Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism

Wölker, A.; Powell, T.E.

DOI
10.1177/1464884918757072

Publication date
2021

Document Version
Final published version

Published in
Journalism

Article

Journalism
2021, Vol. 22(1) 86–103

DOI: 10.1177/1464884918757072
journals.sagepub.com/home/jou

Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism

Anja Wölker
University of Amsterdam, The Netherlands

Thomas E Powell
University of Amsterdam, The Netherlands

Abstract

Automated journalism, the autonomous production of journalistic content through computer algorithms, is increasingly prominent in newsrooms. This enables the production of numerous articles, both rapidly and cheaply. Yet, how news readers perceive journalistic automation is pivotal to the industry, as, like any product, it is dependent on audience approval. As audiences cannot verify all events themselves, they need to trust journalists' accounts, which make credibility a vital quality ascription to journalism. In turn, credibility judgments might influence audiences' selection of automated content for their media diet. Research in this area is scarce, with existing studies focusing on national samples and with no previous research on 'combined' journalism – a relatively novel development where automated content is supplemented by human journalists. We use an experiment to investigate how European news readers ($N = 300$) perceive different forms of automated journalism in regard to message and source credibility, and how this affects their selection behavior. Findings show that, in large part, credibility perceptions of human, automated, and combined content and source(s) may be assumed equal. Only for sports articles was automated content perceived significantly more credible than human messages. Furthermore, credibility does not mediate the likelihood of news readers to either select or avoid articles for news consumption. Findings are, among other things, explained by topic-specific factors

Corresponding author:
Anja Wölker, Department of Communication Science, University of Amsterdam, Nieuwe Achtergracht 166, 1001 NG Amsterdam, The Netherlands.
Email: anjawoel@gmail.com

and suggest that effects of algorithms on journalistic quality are largely indiscernible to European news readers.

Keywords
Algorithms, automated journalism, credibility, mediation, natural language generation, news perception, robot journalism, selectivity

Introduction

Technological advances of the year 2017 might appear to stem right out of a science fiction movie: flying cars are being developed and voice-recognition to detect Alzheimer's and depression already exists (Schulz, 2017). This progress continues unabated but also taps into the field of journalism – computer algorithms are the new employees of various media organizations, autonomously producing journalistic stories. The Associated Press (AP), Forbes, Los Angeles Times, and ProPublica, to name popular examples, already make use of such technology (Graefe, 2016).

Optimists view automated journalism – the application of computer algorithms programmed to generate news articles, also known as robot journalism[1] – as an opportunity. Content can be produced faster, in multiple languages, in greater numbers and possibly with fewer mistakes and bias.[2] This might, for example, enhance news quality and accuracy, potentially countering discussions about 'fake news' (Graefe, 2016; Graefe et al., 2016). Furthermore, individual journalists could concentrate on in-depth or investigative reporting, with routine tasks being covered by algorithms in the meantime. Thus, news media could offer a wide range of stories at minimal costs (Van Dalen, 2012). Pessimists, on the other hand, foresee the elimination of jobs, with human journalists being replaced by their nonhuman counterparts. Furthermore, the automated texts produced can be criticized for their mechanical and insipid style resulting from algorithms being limited to analyzing existing data (Graefe et al., 2016; Latar, 2015). They cannot ask questions, determine causality, form opinions and are, at present, inferior to human writing skills (Graefe, 2016). Importantly, algorithms are also inadequate to fulfill the 'watchdog' function, which assigns journalists the task to oversee the functioning of governments and society (Strömbäck, 2005). Thus, they can never 'become a guardian of democracy and human rights' (Latar, 2015: 79).

The introduction of such software is consequently not only dependent on economic concerns of the newsroom but also bound to news media's key audience: the public. A principal consideration of media companies is whether news readers would want to consume automated content (Kim and Kim, 2016). This is contingent upon the public perceiving automated journalism as credible and individuals' consequent decisions to either select or avoid automated articles for their media diet. As the media's fundamental duty is to foster the public sphere by reporting and reproducing events to the best of their ability, credibility is a necessary quality attribution (Cabedoche, 2015; Harcup, 2014). Its absence fosters an increasing distrust of the press, which leads to the disruption of journalism and can ultimately eliminate its service to the people (Broersma and Peters, 2013).

Due to the novelty of journalistic algorithms, research on the credibility and selection of automated journalism is rather slim. Previous studies have concentrated on assessments of automated journalism made by both media practitioners and their audiences (e.g. Carlson, 2015; Clerwall, 2014; Kim and Kim, 2016). All studies of the public's view focus on quality assessments, the majority of which analyze credibility evaluations. Yet, this work has solely considered automated journalism and quality evaluations in isolation. However, news pieces are also envisioned to be co-authored, such that algorithms produce a preliminary article to which human authors add further information. Thus, the introduction of automation in newsrooms is also likely to comprise stories combining the work of computer and human journalists (Graefe, 2016). Moreover, credibility evaluations of these journalistic formats are likely to trigger certain effects. It is questionable if credibility mediates media behavior and determines news readers' choices in selecting specific journalism forms for consumption. This research provides the first evidence of combined journalism content and potential mediators of its selection.

This study uses an experiment ($N=300$) to investigate whether audiences discern between human and different types of automated journalism. Research participants were exposed to news articles, variously produced by algorithms and human journalists. By sampling a mixed European nationality audience, this study further expands previously solely national research.

The main objectives of this study are to investigate (1) whether there is a difference detectable in news readers' credibility perceptions of automated and human journalism, (2) whether news readers perceive combined automated and human journalism as credible, and (3) how credibility evaluations of automated journalism affect news selection.

Previous research on credibility, selection, and automated journalism

Research on credibility dates back more than 60 years (e.g. Hovland et al., 1953; Hovland and Weiss, 1951). The term can be understood as a trust-related characteristic, which is the perceived believability of a message, source, or medium (Bentele, 1998). It 'results from evaluating multiple dimensions simultaneously' (Tseng and Fogg, 1999: 40), such as bias and accuracy (Flanagin and Metzger, 2000; Meyer, 1988). It is therefore not an inherent, but an ascribed characteristic of a journalistic article (the message), an attributed author (the source), or a newspaper (the medium; Bentele, 1998; Vogel et al., 2015).

Accordingly, message credibility concerns content, source credibility concerns the writer(s), and medium credibility concerns the form through which the message and source is transferred (e.g. television, newspaper). Credible journalism is necessary, as audiences are not able to verify everything that is being accounted for themselves. Instead, citizens need to rely on the news to accurately mediate reality (Harcup, 2015). Thus, credibility studies are at the heart of journalism itself, which need to comprise new journalistic developments as well.

Research on medium credibility (e.g. Kiousis, 2001; Westley and Severin, 1964), source credibility (e.g. Reich, 2011; Sundar, 1998), and message credibility (e.g. Borah,

2014; Hong, 2006) is abundant. This includes studies which analyze the interaction between them. For example, in the absence of a source, as is frequently the case for online news, readers will have to rely on message or medium cues to assess the credibility of a text (Metzger and Flanagin, 2015). However, most often, readers are informed about respective authorships (Graefe et al., 2016), after which the interrelation of message and source dimensions becomes relevant. If a source is evaluated credible, then the message is perceived as credible as well (e.g. Roberts, 2010). While trust in the medium positively influences trust in the source, medium credibility is disregarded in this study (Lucassen and Schraagen, 2012). This is because journalistic algorithms have, until now, only been used within print publications.

The existing studies of automated journalism help guide this study of its perceived credibility. This young field defines automated journalism as 'algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming' (Carlson, 2015: 417). Scholars assessing perceptions of automated journalism have focused on journalists (Carlson, 2015; Kim and Kim, 2016) and how it redefines their own skills (Van Dalen, 2012). Four studies have considered how news audiences perceive automated articles in distinctive national settings: the Netherlands (Van Der Kaa and Krahmer, 2014), Sweden (Clerwall, 2014), Germany (Graefe et al., 2016), and South Korea (Jung et al., 2017). All focus on quality assessments, which the first three studies operationalize as credibility evaluations. Jung et al. (2017) analyzed quality as a broader concept, which includes credibility. For the differing study designs see Supplementary Material A.

Regarding *message credibility*, content produced by algorithms was found to score higher compared to texts written by human journalists (Graefe et al., 2016). While differences in Clerwall's (2014) research were not significant, this could be due to a small sample size comprising 46 respondents. Concerning *source credibility*, all existing research shows algorithm and human authors to be equally credible[3] (Graefe et al., 2016; Van Der Kaa and Krahmer, 2014). Based on these findings, the following can be hypothesized:

H1. News readers perceive message credibility of automated journalism content higher than human content.

H2. News readers perceive source credibility of an algorithm author equal to that of a human author.

Furthermore, researchers argue that future applications of automated journalism will generate content through a 'man–machine marriage' (Graefe, 2016). In these cases, 'algorithms […] provide a first draft, which journalists will then enrich with more in-depth analyses, interviews with key people, and behind-the-scenes reporting' (Graefe, 2016). Thus, 'journalists would have more time available for higher value and labor-intensive tasks' (Graefe et al., 2016: 3), enabling newsrooms to deploy their employees more effectively. For example, automated financial stories published by the AP, can already be updated and expanded by human editors. The accompanying description of the article then indicates that 'elements of the story were automated' (Associated Press,

2016; Automated Insights, n.d.; Graefe, 2016). Existing studies have, however, compared automated and human articles only in isolation. This study expands this research by including a combined condition, where stories are co-authored by an algorithm and a human journalist. As no evidence of the perception of mixed algorithm-human content exists, this study has to refrain from deriving an additional hypothesis. Instead, a research question is introduced:

RQ1. How do news readers perceive combined automated and human journalism with regard to message and source credibility?

Automated journalism and news selection

This research also aims to analyze whether credibility judgments of automated journalistic content mediates the decision of media audiences to select automated journalism articles for their news consumption. This represents another unique exploration of this study, providing the first evidence about whether journalistic automation might affect readership numbers. Literature examining the relation of credibility and selectivity in journalism is limited at large. Whereas credibility often is considered in relation to selective exposure (e.g. Metzger et al., 2015; Wheeless, 1974), which focuses on 'the avoidance of messages likely to evoke […] dissonance', such as those counter to a reader's prior attitudes (Knobloch et al., 2003: 92). Instead, *news selection* within this article is understood as carefully choosing a news article as most suitable and adequate to read (Oxford Dictionaries, n.d.-c). Selecting credible news articles for consumption is based on the premise that information is abundantly available but cannot all be reviewed. Accordingly, media usage patterns of individuals positively relate to content of greater utility (Williams, 2012).

Referring to this, Winter and Krämer (2014) found that users rely on source credibility assessments when selecting journalistic articles to read on online news sites. In their experiment, participants not only selected stories from sources which were evaluated as more credible, 'but also selected more frequently, read for longer, and selected earlier' (p. 451).

Focusing on content, Winter and Krämer (2012) also showed that two-sided blog articles were more frequently chosen than one-sided messages, indicating that message quality – which constitutes credibility assessments – as well as perceived balance, influences selection. Thus, credibility of source and message can be assumed to predict the usage of certain content.

Based on these insights, we hypothesize as follows:

H3a. If the message and attributed source of both automated and combined automated and human journalism is perceived as credible, the likelihood of news readers to select automated journalism articles for consumption increases.

H3b. If the message and attributed source of both automated and combined automated and human journalism is perceived as not credible, the likelihood of news readers to select automated journalism articles for consumption decreases.

Methods

This study used an experimental design administered through an online survey ($N = 300$). Each respondent was randomly assigned to one of the following four conditions, with no statistically significant differences between the groups for age, gender, education, media consumption of news, finance and sports articles, or prior knowledge about automated journalism:

1. Automated texts with assigned computer sources ($N = 85$);
2. Human texts with assigned human sources ($N = 70$);
3. Combined automated-human texts with assigned computer and human sources ($N = 65$);
4. A control condition of automated texts without any assigned source ($N = 80$).

Participants

For participant recruitment, snowball sampling was conducted over the social media platforms Facebook, Twitter, and LinkedIn: an invitation[4] to partake was initially published by the authors, requesting to further share the link, connecting to the experiment, with the participants own network. The sample comprised European news readers and was therefore selected and is characterized by respondents with European nationality and at least occasional news consumption. This selection was made because automated journalism, while increasingly introduced in European media organizations, has not yet reached a comparable national concentration as in the United States and is not applied to the majority of European news publications. Furthermore, as no European laws require algorithmic authorship indications, news readers might have been exposed, but most likely not been aware of reading automated articles. This makes the investigation of the perception of the technology highly significant, as the software is predicted to be prominently applied (e.g. Graefe, 2016; Latar, 2015; Montal and Reich, 2016; Newman, 2017). If European news consumers reject automated journalism, its application by media organizations would be questionable after all.

A total of 456 people answered the questionnaire. The final sample comprised $N = 300$ respondents after the removal of non-European participants and incomplete surveys. The final sample includes 119 male and 180 female participants (one participant answered 'Other'), with an average age of 28 years (standard deviation (SD) = 9.20). Overall, 21 percent of the respondents have an upper secondary education, while 70 percent possess a university degree. The following nationalities are distributed in the sample: Czech, Estonian, Icelandic, Luxembourgian, Romanian, Slovenian, and Turkish (one participant from each country), Belgian, Greek, Hungarian, Portuguese, Swedish, and Ukrainian (two participants from each country), Bulgarian, Finnish, French, and Polish (three participants from each country), Austrian and Swiss (four participants from both countries), Spanish (five participants), Italian (six participants), Danish and British (nine participants from both countries), Dutch (11 participants), German (227 participants), and 'Others' (three participants). Nine respondents indicated the possession of a dual citizenship. While most respondents are German, this sample nevertheless represents a more

diverse one compared to previous studies. It therefore represents a mixed, rather than a national sample.

Procedure

After entering the online survey, respondents were asked about their nationality, media usage patterns, and interest in various kinds of news. Then, exposure to one of the experimental conditions followed. Directly after news exposure, credibility of the message and source(s), as well as the likelihood of selecting respective articles, were measured. Afterwards, a manipulation check question was asked. Finally, demographics and knowledge about automated journalism were recorded. On average, respondents took 9 minutes ($SD = 4.51$) to finish the survey.

Stimuli

Each research participant was presented with one finance and one sports article. The former was an earnings report from the corporation Facebook and the latter was a US professional basketball game recap. Currently, automated journalism is dominantly applied in these data abundant fields, as the availability of structured data is a prerequisite (Graefe, 2016). All computer-generated texts were produced by the natural language generation (NLG) firm Automated Insights. It is one of the two most established NLG organizations operating in the United States (Latar, 2015). The automated finance article was published by the AP, which was enhanced by an AP human author for a separate publication, and used in this study as the combined stimuli. Likewise, the automated sports article was enhanced by the (human) authors of this study with elements of the sports human article. The human articles were taken from the websites of the business magazine Forbes (forbes.com) and the US American National Basketball League (nba.com).

The stimuli material differed on the following characteristics, as they are found in reality: the computer-generated stories include no citations, questions, lack explanations of causal relationships or novel occurrences, and possess poorer writing style. To restate, this is due to the computer's inability to interview people, reliance on existing data, and current underdeveloped sophistication of NLG (Graefe, 2016). Furthermore, the computer stories are comparably shorter.[5] This is because human and combined authorship stories can be substantiated and expanded with narratives computers cannot generate or with information which algorithms cannot access (Automated Insights, n.d.). Thus, computer, human, and combined stories presented to participants varied in style and length, as they would in reality (see Supplementary Material B).

In addition, the authorship of each article was made visible, at the beginning of the texts. To increase external validity, the computer articles were indicated to be produced by Automated Insights, with an additional byline describing the firm, as is being done in actual publications (e.g. Koenig, 2017). Human articles were assigned real names of human journalists; combined stories included both a stated human journalist and Automated Insights. The bylines about the authors were placed at the end of each article. Information about the authorship was thus displayed two times on each text.

Visually, the layout, colors, font, and sizes were kept constant across all conditions. Moreover, author names, source descriptions, keywords, publication dates, and times were identical for finance and sports topics. Facts and numbers, which both human journalists and computers can access, were identical. However, divergent and further information, such as differing statistics, was eliminated. Quotes are not present in the computer and control condition, due to the prescribed inability of algorithms to interview people. This serves to validate that style and length were the only manipulations of stimulus content in this experiment.

An initial pre-test of the stimuli material, conducted with 10 people fulfilling the sample criteria, confirmed the comprehension of authorship. The manipulation check in the actual research resulted in significant differences ($p = .001$) between conditions, for which respondents comprehended the respective author(s) correctly.

Measures

Perceived message and source credibility are operationalized, using Meyer's (1988) and Flanagin and Metzger's (2000) scales, which have been employed in various studies (e.g. Chesney and Su, 2010; Flanagin and Metzger, 2007; West, 1994). Roberts (2010) retested both and found respective scales to be reliable and valid to be used together. Other studies used both scales in combination (e.g. Conlin and Roberts, 2016; Hughes et al., 2014). Message credibility is assessed by five bipolar items: (1) unbelievable or believable, (2) inaccurate or accurate, (3) not trustworthy or trustworthy, (4) biased or not biased, and (5) incomplete or complete (Roberts, 2010). Source credibility is assessed by if the messenger (1) is fair or unfair, (2) is unbiased or biased, (3) tells the whole story or does not tell the whole story, (4) is accurate or inaccurate, and (5) can be trusted or cannot be trusted. Following Clerwall (2014) and Van Der Kaa and Krahmer's (2014) research, these dimensions are operationalized on a 5-point Likert-type scale. Generally, and independent from this study's hypotheses, respondents evaluated both the message ($M = 3.55$, $SD = 0.65$) and source(s) ($M = 3.43$, $SD = 0.64$) in all experimental conditions as credible. For overall credibility scores across topics, see Supplementary Material C.

A principal component factor analysis for message credibility ($KMO = 0.84$; $p = .001$) with Varimax rotation on the five operationalized dimensions resulted in two components with an eigenvalue above 1 (eigenvalue 4.36). All items correlate positively with the components. However, as the reliability of the scale is good (Cronbach's alpha = 0.85), no item was dropped and the study resumed with one component. Corresponding factor analysis for source credibility ($KMO = 0.94$; $p = .001$) with Varimax rotation of its respective five dimensions showed one component with an eigenvalue above 1 (eigenvalue 8.61). Each item correlates positively with the component. Equally, the reliability of the scale is good (Cronbach's alpha = 0.98).

Selectivity ($M = 2.16$, $SD = 0.82$) is operationalized by two complementary questions: asking respondents to rate the likelihood of selecting, or avoiding (reverse scored) the articles they were exposed to for their news consumption. Correspondingly, a 5-point Likert-type scale was also used, ranging from 1 (not at all likely) to 5 (very likely). The questions show a significant, strong positive correlation ($r = 0.58$, $p = .001$).

Analyses

Multiple one-way analyses of variance (ANOVA) are conducted to compare mean ratings of the dependent variables (message and source credibility) among the computer, combined, human and control articles (independent variables). Furthermore, the indirect effect of credibility, mediating selectivity is assessed using Hayes PROCESS, a regression-based path analysis macro for SPSS (Hayes, 2016). For this, the independent variables similarly comprise the article conditions. Message and source credibility are then set as mediator variables, and selectivity forms the dependent variable.

Two repeated measure ANOVA's determined the mean values of message credibility ($F(3, 296) = 4.936$, $p = .002$) and source credibility ($F(3, 296) = 4.715$, $p = .003$) as statistically significant between topics. Thus, this study refrains from merging the data but conducting separated analyses for each article topic, finance and sports.

Results

For the sports articles, there was a significant main effect of the conditions on respondents' message credibility evaluations, $F(3, 296) = 2.80$, $p = .040$, $\eta^2 = 0.028$. As expected, a post hoc test showed message credibility of automated content ($M = 3.83$, $SD = 0.68$) scored significantly higher than human content ($M = 3.55$, $SD = 0.76$; $M_{\text{difference}} = 0.28$, $p = .020$). Also, content of the control group ($M = 3.85$, $SD = 0.78$) scored higher on credibility than the human content ($M_{\text{difference}} = 0.30$, $p = .013$). For the finance articles, no significant main effect was found, thus, credibility perceptions can be assumed equal for all conditions. Hence, the findings partly confirm the first hypothesis (*H1*): News readers perceive message credibility of automated journalism higher than human journalism for sports content. For finance articles, the hypothesis needs to be rejected.

For source credibility, no significant main effects were found for both finance, $F(2, 217) = 0.67$, $p = .513$, and sports articles, $F(2, 217) = 0.66$, $p = .518$. Thus, credibility of all sources can be assumed equal. This supports the expectations of the second hypothesis (*H2*): News readers perceive source credibility of an algorithm author equal to that of a human author for both sports and finance articles.

Considering the combined algorithm and human condition, content is perceived credible for both finance ($M = 3.42$, $SD = 0.82$) and sports articles ($M = 3.66$, $SD = 0.72$). Also, combined sources are perceived credible for the finance ($M = 3.34$, $SD = 0.82$) and sports ($M = 3.53$, $SD = 0.78$) articles. These results answer the first research question (*RQ1*): News readers perceive combined automated and human journalism's message and source as credible. Both combined message and source credibility mean scores fall in-between human and automated credibility means. However, the differences of the combined condition to both human and automated credibility scores are not significant. Therefore, it can be assumed that the co-authorship of algorithm and human journalists have no increasing or decreasing effect on perceived message and source credibility. Mean ratings of message and source credibility for all conditions are summarized in Table 1.

For additional analyses, selectivity was compared as a second dependent variable in addition to credibility. Overall, all articles in each condition received low selectivity ratings, displaying respondents are less likely to select the presented articles for news consumption

Table 1. Credibility perceptions and selectivity evaluations by article topic.

	N	Message credibility	Source credibility	Selectivity
Finance articles	300	3.37 (0.05)	3.30 (0.05)	2.18 (0.06)
Computer condition	85	3.24 (0.09)	3.23 (0.08)	2.02 (0.10)
Combined condition	65	3.42 (0.10)	3.34 (0.10)	2.15 (0.11)
Human condition	70	3.42 (0.10)	3.36 (0.09)	2.39 (0.12)
Control condition	80	3.45 (0.08)	–	2.20 (0.11)
Sports articles	300	3.73 (0.04)	3.56 (0.05)	2.13 (0.06)
Computer condition	85	3.83 (0.07)*	3.64 (0.08)	2.11 (0.13)
Combined condition	65	3.66 (0.09)	3.53 (0.10)	2.00 (0.12)
Human condition	70	3.55 (0.09)*	3.51 (0.09)	2.16 (0.12)
Control condition	80	3.85 (0.09)*	–	2.23 (0.13)

Mean ratings of selectivity, message and source credibility. Credibility variables were measured by five bipolar items on a 5-point Likert-type scale (Message credibility: unbelievable/believable, inaccurate/accurate, not trustworthy/trustworthy, biased/not biased, incomplete/complete; Source credibility: fair/unfair, unbiased/biased, tells the whole story/does not tell the whole story, accurate/inaccurate, can be trusted/cannot be trusted). Selectivity was operationalized by two complementary questions, asking respondents to rate the likelihood of selecting or avoiding the articles they were exposed to for their news consumption. It was measured on a 5-point Likert-type scale, ranging from 1 (not at all likely) to 5 (very likely). Standard errors are reported in parentheses.
*Significant differences between conditions.

(see Table 1). Conducting further one-way ANOVA, no significant main effects were found for both finance, $F(3, 296)=1.94$, $p=.123$, and sports articles, $F(3, 296)=0.54$, $p=.655$.

Mediation analyses of the conditions on selectivity through message and source credibility showed no significant indirect effects. Mediation paths are displayed in Figure 1. All mediation coefficients and corresponding confidence intervals for the indirect effects can be found in Supplementary Material D. Accordingly, credibility evaluations do not mediate the news consumption selection behavior of respondents.

Following, both third hypotheses (*H3a, H3b*) need to be rejected: credibility perceptions of both automated and combined automated and human journalism do not increase or decrease the likelihood of news readers to select automated journalism articles for news consumption.

Supplementary analyses tested whether previous knowledge of automated journalism and media consumption of sports and finance moderate the presumed mediation paths. Thus, knowledge (high vs low) and consumption (high vs low) were introduced as possible moderation variables, of the mediation paths for the conditions on selectivity through message and source credibility. This yielded a single moderation effect of knowledge on source credibility, thus influencing the a-path, in the combined condition and finance topic. A significant indirect effect for the computer condition through message credibility finance moderated by high finance consumption, influencing both the a- and b-path, was found. The results indicated no further significant moderated mediation effects. As this study focused on the not yet explored causal relationship of conditions, credibility, and selectivity, these results will not be further discussed. These extra analyses are included in Supplementary Materials E and F.

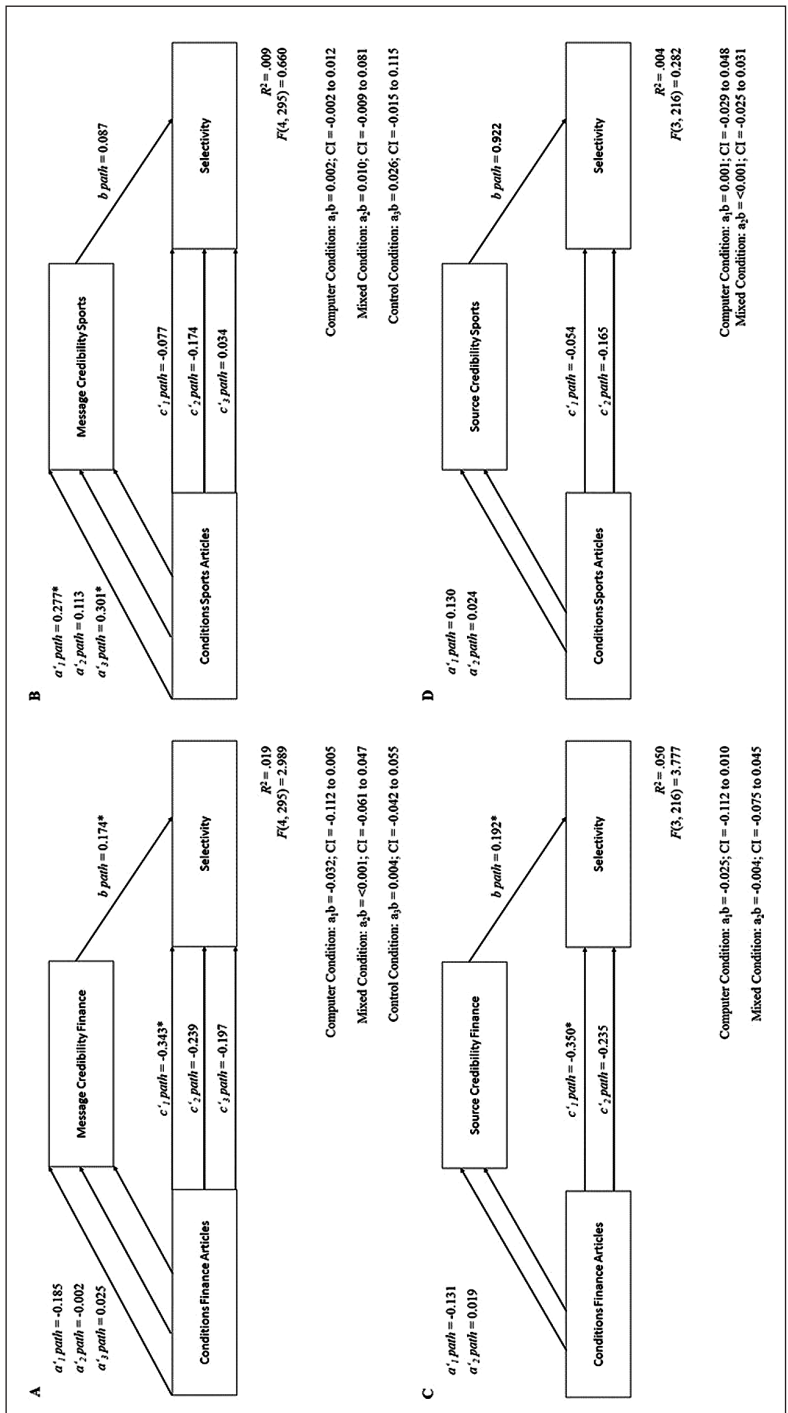

Figure 1. Mediation models for message and source credibility differentiated by article topics.
Unstandardized beta coefficients are shown, *$p < .05$. No indirect effects can be found for (a) finance articles in each condition on selectivity through message credibility, (b) sports articles in each condition on selectivity through message credibility, (c) finance articles in each condition on selectivity through source credibility, and (d) sports articles in each condition on selectivity through source credibility. Non-significance is displayed by 95 percent bias-corrected bootstrap confidence intervals based on 10,000 bootstrap samples.

Discussion

This study investigated European news readers' perception of automated journalism and is the first to study perceptions of novel combined automated and human journalism. Overall, the results show that authors and content of automated and combined journalism are perceived as credible and similarly so to human journalism. This fundamental finding for a mixed nationality sample corroborates previous studies of national audiences, which also found automated journalism articles to be evaluated credible (Clerwall, 2014; Graefe et al., 2016; Van Der Kaa and Krahmer, 2014). Answering the key question of this research, no differences are detectable between automated, combined, and human journalism for source credibility, while small differences are detectable for message credibility (for sports articles).

In line with expectations, all sources were evaluated equally credible. This illustrates a benign view toward the algorithm sources of the automated and combined condition, as computers have only recently been introduced as authors in journalism. One explanation could be what Gillespie (2014) has termed as *the promise of algorithmic objectivity*: 'the technical character of the algorithm is positioned as an assurance of impartiality' (p. 132). News readers might show increased favor toward algorithms, as they appear to guarantee unbiased reporting upon programming. Alternatively, initial low expectations of the algorithm sources might have been surpassed upon consumption, which then may reflect into certain credibility evaluations that are comparable to that of human journalistic content (Graefe, 2016).

Also, consistent with expectations, automated content scored higher on credibility than human content (for sports articles). The algorithmic text included many numbers, which could, compared to written accounts from human journalists, seem more credible (Graefe and Haim, 2016). That said, the observed difference was rather small. In addition, considering finance articles, no differences could be observed – automated and human content were perceived equally credible. The similarity and equality of such credibility ratings may be due to the assessed story topics, sports and finance, which represent routine themes. Human journalists, who are tasked to publish numerous articles at a fast speed, might draft publications by listing facts and figures repeatedly. This could result in the loss of creative writing. Texts produced by algorithms are subsequently comparable, as they follow pre-programmed rules, similarly listing and including the same numbers and details (Graefe, 2016). Hence, 'if automated news succeeds in delivering information that is relevant to the reader, it is not surprising that people rate the content as credible and trustworthy' (Graefe, 2016).

For academia, the findings suggest that future research should broaden their investigations to less 'routine' topics, beyond its current primary application in finance and sports (Graefe, 2016). Synthesizing the differences observed, between sports and finance topics in this study with previous research showing different findings for finance or sports topics, indicates that results can indeed vary by issue (see Graefe et al., 2016; Van Der Kaa and Krahmer, 2014). Thus, this marks an interesting point of departure for additional exploration. For journalists, this study serves to prove that algorithms can compete with their profession for routine reporting, confirming the threat of automation potentially taking away jobs for these areas of coverage (Graefe, 2016; Van Dalen, 2012). The employment of human journalists is comparably costlier, and stories are produced relatively

slower and with a narrower breadth. Whether automation will ultimately transform or deplete the media industry is still uncertain (Carlson, 2015). As Carlson (2015) notes, 'automated journalism harkens to the recurring technological drama between automation and labor present since the earliest days of industrialization' (p. 424).

Moreover, this study's differentiation of message and source credibility assessments allowed control for potential skepticism toward and prior bias against automation. Comparing message credibility mean scores of the automated condition (with an assigned algorithm author) to the control condition (which presented the identical automated text yet with no assigned source) yielded no significant differences. Thus, the visibility of the indicated algorithm source had no effect on the credibility assessment of the automated content alone. The inference from this is that automated journalism can be said to be perceived credible, irrespective of a byline informing about the source. This additional insight can be of interest for the media industry, as no consensus about the full disclosure of authorship between journalism professionals exists (Montal and Reich, 2016; Thurman et al., 2017).

Importantly, this study provided the first empirical exploration of combined human and automated journalism. News readers perceive combined automated and human journalism credible in relation to message and source credibility. Interestingly, no significant differences in credibility assessment of both its message and source, in comparison to the isolated conditions, were found. Theoretically, combined journalism harbors great potential, as it allows human journalists to enhance automated content with original opinions, different from data, which a computer cannot create. This links with positivist scenarios for the news industry, where automation and human labor becomes more integrated, complementing rather than replacing each other (Graefe, 2016; Van Dalen, 2012). Yet, the finding of equal credibility in this study relates back to the 'comparableness' of automated and human articles of routine topics. This is because the presented combined texts, included human-written elements, which were identical to parts used in the human condition. Therefore, forthcoming research may explore how combined journalism will be perceived if human elements *substantially* contrast by providing considerable additional interpretation and reasoning.

Finally, contrary to the expectations, no mediation path of credibility on selectivity was found. This relates to the final aim of this study; credibility evaluations of automated journalism do not affect selectivity. Thus, based on this study, eventual consumption cannot be predicted through the credibility perception of assessed different journalistic articles. Therefore, other variables might be mediating European news readers' decision to select or avoid certain texts for their media diet. The low selectivity scores suggest that other quality variables, such as readability, could have influenced participant's selection ratings (McQuail, 2010). As McQuail (2010) suggests, 'an "information-rich" text packed full of factual information which has a high potential for reducing uncertainty is also likely to be very challenging to a (not very highly motivated reader)' (p. 351). As the presented texts are characterized by a high load of facts, readability could have contributed to low selectivity scores. Research on automated journalism could thus further explore other potential mediation paths and the relationship of additional variables on selectivity.

Certain limitations of this study need to be mentioned. While the sample comprised European news readers, the sports story portrayed a US basketball game. This could have

had a possible decreasing influence on both the dependent variables, credibility and selectivity, because the familiarity with respective sport and described teams cannot be guaranteed. With the expansion of automation in Europe, new research should consider using sports stories that include a more popular European sport such as soccer and choosing coverage of European sports teams. Also, while the editing of the experimental stimuli was kept at a minimum, all changes decreased the external validity of the articles. However, as no modifications concerned the substance, but only the style and length, presented texts came close to reality. Furthermore, the sample is not representative for all European news readers. No proportionate number of respondents per country answered the survey. A majority of German participants predominates the sample. Finally, this study assessed a mixed nationality sample and the questionnaire language was English. Thus, participants answered not necessarily in their mother tongue, which might have affected the results due to potential misunderstanding. Yet, as close to three-fourths of the respondents possess a university degree, comprehension of the respondents can be presumed. In addition, only one participant noted difficulty in understanding the presented texts. Future research should minimize such limitations by including nationally representative samples, who may answer questionnaires in their native language. Furthermore, the study focused only on subsequent selection intentions and can therefore not present long-term effects of content on audiences. In this regard, longitudinal and multiple exposure designs used in competitive framing research (e.g. Lecheler and DeVreese, 2013) can inform further research. Ultimately, this study's findings cannot be generalized to *all* European news readers and can only be compared to the assessment of (US) sports and finance topics.

Normatively, this study's findings show that one principal requirement of computers to become journalists is fulfilled; audiences assign equal credibility to the algorithms and their work. In this respect, they have reached equal status to their human colleagues. Nevertheless, as the results demonstrate, audiences disregard credibility assessments when deciding to select certain articles for consumption and rated the likelihood of selection low for all articles. Thus, when considering the implementation of algorithms in the newsroom, media organizations should carefully deliberate if economic benefits on the production side will ultimately benefit the company in the long run. If readers do not want to select certain articles in the future, it is questionable that high credibility scores are beneficial. This further poses a potentially worrying scenario in light of the increase of fake news, when readers cannot distinguish the credibility of news, which in turn does not have an effect on selectivity. However, given that a difference was detected for the automated sports article, automation could ensure factual and accurate news in the future. Yet, while automated journalism will evolve in sophistication, which could in turn positively affect selectivity, a great weakness persists: computers will never act as the fourth estate. Journalism and democracy are bound through a 'social contract', where journalists take on the role of watchdogs of politicians and society to not misuse their authority (Strömbäck, 2005). Therefore, as journalism is 'under some form of – at least moral – obligation to democracy' (Strömbäck, 2005: 332), automated journalism should best be accompanied by human journalists' contributions. Algorithms alone might never take on journalistic roles, as 'Interpreters' or 'Populist Mobilizers', described by Beam et al. (2009) to 'investigate government claims' (p. 286) or 'set [the] political agenda' (p. 286). This shall be considered in the advancement of automated journalism, which after all is

dependent on how it is handled by the profession (Van Dalen, 2012). Combined journalism poses an ideal adaptation as it is perceived equally credible to automated and human journalism and further merges human journalist's ability of creative writing and interpreting as well as algorithms potential to take over routine tasks. This gives a positive chance for journalism to evolve and for new journalistic role conceptions to emerge, as the industry faces ever-increasing economic pressure and societal scrutiny (Franklin, 2014; Porlezza and Russ-Mohl, 2013).

Taken together, this study has demonstrated both automated and combined journalism as credible alternatives to solely human-created content, with the combined version as a well-rounded ideal for the future of journalism. Automation will not replace but complement human journalism, facilitating the completion of routine tasks and analysis of large data sets.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. While the term 'robot journalism' is dominantly used in media and academia, the terminology is inaccurate. Automated journalism deploys computer algorithms, which are in comparison to robots no machines with 'mechanical power' (Graefe, 2016; Oremus, 2015; Oxford Dictionaries, n.d.-a, n.d.-b).
2. Currently, computer algorithms are programmed by human coders, ultimately introducing a human element to the technology (Linden, 2017; Montal and Reich, 2016). However, as machine learning, which comprises algorithms autonomously developing their own codes, increases in sophistication, so will automation without human influence (Linden, 2017).
3. Van Der Kaa and Krahmer (2014) investigated perceptions of news consumers and journalists. Results of journalists are disregarded because they are professionals within the field and therefore do not compare to this study's sample.
4. The invitation entailed minimum details on the subject of the study: 'It concerns perceptions of journalistic content'.
5. Computer stories summed to an average of 162 words, combined stories summed to an average of 248 words, and human stories summed to an average of 280 words.

Supplementary material

Supplementary material is available for this article online.

References

Associated Press (2016) Home Depot can thank surging house prices for its stunning quarter. Available at: http://www.businessinsider.de/ap-home-depot-tops-street-1q-forecasts-2016-5?internationa l=trueandr=USandIR=T&r=US&IR=T (accessed 10 July 2017).

Automated Insights (n.d.) The future of finance is automated. Available at: https://automatedinsights.com/ap (accessed 10 July 2017).

Bailey C (2014) Butler, Mirotic lead bulls over grizzlies 103–197. Available at: http://www.nba.com/games/20141219/CHIMEM/gameinfo.html (accessed 10 July 2017).

Beam RA, Weaver DH and Brownlee BJ (2009) Changes in professionalism of US journalists in the turbulent twenty-first century. *Journalism & Mass Communication Quarterly* 86(2): 277–298.

Bentele G (1998) Vertrauen/Glaubwürdigkeit. In: Jarren O, Sarcinelli U and Saxer U (eds) *Politische Kommunikation in der Demokratischen Gesellschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 305–311.

Borah P (2014) The hyperlinked world: A look at how the interactions of news frames and hyperlinks influence news credibility and willingness to seek information. *Journal of Computer-mediated Communication* 19(3): 576–590.

Broersma M and Peters C (2013) Introduction. In: Peters C and Broersma MJ (eds) *Rethinking Journalism: Trust and Participation in a Transformed News Landscape*. Abingdon: Routledge, pp. 1–12.

Cabedoche B (2015) New challenges for journalism education: A contribution to UNESCO politics. In: Kramp L, Carpentier N, Hepp A, et al. (eds) *Journalism, Representation and the Public Sphere*. Bremen: Edition Lumière, pp. 57–68.

Carlson M (2015) The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism* 3(3): 416–431.

Chesney T and Su DKS (2010) The impact of anonymity on web log credibility. *International Journal Human-Computer Studies* 68(10): 710–718.

Clerwall C (2014) Enter the robot journalist. *Journalism Practice* 8(5): 519–531.

Conlin L and Roberts C (2016) Presence of online reader comments lowers news site credibility. *Newspaper Research Journal* 37(4): 365–376.

Flanagin AJ and Metzger MJ (2000) Perceptions of internet information credibility. *Journalism and Mass Communication Quarterly* 77(3): 515–540.

Flanagin AJ and Metzger MJ (2007) The role of site features, user attributes and information verification behaviors on the perceived credibility of web-based information. *New Media and Society* 9(2): 319–342.

Franklin B (2014) The future of journalism. *Journalism Practice* 8(5): 469–487.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: The MIT Press, pp. 131–150.

Graefe A (2016) Guide to automated journalism. Available at: http://towcenter.org/research/guide-to-automated-journalism/ (accessed 10 July 2017).

Graefe A and Haim M (2016) Human or computer? Whose stories do readers prefer? Available at: http://en.ejo.ch/digital-news/human-computer-stories-readers-prefer (accessed 10 July 2017).

Graefe A, Haim M, Haarmann B, et al. (2016) Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*. Epub ahead of print 17 April 2016. DOI: 10.1177-/1464884916641269.

Harcup T (2014) Journalism. In: Harcup T (ed.) *A Dictionary of Journalism*. Oxford: Oxford University Press, pp. 148–149.

Harcup T (2015) *Journalism: Principles and Practice*. London: SAGE.

Hayes AF (2016) Supplement to DOI: 10.1111/bmsp.12028. Available at: http://afhayes.com/public/bjmspsupp.pdf (accessed 11 July 2017).

Hong T (2006) The influence of structural and message factors on website credibility. *Journal of the American Society for Information Science and Technology* 57(1): 114–127.

Hovland CI and Weiss W (1951) The influence of source credibility on communication effectiveness. *The Public Opinion Quarterly* 15(4): 635–650.

Hovland CI, Janis IL and Kelley HH (1953) *Communication and Persuasion*. New Haven, CT: Yale University Press.

Hughes MG, Griffith JA, Zeni TA, et al. (2014) Discrediting in a message board forum: The effects of social support and attacks on expertise and trust worthiness. *Journal of Computer-mediated Communication* 19(3): 325–341.

Jung J, Song H, Kim Y, et al. (2017) Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior* 71: 291–298.

Kim D and Kim S (2016) Newspaper companies' determinants in adopting robot journalism. *Technological Forecasting and Social Change* 117: 184–195.

Kiousis S (2001) Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication and Society* 4(4): 381–403.

Knobloch S, Carpentier FD and Zillmann D (2003) Effects of salience dimensions of informational utility on selective exposure to online news. *Journalism and Mass Communication Quarterly* 80(1): 91–108.

Koenig D (2017) FedEx profit rises but misses Wall Street's target. Available at: https://apnews.com/660345848201e490c8723710a2e2cf9b (accessed 1 June 2017).

Latar NL (2015) The robot journalist in the age of social physics: The end of human journalism? In: Einav G (ed.) *The New World of Transitioned Media*. Heidelberg: Springer, pp. 65–80.

Lecheler S and De Vreese CH (2013) What a difference a day makes? The effects of repetitive and competitive news framing overtime. *Communication Research* 40(2): 147–175.

Linden CG (2017) Decades of automation in the newsroom. *Digital Journalism* 5(2): 123–140.

Lucassen T and Schraagen JM (2012) Propensity to trust and the influence of source and medium cues incredibility evaluation. *Journal of Information Science* 38(6): 566–577.

McQuail D (2010) *McQuail's Mass Communication Theory*. London: SAGE.

Metzger MJ and Flanagin AJ (2015) Psychological approaches to credibility assessment online. In: Sundar S (ed.) *The Handbook of the Psychology of Communication Technology*. Chichester: John Wiley and Sons, pp. 445–466.

Metzger MJ, Hartsell EH and Flanagin AJ (2015) Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research*. Epub ahead of print 26 November. DOI: 10.1177/0093650215613136.

Meyer P (1988) Defining and measuring credibility of newspapers: Developing an index. *Journalism Quarterly* 65: 567–574.

Montal T and Reich Z (2016) I, robot: You, journalist. Who is the author? *Digital Journalism* 5: 829–849.

Newman N (2017) Journalism, media and technology predictions 2017. Available at: http://www.digitalnewsreport.org/publications/2017/journalism-media-technology-predictions-2017/ (accessed 11 July 2017).

Oremus W (2015) Why robot? Available at: http://www.slate.com/articles/technology/future_tense/2015/02/automated_insights_ap_earnings_reports_robot_journalists_a_misnomer.html (accessed 27 November 2017).

Oxford Dictionaries (n.d.-a) Machine. Available at: https://en.oxforddictionaries.com/definition/machine (accessed 27 November 2017).

Oxford Dictionaries (n.d.-b) Robot. Available at: https://en.oxforddictionaries.com/definition/robot (accessed 27 November 2017).

Oxford Dictionaries (n.d.-c) Selectivity. Available at: https://en.oxforddictionaries.com/definition/selectivity (accessed 11 July 2017).

Porlezza C and Russ-Mohl S (2013) Getting the facts straight in a digital era: Journalistic accuracy and trustworthiness. In: Peters C and Broersma M (eds) *Rethinking Journalism: Trust and Participation in a Transformed News Landscape*. Abingdon: Routledge, pp. 45–59.

Reich Z (2011) Source credibility and journalism: Between visceral and discretional judgment. *Journalism Practice* 5(1): 51–67.

Roberts C (2010) Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist* 54(1): 43–56.

Schulz T (2017) Zukunft. Zuckerbergs Zweifel. *Der Spiegel* 70(14): 3.

Strömbäck J (2005) In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism Studies* 6(3): 331–345.

Sundar SS (1998) Effect of source attribution on perception of online news stories. *Journalism and Mass Communication Quarterly* 75(1): 55–68.

Thurman N, Dörr K and Kunert J (2017) When reporters get hands-on with robo-writing. *Digital Journalism* 5: 1240–1259.

Tseng S and Fogg BJ (1999) Credibility and computing technology. *Communications of the ACM* 42(5): 39–44.

Van Dalen A (2012) The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice* 6(5–6): 648–658.

Van Der Kaa H and Krahmer E (2014) Journalist versus news consumer: The perceived credibility of machine written news. In: *Proceedings from the computation +journalism conference*, New York. Available at: https://pure.uvt.nl/ws/files/4314960/cj2014_session4_paper2.pdf (accessed 11 July 2017).

Vogel IC, Milde J, Stengel K, et al. (2015) Glaubwürdigkeit und Vertrauen von Online-News. Ein kommunikationswissenschaftlicher Überblick. *Datenschutz und Datensicherheit* 39(5): 312–316.

West MD (1994) Validating a scale for the measurement of credibility: A covariance structure modeling approach. *Journalism Quarterly* 71(1): 159–168.

Westley BH and Severin WJ (1964) Some correlates of media credibility. *Journalism and Mass Communication Quarterly* 41(3): 325–335.

Wheeless LR (1974) The effects of attitude, credibility, and homophily on selective exposure to information. *Speech Monographs* 41(4): 329–338.

Williams A (2012) Trust or bust? Questioning the relationship between media trust and news attention. *Journal of Broadcasting and Electronic Media* 56(1): 116–131.

Winter S and Krämer NC (2012) Selecting science information in web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. *Journal of Computer-Mediated Communication* 18(1): 80–96.

Winter S and Krämer NC (2014) A question of credibility – Effects of source cues and recommendations on information selection on news sites and blogs. *Communications* 39(4): 435–456.

Author biographies

Anja Wölker possesses a joint Master of Arts in Journalism, Media and Globalisation from both Aarhus University, Denmark, and the University of Amsterdam, Netherlands. She previously obtained a Bachelor of Arts in Communication Science from the Westfälische Wilhelms-Universität Münster, Germany.

Thomas E Powell is Assistant Professor of Political Communication at the University of Amsterdam, Netherlands, where he also received his PhD in Political Communication. He obtained his Bachelor of Science in Psychology from the University of Durham, United Kingdom, and his Master of Science in Neuroimaging from Bangor University, United Kingdom.