0	1	0	0	0	1	0	0	0	3
Protected titles would be more useful to distinguish between	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
Total	0	6	1	0	1	2	1	0	0	1	2	1	0	0	15

Table 15: Relevance

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Attributes need to be very specific to the tweet to be useful	0	0	0	0	0	0	0	0	3	0	3	1	0	0	7
Details of personhood less important than truth of expertise	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Displaying domain knowledge helps to make a judgement	3	0	6	4	2	5	4	6	9	3	6	0	7	10	65
Education is not a meaningful attribute	0	0	0	0	0	0	1	0	0	0	0	2	0	0	3
Education is valued as an attribute	0	0	0	2	1	0	4	1	2	0	1	0	0	0	11
The attribute makes the user come across as unqualified	0	0	3	0	0	1	0	0	1	0	0	3	0	0	8
Total	4	1	9	6	3	6	9	7	15	3	10	6	7	10	96

Table 7: Assumption

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Assumes nothing about authenticity unless informed otherwise	1	0	0	0	0	0	1	0	0	0	1	2	0	0	5
Assumes posts are authentic unless informed otherwise	0	0	0	9	2	2	0	0	0	1	0	0	0	0	14
Assumes posts are inauthentic unless informed otherwise	0	0	0	0	0	0	2	0	0	0	1	0	0	0	3
The participant associates the name of this author with a demographic complying with the message content	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Total	1	0	0	9	3	2	3	0	0	1	2	2	0	0	23

Table 8: Authenticity method

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
An unknown authenticity method feels less authentic	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Any authenticity method contributes to making a post feel more authentic	2	0	2	1	0	1	2	0	0	0	0	1	0	0	9
Having the anonymity and freedom of not having an authenticity method at all has a certain value	0	0	0	1	0	0	1	0	1	0	1	0	0	0	3
No authenticity method contributes to lower authenticity	0	0	1	0	0	0	1	0	0	0	1	0	0	0	3
Only valuable to confirm identity (of a famous person)	5	3	2	1	3	4	3	4	1	1	2	1	1	1	32
Using Twid is an extreme contributing factor in users seeming more authentic	0	0	2	1	1	2	3	3	2	1	1	2	3	1	22
Using Verification is an extreme contributing factor in seeming more authentic	5	3	2	2	0	1	2	3	3	4	0	2	0	1	28
Total	13	6	9	5	5	8	11	11	6	7	4	6	4	3	98

Table 9: Content

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Authenticity is irrelevant for messages meant for entertainment	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Opinions always come across as authentic	1	0	0	2	5	3	2	5	0	1	0	0	2	0	21
Total	1	0	0	2	5	3	2	5	1	1	0	0	2	0	22

Table 10: Design elements

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
The profile picture contributes in making the post seem more authentic	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
The Twitter handle contributes to making the post seem more authentic	0	0	1	0	0	0	0	0	0	1	0	1	0	0	3
The Twitter handle makes the post seem less authentic	1	1	0	0	1	0	0	0	0	0	0	0	0	0	3
Total	1	1	2	0	1	0	0	0	0	1	0	1	0	0	7

Table 11: Familiarity

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Already knowing a person contributes in how authentic a post is	0	2	1	1	0	0	0	0	0	0	0	0	0	0	4
An unknown user comes across as less authentic	1	0	0	0	0	2	1	0	1	0	0	0	0	2	7
Total	1	2	1	1	0	2	1	0	1	0	0	0	0	2	11

Table 16: Agenda

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
The message content instils no fear that the user has motivation to misinform	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
The message seems to be an advertisement	0	0	0	0	2	0	0	0	0	1	0	0	1	0	4
Total	0	0	2	0	2	0	0	0	0	1	0	0	1	0	6

Table 17: Foreknowledge

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
The message corresponds with participants pre-existing belief	0	0	0	0	0	1	0	0	0	0	0	0	1	1	3
The message corresponds with participants pre-existing knowledge	0	0	0	0	1	1	1	0	0	1	1	0	0	1	6
The message is agreeable	1	2	0	1	1	1	0	1	0	0	0	0	1	0	8
The message is disagreeable	0	1	1	0	1	0	0	1	0	0	0	0	0	0	4
The message is misinformation based on the users knowledge	0	1	6	1	0	0	0	0	1	1	3	0	3	1	17
The message is old information or common knowledge	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3
Total	1	4	7	2	3	6	1	2	1	2	4	0	5	3	41

Table 18: Importance

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Content is the most important factor in judgements about messages	0	2	0	0	0	0	1	0	0	2	2	0	1	0	8
Total	0	2	0	0	0	0	1	0	0	2	2	0	1	0	8

Table 19: Logical

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
The message is non-sensical	3	1	1	1	1	1	0	0	0	2	0	0	0	0	10
The message seems to make sense	1	0	0	2	0	0	0	0	1	2	2	1	3	1	13
Total	4	1	1	3	1	1	0	0	1	4	2	1	3	1	23

Table 20: Sourcing

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Credibility cannot be determined by message alone	0	3	1	0	0	1	0	0	1	1	2	2	2	1	14
Lack of sourcing	1	1	0	4	4	3	4	1	3	1	1	1	1	3	28
Total	1	4	1	4	4	4	4	1	4	2	3	3	3	4	42

Table 21: Spelling

	Participants:														
Codes:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Avoiding spelling mistakes and other user errors	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Spelling mistakes	1	0	0	0	0	0	1	0	0	0	0	0	0	0	2
Total	3	0	0	0	0	0	1	0	0	0	0	0	0	0	4

Table 22: Abuse sensitivity

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Allows for anonymity, leading to abusive behaviour	1	0	0	0	0	0	1	0	0	0	0	0	1	2	5
Authenticity methods are in essence unsafe	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
Twid seems vulnerable to creating shell organizations/ fake attributes	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Two factor authentication seems safer then alternatives	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Using an ID seems less vulnerable to abuse	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Verification is vulnerable to someone sending a message from your opened device	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Total	1	1	1	0	0	1	1	0	2	0	0	1	1	2	11

Table 23: Design discernibly

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Unintuitive design of an authenticity method	6	8	3	6	7	8	5	1	6	8	0	5	6	2	71
The harshness of the "Not signed" indicator	0	0	3	2	1	0	0	0	1	0	0	0	0	0	7
The obvious presence of an authenticity method based on its design	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2
Total	7	8	6	8	9	8	5	1	7	8	0	5	6	2	80

Table 24: Reputability

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Familiarity with the authenticity method	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Understanding of the workings of verification	0	0	0	0	3	0	1	0	2	0	0	0	0	0	6
Unfamiliar method of authenticity	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Total	1	1	0	0	3	0	1	0	2	1	0	0	0	0	9

Table 25: Authenticity method

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
A sense of authority is given by the authenticity method	0	0	0	0	0	1	0	0	0	2	0	0	0	0	3
Any authenticity already contributes more than nothing at all	1	0	0	0	0	0	1	0	1	0	1	0	1	0	5
Attributes provide no guarantee over content quality giving false authority	0	3	1	0	0	2	1	1	0	1	0	0	0	0	9
Authenticity allows for trusting strangers	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Relevant attributes add a feeling of trustworthiness	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2
The extra effort others expand to us an authenticity method	0	0	0	0	3	0	0	0	1	0	0	0	0	0	4
Twid makes the user seem capable	0	0	2	1	0	0	2	0	0	0	0	0	0	0	5
Unsigned fresh information may still be verified later	0	0	0	0	0	1	0	2	0	0	0	0	0	0	3
Verification makes anything you do a bit more believable	1	0	0	7	0	0	0	2	0	3	0	0	0	0	13
Total	2	3	4	8	3	5	4	5	3	6	1	0	1	0	45

Table 26: Author message relation

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Expertise does not exclude the possibility for authors' to push an agenda	0	0	0	0	1	1	0	0	0	0	0	0	2	0	4
Extremely authentic messages seem more credible	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Politicians are untrustworthy, especially when talking about politics	0	0	0	0	2	0	0	0	0	0	3	1	0	0	6
Statements without sources should only be made by the source of a statement	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
The authenticity method serves as a source	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2
Total	0	0	0	0	3	1	1	0	0	2	4	1	2	0	14

Table 27: Fame

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
No verification implies that a user controls their own account	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Notability requires one to be credible	0	0	2	0	1	0	0	0	0	0	0	0	0	0	3
Verification means a PR-team sometimes manages your account	0	0	0	0	0	1	0	0	0	3	0	0	0	0	4
Total	0	0	2	0	1	2	0	0	0	3	0	0	0	0	8

Table 28: Social proof

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
The post has a lower number of interactions	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
The potential backlash of lying is reason enough for it to be true	0	0	0	0	4	0	2	1	1	2	0	0	0	0	10
Total	0	0	0	0	4	1	2	1	1	2	0	0	0	0	11

Table 29: Comment

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Ask for sources	0	0	0	0	0	1	4	0	0	0	0	0	0	0	5
Correcting a message that is wrong	2	0	3	0	0	0	1	0	0	0	0	0	0	0	6
Total	2	0	3	0	0	1	5	0	0	0	0	0	0	0	11

Table 30: Interacting

Participants:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Codes:															
Any authenticity method	1	1	0	0	1	0	0	2	0	1	0	0	0	0	6
Never or almost never interacts with tweets	1	0	0	1	0	0	0	0	0	0	0	0	0	0	2
The message is not credible	1	0	2	0	1	0	0	0	3	0	0	0	0	0	7
The message is not interesting	4	5	0	5	4	5	2	3	3	4	3	2	3	3	46
There is no authenticity method	0	0	0	0	0	0	2	1	0	0	0	0	0	0	3
Unwilling to further debate	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2
Total	7	7	2	6	6	5	5	6	6	5	3	2	3	3	66

Table 31: Like

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Agreeable message	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2
Interesting message	0	0	1	0	1	0	2	0	0	0	0	0	0	0	4
Propagate to trigger discussion	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total	1	0	1	0	1	2	2	0	0	0	0	0	0	0	7

Table 32: Share

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Acquaintances interested in the subject	0	0	2	0	0	0	0	0	0	0	0	0	2	0	4
Pole acquaintances about statement	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Warning acquaintances about the consequences of the message	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Total	0	0	3	0	0	0	0	0	0	1	0	0	2	0	6

Table 33: General

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
A combination of both authenticity systems seems a good idea	2	0	1	0	0	0	0	0	0	0	0	0	0	0	3
Authenticity methods that are available to everyone, not just a small group	0	1	0	0	0	0	0	0	0	0	0	0	2	1	4
Communicating with real people is important to me	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
These issues are important and good that things are being done to address it	1	3	0	0	1	0	0	1	2	0	0	0	2	1	11
Total	3	4	1	0	1	0	2	1	2	0	0	0	4	2	20

Table 34: Twid

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Expanding extra effort in order to use an authenticity method yourself	0	0	0	0	0	0	0	0	1	0	0	0	1	0	2
Expertise provided has to be from a trusted authority	0	0	1	1	0	0	4	2	2	0	3	1	0	1	15
Mobile availability is a must	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Only useful if the certificates cannot be faked	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2
Only useful with a high adoption rate	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2
Some insight into conflict of interest would be useful	0	0	0	0	0	0	0	0	0	0	1	0	0	1	2
Speed of twitter means that authenticity method should also be quick or useless	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
The participant wants to broadcast their own expertise	1	0	2	2	0	0	0	0	0	0	0	0	0	1	6
Twid encourages a lazy userbase	0	0	0	0	0	4	0	0	0	0	0	0	0	0	4
Twid has potential if implemented well	0	0	1	0	0	5	0	0	0	1	0	0	0	0	7
Twid should allow for some peer review function	0	0	0	0	0	1	2	0	0	3	0	0	0	0	6
Twid's name is confusing	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Would allow participant more ability to interact and trust the information on social media	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Total	3	0	4	3	2	10	7	2	4	5	4	1	2	3	50

Table 35: Verification

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Authenticity methods focussed on identity over content are not helpful	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Providing your ID for an authenticity method	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Twitters method available for everyone would be great	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Verification stays forever which is a negative	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Total	1	1	0	0	1	0	0	0	1	0	0	0	1	0	5

Table 36: Authenticity method

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Any authenticity allows less scrutiny	0	0	0	0	0	0	7	0	0	0	0	0	0	0	7
No authenticity increases scrutiny	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Twid attributes would lower misinformation lowering scrutiny	0	0	0	0	2	0	0	1	0	0	0	0	0	0	3
Total	0	0	0	0	2	0	8	1	0	0	0	0	0	0	11

Table 37: Foreknowledge

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Familiarity with the user	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2
Pre-existing belief of the participant	0	0	0	0	0	0	0	0	1	0	0	0	1	0	2
Total	0	0	0	1	0	0	0	0	2	0	0	0	1	0	4

Table 38: Methodology

Codes:	Participants:														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Always fact checks especially if relevant	0	0	0	0	0	1	1	0	1	0	0	0	0	0	3
Read comments to check the truth to the message	0	0	0	1	0	1	0	0	0	4	0	0	0	0	6
Warrants further research because of interest in tweet	0	2	1	0	0	2	7	0	4	1	2	0	3	0	22
Total	0	2	1	1	0	4	8	0	5	5	2	0	3	0	31

C Appendix C

C.1 Twid project proposal

SIDN fonds 2021 - Getting a grip on disinformation
Twid: Fighting Fake News on Twitter



Twid: Fighting Fake News on Twitter

Main applicant: Bernard van Gastel

Co-applicants: Bart Jacobs and Hanna Schraffenberger

Authors/concept development: Bart Jacobs, Hanna Schraffenberger, Bernard van Gastel, Paul Graßl, Leon Botros, Mariska Kleemans
Radboud University
22/04/2021

Executive summary

With the rapid growth of online information came the exponential spread of disinformation. Recent examples include fake news about Covid-19 and false political advertisements. Disinformation is often counteracted by fact-checking. However, it is difficult and labour-intensive to rectify and remove false information that has already been published. Furthermore, discussions about what is true or not, certainly in a political or morally charged setting with many perspectives, quickly derail. An alternative approach, as proposed by one of the applicants (see <https://ibestuur.nl/weblog/teken-tegen-nepnieuws>), is to concentrate on authenticity of information instead of on truth. This *Twid* project applies this to Twitter. It provides Twitter users – with *Twid*'s browser plugin – with certainty about who is the source of a Twitter message known as tweet (“source authenticity” of tweets) as well as certainty that a message has not been modified or changed since its publication as tweet (“message authenticity”).

The proposed approach is close to traditional approaches, where, for instance, knowing that a message comes from a certain newspaper helps people in their credibility assessment. Unfortunately, on social media platforms, such valuable background information about the source of messages is often lacking. The *Twid* project aims to fill this gap in the context of Twitter, by providing users with verified information about the origin of tweets. Concretely, the project proposes a browser-plugin that allows Twitter-users to link personal verified information to their accounts and sign their tweets. Other users who have the plugin enabled will then be able to see this information about who has posted/signed the tweets, helping them with judging the credibility of the messages. (For instance, knowing that someone is working at a renowned hospital might give credibility to their tweet about COVID-19).

To link personal verified information with Twitter, the *Twid* project utilizes the existing identity management app IRMA, which already allows users to prove properties about themselves and sign digital content. With *Twid*, users can link the personal information that they have collected in IRMA (for instance, their real name from the Dutch Civil Registry, their city of residence or that they have an email-address ending with, e.g., @radboudumc.nl) to their Twitter account and use it to sign their tweets. This means that source and message authenticity is guaranteed via digital (cryptographic) signatures.

The *Twid* project aims to demonstrate the feasibility of this approach with a Proof of Concept (PoC). This PoC is meant to demonstrate the feasibility of guaranteeing authenticity to various message services. The ultimate goal is that Twitter adopts the technology. The PoC concentrates on Twitter since Twitter is very vulnerable to fake news and is itself also exploring options to fight it. The PoC plugin can in principle be used by many people (such as politicians and opinion leaders in The Netherlands, and beyond) and may thus attract much visibility and put pressure on Twitter. At the same time, the applicants are seeking contacts with Twitter about *Twid*.

Problem

In the current online media landscape, the authenticity of messages certainly cannot be taken for granted. Real-world examples that illustrate this are abundant: An extreme case occurred in 2016 when AWDNews (falsely) reported that Israeli Defence Minister Moshe Yaalon was threatening to destroy Pakistan in case it would send troops to Syria. Pakistan Defence Minister Khawaja Asif reacted to this unfounded report on his official Twitter as if it were real, believing that the threat actually had been voiced by Moshe Yaalon and reminding Israel that Pakistan was also a nuclear power (see <https://edition.cnn.com/2016/12/26/middleeast/israel-pakistan-fake-news-nuclear/index.html>). More recently, various high-profile Twitter accounts, including those of Elon Musk, Jeff Bezos, Bill Gates, Barack Obama, Joe Biden, Kanye West, Geert Wilders, Apple, and Uber were hacked and used to present followers with a bitcoin scam. During the short time that the tweets were visible, the associated bitcoin account received hundreds of contributions (see <https://www.bbc.com/news/technology-53425822>) indicating that users believed that the message came from the official holders of the accounts, thus falsely assuming that the claimed source was the actual source.

Traditionally, media outlets such as newspapers have been validating sources and their messages and acted as gatekeepers. However, the current online media landscape, where everyone can easily publish

and share information, leaves consumers with the challenge of assessing whether sources and their messages are authentic. How are consumers supposed to know for sure whether a claimed source is the actual source (e.g., whether a blog post detailing protests in Belarus actually has been published by the independent journalist whose name is featured below the piece)? What guarantees consumers that a tweet about Corona by a supposed specialist in a Dutch hospital who calls herself “Nienke Janssen”, is indeed coming from a medical specialist in the Netherlands? And how can one be sure that the real account holder and no hacker has posted a tweet?

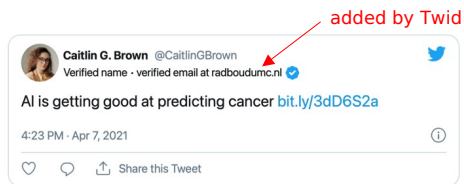
Proposed solution

This project proposes to address these problems via two additions (1) authentication, by linking verified personal information to Twitter accounts and (2) digital signing of individual tweets. Senders of tweets can use only (1), or both (1) and (2). Receivers of tweets will see the authenticated identity and tweet signatures *in the regular Twitter webinterface* when they use *Twid*'s browser plugin. For a consumer who is viewing tweets, the verified account information provides certainty about the person or organization who is behind the account. The signatures that are attached to individual tweets furthermore ensure that the particular tweet has not been altered or posted by a hacker (unless the hacker also has access to the phone and IRMA-app of the account-holder, which is extremely unlikely).

The proposed solution is explained in more detail and illustrated below, using mockups featuring a doctor working at Radboudumc hospital. (The tweets and doctor are made up for illustration purposes.)

Verified account information

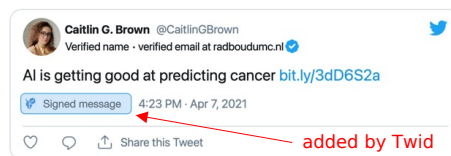
The main functionality of *Twid* is that users can attach verified information about themselves to their Twitter account with IRMA. This involves a one-time enrolment that ensures the provided information is displayed with *all* tweets of the account holder. In the following mockup, a doctor named “Caitlin G. Brown” has linked her name from the Dutch Civil Registry and the ending of her email-address to Twitter (by disclosing this information to *Twid* with IRMA). This information is now displayed when viewers of her tweets have *Twid* enabled. They learn that Caitlin G. Brown is using her real name and has an email-address from Radboudumc Hospital (and thus works there). In particular the information that Caitlin G. Brown has a radboudumc.nl email-address is valuable when interpreting her tweets about medical advances. In current times, one can imagine such background information being particularly relevant when Twitter users make claims about COVID-19 or the situation at certain hospitals. However, doctors are not the only example. Knowing that someone is a working at a certain newspaper, lives in a certain city or is using their real name can likewise be valuable to interpret tweets.



While the verified account information provides viewers certainty about the holder of an account, it does not protect from hacking attacks. In rare cases, the particular account might have been hacked and the tweet might have been posted or altered by hackers. This is where signatures come to the rescue.

Signatures

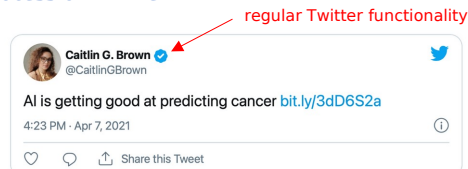
In addition to providing verified account information, *Twid* users also can sign their tweets. This action can be carried out for each individual tweet. If a specific tweet is signed, this is indicated to viewers of the tweet who have *Twid* enabled. The following mockup illustrates a signed tweet.



This helps to counter spreading misinformation by account take-overs (due to lost account credentials, or Twitter hacks). For this PoC, *Twid* users will be signing tweets with the same attributes that they have attached to their account. E.g., if Caitlin G. Brown has attached her real name and her an email-address from Radboudumc Hospital to her account, she will also sign tweets with this information.

Differences to the existing account verification process on Twitter

Twitter already offers some rudimentary account verification (see <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>), using a blue checkmark to indicate so-called “verified accounts”. The following mockup illustrates an account that has been verified by Twitter.



Based on the information provided by Twitter, their verified account program is currently on hold.

However, Twitter has active plans to relaunch the service (https://blog.twitter.com/en_us/topics/company/2020/our-plans-to-relaunch-verification-and-whats-next.html). Our proposed solution differs from and improves the official account verification process currently outlined by Twitter in three major ways:

- Twitter’s account verification is solely available for so-called “notable accounts”. To get verified, one needs to be a “prominently recognised individual or brand” (see <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts> for details). In contrast, *Twid* is for everyone! This is particularly important because a core feature of Twitter is that everyone can post tweets, not just prominent people.
- Twitter’s verification focuses on the identity of the account holder and provides a blue checkmark for all verified accounts. In contrast, *Twid* provides *additional* information about the person in question and shifts the focus from ‘who’ the person is to ‘what’ a person is. To interpret tweets and judge their credibility, a person’s attributes (‘journalist’, ‘doctor’, ‘citizen of city X’) are often more important than a person’s identity.
- Twitter describes its planned account verification to include both automated and human reviews. With *Twid*, there is no need for human reviews as it relies on IRMA and associated issuers to solve the actual authentication problem.

Whereas the first two points make *Twid* very attractive to potential end-users, the first and third point make *Twid*’s approach worthwhile to Twitter and may convince Twitter to adopt the idea and technology. This is one long-term aim of this project.

Limitations of the proposed solution

We present this idea as a partial solution to the complex and multi-faceted disinformation problem and envision it to be used alongside fact-checking. To be clear, *Twid* will not prevent producers from making false claims on Twitter. If, for instance, a user posts false information about possible Covid-19 treatments, this message can still be tweeted. However, when *Twid* is used it is clear who takes responsibility for the message and consumers of this information can use this input in their credibility assessments. Thus, unauthenticated and unsigned messages can be disregarded more easily. In other words, the proposed tool is a means to address the disinformation problem while preserving freedom of speech. The use of *Twid* does not prevent anyone from articulating their opinions and ideas, but it will prevent misleading claims about the origin of the message as well as tampering with online content, thus protecting consumers from mistaking false identities for real ones, falling for imposters or self-proclaimed experts. *Twid* provides consumers of Tweets with valuable information about the source of a tweet, so they can form an informed opinion about the credibility of claims.

Proof of concept

To realize these features as a PoC, the project will develop a browser-plugin (using WebExtensions, supporting both Google Chrome, Mozilla Firefox, and Microsoft Edge) and utilise the existing identity management app IRMA to allow users to digitally sign their tweets and link personal information to their Twitter account. Changes to the Twitter platform itself are not necessary, nor to IRMA.

Identity management with IRMA

For the *Twid* -information and signatures to be meaningful, it is crucial that Twitter users can attach meaningful and verified information to their Twitter account and tweets. The verification of the information is important, because we have to avoid that people sign content with false names or affiliations. This is where IRMA comes in, as an emerging platform for attribute-based authentication and signing. A co-author of this proposal – Prof. Bart Jacobs – has been the driving force behind IRMA: It grew out of earlier research at

Radboud University and is now being rolled out via the non-profit spin-off foundation Privacy by Design (<https://privacybydesign.foundation>), with SIDN now in the driver's seat.

Via the mobile IRMA app, users can reliably prove that a name, an email address, a national identification number ("BSN" in Dutch), a phone number, a bank account number, or a medical registration number ("BIG") is theirs, and thus use this information to login at websites, access online content or sign messages. IRMA is attribute-based in that it allows the user to reveal only relevant attributes (properties) of him/herself. For instance, a user might reveal their BSN to login on a hospital's online portal, reveal that he is older than 16 (and nothing else) in order to play a certain video game, or disclose that he is working at a Dutch University to receive an educational discount. IRMA is getting traction and is being integrated as authentication mechanism in various websites, with (local) government and health care organisations as frontrunners. Besides authentication, the IRMA app also allows users to (freely) generate digital signatures on documents (so far: texts). IRMA currently has over 35.000 registered users, and this number is currently growing with some 100 users per day. This project will give a further boost.

This project will make crucial use of this functionality of IRMA and will link it to Twitter via a browser plugin. *Twid* users will use IRMA to link IRMA-attributes to their account and to sign their Tweets. For instance, one might use IRMA to attach one's real name from the Dutch Civil Registry to a Twitter account and to sign tweets, allowing everyone to see that one is using one's real name and providing them with certainty that a tweet is authentic.

In order to trust the authenticity claims of IRMA, it is important that consumers rely on the attributes that users attach to their accounts and use to sign tweets. The trust chain for IRMA, and thereby also for the certainty about the tweets and their source, is ultimately based on the issuing process for attributes: how certain can one be that IRMA attributes have been issued reliably to a user? This involves an existing process, outside this project, that is operated by the Privacy by Design foundation, in close cooperation with SIDN, the domain main name registrar in the Netherlands. Briefly, issuers of attributes in IRMA need to sign a contract that imposes various diligence duties; in addition, they need to have suitable security (management) certifications. The process of issuing of IRMA attributes is thus in place and not a concern in this project.

Whether the authenticity information provided by *Twid* helps users with their credibility assessment of tweets largely depends on what information is provided. Knowing, for instance, someone has an email-address ending with "@amsterdam.nl" certainly adds credibility to claims made about city planning in Amsterdam. In contrast, knowing that someone has an email-address ending with "gmail.com" carries virtually no additional meaning. On the long term, we envision that professionals in different fields, such as journalists and spokespersons will be able to show attributes describing their role/profession. Furthermore, we envision the Kamer van Koophandel (Chamber of Commerce) will join the IRMA ecosystem. This would allow someone to indicate using KvK attributes that they are speaking on behalf of a company.

For the PoC, the following existing relevant IRMA attributes will be used:

- A user's name and city from the Dutch civil registry ("BRP naam") as issued by the municipality of Nijmegen for everyone in the Netherlands (issued here: <https://services.nijmegen.nl/irma/gemeente/start>)
- The ending of a user's email address ("Domain name") as issued by SIDN (issued here: <https://sidnemailissuer.irmaconnect.nl/issuance/email/>)
- A user's Twitter-URL (issued as an IRMA attribute here: <https://privacybydesign.foundation/issuance/social/twitter/>) will be used for linking the personal information to one's Twitter account.

No changes or additions need to be made to IRMA or Twitter itself. The only tools needed for the PoC is the *Twid* browser plugin and supporting server infrastructure.

The *Twid* browser plugin

The envisioned *Twid* browser plug-in (the client) will be created using WebExtensions to support multiple browsers. It will be validated and developed for Firefox and Chrome. The browser plugin bundles four core functionalities:

- **C1: Authentication (linking verified information from IRMA to Twitter)**

To link personal information from their IRMA app to their Twitter account, users first load their Twitter-account-URL into the IRMA app (as already supported by IRMA). Subsequently, *Twid* will ask users to disclose their Twitter-account-URL together with their name, city and/or domain-name. The attributes that are disclosed alongside the Twitter-account-URL will be stored on the server infrastructure (and made visible to everyone who has the *Twid* browser-plugin installed, see C3+C4).

- **C2: Signing individual tweets**
Twid allows users to sign their tweets. When *Twid*-users write a tweet, an extra 'sign' button will be displayed. When users choose to sign a tweet, the plugin will ask them to disclose the IRMA-attributes that are linked to their account. As a result, everyone who has the *Twid* browser-plugin installed will see that the specific Tweet has been signed.
- **C3: Information check (availability and validity)**
For the consumers/readers/viewers of tweets, the plugin automatically checks whether *Twid* - information is available and valid.
- **C4: Interaction design of plugin**
If valid attributes and/or signatures are associated with an account and/or tweet, this information is displayed to the user (see mockups). This should be easy for the user to interact with, and understand the additional information provided by the *Twid* plugin. Therefore interaction design is needed to understand how the information can best be presented to the user.

Infrastructure

The central infrastructure is the backend for the *Twid* browser plugin. It consists of these parts:

- **S1: Storage of verified information**
If valid attributes and/or signatures are associated with an account and/or tweet, this information is stored in our *Twid* infrastructure. It will consist of a scale-out NoSQL database, that will partition the data to achieve scalability. The infrastructure will be redundant, to achieve availability.
- **S2: Server-side validation of information**
To avoid denial-of-service attacks by injecting false information in the infrastructure, the infrastructure must check new information before storing it. This implies having a IRMA validation service running.

Limitations of the PoC

The browser plugin will be limited to browsers supporting WebExtensions plugins. Hence, the PoC browser plugin will not work on most mobile devices and will require users to install a browser plug-in. However, this is a limitation of the chosen implementation and can be overcome when Twitter chooses to adopt the idea. The infrastructure supporting *Twid* will be designed to scale, however, will initially not be capable of supporting millions of registrations. This will require additional effort and a redesign of the server infrastructure.

Deliverables and planning

Besides the *Twid* browser plugin (numbered C1 to C4 above), and the infrastructure (numbered S1 and S2), there will be an additional deliverable :

- **D1: A demonstration video about *Twid*, including our Proof-of-Concept.**
- **D2: Project website where users can download the plugin and read about our project.**
- **D3: A demonstration Twitter account where users can see how our idea works in practice.**

We foresee a runtime of one year, the indicated months are the expected delivery moments:

- Month 4: Deliverables C1 (authentication), C4 (display of information), and S1 (storage) will be finished, so that users can authenticate themselves, and this authentication information is stored on the server. Information can be displayed inside the Twitter interface, however, this is not yet linked with the backend (see C3).
- Month 6: Deliverable C3 (information check) will be finished and integrated with C4. The core functionality is in place, with information flowing from the client to the server and back, including the authentication flows. The user interface is fully designed.
- Month 8: Deliverable C2 (signing of individual tweets), and D1 (video about *Twid*). Contact Twitter to share D1 with them, and if they are interested, give a demonstration.
- Month 9: Deliverable S2 (server-side validation) will be finished.
- Month 10: Deliverable D2 (public website) will be made public.
- Month 11: Deliverable D3 (Twitter demo) will be made public.

C.2 Interviewguide (Dutch)

Hallo, vandaag ga ik met u een experiment afnemen in een onderzoek wat ik aan het doen ben voor mijn masterscriptie. Voordat wij beginnen is het belangrijk dat ik uw instemming heb met het onderzoek. Daarom ga ik u nu wat uitleggen over het onderzoek, wat we met de data doen en de onderzoeksethiek. Hierna stel ik u wat vragen om zeker te weten dat ik uw volledige toestemming heb waarna ik het onderzoek zal afnemen. De verwachting is dat dit onderzoek 60 minuten duurt. Weet dat u uw toestemming nog kan intrekken tot en met het einde van het experiment en dat u op enig moment kan stoppen met uw deelname zonder daar verdere uitleg voor te hoeven geven. Om u niet te veel te beïnvloeden ga ik u nog niet in detail vertellen waar het onderzoek over gaat behalve dat het met Twitter te maken heeft. Gedurende het onderzoek zal u meer duidelijk worden en aan het einde kan ik uw vragen over het onderzoek in complete openheid beantwoorden. Verder kan ik u zeggen dat ik niet verwacht dat het voor u enige risico's of discomfort mee zou moeten brengen deel te nemen aan dit onderzoek. De data die ik verzamel zal bestaan uit een audio opname van uw deelname die ik tekstueel zal uitwerken. Data die verzameld wordt is naar verwachting relevant voor het in beeld brengen van factoren die invloed hebben op het experiment, of bijdragen aan het beantwoorden van mijn onderzoeksvragen. De data van dit onderzoek zal worden gebruikt om een kwalitatieve analyse te maken van mijn experiment. De data die verzameld wordt zal geanonimiseerd worden en zal conform de eisen van de Radboud universiteit verwerkt en opgeslagen worden. Verder zullen de audio-opnames vernietigd worden binnen 6 maanden na afronding van de studie, de geanonimiseerde transcriptie zullen nog 10 jaar bewaard worden.

- Heeft u nog vragen over hetgeen wat ik u net vertelt heb?
- Heb ik u voldoende geïnformeerd over mijn onderzoek en wat er met uw data wordt gedaan?
- Snapt u dat u op enig moment kan stoppen met de studie en tot en met het einde van het experiment uw toestemming kan intrekken?
- Bent u ouder dan 18 en verkeerd u in tegenwoordigheid van geest?
- Wilt u nog deelnemen aan mijn onderzoek?

Het onderzoek zal beginnen met een aantal vragen die met uw achtergrond te maken hebben. Hierna zal ik u 3 rondes aan tweets laten zien elk gevolgd door een aantal voorbereide vragen. Afhankelijk van uw antwoorden heb ik wellicht nog vervolgvragen.. Elke ronde aan tweets bevat 4 tweets. Terwijl u naar een tweet kijkt zou ik u willen vragen hardop in zo veel mogelijk detail te vertellen wat u voor u ziet en wat u daarover denkt. Bij het onderzoek gaat het vooral over de inzichten en ervaringen die u mij mede kan delen bij wat ik u laat zien, dus wijd vooral zo veel mogelijk als u kan uit bij alles wat we gaan bekijken en wat ik u vraag. Allereerst heb ik een aantal vragen over uw achtergrond die mogelijk invloed zouden kunnen hebben op het onderzoek. Ik start nu de opname.

Pr1: Wat is uw gender?

Pr2: Wat is uw leeftijd?

Pr3: Welk niveau is uw hoogst genoten opleiding?

Pr4: In welke kennisvelden heeft u het meeste expertise?

Pr5: Welk cijfer zou u uw digitale geletterdheid geven op een schaal van 1 tot 5?

Pr6: Welke bronnen van media consumeert u?

Pr7: Hoe vaak bent u op twitter?

Pr8: Hoe bepaalt u of u een tweet wil liken, retweten, delen of een comment wil geven?

We gaan nu kijken naar de eerste 4 tweets. Ik zou u dus willen vragen om hardop te delen wat u voor u ziet en wat u daarover denkt. Indien u een tweet zou liken, retweten, sharen of er een comment op zou geven hoor ik dit ook graag en waarom. Naderhand heb ik wat vragen aan de hand van de tweets, hierbij kunnen we opnieuw kijken naar de tweets.

Po1.1: Wat zijn uw algemene gedachten over de tweets die ik u net heb laten zien?

Zoals u allicht doorhad werden er bij de tweets wat verschillende verificatie methoden gebruikt.

Po1.2: Wat betekent het volgens u om geverifieerd te zijn op sociale media, in het speciaal op Twitter? Hierbij doel ik specifiek op het systeem voor verificatie dat Twitter zelf gebruikt.

Po1.3: Wat is volgens u het doel van verificatie?

Po1.4: Wat zijn uw algemene gedachten over de verschillende verificatie methoden die ik u net heb laten zien?

Po1.5: Welke invloed hadden de verschillende methoden op uw gewilligheid om een post te liken, retweten, sharen of erop te reageren?

Po1.6: Welk verificatie systeem draagt uw voorkeur en waarom? Welk systeem draagt niet uw voorkeur en waarom?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Vervolgens heb ik ik een aantal vragen die gaan over u begrip van de twee verschillende concepten van authenticiteit en geloofwaardigheid, gevolgd door hoe u dat van toepassing vond op de tweets die ik u heb laten zien.

Po1.7: Wat betekent het volgens u als iets authentiek is, in het speciaal in een post op sociale media?

Po1.8: Hoe bepaalt u of een post authentiek is?

Po1.9: Wat is het belang van authenticiteit voor u?

Po1.10: Hoe hadden de verschillende methoden invloed op hoe authentiek een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po1.11: Wat betekent het volgens u om een post geloofwaardig te laten zijn?

Po1.12: Hoe bepaalt u de geloofwaardigheid van een post?

Po1.13: Wat is het belang van geloofwaardigheid voor u?

Po1.14: Hoe hadden de verschillende methoden invloed op hoe geloofwaardig een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po1.15: Is er nog iets anders wat u kwijt zou willen?

U gaat zo kijken naar de tweede ronde van tweets. Deze zal vergelijkbaar zijn met de vorige. Ik zou u weer willen vragen om hardop te delen wat u voor u ziet en wat u daarover denkt. Indien u een tweet zou liken, retweten, sharen of er een comment op zou geven hoor ik dit ook graag en waarom. Verder ben ik ook benieuw of u denkt dat een tweet authentiek is en wat de geloofwaardigheid van de informatie is en waarom. Hierbij mag u de definitie aanhouden dat een post authentiek is als u gelooft dat deze

daadwerkelijk gedeeld is door de persoon van wie de account is. De geloofwaardigheid is het niveau van vertrouwen dat u heeft dat de gedeelde informatie kloppend is. Naderhand heb ik wat vragen aan de hand van de tweets, hierbij kunnen we opnieuw kijken naar de tweets. Indien u geen vragen hierover heeft gaan we naar de tweets kijken.

Po2.1: Wat zijn uw algemene gedachten over de tweets die ik u net heb laten zien?

Po2.2: Wat zijn uw algemene gedachten over de verschillende verificatie methoden die ik u net heb laten zien?

Po2.3: Welke invloed hadden de verschillende methoden op uw gewilligheid om een post te liken, retweten, sharen of erop te reageren?

Po2.4: Welk verificatie systeem draag uw voorkeur en waarom? Welk systeem draagt niet uw voorkeur en waarom?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po2.5: Hoe hadden de verschillende methoden invloed op hoe authentiek een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po2.6: Hoe hadden de verschillende methoden invloed op hoe geloofwaardig een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po2.7: Is er nog iets anders wat u kwijt zou willen?

We gaan zo kijken naar de laatste ronde tweets. Voordat we dit gaan doen wil ik u wat explicieter uitleg geven over de verschillende verificatie methoden. Hierbij pak ik ter illustratie wat tweets uit de vorige ronde erbij. Sommige tweets zijn niet voorzien van enige vorm van verificatie. Dit zijn tweets die u en ik zouden kunnen maken vanaf een account aangemaakt met die naam, zolang het niet in schending is met de gebruikers voorwaarden. Sommige tweets zijn voorzien van de verificatie-methode van Twitter. Dit is enkel beschikbaar voor "notable" accounts. Voorwaarden hiervoor zijn bijvoorbeeld dat mensen politicus zijn, beroemd of dat zij bijdrage leveren aan de publieke discussie. Om van deze verificatie voorzien te worden moet jij op enig moment bewijs leveren aan Twitter dat jij daadwerkelijk de persoon bent die jouw account claimt dat jij bent, bijvoorbeeld door een kopie van je identiteitsdocument op te sturen. Hierna is deze account voor altijd geverifieerd. De laatste methode heet Twid, en is in ontwikkeling op de Radboud universiteit. Het werkt momenteel via een plugin die mensen zelf kunnen installeren op hun browsers, maar zou in principe ook via Twitter of andere sociale media zelf aangeboden kunnen worden. In de huidige implementatie zie je als gebruiker van Twid de ondertekeningen van andere TWid gebruikers. Bij deze methode is het zo dat men individuele tweets ondertekent met een attribuut dat aan u gekoppeld is. Deze ondertekening vergt een extra authenticatie via een losse app genaamd IRMA. Zonder in detail te treden hoe kan ik u wel zeggen dat het niet mogelijk is om een attribuut te hebben waar u niet daadwerkelijk recht op heeft, en dat alle attributen enkel voor u opgeslagen worden. Als iemand ondertekent dat hij Arts is kunt u dus aannemen dat dit ook werkelijk zo is, en is er niet een externe organisatie die bijvoorbeeld de BIG registratie van deze arts bijhoudt. Een IRMA account is altijd gekoppeld aan 1 Twitter account, dus een andere arts kan ook niet voor zijn collega tekenen zonder al zijn inloggegevens. Ook is er in elke ronde een tweet aanwezig van iemand die wel Twid gebruikt maar deze tweet niet ondertekent heeft. Heeft u vragen over deze uitleg? Ik zou u weer willen vragen om hardop te delen wat u voor u ziet en wat u daarover denkt. Indien u een tweet zou liken, retweten,

sharen of er een comment op zou geven hoor ik dit ook graag en waarom. Verder ben ik ook benieuwd of u denkt dat de tweets authentiek is en wat de geloofwaardigheid van de informatie is en waarom. Naderhand heb ik wat vragen aan de hand van de tweets, hierbij kunnen we opnieuw kijken naar de tweets. Indien u geen vragen hierover heeft gaan we naar de tweets kijken.

Po3.1: Wat zijn uw algemene gedachten over de tweets die ik u net heb laten zien?

Po3.2: Wat zijn uw algemene gedachten over de verschillende verificatie methoden die ik u net heb laten zien?

Po3.3: Welke invloed hadden de verschillende methoden op uw gewilligheid om een post te liken, retweten, sharen of erop te reageren?

Po3.4: Welk verificatie systeem draagt uw voorkeur en waarom? Welk systeem draagt niet uw voorkeur en waarom?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po3.5: Hoe hadden de verschillende methoden invloed op hoe authentiek een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po3.6: Hoe hadden de verschillende methoden invloed op hoe geloofwaardig een post voelde?

(a) Wat droeg by aan dat [voorkeurs methode] beter was dan [minder voorkeurs methode]?

(b) Waarom is [genoemde reden] van belang?

Po3.7: Is er nog iets anders wat u kwijt zou willen?

Als aller laatste nog een aantal vragen over wat u van Twid vond.

Po4.1: Wat vond u van Twid?

Po4.2: Zou u deze plugin gebruiken? Waarom?

Po4.3: Zou u nog wat willen veranderen aan Twid om het te verbeteren? Zo ja, wat?

Po4.4: Wilt u nog wat anders kwijt over Twid?

Po4.5: Heeft u nog feedback op dit gehele experiment?

Zoals u misschien al doorhad, is het doel van het onderzoek het in kaart brengen van de perceptie van gebruikers over verschillende verificatie methoden. Behalve uw algemene percepties was ik ook benieuwd naar hoe het u zou beïnvloeden in de interacties die u wilde hebben met de posts en of het invloed had op hoe authentiek en geloofwaardig u de posts vond. Graag wil ik u wel vermelden dat alle tweets die ik gebruikt heb in het onderzoek door mij geconstrueerd zijn, waarbij ik ze zo echt mogelijk heb laten lijken. Verder bevat een deel van de tweets misinformatie, als u wilt kan ik aanwijzen voor welke dat specifiek het geval is, anders raad ik u aan om niet zonder meer te geloven wat u tijdens dit onderzoek gelezen heeft. Heeft u nog verdere vragen voor mij op dit moment? Heb ik nog steeds uw toestemming om de data die ik verkregen heb te verwerken in mijn onderzoek? Dan stop ik nu de opname. Dank u voor uw deelname.

C.3 Interview Guide (English)

The following translation was generated by feeding the existing Dutch interview guide latex file to ChatGTP (OpenAI, 2023) and manually improving upon the result.

Hello, today I will conduct an experiment with you as part of my Master's thesis research. Before we start, it is important that I have your consent for the research. Therefore, I will explain the research, what we will do with the data, and the research ethics. Then, I will ask you some questions to ensure that I have your full consent before conducting the research. The expected duration of this experiment is 60 minutes. Please note that you can withdraw your consent at any time during the experiment, and you can stop participating at any time without any further explanation. To avoid influencing you too much, I will not yet provide detailed information about the research except that it concerns Twitter. During the experiment, you will learn more, and at the end, I can answer any questions you may have about the research in complete transparency. Furthermore, I do not expect any risks or discomfort to you as a participant in this research. The data I will collect will consist of an audio recording of your participation, which I will transcribe into text. The data collected is expected to be relevant for identifying factors that influence the experiment or contribute to answering my research questions. The data from this research will be used to conduct a qualitative analysis of my experiment. The data collected will be anonymized and processed and stored in accordance with the requirements of the Radboud University. Furthermore, the audio recordings will be destroyed within 6 months of the completion of the study, and the anonymized transcripts will be kept for 10 years.

- Do you have any questions about what I just explained?
- Have I provided you with sufficient information about my research and what will be done with your data?
- Do you understand that you can withdraw your consent at any time during the experiment, and that you can stop participating at any time until the end of the experiment?
- Are you over 18 and currently of sound mind?
- Would you still like to participate in my research?

The research will begin with some questions about your background that may have an impact on the experiment. Then, I will show you 3 rounds of tweets, each followed by a number of prepared questions. Depending on your answers, I may have additional follow-up questions. Each round of tweets contains 4 tweets. While you are looking at a tweet, I would like you to tell me out loud in as much detail as possible what you see and what you think about it. The focus of the research is mainly on the insights and experiences that you can share with me regarding what I show you, so please expand as much as possible on everything we will be viewing and what I will be asking you.

First, I have some questions about your background that may have an impact on the experiment. I will start the recording now.

Pr1: What is your gender?

Pr2: What is your age?

Pr3: What is the highest level of education you have completed?

Pr4: In which fields of knowledge do you have the most expertise?

Pr5: On a scale of 1 to 5, what grade would you give yourself for digital literacy?

Pr6: What sources of media do you consume?

Pr7: How often do you use Twitter?

Pr8: How do you decide whether to like, retweet, share, or comment on a tweet?

We will now look at the first 4 tweets. So, I would like to ask you to share out loud what you see and what you think about it. If you would like a tweet, retweet, share, or comment on it, please let me know and explain why. Afterwards, I have some questions based on the tweets, and we can look at them again.

Po1.1: What are your general thoughts on the tweets I just showed you?

As you may have noticed, there were different verification methods used for the tweets.

Po1.2: What does it mean to be verified on social media, specifically on Twitter, according to you? Here, I am referring specifically to the verification system used by Twitter itself.

Po1.3: In your opinion, what is the purpose of verification?

Po1.4: What are your general thoughts on the different verification methods I just showed you?

Po1.5: What influence did the different methods have on your willingness to like, retweet, share, or respond to a post?

Po1.6: Which verification system do you prefer and why? Which system do you not prefer and why?

(a) What made [preferred method] better than [less preferred method]?

(b) Why is [mentioned reason] important?

Next, I have a number of questions about your understanding of the two different concepts of authenticity and credibility, followed by how you found that to be applicable to the tweets I showed you.

Po1.7: What does authenticity mean to you, especially in a social media post?

Po1.8: How do you determine if a post is authentic?

Po1.9: What is the importance of authenticity to you?

Po1.10: How did the different methods influence how authentic a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po1.11: What does it mean to you to make a post credible?

Po1.12: How do you determine the credibility of a post?

Po1.13: What is the importance of credibility to you?

Po1.14: How did the different methods influence how credible a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po1.15: Is there anything else you would like to add?

You will now look at the second round of tweets. This will be similar to the previous round. I would like you to share out loud what you see and what you think. If you would like to like, retweet, share, or comment on a tweet, please let me know why. Additionally, I am curious about whether you think a tweet is authentic and the credibility of the information and why. You can use the definition that a post is authentic if you believe it was actually shared by the person who owns the account. Credibility is the level of trust you have that the shared information is accurate. Afterward, I have some questions based

on the tweets, and we can look at the tweets again. If you don't have any questions about it, we'll move on to looking at the tweets.

Po2.1: What are your general thoughts on the tweets I just showed you?

Po2.2: What are your general thoughts on the different verification methods I just showed you?

Po2.3: What impact did the different methods have on your willingness to like, retweet, share, or comment on a post?

Po2.4: Which verification system do you prefer and why? Which system do you not prefer and why?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po2.5: How did the different methods affect how authentic a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po2.6: How did the different methods affect how credible a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po2.7: Is there anything else you would like to add?

We will now look at the final round of tweets. Before we do that, I want to give you some more explicit explanations of the different verification methods. To illustrate this, I will include some tweets from the previous round. Some tweets are not verified in any way. These are tweets that you and I could make from an account created with that name, as long as it does not violate the user terms and conditions. Some tweets are verified using Twitter's verification method. This is only available for "notable" accounts. The criteria for this include being a politician, famous, or contributing to public discussion. To be verified in this way, you must provide Twitter with proof that you are indeed the person claiming to own the account, for example by sending a copy of your ID. Once verified, the account will be verified forever. The last method is called Twid, and is being developed at Radboud University. It currently works through a plugin that people can install on their browsers, but could in principle also be offered through Twitter or other social media platforms. With Twid, as a user, you can see the signatures of other Twid users. With this method, individual tweets are signed with an attribute that is linked to you. This signature requires extra authentication via a separate app called IRMA. Without going into detail, I can tell you that it is not possible to have an attribute that you are not entitled to, and that all attributes are stored only for you. If someone signs that they are a doctor, you can assume that they really are, and there is no external organization keeping track of the doctor's registration. An IRMA account is always linked to 1 Twitter account, so another doctor cannot sign for his colleague without all his login details. Also, in each round, there is a tweet from someone who uses Twid but has not signed it. Do you have any questions about this explanation? I would like to ask you to share aloud what you see and what you think about it. If you would like to like, retweet, share or comment on a tweet, please let me know and why. Furthermore, I am also interested in whether you think the tweets are authentic and what the credibility of the information is and why. Afterwards, I have some questions based on the tweets, and we can look at the tweets again. If you have no questions about this, we will now look at the tweets.

Po3.1: What are your general thoughts on the tweets I just showed you?

Po3.2: What are your general thoughts on the different verification methods I just showed you?

Po3.3: What influence did the different methods have on your willingness to like, retweet, share, or respond to a post?

Po3.4: Which verification system do you prefer and why? Which system do you not prefer and why?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po3.5: How did the different methods affect how authentic a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po3.6: How did the different methods affect how credible a post felt?

(a) What contributed to [preferred method] being better than [less preferred method]?

(b) Why is [mentioned reason] important?

Po3.7: Is there anything else you would like to add?

Finally, a few questions about your thoughts on Twid.

Po4.1: What did you think of Twid?

Po4.2: Would you use this plugin? Why?


Po4.3: Would you like to change anything about Twid to improve it? If so, what?

Po4.4: Is there anything else you would like to add about Twid?

Po4.5: Do you have any feedback on this entire experiment?

As you may have already noticed, the goal of the research is to map users' perception of different verification methods. In addition to your general perceptions, I was also interested in how it would affect your interactions with the posts and whether it had an impact on how authentic and credible you found the posts. I would like to mention that all the tweets I used in the study were constructed by me to make them as real as possible. Furthermore, some of the tweets contain misinformation, and if you wish, I can point out which ones specifically. Otherwise, I advise you not to simply believe what you read during this study. Do you have any further questions for me at this time? Do I still have your permission to process the data I obtained in my research? Then I will stop the recording. Thank you for your participation.



Robin Q. te Bruggen  @RobinQ · 5h

Net werd bekend dat een zwakte die stilzwijgend gefixt is in de populaire library Node.js er voor zorgde dat er jarenlang miljoenen gegevens gestolen konden worden.

 52  292  755 



Emma Laatbloei @Dokter_Emma · 11h

Rauw water. Sinds kort ook in Nederland te koop. Het komt direct uit de natuur, zonder industriële verwerking, en is een stuk gezonder.

 58  299  503 

Einde tweede ronde

Derde ronde

- Hardop in zoveel mogelijk detail vertellen wat u voor u ziet
- Deel uw gedachtegang
- Vertel ook of u zou liken, retweeten, sharen of commenten
- Waarom is belangrijk
- Deel ook of en waarom het authentiek is
- Deel in hoeverre en waarom het geloofwaardig is



Daan Blom @Daan1337 · 16h

Een vrouw haar tijd vooruit: Ada Lovelace was de eerste computer programmeur ter wereld, decennia voor de eerste computer gebouwd zou worden.

 54  284  740 



ir. Marloes Bakker @ir.Bakker · 5h

Ondanks eerdere angst heeft het Nationaal Cyber Security Centrum tot op heden geen aan de oorlog gerelateerde digitale aanvallen op Nederlandse belangen waargenomen.

 Niet getekend

 67  184  748 



Truus Kort  @Truus73 · 2h

Enkel het opstappen van Rutte zal leiden tot de verbetering die de Nederlander wil.

 74  221  660 



Yigit Demir @YigitDemir · 7h







De aanwezigheid van proanthocyanidine zorgt ervoor dat cranberrysap een blaasontsteking kan genezen.

 Getekend door medewerker Dieet- en voedingsadviesbureau Norma

 98  151  761 

Einde derde ronde

C.5 Study variant 2

<h1>Begin</h1>	<p>Eerste ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk
<p> Emma Laabloei @Dokter_Emma · 11h</p> <p>Rauw water. Sinds kort ook in Nederland te koop. Het komt direct uit de natuur, zonder industriële verwerking, en is een stuk gezonder.</p> <p>58 299 503</p>	<p> Truus Kort @Truus73 · 2h</p> <p>Enkel het opstappen van Rutte zal leiden tot de verbetering die de Nederlander wil.</p> <p>Getekend door medewerker Tweede Kamer</p> <p>74 221 660</p>
<p> ir. Marloes Bakker @ir.Bakker · 5h</p> <p>Ondanks eerdere angst heeft het Nationaal Cyber Security Centrum tot op heden geen aan de oorlog gerelateerde digitale aanvallen op Nederlandse belangen waargenomen.</p> <p>67 184 748</p>	<p> Irene Pardoos @Irene1989 · 7h</p> <p>Een quota voor vrouwen binnen de ICT zou de sector veel goed doen.</p> <p>Niet getekend</p> <p>90 235 716</p>
<h1>Einde eerste ronde</h1>	<p>Tweede ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk• Deel ook of en waarom het authentiek is• Deel in hoeverre en waarom het geloofwaardig is
<p> Maxime Hendrixks @MHendrixks · 14h</p> <p>Het zwaarder straffen van zedendelinquenten is niet een slimme oplossing om recidivisme te voorkomen.</p> <p>63 162 586</p>	<p> Yigit Demir @YigitDemir · 7h</p> <p>De aanwezigheid van proanthocyanidine zorgt ervoor dat cranberrysap een blaasontsteking kan genezen.</p> <p>Niet getekend</p> <p>98 151 761</p>



Daniëlle Wobstra @VoedselcentDaniëlle · 14h

Als men niet meer fruit gaat eten verwacht ik binnenkort veel meer mensen met overgewicht.

Getekend door medewerker Voedselcentrum

80 269 705



Daan Blom @Daan1337 · 16h

Een vrouw haar tijd vooruit: Ada Lovelace was de eerste computer programmeur ter wereld, decennia voor de eerste computer gebouwd zou worden.

54 284 740

Einde tweede ronde

Derde ronde

- Hardop in zoveel mogelijk detail vertellen wat u voor u ziet
- Deel uw gedachtegang
- Vertel ook of u zou liken, retweeten, sharen of commenten
- Waarom is belangrijk
- Deel ook of en waarom het authentiek is
- Deel in hoeverre en waarom het geloofwaardig is



Friso Veringa @Friso_MD · 1h

Chia zaad is een betere bron van Omega-3 dan enige vis, en ook nog eens helemaal vegan.

100 276 714



Robin Q. te Bruggen @RobinQ · 5h

Net werd bekend dat een zwakte die stilzwijgend gefixt is in de populaire library Node.js er voor zorgde dat er jarenlang miljoenen gegevens gestolen konden worden.

Getekend door Master of Science in Software Science

52 292 755



Jan-Kees Overmeer @Kamerlid_overmeer · 12h

Uit een recente poll blijkt een meerderheid van ondervraagden terug naar de gulden te willen. Waarom gaan we nog door met de euro?

99 253 774



Ton Heijdem @Ton_NieuwigNieuws · 3h







Ouders verwickeld in de toeslagenaffaire hebben bijna 1700 uithuisplaatsingen van hun kroost meegemaakt, leed dat geen mens gund is.

Niet getekend

91 262 535

Einde derde ronde

C.6 Study variant 3

<h1>Begin</h1>	<p>Eerste ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk
<p> Ton Heijema @Ton_NieuwigNieuws · 3h</p> <p>Ouders verwikkeld in de toeslagenaffaire hebben bijna 1700 uithuisplaatsingen van hun kroost meegemaakt, leed dat geen mens gegend is.</p> <p>91 262 535</p>	<p> Friso Veringa @Friso_MD · 1h</p> <p>Chia zaad is een betere bron van Omega-3 dan enige vis, en ook nog eens helemaal vegan.</p> <p>Getekend door geregistreerd arts</p> <p>100 276 714</p>
<p> Yigit Demir @YigitDemir · 7h</p> <p>De aanwezigheid van proanthocyanidine zorgt ervoor dat cranberrysap een blaasontsteking kan genezen.</p> <p>98 151 761</p>	<p> Robin Q. te Bruggen @RobinQ · 5h</p> <p>Net werd bekend dat een zwakte die stilzwijgend gefixt is in de populaire library Node.js er voor zorgde dat er jarenlang miljoenen gegevens gestolen konden worden.</p> <p>Niet getekend</p> <p>52 292 755</p>
<h1>Einde eerste ronde</h1>	<p>Tweede ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk• Deel ook of en waarom het authentiek is• Deel in hoeverre en waarom het geloofwaardig is
<p> ir. Marloes Bakker @ir.bakker · 5h</p> <p>Ondanks eerdere angst heeft het Nationaal Cyber Security Centrum tot op heden geen aan de oorlog gerelateerde digitale aanvallen op Nederlandse belangen waargenomen.</p> <p>67 184 748</p>	<p> Jan-Kees Overmeer @Kamerlid_overmeer · 12h</p> <p>Uit een recente poll blijkt een meerderheid van ondervraagden terug naar de gulden te willen. Waarom gaan we nog door met de euro?</p> <p>Getekend door medewerker Tweede Kamer</p> <p>99 253 774</p>



Irene Pardoos @Irene1989 · 7h

Een quota voor vrouwen binnen de ICT zou de sector veel goed doen.

...

90 235 716



Truus Kort @Truus73 · 2h

Enkel het opstappen van Rutte zal leiden tot de verbetering die de Nederlander wil.

...

Niet getekend

74 221 660

Einde tweede ronde

Derde ronde

- Hardop in zoveel mogelijk detail vertellen wat u voor u ziet
- Deel uw gedachtegang
- Vertel ook of u zou liken, retweeten, sharen of commenten
- Waarom is belangrijk
- Deel ook of en waarom het authentiek is
- Deel in hoeverre en waarom het geloofwaardig is



Daan Blom @Daan1337 · 16h

Een vrouw haar tijd vooruit: Ada Lovelace was de eerste computer programmeur ter wereld, decennia voor de eerste computer gebouwd zou worden.

...

Getekend door Master of Science in Data science

54 284 740



Maxime Hendrixks @MHendrixks · 14h

Het zwaarder straffen van zedendelinquenten is niet een slimme oplossing om recidivisme te voorkomen.

...

63 162 586



Daniëlle Wobstra @VoedselcentDaniëlle · 14h

Als men niet meer fruit gaat eten verwacht ik binnenkort veel meer mensen met overgewicht.

...

80 269 705



Emma Laatbloei @Dokter_Emma · 11h

Rauw water. Sinds kort ook in Nederland te koop. Het komt direct uit de natuur, zonder industriële verwerking, en is een stuk gezonder.







...

Niet getekend

58 299 503

Einde derde ronde

C.7 Study variant 4

<h1>Begin</h1>	<p>Eerste ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk
<p> Ton Heijdemā @Ton_NieuwigNieuws · 3h</p> <p>Ouders verwickeld in de toeslagenaffaire hebben bijna 1700 uithuisplaatsingen van hun kroost meegemaakt, leed dat geen mens gegend is.</p> <p>91 262 535</p>	<p> Yigit Demir @YigitDemir · 7h</p> <p>De aanwezigheid van proanthocyanidine zorgt ervoor dat cranberrysap een blaasontsteking kan genezen.</p> <p>98 151 761</p>
<p> Daan Blom @Daan1337 · 16h</p> <p>Een vrouw haar tijd vooruit: Ada Lovelace was de eerste computer programmeur ter wereld, decennia voor de eerste computer gebouwd zou worden.</p> <p>Niet getekend</p> <p>54 284 740</p>	<p> Emma Laatbloei @Dokter_Emma · 11h</p> <p>Rauw water. Sinds kort ook in Nederland te koop. Het komt direct uit de natuur, zonder industriële verwerking, en is een stuk gezonder.</p> <p>Getekend door geregistreerd arts</p> <p>58 299 503</p>
<h1>Einde eerste ronde</h1>	<p>Tweede ronde</p> <ul style="list-style-type: none">• Hardop in zoveel mogelijk detail vertellen wat u voor u ziet• Deel uw gedachtegang• Vertel ook of u zou liken, retweeten, sharen of commenten• Waarom is belangrijk• Deel ook of en waarom het authentiek is• Deel in hoeverre en waarom het geloofwaardig is
<p> Robin Q. te Bruggen @RobinQ · 5h</p> <p>Net werd bekend dat een zwakte die stilzwijgend gefixt is in de populaire library Node.js er voor zorgde dat er jarenlang miljoenen gegevens gestolen konden worden.</p> <p>52 292 755</p>	<p> Friso Veringa @Friso_MD · 1h</p> <p>Chia zaad is een betere bron van Omega-3 dan enige vis, en ook nog eens helemaal vegan.</p> <p>Niet getekend</p> <p>100 276 714</p>



Daniëlle Wobstra ✓ @VoedselcentDaniëlle · 14h

Als men niet meer fruit gaat eten verwacht ik binnenkort veel meer mensen met overgewicht.

...

80 269 705



Maxime Hendrixks @MHendrixks · 14h

Het zwaarder straffen van zedendelinquenten is niet een slimme oplossing om recidivisme te voorkomen.

...

Getekend door Master of Science Criminaliteit en Rechtshandaving

63 162 586

Einde tweede ronde

Derde ronde

- Hardop in zoveel mogelijk detail vertellen wat u voor u ziet
- Deel uw gedachtegang
- Vertel ook of u zou liken, retweeten, sharen of commenten
- Waarom is belangrijk
- Deel ook of en waarom het authentiek is
- Deel in hoeverre en waarom het geloofwaardig is



ir. Marloes Bakker @ir.Bakker · 5h

Ondanks eerdere angst heeft het Nationaal Cyber Security Centrum tot op heden geen aan de oorlog gerelateerde digitale aanvallen op Nederlandse belangen waargenomen.

...

Getekend door medewerker ICT Nieuws

67 184 748



Irene Pardoos ✓ @Irene1989 · 7h

Een quota voor vrouwen binnen de ICT zou de sector veel goed doen.

...

90 235 716



Truus Kort @Truus73 · 2h

Enkel het opstappen van Rutte zal leiden tot de verbetering die de Nederlander wil.

...

74 221 660



Jan-Kees Overmeer @Kamerlid_overmeer · 12h

Uit een recente poll blijkt een meerderheid van ondervraagden terug naar de gulden te willen. Waarom gaan we nog door met de euro?

...

Niet getekend

99 253 774

Einde derde ronde