Volume 13 Number 8 October 2025

DIGITAL JOURNALISM

Digital Journalism

Routledge Taylor & Francis Group

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: www.tandfonline.com/journals/rdij20

The Effects of AI Attribution, Source Priming, and Story Topic Polarization on News Credibility

T. Franklin Waddell

To cite this article: T. Franklin Waddell (03 Sep 2025): The Effects of AI Attribution, Source Priming, and Story Topic Polarization on News Credibility, Digital Journalism, DOI: 10.1080/21670811.2025.2551628

To link to this article: https://doi.org/10.1080/21670811.2025.2551628

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2025.2551628

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

The Effects of AI Attribution, Source Priming, and Story Topic Polarization on News Credibility

T. Franklin Waddell

College of Journalism and Communications, University of Florida, Gainesville, FL, USA

ABSTRACT

The effects of automated authorship on news perceptions have been inconsistent in past research in both directionality and statistical significance. Could null or opposite direction effects found across prior studies be attributed to differences in manipulation recall or variability in news story polarization? What about inconsistencies in the measurement of key outcome variables like message credibility? An online experiment ($N = 820$) tested these possibilities using a 2 (declared source: human vs. AI) × 2 (source priming: present vs absent) × 2 (news topic: data-focused vs polarizing) between-subjects design. The effect of declared source was not statistically significant across most tested variables, even varying across different measurement approaches for the same outcome. The two exceptions out of eleven tested outcomes showed a preference for human authorship, such that news attributed to human authors was perceived as more authentic and trustworthy than news attributed to AI authors. Theoretical and practical implications of these findings are discussed.

KEYWORDS

Automated news; credibility; machine heuristic; MAIN model; journalism studies; topic polarization

The use of algorithms for the writing of data-focused news like sports, weather and finance has become more common in the past 10 years (e.g., Graefe and Bohlken 2020). Academic research on the topic of automated journalism has grown as well, often asking the following question: how do readers perceive news written by an automated author compared to a human author? (e.g., Graefe and Bohlken 2020). A large body of experimental studies have been conducted to answer this question, typically by examining differences between news written by (or attributed to) a human or automated author (e.g., Graefe et al. 2018). Although many experimental studies have been conducted on this topic, the results yielded by these past studies have not been consistent, ranging from positive effects (such that automated authors are preferred relative to human authors; e.g., Clerwall 2014) to either negative effects (such that human authors are preferred; e.g., Waddell 2018) or null effects (such that there is mostly no difference when comparing the two authors; e.g., Kim et al. 2020).

One of the most frequently used theories in past research on automated journalism is the MAIN model/machine heuristic (e.g., Cloudy, Banks, and Bowman 2023; Jang et al. 2023; Molina et al. 2023). As Sundar (2008) explains, the source of a news article

CONTACT T. Franklin Waddell frank.waddell@ufl.edu

(such as the difference between an AI or human author) can activate heuristics, or mental rules of thumb, that influence how readers evaluate the quality of news. In the context of automation, the heuristic theorized to be activated by automated authorship is the machine heuristic: "if a machine chose the story, then it must be objective in its selection and free from ideological bias" (Sundar 2008, p.83). The theoretical guidance of the MAIN model and the machine heuristic would thus predict that automated news should score lower in outcomes like bias or credibility relative to human written news. However, scholars employing a heuristic-driven approach like the MAIN model must remember that heuristics vary in their influence depending on additional factors related to the context of the judgement and/or based on characteristics of the individual (e.g., Khalifa 2022). Applied to the context of automated journalism, although the machine heuristic might predict that automated authors are perceived as objective and free from ideological bias, the machine heuristic may not always be applied to one's news related judgement across every possible news topic. Fortunately, dual process models of persuasion like the elaboration likelihood model (ELM) and heuristic systematic model (HSM) both offer clear predictions when heuristics are most likely to be used (e.g., Khalifa 2022). A consideration of the moderating factors identified by dual process theories might be especially helpful for reconciling past research that has yielded inconsistent results because the machine heuristic is frequently used in past work when theorizing about the prospective effects of automation (e.g., Molina et al. 2023; Waddell 2019). More specifically, the following possibility arises: has there been variability in the directionality of automated effects because known moderators of heuristics have not been fully accounted for when testing the effects of automated journalism in past research? While comparing past research involves considering a variety of factors that likely co-occur across studies (such as variability in study context, operationalization or analysis approach), the guidance of the theories that informed the work are a starting point for accounting for diverging results, especially when a field has regularly relied upon the theory in question (as is the case for many past studies of automated journalism, which have frequently used the MAIN model and machine heuristic; e.g., Cloudy, Banks, and Bowman 2023; Jang et al. 2023; Waddell 2019).

In short, if the machine heuristic is theorized to be one reason why automated news should differ from human written news, then variables identified as moderators by the theoretical framework used to inform the research in question must be considered. With this theoretical logic in mind, the present study focuses on two variables from the dual processing literature (e.g, Isaac and Calder 2025) that are theorized to moderate the effects of AI authorship on news perception. First, this study tests the effect of priming the salience of news source through directions that call attention to the purported authorship of the article. Such an approach is recommended when heuristics are used as an explanatory framework (Bellur and Sundar 2014), especially when approaching past research that has frequently found null results (Graefe and Bohlken 2020) so that a potential replication of null findings can be distinguished from the alternative possibility of type I error. Second, this study also tests the moderating effect of news topic, theorizing that the difference between AI authors and human authors will be greater when the news topic is polarizing rather than data driven. Such comparisons across data driven and polarizing news topics hold

theoretical utility for reconciling conflicting findings from past work because many different news topics have been tested in past research, which have varied in their coverage of polarizing and non-polarizing topics (e.g., Jung et al. 2017; Liu and Wei 2019).

To answer these questions, an online experiment was conducted using a 2 (declared source: human author vs. algorithm) x 2 (source priming: present vs absent) x 2 (news topic: polarizing vs control) between-subjects design.

Literature Review

The effects of automated authorship on news evaluation is a widely studied topic (e.g., Cloudy, Banks, and Bowman 2023; Graefe and Bohlken 2020; Waddell 2019). One of the first experimental studies to explore this topic preceded the proliferation of automated news, but has been highly influential to the subsequent development of the field. Specifically, Sundar and Nass (2001) tested how varying the declared source of a news article affected subsequent evaluation of the article's perceived credibility, quality, liking and representativeness. The study found that news purportedly selected by a computer was perceived as higher quality than news selected by a human editor, although no similar effect was found for credibility, liking or representativeness between the computer and human editor conditions. Guided by this finding and theoretical insight of dual process models of persuasion, Sundar (2008) introduced the machine heuristic, predicting that users may automatically evaluate news curated by automated authors as objective and free of bias. Of course, there is an inherent difference between trust in algorithms and beliefs about automated authorship. Likewise, Sundar and Nass (2001) focused specifically on news selection by a computer, which is distinct from contemporary algorithm-driven news creation. Nonetheless, past literature shows that the machine heuristic has frequently been applied to the context of automated news (e.g., Cloudy, Banks, and Bowman 2023; Waddell 2018), with the expectation derived from it typically being that news written by AI is evaluated more favorably than news written by a human author.

With this background in mind, what has research on the topic of automated journalism generally studied, and what has been the typical pattern of findings? In regards to commonly measured variables, message credibility is a widely studied outcome across many published studies (e.g., Clerwall 2014; Jang et al. 2023; Waddell 2018). News related judgements such as perceived bias (e.g., Waddell 2019), quality (e.g., Melin et al. 2018), and readability (e.g., Kim et al. 2020) have also been studied. The impact of automated authorship on these variables, however, has been inconsistent from study to study. In some cases, the distribution of significant and non-significant results across outcomes trends towards mostly null, such as Clerwall (2014) who found no significant difference among eleven out of twelve indicators of news credibility/quality with the one exception being that human-written news was perceived as more pleasant than AI written news. One meta-analysis with a sample of 12 experimental studies on AI authorship found no difference across studies for scores related to credibility but more consistent effects for readability (Graefe and Bohlken 2020). With that said, some studies find results in favor of automation, such as two studies by Jung et al. (2017), both of which found automated news was perceived as higher

in quality than human written news across both the general public and a sample of journalists. Likewise, Liu and Wei (2019) found that news purportedly written *via* automation was perceived as more objective than news attributed to a human author, although perceived expertise was higher for purportedly human-written content. Finally, some studies find evidence consistently in favor of human authors, such as Melin et al. (2018) who found that human written news was perceived as higher in credibility, liking, quality and representativeness. Likewise, Waddell (2018) found across two studies that news purportedly written *via* automation was perceived as less credible and less newsworthy than news written by a human author.

As the above summary shows, the results of past work regarding the influence on AI authorship on news perception have not converged towards a consistent pattern of results. One commonality across many of these studies (e.g., Cloudy, Banks, and Bowman 2023; Jang et al. 2023; Waddell 2018), however, is that they frequently use dual process models of persuasion like the HSM and ELM as an explanatory framework for predicting the effects of automated authorship (e.g., Isaac & Calder, 2024; Khalifa 2022). This is relevant to possibly reconciling divergent results from past research because these theories offer predictions for when the effects of AI authorship are more or less likely to be observed. In particular, theories like the HSM and ELM would explain that heuristics associated with whether a source is human-like or automated should vary based on the salience and subsequent recall of the authorship manipulation. Despite this important theoretical consideration, some studies testing the effects of automated news do not include (or at least report in-text) a manipulation check to assess whether participants are able to accurately recall the listed author of the article (e.g., Clerwall 2014; Jang et al. 2023; Jung et al. 2017). Without such information, if null effects have been found, it becomes difficult to know if null results are because authorship had no effect, or alternatively, because subjects were merely not aware of the authorship manipulation. This is a critical issue for studying the effects of automation, particularly if a dual processing theory like the MAIN model/machine heuristic (Sundar 2008) is adopted as the explanatory framework, since a heuristic can only impact news related judgements if it is first activated by the corresponding authorship manipulation (e.g., Bellur and Sundar 2014). Without a manipulation check, activation cannot be assumed since no awareness of the cue can be confirmed. For example, Bellur and Sundar (2014) explain that it is necessary to prime (e.g., to activate with a preceding stimulus) a heuristic before stimulus exposure. In many cases, it is difficult to determine based on methodological details shared in past studies if aspects of different study designs may have primed the salience of authorship prior to news exposure. Furthermore, some studies do not disclose that the stimulus was produced *via* automation (e.g., Clerwall 2014; Kim et al. 2020; Melin et al. 2018), thus removing the source cue entirely and therefore decreasing the probability of observing an effect of automation if heuristics are theorized to be the mechanism responsible for the effect.

To address this gap in prior research, the present study tests the impact of source priming by varying the presence of directions that call attention to the declared author of the news article. Such an approach allows this study to ensure that if null effects are observed, the alternative explanation of inadequate source recall can be excluded among participants in the priming condition. Furthermore, it follows the

best practices for studying heuristics recommended in past research (Bellur and Sundar 2014).

H1: There will be a main effect of source priming on source recall, such that recall will be higher in the source priming condition than the no priming control condition

H2: Source priming will moderate the effect of declared source on news perception, such that the difference in news judgements between the automated and human author conditions will be greater in the source priming condition than the no priming control

The Moderating Role of News Topic Polarization

The second variable that the ELM and HSM would theorize should alter the effects of automation is the content/topic of news purportedly being written *via* automation. Many different news topics have been tested when considering the effect of automated authorship, varying from data driven news such as sports (e.g., Clerwall 2014; Jang et al. 2017) and finance news (e.g., Graefe et al. 2018; Jia 2020) to more polarizing topics covering a variety of social issues (e.g., Cloudy, Banks, and Bowman 2023;; Jia 2020). Dual process theories such as the ELM and HSM (and the technologically focused perspectives based upon them, like the MAIN model) theorize that news topic is likely to determine the likelihood of heuristic processing, and by extension, the differential effects of cognitive rules of thumb like the machine heuristic. More generally, news topic also matters because many of the most common outcome variables from the automated journalism literature like credibility and bias (e.g., Graefe and Bohlken 2020; Waddell 2019) might have relatively little variation when the testing context are news topics such as simple sports summaries or weather reporting because such data-driven reporting has little room for challenges to their credibility/veracity and also have less potential for perceived slant (e.g., bias). In sum, measuring variables like bias or credibility might be less likely to find a difference between automated and human authors due to floor effects (a lack of variability) when the primary reporting is based in numerical values and the topic itself is not polarized.

Considering news topic as a moderator is thus potentially a theoretically meaningful endeavor because it could illuminate one possible reason (among admittedly a host of other possibilities) why past studies of automation have been mixed. For example, the literature on automated journalism summarized above shows that a variety of topics have been tested, some of which have been data-driven and benign but others which have been more polarized. This is important because variations like this between neutral and polarizing topics are theorized to impact the differential effects of variables that operate due to heuristics (e.g., Isaac & Calder, 2024; Khalifa 2022), which would include the machine heuristic. Available evidence from past research tentatively supports this possibility in several cases, such that some studies have found an effect of automated authorship when the context of news has been a polarizing topic. For example, Cloudy, Banks, and Bowman (2023) tested the effects of purported automated news writing in the context of abortion news and found that declared automated authorship decreased perceived hostile media bias through perceived source bias (which was assumed to fluctuate because of the machine heuristic). Likewise, another study (Liu and Wei 2019) that tested the effects of automation with three different

polarizing topics (e.g., refugees, LGBTQ, universal healthcare) across polarizing news organizations (e.g., Fox News, MSNBC) found a difference between declared automated and human authors, such that purported automated news was perceived as more credible than human declared news. Waddell (2019) tested the effects of automation with news stories covering politics across Fox News and MSNBC, finding that presumed automated authorship increased perceived credibility relative to declared human authorship through the indirect pathway of perceived news bias. A similar study (Waddell 2019) that focused on fiscal politics found that news was perceived as less biased when written by automation and human authorship in tandem, but only among respondents who identified as strong conservatives. In sum, while not every study follows this pattern, one could still speculate that the aforementioned studies which found effects share the commonality of studying topics that are more polarizing, whereas studies that have failed to find an effect have found these null results because they studied topics that were less polarizing. However, these past studies did not deliberately manipulate news topic to vary news topic polarity, thus necessitating an original study where news context is carefully manipulated with the explicit goal of being either relatively data-driven or polarizing in regards to content.

The present work aims to fill this gap in research. Specifically, this work theorizes that the effects of automation are more likely to be found when considering news topics that are more polarizing. Furthermore, we expect this difference because effects derived from outcome variables that are focused on veracity or objectivity (such as credibility or bias) are more likely to be found when studying topics that have room for response variability due the potential for the topic/content in question to challenged for its authenticity and contain a position that has multiple sides (and thus the potential for someone to feel as though the journalist has taken a side/introduced subjectivity into the writing). To that end, this study tests the effects of automation while varying the degree to which the news topic is polarizing or merely data-focused/benign.

H3a/b: News content and pre-existing issue partisanship will moderate the effect of declared source, such that the difference between the automated and human author conditions will be greater in the polarizing content condition than the data-focused control (H3a), particularly among issue partisans (H3b)

Mediators Responsible for Automation Effects

Although several past studies have cited the machine heuristic as the theorized reason why a difference between automated and human authors is expected (e.g., Cloudy, Banks, and Bowman 2023; Molina et al. 2023; Waddell 2018), most studies do not directly test mediation effects through the machine heuristic as a mediator, instead assuming that a difference between conditions occurs because of the heuristic (e.g., Jang et al. 2023). To empirically test that heuristics are actually the theoretical mechanism, we formally measure agreement with statements that correspond to the theoretical definition of the machine heuristic offered by the MAIN model (Sundar 2008). This is a necessary step for the development of theoretical logic in the field of automated journalism because mediation testing provides empirical evidence that heuristics are a mechanism through which the effects of automation occur. In line with this point, Bellur and Sundar (2014) explain that Authors should formally measure

the heuristic responsible for the heuristic assumed to operate, typically through the formulation of "if, then" statements that capture one's agreement with the heuristic in question, which dual process theories predict should higher when a heuristic has been activated (e.g., Isaac & Calder, 2024). Admittedly, this does not directly measure retrospective recall of the heuristic's use, but dual process theory would likely not expect such retrospective recall of heuristic use nor awareness of the mental rule of thumb given the automaticity of the processes in question (Isaac & Calder, 2024; Khalifa 2022). Put another way, although measurement of agreement with a heuristic is not the same as measuring its actual use, recalling heuristic use is perhaps theoretically unlikely in the case of heuristic processing given the automaticity of the theoretical processes in question that operate below a level of conscious awareness. Nonetheless, measurement of machine heuristic agreement (with the assumption that agreement should be higher when the heuristic has been made active) offers a more direct approach than abstaining from any measurement of the heuristic at all.

Therefore, the present study tests whether automation effects on news-related judgments are mediated by agreement with the machine heuristic:

> **H4:** There will be an indirect effect of declared source on news judgement through the machine heuristic that is moderated by news topic, such that the indirect effect will be statistically significant when the news topic is polarizing but not statistically significant when the news topic is data-focused

Theoretically Distinguishing between Heuristic Agreement and Message Perceptions Theoretically Related to Heuristics

Lastly, as mentioned previously, there has been a tradition in past research for the machine heuristic to be tested by measuring outcome variables assumed to be impacted by variations of the machine heuristic (Bellur and Sundar 2014) without actually measuring the heuristic as a mediator. As noted before, measuring an outcome caused by a heuristic is not the same as measuring the heuristic itself, particularly compared to alternative approaches where a mediation model can be used to specially isolate a pathway through measurement of the heuristic theorized to be associated with the effect in question. For example, scholars have sometimes tested perceptions of source bias or message bias, presuming that both of these outcomes should be lower if the machine heuristic is the theoretical mechanism responsible for the effect. Although the machine heuristic should influence the perceived bias of automated authors (Sundar 2008), measuring either source or message bias is not equivalent with direct measurement of the underlying heuristic, which is the general rule of thumb regarding the presumption that news produced by automation should be free of bias and objective. This explication between heuristic and message judgement is consistent with the MAIN model and its specification of heuristics as mediators, where the theoretical logic of model specifies that heuristics and message judgements are distinct variables that both therefore necessitate separate measurement and testing (Sundar 2008). Notably, we test whether a parallel effect through the machine heuristic is found while also testing for perceived bias through a second parallel pathway; while a serial pathway that operates through the machine heuristic

and perceived bias in sequence is theoretically defensible based on the MAIN model (Sundar 2008), the primary goal of this study is to demonstrate the utility of measuring the machine heuristic directly relative to mediation effects that operate through a proxy of the heuristic's activation (in this case, perceived bias) rather than demonstrating a sequential effect of their combined indirect influence.

> **H5:** There will be an indirect effect of declared source on news judgements *via* perceived bias, such that news written by automated authors will be perceived as less biased, which in turn will be negatively related with perceived credibility

Methods

The full questionnaire, clean data set and the study's pre-registration plan (including measures, pre-registered hypotheses and planned analyses) are publicly available on the open science framework at the following link: https://osf.io/9hfaw/?view_only=9ec8963763244df5ae79f9c1c6fa5d93.

Sample

An a priori power analysis determined that approximately 787 participants would be needed to achieve 80% power assuming $f = .10$ and df = 1. To account for possible exclusions, an additional 10% of cases ($N = 78$) were collected to maintain adequate statistical power if cases were removed. Data were collected in July 2023 with participants from the United States *via* Prolific who were at least 18 years old. Attention checks and time of completion were monitored to ensure data quality. Any subject who failed an attention check or whose time of completion was considered improbably fast based on the average response time for the study was excluded. The final sample exceeded the necessary number of cases for the desired level of statistical power after data exclusions were applied ($N = 820$). In regard to sample demographics, the sample was evenly divided between by sex with 50.9% of participants self-reporting as female (50.9%, $n = 402$) with a mean age of 38.25 ($SD = 13.02$). When asked to self-report their race, 67.8% ($n = 556$) identified as "White/Caucasian," 10.1% identified as "Asian/Asian American," 9.6% identified as "Black/African American," 7.8% identified as "Hispanic/Latino/Latina," 3.9% identified as "Bi-racial" and 0.7% identified as "Other."

Stimuli

Participants were randomly assigned to read a news article that was attributed to either a human author or an algorithm. To heighten external validity, stimulus sampling was employed such that all manipulations were conducted across one of two news topics: either a news story about weather or a news story about the stock market. In addition, each news topic was varied such that the news article either merely reported numerical values like stock price changes or anticipated temperatures or included additional statements alongside the numbers that interpreted the meaning of the data. For example, the weather news condition (a) described recent temperatures, noting trends in weather patterns based on year-to-year data or (b) included the same

content, along with claims that the temperature highs were evidence of climate change. Likewise, the finance news condition (a) described recent financial trends, noting patterns based on recent data or (b) included the same content along with commentary that President Biden's financial policies were responsible for the positive trends in question. No masthead was provided to methodologically control for bias towards pre-existing news organizations. The name of the human author ("Blake Campbell") was selected with the intent of remaining ambiguous regarding presumed author sex or gender to avoid any possible confounds related to authorship and sexism.

Independent Variables

Declared Authorship

Participants were randomly assigned to read a news article purportedly written by either a human author or an automated author. Notably, content was held the same across conditions (aside from varying topic) to avoid confounding the effect of source with the effects of content produced by human or automated means. The byline in the automated condition read, "Automated Insights, algorithm reporter" while the byline in the human author condition read "Blake Campbell, reporter."

Priming

Participants were randomly assigned to view directions before exposure to the news story that explicitly mentioned the author of the story (e.g., Please note that this news story was written by a human journalist). By comparison, participants in the control condition were not given information about the Author of the article before exposure, simply being told they would read a news article without any further priming about the declared author. Participants in the directions condition were required to view the directions for five seconds before proceeding to the stimulus.

News Article Content

In the control condition, participants read a news article that focused exclusively on data and trends that made no claims or assertations based on the underlying data (specifically, either a weather report that described trends in temperature or a finance report that described trends in the stock market; see the pre-registration anonymized OSF link for exact stimuli). By comparison, participants in the polarizing condition read the same news article as featured in the control, except that the article also included interpretations about why the data patterns were occurring, such as asserting that a trend in temperature is evidence of climate change, or that a trend in the stock market is evidence of a successful political policy.

Outcome Variables

Source Recall

A single categorical question asked participants to indicate whether the news article was written by either a human journalist, an algorithm or the option to indicate that they did not recall the source of the story.

Perceived News Article Credibility

Three items were adapted from prior research (Appelman and Sundar 2016) to measure the perceived message credibility of the news article including the extent to which the article was perceived as "accurate," "authentic" and "believable." An index was formed by averaging the items, which was internally consistent ($M = 5.13$, $SD = 1.35$; Cronbach's alpha = .91).

Machine Heuristic Agreement

Two Likert-type items (1 = strongly disagree, 7 = strongly agree) asked participants the extent to which they agreed with the following statements: "if a news story is written by an algorithm, then it must be objective" and "if a news story is written by an algorithm, then it must be free of bias," both of which are consistent with the original explication of the machine heuristic (Sundar 2008). Notably, these items measure agreement with the heuristic (theoretically predicted to fluctuate because of underlying heuristic activation) but does not capture retrospective use of the heuristic. An index was formed by averaging the items, which were significantly and strongly correlated as expected ($M = 2.89$, $SD = 1.68$; $r = .86$).

News Judgement

Although message credibility is a commonly studied outcome variable in past research on automated news (e.g., Graefe et al. 2018), the indicators actually employed for the measurement of credibility vary from the use of single items that measure credibility directly to multiple item scales asking several different questions intended to capture message credibility indirectly (e.g., Appelman and Sundar 2016). This makes comparisons across previous studies difficult since the measures used are often not consistent, even when purportedly measuring the same outcome. There could be variation in effects depending on the measurement approach that is used, so multiple Likert-type items (1 = strongly disagree, 7 = strongly agree) were adapted to measure the extent to which the content of the news article was perceived as biased, believable, authentic, well written, clear, coherent, fair, trustworthy and interesting. A single item that measured perceived credibility directly was also included for the sake of comparison with the three-item credibility index. Descriptive statistics for each indicator are displayed in Table 1.

Table 1. Descriptive statistics for news perception indicators.

Variable	*M* (*SD*)
Accurate	5.08 (1.43)
Authentic	4.85 (1.52)
Believable	5.47 (1.45)
Biased	3.18 (1.79)
Clear	5.63 (1.20)
Coherent	5.68 (1.17)
Credible (single item)	4.92 (1.50)
Fair	5.04 (1.38)
Interesting	4.31 (1.73)
Trustworthy	4.82 (1.49)
Well-written	4.90 (1.42)

Other Measured Variables

Topic Perception and Stimuli Topic Recall

Participants were asked about their prior attitude towards several social issues (including the topics from the study). Specifically, a series of seven-point Likert-type (1 = strongly disagree, 7 = strongly agree) questions were used such as "I am a strong believer in climate change" or "I am a strong supporter of President Biden's economic policies" alongside several questions about other social issues (gun control; same sex marriage) to avoid sensitizing participants. Two indices were formed by combining the two questions about climate change (to subsequently calculate an index about prior attitudes regarding climate change; $M = 5.77$, $SD = 1.67$) and the same was done with two questions related to Biden's economic policy ($M = 4.26$, $SD = 1.72$). Participant were also asked to recall the topic of the news story with the options, "weather," "finance," or "sports." Recall of the news topic was nearly perfect across both the finance (100%, $n = 406$) and weather (99.5%, $n = 408$) conditions.

Results

Source Recall

Two chi-square tests were run to assess the efficacy of the author type manipulation between the priming condition and the no priming control. There were significantly different patterns of recall in both groups, but recall only reached an adequate level of success in the priming condition, such that 97.6% of participants correctly identified the Author as an algorithm in the algorithm condition and 98.5% of participants correctly identified the author as human in the human condition. By comparison, recall accuracy did not exceed 53% without priming in the control groups for either the human or algorithm conditions, primarily due to the increased rate of participants who indicated that they were unable to recall the declared author of the news article. Given these findings, H1 was supported. All subsequent analyses adjust for whether priming was employed, thus allowing comparisons to be made between participants who scored highly in recall relative to those who did not.

Main Analyses

Analyses are first conducted with a commonly employed three question index of message credibility (Appelman and Sundar 2016). An exploratory analysis is presented afterwards which assesses the consistency of these results across eleven additional indicators of news credibility and related judgements.

Credibility

A three-way ANOVA was conducted with declared authorship, priming and news type as the independent variables and the credibility index as the dependent variable. A main effect of author type was found, $F(1, 812) = 5.13$, $p = .02$, partial eta squared = .01, such that news attributed to a human author ($M = 5.24$, $SE = .07$) was perceived as more credible than news attributed to an automated author ($M = 5.04$, $SE = .06$).

The interaction between declared authorship and directions was also significant, $F(1, 802) = 5.67$, $p = .02$, partial eta squared = .01, with post-hoc comparisons using the Sidak correction revealing that the difference between the human and algorithm condition was significant in the priming condition ($p = .001$) but not statistically significant in the no priming control condition ($p = .94$). The interaction between declared authorship and news type was not statistically significant, $F(1, 802).02$, $p = .88$, partial eta squared = .00. Given these findings, H2 was supported.

As for the moderating effect of self-reported partisanship, model 3 of the PROCESS macro was utilized to test the three-way interaction between source type, news topic, and self-reported partisanship. The highest order interaction was not statistically significant, $B = .07$, SE = .09, $p = .43$. Another three-way interaction also tested whether the relationship between source type and priming was possibly moderated by self-reported issue partisanship, but the highest order interaction was again not significant, B = −0.11, SE = .09, $p = .23$. Given these findings, H3a and H3b were not supported.

H4 and H5 both asked about potential mediation effects of declared author type on credibility through agreement with the machine heuristic and message bias, respectively. Model 4 of the PROCESS macro was employed to test these hypotheses with 5,000 bootstrapped samples and 95% bias-adjusted confidence intervals. The indirect effect of author type on credibility through the pathway of the machine heuristic was not statistically significant because the 95% bias adjusted confidence interval for the effect included zero, 95% CI [.02; .01, −0.00]. While the machine heuristic and credibility were positively correlated as expected ($b = .10$, $p < .001$), author type did not significantly affect the machine heuristic ($a = .20$, $p = .09$). Likewise, the indirect effect of author type through the pathway of perceived bias was also not statistically significant, 95% CI [.04; −0.07, .15]. Similar to before, the mediator of bias was significantly correlated with credibility as predicted ($b = -0.43$, $p < .001$), but source type did not significantly affect perceived bias ($a = -0.09$, $p = .47$). The direct effect independent of the machine heuristic and perceived bias was statistically significant, $c' = -0.27$, $p = .00$, such that human authors were perceived as more credible than AI authors (the same pattern that was found by the preceding ANOVA test). Given these findings, H4 and H5 were not supported.

Model 11 of the PROCESS macro was also conducted to determine whether the aforementioned indirect effects were conditioned by either news topic or priming. Results revealed that the highest order interaction was not significant for either of the mediating pathways including through the machine heuristic, B = −0.30, SE = .23, $p = .19$, or perceived bias, B = −0.50, SE = .45, $p = .27$. Given these data, it does not appear that the significance of the indirect effect varied based on the differential presence of priming alongside a manipulation for the content of the news article. Again, H4 and H5 were not supported.

Exploratory Analyses

Although the main analyses were conducted using a commonly employed three question index for measuring message credibility, the present study also assessed several additional indicators of news judgement/perceptions that are commonly

measured in the literature (e.g., Clerwall 2014; Graefe et al. 2018; Jia 2020) for the sake of assessing whether differences in results from study to study could be attributed to varying measurement approaches for assessing credibility.

Specifically, a series of three-way ANOVAS were conducted (the same analysis approach used in the main study) to examine whether the pattern of findings observed in the main analysis (when using the three item index of message credibility) were comparable across other commonly used indicators of credibility. The test and group statistics for each analysis are summarized in Table 2.

Interestingly, only one of the three items used in the message credibility index was statistically significant when analyzed individually: the perception that the message was authentic, such that the news article was perceived as more authentic when attributed to a human author (M=5.12, SE = .07) than an AI author (M=4.59, SE = .07). The effect of declared source on authenticity was also moderated by directions, $F(1, 812) = 11.90$, $p < .001$, partial eta squared = .01, such that the difference between the automated (M=4.40, SE = .10) and human condition (M=5.19, SE = .10) was only significant in the directions condition, $p < .001$. As for the other 8 indicators, only one item was found to be statistically significant, such that the message was perceived as more trustworthy when attributed to a human author (M=4.96, SE = .07) than an automated author (M=4.69, SE = .07). The interaction between source and directions was also significant in the case of trustworthiness, $F(1, 811) = 6.67$, p = .01, partial eta squared = .01, such that the difference between the human (M=5.10, SE = .10) and automated conditions (M (4.57, SE = .10) was only significant when priming preceded the news article compared the control condition where the difference between conditions was not statistically significant. In sum, out of the eleven indicators that were tested, nine of the indicators were not statistically significant, including a single item indicator that simply asked, "is the news article credible?" which was merely at the threshold of statistical significance (p = .05).

Discussion

Theoretical Implications for Dual Process Models and Automated Journalism

This study's results offer a variety of theoretical implications for the study of automated journalism. First, our results find that source recall was significantly higher

Table 2. Main effect of declared source on credibility index and news perception indicators.

	F	*p*	ηp2	AI	Human
Accurate	1.10	.29	.00	5.03 (.07)	5.13 (.07)
Authentic	18.17	< .001	.02	4.64 (.07)	5.07 (.07)
Believable	.65	.42	.00	5.44 (.07)	5.52 (.07)
Bias	.84	.36	.00	3.11 (.08)	3.22 (.08)
Clear	.39	.53	.00	5.65 (.06)	5.60 (.06)
Coherent	.75	.39	.00	5.72 (.06)	5.64 (.06)
Credible	3.73	.05	.01	4.83 (.07)	5.02 (.07)
Credibility (index)	5.13	.02	.01	5.04 (.07)	5.24 (.07)
Fair	2.19	.14	.00	4.98 (.07)	5.12 (.07)
Interesting	.61	.44	.00	4.26 (.09)	4.35 (.09)
Trustworthy	6.99	.01	.01	4.69 (.07)	4.96 (.07)
Well-written	.41	.52	.00	4.93 (.07)	4.87 (.07)

Note: Test statistics for Main Effects of Declared Source are Derived from Three-Way Factorial ANOVAs.

when preceding directions explicitly informed participants that they were about to read a news article written by automation. By comparison, source recall failed in the control condition such that only about half of participants were able to accurately recall the news article's declared author. This finding holds important theoretical implications for the MAIN model because it shows many people are unable to accurately recall the listed author of an article unless they are primed about authorship before exposure to the stimulus. Therefore, it appears that prominence of the authorship manipulation should be integrated within the MAIN model framework, with the assumption that a cue must be prominent in order to be memorable and thus psychologically influential to downstream judgements specified by the theory. By extension, it shows that experimental studies on automated news authorship must strongly consider bolstering the salience of automation-related information (e.g., Bellur and Sundar 2014). By extension, it is imperative for future research on automated authorship to clearly explain all relevant methodological details that could influence participant attention to source information (like bylines) given that variation in study directions that precede stimulus exposure could determine the likelihood of successful recall.

Second, it should also be noted that the difference between the human and AI conditions was not statistically significant across eight out eleven tested outcomes, even in the priming condition where source recall was high. This finding speaks directly to an open question raised by past literature on automated journalism, namely, whether readers discern between human-written and AI written news (e.g., Clerwall 2014; Graefe et al. 2018; Waddell 2018). The mostly null pattern of results in this study is consistent with the findings of an earlier meta-analysis that found the effects of AI authorship were either small or mainly null, especially for perceived credibility (e.g., Graefe and Bohlken 2020). In our study, the only exceptions to this general trend were results were significant was when analyses were conducted with a three item index of credibility or when evaluating message believability or trustworthiness. In these three instances, human authorship was consistently preferred relative to AI authorship such that declared human authorship was perceived as more credible (when using a three item scale), more believable and more trustworthy than declared automated authorship. This pattern of results is not consistent with the machine heuristic (Sundar 2008), which would predict the opposite pattern of results (a preference for automated authors). To the contrary of the machine heuristic, these results show a completely different pattern: that participants do not believe the messages produced by AI are more credible, trustworthy or believable relative to the same content attributed to human authors. This finding for these three specific outcomes aligns with past research that has previously found a preference for human authors relative to automated authors (e.g., Waddell 2019). More generally, this mostly null pattern could be indicative of shifting attitudes towards automation such that ontological differences between human and AI authors are converging in the last 15 years since the initial explication of the machine heuristic (e.g., Guzman 2020). As Lewis, Guzman, and Schmidt (2019) explain, our theorizing regarding the effects of automation must "be responsive to what is, not what was, even when it runs counter to what we think we know about nature of machines and ourselves (p.422). This might be especially true for the original explication of the machine heuristic (Sundar 2008),

which more than 15 years later is perhaps no longer consistent with the mind perception (Gray, Gray, and Wegner 2007) of news consumers when exposed to content attributed to AI.

Therefore, the takeaway from this research (given that eight of eleven findings were non-significant) is that AI authorship mostly does not affect news judgements. Notably, this pattern of null results held even when participants were primarily successful in their recall of the authorship manipulation (thus fulfilling best practices for priming heuristics, per Bellur and Sundar 2014). Our results thus show that the machine heuristic as previously explicated and theorized does not explain this mostly null pattern of results, nor would it support a pattern of results where humans are preferred relative to AI. Given the divergence of these findings from the predictions of the machine heuristic, it may be time to consider alternative theoretical frameworks, such as those related to the dimensions of mind perception (e.g., Gray, Gray, and Wegner 2007) which shows that automated agents are distinguished from human authors along both indicators of agency and experience. This could be meaningful for the study of automated journalism, as a lack of mind perception to automated agents could potentially impact perceptions of the content produced *via* automation among other social judgements (e.g., Waytz et al. 2010). In the case of authenticity and trustworthiness, it could be that the automated news was perceived as less authentic because it did not contain stylistic indicators normally associated with automated writing (such as the lack of a human voice). Alternatively, a competing interpretation of the current findings is that the significant results we do find for authenticity and trustworthiness are simply type I error given that the other eight tested outcomes (despite their theoretical overlap with the other tested outcomes) were not significant. An isolated set of findings that pertain specifically to authenticity and trustworthiness compared to the non-significant results for related outcomes like believability or bias is not consistent with the theory that initially guided this research (e.g., Sundar 2008).

Third, this study also tested the possibility that effects of automation on credibility across data-focused news relative to polarizing news content. Results revealed that news-related judgements were not statistically different between the human and AI conditions regardless of news articles content. Therefore, it does not appear that our results offer support for the conclusion that the effects of AI authorship are contingent on news topic, at least based on variations in the degree to which the news topic is more or less polarizing. This null finding shows that accounting for variations in news topic polarization is not a key to unlocking the theoretical utility of dual process frameworks when applied to automated journalism. To the contrary, it actually shows a lack of fit for these theories to the automated journalism domain since these theories typically find effects that are moderated by topic-based differences (e.g., Isaac & Calder, 2024; Khalifa 2022). As a result, this again might suggest such theories have less explanatory potential for the area of automated journalism given the divergence of the present findings from the results of foundational studies on dual process frameworks (e.g., Isaac & Calder, 2024; Khalifa 2022). Compared to past research, this is consistent with at least several past studies that have failed to find differences of AI authorship across multiple news topics, such as Jia (2020; study 1) who found null results of automated authorship across four different topics. With that said, some past

studies have found effects that were isolated to some topics but not others, such as Jia and Johnson (2021) that found human-authored news was seen as more credible when writing about gun control but not other social issues. More research is thus clearly needed to understand why news topics sometimes moderate authorship effects, ideally taking a theoretical approach that allows a moderator like news topic to be adequately problematized beyond simply benefitting methodological goals like stimulus sampling.

Finally, the present study tested two possible mediators including agreement with the machine heuristic along with the competing mediator of perceived message bias (which is often used as a proxy for the machine heuristic's operation; e.g., Cloudy, Banks, and Bowman 2023). No indirect effect of automated authorship on news judgement was found *via* either of the tested mediators. Notably, participants generally reported low scores for the machine heuristic across both conditions. Given these low scores for agreement with the machine heuristic across both human and automated conditions, it is worth considering whether the original explication of the machine heuristic still reflects the contemporary mind perception of news readers, which perhaps has changed in the last 16 years since the machine heuristic and broader MAIN model was first introduced. Compared with past research, these results are consistent with at least one prior study that also did not find a mediating effect through the machine heuristic (Waddell 2018). The results also do not show an effect of automated authorship on message bias, which does diverge from a similar yet different study where source bias (rather than message bias) was measured as a proxy for activation of the machine heuristic (Cloudy, Banks, and Bowman 2023). It should also be noted that the present study tested agreement with the machine heuristic, but it could be argued that an even more direct approach would be retrospective recall of heuristic use/application to the news judgement in question. Given that heuristic processing is an automatic process that is typically assumed to operate without an individual's conscious awareness, it may not be expected based on dual process theories (e.g., Isaac & Calder, 2024; Khalifa 2022) that recall of heuristic use would be likely or even possible since an awareness of the heuristic would impede the effort conserving goals that are aided by the heuristic/peripheral routes. Nonetheless, future research might benefit from comparing whether agreement with heuristics is consistent with fluctuations in their differential activation or subsequent use in the judgement formation process regarding news.

Practical Implications

The present study holds clear practical implications for news organizations that are publishing AI-written work. Namely, our results show that nearly half of readers may not even notice that a news story is written by automation unless they are notified beforehand about the declared source of the news article. Controversies have arisen where news organizations have deliberately obscured when news is written by automation. These types of crises can perhaps be avoided without any clear downside caused by transparent attribution since our results show that news readers mostly do not notice that the news has been declared as written by automated even when listed on the byline at the top of the page, at least when the content itself does not

portend any evidence of automated writing relative to what a human author otherwise might produce. Our results also show that news readers who are aware of automated authorship mostly evaluate such news as mostly equal to human-authored content with just a few exceptions, in which cases human-authors are still preferred but the difference/size of effect is not particularly large. Therefore, news organizations can publish automated content with full attribution without much concern that they might undermine many of the perceptual outcomes that typically matter to news organizations such as believability, bias or newsworthiness.

Limitations and Future Research

There are several contextual factors that should be considered when interpreting these findings. First, it is worth noting that these results are relevant to the effect of declared authorship while holding the actual content of the news article constant between conditions. Declared authorship is a common manipulation used in the literature (e.g., Graefe and Bohlken 2020), but there do remain stylistic differences between human and AI generated content that might produce different effects when studying actual AI authorship rather than declared AI authorship. Second, the present study tested the effects of news topic polarization in two ways with news related to either climate change or the economy. Although climate change and the economy are both highly polarizing issues for American news readers (at least at the time the study was conducted, circa 2023), there are still many more topics that could be tested to further expand upon the results of the present study. Finally, the typical limitations of experimental research should also be noted, including the use of a non-probability sample and the relatively artificial nature of the stimulus exposure, particularly for participants in the priming condition.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Appelman, A., and S. S. Sundar. 2016. "Measuring Message Credibility: Construction and Validation of an Exclusive Scale." *Journalism & Mass Communication Quarterly* 93 (1): 59–79. https://doi.org/10.1177/1077699015606057

Bellur, S., and S. S. Sundar. 2014. "How Can we Tell When a Heuristic Has Been Used? Design and Analysis Strategies for Capturing the Operation of Heuristics." *Communication Methods and Measures* 8 (2): 116–137. https://doi.org/10.1080/19312458.2014.903390

Clerwall, C. 2014. "Enter the Robot Journalist: Users' Perceptions of Automated Content." *Journalism Practice* 8 (5): 519–531. https://doi.org/10.1080/17512786.2014.883116

Cloudy, J., J. Banks, and N. D. Bowman. 2023. "The Str (AI) Ght Scoop: Artificial Intelligence Cues Reduce Perceptions of Hostile Media Bias." *Digital Journalism* 11 (9): 1577–1596. https://doi.org/10.1080/21670811.2021.1969974

Graefe, A., and N. Bohlken. 2020. "Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News." *Media and Communication* 8 (3): 50–59. https://doi.org/10.17645/mac.v8i3.3019

Graefe, A., M. Haim, B. Haarmann, and H.-B. Brosius. 2018. "Readers' Perception of Computer-Generated News: Credibility, Expertise, and Readability." *Journalism* 19 (5): 595–610. https://doi.org/10.1177/1464884916641269

Gray, H. M., K. Gray, and D. M. Wegner. 2007. "Dimensions of Mind Perception." *Science (New York, N.Y.)* 315 (5812): 619–619. https://doi.org/10.1126/science.1134475

Guzman, A. L. 2020. "Ontological Boundaries between Humans and Computers and the Implications for Human-Machine Communication." *Human-Machine Communication* 1: 37–54. https://doi.org/10.30658/hmc.1.3

Isaac, M. S., and B. J. Calder. 2025. "Thirty Years of Persuasion Knowledge Research: From Demonstrating Effects to Building Theory to Increasing Applicability." *Consumer Psychology Review* 8 (1): 3–14. https://doi.org/10.1002/arcp.1107

Jia, C. 2020. "Chinese Automated Journalism: A Comparison between Expectations and Perceived Quality." *International Journal of Communication* 14: 1–22.

Jia, C., and T. J. Johnson. 2021. "Source Credibility Matters: Does Automated Journalism Inspire Selective Exposure?" *International Journal of Communication* 15: 1–22.

Jung, J., H. Song, Y. Kim, H. Im, and S. Oh. 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. https://doi.org/10.1016/j.chb.2017.02.022

Khalifa, H. K. H. 2022. "A Conceptual Review on Heuristic Systematic Model in Mass Communication Studies." *International Journal of Media and Mass Communication* 04 (02): 164–175. https://doi.org/10.46988/IJMMC.04.02.2022.007

Kim, J., S. Shin, K. Bae, S. Oh, E. Park, and A. P. del Pobil. 2020. "Can AI be a Content Generator? Effects of Content Generators and Information Delivery Methods on the Psychology of Content Consumers." *Telematics and Informatics* 55: 101452. https://doi.org/10.1016/j.tele.2020.101452

Lewis, S. C., A. L. Guzman, and T. R. Schmidt. 2019. "Automation, Hournalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News." *Digital Journalism* 7 (4): 409–427. https://doi.org/10.1080/21670811.2019.1577147

Liu, B., and L. Wei. 2019. "Machine Authorship in Situ: Effect of News Organization and News Genre on News Credibility." *Digital Journalism*, 7 (5): 635–657. https://doi.org/10.1080/21670811.2018.1510740

Melin, Magnus, Asta Back, Caj Sodergard, Myriam D. Munezero, Leo J. Leppanen, and Hannu Toivonen. 2018. "No Landslide for the Human journalist-An Empirical Study of Computer-Generated Election News in Finland." *IEEE Access*. 6: 43356–43367. https://doi.org/10.1109/ACCESS.2018.2861987

Molina, M. D., J. Wang, S. S. Sundar, T. Le, and C. DiRusso. 2023. "Reading, Commenting and Sharing of Fake News: How Online Bandwagons and Bots Dictate User Engagement." *Communication Research* 50 (6): 667–694. https://doi.org/10.1177/00936502211073398

Sundar, S. S. 2008. "The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility." In *Digital Media, Youth, and Credibility. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, edited by M. J. Metzger & A. J. Flanagin, 73–100. Cambridge, MA: The MIT Press.

Sundar, S. S., and C. Nass. 2001. "Conceptualizing Sources in Online News." *Journal of Communication* 51 (1): 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Waddell, T. F. 2018. "A Robot Wrote This? How Perceived Machine Authorship Affects News Credibility." *Digital Journalism*, 6 (2): 236–255. https://doi.org/10.1080/21670811.2017.1384319

Waddell, T. F. 2019. "Can an Algorithm Reduce the Perceived Bias of News? Testing the Effect of Machine Attribution on News Readers' Evaluations of Bias, Anthropomorphism, and Credibility." *Journalism & Mass Communication Quarterly* 96 (1): 82–100. https://doi.org/10.1177/1077699018815891

Waytz, A., K. Gray, N. Epley, and D. M. Wegner. 2010. "Causes and Consequences of Mind Perception." *Trends in Cognitive Sciences* 14 (8): 383–388. https://doi.org/10.1016/j.tics.2010.05.006