FACULTEIT POLITIEKE EN SOCIALE WETENSCHAPPEN

The Algorithm Paradox: knowledge of versus attitudes towards curation algorithms

An Instagram case study experiment

Wetenschappelijk artikel
Aantal woorden: 9925

Thibault Fouquaert
Stamnummer: 01710179

Promotor: dr. Peter Mechant
Commissaris: Mathias Van Compernolle

Masterproef voorgelegd voor het behalen van de graad master in de richting Communicatiewetenschappen afstudeerrichting Nieuwe Media en Maatschappij

Academiejaar: 2018 – 2019

UNIVERSITEIT GENT

Abstract
De toenemende prevalentie van selectie algoritmes op allerlei sociale netwerksites (SNSs) brengt problemen met zich mee zoals een beperktere nieuwsdiversiteit en *filter bubbles*. Ondanks de alomtegenwoordigheid van selectie algoritmes, zijn ze vaak moeilijk waarneembaar en ook niet evident om te begrijpen voor doorsnee gebruikers aangezien SNSs weinig inzicht gegeven in de feitelijke werking van deze algoritmes. Om deze problemen aan te pakken stelt dit artikel het concept van 'de algoritme paradox' voor, waarin wordt aangenomen dat de attitudes en kritische bezorgdheid van gebruikers niet in lijn is met hun online SNS-gebruik vanwege de onwetendheid over de impact en werking van deze selectie algoritmes. Om deze tegenstrijdigheid tussen attitudes en gedrag te onderzoeken ontwikkelden we Instawareness, een online interface met als doel de onwetendheid over het Instagram selectie algoritme te verminderen. Deze tool werd getest in een quasi-experiment, wat onze voorgestelde algoritme paradox empirisch bevestigde. Een eerste bevinding is dat cognitieve mediawijsheid omtrent selectie algoritmes op SNSs niet essentieel blijkt, maar dat loutere bewustwording van deze algoritmes voldoende is voor mensen om verhoogde kritische bezorgdheid ten aanzien van SNSs te rapporteren. Een verandering in gebruiksgewoonten naar aanleiding van deze bekommernissen bleef, zoals verondersteld, echter uit. Daarnaast blijken visuele feedback tools (e.g., Instawareness) efficiënt te zijn in het verhogen van cognitieve mediawijsheid en stimuleren ze zo ook indirect de kritische attitudes ten aanzien van sociale netwerken. Hiermee levert deze paper een unieke bijdrage omtrent hoe indirect om te gaan met *filter bubbles* en kan het ondersteuning bieden bij verdere ontwikkelingen van het huidige beleid inzake mediawijsheid.

The Algorithm Paradox: knowledge of versus attitudes towards curation algorithms

An Instagram case study experiment

Thibault Fouquaert[*]

[*] Communication Sciences New Media & Society, Ghent University, Ghent, Belgium

ARTICLE HISTORY
Compiled 14th August 2019

Abstract
The increasing prevalence of curation algorithms on everyday social network sites (SNSs) comes with issues such as a decrease in news diversity and filter bubbles. Despite curation algorithms' popularity, they often are imperceptible and difficult for user to become knowledgeable about since little insights in the actual working of these algorithms are given. To address these issues, this paper proposes the concept of "the algorithm paradox" in which it assumes that users' attitudes and critical concerns are contradictory to their beliefs and habits due to a lack of cognitive understanding about curation algorithms. To examine this discrepancy between concerns and behaviour, we developed Instawareness, an online interface as visual feedback tool aiming to decrease the ignorance about the Instagram curation algorithm. Instawareness was tested in a quasi-experiment, which confirmed our proposed algorithm paradox. However, it is not cognitive understanding about SNSs' algorithms but solely awareness about them that appears to be sufficient for people in order to indicate increased critical concerns towards SNSs. As presumed, these people did not indicate any change in their habits to act accordingly upon their concerns. Furthermore, visual feedback tools prove to be efficient in increasing cognitive media literacy and indirectly stimulate critical concerns towards SNSs. Overall, this provides a unique contribution towards how to indirectly cope with filter bubbles and assist in the adjustment and further development of currently deployed policies on media literacy.

KEYWORDS
Algorithms, Algorithm Awareness, News Feeds, Media Literacy, Paradox, Social Network Sites, Instagram

CONTACT Thibault Fouquaert. Email: thibault.fouquaert@ugent.be

1. Introduction

In the past decades, the theory of gatekeeping has been a touchstone for research in communication sciences. With the emergence of the Internet and new technologies such as social network sites (SNSs), gatekeeping continues to exist although new players entered the field (Roberts, 2005). One of these new entrants are curation algorithms, which arrange online delivered content by prioritising, classifying and filtering information. This new instrument as a means to gatekeeping is shaped by many actors (i.e. developers, engineers, end users, regulation, industry) and by the datasets and digital traces left behind from everyday use (Bucher, 2012). Particularly the habits and data generated by SNSs' users is becoming an important source of input for these kind of algorithms, which gatekeeping theory addresses in the changing an increasing role of end users as a sort of gatekeepers for each other, whereas journalist are becoming rather gatewatchers than gatekeepers (Bruns, 2008). Existing theory, however, treats SNSs services often mistakenly as solely an algorithm thereby creating the incorrect impression that algorithms operate objectively and free of (de-)selection (Beer, 2017; Bozdag, 2013; Tandoc Jr, 2014). Similar to pre-digital gatekeeping, this could raise problems such as distortion of reality. A well known manifestation of this distortion is the *filter bubble*, where users are presented with mostly content which interests them and fits their discourse or way of thinking (Pariser, 2011).

Curation algorithms are one mechanism which contribute to the monoculture of a filter bubble. In fact, their contribution towards the information diversity of our digital society might be remarkably higher than expected, since these curation algorithms are the backbone of many news feeds in everyday SNSs. Instagram, for example, makes use of such algorithms to personalise the news feed of each individual user based on data resulting from one's online activities, thereby 'showing the moments one cares about the most' (Cotter, 2019; Instagram, 2016).

The Instagram news feed is used by over 500 million daily active users as of 2019, while the *#metoo* hashtag was used over 1.5 million times in 2018 (Instagram, 2019). In Flanders, the platform is becoming increasingly popular in younger segments with

respectively a 73% and 60% active monthly usage amongst the 16-24 and 25-34 year old segment. Even 70% of them even claim to 'stay up to date with what's going on through social media'(Vanhaelewyn & De Marez, 2018, p. 56). Moreover, Instagram was marked as the fastest growing SNS in the past year with a remarkable increase in monthly usage of 12 percentage points amongst the 25-34 age group (Vanhaelewyn & De Marez, 2018).

The increasing popularity of platforms that use opaque curation algorithms along with the potential consequences of filter bubbles might also raise question about the public awareness of such algorithms. Recent studies argued that users' literacy about curation algorithms and experiences with them, might affect their attitudes towards how the platform should be used (Beer, 2017). As Boyd (2014) points out in her work, for instance, teens attempt to manipulate the technology (i.e. the algorithms at work) to attract more attention from their peers by sharing relevant content, even if the subject might be more hurtful than enriching. With personal or financial gain at stake, people find creative ways in attempting to manipulate the hidden algorithms and increase the visibility of their posts (Boyd, 2014; Bucher, 2012; Rader & Gray, 2015). Despite the common deployment of curation algorithms on SNSs, only few SNSs offer insights into their algorithms' outcomes. With no possibility of such feedback, it can be difficult for users to become knowledgeable about these curation algorithms, assess the personal news feed from a critical angle and change attitudes accordingly. Yet, it is this feedback mechanism that might ultimately be required to prevent the potential negative effects of algorithmic selection and curation such as filter bubbles, as argued by Just and Latzer (2017): 'Algorithmic selection shapes the construction of individuals' realities, that is, individual consciousness, and as a result affects culture, knowledge, norms, and values of societies, that is, collective consciousness, thereby shaping social order in modern societies. [...] the design of these institutions needs democratic legitimation [...]' (p. 246).

To tackle these issues, this paper sheds light on the awareness of curation algorithms used on SNSs. The aims is twofold: first, to examine the relationship between media literacy and attitudes towards curated news feeds; second, to assess if and how *Inst-*

awareness[1], a self-developed visual feedback tool, increases awareness and media literacy about curation algorithms. Instawareness (see section 3) allows people to log in with their personal Instagram account, whereafter it extracts information from their news feed in order to reveal the mechanisms behind the hidden algorithm. Using this tool, people are offered a side-by-side comparison of their news feed with and without curation algorithm, as well as other insights such as highest or lowest ranked friends and 'hidden' posts.

In what follows, this paper starts with outlining relevant literature and conceptualisations in section 2 along with the central research question and hypotheses derived from this literature review. In section 3, the study design is explained and its results are discussed in section 4. The conclusions along with debating points are presented in section 5.

2. Literature and conceptualisations

A wide range of research has already investigated various kinds of invisible dimensions when interacting with technology. Human-Computer-Interaction (HCI) in cognitive science research, for example, studies the mental models people develop while perceiving and interacting with different technologies (Mantovani, 1996). Other fields of studies, specifically concerning SNSs, have examined invisible components related to daily life such as imagined audiences. On SNSs, the audience is not spatially present or bounded, nor is there any participation. This seems a difficulty in gaining a full understanding of this imagined audience, leading to users who are often unaware of their entire connected network (Boyd, 2014; Tufekci, 2008). Marwick and Boyd (2011) elaborated on this idea and its impact on daily life under the name of — also invisible — networked audiences in our networked society, while Bernstein, Bakshy, Burke and Karrer (2013) claimed that '[...] social media users consistently underestimate their audience size for their posts, guessing that their audience is just 27% of its true size.' (p. 21). Moreover, the latter considered curation algorithms, in this case Facebook's *EdgeRank* algorithm, as one of the reasons why users have lower estimates on their audience size. Others, then

[1] https://instawareness.ugent.be, the name is a concatenation of Instagram and awareness

again, have taken a closer look at the role of a new but hidden curator (i.e. 'algorithms designed by site maintainers') in our self-presentation on SNSs (Hogan, 2010, p. 381).

2.1. Perspectives on algorithmic selection and curation

To assess the role and impact of technology on society, not to mention the development of recommendations for policy makers, an analytical lens must be chosen. This paper adopts the lens of co-evolutionary theory coupled with an institutional view on technology, which could be seen as a specific approach to the mutual shaping of technology. By that, it refutes the more technological deterministic approaches (i.e. technology as a driver and direct cause of behaviour) while not fully committing to the more social deterministic approaches (i.e. technology is mainly shaped by social forces).

According to co-evolutionary theory, technological innovation processes are evolutionary processes (Nelson & Winter, 2002). Although not fully comparable with biological evolution, technological evolution has similarities to it. Both systems have a kind of collective behaviour (i.e. values and norms, adoption) which are in fact simple interdependent rules at a more individual level (i.e. actions, implementations), characterised by the skill to adapt via trial-and error learning. In this way, co-evolution is the process where technology is simultaneously designing and being designed. It focuses on the interaction and reciprocity between social forces such as technical, economic, political, cultural and end-users processes (Just & Latzer, 2017).

Besides the shaping of and simultaneously being shaped by social forces, another eminent aspect being applied to co-evolutionary theory in this paper, is that it acts as a structure for users to negotiate with. That is, making the structure coupled with the 'rules' described therein into an instrument for achieving certain goals. This is a more institutional perspective being combined with the co-evolutionary approach (Napoli, 2014), and helps to better reflect upon the governing role of algorithms, as seen in other studies (Cotter, 2019; Just & Latzer, 2017). Accordingly, algorithms can be seen as the enabling technologies which act as governance mechanisms and could have unforeseen effects on social issues.

This combined approach of co-evolutionary and instrumentality is also being used by Cotter (2019) in her work on 'playing the visibility game on Instagram'. The adopted perspective on 'playing the game' acknowledges the authority of SNSs' owners to set constraints on how the SNS could be used, although not neglecting the autonomy of their users to interpret these limitations, and act upon them according to their affordances. This builds on the analogy of video games as objects of algorithmic culture by (Galloway, 2006, pp. 90–91): 'To play the game means to play the code of the game. To win means to know the system. And thus to interpret a game means to interpret its algorithm'. In line with this, the perspective throughout this paper is one where algorithms are more structural and instrumental elements to which users can adapt even if they do not know, nor understand, the complete 'rulebook'. As already mentioned in section 1, people tend to play a visibility game on everyday platforms to avoid 'the threat of invisibility' (Bucher, 2012), thereby consciously engaging with and interpreting the rules set by the SNSs.

2.2. Algorithms

At present, algorithms on SNSs have been discussed in a variety of ways. Research ranges from gatekeeping (Bozdag, 2013; Bucher, 2012; Introna & Nissenbaum, 2000; Tandoc Jr, 2014) to ways of knowing the (black box) algorithms (Bucher, 2016; Sandvig, Hamilton, Karahalios & Langbort, 2014). This paper' focus is on research about the perception, understanding and awareness of algorithms on SNSs, which several authors have already explained in great detail (for example, Bucher, 2017; Eslami et al., 2016; Eslami, Rickman et al., 2015; Eslami, Vaccaro, Karahalios & Hamilton, 2017; Hamilton, Karahalios, Sandvig & Eslami, 2014; Rader & Gray, 2015; Verdegem, Haspeslagh & Vanwynsberghe, 2014).

While the exact definition of '*algorithm*' is hard to provide, it can be mainly described as a finite set of precisely defined rules and processes to achieve a certain outcome. Algorithms take input and transforms this through its computational rules (throughput) into output (Cormen, Leiserson, Rivest & Stein, 2009). First, the input is characterised by a user request including available user characteristics. Second, this input gets pro-

cessed in the throughput stage so that relevance can be assigned to selected elements of a data set. Last, certain elements (i.e. often the most relevant) are selected to act as output (Latzer, Hollnbuchner, Just & Saurwein, 2016). Subsequently, and in line with Latzer et al. (2016), *algorithmic curation* (also frequently labelled as algorithmic selection) is conceptualised as the process where relevance is assigned to information elements of a data set by a computational assessment of its input, i.e. generated data signals. *algorithmic ranking* is then seen as a subset of algorithmic curation: a process that is free of any filtering that might hide content from the possible output list, but still performs practises such as ranking, aggregation, recommendation and more (for a complete functional typology, see Latzer et al., 2016, p. 6).

Curation algorithms' input can occur in multiple formats. Big data is often used as one of the possible inputs, although questions may arise on the use of big data in combination with algorithmic selection. Boyd and Crawford (2012), for instance, highlight six provocations on the use of big data in a social media context, where primarily objectivity, potential bias, erosion of context, practices and ethics of big data were disputed. Furthermore, Yeung (2017) addresses the characterising of big data throughput mechanisms (e.g., curation algorithms) as a form of *nudge*, which '[...] channel user choices in directions preferred by the choice architect through processes that are subtle, unobtrusive, yet extraordinarily powerful.' (p. 2). Despite the notable interdependence of big data and algorithmic curation, an analysis of the role and implications of big data in the chain of opaque selection mechanisms is beyond the scope of this paper.

To further define the conceptualization of selection algorithms, the taxonomy of analytics from Delen and Demirkan (2013) is used. This taxonomy has three categories based on algorithms' capabilities. (1) Descriptive analytics is used to describe data, identify opportunity or outcomes, and answers the question 'what happened and/or what is happening?'. (2) Predictive analytics is used to discover patterns which might explain input-output relationships and answers the question 'what will happen and/or why will it happen?'. (3) Prescriptive analytics is used to determine a set of best course actions for a given situation, from which one of these could then be used autonomously. It also answers the question 'What should I do and/or why should I do it?'. Even though this

taxonomy was designed for business analytics, we are convinced it can support in the further outlining of curation algorithms' characteristics. Therefore, curation algorithms are seen as prescriptive algorithms throughout this paper.

2.2.1. Work on revealing hidden selection algorithms

An eminent study on understanding users' awareness of curation algorithms is the work by Eslami, Rickman et al. (2015). These researchers developed *FeedVis*, a visual feedback tool which allowed them to present 40 participants a side-by-side comparison of their curated vs. uncurated (i.e. all possible content in reverse chronological orderer) Facebook news feed, who were followed up a few months later. Major findings were that 62.5% of participants were unaware of the curation algorithm and initially developed negative attitudes with consequences such as feelings of betrayal or doubts about real life relationships because of missed posts. Over time, some participants started to manipulate the algorithm with newly developed habits including an increasing use of the 'most recent' view, setting goals as to who appears, liking friends' posts and more. Participants also started to developed more positive feelings over time since they became more knowledgable about the algorithm. For example, participants were more satisfied, understood its importance in order to receive a more relevant feed, or realised that a post with little attention (i.e. likes or comments) might be because of an algorithm instead of due to their friends. Overall, these findings were the basis for *Fiddler*, a newly developed visualisation interface by these researchers (Zhang, 2015).

By contrast, Rader and Gray (2015) found that only 22% of their total sample (n=464) was in fact unaware of the Facebook curation algorithm. These contradictory findings are likely due to the recruitment of Rader and Gray, since they aimed for a generally more aware population who thought about curation before. More interestingly is the wide range of users' beliefs on curation and the effect on their attitudes and habits. For example, there was the trend of passive and uncritical consumption at which users did not reflect all too often about why they see the post they do. Moreover, more than half of these respondents were aware of the algorithm while exerting this uncritical behaviour, indicating one might not have thought about possible issues and side-effects

such as filter bubbles or imagined audiences (see section 1). Other common beliefs were that (1) the curation helped them by displaying what they wanted to see, (2) they are missing posts because of the curation and (3) personal behaviour was used together with factors such as popularity to prioritise posts. Although this study could not provide any feedback mechanism to understand the curation, it reported insightful findings similar to Eslami, Rickman et al. (2015) and Eslami et al. (2016).

2.3. Media Literacy

The term *media literacy* is used to address how knowledgeable one is about different aspects of their own media consumption. It originates from earlier research which focused on the digital divide, that is, a binary classification of physical access to computers. Further research recognised the limitations of this conceptualization as the gap between the 'haves' and the 'have-nots', arguing that more attention should be paid to socio-economic backgrounds and capabilities to effectively use these new technologies (van Deursen & van Dijk, 2011). As a result, a 'second-level digital divide' (Hargittai, 2002) revealed the emergence of widening gaps regarding digital skills which were also mainly unaddressed, unlike the gaps in pure digital access (van Dijk, 2005). Since the unequal divide of digital skills may lead to an exacerbation of existing societal inequalities (van Dijk, 2005), continued research focused on ways to define, measure and minimise the disparity of these digital skills.

Based on extensive prior research, Livingstone, Van Couvering and Thumin (2008) defined a more traditional approach towards media literacy in that it exists out of both (1) technical and (2) cognitive competencies. With regard to this distinction, Helsper and Eynon (2013) defined four broad skill categories: operational competencies including technical, creative skills and strategic competencies including social, critical skills. An overlapping definition based on a range of studies conducted in the Netherlands suggested and validated four similar types of skills: (1) Operational, (2) Formal, (3) Information and (4) Strategic (van Deursen, Courtois & van Dijk, 2014; van Deursen & van Dijk, 2009, 2010, 2011, 2015; van Deursen, van Dijk & Peters, 2012). The first two, operational and formal skills, account for technical or medium-related aspects whereas

the last two, information and strategic skills, account for cognitive or content-related aspects of media consumption. This distinction is thereby also in accordance with the earlier provided approach of technical versus cognitive competencies by Livingstone et al. (2008). An additional two skills, (5) content creation and (6) communication, were later added and validated which resulted in the current sixfold typology (van Deursen, Helsper & Eynon, 2016; van Dijk & van Deursen, 2014). Admittedly, this typology might be incomplete to fully conceptualise (new) media literacy since it has been argued to also include cultural competencies such as play, judgement, networking and negotiations (Jenkins, 2009). An overview of the aforementioned skills used in this paper is given in table 1.

Despite the distinct categorisation applied in this typology, most skills are in fact interdependent and have some overlap. For instance, technical media literacy is found to be required for performances on cognitive media literacy (van Deursen & van Dijk, 2015). In addition, van Deursen and van Dijk also found that an increase in technical media literacy resulted in a better performance on cognitive media literacy, especially among older people. van Deursen et al. (2016) this link between technical and cognitive media literacy, as well as the stronger link to be found for older age groups compared to younger groups.

On a final note, research conducted by Vanwynsberghe, Boudry and Verdegem (2015) that specifically focused on *social media literacy* argued that more traditional definitions of media literacy (e.g., the aforementioned typology) are only partly applicable to *social* media, due to the higher degree of participation required on SNSs. For this reason, social media literacy was defined as including the competencies to actively participate online requiring skills such as communicating and content creation (cf. supra), as well as the more traditional technical and cognitive competencies (cf. supra). This twofold definition is adopted in this paper whereby the sixfold typology of van Dijk and van Deursen was used to operationalise the concept of both technical and cognitive social media literacy.

Table 1. Conceptual definitions and operationalisation of the media literacy competencies typology as proposed by van Dijk and van Deursen (source and complete definitions: van Deursen, Courtois & van Dijk, 2014; van Deursen, Helsper & Eynon, 2016; van Deursen & van Dijk, 2009, 2011; van Dijk & van Deursen, 2014).

Technical media literacy medium related		***Cognitive media literacy***[a] ***content related***	
Operational	**Formal**	**Information**[a]	**Strategic**
Skills to operate digital media, 'button knowledge'	Skills to orient oneself within non-linear medium specific structures	Skills to find, select and evaluate sources of digital information	skills to use digital sources to reach a personal or professional goal
Download files	Navigate between menus	Deciding keywords	Orientation towards a particular goal
Open files	Following hyperlinks	Evaluating information sources	Taking the right actions to reach this goal
Using shortcuts	No disorientation when navigating	Check correctness of sources	Making the right decisions to reach this goal
Using bookmarks	Understanding the design flow	Examine not only top results	Gaining benefits resulting from this goal
Connect to Wi-Fi		Understanding filter mechanisms	

[a]The research presented in this paper puts more emphasis on the correct evaluation of information or 'the art of critical thinking' (van Deursen & van Dijk, 2009, p. 395) rather than the ability to find it. This is to avoid high factor similarity between formal and information skills as found by van Deursen, Helsper and Eynon (2016), who suggested this might be because of the fact that navigational issues primarily rise when looking for information. They handled this by combining formal and information skills into the factor *Information Navigation skills* after data collection.

2.3.1. Media literacy on curation algorithms

The general level of cognitive media literacy might be insufficient since a minority of studied Facebook users was aware of its curation algorithm (see section 2.2.1). Similarly, a longitudinal cross-sectional in the Netherlands revealed that information and strategic skills appeared to be low for most people, whereby the level of these skills virtually stagnated between 2010 and 2013 (van Deursen & van Dijk, 2015).

In Flanders, the cognitive social media literacy on selection algorithm is also rather low. Work from Verdegem et al. (2014) states that little is known about the existence of curation algorithms since 70% of their Facebook-using participants still think they were shown all possible online activities. This number even increases for Twitter, where 90% of the participants were unaware that not necessarily all tweets from people they follow are displayed. These results may confirm earlier findings about media literacy in Flanders which identified 59% as 'advanced users' yet admittedly addressing that these results only reflected the operational and formal skills (Paulussen, Courtois, Vanwynsberghe & Verdegem, 2011). The same fact was found by van Deursen and van Dijk (2011),

who indicated the absence of a correlation between Internet usage and cognitive skills. Given that these technical skills are required to acquire cognitive skills (van Deursen & van Dijk, 2015), it should be safe to assume that cognitive skills are indeed insufficient in Flanders. Nonetheless, it is to note that the Flemish government, as well as other governments such as the Dutch one, carries out a proactive policy for media literacy with several running projects to make both youth and elderly more media literate (Gatz, 2018).

H_1 The average cognitive media literacy is significantly higher for people who used the visual feedback tool Instawareness[2] compared to those who did not.[3]

2.4. Knowledge versus attitudes: the algorithm paradox

Based on cognitive dissonance theory, we might expect users to adapt their attitudes or expectations accordingly to their knowledge in order to reduce dissonance. Still, we often see this is not always the case. The information privacy paradox, for instance, is the phenomena where users claim to value their personal information while actual behaviours are contradictory (Norberg, Horne & Horne, 2007). This inconsistency has also been found between the attitudes of people towards information transparency features and the actual partake in personalisation (Awad & Krishnan, 2006). New research suggested, however, that the discrepancy between privacy attitudes and privacy behaviours should not be any longer regarded as a paradox, given that continued research provided several explanations. For example, perceived benefits of participation on SNSs (e.g., attention) seem to outweigh observed risks (Kokolakis, 2017).

Not only privacy but also algorithmic curation appears to have such discrepancy. As mentioned in section 2.3.1, only 30% of the questioned were aware of the selection algorithm behind Facebook. Despite this ignorance about curation algorithms, many respondents (83% - 56%, depending on the activity) actually claimed they would mind

[2]See section 3.
[3]Only alternative hypotheses are listed throughout this paper.

if SNSs carried out certain activities such as filtering or selling their personal or user data (Verdegem et al., 2014).

> In particular, users are unaware that their data is being sold, of the extent of information filtered out [due to algorithmic curation] before it is shown on their Twitter or Facebook feeds [...]. This lack of knowledge is closely related to users' attitudes; they claim to be bothered by the selling of data [and] filter algorithms [...] but their ignorance of whether these things actually occur means they fail to evaluate the social media sites as critically as may be required. (Verdegem et al., 2014, p. 29)

Put differently, users have a vague understanding of what curation algorithms are and hence their existence, although they claim to be bothered if algorithmic curation were to happen. These findings may be insightful when compared to previously identified contradictory relationships such as the privacy paradox. Similar to this privacy paradox, technical skills to tackle algorithmic curation appear to be present (see section 2.3.1). In this regard, one could assume that if people knew about the existence of curation algorithms, they would claim to be bothered by these algorithms while not acting accordingly, that is, actively altering the effects of the curation algorithm and critically assessing online information. Therefore, this paper proposes the concept '*The Algorithm Paradox*' to address the assumption of this contradictory relationship.

Building upon the concluding remarks of Kokolakis (2017), this algorithm paradox might similarly be a complex phenomenon rather than a paradox, thus having a logical explanation to no longer be regarded as a paradox. Based on all the preceding literature, one explanation might be that the cognitive skills to understand what algorithms are really able to do is a required asset to properly take subsequent actions, such as a more critical assessment of what is seen on SNSs. In fact, studies of Eslami, Rickman et al. (2015) and Eslami et al. (2016) revealed that once people understood what effect algorithms have on *their* news feed (i.e. increasing cognitive competencies specifically on algorithms), they acted upon this accordingly (more details in section 2.2.1). However, these cognitive skills seem to be insufficient fir now as already mentioned in section 2.3.1. Based

on this assumption, cognitive development of curation algorithms' actual working in relation towards attitudes is the subject matter of this paper.

H_2 The average general feelings are significantly higher for people who used the visual feedback tool Instawareness compared to those who did not.

H_3 The average critical concern is significantly higher for people who used the visual feedback tool Instawareness compared to those who did not.

RQ What is the relation between both technical and cognitive media literacy about selection algorithms on social network sites and attitudes towards social network sites for Flemish adults aged 18 to 35?[4]

2.5. An Instagram case study

An inspiring source for this paper was the case study on Facebook News Feed conducted by Eslami, Aleyasen, Karahalios, Hamilton and Sandvig (2015). This case study was assisted by the FeedVis tool, which revealed the EdgeRank algorithm to participants by extracting their news feed data. Likewise, we wanted to use such a visual feedback tool to explain the working of algorithms to participants. The use of FeedVis or Fiddler for this paper, however, was presumably unachievable since (1) it was not open source and (2) it was based on an API endpoint that was closed by Facebook as of October 2015 (Facebook, 2015). Moreover, Instagram's popularity in terms of usage is increasing in comparison to other SNSs' popularity, making it an interesting matter of subject (Vanhaelewyn & De Marez, 2018). For these reasons, and the fact that a visual feedback tool for Instagram was technically feasible (i.e. API endpoint access) while no researchers to our knowledge developed such a system, Instagram was chosen as a case study.

A final interesting remark is that, according to Instagram, content is never hidden but solely reordered based on its relevancy, given one keeps scrolling long enough (Constine, 2018). Therefore, the Instagram algorithm is a ranking algorithm as mentioned in section 2.2. Six factors are currently used to algorithmically determine the ranking

[4] Although mediation analysis is used in presenting the results, Hayes (2017, p. 119) argues against hypotheses subjected to finding full or partial mediations since neither are 'better' and it are 'empty' concepts that have no substantive or theoretical meaning. Therefore, no hypothesis but only a RQ was proposed to elucidate these findings.

of posts shown in the news feed each time the feed is loaded: (1) interest, (2) recency, (3) relationship, (4) frequency, (5) following and (6) usage (Constine, 2018). Although these factors were officially communicated by the Instagram product team, there is still little understanding on how these six factors get computed.

3. Study design

In order to examine our proposed assumption of the algorithm paradox, a twofold approach was undertaken. First, the visual feedback tool Instawareness was developed, which allowed people to log in with their personal Instagram account and subsequently reveal them the mechanisms behind the hidden algorithms. Second, a pre-post mixed-design quasi-experiment consisting of three phases was conducted to answer the aforementioned research question and hypotheses. In the first phase, the level of technical media literacy (TML), cognitive media literacy (CML) and attitudes (ATT) was measured for each participant using an online questionnaire in Qualtrics. Next, participants were randomly selected for phase two and asked to use Instawareness, preferably a few times per week. After two to three weeks, phase 3 began in which participants were asked to retake the online questionnaire which again measured their levels of media literacy and attitudes. Finally, these two sets of collected data, both existing of two separate groups, were combined and used for further data analysis. An overview of this pre-post setup is found in table 2 with an indication of which groups were used to test each hypothesis.

Table 2. Study design overview

Phase	**Group A (n=32)**	**Group B (n=32)**	**Union ($A + B$)**
Pre-test	Measure TML, CML and ATT		—
Intervention	Control group without placebo Follow-up: 14-18 days	Group using Instawareness Follow-up: 14-21 days	—
Post-test	Measure TML, CML and ATT		RQ (N=64)
Mixed-design[a]	H_1, H_2, H_3		—

[a]Test were run between-subject groups A and B using the within-subject pre-test measurements as covariates.

3.1. Pre-assessment

The quasi-experiment started by an initial measurement of three latent concepts which addressed participants' familiarity with Instagram including its mechanisms such as curation algorithms, as well as their attitudes towards Instagram. To make sure participants could answer these questions in a meaningful way, a weekly use of Instagram was set as a prerequisite to partake in the experiment. The actual active minutes spent daily on Instagram was on average 43($= M,\ SD = 20.16$) minutes with an interquartile range of 23 to 52 minutes ($n = 52$), which is based on the app's own indication[5].

First, the latent concept of technical media literacy was measured by looking at (1) frequency and (2) familiarity. Frequency encompasses the number of performed online activities and how frequently these are done. Although this proxy actually measures media use, it is still found to properly grasp actual technical media literacy. This is suggested by Vanwynsberghe and Haspeslagh (2014), whose research indicated that people who already accomplished most online activities were found to be higher technical media literate. Items used in the questionnaire were adopted from Vanwynsberghe and Haspeslagh (2014) and included questions about the frequency of practises such as 'creating a story', 'using hashtags', 'tagging people' and more. Familiarity, on the other hand, encompasses the understanding of various digital related terms and has found to be an even stronger predictor for technical media literacy than frequency or self-efficacy (Hargittai, 2005), but seems to be underused despite its positive outcomes (Vanwynsberghe & Haspeslagh, 2014). This scale was also adopted from Vanwynsberghe and Haspeslagh (2014) and included items about the familiarity with 'tagging', 'search function', 'unfollowing' and more. Overall, both the frequency and familiarity scale were found to have a good internal validity with a Cronbach's Alpha of respectively .79 and .85 after three items with a corrected item-total correlation less than .3 were removed.

Alternatively, self-efficacy (i.e. the self-evaluation of own knowledge and skills) is a frequently used proxy to measure technical media literacy. This is because it has been indicated that people with a higher belief in their own skills and knowledge are more

[5]This can be seen under 'profile > my activities'

likely to complete tasks online successfully (Livingstone & Helsper, 2010; van Deursen, 2010). Criticism, however, argues that the proxy of self-perceived skills might be an unreliable measure for actual media literacy, since it always is context-dependent. For example, whom users compare themselves with or how they are feeling when taking the questionnaire may alter one's perception of their competencies (Talja, 2005). For this reason, self-efficacy was excluded as a scale to measure technical media literacy.

Second, the latent concept of cognitive media literacy includes the measurement of (1) knowledge and (2) critical thinking. The latter differentiates itself from (critical) attitudes by measuring respondents' awareness — or rather ignorance — instead of concerns, general feelings and habits (Vanwynsberghe & Haspeslagh, 2014). Even though the trust measure of Hargittai, Fullerton, Menchen-Trevino and Thomas (2010) has been used to assess cognitive competencies, this is often limited to information searching and neglects trust in peers on SNSs despite the importance of this factor in a SNS context. Therefore, it has been argued that asking directly about users' knowledge and awareness is better than to measure users' (dis)trust in SNSs (Dwyer, Basu & Marsh, 2013). Based on these insights, Vanwynsberghe and Haspeslagh (2014) proposed a scale to measure cognitive media literacy which was adopted in this paper. The scale included items focused on filtering mechanisms and ranged from 'changing your feed based on the amount of people you follow' to 'thinking about how many people will see your post'. Overall, the scale was found to have good internal validity with a Cronbach's Alpha of .74 after one item with a corrected item-total correlation less than .3 was removed.

Last, the latent concept of attitudes overspans three measurements: (1) critical concern, (2) general feelings and (3) critical habits. Based on the research about emotional competencies from Vanwynsberghe and Haspeslagh (2014), respondents were asked about the extent to which they critically reflect upon and have concerns about actions the Instagram algorithm performs like 'altering your feed based on your usage data', 'keeping deleted data' and more. Attitudinal items (i.e. happy/annoyed, positive/negative, untrustworthy/trustworthy) as well as cognitive attitudinal items (opaque/transparent, beneficial/harmful) from Yang and Yoo (2004) were used to measure general feelings towards the SNS. According to the changes in habits found in the study of Eslami,

Rickman et al. (2015), critical habits was measured by asking respondents they had performed or will perform actions such as 'changing interaction with friends', 'assigning their number likes partly to an algorithm' and more. A Cronbach's Alpha of respectively .88 and .81 was found for the critical concern and general feelings scales after two items with a corrected item-total correlation less than .3 were removed. The critical habits scale had too low internal validity to be used as a continuous scale and was therefore further used as ordinal statements.

At the end of the questionnaire, participants answered some socio-demographic questions and were randomly divided into two groups: (1) group A, who received no placebo effect and acted as the control group; (2) group B, who were prompted to utilise Instawareness which unknowingly acted as a manipulation. As the actual use of Instawareness formed a kind of barrier to continue in the experiment, there was some dropout in group B which was accounted for by stratifying the random assignments for equal group size. In the upcoming weeks, participants were send personal emails as a stimulus to use Instawareness, and preferably use it more than once for those who already did. This resulted in a usage of Instawareness ranging from 1 to 5 times ($Median = 2$ times) for each participant of group B before they started the post-assessment.

3.2. Intervention: using Instawareness

Since SNS users are generally unaware of curation algorithms' capabilities, a generic explanation may still be too distant to achieve full understanding (Eslami, Rickman et al., 2015). In fact, Hamilton et al. (2014) suggested that one approach of studying the effects of algorithm awareness is to reveal the algorithms at work to user through an interface (i.e. what we call visual feedback tool). Therefore, we developed Instawareness as a means to personalise the explanation of curation algorithms' capabilities. Additionally, participants were also able to watch a video covering the key factors that contribute to its ranking (see section 2.5). To start, participants were asked to log in with their Instagram account by copying cookies from Instagram's website so that their personal data could be fetched. This data could then be analysed in order to visually reveal the

effects of Instagram's curation algorithm on *their* news feed with the help of a guided tour through the tool.

After the log in, our tool began to fetch the first 50 posts as ranked by the curation algorithm to be shown in the news feed (hereafter 'curated posts')[6]. Subsequently, our tool continued to fetch all posts that would have appeared before the least recent post from the top posts if no curation algorithm had been present (hereafter 'uncurated posts'). The main difference between these curated and uncurated posts lists is the sequence in which the posts appear in each list. This sequencing is given by the Instagram ranking algorithm (see section 2.2) and is based on its seven main factors (see section 2.5). Since this sequencing is personalised for each user and thus hard to provide general insights or metrics about — if any even exist —, we computed the Kendall Tau coefficient and amount of 'hidden' posts (see section 3.2.3) for each visitor of Instawareness in order to gain and share insights about the algorithm's dynamics. The average Kendall Tau (τ_b) was 0.48[7] and ranged from -0.19 to 0.98, which gives an indication that most users' news feeds are significantly reordered while largely still in line with the posts' recency (i.e. the positive correlation between the curated and uncurated reverse chronological posts lists). Furthermore, the amount of filtered out posts as a consequence of its lower ranking (i.e. hidden posts), which gives an indication of how substantially posts get reordered, had great variability as can be seen in figure 6 and ranged from zero hidden posts for some, to a few dozen for most, up to hundreds for others ($M_{\text{hiddenPosts}} = 119.36,\ SD = 151.15,\ Median = 43,\ n = 83$).

One final note is that this fetching logic is used under the premise that a news feed with no curation algorithm is a reverse chronological news feed, as was the case with earlier versions of Instagram's news feed (Constine, 2018). Any further calculations that were displayed were made with both the uncurated posts list and its subset list of curated posts. Also, consider that these curated posts and uncurated posts lists, as well as the output depending on them, will highly likely be different each time Instawareness is (re)loaded because Instagram's curation algorithm is designed to provide different posts

[6] See limitations in section section 6 for further details.

[7] 60 of the 83 correlation tests were significant at the $p < .05$ level. Only four entries had a negative correlation.

at the top of its feed since one's last visit. The views resulting from these calculations are briefly described next.

3.2.1. View one: disclosing personalised rankings

The first view's main purpose was to disclose the main effect of the Instagram curation algorithm, that is, ranking posts differently based on one's prior behaviour. This view was set up with two columns next to each other, allowing users to compare them (see figure 1). On the left, the uncurated posts list was shown in a reverse chronological order on a blue background and labelled to the participants as their 'news feed in absence of a curation algorithm'. This is the aforementioned uncurated posts list and was shortened to the same size (i.e. 50 posts, see section 6) of the curated posts list so one could imagine this list of post as their algorithm-free news feed. On the right, the curated posts list was shown in the identical order as ranked by Instagram's curation algorithm on a yellow background and labelled to the participants as their 'news feed in the presence of a curation algorithm' or 'as it would appear in the app'.

Additional information was provided for enhanced comparison of the algorithm-free and curated feed, as can be seen in figure 1. For instance, next to each post of the curated posts lists a number indicated how many positions the post was 'lower ranked' or 'higher ranked' as a result of Instagram's curation algorithm. Posts that had the same ranking were indicated as 'equally ranked'. Each post of both lists conveyed the picture itself, the account that posted it, and the amount of likes.

As a final point, most of these features were pointed out to the participants in a step-by-step guided tour. In doing so, one of the steps covered the highest ranked post and asked participants to reflect upon the questions 'Does this post indeed deserved a higher spot?' and 'Why does or doesn't the content of [post's author] appeals more to you?'. The next step covered the lowest ranked post with the questions 'Do you mind this post is ranked lower or is it okay?' and 'Do you think that the algorithm ranks your posts lower in your followers' feed from time to time?'. These critical questions, as well as those in any subsequent views (see infra), were used to stimulate reflection upon the

algorithm's consequences and thus its answers were not written down or collected in any way.

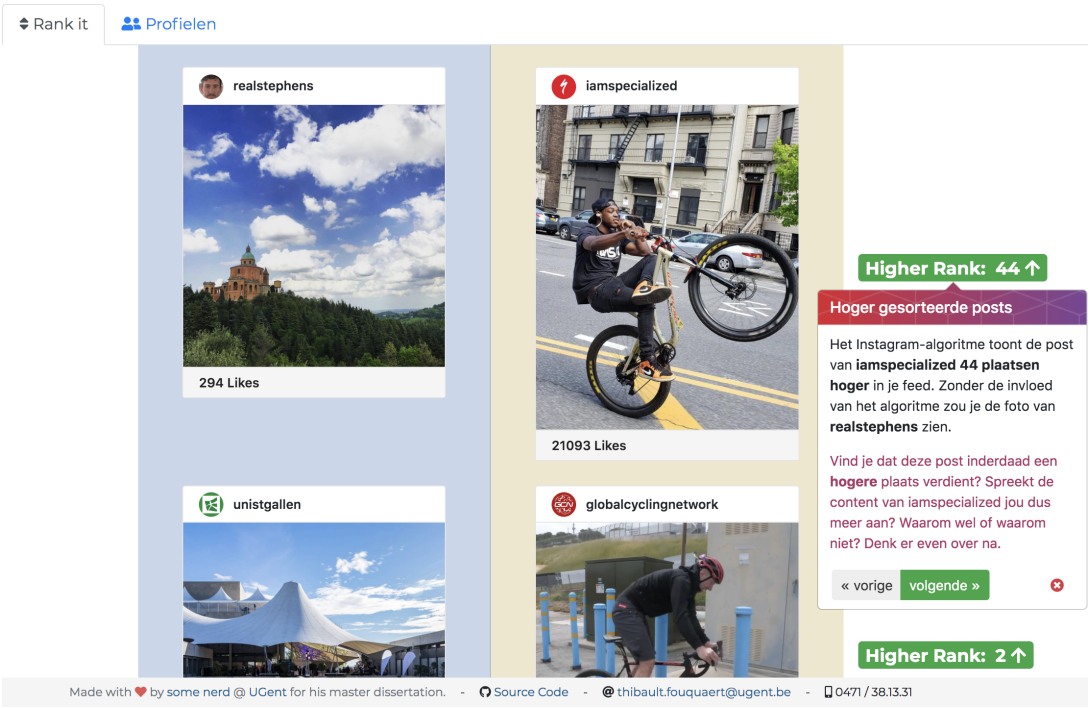

Figure 1. Instawareness view one: side-by-side comparison of how the uncurated (i.e. reverse chronological order) feed would be (in blue on the left) versus how the current curated view is (in yellow on the right). Accessible at https://instawareness.ugent.be

3.2.2. View two: disclosing estimated affinity

The second view's main purpose was to disclose any patterns by the Instagram curation algorithm. SNS users may have a distorted reality of who utilises an SNS more frequently than others (e.g., imagined audiences from section 2) or which type of content is commonly posted as a consequence of curation algorithms filtering and ranking. To help users understand the existence of this distortion and its causes, we built a view with an overview of who is overall higher, equally or lower ranked.

As can be seen in figure 4 (Appendices), the view was divided into three categories to which followings were assigned, coupled with their cumulative rank. The first category (1) 'lower ranked' on the left included followings whose sum of rankings in the curated posts list were negative and thus overall placed lower in participants' news feed by the algorithm. The second category (2) 'equally ranked' in the centre included followings

whose posts were for the majority equally ranked posts with or without algorithm, or had a ratio of total higher ranks to total lower ranks that ranged between .66 and 1.5. The last category 'higher ranked' on the right is opposite to the 'lower ranked' category and includes followings whose sum of rankings in the curated posts list were positive, or in other words, overall placed higher in the participants' news feed.

Similar to view one, some critical questions were posed in the guided tour concerning the algorithm's working. First, the category 'higher ranked' was explained where the questions 'Do you really want to see [name following] more frequently in your news feed?' and 'In whose news feed would you appear more on top?' were posed. Next, the category 'lower ranked' was explained and participants were again asked to reflect upon questions such as 'Why do you think [name following] is overall lower ranked than others in your feed?', 'Does [name following] interests you less or not?', 'Do you agree with this lower ranking?' and 'Do you feel like you can often miss out on posts because of this ranking mechanism?'.

3.2.3. View three: disclosing hidden posts

The third view's main purpose was to disclose whom one could overlook some of their followings and how this absence is possible. In order to explain this, a side-by-side comparison of the uncurated posts list on the left and the curated posts lists on the right was presented (see figure 5, Appendices). The differences with view one, however, are (1) a full-length uncurated posts list that is not shortened to the same size as the curated posts list and (2) a reverse chronological order for both lists which allowed for easier comparison. Next to each post that appeared in the uncurated posts list but not in the curated posts list, the label 'hidden' was added.

Participants were in the guided tour initially informed that Instagram deploys a ranking algorithm and therefore does not hide any posts if one keep scrolling far enough (see sections 2.2 and 2.5). However, posts can be overlooked if one only peeks at some of the first posts in their news feed. Therefore, the explanation started by asking participants to imagine only looking at the first 50 posts in their news feeds instead of keep scrolling

until no new posts appeared. We then explained that some of these posts would fall outside this top 50 curated posts list due to algorithmic curation (e.g., on place 55 instead of 40). Accordingly, posts that were unseen because a lower ranking but would be seen in the absence of a ranking algorithm were labelled as hidden posts.

Similarly, some critical questions were posed. The first question addressed the fact that content could be overlooked by asking 'Do you have the feeling that you frequently miss out on important posts?'. Next, the second question attempted to provoke thought about being filtered oneself and therefore attaining less attention by asking 'Do you think that the algorithm "hides" your posts for your friends?'.

3.3. Post-assessment

In order to understand how participants evolved in terms of cognitive media literacy and attitudes after utilizing Instawareness, as well as the effectiveness of using a visual feedback tool to raise these competencies, participants were assessed in a post-test. Two to three weeks after the pre-test, participants were invited to fill in the same questionnaire as described in section 3.1, which was slightly altered for group B by adding 'After the use of Instawareness, [question]'. After this, each response was linked back to the participant's corresponding pre-test response. Although group A could be influenced by learning effects from reading and answering the questions in the pre-test, no significant differences ($p>.05$) were found within this group between the pre- and post-test for each measured concept as assessed by a paired samples t-test, which allowed them to be used in further between-subject comparison.

3.4. Participants

Due to pragmatic reasons, participants were recruited using the snowball sampling method. Recruitment was done on social media through the personal network of the researchers and mostly reached students from Ghent University (Flanders, Belgium). Initially, 93 respondents started the pre-test of which 70 completed the entire questionnaire. Ultimately, 64 respondents (=N) completed all three phases and were used

for further data analysis. No incentive was granted to participate. The sample existed out of 64% women and 36% men who ranged between 18 and 35 years of age ($M_{age} = 23.97,\ SD = 2.81$). Participants were largely (86 %) highly educated with 31% having obtained a bachelor's degree while an additional 55% had obtained a master's degree. Most of them were also familiar with one or more other SNSs such as Facebook (63 participants), YouTube (45 participants), LinkedIn (34 participants) and more.

3.5. Data analysis

Data analysis was done in R (R Core Team, 2018) and by use of the PROCESS v3 macro for SPSS from Hayes (2017). First, summation scales were calculated for each measured concept (see section 3.1). Second, a simple mediation analysis was conducted using ordinary least squares path analysis to interpret the mediation model (Hayes, 2017, model 4) which helped to partly answer the RQ. Post-test score were used to analyse this model, as recommended by Hayes (2017, p. 544). Next, a one-way analysis of covariances (ANCOVA) was conducted using pre-test scores as covariates to analyse differences in cognitive media literacy, general feelings and critical concern between group A and B for H_1, H_2 and H_3 respectively. Subsequently, Spearman correlation tests were conducted to detect covariates (e.g., technical media literacy) but yielded no positive results. In some cases, we followed up with a two-way mixed analysis of variances (MANOVA). Last, each of the three analyses were preceded by testing its underlying assumption (for details, see section 8.1, Appendices) and confirmed the data's model fit for every used test.

Even though a MANOVA could also be conducted to initially test the hypotheses H_1, H_2 and H_3, a one-way ANCOVA analysis allows for detecting and comparing (post-test) differences rather than analysing the amount of gain for each group (Wright, 2006). Overall, it is argued that a one-way ANCOVA is the preferred method for randomised pre- post-test designs as it allows for a higher degree of external validity by dint of reduced error variance (Dimitrov & Rumrill, 2003). In this regard, a one-way ANCOVA appeared to be the most appropriate test to answer H_1, H_2 and H_3 while a MANOVA allowed for appropriate follow-up testing when the ANCOVA results were inconclusive.

Table 3. Adjusted and unadjusted means and variability for post-test cognitive media literacy with pre-test cognitive media literacy as covariates, split by control group A and group B that used Instawareness.

		Unadjusted		*Adjusted*	
	n	*M*	*SD*	*M*	*SE*
Control (A)	32	10.69	3.65	10.82	0.55
Instawareness (B)	32	13.31	4.04	13.18	0.55

Control = Control group A that received no treatment, Instawareness = Group B that used Instawareness. Range cognitive media literacy: $0 - 20$.

Outcomes of these analyses are presented in section 4. Results discussed for H_1, H_2 and H_3 concern the use of Instawareness for our sample, but might give an indication for other, yet to be developed or existing, visual feedback tools (e.g., FeedVis by Eslami, Aleyasen et al., 2015) when applied to a similar sample.

4. Results

4.1. H_1: visual feedback tools' effectiveness in raising algorithm awareness

In hypothesis one, we test whether or not people have gained some understanding of the actual working of Instagram's algorithm after they used a visual feedback tool such as Instawareness. After adjusting for the pre-test cognitive media literacy scores, positive differences with a large effect size in the post-test (i.e. after the use of Instawareness) cognitive media literacy scores are found for those who used Instawareness, $F(1, 61) = 8.93,\ p = .004,\ \eta_p^2 = .13$. These participants thus gained significant better understanding about algorithms and their capabilities, as can be seen in figure 2 and table 3.

In short, the cognitive media literacy is found to be 2.36 points ($= M_{\text{diff}},\ 95\,\%$ CI $[0.81, 3.91]$) higher after the intervention on a 0 to 20 scale for group B ($M_{\text{b}} = 13.18,\ 95\,\%$ CI $[12.06, 14.30]$) compared to control group A ($M_{\text{a}} = 10.82,\ 95\,\%$ CI $[9.70, 11.94]$). Furthermore, a paired sample t-test shows that the gain score in mean difference from pre- to post-test is significant for solely group B with an increase of 2.69, $t(31) = 3.56,\ p = .001$, and not for group A, which had a gain score

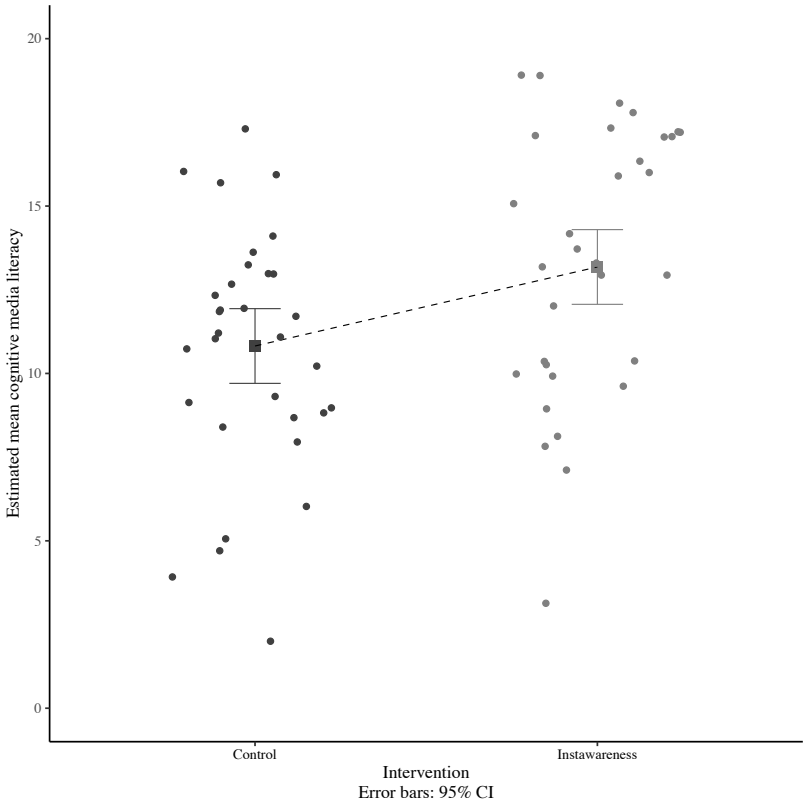

Figure 2. Hypothesis H_1: difference in estimated mean media literacy score between control group A and manipulation group B.

in mean difference of 0.53, $p = .28$. This paired sample t-test additionally supports the already indicated effectiveness of Instawareness, as only group B gained a significant increase over time. Adjusted means have been presented, unless otherwise stated.

4.2. H_2: visual feedback tools' effect on general feelings towards SNSs

Hypothesis two tests if people who have used Instawareness had more positive general feelings towards Instagram than those who did not use Instawareness. After adjusting for the pre-test general feelings scores, those who used Instawareness appear to have more or less the same feelings towards Instagram than those who did not use the tool, $F(1, 60) = 1.18,\ p = .28,\ \eta_p^2 = .02$. Manipulation group B had a mean score of general feelings of 13.59 for the post-test, which is a small increase of 0.53 points ($= M_{\text{diff}},\ 95\,\%$ CI $[-0.42, 1.48]$) on a scale of 0 to 24 compared to the mean general feelings score of 13.06 for control group A during the post-test.

More importantly, however, is the observed overall decrease in general feelings towards Instagram scores from pre- to post-test as indicated by a follow-up MANOVA. Even though we confirmed a slight increase for the post-test general feelings scores for group B compared to group A, both groups do actually have less positive feelings towards Instagram after the intervention. As mentioned above, people who did fill in the post-test questionnaire after having used Instawareness (i.e. group B) postulate less negative feelings ($M_{\text{diff}} = 0.53$) compared to those who only filled in the post-test questionnaire and did not use Instawareness (i.e. group A). This can also be seen in figure 7 (Appendices), the mean general feelings score for control group A decreased pre to post from 13.75 to 12.81 ($M_{\text{diff}} = -0.94$) and for group B from 13.84 to 13.66 ($M_{\text{diff}} = -0.19$). This decrease in general feelings is not significant for the interaction effect between intervention and time ($F(1, 62) = 1.40,\ p = .24,\ \eta_p^2 = .02$), nor for the simple main effect of time ($F(1, 61) = 3.15,\ p = .08,\ \eta_p^2 = .05$).

4.3. H_3: visual feedback tools' effect on critical concern towards curation algorithms

In hypothesis three, we ask ourselves if people who have used Instawareness would pose increased critical concerns towards Instagram compared to those who did not use Instawareness. After adjusting for pre-test scores of critical concerns, users of Instawareness do not significantly pose increased critical concerns towards Instagram compared to the participants who did not use Instawareness, $F(1, 61) = 1.63,\ p = .21,\ \eta_p^2 = .03$. The post-test mean critical concern score is 27.89 for group B and higher by 1.56 points ($= M_{\text{diff}},\ 95\,\%$ CI $[-0.83, 3.94]$) on a scale from 0 to 40 compared to the mean critical concern score of 26.33 from group A.

Even though no significant increase in critical concerns towards Instagram can be attributed to solely the use of Instawareness, we do nevertheless provide proof that our experiment impacts participant's critical concerns towards Instagram. The simple main effect of time in both groups taken together, that is, either completing the questionnaires or completing the questionnaires using Instawareness in between, allowed our participants to significantly pose increased critical concerns, $F(1, 62) = 8.52,\ p = .005,\ \eta_p^2 = .12$.

This growth of critical concerns when becoming aware and knowledgeable about curation algorithms is moreover an important first step in the illustration of our algorithm paradox.

A visual overview of the aforementioned effect from our experiment on participant's critical concern is given in figure 8. From this, we observe an increase in mean critical concerns from pre- to post-test for control group A from 15.38 to 16.44 ($M_{\text{diff}} = 1.06$) while there is a slightly bigger increase in mean critical concerns score for group B from 15.09 to 17.78 ($M_{\text{diff}} = 2.69$). The interaction effect between time and intervention appears to be insignificant, $F(1, 62) = 1.60,\ p = .21,\ \eta_p^2 = .03$.

This suggests that becoming aware (i.e. through learning effects from the questionnaire) is sufficient to pose increased critical concerns towards SNS while becoming knowledgeable about curation algorithms' capabilities (i.e. increased cognitive media literacy besides becoming aware by the use of Instawareness) does not further increase critical concerns significantly. Overall this confirms previous studies described in section 2.4 as well as the first stage of our proposed algorithm paradox, that is, people who know about the existence of curation algorithms claim to be bothered by it. Even though people who utilised Instawareness do not mind activities such as feed curation significantly more than others who solely became aware of this, the overall findings are still insightful in evaluating the relationship between media literacy and attitudes as posed in our main research question further described in section 4.5.

4.4. Overview hypotheses testing

Table 4. Overview of the hypotheses testing results.

Alternative hypothesis	*result*	*interpretation*
H_1	accepted**	The use of Instawareness significantly increases cognitive media literacy.
H_2	rejected	The use of Instawareness reduces the decrease in general feelings towards SNS. The reduction and decrease in general feelings is, however, not significant.
H_3	rejected	The effect of using of Instawareness is insufficient to significantly increase critical concerns towards certain activities SNSs carry out, although becoming aware of solely algorithms' presence is sufficient to pose increased critical concerns.

$* = p < .05, ** = p < .01$, only the alternative hypotheses are listed as was done in section 2.

Table 5. Overview mediation model

		Consequent						
		M(TML[a])				Y(CCONCERN)		
Antecedent		Coeff.	SE	p		Coeff.	SE	p
X(CML)	a	0.29	0.15	.06	c'	0.50	0.19	.009
M(TML[a])		—	—	—	b	0.61	0.15	$< .001$
constant	i_M	13.10	1.91	$< .001$	i_Y	1.27	3.04	.68
		$R^2 = .06$ $F(1, 61) = 3.78,\ p = .06$				$R^2 = .33$ $F(1, 61) = 14.96,\ p < .001$		

[a]Frequency was used as a proxy.
The analysis was conducted with a one-tailed α of .05 as we expected a positive correlation, which is equal to running the mediation analysis at two-tailed α of .1. Unstandardised coefficients have been reported.

4.5. RQ: cognitive media literacy's role in explaining attitudes towards SNSs

Users' cognitive understanding and competencies in consumption and evaluation of SNSs' content and mechanisms has a direct and indirect influence on their critical concerns towards certain activities SNSs carry out, such as filtering or selling and using personal behavioural data. The indirect influence arises through cognitive media literacy's effect on users' technical understanding and competencies in operating SNSs.

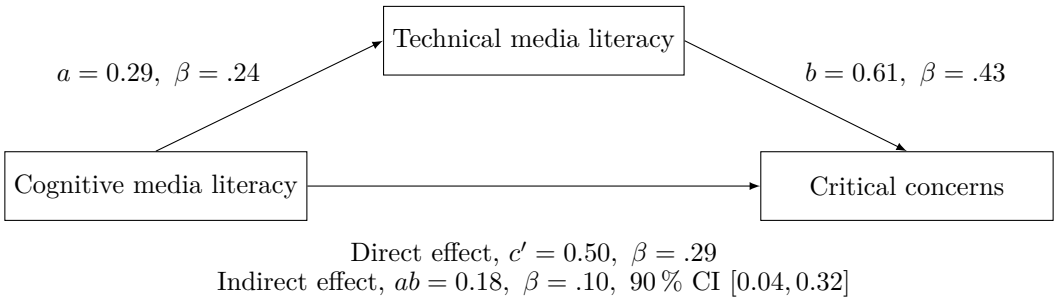

Figure 3. Partial mediation through technical media literacy for the effect of cognitive media literacy on critical concerns. Frequency was used as a proxy for technical media literacy. β reports the standardised coefficients which represent the expected difference in the dependent variable (e.g., critical concerns for effect c'), *in standard deviations*, between two participants that differ by one standard deviation on the independent variable (e.g., cognitive media literacy for effect c'). β can be used to compare relative importance and effect size.

As can be seen in figure 3 and table 5, participants who have higher cognitive understanding indicate to have raised critical concerns as direct effect ($c' = 0.50$). Besides this, an additional indirect effect on critical concerns appears to be present. Participants who had such higher cognitive understanding show greater technical under-

standing ($a = 0.29$), and those who had greater technical understanding indicate to have raised critical concerns ($b = 0.61$) as well. This indirect effect is entirely above zero ($ab = 0.18,\ 90\,\%$ CI $[0.04, 0.32]$) based on 5000 bootstrap samples. Finding this additional mediation effect of technical media literacy on critical concerns can be one of the explanations why hypothesis H_3 was rejected, as the influence of technical media literacy could not be manipulated through Instawareness despite its effect on critical concerns.

Remarkable is the difference in using frequency as a proxy instead of familiarity (see section 3.1). The model presented above indicates a strong link between frequency as measure of technical media literacy and critical concerns. In contrast, using familiarity as a proxy for technical media literacy results in no link at all with critical concerns when applied to the same model ($b_{\text{familiarity}} = -0.01,\ p = .91$). This discrepancy clearly implies that frequency and familiarity do not measure the exact same concept. Although they both measure technical media literacy, frequency inevitably also incorporates the side effect of the intensity in using SNS, which may in turn be one of the many confounds that explains this observed difference in having raised critical concerns. Despite this discrepancy, frequency of use does not seem to influence cognitive media literacy more or less than familiarity does, as both measures were equally correlated and insignificant.

Lastly, we like to give some reasoning on why the relationships in this mediation model (i.e. $a,\ b,\ c'$), apart from its empirical justification or natural correlation, constitute a sensible causal process. First, since technical understanding is considered to be required for the development of cognitive understanding (section 2.3), it is reasonable for cognitive media literacy to be an indicator of technical media literacy (a). Second, users' cognitive understanding can further develop through consequent frequent use or oppositely diminish when they lack the ability to make appropriate use of their technical understanding. In this way, we believe that technical media literacy affects critical concern (b). Third, earlier studies indicated changes in attitudes and habits as a result of change in cognitive understanding (see section 2.2.1), which makes it reasonable to believe that this will equally effect users' critical concerns (c').

4.5.1. The discrepancy between concerns versus behaviour

Nearly all participants (95 %) are aware of at least one activity that encompasses some form of algorithmic curation at the start of the experiment. Even if one argues against the supposition that knowing only one feature can be considered aware of algorithmic curation, the vast majority (80 %) of participants could still be considered aware since they know at least 3 (out of 7) of curation algorithms' features. This is in contrast to previous studies that indicate far lower awareness of mechanisms such as filtering and curation. One explanation might be because participants had to indicate the activities thought to be present on SNSs, instead of answering to questions directly probing about some 'algorithm'. Moreover, more than half of the participants (63 %) indicated to have the feeling they sometimes miss posts and nearly all (98 %) of these participants indicated to have the feeling of missing posts 'due to some filtering of Instagram itself'.

Furthermore, this high level of participants' awareness seems in line with their level of technical media literacy. The majority indicates to be highly familiar with Instagram as SNS ($M_{\text{familiarity}} = 30.66,\ SD = 7.57,\ \text{scale} = 0 - 45$) even while indicating to use different features such as 'posting a story' or 'commenting on a post' only monthly or less frequently on average ($M_{\text{frequency}} = 9.42,\ SD = 4.77,\ \text{scale} = 0 - 28$). Participants' above average operational and formal skills are reflected in their ability to find, evaluate and utilise the presented content as seen in their fair to high level of cognitive media literacy ($M_{\text{cml}} = 10.39,\ SD = 3.95,\ \text{scale} = 0 - 20$). We argue this is a fair to high level of cognitive media literacy given that participants had no prior formal education in this and existing knowledge was gained through (experimental) usage. Participants' concerns, on the other hand, appear present but rather low ($M_{\text{concerns}} = 15.23,\ SD = 7.05,\ \text{scale} = 0 - 40$) for the given level of average cognitive media literacy prior to the use of Instawareness.

Given these points, the level of technical media literacy that can be used to tackle people's potential concerns about curation algorithms appear to be sufficiently present. In addition, the level of cognitive media literacy and critical concerns is mediocre but certainly not absent, which allows to develop certain beliefs (i.e. concerns) and act ac-

cordingly (i.e. change in habits). Furthermore, H_1 and H_3 confirmed the significant gain in participants' level of cognitive media literacy and concerns by making them aware, supporting the fact that our participants do adapt their attitudes and expectations accordingly to their knowledge. However, no change in habits to act accordingly to their increased concerns was identified as examined by McNemar's test for neither group A nor B from pre to post. Likewise, no correlation was found between participants level of concerns and their habits as examined by a point-biserial correlation, except for fear of missing out (FOMO), $r(62) = .36,\ p = .004$, and blaming the algorithm when having fewer likes, $r(62) = .32,\ p = .01$.

In conclusion, these findings support our argumentation of the algorithms paradox. If people know about the existence of curation algorithms, they claim to be bothered by its mechanism while not acting accordingly, that is, altering settings, changing interaction or visiting certain profiles.

5. Conclusion and discussion

This paper addresses the influential yet subtle and often hidden side effects of algorithmic curation on SNSs. One such eminent side effect is the filter bubble, in which prescriptive algorithms personalise all sort of results (e.g., search results, news feeds, ads) based on estimates about what fits people's beliefs and likings. The filter bubble effect is exacerbated and especially difficult to tackle due to people's ignorance about these personalisation mechanisms and high faith in the veracity of its results. We argue, however, that yet another — and maybe even greater — difficulty in tackling the issues of the filter bubble is our proposed algorithm paradox.

Based on prior research and the empirical results presented in this paper, people appear to be unable to escape their filter bubble because of the ignorance about personalisation mechanisms and its side effects, which in turn impedes the — required — change in critical attitudes. Following up on this, we observed a phenomenon what we like to call '*the algorithm paradox*'. Herein, reducing people's ignorance leads to higher critical concerns and claims about being bothered by algorithms' mechanisms, although the habits

of these people in using SNSs do not change accordingly. Yet, it is this change in habits, such as checking different sources, following a diversity of profiles or acknowledging greater prevalence of like-minded opinions, that is ultimately required to overcome the negative side effects of filter bubbles. Evidencing this discrepant relationship is our first contribution towards paths of understanding and measuring the perception, media literacy and attitudes towards algorithmic processes on SNSs.

Besides this, these findings also revealed unexpected differences regarding technical media literacy's conceptualisation. In short, it is fair to say that the effect of frequency of use stimulates critical concerns, whereas the effect of self-reported familiarity on critical concerns seems to be completely absent. These conceptualisations were nonetheless thought to both measure the same operational and formal skills (i.e. technical media literacy). That being so, we state that more intensive users become aware of and knowledgeable about mechanisms such as personalisation through their frequent use of SNSs. The consequent increase in critical concerns could, for instance, originate from people's distress about the impact of known mechanisms that could in turn be amplified by their intensive use combined with their ignorance about the actual working of such mechanisms.

A second contribution is our approach in increasing algorithmic awareness about curation algorithms, cognitive media literacy and critical attitudes towards SNSs. This paper overcomes the limitations of previous studies (e.g., Eslami, Rickman et al., 2015) and fulfills their suggestions for new research to quantitatively confirm their findings about the effects and effectiveness of exposing hidden algorithmic processes to users with the help of what we call visual feedback tools. Instawareness, our self-developed visual feedback tool covering Instagram, is effective in properly exposing and explaining the working of invisible curation algorithms thereby directly increasing the cognitive media literacy for those who use such intervention tools. However, gaining cognitive understanding about the actual working of curation algorithms on SNSs through Instawareness did not seem to be necessary in order for people to pose increased critical concerns. Instead, solely making participants aware of the curation algorithms' existence (e.g., through a questionnaire) already appeared to be sufficient in raising people's critical concerns.

Contrary to prior research, Instawareness provoked a slight (insignificant) decrease in satisfaction and general feelings towards SNSs for those who used it, while people who were aware but still ignorant about the algorithm's actual working imposed an even greater decrease. While contrary on the first sight, these results could however also reflect the development of attitudes as described in section 2.2.1 in that aware but unknowledgeable users (i.e. control group A) develop negative feelings while aware and knowledgeable users (i.e. manipulation group B) begin to develop more positive feelings. On a final note, we might suspect these results to be caused by our admittedly strong focus on the possible negative side effects of curation algorithms as pointed out by Instawareness, instead of also focussing on possible positive effects.

Combining these two insights, we argue that both users and companies can benefit from the use of visual feedback tools if properly integrated into SNSs. Even though our results point out that this in-depth approach using Instawareness is no necessity towards making people raise concerns, the benefits that come along with using a visual feedback tool make this the preferred approach. First, using tools such as Instawareness allows users to become far more knowledgeable about the consequences of contemporary online mechanisms such as personalisation and assess its actual impact, rather than raising concerns based on their own beliefs. Second, these tools can positively impact users' self-development by stimulating their capabilities to better estimate the impact on their self-presentation. Understanding that less online attention might be due to an algorithm instead of their peers, can be a major deal to some. Third, this approach educates users about the 'rulebook' they can adapt to and how to play their 'visibility game' accordingly, which we currently value most since users can exert little direct influence themselves on SNSs' algorithms. Similarly, companies can benefit from this approach when integrating feedback mechanisms into their SNSs or when providing an external platform. Users who are offered a decent explanation and transparency into the hidden mechanisms, develop less negative feelings when becoming aware and might even become more satisfied in the long term when potential benefits are pointed out. Company's main goal should be to allow for similar gains in cognitive understanding while still

maintaining a continued seamless interaction on the SNS itself, that is, uninterrupted use with this background knowledge.

Most importantly, we argue that this seamless interaction is one of the biggest challenges still to overcome and we even believe this is one of the explanatory factors of our said algorithm paradox. Users are offered too little ways in which they are able to change their habits according to their concerns, in a way that is above all seamless, effortless and moderate. Instead, the nature of most measures against the side effects curation algorithms at this point are anything but the aforementioned ways. For instance, users must manually visit different sources or profiles, make own estimates about planned interactions or check the recency of newly appearing posts, which all appears to be too drastic or time-consuming to actually perform. Hence, research on the explaining factors of our proposed algorithm paradox, as well as research on interaction design that allows for more critical SNSs use, deserve appropriate attention and will require further contributions from research in this field.

6. Limitations

Some technical issues arose during the development of Instawareness. First, the amount of requests to be made while fetching the uncurated posts could quickly become extensive. To limit this, the amount of fetched curated posts was reduced to a total of 50 posts as an attempt to reduce the timespan that had to be fetched back in time. This limited list of 50 curated posts as the base for further calculations was found to still reveal quality information while not all too often reaching the API's limits. Despite this measure, participants with a substantial amount of followings[8] could still reach the API's limit if the news feed was significantly reordered. These participants were therefore served calculations with the uncurated posts up to the point of request limit.

Second, there was no way possible to be completely certain that the uncurated posts list was exhaustive since no guarantees exists of the lowest possible rank a post can get by the Instagram curation algorithm. Even though near complete certainty could

[8]The Instagram accounts one follows and therefore appear in their news feed.

be achieved by fetching an extensive amount of posts that were made earlier than the least recent post from the curated posts list, this would by far exceed the request limit. Therefore, the uncurated posts list was assumed as complete once a subsequent request, which exists out of 50 posts, contained no posts that belonged to the uncurated posts (i.e. more recent than the oldest post in the curated posts list). Based on trial and error, this appeared as a valid cut-off criteria to continue with.

One final issue was the unavailability of an official API to fetch one's news feed so that the Instagram website endpoints had to be used to implement the retrieval of one's news feed. During this process no personal data was being saved except for (1) the rankings produced by the curation algorithm in a reverse chronological order and (2) the amount of hidden posts (see section 3.2.3). Data was processed anonymously and transferred with encryption under any circumstances. To account for this, the whole Instawareness project was open source available[9] during the experiment and will remain so in the future.

Regarding data analysis, the simple mediation analysis could have been optimised if the pre-test scores were used as covariates when interpreting the model (Hayes, 2017, p. 544). This was not done at the time the analysis was conducted because considering our data for longitudinal mediation models was beyond the scope of this paper and also required additional specific expertise in structural equation modeling.

Acknowledgements

This master's thesis would not have been there without the guidance and support of some people. To this end, I would like to thank my adviser and promoter, dr. Peter Mechant, for his thoughtful advice, feedback and encouragement. This has brought me (career) opportunities I would not have came across otherwise for which I am truly grateful. On a personal note, I would like to thank both my parents for their support throughout my studies.

[9] https://github.com/Thibaultfq/Instawareness

Notes

Instawareness is considered to be included in the Databuzz project. This is one of the current flagship projects of our colleagues at imec-SMIT-VUB and aims to increase the data literacy of youth aged 10 to 18 by means of interactive educational activities.[10]

7. References

Awad, N. F. & Krishnan, M. S. (2006). The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, *30*(1), 13–28.

Beer, D. (2017). The social power of algorithms. *Information Communication and Society*, *20*(1), 1–13.

Bernstein, M. S., Bakshy, E., Burke, M. & Karrer, B. (2013). Quantifying the invisible audience in social networks. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 21–30).

Boyd, D. (2014). *It's complicated: the social lives of networked teens*. Yale University Press.

Boyd, D. & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*(5), 662–679.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227.

Bruns, A. (2008). Gatewatching, gatecrashing: futures for tactical news media. In M. Boler (Ed.), *Digital media and democracy: Tactics in hard times* (pp. 247–270). Cambridge, Mass.: The MIT Press.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180.

Bucher, T. (2016). Neither black nor box: ways of knowing algorithms. In *Innovative methods in media and communication research* (pp. 81–98). London: Palgrave Macmillan.

Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information Communication and Society*, *20*(1), 30–44.

Constine, J. (2018). How Instagram's algorithm works - TechCrunch. Retrieved October 22, 2018, from https://techcrunch.com/2018/06/01/how-instagram-feed-works

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2009). *Introduction to algorithms, Third Edition*. The MIT Press.

Cotter, K. (2019). Playing the visibility game: how digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, *21*(4), 895–913.

Delen, D. & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems*, *55*(1), 359–363.

Dimitrov, D. & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159–165.

Dwyer, N., Basu, A. & Marsh, S. (2013). Reflections on measuring the trust empowerment potential of a digital environment. In C. Fernández-Gago, F. Martinelli, S. Pearson & I. Agudo (Eds.), *Trust management vii. ifiptm 2013. ifip advances in information and communication technology* (Vol. 401, pp. 127–135).

Eslami, M., Aleyasen, A., Karahalios, K., Hamilton, K. & Sandvig, C. (2015). FeedVis: a path for exploring news feed curation algorithms. In *Proceedings of the 18th acm conference companion on computer supported cooperative work & social computing* (pp. 65–68).

[10] More information about the Databuzz project can be found on http://smit.vub.ac.be/policy-brief-28-the-databuzz

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K. & Kirlik, A. (2016). First I like it, then I hide it: folk theories of social feeds. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 2371–2382).

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., ... Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": reasoning about invisible algorithms in news feeds. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 153–162).

Eslami, M., Vaccaro, K., Karahalios, K. & Hamilton, K. (2017). "Be careful; things can be worse than they appear": understanding biased algorithms and users' behavior around them in rating platforms. In *Eleventh international aaai conference on web and social media*.

Facebook. (2015). Graph API reference /{user-id}/home. Retrieved May 7, 2019, from https://developers.facebook.com/docs/graph-api/reference/v3.3/user/home

Galloway, A. R. (2006). *Gaming: essays on algorithmic culture*. Minneapolis, MN: University of Minnesota Press.

Gatz, S. (2018). *Beleidsbrief media 2018-2019*. Vlaams Parlement.

Hamilton, K., Karahalios, K., Sandvig, C. & Eslami, M. (2014). A path to understanding the effects of algorithm awareness. In *Proceedings of the extended abstracts of the 32nd annual acm conference on human factors in computing systems* (pp. 631–642).

Hargittai, E. (2002). Second-level digital divide: differences in people's online skills. *First Monday*, *7*(4).

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, *23*(3), 371–379.

Hargittai, E., Fullerton, L., Menchen-Trevino, E. & Thomas, K. Y. (2010). Trust online: young adults' evaluation of web content. *International Journal of Communication*, *4*, 468–494.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Publications.

Helsper, E. J. & Eynon, R. (2013). Distinct skill pathways to digital engagement. *European Journal of Communication*, *28*(6), 696–713.

Hogan, B. (2010). The presentation of self in the age of social media: distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, *30*(6), 377–386.

Instagram. (2016). See the moments you care about first. Retrieved February 25, 2019, from https://instagram-press.com/blog/2016/03/15/see-the-moments-you-care-about-first/

Instagram. (2019). Our story. Retrieved February 26, 2019, from https://instagram-press.com/our-story/

Introna, L. D. & Nissenbaum, H. (2000). Shaping the web: why the politics of search engines matters. *The Information Society*, *16*(3), 169–185.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: media education for the 21st century*. The MIT Press.

Just, N. & Latzer, M. (2017). Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, *39*(2), 238–258.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134.

Latzer, M., Hollnbuchner, K., Just, N. & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In *Handbook on the economics of the internet.* (p. 395).

Livingstone, S. & Helsper, E. (2010). Balancing opportunities and risks in teenagers' use of the internet: the role of online skills and internet self-efficacy. *New Media & Society*, *12*(2), 309–329.

Livingstone, S., Van Couvering, E. & Thumin, N. (2008). Converging traditions of research on media and information literacies. In *Handbook of research on new literacies* (pp. 103–132). Routeledge.

Mantovani, G. (1996). Social context in HCI: A new framework for mental models, cooperation, and communication. *Cognitive Science*, *20*(2), 237–269.

Marwick, A. E. & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133.

McClelland, G. H. (2014). Nasty data: unruly, ill-mannered observations can ruin your analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 608–626). Cambridge University Press.

Napoli, P. M. (2014). Automated media: an institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, *24*(3), 340–360.

Nelson, R. R. & Winter, S. G. (2002). Evolutionary theorizing in economics. *Journal of Economic Perspectives*, *16*(2), 23–46.

Norberg, P. A., Horne, D. R. & Horne, D. A. (2007). The privacy paradox : personal information disclosure intentions versus behaviors. *The Journal of Consumer Affairs*, *41*(1), 100–126.

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you.* Penguin UK.

Paulussen, S., Courtois, C., Vanwynsberghe, H. & Verdegem, P. (2011). Profielen van mediageletterdheid: een exploratie van de digitale vaardigheden van burgers in Vlaanderen. In M.-A. Moreas & J. Pickery (Eds.), *Mediageletterdheid in een digitale wereld* (Vol. 1, pp. 61–76). Studiedienst van de Vlaamse Regering.

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria.

Rader, E. & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 172–182). ACM.

Roberts, C. (2005). Gatekeeping theory : an evolution. *Paper presented at Communication Theory and Methodology Division, Association for Education in Journalism and Mass Communication, San Antonio, TX.*

Sandvig, C., Hamilton, K., Karahalios, K. & Langbort, C. (2014). Auditing algorithms: research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, *22*.

Talja, S. (2005). The social and discursive construction of computing skills. *Journal of the American Society for Information Science and Technology*, *56*(1), 13–22.

Tandoc Jr, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media and Society*, *16*(4), 559–575.

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, *28*(1), 20–36.

van Deursen, A. J. A. M. (2010). *Internet skills. Vital assets in an information society* (Doctoral dissertation, Enschede, the Netherlands: University of Twente).

van Deursen, A. J. A. M., Courtois, C. & van Dijk, J. A. G. M. (2014). Internet skills, sources of support, and benefiting from Internet use. *International Journal of Human-Computer Interaction*, *30*(4), 278–290.

van Deursen, A. J. A. M., Helsper, E. J. & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). *Information Communication and Society*, *19*(6), 804–823.

van Deursen, A. J. A. M. & van Dijk, J. A. G. M. (2009). Using the Internet: skill related problems in users' online behavior. *Interacting with Computers*, *21*(5-6), 393–402.

van Deursen, A. J. A. M. & van Dijk, J. A. G. M. (2010). Measuring Internet skills. *International Journal of Human-Computer Interaction*, *26*(10), 891–916.

van Deursen, A. J. A. M. & van Dijk, J. A. G. M. (2011). Internet skills and the digital divide. *New Media and Society*, *13*(6), 893–911.

van Deursen, A. J. A. M. & van Dijk, J. A. G. M. (2015). Internet skill levels increase, but gaps widen: a longitudinal cross-sectional analysis (2010–2013) among the Dutch population. *Information, Communication & Society*, *18*(7), 782–797.

van Deursen, A. J. A. M., van Dijk, J. A. G. M. & Peters, O. (2012). Proposing a survey instrument for measuring operational, formal, information, and strategic Internet skills. *International Journal of Human-Computer Interaction*, *28*(12), 827–837.

van Dijk, J. A. G. M. (2005). *The deepening divide: inequality in the information society*. Sage Publication.

van Dijk, J. A. G. M. & van Deursen, A. J. A. M. (2014). *Digital Skills: unlocking the information society*. New York: Palgrave Macmillan.

Vanhaelewyn, B. & De Marez, L. (2018). *Digimeter 2018: measuring digital media trends in Flanders*. Imec.

Vanwynsberghe, H., Boudry, E. & Verdegem, P. (2015). De impact van ouderschapsstijlen op de ontwikkeling van sociale mediageletterdheid bij adolescenten. *Tijdschrift voor communicatiewetenschappen*, *43*(1), 84–100.

Vanwynsberghe, H. & Haspeslagh, L. (2014). *Getting started measuring social media literacy*. iMinds-mict-UGent.

Verdegem, P., Haspeslagh, L. & Vanwynsberghe, H. (2014). *EMSOC survey report: social media use and experience of the Flemish population*. iMinds-mict-UGent.

Wright, D. B. (2006). Comparing groups in a before-after design: shen t-test and ANCOVA produce different results. *British Journal of Educational Psychology*, *76*(3), 663–675.

Yang, H.-d. & Yoo, Y. (2004). It's all about attitude: revisiting the technology acceptance model. *Decision Support Systems*, *38*(1), 19–31.

Yeung, K. (2017). "Hypernudge": Big Data as a mode of regulation by design. *Information Communication and Society*, *20*(1), 118–136.

Zhang, J. (2015). *Fiddler: A visualization prototype interface for making sense of newsfeeds* (Master's thesis, University of Illinois).

8. Appendices

8.1. Assumptions underlying the analyses

8.1.1. Multiple regression (RQ)

For the main research question, assumptions underlying multiple regression were checked as this was used to perform the simple mediation analysis. First, outliers were checked on the basis of having studentised deleted residuals $> \pm 3$ standard deviations. One outlier was found with a studentised deleted residual of -3.07 and was removed from further analysis since it did not determine the model's meaning or significance. In such case, this is claimed to be an appropriate method as observations with high leverage will substantially distort the model and actually tell a different 'story' (i.e. model) (McClelland, 2014). Moreover, McClelland (2014) argues that similar to explained variance, one may admit finding a model that applies to, for example, 95 % of the observations while having no clue about the 5 % other entries (i.e. outliers) as 'A good model for most of our data is better than a poor model for all of our data.' (p. 2479).

Second, no leverage points greater than 0.2 or Cook's distances greater than 1 were found. Inspection of a scatter plot of studentised residuals against the predicted values, as well as partial regression plots of technical media literacy and cognitive media literacy against critical concerns, indicated the data's linearity and homoscedasticity. Next, independence of residuals appeared to be present by a Durbin-Watson test of 1.78. Also, there was no evidence of multicollinearity with VIF values of 1.06 and no correlations of between the model's variables greater than .7. Last, normality was assured by Shapiro-Wilk's test ($p = .25$) and inspection of a Q-Q plot of the studentised residuals. Overall, all assumptions for the simple mediation model were met.

8.1.2. ANCOVA (H_1, H_2, H_3)

Before the ANCOVAs for H_1, H_2 and H_3 were conducted, assumptions underlying analysis of covariance were checked. To begin, no outliers were found having studentised residuals $> \pm 3$ standard deviations regarding cognitive media literacy (CML) (H_1) and

critical concerns (CC) (H_3). For general feelings (GF) (H_2), however, one outlier was identified having a studentised residual of -4.11 ($CA_{pre} = 17$, $CA_{post} = 7$). Since analysis of (co)variances is sensitive for outliers, the outlier was removed for the ANCOVA assessment of H_2 because it did not determine any meaningful levels of significance and so that obtained results had a better model fit (see section 8.1.1).

Continuing with outliers removed only for H_2, a positive linear relationship was present between pre- and post-test CML scores (H_1) in both group A and B ($p_A < .001,\ p_B = .024$). This was also the case for GF scores (H_2) and CC scorers (H_3) in both group A and B (each $p_A < .001,\ p_B < .001$) as assessed by Pearson correlation coefficients and visual inspection of scatter plots. Next, no significant interaction was found between pre-test CML and the use of Instawareness ($F(1, 60) = 1.40,\ p = .24$), GF and the use of Instawareness ($F(1, 59) = 0.54,\ p = .46$), nor CC and the use of Instawareness ($F(1, 60) = 3.15,\ p = .08$). Hence, homogeneity of regression slopes was checked and confirmed for all three hypotheses. Following, Shapiro-Wilk's test and visual inspection of Q-Q plots revealed both normally distributed within-group and overall standardised residuals for CML, GF and CC. Also, homoscedasticity appeared to be present for all three constructs upon visual inspection of the studentised residuals and homogeneity of variances was found as well for CML ($p = .35$), GF ($p = .56$) and CC ($p = .09$) using Levene's test. Last, No significant differences were found between group A and B for CML covariates (i.e. pre-test scores) ($p = .64$), GF covariates ($p = .81$), nor CC covariates ($p = .87$) using Welch's t-test. Overall, all assumptions regarding H_1, H_2 and H_3 were met.

8.1.3. MANOVA (H_2, H_3)

Before conducting the follow-up tests in H_2 and H_3, assumptions underlying MANOVA were checked except for those already met in order to conduct the ANCOVA (see section 8.1.2). First, no outliers were detected having studentised residuals $> \pm 3$ standard deviation for both general feelings (GF) (H_2) and critical concerns (CC) (H_3). Second, homogeneity of variances was confirmed between groups A and B at both pre- and post-measurements for GF ($p_{pre} = .50,\ p_{post} = .37$) and for CC ($p_{pre} = .91,\ p_{post} = .09$).

Last, Box's test for equality of covariances matrices revealed homogeneity of covariance for both GF (Box's M = 4.21, $p = .25$) and CC (Box's M = 4.09, $p = .27$).

8.2. *Instawareness Views*

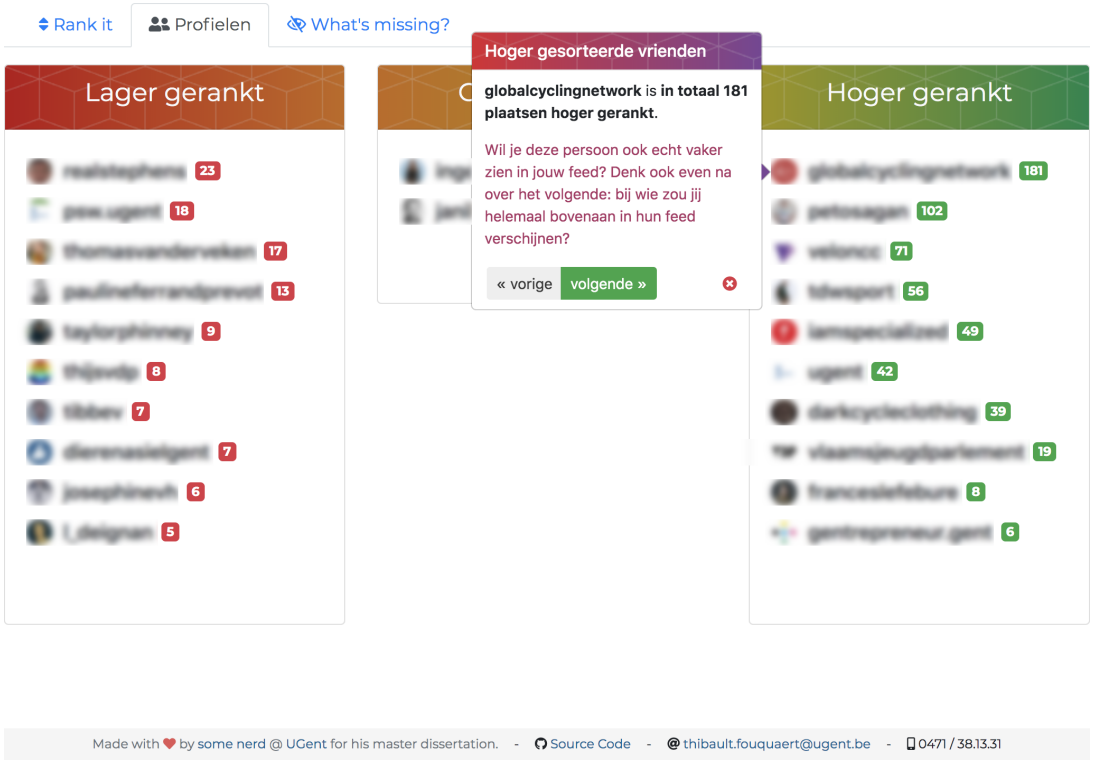

Figure 4. Instawareness view two: followings lower ranked (left), equally ranked (centre) and higher ranked (right) and how many places lower of higher ranked.

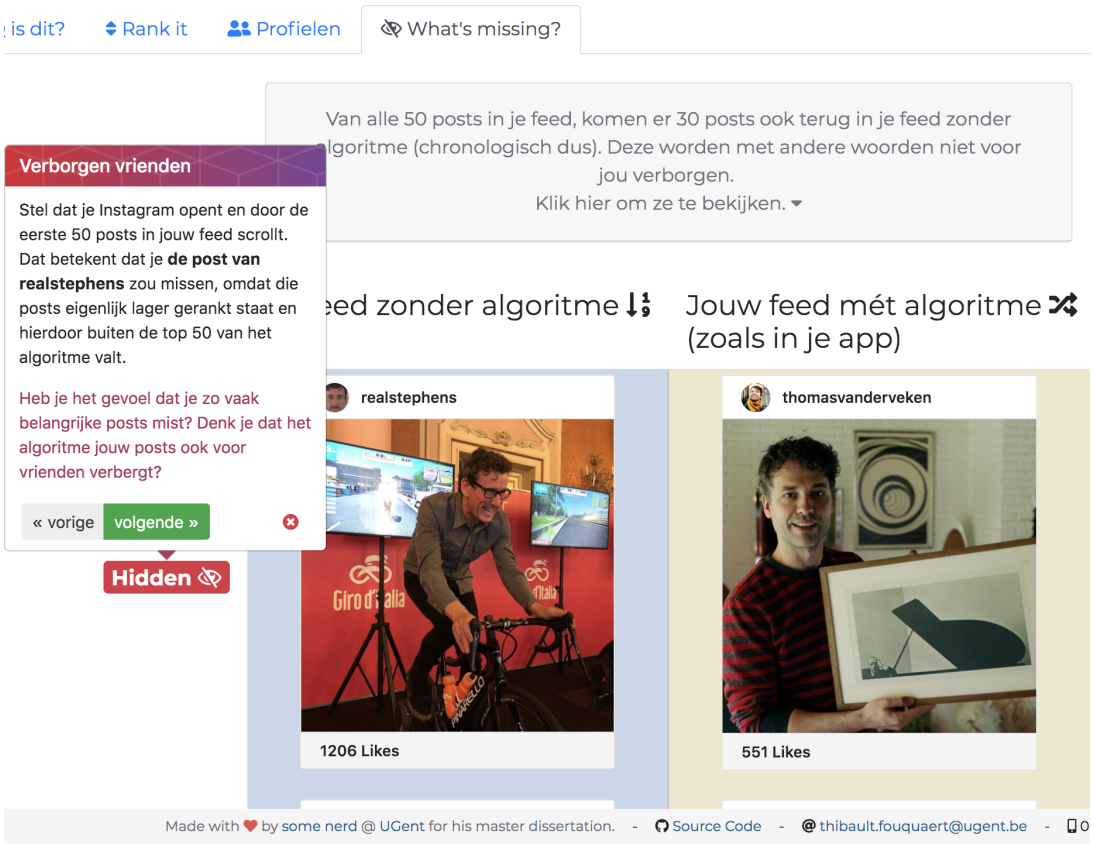

Figure 5. Instawareness view three: which posts are hidden in the current session, that is, not to be seen provided that one only looks at the first 50 posts on their Instagram new feed per session.

8.3. Descriptive statistics and hypothesis visualisation

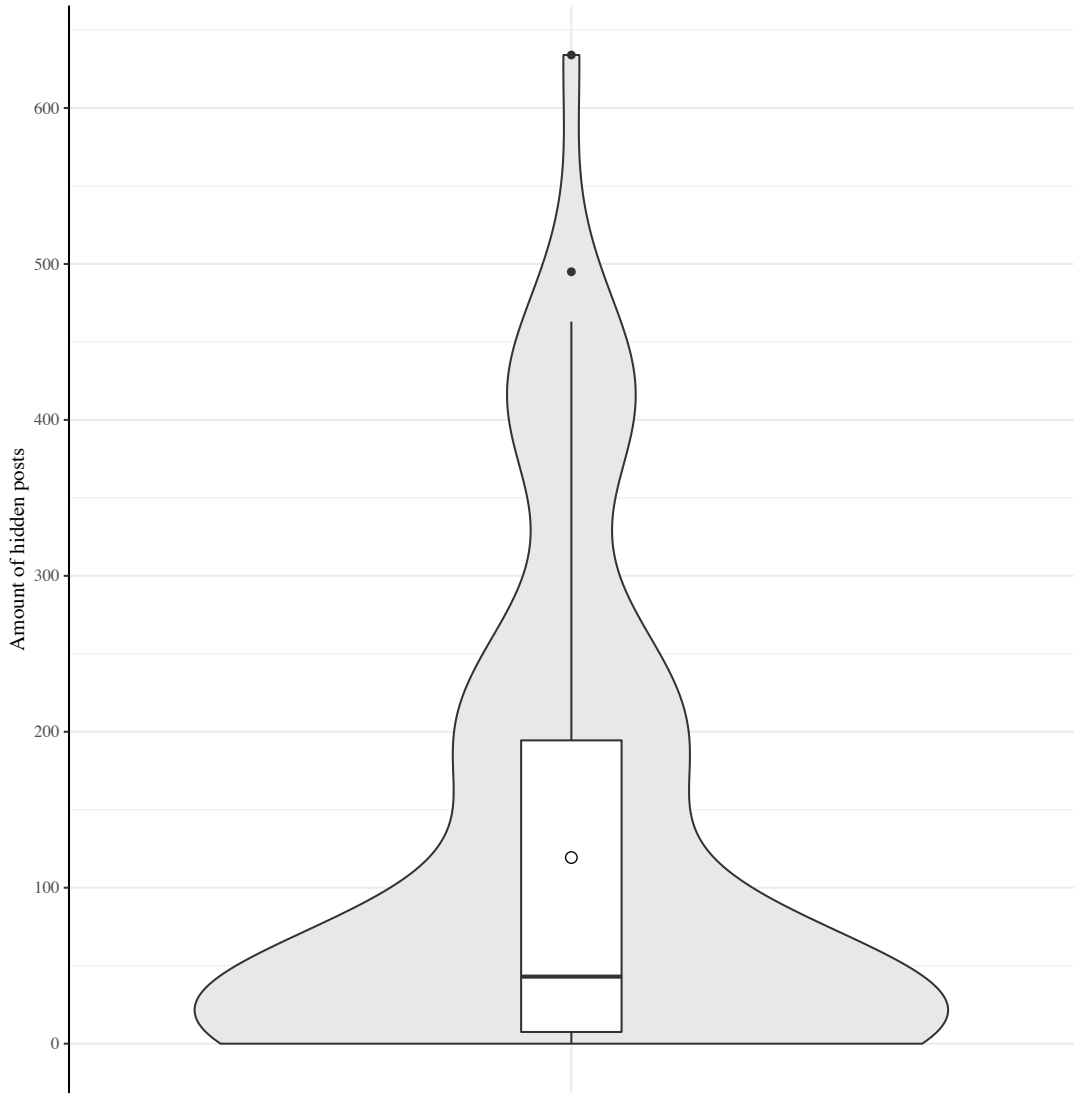

Figure 6. Violin plot of the amount of 'hidden' posts (see operationalisation in section 3.2.3) drawn from all visits to Instawareness from the intervention group (B) ($n = 83$).

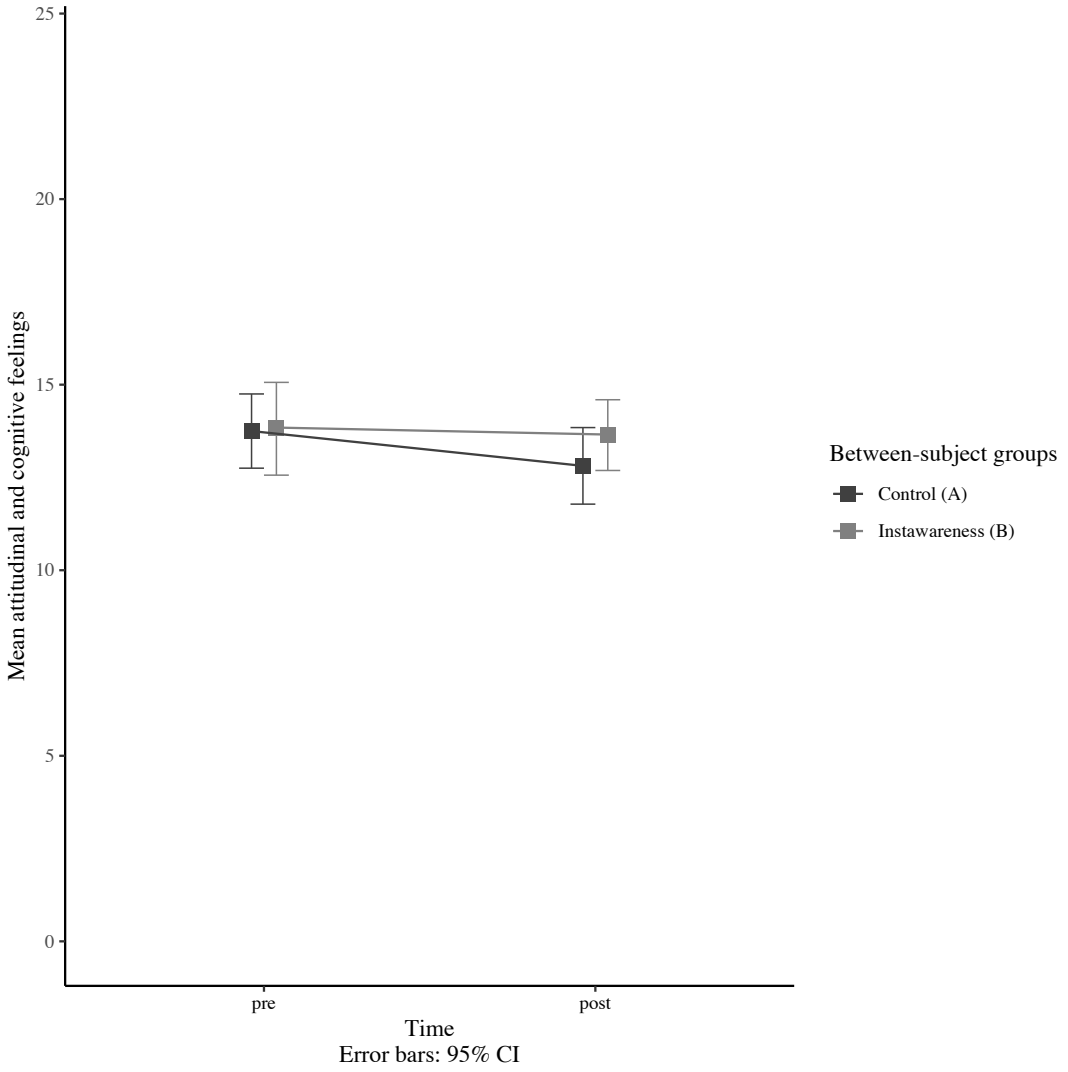

Figure 7. Hypothesis H_2: decreasing evolution of attitudinal and cognitive feelings (e.g., positive/negative, untrustworthy/trustworthy, opaque/transparent) over time (pre- and post-intervention) between control group A and group B that used Instawareness. Scale: $0 - 24$.

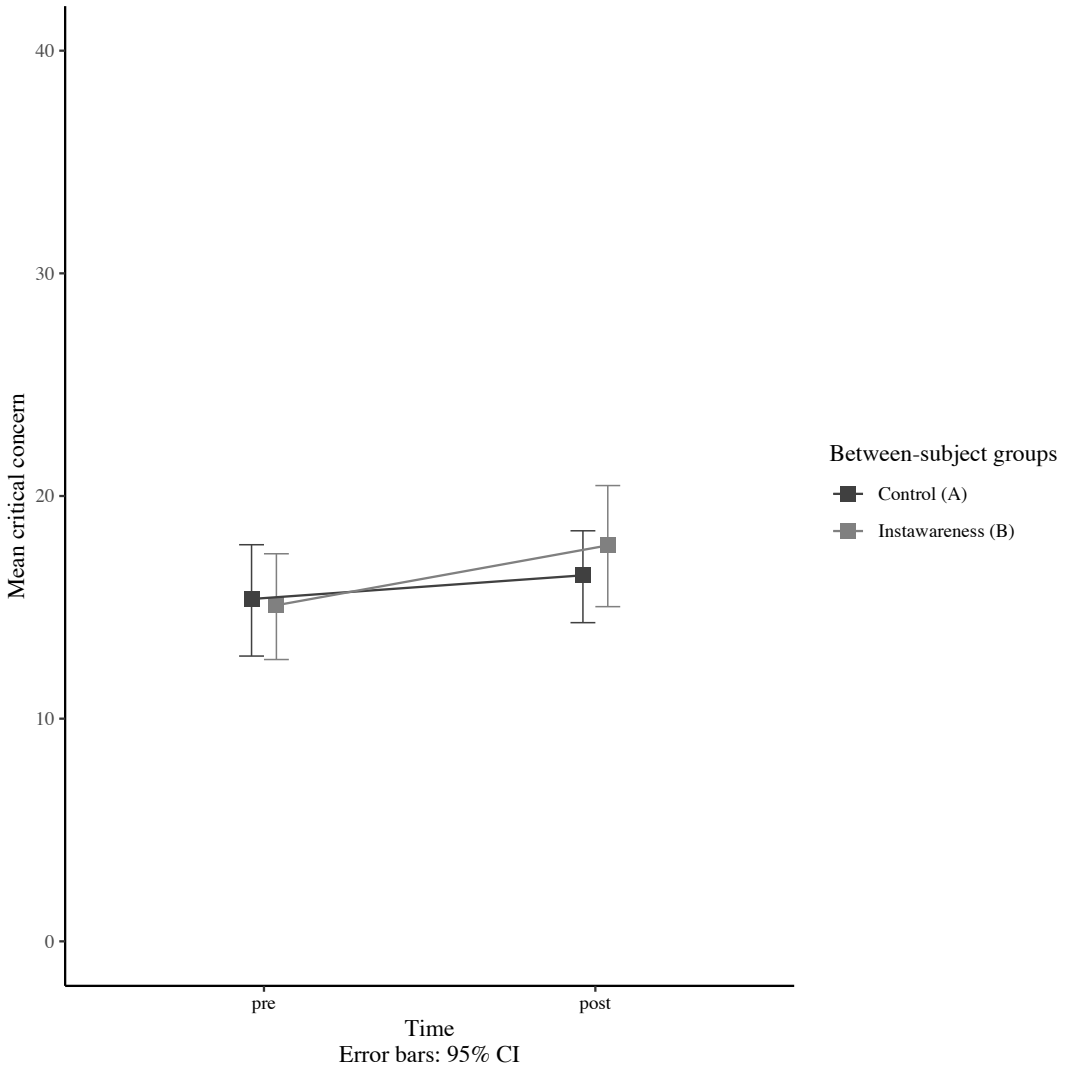

Figure 8. Hypothesis H_3: increasing evolution of critical concern over time (pre- and post-intervention) between control group A and group B that used Instawareness. Scale $0 - 40$.